GreeNets 2021

Proceedings of the 8th EAI International Conference on Green Energy and Networking

Dalian, People's Republic of China

6-7 June 2021

EDITORS

Peng Li

Jun Mou

Lei Fan

Yinghong Cao

CORE

EAI

RESEARCH MEETS INNOVATION

Proceedings of the 8th EAI International Conference on Green Energy and Networking

June 6-7, 2021, Dalian, People's Republic of China

GreeNets 2021

General Chairs

Peng Li, Dalian Polytechnic University, China

Fanyi Meng, Dalian Polytechnic University, China

Technical Program Committee Chairs

Jun Mou, Dalian Polytechnic University, China

Zhenyu Na, Dalian Maritime University, China

Preface

We are delighted to present the proceedings of the eighth edition of the 2021 European Alliance for Innovation (EAI) International Conference on Green Energy and Networking (GreeNets). This conference aimed at establish a multidisciplinary scientific meeting to discuss complex societal, nonlinear dynamical systems, technological, and economic problems of Green Communication and Green IoT for researchers, developers and practitioners around the world. All of the topics related to these subjects were addressed during the GreeNets 2021 conference.

The technical program of GreeNets 2021 consisted of 31 full papers in oral presentation sessions at the main conference tracks. The conference tracks were: Track 1 - Green Energy; Track 2 - Green Communication and Networking; Track 3 - Intelligent Lighting Control; Track 4 - Machine Learning; Track 5 - Nonlinear system and circuits and Track 6 – Image Encryption. Aside from the high quality technical paper presentations, the technical program also featured two keynote speeches. The two keynote speeches were Dr. Yingqian Zhang from the School of Information and Science Technology at Xiamen University Tan Kah Kee College of China and Dr. Xin Liu from the school of information and communication engineering at Dalian University of Technology of China.

It was a great pleasure to work with the excellent organizing team of the EAI, which was absolutely essential for the success of the GreeNets 2021 conference. In particular, the peer-review process of technical papers have completed and made a high-quality technical program by the Technical Program Committee. We are also grateful to all the authors who submitted their papers to the GreeNets 2021 conference.

We strongly believe that the GreeNets 2021 conference provided a good forum for all researchers, developers, and practitioners to discuss all science and technology related to Green Energy and Networking. We also expect that the GreeNets 2021 conference will be as successful and stimulating, as indicated by the contributions presented in this volume.

June 2021

Peng Li
Fanyi Meng

Conference Organization

Steering Committee

Imrich Chlamtac	University of Trento, Italy
Peng Li	Dalian Polytechnic University, China

Organizing Committee

General Chair

Peng Li	Dalian Polytechnic University, China
Fanyi Meng	Harbin Institute of Technology, China

TPC Chair and Co-Chair

Jun Mou	Dalian Polytechnic University, China
Zhenyu Na	Dalian Maritime University, China

Sponsorship and Exhibit Chair

Guan Gui	Nanjing University of Posts and Telecommunications, China
Jiyu Jin	Dalian Polytechnic University, China

Local Chair

Zijun Gao	Dalian Polytechnic University, China
Nan Xia	Dalian Polytechnic University, China

Workshops Chair

Jun Mou	Dalian Polytechnic University, China
Shaohua Chen	Dalian Jiaotong University, China

Publicity & Social Media Chair

Xiaolin Jiang	Jinhua Advanced Research Institute, China
Jinpeng Wang	Dalian Polytechnic University, China

Publications Chair

Lei Fan	Dalian Polytechnic University, China
Yinghong Cao	Dalian Polytechnic University, China

Web Chair

Xinzhe Wang	Dalian Polytechnic University, China
Weidang Lu	Zhejiang University of Technology, China

Technical Program Committee

Zhenyu Na	Dalian Maritime University, China
Jun Mou	Dalian Polytechnic University, China
Fanyi Meng	Harbin Institute of Technology, China
Xiaolin Jiang	Jinhua Advanced Research Institute, China
Guan Gui	Nanjing University of Posts and Telecommunications, China
Xueyan Zhang	Dalian University of Technology, China
Zhian Deng	Harbin engineering university, China
Yongliang Sun	Nanjing Tech University, China
Fei Yu	Changsha University of Science & Technology, China
Huihai Wang	Central South University, China
Hongjun Liu	University of Jinan, China
Jian Liu	University of Jinan, China
Li Xiong	HeXi University, China
Shaobo He	Central South University, China
Weidang Lu	Zhejiang University of Technology, China
Lin Ma	Harbin Institute of Technology, China
Jiyu Jin	Dalian Polytechnic University, China
Zijun Gao	Dalian Polytechnic University, China
Shaohua Chen	Dalian Jiaotong University, China
Lei Fan	Dalian Polytechnic University, China

Contents

Image Encryption

Wind power prediction based on meteorological data visualization

Shengchi Liu[1], Shuangyue Xiao[2], Li Liu, Junqiao Liu(✉)
{1830997169@qq.com[1], 1546536983@qq.com[2],dalianliujunqiao@126.com (✉)}

Department of Information Science and Engineering Dalian Polytechnic University
Dalian, P. R. China

Abstract. With the development of clean energy, wind power generation has become one of the most important power generation methods. However, the output power of wind power generation system is characterized by uncertainty, so the effective interval prediction of wind power is an effective method to reduce the uncertainty.In this article, through multi-channel multi-dimensional meteorological data, visual correlation analysis, and in-depth analysis of the main factors affecting wind power, put forward based on the extreme gradient promotion (XGB) improved LGB model to forecast. In addition, in order to improve the model calculating speed and accuracy, using principal component analysis was carried out on the original data dimension reduction analysis and visualization processing, then predicted the results compared with the actual situation, to verify the validity of the established model, it shows that this method can be applied to the era of big data of wind power prediction in the future.

Keywords: decision tree; wind power prediction; meteorological data visualization; correlation analysis

1 Introduction

In order to ensure the safe production, reliable operation and regulation of the wind power generation system [1], the micro-grid department must make reasonable plans of energy storage and output [2], so it is necessary to predict the wind power generation system [3].For the power system with access to wind power generation, due to the randomness of the output power of wind power generation system [4], the prediction of the output power of wind power generation system has been a research hotspot at home and abroad.

The world's most developed countries have stronger ability to predict the wind power [5], Denmark is one of the first countries to use wind power in the enhancement power prediction [6], in the early 1990 s, Denmark began to use the wind power prediction system for wind power prediction, the Danish national grid by Danish meteorological administration, the European centre for the weather forecast and ConWX numerical weather prediction products [7] of the company, and then through the national statistical prediction model for wind power forecast, finally to realize the integration of the forecast[8].The United States takes a leading position in

✉ Corresponding author: Junqiao Liu
Fund Project: Natural Science Foundation of Educational Department of Liaoning Province (Grant: J2020053), and Danlian Science and Technology Innovation Fund (2020JJ26GX029)

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307765

wind power prediction and microgrid research and practice[9], and is one of the countries with the largest number of microgrid demonstration projects in the world. Among them, the microgrid project of the University of Wisconsin-Madison is a strong representative.

In wind power prediction in China started late, the micro grid wind power prediction research is less, the current domestic have based on time series method and the neural network study of wind power [10], on the basis of time series short-term prediction, for a long time prediction effect is poorer, numerical weather forecast data, forecast the time scale of the general is limited to six hours, for micro power dispatching operation is of great significance, the research questions for emergency micro power grid scheduling needs to be improved.

At present, there are two commonly used methods at home and abroad. One is the direct method, that is, the wind power generation environment is not predicted [11], but the historical meteorological data is directly used to predict the power.The other is the physical method, which takes terrain, roughness and other information into consideration and uses physical equations to predict and establish a short-term prediction model.

In this paper, through multi-dimensional historical meteorological data, correlation analysis will be carried out on factors affecting air pressure, temperature, humidity, visibility, , wind direction, wind speed and other factors, so as to obtain further in-depth analysis of the main factors affecting wind power [12].On the basis of the original data, further wind direction, wind speed are the main influencing factors of extraction of the data processing [13], and then respectively the average power and total power of the wind direction visualization, According to the data visualization effect [14], the characteristics of local wind can be observed more intuitively, and then through the direction of the wind power curve, clear to observe the influence of the wind to the wind power, and by comparing the wind direction, wind speed 3 d histogram [15], and the wind rose, intuitive to observe the wind direction, wind speed distribution characteristics. Through visualization, the optimal prediction model can be established more clearly, so as to improve the accuracy of prediction.

Finally, XGB and LGB predictions [16] are selected to achieve better results.XGB model cost function adds regular term to control model complexity.The regular term contains the number of leaf nodes of the tree and the sum of squares of L2 module of score output on each leaf node.From the perspective of biase-variancetradeoff, the regular term reduces the variance of the model, makes the learned model more simple and prevents overfitting, which is also a feature of XGBoost over traditional GBDT.Pruning is available, and leaf node output L2 smoothing is new.It has good effect.

2 Data preprocessing

The data used in this study were from DarkSky's historical data, including cloud cover, dew point temperature, humidity, ozone, atmospheric pressure, temperature, vuindex, intensity, visibility, wind direction, gust, wind speed and power multidimensional characteristics.The data accuracy used in this study can reach the level of hours, and the corresponding model can be established with hours as the time sampling interval.Finally get the full year 2018 data, will obtain the multidimensional feature information of pretreatment, the unified data format, fill in the missing data, data visualization and convenient prediction model is established, in these data visualization using annual data, then by further processing data, the prediction model [17] is set up when the data from January 2018 to November as training set of 0forecasts [18], in December as the data set of predictionv [19].Compare the results of prediction model analysis [20].

2.1 correlation analysis

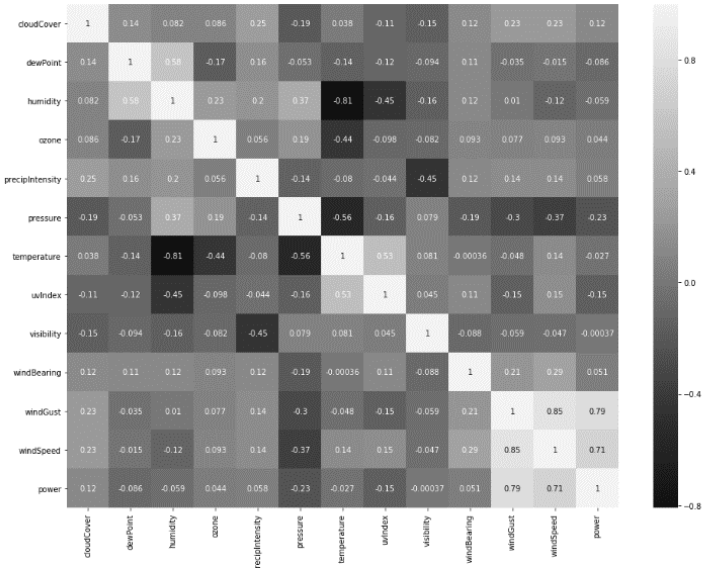


FIG. 1 Multi-dimensional correlation analysis diagram

Based on the obtained data, in python environment, the heatmap function in seaborn library was called to draw the correlation analysis diagram with multidimensional feature information. and the heatmap was obtained as shown in figure 1.

As shown in the multi-dimensional correlation analysis diagram, the power is taken as the dependent variable and the weight of the influencing factors is compared. At this time, the main factors affecting the power are wind direction and wind speed [21].Other factors have very little effect on wind power.The main factors affecting wind power are wind direction, wind speed and gust wind.Further process the visual data, and extract the wind direction and speed data. At the same time, remove other factors affecting wind power, such as air pressure, temperature, humidity, visibility and uvindex , etc. to further visualize the new data.

3 Visual data analysis

Smaller portions of removal influence factors and the new data by the theory of wind direction, wind speed, power, actual power, and wind direction data further processing, converting wind 0-360 degrees 12 wind direction, respectively, using N, NNE, NEE, E, SEE, SSE, S, SSW, SWW, W, NWW and NNW said, according to the different wind data, draw the theory of the average power of the actual power and power comparing histograms, as shown in figure 2.

According to the data of different wind direction, draw the comparison diagram of the actual power and theoretical power of the total power, as shown in FIG. 3.

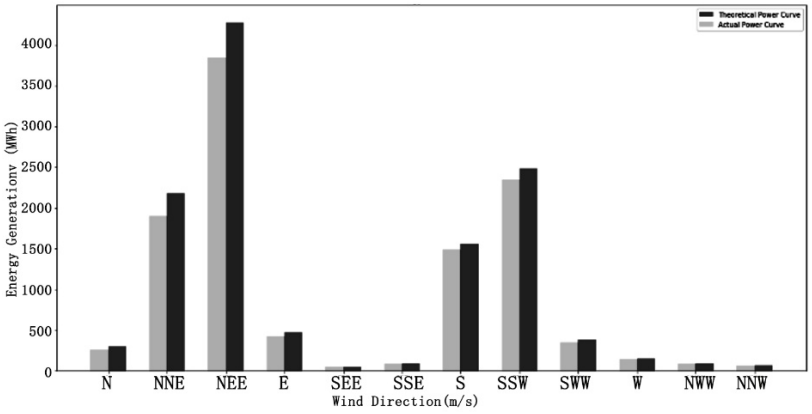


FIG. 2 Bar chart of average power of each wind direction

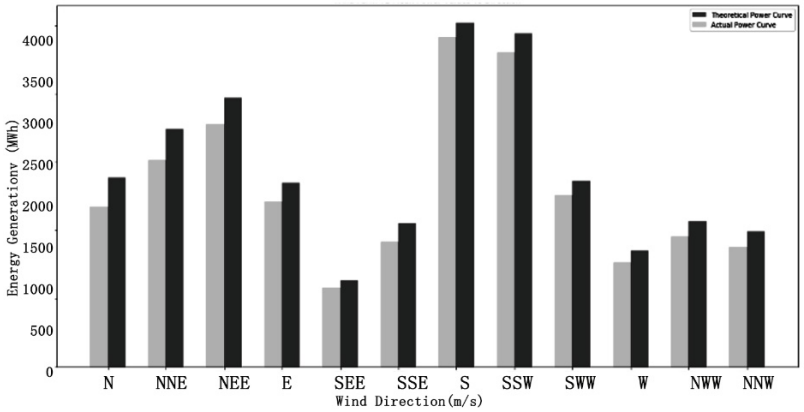


FIG. 3 Histogram of total power at each wind direction

According to Figure 1, the average power distribution of each wind direction can be seen. By comparison, the total power of northeast wind and southwest wind is relatively large, while that of southeast wind and southwest wind is relatively small. Moreover, the theoretical power value of EAST-northeast wind differs greatly from the actual power value, and the characteristics of the average power and total power of each wind direction can be seen by comparing Figure 2 and Figure 3.According to the comparison of total power and average power, it can be seen that there are more south and southwest winds in this region, and the wind speed and generating power are relatively high; the average power of northeast wind is particularly high, and the total power is medium; the northeast wind is relatively less, but the wind is very strong.The characteristics of local wind direction can be preliminarily obtained, and the spatial layout of wind turbines can be optimized.Improve the efficiency of electricity generation [22].

3.1 Draw the relationship curve between wind speed and wind power

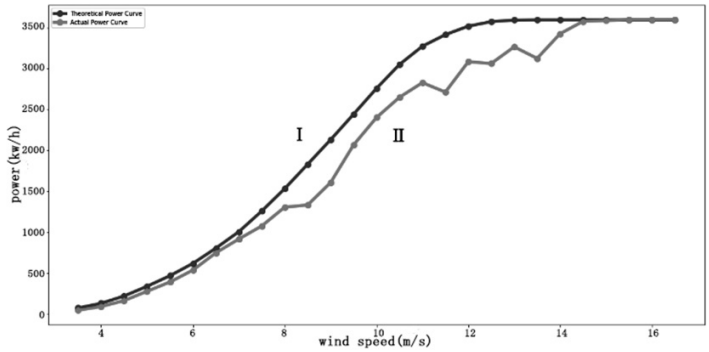


FIG. 4 Wind speed total power curve

In order to further analyze the characteristics of wind data in this region, the actual and theoretical power curves of each wind direction can be drawn to compare the actual and theoretical power curves of different wind directions and observe the power changes under different wind directions. First draw the wind speed and power curve, as shown in the following figure 4. Which | theoretical value, says || said the actual value.

Then, for 12 different wind directions, draw main influence wind power generation of the main direction,such as NNE,NEE,SSW,SWW,draw the curve relation diagram of actual power and theoretical power of different wind directions, as shown in the figure below.

According to the relationship between wind direction and wind power, the influence of different wind directions on wind power output can be seen. Take the wind direction of SSE and SEE for example, when the wind speed is relatively high, the power generation attenuation is relatively large.

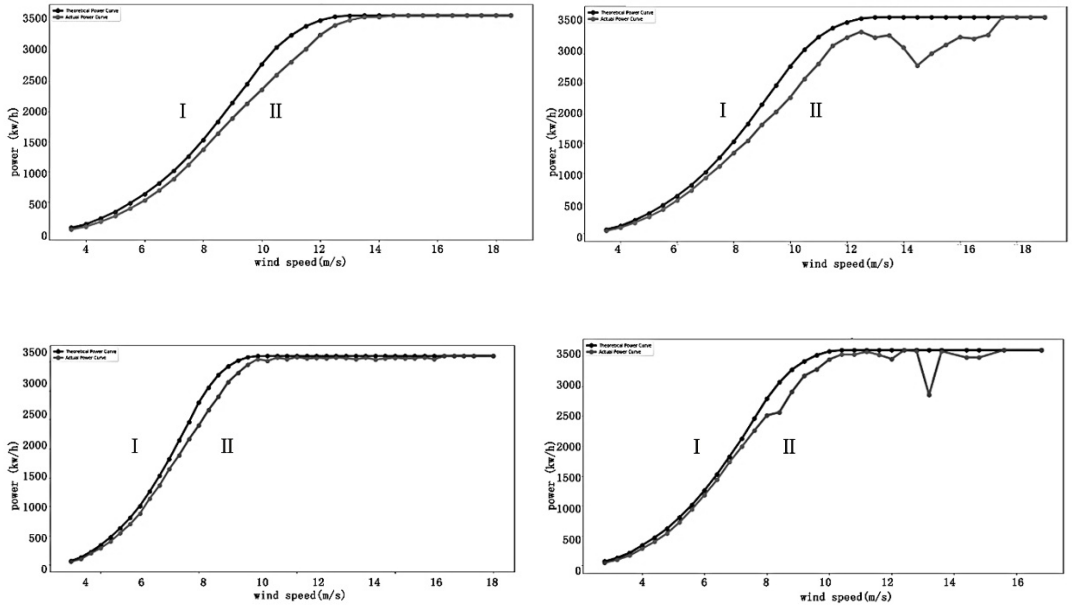


FIG. 5 Different wind direction curves

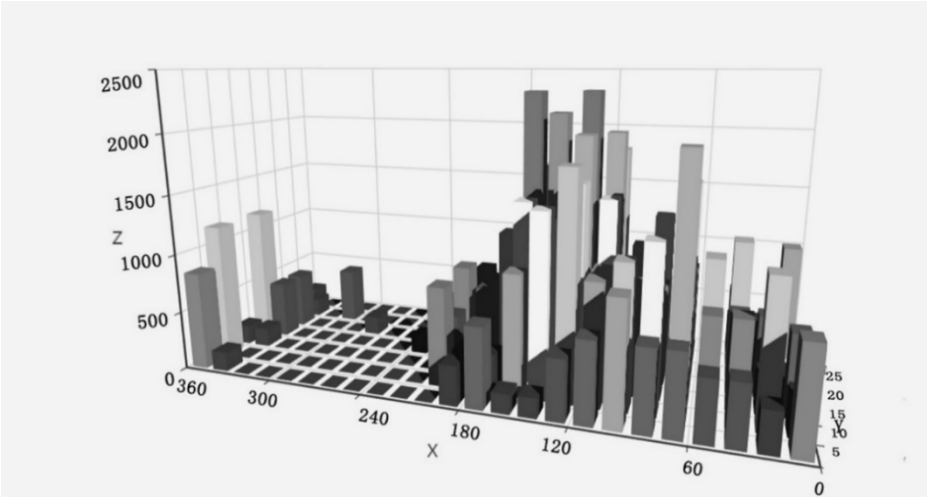


FIG. 6 3D histogram of wind direction and speed distribution

3.2 Visualization of wind direction and speed distribution

By comparing the wind direction of NEE and SSW, the theoretical value of wind power curve is close to the actual value, showing good power generation.And according to the wind direction total power diagram, we can preliminarily judge the rationality of wind turbine installation in this area.

Wind direction and speed have special properties. According to the change of wind direction and speed, the distribution of detected data is converted into a visual 3D rendering using the drawing tool, as shown in Figure 6.

In the 3D figure, the direction of X axis represents the wind direction, the direction of Y axis represents the size of the wind speed, and the Z axis represents The Times of data appearing in the corresponding X and Y axes. Raw data further processing in the visual process, first of all for the sampling, will have a uniform wind direction, wind speed and height of histogram said corresponding to the number of occurrences of wind direction, wind speed, through the intuitive see above the wind distribution characteristics of concentration distribution between 30 degrees - 90 degrees, which are mainly concentrated in NEE, NNE direction, the distribution of wind speed are mainly concentrated in the range of 5 to 15 m/s, according to the characteristics of the wind can be optimized in the region of the wind power equipment installation program.

The data is divided into 16 directions according to the wind direction, as shown in the corresponding direction table 1.

Table 1. Corresponding azimuth chart of wind direction rosette.

Number	1	2	3	4	5	6	7	8
Direction	N	NNE	NE	ENE	E	ESE	SE	SSE
Number	9	10	11	12	13	14	15	16
Direction	S	SSW	SW	SWS	W	WNW	NW	NNW

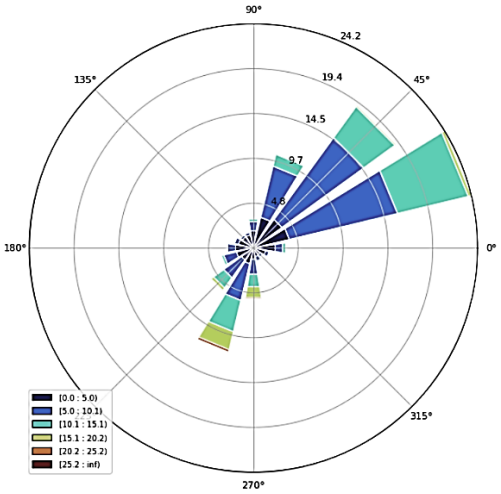


FIG. 7 Wind direction roses map

Then correspond the data to the wind direction rose chart, as shown in **Figure 7**.Draw wind direction roses according to the processed data;Can see clearly the main distribution of wind direction, wind speed, and the percentage of each wind direction and wind speed. As shown in the figure, the concentric circles represent the frequency of different wind directions. The farther away from the center, the greater the frequency. The different colors of the chart indicate the difference of wind speed. The wind speed range is 0 to 25.2m/s. According to the data statistics, the frequency of various wind directions during this period, Calculation formula:

$$g_n = \frac{f_n}{\sum_{n-1}^{16} f_n} . \tag{1}$$

Where $g_n$is the frequency of wind direction in n direction and $f_n$is the frequency of wind direction in n direction in this period.According to the pie chart can clearly see that wind speed are mainly concentrated in 5 to 15 m/s, and the larger wind speed generally appeared between 180 degrees to 270 degrees, namely SSW, SWW direction, combining the characteristics of figure 1, figure 2 power, in order to establish the prediction model in this paper, to provide better modeling scheme, after the visualization of data can be applied to meteorological data statistics, industrial layout and other applications.

4.Building prediction model

Decision tree [23] is a very common classification method.It is based on the known probability of occurrence of various situations,The probability that the expected value of NPV is greater than or equal to zero is obtained by constructing a decision tree.Compared with the XGBoost method, XGBoost uses the method of pre sorting [24], the calculation process is sorted according to the value, one by one data samples to calculate the division of income, this algorithm can accurately find the best division value, but the cost is similar, there is no good generalization.

LGB has stronger adaptability and higher efficiency. LGB does not use the traditional idea of pre sorting, but divides each value into a series of discrete fields, that is, silos. Take floating-point data as an example, the value of an interval will be taken as a cylinder, and then the histogram of these cylinders as the precision unit will be used.

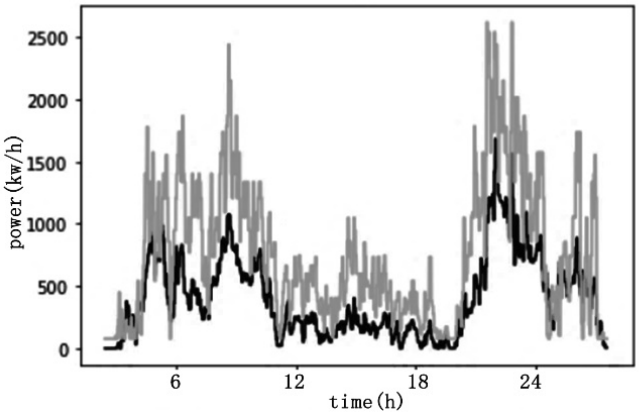


Fig.8. Prediction effect diagram based on LGB model

In this way, the data expression becomes more simplified, the use of memory is reduced, and the histogram brings a certain regularization effect, which can make our model avoid over fitting and has better generalization.When processing feature dimension is high and data volume is large, LGB has faster training speed on the basis of ensuring accuracy and precision.

In this paper, the data from January to November 2018 is used as the training data of the model [25], and the data of December is used as the prediction data for prediction. First, the training data is input into the prediction model for training [26], and then the test data set is used to compare the prediction results, which are shown in **Figure**. 8 below.

As shown in figure 8, compared to forecast power and real power, according to the test data for predicting power value is generally greater than the actual power value, but the trend of wind power prediction and the actual value, which have kept broadly in step with good analysis of the effect, the prediction results can optimize energy storage and wind power system dispatching plan, realize the adjustment of wind power generation system in advance to maximize the use of energy, reduce more economic expenditure.

5.Conclusion

In this paper, multi-dimensional correlation analysis is provided based on the multi-dimensional historical meteorological data. Visual tools are used to sort out the statistical historical data to intuitively display the characteristics of wind power data in this region. LGB model, a learning algorithm based on decision tree, is proposed to predict wind power.First by pretreatment of data, the multidimensional characteristics data correlation analysis, analysis of the main factors influencing the wind power, wind direction, wind speed is put forward the main factors affecting the data further processing, the processed data visualization analysis, respectively, analyze the impact of wind power and the influence of wind speed of power, so as to provide more choice model and optimization model of optimal solution.The main contribution of this paper is to use visualization tools, to establish a prediction model and analyze the characteristics of wind power in the region with the intuitive data support, to comprehensively consider the effect of different influence factors on wind power, the selection of the optimal data of wind power prediction model, finally compare prediction model, the LGB model based on decision tree has better prediction accuracy and faster calculation speed.The advanced prediction of wind power can greatly help wind power generation system to save more

energy and realize the maximization of economic benefits.

Acknowledgment

The authors acknowledge financial support from the Natural Science Foundation of Educational Department of Liaoning Province (Grant: J2020053), and Technology Innovation Fund (Grant: 2020JJ26GX029) and would like to express many thanks to the support of Dalian Key Laboratory of Smart Micro-grid and Green Recycling Industry.

References

[1] González-Sopeña, J, M.: An overview of performance evaluation metrics for short-term statistical wind power forecasting. Renewable and Sustainable Energy Reviews. pp. 245-252 (2020)

[2] Zheng, X.: Deep Neural Network for Short-Term Offshore Wind Power Forecasting. IEEE Transactions on Power Systems. pp. 101-106 (2018)

[3] Electronics.: Combining K-Means and XGBoost Models for Anomaly Detection Using Log Datasets. IEEE Transactions on Industrial Informatics. pp. 221-225 (2020)

[4] Shi, Z.: Direct interval forecast of uncertain wind power based on recurrent neural networks. IEEE Transactions on Sustainable Energy. pp. 1177-1187 (2017)

[5] Fu, X.: Spatiotemporal Attention Networks for Wind Power Forecasting. International Conference on Data Mining Workshops. pp. 149-154 (2019)

[6] Kong, W.: Short-term residential load forecasting based on resident behaviour learning. IEEE Transactions on Power Systems. pp. 1087-1088 (2017)

[7] Zhao, Y.: Research on Photovoltaic Power Generation Output Forecasting Model Based on Wavelet SVM and Micrometeorological Influencing Factor Analysis.Innovation in Artificial Intelligence. pp. 863-868 (2020)

[8] Chen, G.: An ADMM-based distributed algorithm for economic dispatch in islanded microgrids. IEEE Transactions on Industrial Informatics. pp. 3892-3903 (2017)

[9] Meng, W.: Distributed energy management in smart grid with wind power and temporally coupled constraints. IEEE Transactions on Industrial Electronics. pp. 6052-6062 (2017)

[10] Köksoy.: Improved wind power forecasting using combination methods. IEEE 14th International Conference on Machine Learning and Applications. pp. 26-32 (2015)

[11] Peng.: A very short term wind power forecasting approach based on numerical weather prediction and error correction method. IEEE Transactions on Power Systems. pp. 303-308 (2016)

[12] Gupta, S.: Wind ramp event prediction with parallelized gradient boosted regression trees. IEEE. pp. 5296-5301 (2016)

[13] Afrasiabi, M.: Advanced deep learning approach for probabilistic wind speed forecasting. IEEE Transactions on Industrial Informatics. pp. 720-727 (2020)

[14] Khodayar, M.: Rough deep neural architecture for short-term wind speed forecasting. IEEE Transactions on Industrial Informatics. pp. 2770-2779 (2017)

[15] Zhou, Q, Y.: A modern library for 3D data processing. IEEE Transactions on Industrial Informatics. pp. 236-242 (2018)

[16]Jun, M.: Identification of the most influential areas for air pollution control using XGBoost and Grid Importance Rank. IEEE Transactions on Industrial Informatics. pp. 463-469 (2020)

[17] Ozkan, M, B.: Statistical hybrid wind power forecast technique. IEEE Transactions on Industrial Informatics. pp. 375-387 (2015)

[18] Lemenkova, P.: Processing oceanographic data by Python libraries NumPy, SciPy and Pandas. IEEE Transactions on Knowledge and Data Engineering. pp. 121-125 (2019)
[19] Jiang, Z.: Focal-test-based spatial decision tree learning. IEEE Transactions on Knowledge and Data Engineering. pp. 1547-1559 (2014)
[20] Wang, L.: Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. Expert Systems with Applications. pp. 855-863 (2015)
[21] Yang, W.: Load frequency control strategy for wind power grid-connected power systems considering wind power forecast. 2019 IEEE 3rd Conference on Energy Internet and Energy System Integration. pp.763-768 (2019)
[22] Wang, Y.: Deterministic and probabilistic wind power forecasting using a variational Bayesian-based adaptive robust multi-kernel regression model. Applied Energy. pp. 1097-1112 (2017)
[23] Wang, Y.: Optimal wind power uncertainty intervals for electricity market operation. IEEE Transactions on Sustainable Energy. pp. 199-210 (2017)
[24] Alessandrini, S.: A novel application of an analog ensemble for short-term wind power forecasting. Renewable Energy. pp. 768-781 (2015)
[25] Zjavka, L.: Direct wind power forecasting using a polynomial decomposition of the general differential equation. IEEE Transactions on Sustainable Energy. pp. 1529-1539 (2018)
[26] Perry, R.: Multiview Machine Learning in Python. IEEE Transactions on Knowledge and Data Engineering. pp. 356-360 (2019)

Short-term Ultraviolet Index Forecasting Using ARIMA Model

Shuangyue Xiao[1], Shengchi Liu[2],Li Liu(✉)
{ 1546536983@qq.com[1], 1830997169@qq.com[2], link_liuli@hotmail.com ✉}

Department of Information Science and Engineering Dalian Polytechnic University Dalian, P. R. China

Abstract. Solar energy is recognized as an ideal renewable energy. Solar photovoltaic power generation is an important way to use solar energy. Photovoltaic power generation can alleviate the existing energy crisis and alleviate various environmental problems. As a clean and renewable energy, photovoltaic power generation is playing a prominent role in smart micro-grid. The intensity of solar radiation is fluctuating, so the successful grid connection of solar power stations requires accurate power prediction. This paper focuses on the prediction of ultraviolet power index to solve the problem of power prediction of photovoltaic power station. Based on the analysis of cumulative autoregressive moving average model, the stationary of ultraviolet index time series was detected, the order of ultraviolet index model was estimated, and the ARIMA model of ultraviolet index was determined. The prediction accuracy of the model is determined by the root mean square error (RMSE) and mean absolute error (MAE).

Keywords: ARIMA model; ultraviolet index forecasting; the smart micro-grid.

1 Introduction

Smart microgrid is a kind of localized smart grid system [1]. With the rise of distributed generation technology, it is committed to solving the two major problems of distributed generation power quality control and grid reliability control. Microgrid system integrates power generation, energy storage, load and control devices [2], aiming to convert distributed energy into electric energy locally, and then supply local load nearby [3]. Microgrid itself is either independent of large power grid, or interactive with large power grid, or both, which has the characteristics of flexible distribution and complex control [4].

Based on solar radiation, photovoltaic power generation, which realizes the direct conversion of light energy and electric energy, is an important distributed generation technology [5]. In terms of decentralized energy supply, it has obvious advantages over wind power, hydropower and biomass power generation technologies [6]. Photovoltaic power generation module is an indispensable power source category for microgrid system which emphasizes local clean energy utilization [7]. Since the concept of "microgrid" was proposed at the beginning of this century, the microgrid systems developed by various countries or

✉ Corresponding author: Li Liu
Fund Project:Danlian Science and Technology Innovation Fund (2020JJ26GX029)
The Natural Science Foundation of Educational Department of Liaoning Province (Grant: J2020053)

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307719

organizations all contain photovoltaic modules [8]. However, due to the natural intermittent characteristics of photovoltaic power generation, microgrid control is more complex than traditional grid control. As the main force of distributed generation, photovoltaic power generation can accurately predict photovoltaic power, which has important practical significance for the smooth operation of smart microgrid [9].

The common statistical methods include multiple linear regression algorithm, artificial neural network (ANN) algorithm, support vector machine (SVM) algorithm and grey theory algorithm. When the multiple linear regression algorithm is used to predict the photovoltaic power generation system, only the influence of radiation and temperature on the system is considered, and the temperature model is relatively simple, and the influence of wind speed on the temperature of photovoltaic modules is not reflected in the model, so the prediction result error of the model is large; the artificial neural network algorithm is used to predict the photovoltaic system [10], and the error is minimum, and the artificial neural network model can be used the accuracy of test data can be improved continuously, and new samples can be changed at any time. But the artificial neural network algorithm is easy to fall into the local minimum problem, so that the problem can not get the optimal solution, and the generalization ability is not strong, which has a certain impact on the prediction accuracy; the least squares support vector machine algorithm, Vapnik et al [11.12]. proposed the standard support vector machine (SVM) algorithm, which has been successfully applied in many fields, but the standard SVM algorithm has the disadvantages of slow calculation speed and robustness The gray theory algorithm is used to predict the power of photovoltaic power generation system [13], which only considers the data fitting of the output power and re fitting on the residual to reduce the error of power prediction. When the weather is unstable, the prediction error is large.

Generally speaking, multiple linear regression and grey theory are used in photovoltaic output prediction. Although the method is relatively simple, the prediction error is also large, and the prediction ability is weak when the weather changes. The neural network model has strong randomness and accuracy in predicting the output of photovoltaic system, but the process is relatively complicated.

In this paper, a method based on ARIMA is proposed the ultraviolet index prediction of the model, according to the historical ultraviolet index data, samples and analyzes the ultraviolet index according to the time series, and uses the historical data to model and predict the future changes. First, test the stationarity of the historical ultraviolet index data time series. Second, estimate the order of the model on the basis of the stationarity of the ultraviolet index, and get the ARIMA model, it can simplify the complexity of modeling, reduce the complexity of system design procedures, simplify the cumbersome operation, improve the system operation speed and improve the prediction accuracy.

2 ARIMA Modeling Principles and Steps

2.1 Introduction to the ARIMA Model

In this paper, ARIMA model is used in R software for predicting the wind speed [14].

AR in the ARIMA (p, d, q) model is called the autoregressive component; p is the order of the autoregressive component; MA is the moving average component; q is the order of the moving average component; I is Difference; d is the number of differences in the time series that makes the time series stationary [15].

To see that if the Time Series is Stationary or not we take help of ACF (Auto Correlation Function) and PACF (Partial Auto Correlation Function). Time series is a set of interpretations ordered (sequentially) according to the time they were observed.

2.2 Modeling Steps

ARIMA modeling generally has the following 4 steps:

(1) Inspection and processing of data series.

Before modeling, the data series must be tested, analyzed, and preprocessed to meet the requirements of normal and stable zero mean. After processing, the data sequence meets the modeling requirements, and then the next step is analyzed.

(2)Model ordering.

Determine p and q according to the AIC criterion, and calculate the AIC value corresponding to each order. The smaller the AIC value, the better. AIC is known as the Akaike's information criteria (AIC) is outlined as [16]:

$$AIC = -2\ln(L) + 2(p + q + k + 1). \quad (1)$$

where L is the likelihood function, k is the number of parameters.

(3) The estimation of parameters $\varphi_1, \varphi_2, \cdots \varphi_p$ and $\theta_1, \theta_2, \cdots \theta_q$.

This study uses Marquardt nonlinear least squares method.

(4) Model diagnosis.

Check whether the residual sequence is a white noise sequence. If it is, it means that the residual is purely caused by random interference, then the model is available, otherwise it needs to be re-modeled.

3 Experimental Procedure and Result

This paper takes the actual ultraviolet index data of Shanghai, China in the dark sky website of the United States as an example, selects the three-day ultraviolet index data from January 1, 2020 to January 3, 2020, and establishes ARIMA (p, d, q) model. The ultraviolet index in the first 9 hours before January 4, 2020 is predicted, as shown in **Figure 1**.

From the **Figure 1**, we can find that although there are some outliers in the sudden decrease of the value, the time series of the overall ultraviolet index is basically stable, but we still need to check whether it is really stable.

3.1 Data Preprocessing

The establishment of the ARIMA model requires that the time series be a stationary random process. Therefore, before modeling, the basic statistical features such as independence and stationarity of time series data need to be tested to prepare for modeling.

- Data Preprocessing

The establishment of the ARIMA model requires that the time series be a stationary random process. Therefore, before modeling, the basic statistical features such as independence and stationarity of time series data need to be tested to prepare for modeling.

- Independence Test

The overall test method proposed by Box and Pierce is used to test whether the sequence is a white noise sequence. After testing, $p<0.05$, the data meets the independence requirements, so it can be processed with time series.

- Stationarity Test

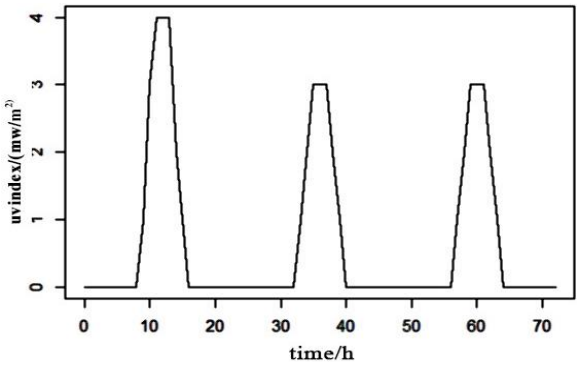


Fig. 1. The actual ultraviolet index in Shanghai from January 1th to 3th

The overall test method proposed by Box and Pierce is used to test whether the sequence is a white noise sequence. After testing, $p<0.05$, the data meets the independence requirements, so it can be processed with time series.

The nonparametric Daniel test is used to test whether the data series have potential trends. The test method is to calculate the Spearman rank correlation coefficient σ of the time series , the formula is as follows:

$$\sigma = 1 - 6\sum_{i=1}^{n} \frac{dt^2}{n(n^2-1)} . \tag{2}$$

The rank of $dt = t - x_i$ is , $t - R(x_i)$ for large samples, statistics $Z = (\rho - \mu_\rho)/\sigma_\rho$ .In the formula $\mu_P = 0$; $\sigma_\rho = 1/\sqrt{n-1}$.

Then determine the significance level α, and calculate the significance level p according to the statistics. If $p<\alpha$, it means that the time series has an upward or downward trend; $p>\alpha$, means that there is no trend in the series, here α=0.05. According to Daniel test p= 0.09825>0.05, the time series is not stable. In order to achieve the requirements of zero equalization and the best fitting effect, the ultraviolet index series needs to be differentiated.

3.2 Model Ordering and Model Determination

First, make a first-order difference to the ultraviolet index data. The difference results are shown in **Figure 2**.

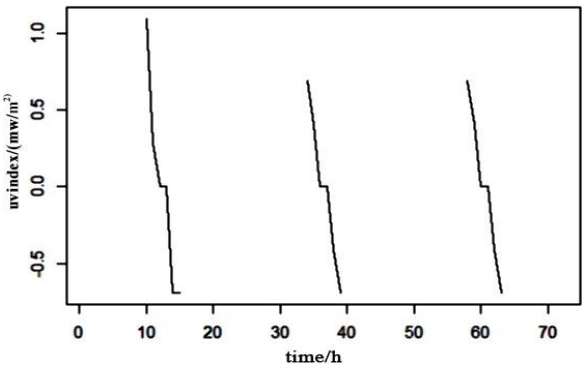


Fig. 2. Ultraviolet index data after first-order difference

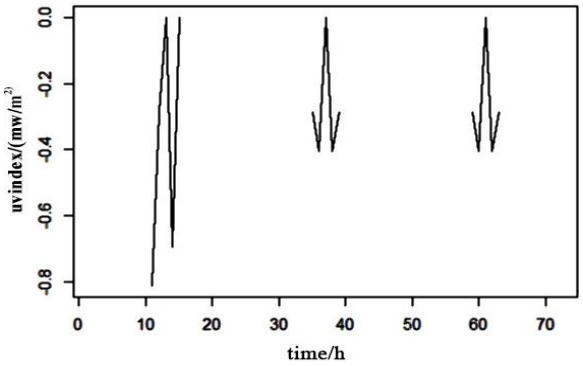


Fig. 3. Ultraviolet index data after second-order difference

Figure 2 looks much more stable than the **Figure 1**. The stationarity test of the ultraviolet index data after the first-order difference is p=0.3628 >0.05, the test result does not meet the standard, the data shall be second-order differenced and the stationarity test shall be performed on the differenced result.

The second-order difference is shown in the **Figure 3**.

The stationarity test of the ultraviolet index data after the second-order difference is p= 0.6504 >0.05, it still does not meet the standard, so we need to carry out the third-order difference.

The third-order difference is shown in the **Figure 4**.

The stationarity test of the ultraviolet index data after the third-order difference is p= 0.01<0.05, which meets the standard, the sequence is stable, and autocorrelation analysis can be performed.

After the time series data is stabilized, first calculate the original series autocorrelation function (ACF), partial autocorrelation function (PACF), and draw the autocorrelation and

partial correlation function graphs. According to the autocorrelation analysis method, make a preliminary series model type and model order preliminary identify, refer to Table 1, and then judge by the AIC criterion, find the p, q corresponding to the minimum value of the model's one-step error, and select the optimal model.

For stationary time series y_t there is

Self-covariance:

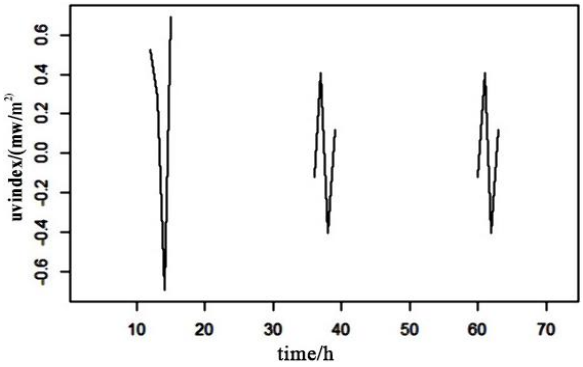


Fig. 4. Ultraviolet index data after third-order difference

Table 1. Stationary time series model structure identification rule

Category	Model recognition		
	AR(p)	**MA**(q)	**ARMA**(p,q)
Autocorrelation function(ACF)	Tailing	Censored	Tailing
Partial autocorrelation function Model recognition(PACF)	Censored	Tailing	Tailing

$$\hat{\gamma}_k = \frac{1}{N}\sum_{j=1}^{N-K} y_k y_{t+k} \ . \tag{3}$$

Autocorrelation function:

$$\hat{\rho} = \frac{\gamma_k}{\gamma_0} \ . \tag{4}$$

The partial correlation function:

$$\begin{cases} \hat{\alpha}_{11} = \hat{\rho}_1 \\ \hat{\alpha}_{k+1,k+1} = \left(\hat{\rho}_{k+1} - \sum_{j=1}^{k} \hat{\rho}_{k+1-j}\hat{\alpha}_{kj} \right) \times \left(1 - \sum_{j=1}^{k} \hat{\rho}_j \hat{\alpha}_{kj} \right) \\ \hat{\alpha}_{k+1,j} = \hat{\alpha}_{kj} - \hat{\alpha}_{k+1,k+1} \times \hat{\alpha}_{k,k-j+1} \end{cases} . \tag{5}$$

The order of the model can be preliminarily determined by judging the truncation of $\hat{\rho}_k$, $\hat{\alpha}_k$.

Make autocorrelation and partial correlation function graphs based on ultraviolet index data. The autocorrelation function graph of the ultraviolet index sequence is shown in **Figure 5**.

According to the **Figure 5**, it can be seen that the autocorrelation coefficients after the second order fluctuate around 0, that is, they are all within the confidence boundary.

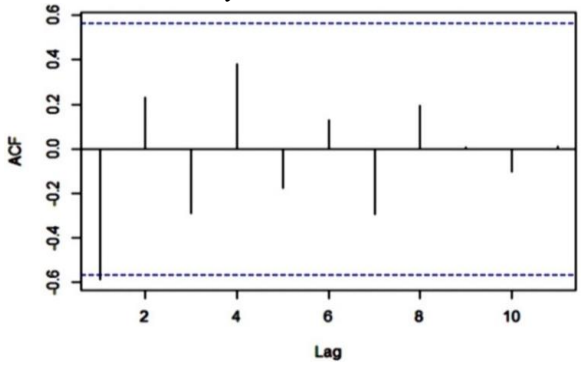


Fig. 5. The autocorrelation function graph of the ultraviolet index sequence

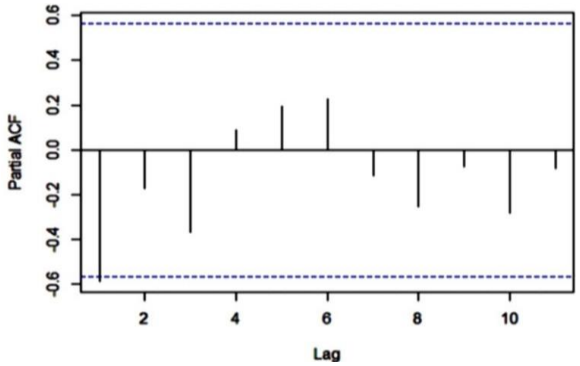


Fig. 6. The partial autocorrelation function of ultraviolet index sequence

Carry out root test (p=0.01<0.05) and white noise test (p=0.02168<0.05) on the data after autocorrelation function calculation. The test results meet the requirements for partial autocorrelation function calculation.

Make a graph of partial autocorrelation function of ultraviolet index sequence, the result is shown in **Figure 6**.

According to the partial autocorrelation graph, it can be seen that the partial autocorrelation coefficients are all within the confidence boundary after order 2.

According to the autocorrelation function graph and part of the autocorrelation function graph, the following models can be known:

•ARMA(0,2) model: that is, the autocorrelation graph is reduced to 0 after lagging by 2 order, and the partial autocorrelation is reduced to 0, it is a moving average model (MA model) with order q=2;

•ARMA(2,0) model: that is, the partial autocorrelation graph is reduced to 0 after the lag 2 order, and the autocorrelation is reduced to 0, it is an autoregressive model (AR model) with level p=2;

•ARMA(2,2) model: Even the autocorrelation and partial autocorrelation are reduced to zero. It is a mixed model (ARMA model).

It is highly subjective to directly determine the order of the model through tailing and censoring. For the estimation of model parameters, we need to perform a weighted evaluation on the loss and regularization terms. When selecting parameters, we need to balance the prediction error and model complexity. So we need to determine the order of the model according to the information criterion function method (AIC).

Finally, according to the AIC criterion and the AIC results of each model in Table 2, the optimal model is determined to be ARIMA (2,3,2), AIC=112.69.

3.3 Model Verification and Model Application

Table 2. The AIC results of each model

Model	AIC
ARIMA(0,3,2)	113.76
ARIMA(2,3,0)	130.30
ARIMA(2,3,2)	112.69

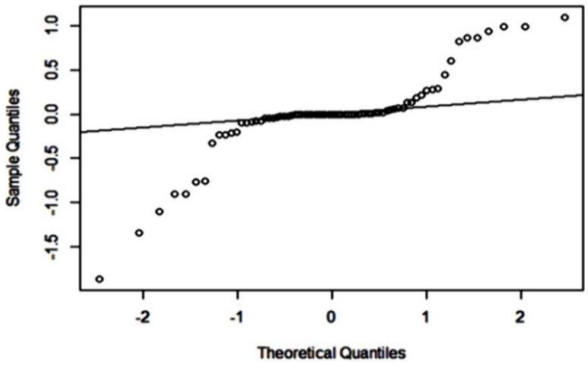


Fig. 7. Residual plot of ARIMA model

See if the model's residuals satisfy an independent normal distribution with a mean of zero.

From the **Figure 7**, it can be seen that the scatter is basically at the two ends of the straight line, so the residuals meet the normal distribution.

Significance test on the model:

p=0.6404>0.05, reflecting that the residual is a white noise sequence, so the residual white noise test passes.

Significance test on the parameters:

P_1= 0.06897761>0.05 ,

P_2= 0.8891828>0.05,

P_3= 0.09903509>0.05,

P_4= 0.06141894>0.05, parameter significance test also passed.

Calculate model accuracy:

Bring in the parameters and after fitting, calculate the model accuracy. Root mean square error (RMSE) and mean absolute error (MAE) are used to evaluate the prediction accuracy of the method [17].

MAE is a common measure of forecast error in time series analysis. It measures the average magnitude of a set of forecast errors:

$$MAE = \frac{1}{N}\sum_{t=1}^{n}\left|y_t - f_t\right|. \tag{6}$$

where n is the number of observations during the total evaluation period, is the observed value at time t, and is the predicted value.

Equation (6) shows that MAE is the average value of the absolute deviation between the predicted value and the corresponding observation value. RMSE is a secondary scoring rule used to measure the average magnitude of error.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - f_{(t)})^2}\ . \tag{7}$$

The difference between the predicted value and the corresponding observed value is squared, summed, and the average number of samples is averaged. Finally, take the square root of the average. Since the error is squared before averaging, RMSE gives higher weight to larger errors. This means that RMSE is most useful when big mistakes are not needed.

Table 3. Forecast results

time	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
0:00	0	-0.642477	0.6577462	-0.9866253	1.001895
1:00	0	-1.457045	1.4794547	-2.2342892	2.256699
2:00	0	-2.557124	2.5944630	-3.9206667	3.958006
3:00	0	-3.825297	3.8719525	-5.8626362	5.909291
4:00	0	-5.290319	5.3512132	-8.1069630	8.167857
5:00	0	-6.901521	6.9731178	-10.5739160	10.645513
6:00	0	-8.673376	8.7588861	-13.2874170	13.372927
7:00	0	-10.579488	10.6767307	-16.2056696	16.302912
8:00	5.048e-13	-12.626980	12.7380799	-19.3407059	19.451806

In this model, MAE=0.2604444,$RMSE$=0.4790116, indicating that the sequence is successfully modeled.

Use the model to predict the value of the next 9 hour:

As the value of ultraviolet index can not be negative, the final accurate result is obtained by processing the data of prediction result as shown in Table 3.

Draw a prediction map, the result is shown in **Figure 8**.

The thin dashed line represents the original data, and the solid line represents the fitted value. It can be seen from the figure that the fitting effect is very good. The prediction result is displayed as a thick dotted line, dark gray indicates a confidence interval with 80% confidence, and light gray indicates a confidence interval with 95% confidence.

3.4 Conclusion

It is feasible to predict the ultraviolet index by applying ARIMA(2,3,2) model. After verification, the model can better fit the trend of the ultraviolet index in Shanghai, China. The research results have very high value for the control of solar energy in the smart micro grid.

The implemented technique is effective and precise for ultraviolet index prediction but the ultraviolet index is affected by altitude, weather conditions and Ozone layer thickness, etc. The impact and the constraints of the model itself can only be short-term forecasts. If long-term prediction is required, the model can be optimized or combined with other models to establish a hybrid model. In the future, we plan to develop our own fuzzy logic techniques for ultraviolet index forecasting. Finally, we need to compare those techniques along with ARIMA model to find the best one.

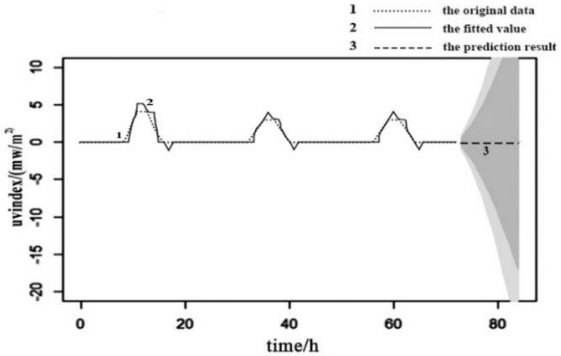


Fig. 8. Predicted value through ARIMA model

Acknowledgment

The authors acknowledge financial support from the Natural Science Foundation of Educational Department of Liaoning Province (Grant: J2020053), and Technology Innovation Fund (Grant: 2020JJ26GX029) and would like to express many thanks to the support of Dalian Key Laboratory of Smart Micro-grid and Green Recycling Industry.

References

[1] Rana, M., Li, L.: An Overview of Distributed Microgrid State Estimation and Control for Smart Grids. Sensors. pp. 4302-4325 (2015)

[2] Pavan Kumar, Y, V., Bhimasingu, R.: Key Aspects of Smart Grid Design for Distribution System Automation: Architecture and Responsibilities. Procedia Technology. pp. 352-359 (2015)

[3] Venkataramanan, G., Marnay, C.: A larger role for microgrids. IEEE Power & Energy Magazine. pp. 78-82 (2008)

[4] Vaikund, H., Srivani, S, G.: Trends in Energy Management System for Smart Microgrid-An Overview. Advances in Signal and Data Processing. pp. 15-28 (2021)

[5] Fan, R., Xu, N., Wang, W., Li, Q., Meng, T.: Research on Photovoltaic Power Generation Efficiency Detection & Case Analysis. Vol. 6, pp. 244-249. Proceedings of the 2017 2nd International Conference on Machinery, Electronics and Control Simulation (MECS 2017), China (2016)

[6] Qt, A., Peng, W, A., Wei, T, A., et al.: Benefit allocation model of distributed photovoltaic power generation vehicle shed and energy storage charging pile based on integrated weighting-Shapley method. Global Energy Interconnection. pp. 375-384 (2020)

[7] Zhao, Y., She, S., Chen, Y., Lin, P., Zhao, J., Zhang, M., Wang, L., Zhang, W., Liu, Y.: Research on Photovoltaic Power Generation Output Forecasting Model Based on Wavelet SVM and Micrometeorological Influencing Factor Analysis. Proceedings of the 2020 the 4th International Conference on Innovation in Artificial Intelligence. pp. 250-256 (2020)

[8] Nikkhajoei, H., Lasseter, R, H.: Distributed Generation Interface to the CERTS Microgrid. IEEE transactions on power delivery pwrd. pp. 1598-1608 (2009)

[9] Yona, A., Senjyu, T., Funabashi, T.: Application of recurrent neural network to short-term-ahead generating power forecasting for photovoltaic system. IEEE Power Engineering Society General Meeting. pp. 24-28 (2007)

[10] Vapnik, V.: An overview of statistical learning theory. IEEE Transactions on Neural Networks. pp. 988-999 (1999)

[11] Farhangi, H.: Intelligent Micro Grid Research at BCIT. IEEE Electrical Power and Energy Conference. pp. 18-28 (2008)

[12] Platt, J.: Fast training of support vector machines using sequential minimum optimization. Advance in Kernel Methods-support Vector Learning. pp. 185-208 (1999)

[13] Xu, J.: Application of Grey Theory in the Load Prediction of Power System. JiangXi Hydraulic Science and Technology. pp. 216-219 (2005)

[14] Wang, H., Huang, J., Zhou, H., Zhao, L., Yuan, Y.: An Integrated Variational Mode Decomposition and ARIMA Model to Forecast Air Temperature. Sustainability. pp. 1-11 (2019)

[15] Zheng, P.: Telecommunication Data Forecasting based on ARIMA Model. International Journal of Electrical and Electronics Engineers. pp. 451-458 (2016)

[16] Han, P., Wang, P, X., Shu, Y, Z., et al.: Drought forecasting based on the remote sensing data using ARIMA models. Mathematical and Computer Modelling. pp. 1398-1403 (2010)

[17] Wang, W., Lu, Y.: Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. IOP Conference Series: Materials Science and Engineering. pp. 12-49 (2018)

Research on Optimal MPPT of Photovoltaic System Based on Disturbance Observation Method

Jianhua Deng[1], Yanping Wang[2,*]

{djhisgood@yeah.net[1], wangyp@dlpu.edu.cn[2]}

[1]Department of Information Science and Engineering Dalian Polytechnic University Dalian, P. R. China

[2]Dalian Key Laboratory of Smart Micro-grid and Green Recycling Industry, Liaoning province

Abstract. The biggest problem with the photovoltaic power generation system is that the energy utilization rate is too low due to the instability of its power. Therefore, a reasonable method must be developed to track its maximum power so that the photovoltaic cell can output the maximum power at all times. Compared with the large oscillation at the maximum power point of the traditional disturbance observation method, an improved disturbance observation method with PID control duty cycle is proposed. The duty cycle of the boost circuit is controlled by the PID algorithm and the disturbance step length of the *P-D* characteristic is changed. Match the load impedance with the input impedance of the photovoltaic cell, so the photovoltaic cell outputs the maximum power, increasing the tracking efficiency.

Keywords: Photovoltaic system; maximum power tracking; PID control

1 Introduction

Since the output power of photovoltaic cells will change with changes in the external environment and load, it must be tracked and controlled so that photovoltaic cells always output at maximum power. At present, the commonly used maximum power tracking (MPPT) methods for photovoltaic systems include constant voltage method, disturbance observation method, conductance increment method, etc. [1-3]. Literature [3] proposes a tracking method that directly perturbs the duty cycle. Literature [9] verifies the effectiveness of directly perturbing the duty cycle through simulation, but there is a certain deviation, which is different from the initial value of the duty cycle.

Select the relevant, this article introduces a method based on the perturbation observation method, which is an optimized method in the literature [8]. The general process of this method is to add an MPPT module between the photovoltaic cell and the load.

The function of this module is to sample the duty cycle and output power, and then compare the power of equidistant points on both sides of the duty cycle, and finally determine the direction of the disturbance.

* Corresponding Author: Yanping Wang

Fund Project: Dalian Science and Technology Innovation Fund (2020JJ26GX029)
The Natural Science Foundation of Educational Department of Liaoning Province (Grant: J2020053)

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307652

The function of this module is to sample the duty cycle and output power, and then compare the power of equidistant points on both sides of the duty cycle, and finally determine the direction of the disturbance. Compared with the traditional disturbance observation method, the tracking efficiency is improved. Effectively avoid the oscillation phenomenon of power tracking. By controlling the duty cycle of the switching element in the Boost circuit, the load impedance is matched with the input impedance of the photovoltaic cell, so the maximum power is output and the photovoltaic cell is boosted at the same time.

According to the characteristics of the photovoltaic cell, the simulink model is established, and the improved disturbance observation method is combined to simulate it, and the result of MPPT is obtained.

2 Photovoltaic cell modeling and output characteristic simulation

2.1 Photovoltaic cell equivalent circuit

There are three types of equivalent circuits for photovoltaic cells [4]. The more commonly used equivalent circuits are shown in **Figure 1**.

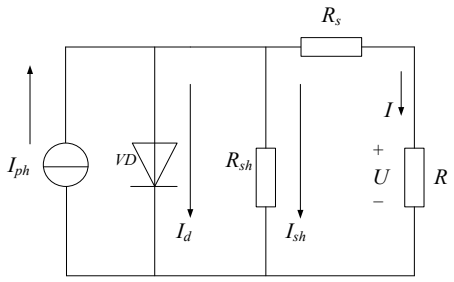


Fig. 1. Solar photovoltaic cell equivalent circuit.

The *I-U* characteristic curve of an ideal photovoltaic cell is obtained by subtracting the current of the diode from the photovoltaic cell photocurrent, but the photovoltaic cell in the experiment should consider the actual internal loss of the battery. Therefore, the model shown in the above **[1]**figure not only considers the series resistance R_s, the resistance value is generally low, which means the loss resistance of the line [5], but also considers the parallel resistance R_{sh}, which means the bypass resistance, and the resistance value is generally higher. , The order of magnitude is $10^3\Omega$, I_{sh} is the leakage current caused by the process, I_{ph} is the photogenerated current proportional to the light intensity, I is the output current, and I_d is the diode saturation current [6]. According to the circuit diagram, Kirchhoff's theorem can be obtained:

$$I = I_{Ph} - I_d - I_{sh} \tag{1}$$

The expression of I_d and I_{sh} in formula (1) is (2).

$$I_d = I_0\{\exp\left[\frac{q(U+I\times R_S)}{nKT}\right] - 1\} \qquad I_{sh} = \frac{U+IR_S}{R_{sh}} \tag{2}$$

In the formula, I_0 is the saturated reverse current, K is Boltzmann's constant, and the value is 1.38x 10^{-23}*J/K*; n is the ideal factor of the diode, usually about 1.3, and q is the charge of the

electron, which is $1.6\text{x}10^{-19}C$.

Substituting (2) into (1) can obtain the characteristic expression (3) of photovoltaic cells:

$$I = I_{ph} - I_0\left\{\exp\left[\frac{q(U+I\times R_S)}{nKT}\right]-1\right\} - \frac{U+IR_S}{R_{sh}} \tag{3}$$

Since the value of the bypass resistor R_{sh} is particularly large, the bypass current I_{sh} is small enough to be ignored, so the above equation can be rewritten as (4).

$$I = I_{ph} - I_0\left\{\exp\left[\frac{q(U+I\times R_S)}{nKT}\right]-1\right\} \tag{4}$$

Because R_s in the equivalent circuit diagram of the photovoltaic cell is much smaller than the forward conduction resistance of the diode, it can be approximated as $I_{ph}=I_{sc}$.

According to the *I-U* relationship of the photovoltaic cell, the output characteristic equation can be obtained. At the same time, the engineering mathematical model is established, and $I_0=A_1I_{sc}$, $\frac{nKT}{q}=A_2U_{oc}$.Available (5):

$$I = I_{sc} - A_1I_{sc}\left\{\exp\left[\frac{U}{U_{oc}A_2}\right]-1\right\} \tag{5}$$

Solve A_1 and A_2 by formula (5):

$$A_1 = (1-\frac{I_m}{I_{sc}})\exp\left(\frac{-U_m}{A_2U_{oc}}\right) \tag{6}$$

$$A_2 = \left(\frac{U_m}{U_{oc}}-1\right)\left[\ln\left(1-\frac{I_m}{I_{sc}}\right)\right]^{-1} \tag{7}$$

In practical applications, the short-circuit current of photovoltaic cells under standard conditions (light intensity $S_c=1000W/m^2$, battery temperature T=25 degrees Celsius) is I_{sc}, the open circuit voltage is U_{oc}, the maximum power point U_m, and the maximum power point current I_m. These parameters are fixed values, so A_1, A_2 can be obtained as a constant. When the temperature or light intensity changes, it is necessary to recalculate the 4 parameters under the new conditions as $I_{sc\text{-}n}$, $I_{m\text{-}n}$, $U_{oc\text{-}n}$, $U_{m\text{-}n}$. And recalculate $A_{1\text{-}n}$, $A_{2\text{-}n}$, under the new conditions:

$$\Delta T = T - T_c \tag{8}$$

$$\Delta S = \frac{S}{S_c} - 1 \tag{9}$$

$$I_{sc-n} = I_{sc}(\frac{S}{S_c})(1+a\Delta T) \tag{10}$$

$$I_{m-n} = I_m(\frac{S}{S_c})(1+a\Delta T) \tag{11}$$

$$U_{oc-n} = U_{oc}[(1-c\Delta T)\ln(e+b\Delta S)] \tag{12}$$

$$U_{m-n} = U_m[(1-c\Delta T)\ln(e+b\Delta S)] \tag{13}$$

The coefficients in the formula are typical values: a=0.0025, b=0.5, c=0.00288. To sum up, build a simulation model in simulink according to the engineering mathematical model of photovoltaic cells as shown in the figure. The parameter settings select the photovoltaic cell model CN-200S on the market: $P_m=120W$, $I_{sc}=7.68A$, $I_m=6.94A$, $U_{oc}=17.5V$, $U_m=28.8V$. The coefficients a, b, c in the formula are typical values. The model of the photovoltaic cell is shown

in the **Figure 2**.

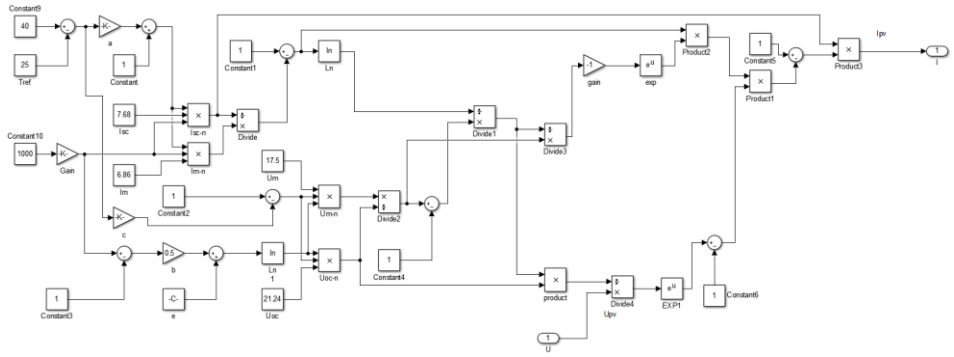


Fig. 2. Photovoltaic cell model.

2.2 Photovoltaic cell simulation results

Run the simulation of the photovoltaic cell model, and finally get the output characteristic curve of the photovoltaic cell. As shown **Figure 3**.

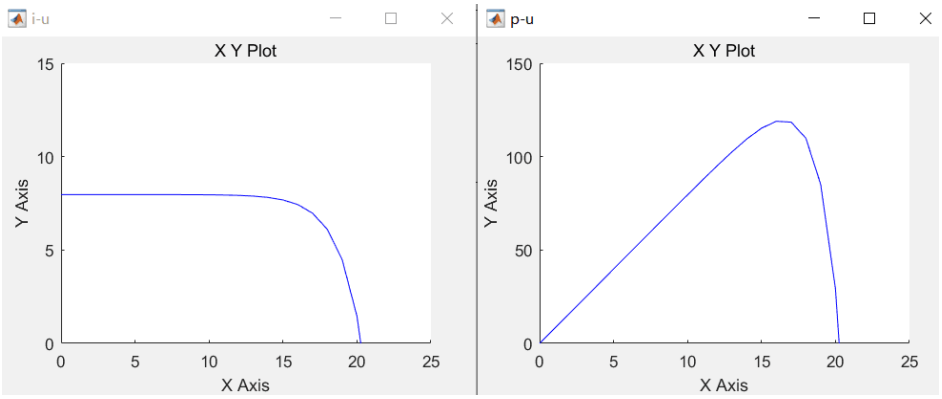


Fig. 3. *U-I*, *U-P* characteristic curves of photovoltaic cells.

Figure 4 is the *P-U* characteristic curve diagram of the output of the photovoltaic cell when the temperature is constant and the light intensity is different. The image shows that the output power of the photovoltaic cell changes with the voltage. Under certain external conditions, the photovoltaic cell has a maximum output power.

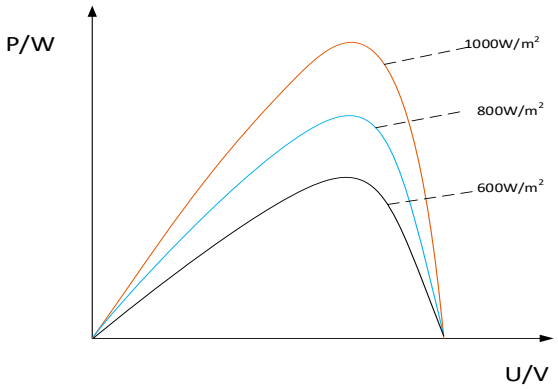


Fig. 4. Photovoltaic cell *P-U* diagram.

It can be seen from **Figure 4** that when the temperature of the photovoltaic cell is constant, as the light intensity increases, the short-circuit current of the photovoltaic cell will increase with the increase of the light intensity; the output power of the photovoltaic cell will increase with the increase of the light intensity. And the maximum output power also increases with the increase of light intensity. It is worth mentioning that their changes are very obvious.

Therefore, we can get a conclusion, that is the power-voltage relationship and the current-voltage relationship of photovoltaic cells is not a linear relationship, but an obvious non-linear relationship.

At the same time, we can also get that the short-circuit current and output power of photovoltaic cells are greatly affected by the light intensity, and the open circuit voltage is relatively small.

3 Disturbance Duty Cycle Control Method

3.1 Traditional disturbance observation

During the entire process of photovoltaic power generation grid-connected system, the efficiency of photovoltaic cells is worth considering. According to statistics, the efficiency of photovoltaic cells is only about 10%, which limits the operation of photovoltaic power generation systems to a certain extent. It is known that the output characteristics of photovoltaic cells are affected by light intensity and temperature, and it has been known that the output characteristic curve of photovoltaic cells is not a straight line, so they are in a nonlinear relationship.

When the external environmental conditions change, the output of photovoltaic cells Voltage and current will change, output power will also change, but no matter how it changes, there is only one maximum power point.

The process of traditional disturbance observation method is to artificially apply a disturbance amount ΔU to the output voltage of the photovoltaic cell, observe the power change, if the output power increases, continue to impose disturbance in this direction, if the power decreases, impose disturbance in the reverse direction.

When the power is reduced again, it means that it is near the maximum power point, and it continues continuously, and finally the power has been oscillating near the maximum power point. This method is relatively simple to implement, but there will be a lot of power loss at the maximum power point. The traditional disturbance observation method MPPT model is shown in the **Figure 5**.

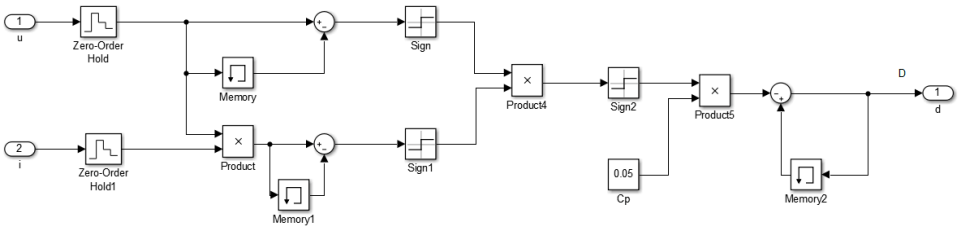


Fig. 5. Traditional disturbance observation method MPPT model.

3.2 PID disturbance duty cycle observation method

Because the *P-D* curve of the photovoltaic system is basically similar to the *P-U* image, the disturbance observation method is also applicable. This method tracks the power by controlling the duty cycle of the switching element in the Boost circuit. The specific method is as follows [7].

As shown in **Figure 6**, it is a Boost circuit with a period of T and a duty cycle of d. According to the circuit, the relationship between the duty cycle and the voltage can be obtained.

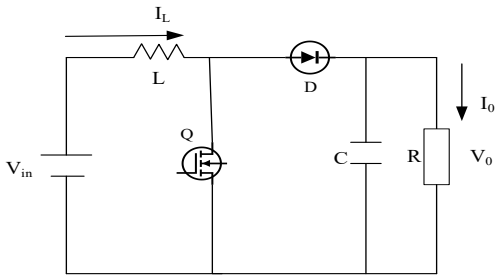


Fig. 6. Boost circuit.

When the switching element is turned on, the photovoltaic power supply charges the inductor and the capacitor discharges the resistor. The current that the inductor increases during this process is $\Delta I_L = \frac{V_{in}}{L} dT$.

When the switching element is turned off, the inductor and the photovoltaic power supply both charge the capacitor, and the current reduced by the inductor is $\Delta I_L' = \frac{V_{in}-V_0}{L} T(1-d)$.

In a cycle, when a continuous current flows through the inductor, the charging current and

the discharging current of the inductor are equal, so you can get ΔI_L=$\Delta I_L'$, so $V_0 = \frac{V_{in}}{1-d}$, $I_0 = I_L(1-d)$.

Therefore, the input impedance of the photovoltaic cell can be obtained R_{in}=R(1-d)^2, so only the duty cycle d needs to be adjusted so that the input impedance of the photovoltaic cell is equal to the load impedance, and the maximum power can be obtained. Improved tracking efficiency.

3.3 How to change the duty cycle through PID control

Track the maximum power of the photovoltaic cell and then output the maximum power. This paper proposes an MPPT strategy based on the perturbation and observation method. The method is to obtain a duty cycle that can be automatically changed through the perturbation observation method, and then PID control the duty cycle Then output the final duty cycle. The specific method is as follows.

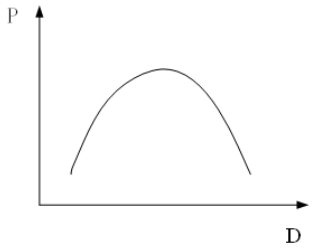


Fig. 7. Photovoltaic cell *P-D* diagram

The *P-D* relationship diagram of the photovoltaic cell can be obtained by calculation as shown in the **Figure 7**, so any point in the figure above indicates that the size of *D* at a certain moment is D_k, and then two equidistant points are taken before and after this point.

They are D_{k-1}, D_{k+1} and D_{k+1}=D_k+ΔD， D_{k-1}= D_k -ΔD, and finally 9 different situations of three point powers can be obtained according to different positional relationships [8].

Suppose that point *A* on the *P-D* diagram is taken as the original operating point, points *B* and *C* are two sampling points with the same step distance from *A*, and the power of the three points is P_k for point *A* and P_{k-1} for point *B*. Point *C* is P_{k+1}. Then the direction of duty cycle disturbance can be determined by comparing the power relationship between *A* and *B*, *C* and *A*.

If both comparisons are greater than, it can be determined that the photovoltaic cell is working on the left side of the maximum power point at this time, and the duty cycle should be increased at this time. If both comparisons are less than, it can be determined that the photovoltaic cell is working on the right side of the maximum power point at this time, and the duty cycle should be reduced at this time. If one of the two comparisons is greater than one and the other is less than, it can indicate that the photovoltaic cell is working near the maximum power point at this time, and the duty cycle is not adjusted.

Through the experimental simulation results, it can be seen that the method of pairwise comparison to determine the disturbance direction proposed in this paper has certain advantages. Its control logic includes every situation, and the tracking speed and tracking efficiency have a certain improvement compared with traditional methods.

First set the reference quantities *H* and *Y*, [9] when P_k> P_{k-1}, *Y*=1, otherwise *Y*=-1. When P_{k+1}>P_k, *Y*=1 otherwise *Y*=-1. when P_k=P_{k-1}, P_{k+1}=P_k, *Y*=0. *H*=Y_1+Y_2. The MPPT control process

introduced in this article is as follows:

If $P_k>P_{k-1}$, $P_{k+1}>P_k$, then H=2
If $P_k>P_{k-1}$, $P_{k+1}<P_k$, then H=0
If $P_k<P_{k-1}$, $P_{k+1}<P_k$, then H=-2
If $P_k>P_{k-1}$, $P_{k+1}=P_k$, then H=1
If $P_k=P_{k-1}$, $P_{k+1}<P_k$, then H=-1
If $P_k=P_{k+1}$, $P_{k-1}=P_k$, then H=0

The above six situations do not include misjudgment. If a misjudgment occurs, it means that the system has made an error. Therefore, they are all judged as H=0, and the existing duty cycle remains unchanged [10].

This control method of changing the duty cycle based on the disturbance observation method can significantly reduce the oscillation of the output power of the photovoltaic cell at the maximum power point.

The duty cycle is sampled and the relationship between the power is compared and then an appropriate duty cycle is output. Finally, the output duty cycle is PID adjusted and the final duty cycle is output to the switching element in the Boost circuit.

The photovoltaic cell can output the maximum power, so that the cell always works near the maximum power point to achieve the purpose of tracking the maximum common power. The control principle diagram and control flowchart are shown in the **Figure 8** and **Figure 9**.

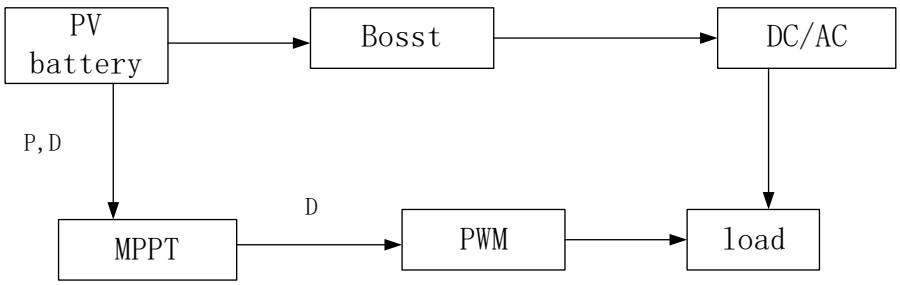


Fig. 8. Schematic diagram of duty cycle disturbance.

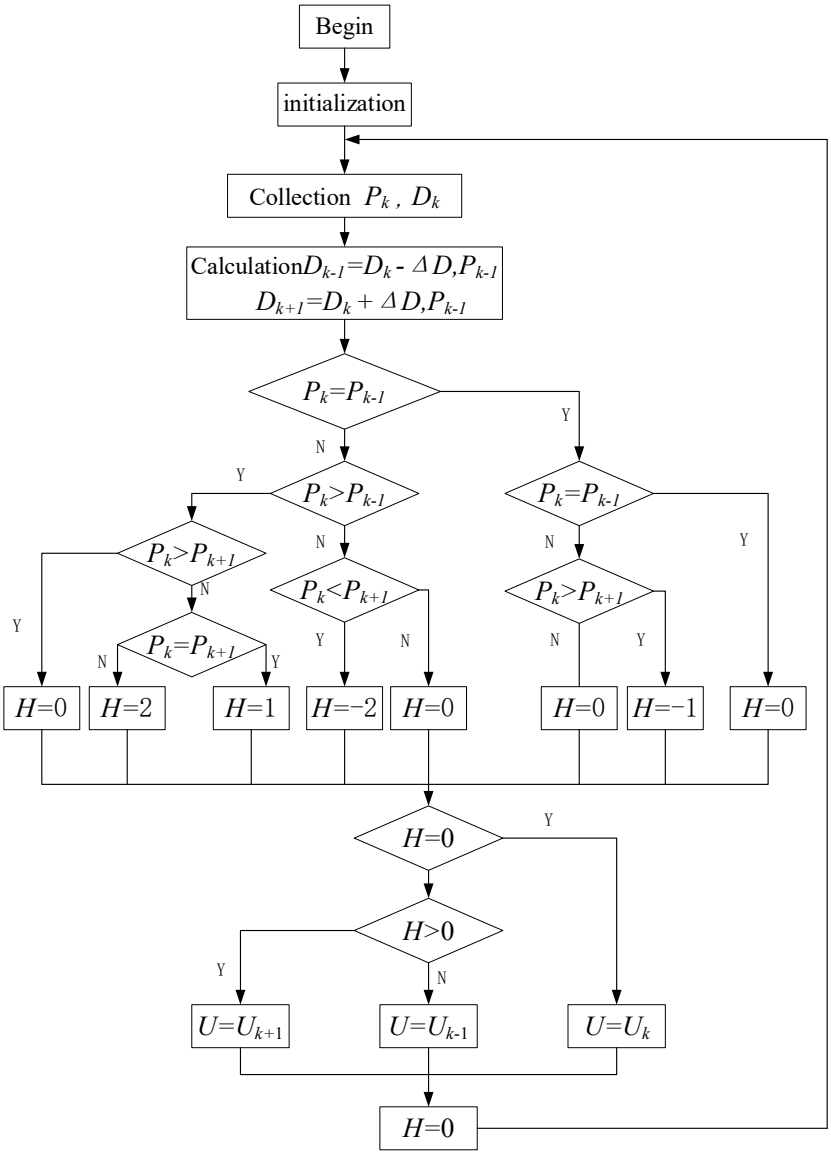


Fig. 9. Duty cycle disturbance control flowchart

The overall model of photovoltaic cells can be seen as a combination of several modules, first is the photovoltaic cell module, then the Boost circuit module, MPPT control module, and PWM pulse width modulation module. Combine various modules together to form the entire photovoltaic power generation system [11]. The simulation model is shown in the **Figure10**.

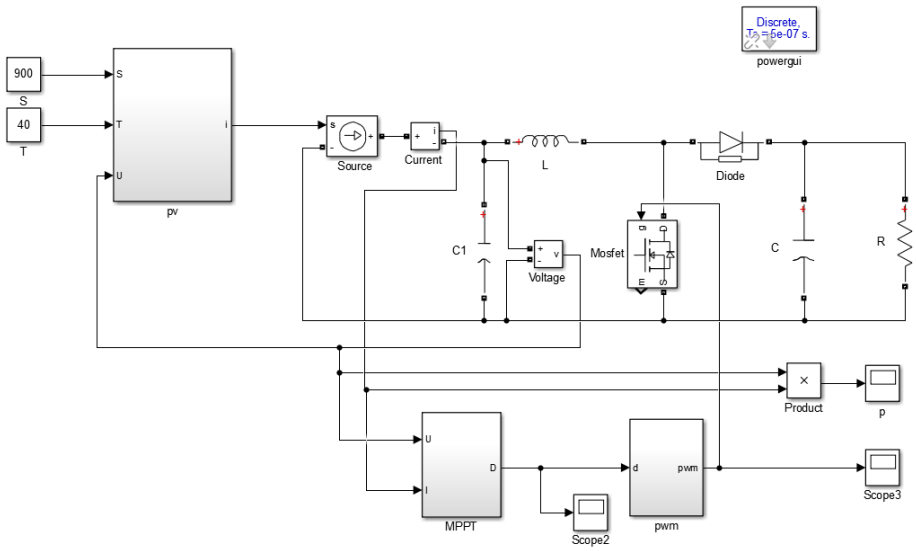
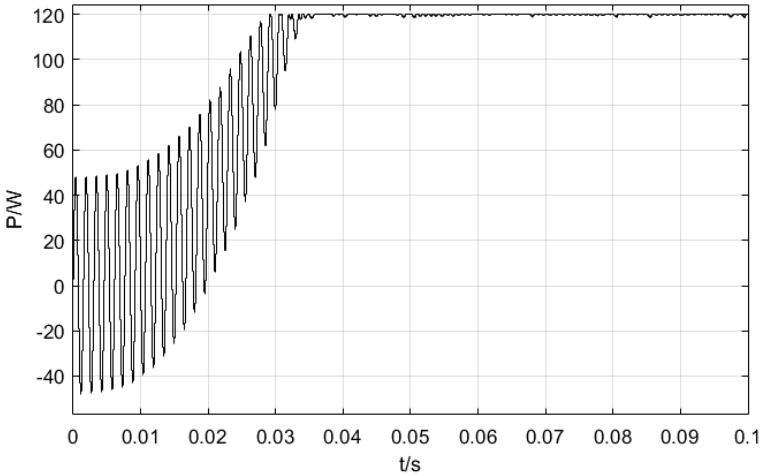


Fig. 10. Overall model of photovoltaic cells.

The power tracking results obtained by running the above model in Matlab/Simulink are shown in **Figure 11** and **Figure 12**.



$1000W/m^2$,25℃

Fig. 11. Power tracking graph.

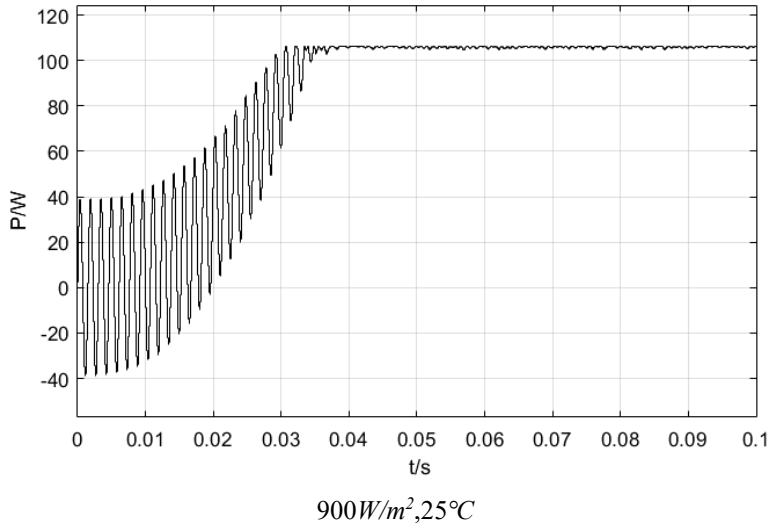


Fig. 12. Power tracking graph.

4 Conclusion

At present, there are many methods for maximum power tracking of photovoltaic systems. The traditional methods can no longer meet people's requirements due to their imperfect tracking efficiency and tracking accuracy. Therefore, this article proposes an improved disturbance observation method. And the feasibility of the method is verified through simulation experiments. This method avoids the problem of photovoltaic cells oscillating at the maximum power point, reduces power loss and increases the efficiency of power tracking. For the power tracking research of photovoltaic systems, A certain reference price.

Acknowledgments

The authors acknowledge financial support from the Natural Science Foundation of Educational Department of Liaoning Province (Grant: J2020053), and Technology Innovation Fund (Grant: 2020JJ26GX029) and would like to express many thanks to the support of Dalian Key Laboratory of Smart Micro-grid and Green Recycling Industry.

References

[1] Lin Zhou., Jian Wu., Ke Guo.: Overview of the maximum power point tracking control method for photovoltaic arrays[J]. High voltage technology,2008(06):1145-1154.
[2] Liu F., Duan S., Liu B., et al.: A Variable Step Size INC MPPT Method for PV Systems[J]. IEEE Transactions on Industrial Electronics, 2008, 55(7):2622-2628.
[3] Koutroulis., Eftichios., Kalaitzakis., et al.: Development of a Microcontroller-Based Photovoltaic Maximum Power Point Tracking Control System[J]. IEEE Transactions on Power Electronics, 2001.
[4] Xuemei Yao.: Research on Optimal Control of Solar Energy (Photovoltaic) Application System[D]. Qingdao University, 2009.
[5] Xiangyu E., Jianjun Tan., Lei Zou., et al.: Improved photovoltaic MPPT algorithm based on disturbance observation method[J]. Journal of Hubei University for Nationalities (Natural Science Edition), 2018.
[6] Yibin Zheng.: Maximum power tracking and grid connection control method of photovoltaic power generation system[D]. Southwest Jiaotong University, 2016.
[7] Yidan Li., Duting Wang., Xianhui Zhu., Hongming Kang., Sanmin Wei.: Maximum power point tracking control based on adaptive duty cycle disturbance method[J]. Heilongjiang Electric Power,2015,37(02):102-104+115.
[8] Shengyun Jing.: New type of boost converter and its application in photovoltaic power generation system[D]. North China University of Technology, 2018.
[9] Bo Fang., Shuke Luo., Longyun Kang.: Research on MPPT and Simulation of Photovoltaic Duty Cycle Disturbance Control[J]. Renewable Energy,2013,31(10):5-9.
[10] Wu Yuwei., Shi Bin., Zhu Haiyong., Zhang Jialing.: Maximum power point tracking of photovoltaic cells based on variable step disturbance observation method[J]. Electrical Technology, 2014(06): 23-25.
[11] Huang Yao., Huang Hongquan.: Incremental conductance method to realize the maximum power point tracking control of photovoltaic system[J]. Modern Electronic Technology, 2008(22): 18-19.

Research on MPPT Control of Photovoltaic Cells Based on Fuzzy Control

Fengshan Liu[1], Yanping Wang[2*]
{824162388@qq.com[1], wangyp@dlpu.edu.cn[2] }

Dalian Polytechnic University ,Liaoning province [1], Dalian Polytechnic University, Dalian Key Laboratory of Smart Micro-grid and Green Recycling Industry, Liaoning province [2]

Abstract. Changes in factors such as light intensity and temperature can cause fluctuations in the power of photovoltaic cells. To solve this problem, this paper proposes a maximum power point tracking strategy (MPPT) based on fuzzy control. According to the fuzzy control algorithm, this paper determines the membership function and fuzzy rules, builds a fuzzy controller, sets up a photovoltaic system, and performs simulation verification in MATLAB/Simulink. The results show that, in terms of tracking speed, compared with the commonly used Perturbation and Observation method, the fuzzy control is improved by five times faster, and the method also has better smoothness and anti-interference.

Keywords: Photovoltaic cells; MPPT; Fuzzy logic control; Perturbation and Observation.

1 Introduction

Due to the development of society and technology, people's demand for electricity is increasing, coupled with the excessive consumption of traditional fossil energy and the aggravation of environmental problems, people are focusing on clean energy, such as wind, solar, tidal energy, etc.[1]. In recent years, with the development of technology, the proportion of clean energy power generation has increased, and the proportion of photovoltaic power generation has reached 3.1% at the end of 2019.

In photovoltaic power generation systems, the output characteristics of photovoltaic cells are not only related to their own materials, but also to environmental factors such as temperature and light intensity. Therefore, making the system run stably at the MPP is a key issue. The current MPPT algorithms include Open Circuit Voltage method, Short Circuit Current method, Perturbation and Observation method, Increment Conductance method, neural network control method, fuzzy control and other intelligent algorithms [2-5].

At present, Perturbation and Observation method is widely used because of its tracking speed is good and the simple structure, while the control accuracy is not high enough in some situations and the tracking speed is not fast enough either.

The better improvement is the variable step size disturbance observation which has high convergence stability [6].This algorithm is generated based on the mathematical principle that the derivative of the MPP power point is 0. Although the performance of the algorithm is optimized, the calculation of the maximum step size and speed factor is difficult, and sudden

* Corresponding author: Yanping Wang
Fund Project: Dalian Science and Technology Innovation Fund(2020JJ26GX029)

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307606

changes in the environment produce dead zone convergence. Literature [7, 8] proposed an optimization algorithm for the variable step size perturbation and observation method, but it is unavoidable that the model is too complex and affects real-time response.

However, currently, most MPPT strategy use the same method to track the point. It is to control the DC/DC convert circuit by changing the duty cycle, the output power can stable at MPP [9, 10]. The fuzzy control method has high control accuracy and fast tracking speed, but the complex structure of the algorithm is also complicated [11].Recently, fuzzy control has become a popular research in the control engineering. The fuzzy logic controller has made itself available not only in the laboratory work but also in industrial applications, such as the cement kiln control and the servo hydraulic cylinder position control [12]. The controller can be regarded as a set of heuristic decision rules derived from the experienced operator.

The main advantages of fuzzy control are no mathematical formulation of the system is needed and linguistic variables and approximate reasoning are used to describe the inexact objects and achieve multi-objective control [13].The main work this paper done is design a fuzzy logic controller to adjust the duty cycle so as to control the power electronics. The core of a fuzzy logic controller is to establish fuzzy tables and fuzzy rules [14, 15]. Fuzzy rules are usually summed up by expert experience, engineering experience and practice [16]. This paper improves the membership function and the fuzzy rules based on the traditional fuzzy control method.

The overall methodology for planned research is organized as follows: Section 2 introduces the construction of photovoltaic cells and the system, Section 3 introduces the fuzzy control method, Section 4 introduces the disturbance tracking method, and the final section compares and draw conclusion.

2 Photovoltaic cells and the system

2.1 Photovoltaic cells

Photovoltaic power generation is a power generation system that uses the photoelectric effect to convert solar energy into electrical energy. It has the advantages of high reliability, cleanliness, and grid connection, and is favored by all over the world. The structure of the photovoltaic cell is similar to the PN junction. When there is light, it will generate voltage. According to its principle, the equivalent circuit of the photovoltaic cell can be obtained, as shown in **Figure 1**.

According to the paper [17, 18] and the equivalent circuit diagram, the I-V equation can be obtained：

$$I = I_{SC}[1 - C_1(exp^{V/C_2V_{OC}} - 1)] \tag{1}$$

$$C_1 = (1 - \frac{I_m}{I_{SC}}) \exp^{-V_m/C_2V_{oc}} \tag{2}$$

$$C_2 = (\frac{V_m}{V_{oc}} - 1)[\ln(1 - \frac{I_m}{I_{sc}})]^{-1} \tag{3}$$

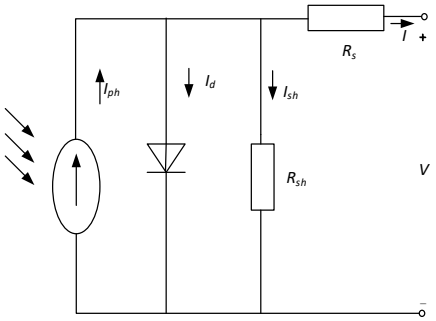


Fig. 1. Photovoltaic cell equivalent circuit diagram

The above is an engineering model, which will be affected in the actual environment, so it needs to be corrected.

$$V_{OC}^{'} = V_{OC}(1 - c\Delta T)(1 + b\Delta S) \tag{4}$$

$$V_{m}^{'} = V_{m}(1 - c\Delta T)(1 + b\Delta S) \tag{5}$$

$$I_{SC}^{'} = \frac{I_{sc}S}{S_0}(1 + \alpha\Delta T) \tag{6}$$

$$I_{m}^{'} = \frac{I_m S}{S_0}(1 + \alpha\Delta T) \tag{7}$$

Standard battery temperature: T_0 = 25℃; standard light intensity: S_0=1000W/m^2; constant a = 0.0025/℃, b = 0.5, C = 0.00288/℃, the parameters of the photovoltaic cell selected in this article are: V_{oc} = 38.3V, V_m =31.2V, I_{oc} =9.29A, I_m=8.70A, P=270W. Based on the formula and these parameters, the photovoltaic cell is modeled shown in **Figure 2**. And the simulated P-U curve under different light intensity and the same temperature is shown in **Figure 3**.

2.2 Photovoltaic system

In the smart micro-grid, distributed power generation is the core, and there are many distributed power sources. Nowadays, photovoltaic systems are widely used. In addition to the structure of the photovoltaic cell, the external environment has a great influence on it. The main influencing factors are light intensity and temperature. Under different external conditions, there will be different characteristic curves, and in each case there will be a maximum power point. The power point is very important to ensure that the system runs at the maximum power point as much as possible [19]. The photovoltaic system studied in this paper is mainly based on changing the duty cycle of the power electronic devices of the DC/DC converter. The simulation model of integrated MPPT control is mainly composed of photovoltaic cell modules, DC/DC convert circuit, MPPT control module and PWM module, and the voltage and current detection circuit. When the output voltage and current of the photovoltaic cell are measured, they are sent to the MPPT module. The MPPT controller performs optimal output and serves as the input of the PWM module to make the PWM adjust the duty cycle of the switching device of the DC/DC circuit to stabilize the photovoltaic cell at the maximum power output state.

The composition of the photovoltaic system is shown in **Figure 4**. And the simulation in SIMULINK of the system is shown in **Figure 5**.

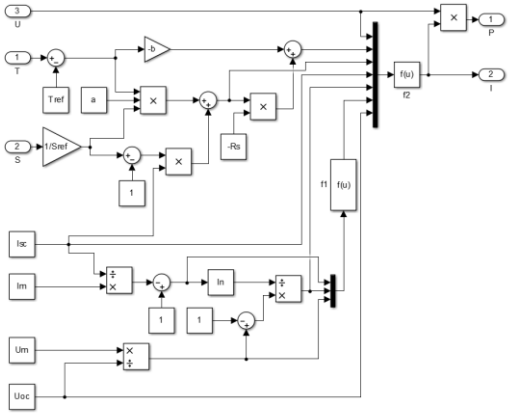


Fig. 2. Photovoltaic cell model

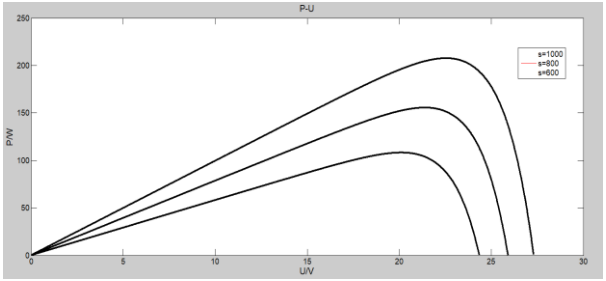


Fig. 3. P-U curve under different light intensity

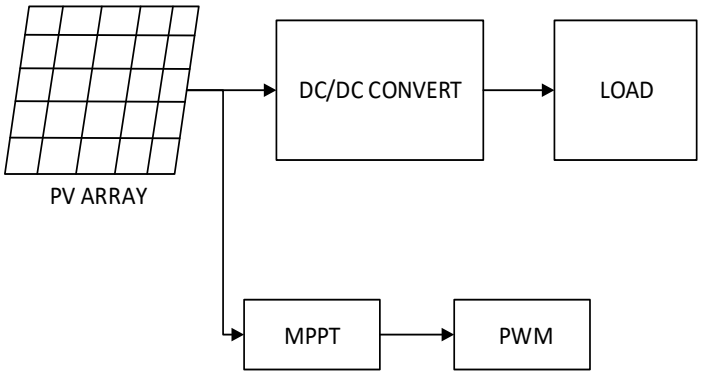


Fig. 4. The composition of the photovoltaic system

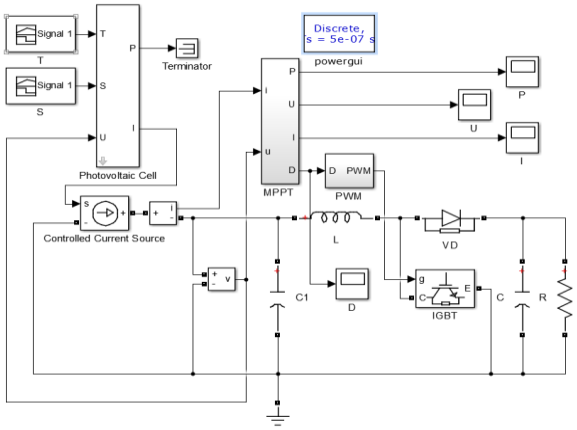


Fig. 5. The simulation of the photovoltaic system

3 Perturbation and Observation

The principle of the perturbation and observation method is to perturb the voltage, and then control the system according to the change of the power before and after the disturbance point [20].If power increases, the disturbance continues in the same direction; if the power decreases, the disturbance is applied in the opposite direction. **Figure 6** shows the model of the P&O control method with fixed step in MATLAB/Simulink.

The perturbation and observation method used to increase or decrease the voltage at a fixed step length, and constantly compare the power after the voltage change. However, due to the inherent contradiction between the response speed and stability caused by the setting of the step size, the variable step size perturbation and observation method like **Figure 7** was developed later to collect the current and voltage values, calculate the power difference between the power value and the sampling points before and after, set the disturbance coefficient, In the early stage, use a large step size to approach quickly, change

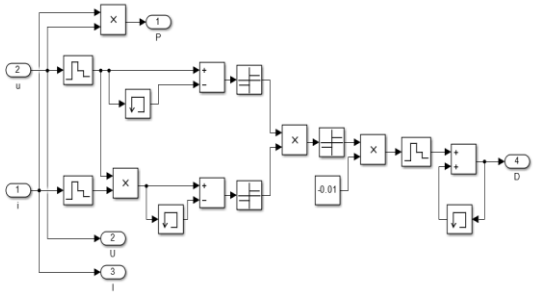


Fig. 6. The model of the P&O with fixed step

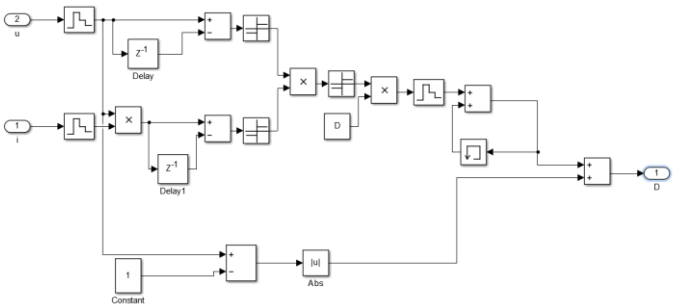


Fig. 7. The model of the P&O with variable step

the step size in the later stage, use a small step size, and then use iteration to make the step size smaller and smaller, which can quickly stabilize the MPP and minimize the amplitude.

4 Fuzzy Logic Control

Human thinking mainly relies on conceptual patterns and thinking images, rather than quantity, but the basis of computer work is just the opposite. Natural language, which is the shell of human thought, naturally has a silent house type, which computers cannot understand. Fuzzy control is based on the combination of fuzzy theory and fuzzy logic reasoning. It converts the knowledge and control experience expressed in expert language into mathematical functions through fuzzy theory and then uses computers for processing.

As the complexity, non-linearity, hysteresis and coupling of control objects increase, people's ability to acquire accurate knowledge is relatively reduced, and the possibility of using traditional precise control is also reduced, and fuzzy control has emerged. Fuzzy logic control (FLC) is a kind of language logic controller, which includes: Data acquisition- Fuzzification - Fuzzy rule logic reasoning- Defuzzification-output and other processes [21] shown in **Figure 8**.

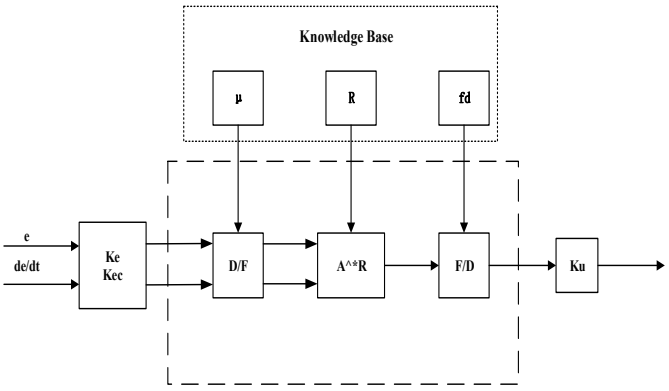


Fig. 8. Fuzzy controller schematic diagram

In physical systems, it is sometimes difficult to describe a complex system using the linearization or identification techniques. Takagi and Sugeno used fuzzy implications to express a complex system which is called the fuzzy model. In general, the construction of the fuzzy model is based on physical properties of the system, the input-output data, and the corresponding empirical knowledge and so on.

When designing the fuzzy controller, dP/dV,ΔdP/ΔdV is obtained as the deviation E and the deviation change rate EC through continuous sampling of the output voltage and current of the photovoltaic cell:

$$e(k) = \frac{P(k) - P(k-1)}{U(k) - U(k-1)} \tag{8}$$

$$e_c(k) = e(k) - e(k-1) \tag{9}$$

After quantizing factors Ke and Kec, the input variables are mapped to the fuzzy controller, and the output is obtained. Define 7 fuzzy subsets respectively to represent negative large, negative medium, negative small, zero equal, positive small, positive middle, and positive large. Their membership function is shown in the **Figure 9** and **Figure 10**.

$$\begin{cases} e = \{NB, NM, NS, ZE, PS, PM, PB\} \\ e_c = \{NB, NM, NS, ZE, PS, PM, PB\} \\ u = \{NB, NM, NS, ZE, PS, PM, PB\} \end{cases}$$

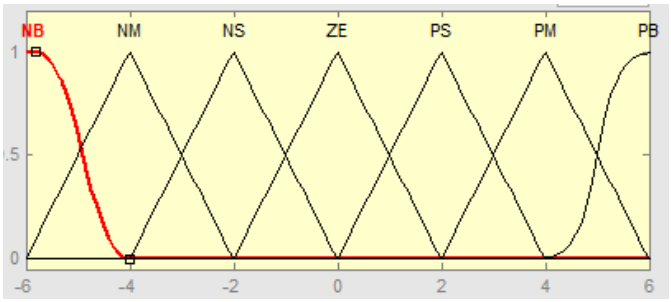


Fig. 9. The membership function of dP/dV,ΔdP/ΔdV

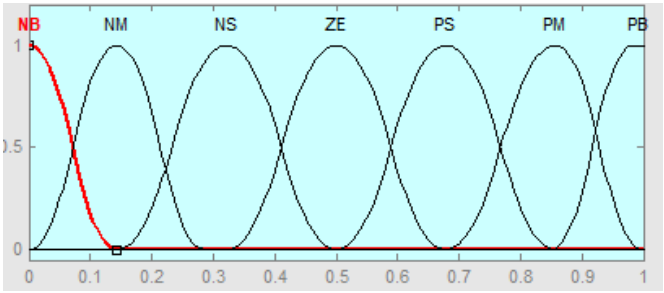


Fig. 10. The membership function of D

Table 1. Fuzzy rules

E \ EC	NB	NM	NS	ZE	PS	PM	PB
NB	PM	PB	PM	NB	NM	NB	NM
NM	PM	PM	PS	NM	NS	NM	NM
NS	PS	PS	PS	NS	NS	NS	NS
ZE	ZE	ZE	ZE	ZE	ZE	ZE	ZE
PS	NS	NS	NS	PS	PS	PS	PS
PM	NM	NM	NS	PM	PS	PS	PM
PB	NM	NB	NM	PB	PM	PB	PM

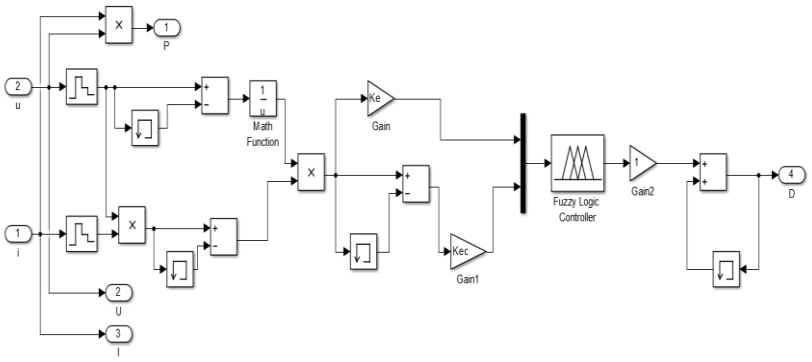


Fig. 11 The model of MPPT with fuzzy controller

The core of the fuzzy controller is the fuzzy control rules. This article establishes a fuzzy rule table as shown in the **Table 1**. These rules are based on when the change of *E* and *EC* causes the output power to change, according to the fuzzy rule table, the output power returns to the MPP. The establishment of the fuzzy control table is mainly based on experience, theoretical analysis, and expert experience, among which expert experience is the leading source.

In this paper, the Mamdani model is selected as the fuzzy controller model, the input membership function chooses the triangle, and the output membership function chooses the Gaussian type, because the triangle and the Gaussian membership function have good adaptability, and the operation is convenient and the performance is familiar. The input variable selects the triangular membership function, and the output selects the Gaussian membership function because of the convenience of calculation, which can improve the resolution, so that the accuracy becomes higher and the error is reduced. The method clear to fuzzy conversion chooses the larger and the smaller synthesis method, because the calculation is small and the calculation speed is fast, the area-gravity method is used for the method of blur to clear conversion, because although the calculation is larger, it is intuitive and reasonable, and it is well-founded, more accurate.

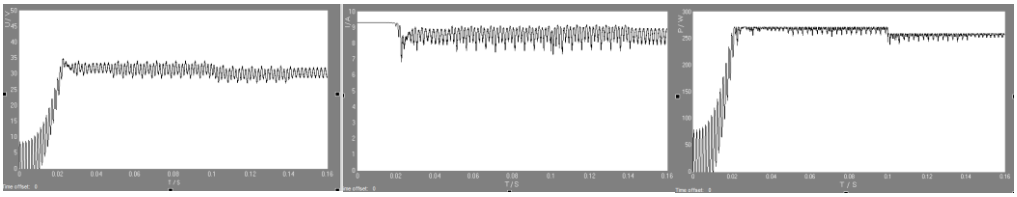


Fig. 12. The simulation of P&O

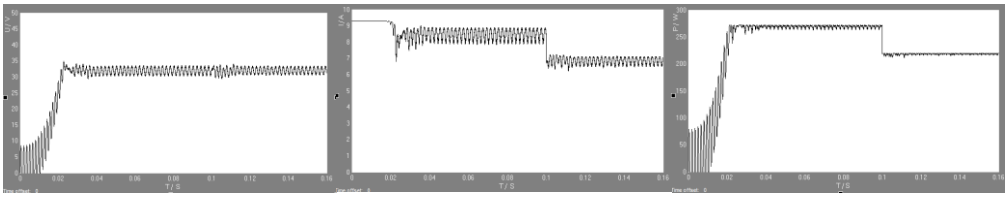


Fig. 13. The simulation of FLC

Build a model base on these parameters and math formula shown in **Figure 11**.

In order to compare the optimization ability of the fuzzy control method for the system, the simulation model is used for simulation. Set the same simulation time to 0.16s, keep the temperature constant at 25, and the initial light intensity to 1000, which will decrease to 800 at 0.1s for simulation. The **Figure 12** shows the simulation of voltage, current and power of P&O and **Figure 13** shows the simulation of voltage, current and power of FLC (Fuzzy logic control).

The simulation results are shown in the figure. Two methods to track to the maximum power point. The voltage fluctuation of the perturbation and observation tracking method is 5 volts, the current fluctuation is 1.8 A, and it takes 0.022 seconds to regain the stable state; the Fuzzy logic control tracking method, the voltage fluctuation is 3 volts, the current fluctuation is 1 A, and it takes 0.005 seconds to regain the stable state.

5 Conclusion

This article uses the same photovoltaic model and only changes the control strategy for simulation. It can be seen from the simulation results that both methods can finally reach the maximum power point, indicating that the two algorithms meet the functional requirements, but the performance is quite different. The results show that the output current, voltage and power using the perturbation and observation method will fluctuate after being stabilized, and it will be slow to reach a stable state. This is caused by its own structure, but the output current, voltage and power obtained by the fuzzy control method although there are fluctuations, the fluctuations are relatively small, that is, the so-called smoothness is good, and the speed to reach the optimal solution is also shorter, which shows that the tacking speed is good. When encountering a sudden situation, in this article, which is, when the external conditions suddenly change, it takes longer for the perturbation and observation tracking method to reach a stable state again than using the fuzzy logic control method, so the anti-interference ability of the fuzzy control method is much stronger.

It can be concluded that fuzzy control is more suitable for nonlinear time-varying systems. Considering the actual project, the external conditions of the photovoltaic system are changing

all the time, so the fuzzy control method is more suitable for actual projects. Although the fuzzy logic control system has the above advantages, the design of a fuzzy controller is based on trial-and-observation procedures and there is still no analytical tool to analyze the stability of the fuzzy control system. And the actual cost of the fuzzy control method is relatively higher, which is a problem that needs to be solved in the future. These factors hinder the engineering application of fuzzy control.

Acknowledgments

Authors acknowledge the financial support of Dalian Science and Technology Innovation Fund (Grant: 2020JJ26GX029) and would like to express many thanks the support of Dalian Key Laboratory of Smart Micro-grid and Green Recycling Industry.

References

[1] Ahmed J., Salam Z.: An accurate method for MPPT to detect the partial shading occurrence in a PV system. IEEE Trans And Inform. Vol. 13. pp. 2151–2161 (2017)

[2] Patel, A (Patel, Amit)., Tiwari, H (Tiwari, Harpal).: Implementation of INC-PI MPPT and Its comparison with INC MPPT by Direct Duty Cycle Control for Solar Photovoltaic Employing Zeta Converter. International Conference on Information, Communication, Instrumentation and Control. pp. 1-6(2017)

[3] Jain, K (Jain, Kriti)., Gupta., M (Gupta, Manju)., Bohre., AK (Bohre, Aashish Kumar).: Implementation and Comparative Analysis of P&O and INC MPPT Method for PV System. India International Conference on Power Electronics. pp. 1-6(2018)

[4] A Rajavel., N Rathina Prabha.: Fuzzy logic controller-based boost and buck-boost converter for maximum power point tracking in solar system. Transactions of the Institute of Measurement and Control. pp. 1–13(2020)

[5] Farayola., AM (Farayola, Adedayo M.).,Sun, YX (Sun, Yanxia) ., Ali A (Ali, Ahmed).: ANN-PSO Optimization of PV Systems Under Different Weather Conditions. International Conference on Renewable Energy Research and Applications. pp. 1363-1368 (2018)

[6] Fengcong Lan., Wenjie Li., Zhijie Li.: Deformation response of squeeze load and internal short circuit failure analysis of vehicle power battery, Journal of South China University of Technology. Vol. 46. pp. 65-72(2018)

[7] Huifeng Li., Jing Pang., Shigang Lu.: Research on the safety of lithium-ion batteries under abuse conditions. Power technology. Vol. 37(12). pp. 2235-2238(2013)

[8] He Li., Shenju Yu., Zhikui Chen.: Reactor for internal short-circuit failure of lithium ion battery Scientific research. Electrochemistry. Vol. 16(2). pp. 185-192(2010)

[9] Bruno W França.， Mauricio Aredes.，MAYNARA Aredes.: Fuzzy adaptive P&O control for MPPT of a photovoltaic module. Journal of power & energy engineering. Vol. 2 (4). 120-129(2014)

[10] Huan-Liang., T.: Insolation-oriented model of photovoltaic module using MATLAB/Simulink. Solar Energy. Vol. 84(7). pp. 1318-1326(2010)

[11] Walker., G.: Evaluating MPPT converter topologies using a MATLAB PV model. Journal of Electrical & Electronics Engineering, Australia. Vol. 1(1). pp. 49-56(2001)

[12] Umbers and P. King.: Analysis of human decision-making in cement kiln control and the implication for automation. Internet Man-Machine Studies. Vol. 12. pp. 9-64(1976)

[13] P.C. Chen and M.C. Shih.: An experimental study on the position control of a hydraulic cylinder using a fuzzy logic controller. JSME Series. Vol. 34(4). pp. 481-489(1991)

[14] Patel R., Panda AK.: Real time implementation of PI and fuzzy logic controller based 3-phase 4-wire interleaved buck active power filter for mitigation of harmonics with id–iq control strategy. International Journal of Electrical Power & Energy Systems. Vol. 59. pp. 66–78(2014)

[15] F.Bouchafal., D.Beriberl, M.S.Boucherit.: Modeling and simulation of a gird connected PV generation system With MPPT fuzzy logic control. International Multi-conference on

Systems, Signals & Devices. pp. 1-7(2010)
[16] Dorin Petreus., Toma Pa tara u., ete.: A novel maximum power point tracker based on analog and digital control loops. Solar Energy. pp. 588-600(2010)
[17] Piegari L., Rizzo R.: Adaptive perturb and observe algorithm for photovoltaic maximum power point tracking. Renewable Power Generation. Vol. 4(4). pp. 317-328(2014)
[18] Zhao Y., Zhao X., Zhang Y.: MPPT for photovoltaic system using multi-objective improved particle swarm optimization algorithm. TELKOMNIKA Indonesia Journal of Electric Engineering. Vol. 12(1). pp. 261-268(2014)
[19] Yang P.，He T.，Xu Z R.，et al.: A novel region partition MPPT method based on variable step-size INC. IEEE Innovative Smart Grid Technologies - Asia. Melbourne. IEEE. pp. 299-704(2016)
[20] Patcharaprakiti N., Premrudeepreechacharn S., Sriuthaisiriwong Y.: Maximum power point tracking using adaptive fuzzy logic control for grid-connected photovoltaic system. Renewable Energy. Vol. 30(11). pp. 1771-1788(2015)
[21] Ze Cheng., Hongzhi Yang., Yanli Liu.: Self-adjusting Fuzzy MPPT PV System Control by FPGA Design. IEEE. pp. 1-4(2015)

Study on low-power Sensor and Heater for PCR microchip

Wenbo He*, Jinyu Liu[0000-0003-0233-3493], Lei Fan[0000-0003-4654-6511]

School of Information Science and Engineering, Dalian Polytechnic University, Dalian 116034, China
hewb@dlpu.edu.cn

Abstract. Low-power temperature sensor and heater were designed and fabricated by using chromium, nickel chromium alloy (Ni: Cr = 80%: 20%) and nickel materials respectively. Through the analysis of the working characteristics of the Low-power temperature sensor and heater, it is shown that the Ni/Cr film resistance heater and the nickel film resistance temperature sensor can meet the requirements of the heating rate and temperature consistency of the reaction chamber for PCR amplification.

Keywords: PCR microchip, Low-power temperature Sensor, Low-power heater, micro-fabrication, film-resistance.

1 Introduction

In 1985, K. B. Mullis and R. K. Saiki of Cetus Company, USA, invented a specific DNA amplification technique in vitro. The essence of this technique is the enzymatic synthesis of specific DNA fragments in vitro. This technique is called Polymerase Chain Reaction, or PCR for short. PCR is an enzymatic synthesis reaction dependent on DNA polymerase in the presence of polymerase, primers and four deoxyribonucleotides. It works through three steps: high temperature denaturation of double-stranded DNA (92-94℃), low temperature annealing of primers and templates (37-55℃), and primer extension under suitable temperature conditions (68-72℃). Through repeated cycles of these three steps, PCR can achieve exponential amplification of specific DNA sequence yield with the number of cycles, thus achieving the goal of rapid and large-scale amplification of template DNA [1-2]. Not long after the advent of PCR amplification technology, it quickly became the basic technology and powerful tool of molecular biology research because of its simple, fast and efficient characteristics, and has been widely used in medicine and biology.

Subsequent The traditional PCR chip amplification reaction vessel and heater are separated [3], the volume is huge, the operation is inconvenient, and the sample consumption is large. Therefore, the traditional PCR chip can not meet the needs of modern biochemical research. In addition, mature thin-film technology and microfabrication technology make it possible to integrate reaction vessel, heater, temperature sensor and other components into a chip. PCR biochip is an integrated PCR microreactor,

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2306783

which is made on silicon, glass and organic polymers by using the technology of micro-electro-mechanical system (MEMS) [4]. It can control the temperature of reaction zone accurately by using heaters and temperature sensors integrated on the chip. It fully embodies the advantages of microchip, such as high efficiency, rapidity and saving reagents.

In this paper, the characteristics of a self-developed low-power temperature sensor and low-power heater for integrated PCR chip are analyzed.

2 Development of Low-power temperature Sensor and heater

There are two main types of PCR microchip fabricated by MEMS technology: the first is the single reaction zone structure, in which the whole process of amplification is carried out in one reaction chamber, and the continuous cycle of PCR amplification is realized by periodically changing the temperature in the reaction cell. The second is the three reaction zone structure, i.e. the amplification reaction is carried out in three different reaction zones, each of which maintains a constant temperature. Under the action of external forces, the sample flows through the channel on the chip in three reaction zones to realize the transformation of the sample in three temperature zones, thus realizing the cycle of PCR. In single reaction region PCR chips, the rate of temperature change may affect the amplification effect, but there is no such problem on three reaction region PCR chips. However, the fabrication process of three reaction zone PCR chip is much more complicated than that of single reaction zone PCR chip. How to achieve thermal isolation between different temperature zones and how to accurately control the temperature of the reaction zone are the problems that must be solved by the three reaction zone PCR chip.

On the basis of consulting the literature [5-6], according to the working requirement of the integrated PCR chip and the problems faced in the current research, we have developed a single reaction region integrated PCR chip. Fig.1 shows a schematic diagram of an integrated PCR chip.

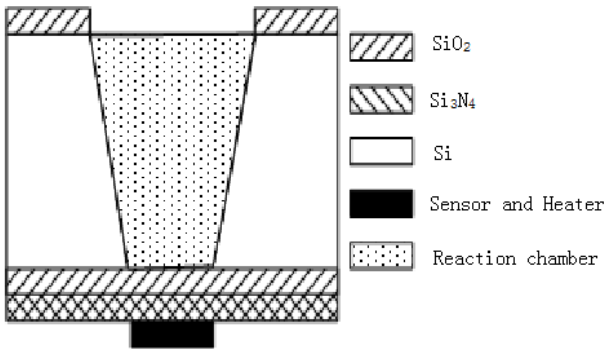


Fig. 1. Schematic structure of integrated PCR chip.

As the main device for DNA sample amplification, the PCR chip must be able to meet the thermal uniformity, the accuracy of temperature control and the accuracy of physical size. Therefore, the material selection and structure design of low-power temperature sensor and heater are important factors to determine the performance of the PCR chip.

2.1 Material selection of Low power temperature Sensor and heater

At present, the most suitable material for making thin film resistance temperature sensor and heater is platinum, which has good heating and temperature characteristics [7]. But because platinum is a precious metal, the cost is too high, so this paper hopes to use some cheaper metals to make temperature sensor and heater.

Based on the literature [8-10], it is found that chromium and nickel-chromium alloys are good heating materials, and nickel is a good temperature sensor material. In this paper, we hope to use the same material to make temperature sensor and heater, because the reduction of process means the increase of chip yield. Therefore, chromium (Cr), nickel-chromium alloy (Ni: Cr=80%:20%) and nickel (Ni) are used to make heaters and temperature sensors respectively.

2.2 Design of Low-power temperature Sensor and heater

In this paper, four independent PCR reaction chambers are etched on a single crystal silicon of type N with a crystal orientation of [100] and a size of $20\times15mm^2$ at one time.

In order to improve heating efficiency and ensure accurate temperature control, the area of Low-power heater at the bottom of the reaction chamber should be enlarged as much as possible so that the reaction chamber can be heated uniformly. Meanwhile, the Low-power temperature sensor should be placed in the middle of the reaction chamber bottom.

Fig. 2 shows the structure of the low-power temperature Sensor and heater designed in this paper. The longer resistance is the heater and the shorter resistance is the temperature sensor. The line width of the thin film resistance is 50μm, the distance between the lines is 100μm, and the area of the welding pad at the top of the resistance bar is $200\times200\mu m^2$.

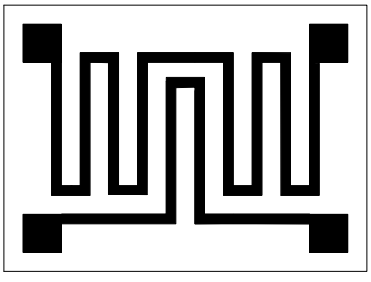


Fig. 2. Structure of Low-power temperature Sensor and heater.

2.3 Fabrication of Low-power temperature Sensor and heater

The fabrication of low-power temperature Sensor and heater is combined with the fabrication of PCR chip. Based on a large number of references and previous work experience in the laboratory, wet etching and sputtering process is adopted. The process flow is as follows:

According to the design scheme, the required photolithographic mask are obtained by drawing graphics on coordinate paper, carving red film, exposure, development and stereotyping.

SiO_2 thin films with thickness of about 1μm are grown on both sides of single crystal silicon by thermal oxidation method, and then a Si_3N_4 thin film with thickness of about 1840 nm is deposited on one side of SiO_2 by LPCVD method as an insulating layer.

On the side without Si_3N_4 film, the window of PCR reaction chamber is engraved on SiO_2 using photolithographic mask.

Using SiO_2 as mask and KOH solution as anisotropic etchant, the reaction chamber of PCR is etched on silicon wafer.

Because of the incompatibility between silicon and PCR, a thin SiO_2 film is grown on the inner wall of the reaction chamber by thermal oxidation.

At the bottom of the reaction chamber, a thin film heater and a thin film temperature sensor are fabricated as follows: first, a thin metal film is sputtered on Si_3N_4 by vacuum sputtering method, then the metal film is lithographed by photolithographic mask, then corroded by corrosive solution, and finally annealed in a nitrogen-filled environment at 400℃ for 30 minutes.

Pressure solder joint: a layer of metal aluminum is grown on the metal layer by vacuum evaporation coating machine, photolithographic mask is used to photolithograph it, then corrode it with corrosive solution. Finally, it is placed in a nitrogen-filled environment at 500℃ and alloyed for 10-15 minutes to ensure good ohmic contact between silicon and metal.

The alloy film is bonded by hot pressing with a gold wire ball welding machine to connect the external lead.

3 Characteristic analysis of Low-power temperature Sensor and heater

According to the application requirements, it is hoped that the thin film resistance heater has high heating efficiency, and the resistance of the thin film temperature sensor varies sensitively with the change of the external environment. The working characteristics of the low-power temperature sensor and the low-power heater made of three materials will be analyzed separately.

3.1 Analysis of temperature characteristics

According to the temperature requirement of PCR amplification reaction, the analysis range of heater and temperature sensor is chosen in the range of 30 ~ 110℃.

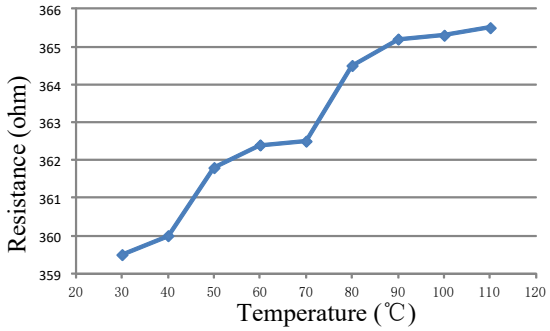


Fig. 3: Temperature Characteristic of integrated Cr resistance.

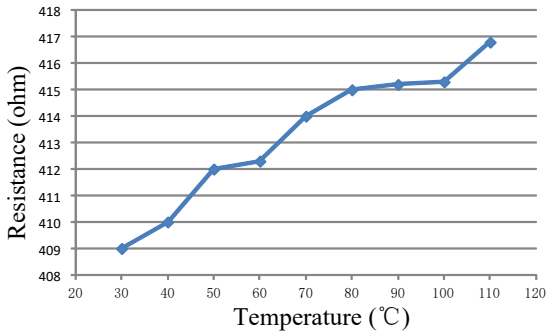


Fig. 4: Temperature Characteristic of integrated Cr resistance.

Fig. 3 and Fig. 4 show the temperature characteristic curves of two Cr film resistances. The abscissa is the test temperature and the ordinate is the resistance value. It can be seen from the figure that the Cr film resistance has no linear temperature characteristic in the temperature range of 30-110℃, and is insensitive to temperature. The resistance value changes within several ohms with the change of 10℃, so the chromium film resistance can not be used as temperature sensor and heater.

Fig. 5 and Fig.6 show the temperature characteristic curves of the two nickel thin film resistances. The abscissa is the test temperature and the ordinate is the resistance value. It can be seen that the resistances of nickel film has good linear temperature characteristics in the temperature range of 30-110℃, and is very sensitive to temperature. The variation of resistor is tens of ohms for every 10℃ change of temperature. This shows that the resistance of nickel film is a good temperature sensor, but it is too sensitive to temperature, therefore, it can not be used as a heater.

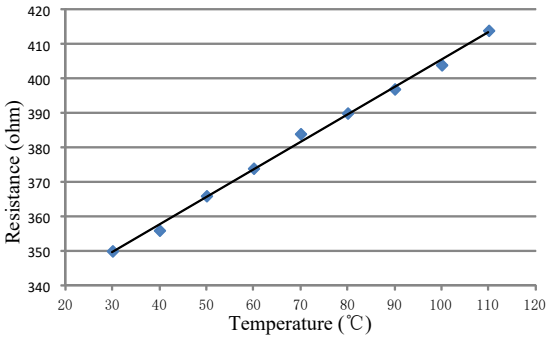


Fig. 5. Temperature Characteristic of integrated Ni resistance.

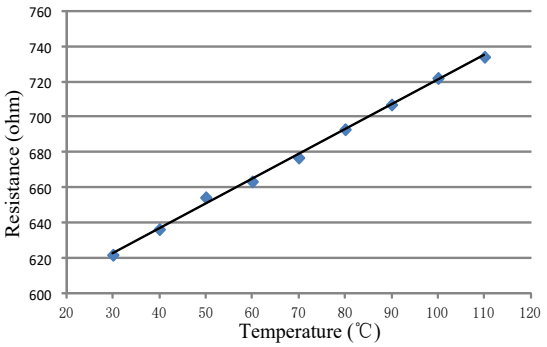


Fig. 6. Temperature Characteristic of integrated Ni resistance.

The temperature characteristic curve of Ni-Cr alloy film resistance is no longer given here as its resistance value hardly varies with temperature. The values of two thin film resistances are 1100 ohms and 428 ohms at 20℃ respectively. The values of two thin film resistances are 1102 ohms and 430 ohms at 100℃ respectively. This shows Ni-Cr alloy is a good heating material, but it can not be used to make temperature sensor.

3.2 Analysis of Heating Characteristics

In order to verify whether the nickel chromium alloy thin film resistance heater can meet the needs of PCR amplification reaction, the heating effect of its PCR chip is tested by micro thermocouple.

Fig. 7 shows the Heating characteristic of the integrated Ni/Cr resistance. The resistance value tested is 1281 ohms. The abscissa is the applied voltage, and the ordinate is the highest temperature that can be reached in the reaction chamber of the PCR. It can be seen from the figure that the heating effect is very poor when the applied voltage is low. However, when the applied voltage exceeds a certain critical

value, the heating effect varies greatly, and the maximum temperature in the reaction chamber increases approximately linearly with the increase of the applied voltage. When the applied voltage is 48V, the highest temperature in the reaction chamber reaches 116℃, while the temperature required for PCR amplification is below 100℃, which indicates that the heater can meet the requirements of the temperature of PCR amplification.

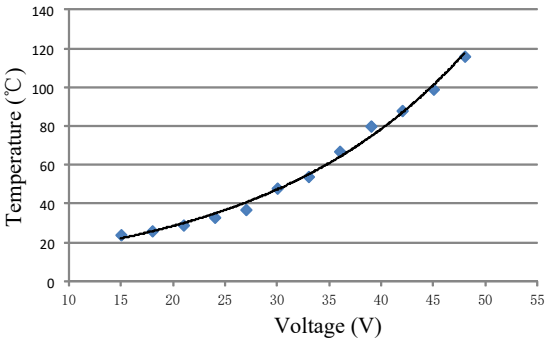


Fig. 7. Heating Characteristic of Ni/Cr resistance.

The amplification reaction of PCR is repeated in three temperature zones, so the heating speed of single reaction zone PCR chip is very high, and generally requires heating speed not less than 0.5℃/s. In order to verify whether the resistance of nickel-chromium alloy thin film can meet the requirement of heating rate for PCR amplification reaction, the heating states of two thin film resistors are tested. Fig. 8 and Fig. 9 show temperature-time characteristic curves of thin film resistors with resistance values of 1281 ohms and 504 ohms respectively. The abscissa is the heating time and the ordinate is the temperature reached.

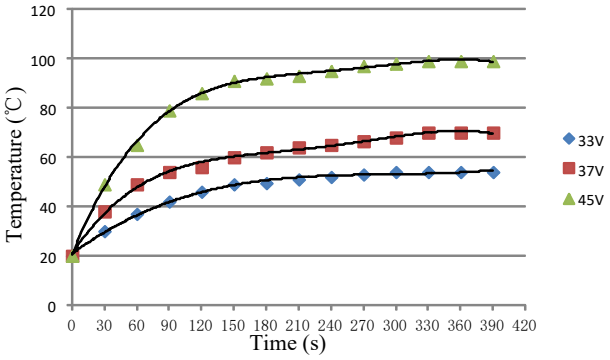


Fig. 8. Heat characteristic of Ni/Cr resistance.

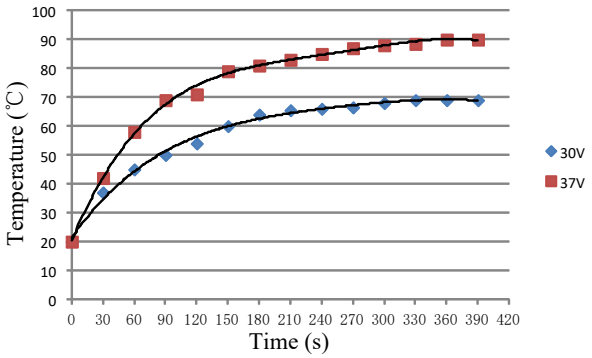


Fig. 9. Heat characteristic of Ni/Cr resistance.

As can be seen from Fig. 8 and Fig. 9, the heating characteristics of the two heaters are comparatively consistent. The temperature rises rapidly in the first 30 seconds. After 30 seconds, the temperature rises slowly. After 90 seconds, the temperature changes more slowly, and gradually tends to a constant temperature. At the same time, with the increase of the applied voltage, the maximum temperature of the reaction chamber can also change significantly. In addition, when the applied voltage is the same, the heating effect of the heater with small resistance is obviously better than that of the heater with large resistance, and the stable temperature can be achieved is also high. For example, when the applied voltage is 37V, the final constant temperature of small resistance is 94℃, while the final constant temperature of large resistance is 74℃. Within 30 seconds, the average temperature change speed of small resistance is 0.75℃/s, while that of large resistance is 0.5℃/s. This also provides a train of thought for future research, which means we can use multiple small resistances to replace the large snake resistance.

Based on the above analysis, the Ni-Cr alloy film resistance heater can meet the needs of the temperature change of the PCR amplification reaction, and can be used as the heater of the PCR microchip.

3.3 Analysis of Temperature Uniformity

In order to test the temperature uniformity in the reaction chamber, the temperature is measured by sticking four micro thermocouples to the bottom of the reaction chamber. The temperature-time characteristics of the four test points are shown in Fig. 10. The film resistance value used is 1281 ohms, and the applied voltage is 45V. The abscissa is the heating time and the ordinate is the temperature reached.

From Fig.10, it can be seen that the temperature uniformity in the reaction chamber is very good, and the temperature of each point is comparatively consistent. This shows that the heater designed in this paper is reasonable and can meet the requirements of temperature uniformity for PCR amplification reaction.

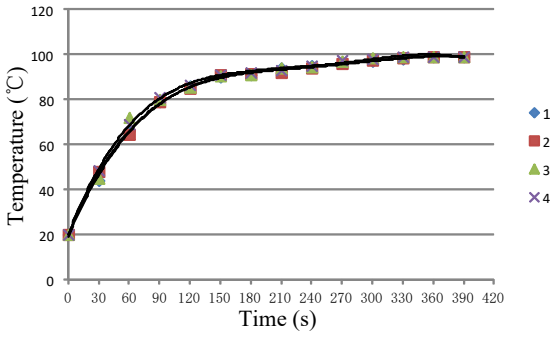


Fig. 10. Temperature characteristics of the four test points.

4 Conclusions

In this paper, a single reaction zone integrated PCR chip is designed and fabricated. The low-power temperature Sensor and heater are made of chromium, Ni-Cr alloy (Ni: Cr = 80%: 20%) and nickel respectively. The low-power temperature Sensor and heater are integrated at the bottom of the PCR reaction chamber. The low-power temperature Sensor and heater depend on the process to some extent. The resistance deviation and consistency can be further controlled by optimizing the design parameters and process parameters. The results show that the nickel-chromium alloy film resistance produced in this paper can meet the requirements of heating speed and temperature consistency of PCR reaction chamber, and can be used as heater of PCR microchip. Nickel film resistance can be used as temperature sensor of PCR microchip.

In the process of chip fabrication, using different materials to make heaters and temperature sensors means the process is more troublesome and material consumption is greater. Therefore, we still hope to find the same material to fabricate heaters and temperature sensors.

References

1. Giordano, B. C. J., Swedberg, S., Huhmer, A. F. R., Landers, J. P.: Polymerase Chain Reaction in Polymeric Microchips: DNA Amplification in Less Than 240 Seconds, Analytical Biochemistry, vol. 291, pp. 124-132(2001)
2. Cheng, J. , Shoffner, M. A., Hvichia, G. E.: Investigation of different PCR amplification system in microfabricated silicon-glass chips, Nucleic Acids Research, 24(2): 380-385(1996)
3. Nagai H., Murakami, Yuji, Yokoyama K., Tamiya E, : High-throughput PCR in silicon based microchamber array, Biosensors and Bioelectronics, 16 (9-12): 1015-1019(2001)
4. Sun, K., Yamaguchi, A., Ishida, Y.: A heater-integrated transparent microchannel chip for continuous-flow PCR [J], Sensors and Actuators: B, 87(2/3): 612-619 (2002)

5. Yu. C. Y., Liang. W. S.: Fabrication and characterization of a flow-through PCR device with integrated chromium resistive heaters[J].Journal of the Chinese Institute of Chemical Engineers(3/4), 333-339(2007)
6. Lai, L. F., Sun, R.: Processing technology of embedded thin-film resistor materials.In:2011 International Symposium on APM, pp.60-65.XiaMen(2011)
7. Lai, L. F., Su, X.S., Fu, X. Z., Sun, R.: The microstructure and properties of C and W co-doped Ni/Cr embedded thin film resistors, Surface Coat. Technol, vol. 259, pp. 759-766(2014)
8. Yan, J. W., Zhou, J. ch.: The oxidation and the electrical properties of Ni-Cr thin film after rapid thermal annealing, Int. J. Mod Phys B, Vol. 21, No. 26, pp. 4561-4574(2007)
9. Danisman, M., Cansever, N.: Effect of Cr content on mechanical and electrical properties of Ni-Cr thin film, J. Alloys Compd, vol. 493, pp. 649-653(2010)

A Novel Solar Glass Window with Photovoltaic Power Generation and Automatic Temperature Regulation and Dust Removal Functions

Teng Zhang[1], Jinpeng Wang[1,*], Yang Zhou[1], Xin Guan[1], Wei Li[1], Xin Zhao[1]

[1]Dalian Polytechnic University, Dalian 116034, China
wangjp@dlpu.edu.cn

Abstract. Solar energy is a new clean energy with many advantages compared with traditional energy, and the range of its application is becoming wider. This paper aims to combine the power technology based on solar energy with windows, so as to alleviate stuffiness indoors in summer and save energy. For this purpose, a novel solar glass window is designed, which reduces indoor temperature through air flow, a solar energy panel is installed in the window frame to provide power to the indoor air conditioner used for auxiliary cooling, and the air-conditioning switch is automatically controlled according to room temperature, dust on the solar energy panel is removed according to dust deposition, so as to improve the photoelectric conversion efficiency.

Keywords: Photovoltaic glass, Solar power supply, Intelligent temperature control, Automatic dust removal

1 Introduction

In society today, energy has become a hot topic. Fossil fuel, as a traditional non-renewable energy source, is still one of the major energies on which we depend. Its excessive consumption not only accelerates the depletion of non-renewable energy, but also aggravates environmental pollution. Nevertheless, building energy is an important portion of total energy, accounting for about 40% [1], which consumes a large part of fossil energy. This problem shall never be ignored.

Compared with non-renewable energy, solar energy is advantageously characterized in pollution free, large reserve and long-lasting usage, etc. and a main way of generating power from solar energy is photovoltaic (PV) power generation. By combining windows with PV power generation, part of the energy crisis can be alleviated. In Japan, the government implemented “Sunshine Project” and by the same built several typical solar heating and air conditioner test buildings, such as Yazaki test solar house [2]. Kai Li established the overall relative power loss model of non-uniform width bus grid solar cells and found that when the number of bus grid lines is 2-6, compared with the typical constant width grid line structure solar cells, the photovoltaic Conversion efficiency increased by 0.10%, 0.09%, 0.10%, 0.09% and 0.09% respectively [3]. Sonveer carried out an optimized analysis on the

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2306766

thermoelectric properties of translucent PV glass with hybrid double channels by using the fuzzy genetic algorithm, the power generation efficiency was improved by 12.21%, and the mean temperature of PV glass was decreased by 2.28℃ [4]. Antika team analyzed the relationship between temperature and power generation efficiency by combining experiments and models [5]. Ye Huang built an experimental platform for the Concentrated Photovoltaic Photothermal (CPV/T) system using dual-axis tracking, and found that the average photoelectric conversion efficiency of the system reached 13.0% [6]. Through experiment, Xiaoli WU analyzed the influence of the temperature of hollow PV glass component on its electrical property. The results indicated that the negative effect of temperature increase on conversion rate shall be taken into consideration when the outdoor temperature is high [7].

An important parameter which can denote photovoltaic power generation is the photoelectric conversion rate. The papers by the above and others has contributed to the field of using sunlight for power generation, and different methods have been used to improve the photoelectric conversion rate. However, these papers did not consider the form of photovoltaic power generation combining multiple functions, and did not provide an exact solution for dust, an important factor affecting the photoelectric conversion rate. Compared with uncomplicated power generation function, the novel solar glass window in this paper realizes intelligent temperature control while supplying power to indoor air conditioner, and automatically control the turn-on and turn-off of cooling equipment according to the indoor temperature. However, the dust accumulated on solar energy panel can reflect, scatter and absorb solar radiation to decrease the transmittance of solar radiation, so the output power decreases accordingly, which severely affects its photoelectric conversion efficiency. In order to solve the problem of low photoelectric conversion efficiency due to dust accumulation, this solar glass window is equipped with photoelectric sensor and dust removal device, which enables functions of automatic dust removal according to the dust accumulation on the solar energy panel.

2 System overview

This design first installs the suitable size solar panel on the window frame, absorbs the solar energy to supply the indoor power. Different from the single-layer glass used for Windows of most buildings, this design adopts double-layer glass structure, and combines the flexible thin film solar cells with semi-transparent nature on the outer glass [8-9]. It is also used to generate electricity for indoor air conditioning without affecting indoor lighting requirements. Some small holes are drilled in the upper and lower parts respectively to facilitate air flow in and out. The inner glass is made of ordinary glass. Photoelectric sensor and dust removal device are installed on the outside of the window frame. According to the degree of ash accumulation, the photoelectric sensor sends electrical signals to the MSP430 single-chip microcomputer after amplification and shaping, this single chip microcomputer controls the operation of dust removal device to realize automatic dust removal. The

elevation view and side view profile of the solar glass window are shown in Fig. 1. and Fig. 2. respectively.

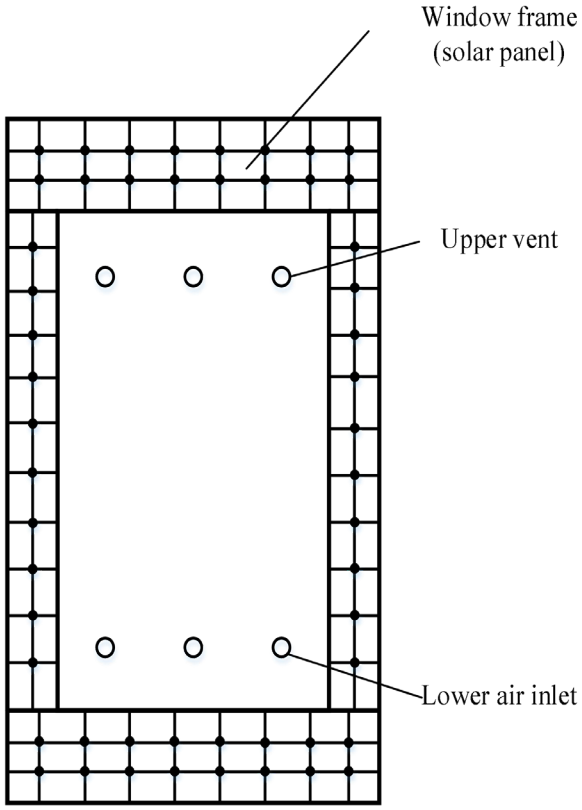


Fig. 1. Elevation view

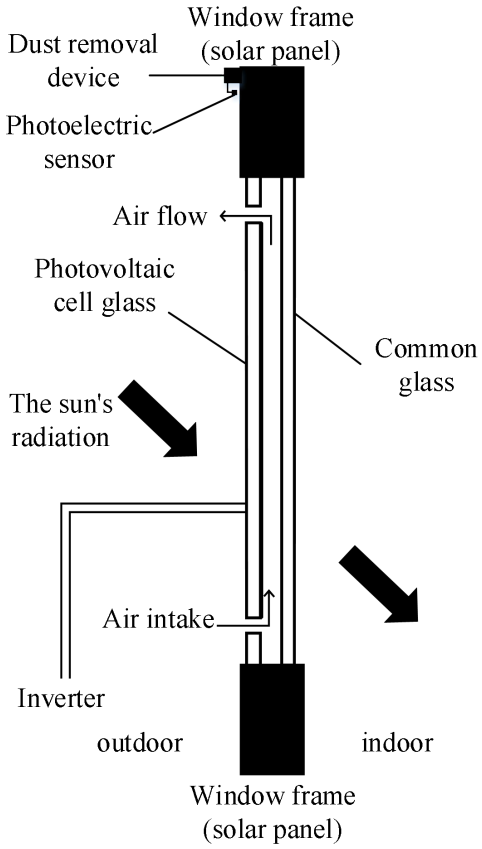


Fig. 2. Side view profile

3 Functional principle

3.1 Reduce the room temperature

The outdoor air enters the area between the two layers of glass through the air inlet under the outer glass. When the indoor temperature is too high, the indoor heat will be transferred to the air in the middle area through the ordinary glass in the inner layer, a flexible thin-film solar cell attached to the outer glass also generates heat from the sun, which is also transferred to the air in the middle layer, the heated air becomes denser and rises, passing out through the upper vent of the outer glass, in a process that reduces the room's temperature.

3.2 Photovoltaic power generation

The composition of photovoltaic power generation system needs to take into account generation, storage, transmission and automatic control. In the new solar glass window designed in this paper, the power generation system is a flexible thin-film solar cell on the outer glass and a solar panel installed on the window frame. They can convert the solar energy shining on its surface into electricity for indoor cooling equipment, such as air conditioning, The storage system consists of a photovoltaic controller and a battery bank. The battery bank is composed of a single battery in series and in parallel. Its function is to store the electric energy converted into light energy for overall deployment, and the unit is relatively independent. The function of the photovoltaic controller is to control the charging process of the battery, let the photovoltaic array charge the battery, and carry out overcharge and overdischarge protection on the battery, so as to protect the normal use of the battery and ensure the normal operation of the system [10]. While the electricity converted from solar energy is direct current, the air conditioner is powered by alternating current. In order to make the air conditioner work normally, the inverter is needed to change the direct current in the battery into alternating current. In order to optimize the utilization of the whole photovoltaic power generation system, it is very important to calculate the conversion efficiency of the inverter and select the inverter with suitable specifications.

The conversion efficiency of the inverter is the ratio of output power to input power. The input power of the inverter is calculated as follows:

$$P_{DC} = I_{DC} \times V_{DC} \tag{1}$$

Among them, P_{DC} is direct current power, I_{DC} and V_{DC} are DC current value and DC voltage value respectively, which are generated by photovoltaic cells. According to Kirchhoff's law (KCL), the current I_{DC} generated by the photovoltaic cell can be obtained as:

$$I_{DC} = I_{ph} - I_D - I_{sh} \tag{2}$$

Among them, I_{ph} is the photo-generated current and I_D is the current generated by the P-N junction of the photovoltaic cell itself, which can be further calculated to obtain:

$$I_{DC} = I_{ph} - I_o[e^{\frac{q(V_{DC}+I_{DC}R_s)}{AkT}} - 1] - \frac{V_{DC}+I_{DC}R_s}{R_{sh}} \tag{3}$$

Among them, I_o is the reverse saturation current, q is the charge of electrons, A is the quality factor of the diode, T is the temperature of the photovoltaic cell, and k is the Boltzmann constant, whose value is about 1.3806×10^{-23}J/K.

In the case of an open circuit, due to the large resistance of the parallel parasitic resistance R_{sh} in the photovoltaic cell, the open circuit voltage can be obtained as:

$$V_{OC} \approx \frac{AkT}{q} \ln\left(\frac{I_{ph}}{I_0} + 1\right) \tag{4}$$

In practical applications, I_{sh} and I_D are generally ignored, so that I_{ph} is approximately equal to I_{sc}. In the case of an open circuit, the output voltage of the photovoltaic cell $V_{DC}=V_{OC}$, so:

$$I_{DC} = I_{sc}(1 - C_1(e^{\frac{V_{DC}}{C_2 V_{OC}}} - 1)) \tag{5}$$

$$C_1 = (1 - \frac{I_m}{I_{sc}})e^{(-\frac{V_m}{C_2 V_{OC}})} \tag{6}$$

$$C_2 = (\frac{V_m}{V_{OC}} - 1)[\ln(1 - \frac{I_m}{I_{sc}})]^{-1} \tag{7}$$

Among them, I_m and V_m are the output current and output voltage of the photovoltaic cell operating at maximum power. The output voltage of the photovoltaic cell is affected by the sunlight intensity and the external environment temperature. Taking into account the changes of these two points, the temperature difference ΔT and the light intensity difference ΔS can be set as:

$$\Delta T = T - T_{ref} \tag{8}$$

$$\Delta S = \frac{S}{S_{ref}} - 1 \tag{9}$$

Among them, T_{ref} is the ambient temperature under standard conditions, which is 25° C, and S_{ref} is the sunlight intensity under standard conditions, which is 1000W/m^2.

Finally, the values of C_1 and C_2 can be determined and obtained:

$$I_{DC} = I_{sc} \frac{S}{S_{ref}}(1 + a\Delta T) \tag{10}$$

$$V_{DC} = V_{OC}(1 + c\Delta T \ln(1 + b\Delta S)) \tag{11}$$

Among them, a, b, and c are correction coefficients, and their values are 0.00225, 0.5, and 0.00288, respectively [11].

The output power of the inverter can be calculated as follows:

$$P_{AC} = I_{AC} \times V_{AC} \tag{12}$$

Among them, P_{AC} is AC electric power, I_{AC} and V_{AC} are AC current value and AC voltage value respectively, which are obtained by the inverter. Therefore, the conversion efficiency of the inverter can be calculated as follows:

$$\eta = P_{AC} / P_{DC} \tag{13}$$

3.3 Intelligent temperature control

When sunlight hits the outer glass attached to a flexible thin-film solar cell, the energy is divided into four parts. Some of the energy enters the room through the outer glass, Some of the energy is reflected back into the environment;Some of the energy is absorbed by photovoltaic cells and converted into electricity, The last part of the energy is converted into the thermal energy generated when the photovoltaic cell works [12]. The energy analysis on the outer glass is shown in Fig. 3.

Due to the inner layer of glass and the use of ordinary glass, through the outer layer of the sun to the inner glass on the surface of most of the energy will return through reflection and transmission effect to the surrounding environment, only a small percentage of the energy absorbed by photovoltaic cells and by convection and radiation heat transfer back to the surrounding environment, without considering the energy loss caused by reflection and transmission, the inner glass of heat exchange with the environment as shown in Fig. 4.

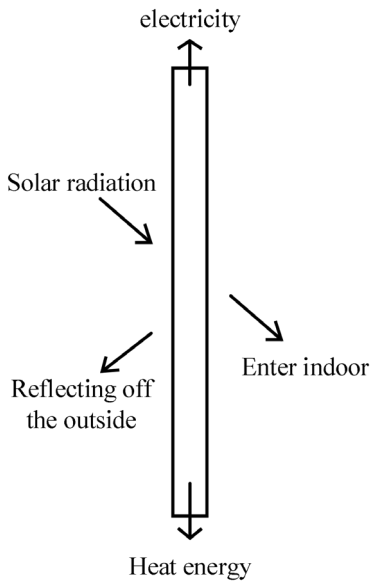


Fig. 3. Energy analysis diagram of outer glass

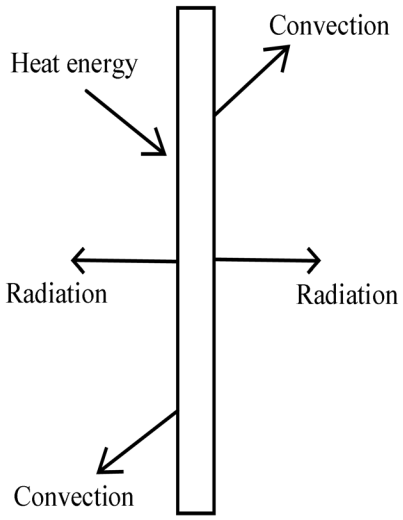


Fig. 4. Heat exchange diagram of inner glass with surrounding environment

The new type of solar glass window designed in this paper has limited ability to cool the indoor temperature through air circulation between two layers of glass. In hot summer, when the indoor temperature is too high, people need to turn on the air conditioner to help cool down. A temperature sensor is installed in the room to detect the indoor temperature. When the indoor temperature rises to a certain degree Celsius, the air conditioner can start itself to cool down the room, so as to realize the intelligent temperature regulation function of controlling the opening and closing of the air conditioner according to the room temperature.The principle of the Temperature sensor is that its resistance will decrease with the increase of Temperature, so the change of resistance value can show the change of indoor Temperature. The switch operation of the air conditioner can be controlled by the change of resistance value [13]. In this paper, the Negative Temperature Coefficient (NTC) thermistor is selected as the Temperature sensor of this system. Its resistance value is calculated as follows:

$$R_t = \mathrm{Re}^{B(1/T_1 - 1/T_2)} \tag{14}$$

In the formula, R_t is the resistance value of thermistor at T_1 temperature, R is the nominal resistance value of thermistor at T_2 room temperature, and B is an important parameter of thermistor, The natural logarithm of both sides of Equation (14) is obtained:

$$\ln R_t = \ln R + B\left(\frac{1}{T_1} - \frac{1}{T_2}\right) \tag{15}$$

After transforming it, we get:

$$\ln R_t = \frac{B}{T_1} + (\ln R - \frac{B}{T_2}) \tag{16}$$

It can be seen that $\ln R_t$ has a linear relationship with $1/T_1$, and the slope obtained by linear fitting is the B value of the thermistor [14]. It reflects the sensitivity of the thermistor's resistance to changes in temperature, It is defined as the ratio of the difference between the natural logarithm of the zero-power resistance value at two temperatures and the difference between the reciprocal of the two temperatures. The formula is as follows:

$$B = \frac{\ln(R_{T1} / R_{T2})}{(1/T_1) - (1/T_2)} \tag{17}$$

Among them, R_{T1} refers to the zero power resistance value at temperature T_1, and R_{T2} refers to the zero power resistance value at temperature T_2. Considering the accuracy of measurement problems in this system, 10k value of 3950 B is finally determined Ω thermistors.

However, in real life, people may be dissatisfied with the set temperature value that controls the opening of the air conditioner and choose to turn on the air conditioner manually. Therefore, manual mode is added in the process of controlling the air conditioner. On this basis, the timing mode of air conditioning operation time is added to make it more flexible and convenient. The circuit schematic diagram of using solar energy to supply power for air conditioning is shown in Fig. 5.

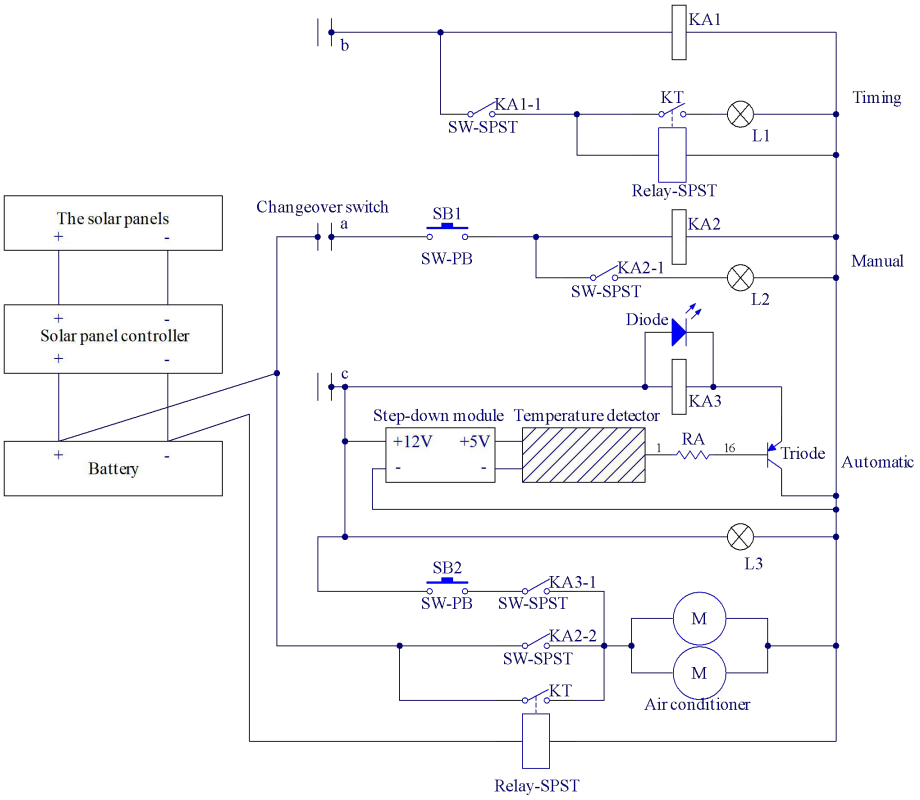


Fig. 5. Schematic diagram of intelligent temperature regulating circuit for solar power supply

3.4 Automatic dust removal

At present, the photoelectric conversion efficiency of solar cells is still not very ideal, usually only 10% to 20%, and the conversion efficiency needs to be further improved. The accumulation of dust on the battery surface is an important factor affecting the photoelectric conversion efficiency. New solar glass window in the long-term use of the process will be affected by all kinds of weather factors, it is easy to form ash on its surface, especially in some harsh environment in the region, the speed of ash will be accelerated, whether the ash is timely treatment is a problem that cannot be ignored. The accumulated dust will not only reduce the daylighting rate of solar cell modules, affect their power generation efficiency, but also accelerate the aging speed of their internal structures, thus plunging the whole system into a vicious cycle [15].

The new solar glass window designed in this paper has an automatic dust removal function. The principle of the dust removal system is shown in Fig. 6. and the principle diagram of the circuit is shown in Fig. 7.

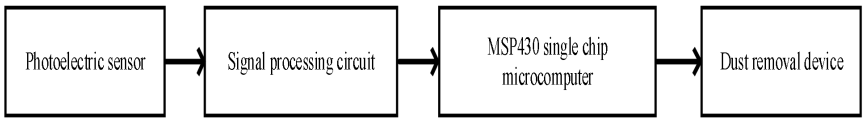


Fig. 6. Schematic diagram of automatic dust removal system

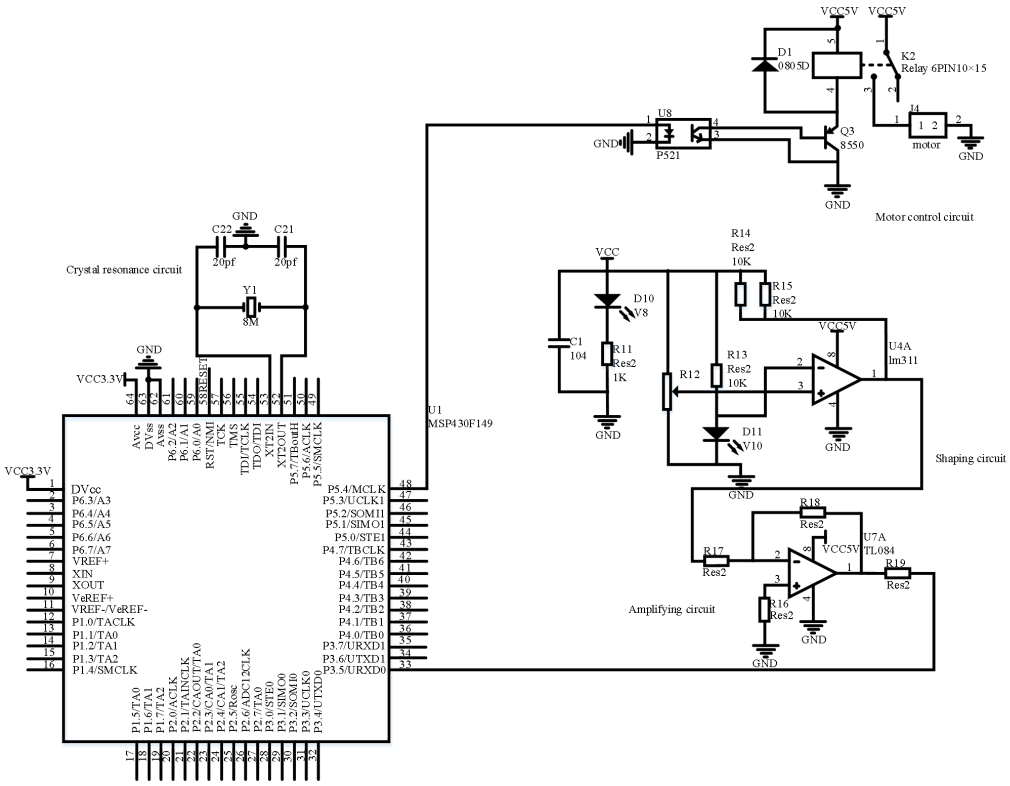


Fig. 7. Circuit schematic diagram of automatic dust removal system

Photoelectric sensor has the technical advantages of high resolution, high response speed and large information capacity, etc. In addition, its non-contact characteristics will not cause damage to the detection target and the sensor itself, which makes it can be used for a long time [16]. Because the photoelectric sensor has been working for a long time on the side outside the new glass window, it will be affected by some environmental factors. Therefore, the photoelectric sensor with higher sensitivity, lower energy consumption and easy maintenance should be selected. The photoelectric sensor selected by this system is E3F-DS30C4 adjustable diffuse reflection photoelectric switch, which is of NPN type with normally open structure. The detection distance is appropriate (30cm), and the suitable ambient temperature

range is wide (-25℃ ~ 70℃), meeting its due working conditions [17]. Photoelectric sensor in this system is the function of testing dust accumulation on photovoltaic modules, then send electrical signals to the system part of follow-up, when the dust on the solar panel is large, light irradiation to the solar panel is absorbed, and so cannot be reflected, the photoelectric switch into a closed state, sends out electric signals.

Because the electrical signal from the photoelectric sensor is very weak and unstable, it needs to be amplified and reshaped, so the signal processing circuit consists of operational amplifying circuit and reshaping circuit.The signal amplification circuit as shown in Fig. 8. is designed with TL084 as the operational amplifier. The TL084 chip has the advantages of small size, low power consumption, high conversion efficiency, low bias current and strong anti-interference [18].

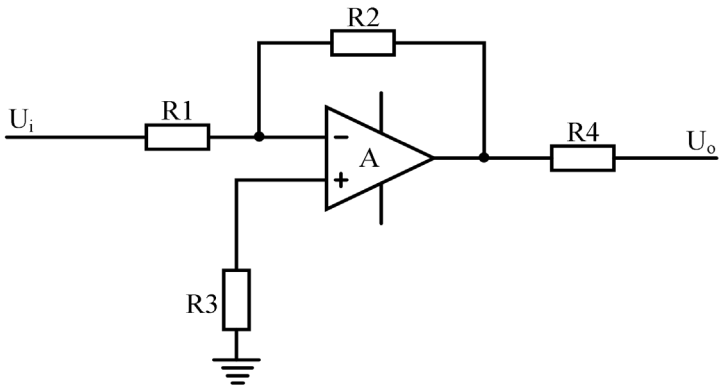


Fig. 8. Amplifying circuit schematic

The amplified signal is not a standard square wave and needs to go through a shaping circuit. Fig. 9. is a TTL shaping circuit with LM311 as the voltage comparator. The output pin of the amplifier is connected to an input pin of the voltage comparator, and the other input pin is connected to a potentiometer that can adjust the amplitude of the reference voltage. Adjusting the resistance value of R_t can change the size of the reference voltage. Depending on the operating cost of photovoltaic power generation, different regions have different tolerances for dust accumulation, so the reference voltage will be different.

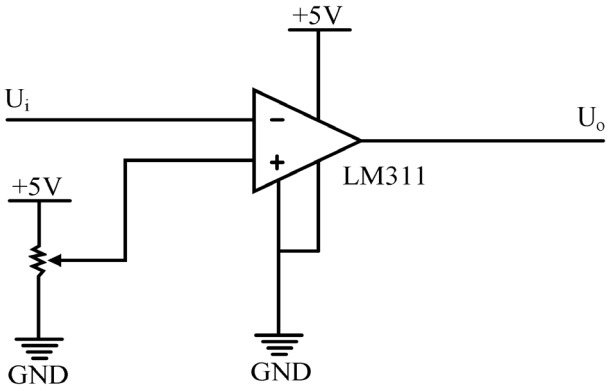


Fig. 9. TTL shaping circuit schematic diagram

After amplification and shaping, the signal is transmitted to the MSP340 MCU. The MSP430 MCU has its own ADC12 module, which is composed of a 12-bit SAR kernel, a sampling hold circuit, a reference voltage circuit and a clock circuit. With a low-power design, it only needs to set the corresponding register and power off independently to realize the conversion quickly and conveniently [19]. MSP430 microcontroller needs 3.3V DC power supply to work, and the sensor and other peripheral circuits need 5V power supply, so a 12-V rechargeable battery can be used for power supply. The 2V DC power supply is converted to a 5V power supply through step-down switching voltage regulator LM2576 for peripheral circuits such as sensors and relays. The 5V voltage is then converted to a 3.3V voltage through voltage regulator LP2985 for use by MSP430 single chip microcomputer. The power supply circuit is shown in Fig. 10.

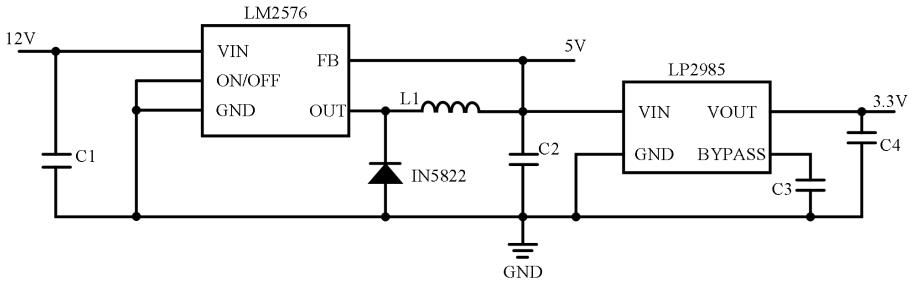


Fig. 10. Schematic diagram of power supply circuit

Since the MCU is a weak current system, it needs to be isolated from the strong current side to ensure safety and prevent the voltage backflow on the strong current side from burning out the MSP430. Therefore, the switch optocoupler is first used for photoelectric isolation, and then the relay is driven by the audion. The relay controls

the motor in the dust removal device for dust removal, and the driving circuit is shown in Fig. 11.

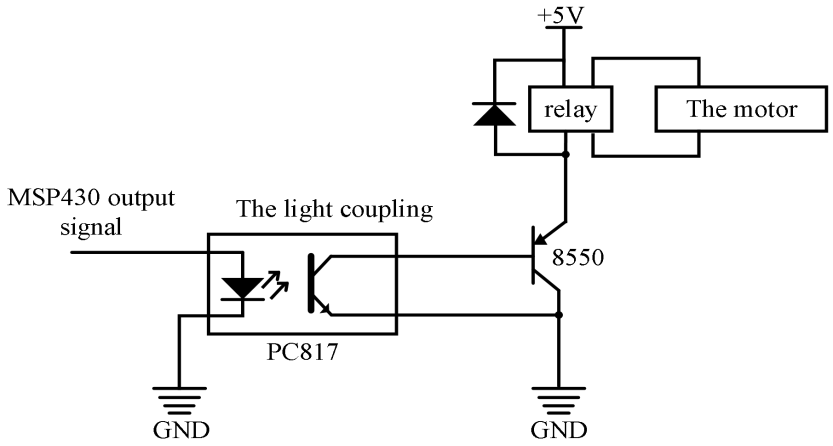


Fig. 11. Schematic diagram of drive circuit

The dust removal device is composed of brush head, DC motor and filter bag. The DC motor rotates to drive the brush head to work, breaking the binding force between dust and the surface of the solar panel, making the dust attached to the surface become suspended dust particles. Due to the high speed operation of the motor, the negative pressure difference is formed between the internal pressure of the main engine and the external atmosphere. Under the action of pressure difference, the suspended dust particles are inhaled into the filter bag, and the filter bag can be taken out for cleaning after being filled. Dust removal efficiency can be calculated by the following formula:

$$\eta = \frac{C_{in} - C_{out}}{C_{in}} \times 100\% \tag{18}$$

Among them, C_{in} is the dust mass concentration of the inlet flow and C_{out} is the dust mass concentration of the outlet flow. The calculation expression of C_{in} is as follows:

$$C_{in} = \frac{1000W}{60v_{in}S} \tag{19}$$

Among them, W is the amount of powder given per unit time (g/min), V_{in} is the inlet velocity of airflow (m/s), which can be measured by anemometer, and S is the cross-sectional area of airflow (m^2). Outlet concentration C_{out} can be measured by the HC-9001 smoke and dust automatic tester according to the principle of constant velocity sampling of particulate matter [20]. The efficiency of the dust collector exceeds 99.5% by calculation.

4 Experimental analysis

Fig. 12. is the new type of solar glass window designed in this paper. Its overall size is 37cm*30cm.

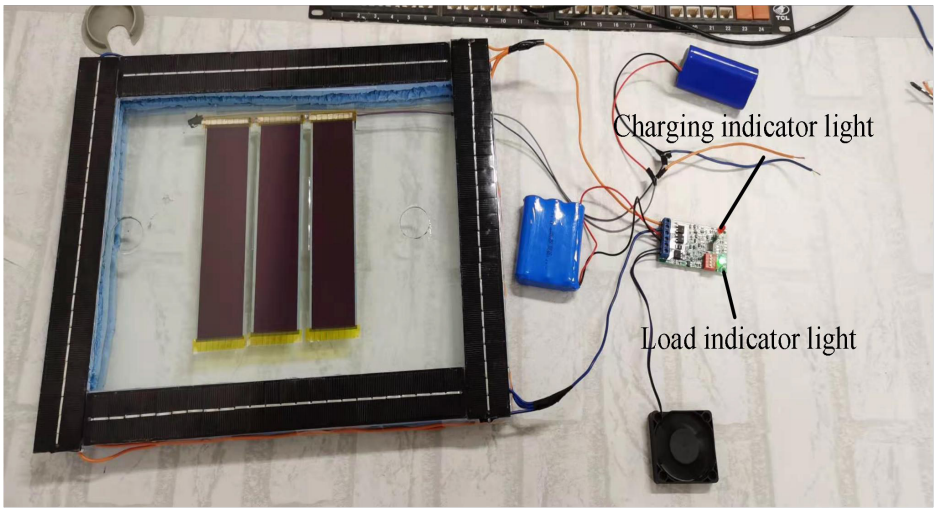


Fig. 12. Physical view of new solar glass Windows

After testing, when the glass window is not placed in the sun and the sun is not enough, the solar controller selects the normally open mode. At this time, the charging indicator light is off but the load indicator light is on, and the battery supplies power for the small fan, as shown in Fig. 13. Show.

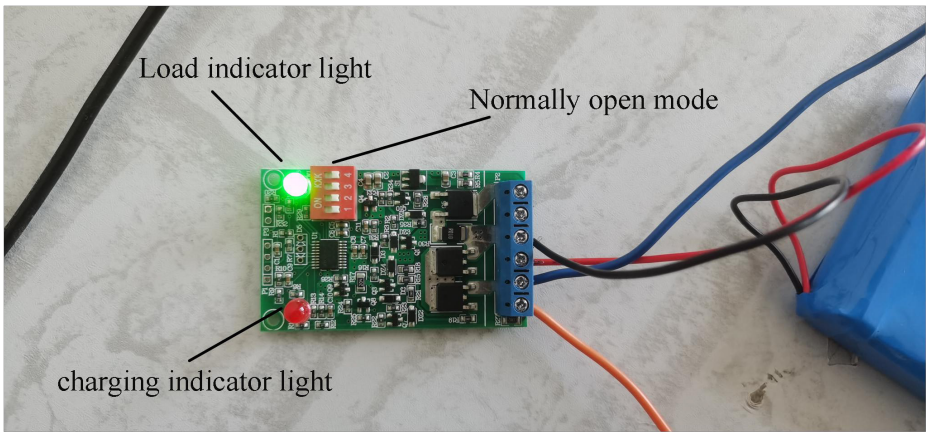


Fig. 13. Diagram of the light on when the light is insufficient

When it is placed in the sun and the sunlight is sufficient, the charging indicator and load indicator will all light up. The solar panel and flexible thin-film solar cell

generate electricity to charge the battery. The battery can drive a small DC 12V fan, as shown in Fig. 14. Show.

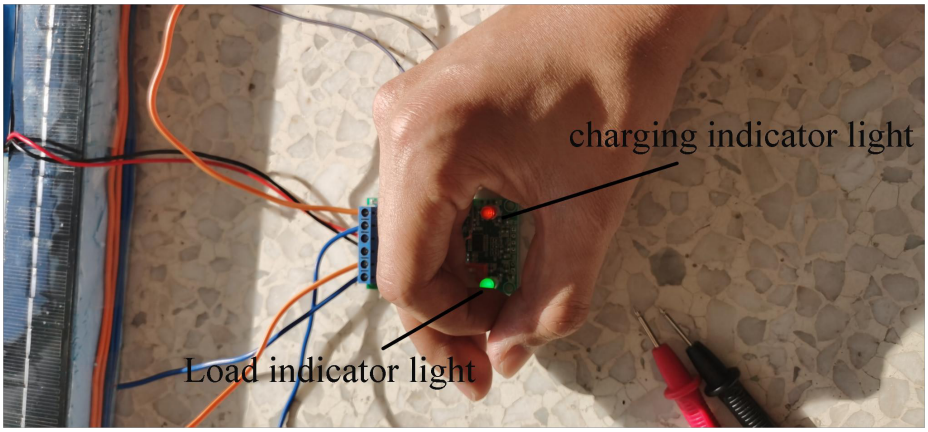


Fig. 14. Picture of the indicator light is on when the light is sufficient

The specific performance and value are shown in Table 1.

Table 1. New solar glass window performance table

Specific performance	Specific numerical
The voltage produced by the solar panel	11.49V
The voltage produced by a flexible thin film solar cell	3.88V
The storage capacity of an accumulator (connected to a solar panel)	2000mAh
Storage capacity of a battery (connected to a flexible thin film solar cell)	6000mAh
Improve the energy saving effect	8.3%
Dust removal efficiency	99.5%
Photoelectric conversion efficiency	1.91%

5 conclusion

In this paper, aiming at the problems such as excessive indoor temperature in hot summer and waste of electricity caused by the use of air-conditioning and other cooling equipment, a new type of solar glass window with photovoltaic power generation and intelligent control of indoor temperature is designed. Flexible thin-film solar cells combined with outer glass and solar panels mounted on window frames can generate electricity for indoor air conditioning, according to the temperature sensor intelligent control of the air conditioning on and off, to solve the problem of waste of air conditioning energy consumption, at the same time, in order to improve the photoelectric conversion efficiency of the whole system and reduce the influence of ash accumulation on the solar cells, an automatic dust removal device is installed on

the outside of the glass window, and the solar cells are cleaned automatically according to the photoelectric sensor. The new solar glass window designed in this paper is of great significance to indoor cooling and energy saving.

References

1. Wang Jinpeng., Ye Zhengpeng., Jeremy Gillbanks., Tarun M. Sanders and Zou Nianyu.: A Power Control Algorithm Based on Chicken Game Theory in Multi-Hop Networks. Symmetry 2019, 11(5),718.
2. Zhang Ying.: Research on the Application of Solar Energy in Rural Residential Buildings in the Sichuan Basin[D]. Southwest Jiaotong University, 2018.
3. Li Kai., Zhang Xianmin., Zhang Xiang.: Optimum design of front electrode grid line of non-equal width main grid of solar cell[J]. Acta Solar Energy,2021,42(01):69-76.
4. Singh S., Agrawal S.: Efficiency maximization and performance evaluation of hybrid dual channel semitransparent photovoltaic thermal module using fuzzyfied genetic algorithm [J]. Energy Conversion & Management, 2016, 122: 449-461.
5. Gaur A., Tiwari G N., Ménézo C., et al.: Numerical and experimental studies on a Building integrated Semi-transparent Photovoltaic Thermal(BiSPVT)system: Model validation with a prototype test setup [J]. Energy Conversion & Management, 2016, 129: 329-343.
6. Huang Ye., Zhang Gaoming., Gao feng., Liu Zhibing., Wang Zexin., Wei Jinjia.: The thermoelectric performance of dual-axis tracking and single-axis north-south tracking CPV/T system[J]. Solar Energy,2021(02):66-73.
7. Xiaoli Wu.: Influence of temperature of photovoltaic insulating glass modules on electrical performance [D]. Xi 'an University of Technology,2013.
8. Wang Jinpeng., Cao Fan., Zou Nianyu.: Multi carrier system joint receiving method based on MAI and ICI. [J]. Jilin Daxue Xuebao (Gongxueban), 2018, 41(6): 301-305.
9. Wang Jinpeng., Cao Fan., Zou Nianyu.: Cooperative distributed antenna transmission method based on co-channel interference in 5G mobile communication system[J]. Jilin Daxue Xuebao (Gongxueban),2020，48（1）：333-341.
10. Jun Gao., Hongyun Wang.: Application research of Photovoltaic power generation System in building Energy Saving [J]. Equipment Maintenance Technology,2020(02):64.
11. Wang Jinpeng., Ye Zhengpeng., Tarun M., Bo Li and Zou Nianyu.: A Novel Linear Antenna Synthesis for Linear Dispersion Codes Based on an Innovative HYBRID Genetic Algorithm. Symmetry 2019, 11(9), 1176.
12. Chen Chang.: research on the application of solar photovoltaic ventilated glass Windows in architecture based on CFD simulation [D]. Inner Mongolia: Inner Mongolia University of Technology. 2015, 1-63.
13. Lan Yu., Jie Zhang., Xiaoman Zhai., Qiming Zhao., Huijie Guo.: Design of intelligent Ventilation System powered by Solar Energy [J]. Science and Technology Economy Guide. 2017, (18) : 107.
14. Su Jun., Lin Yangfan., Yu Guanxia., Luo Min.: Thermistor characteristics and the design of temperature control experimental instrument[J]. Physics Experiment of College, 2019, 32(05): 60-63.
15. Yu Duan.: Research on the function and progress of battery modules in photovoltaic power generation unit system [J]. Smart city,2019, 5(20): 80-81.
16. Tao Wang.: Principle and Application of Photoelectric Sensor [J]. Computer Products and Circulation, 2018(07): 64+122.

17. Jinpeng Wang., Zhengpeng Ye., Fan Cao., Nianyu Zou.: Collaborative Distributed Antenna Transmission System based on Co-frequency Interference distribution in 5G Mobile Communication [J]. Journal of jilin university (engineering science), 2020, 50(01): 333-341.
18. Wang Jinpeng., Cao Fan., Zou Nianyu.: Cooperative distributed antenna transmission method based on co-channel interference in 5G mobile communication system[J]. Jilin Daxue Xuebao (Gongxueban), 2020, 48(1): 333-341.
19. Shoufeng Zhang., Xiaoyan Liu.: Design of Photovoltaic Dust removal Detection System based on MSP430 [J]. Information Technology and Informatization, 2017(11): 17-19.
20. Zhenmin Yang., Chao Huang., Jiahao Wang., Wenting Geng.: Surfactant influence on wet electrostatic precipitator performance [J/OL]. Environmental engineering: 1-10 (2020). http://kns.cnki.net/kcms/detail/11.2097.X.20201124.1749.010.html.

The simulation of coherent optical communication technology

Mingyang Li[1], Man Feng[2],Pengyuan Chen[3],Zhenping Lan[4],Ping Li(✉)
{ 523155344@qq.com[1], fengman2020@163.com[2], 554855487@qq.com[3],
lanzp@dlpu.edu.cn[4],liping@dlpu.edu.cn(✉)}

Department of Information Science and Engineering Dalian Polytechnic University Dalian, P. R. China

Abstract. With the advent of the era of big data and the development of Internet technology, various emerging businesses have emerged. Due to the lack of optical coherence, the essence of optical fiber communication is a noise carrier communication system. In order to be able to communicate with high capacity over long distance, various modulation methods and heterodyne reception are used to construct coherent optical communication system. Heterodyne/coherent fiber communication is being accelerated in its research and development due to the possibility of increasing the receiving sensitivity by 20-30dB[1]. The coherent optical communication system has good selectivity and large communication capacity. The receiving sensitivity of the coherent balanced detection technology is about $10\ dB \sim 20\ dB$ higher than that of the incoherent optical communication system, which greatly extends the transmission distance [2]. This paper will study the optical coherence communication system and its simulation, the main components are as follows:

1. General information and importance of the development status and future trend of coherent optical communication.

2. A review of coherent optical communication system.

3. Based on the understanding and analysis of the previous chapter, the OptiSystem software platform is used for system simulation. Designed the core algorithm of DSP processing unit and simulated the coherent optical communication system.

Keywords: Coherent optical communication system; the key technology of coherent optical communication; OptiSystem simulation.

✉ Corresponding author: Ping Li(1969-), female, Professor, master, main research directions for optical communications and intelligence systems.
E-mail:liping@dlpu.edu.cn
Room 312, Photonics Institute, Dalian Polytechnic University, No. 1, Light Industry Garden, Ganjingzi District, Dalian.

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307403

1 Introduction

1.1 Background and significance of the project

Information is booming rapidly. With the development of the global Internet, people have made great efforts to catch up with the huge increase in global traffic demand [3].Heterodyne detection is the main method to realize information transmission detection by using coherent optical communication technology, which can ensure that communication managers can find out problems such as frequency conversion difference in information transmission, so that managers can modify information transmission parameters and optimize the quality of information transmission.

In the 21st century, there has been an explosive growth of broadband information services, and the existing optical communication systems are facing serious data transmission problems.A communication system is urgently needed to make the information transmission faster, stronger and farther.In recent years, coherent optical communication technology has developed rapidly, because coherent optical communication makes full use of the mixed frequency gain, good channel selectivity and wavelength tunability of coherent communication [4].

1.2 Development status and future trend of coherent optical communication

Optical fiber technology is widely used in communication construction. With the extensive use of communication, the application of optical fiber technology is gradually expanding [5].The ultimate goal of scientists is always to increase the transmission bandwidth, increase the transmission distance and improve the reception sensitivity as the research direction of optical fiber communication.In recent years, with the spread of Internet video, the information generated by them has increased dramatically. The application of these communication technologies is huge, which requires the stability and efficiency of the communication system.The transport performance of the physical layer becomes more and more important.Therefore, wave-multiplexing technology no longer meets the requirements of optical fiber communication for transmission distance and bandwidth.Although erbium-doped fiber amplifier (EDFA) is also in use, it still fails to meet the requirements of optical fiber communication researchers [6].The emergence of new technology inevitably requires the emergence of new equipment.Optical equipment has always been expensive, if large-scale replacement of communication equipment.Operators can't afford to spend so much.Therefore, for equipment manufacturers, the development of new optical communication is also risky.We now need to improve the performance of optical communication systems on existing equipment.Coherent optical communication system is born in this case.

2 A Review of coherent optical communication systems

2.1 Basic principles of coherent light detection

Coherent modulation and heterodyne detection are the main means of coherent optical communication [7].The basic principle is that the signal light first acts on the optical carrier by means of external light modulation, which has various modulation modes, including amplitude, frequency or phase modulation, and then passes through the optical detector.All above belong to the optical transmitter part.The signal is transmitted to the optical fiber, belonging to the optical fiber part.At the receiving end, the signal optical signal is coherently coupled with the local optical oscillator at the mixer, and then the signal is output to the optical detector for

detection.In order to ensure a good match between the modulated light wave output from the photodetector and the fundamental mode HE_{11} in the single-mode fiber, and to ensure that the polarization state of the modulated light wave modulated by external light is consistent with the local vibration state in the single-mode fiber, it is necessary to detect at the transmitter part.

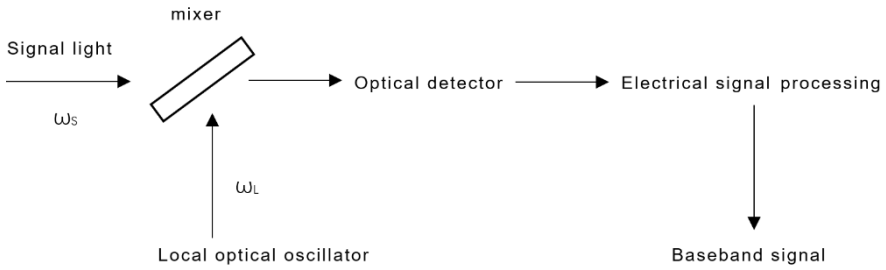


Fig. 1. Principle block diagram of optical coherence detection

2.2 Composition of coherent optical communication system

Two beams of light vibrate in the same direction, have the same oscillation frequency, the same phase or phase difference.Such two beams of light meet the conditions of coherence, known as coherent light.Coherent optical communication system produces coherent optical transmitter, coherent optical receiver and intermediate frequency signals generated by optical transmission link.After photoelectric detection and enhanced if signal, the demodulator recovers the output signal required by the user [2].Compared with traditional intensity modulation techniques, the receiver is more sensitive, can maintain a higher modulation level, and can extract all information about electric field light [8].Coherent optical communication system consists of three parts: coherent optical transmitting terminal, optical transmission link and coherent optical receiving terminal.

The coherent optical transmitter receiver includes channel coding, external modulator, automatic bias control circuit of external modulator, narrow line-width laser light source, optical splitter, polarization control, etc [2].The optical transmitter assembly converts the HDB3 signal encoding from a multistage device to an NRZ code, which is converted to the code available on the photoelectric cable, and then converts the electrical signal to an optical signal and enters the optical fiber for communication.

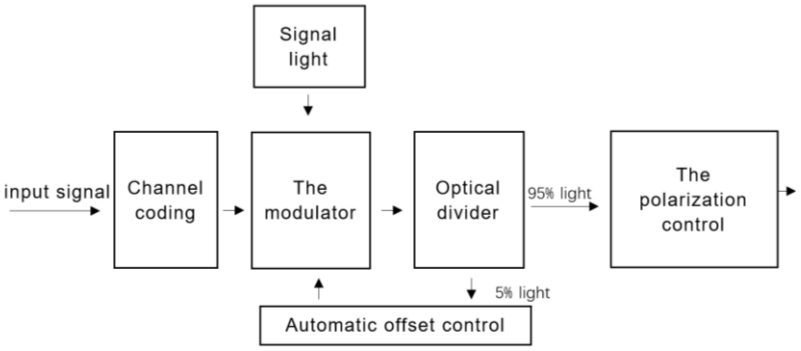


Fig. 2. Coherent optical transmitter

The function of optical transmission link is to transmit modulated modulated light waves from coherent optical transmitter to coherent optical receiver. The transmission mode is HE_{11} mode.Attenuation of the amplitude, phase delay and possible change of polarization direction of the light wave occur during the whole transmission process.

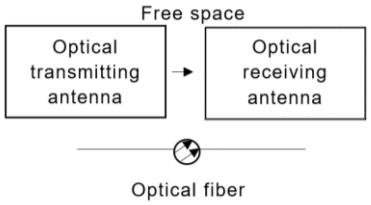


Fig. 3. Optical transmission link

The light transmitted from the optical transmission link goes into a matching device that matches the transmitter.The device also allows receiving the spatial distribution and polarized light state based on the laser's output of its own vibration.The mixer is used to mix the local light wave (frequency f_1) and the received light wave (frequency f_S) with each other, and the photoelectric detector is used for detection, and then the difference frequency signal (frequency $f_S - f_1$) between them is detected and amplified with an intermediate frequency amplifier.

2.3 Key technology of coherent optical communication

It mainly includes the following: signal source and local vibration source must use laser with stable frequency and high spectral purity.Ordinary lasers cannot meet this requirement, so it is necessary to study frequency stabilization technology.In order to obtain high mixing efficiency, it is necessary for coherent optical communication system to carry out very strict matching conditions when signal light and local oscillator light are mixed.

If it is directly modulated by optical laser parameters.Then the other parameters will always cause parasitic vibrations, just as the modulation frequency is when the phase changes.In addition, they may also encounter problems such as unstable loose frequencies. Therefore, in addition to the current directly injected modulation frequency, other transmitted light wave modulation adopts external light modulation [14].External light modulation depending on physical phenomena, some characteristics of photocrystals or optical crystals differ from external factors.For example, a diffused LiNbO3 Mach interferometer or a directionally coupled modulator can be used for ASK modulation, and a semiconductor quantum well phase external modulator or LiNbO3 phase modulator can be used for PSK modulation [14].

In coherent optical communication system, coherent control requires the signal source to be polarized in the same direction, that is, the vector direction of two light sources is the same, in order to obtain high coherent reception sensitivity, otherwise the sensitivity of coherent detection will be greatly reduced.Because the polarization direction of the signal light source and the local oscillator light source is the same, only the projection of the electric signal vector can truly generate the intermediate frequency current [14].Receiver sensitivity can be difficult to ensure if the Angle misalignment exceeds 60°, which may lead to the loss of coherent reception advantage.Therefore, in order to make full use of the benefits of coherent reception, measures are being taken to stabilize polarization.

Polarization-maintaining fiber: in order to let the process of optical wavelength polarization state remains unchanged (ordinary single-mode fiber will change the light

polarization state, reason is due to the mechanical vibration of the fiber or temperature change and other factors), but compared with the single mode fiber, "polarization-maintaining fiber with high loss and price on the high side [14].Polarization classification: The use of a common single-mode fiber, using polarization classification technology at the receiving end of the coherent light.After the signal source is mixed with the local oscillator, it is first divided into two channels for balanced reception.For each channel signal, polarizing beam splitter is used to divide them into orthogonal polarized signals of two channels for detection, and then the sum of squares is calculated.Finally, the two receiving channels are judged, and the better channel is selected as the output signal.However, the output signal is independent of the polarization state of the received signal, which eliminates random variations in the polarization mode of the signal during transmission.

Laser frequency stabilization is very important in coherent optical communication.For example, when the frequency (or wavelength) of the laser changes under different working conditions, it is difficult to determine the zero deviation by the relative frequency stability of the light source in the coherent optical communication system.This also applies to heterodyne coherent optical communication systems.In general, the average heterodyne frequency is between 0.2 and 2GHz.If the optical carrier is 1.5 m, she will have a 200THz intermediate frequency of 10^(-6)~10^(-5).If the optical frequency carrier changes, there will be a large average frequency change.Therefore, it is necessary to ensure the stability of the high-frequency light wave generator and light wave, so that the coherent light communication system can work normally [14].

The amplitude noise of the semiconductor laser and the influence of frequency modulation on the sensitivity of the receiver can be overcome when the narrowband of the light wave is ensured.The narrower the line width is, the smaller the phase shift generated by the noise will be[14].There are two ways to do this :(1) this is a method of blocking patterns.In other words, the intensity of the laser is the stable frequency and narrow spectral line, and the single-type laser introduces the laser with spectral width, similarity and frequency stability, and the laser is mainly the compression width.(2) External cavity feedback method.The external cavity feedback is that the laser output passes through the mirror cavity and then through the external, grating and other dispersed elements, so as to obtain the external cavity with dynamic single-mode application and select the characteristic using the mode, and the width of the compression spectral line depends on the Q value of the external cavity [14].

Coherent optical communication often uses a strong frequency separation method, so nonlinear effects in optical fiber communication may cause the signal strength and phase strength of a single optical communication channel to depend on the conduction frequency of another channel and nonlinear continuous interference.The effect of the nonlinear refractive index fiber, and thus the phase modulation of phase noise, should be considered when the signal power is greater than 10mW in coherent optical communication systems or when the signal is transmitted over long distances through optical amplifiers.When the channel space and fiber dispersion are small enough, the phase condition of FWM can be satisfied.FWM is an important component of nonlinear intervention.FWM limits the performance of the system and reduces the energy of the channel to block the channel.When the channel power is less than a certain value, the influence of FWM on the system can be avoided [14].

Signal light and local oscillator light need to be provided by coherent optical communication.It requires narrow band of a longitudinal module, high frequency stability, dynamic single frequency characteristic of dynamic laser signal in dynamic modulation, width of modulation band meeting coding speed requirement.At the same time, it is also necessary to design a good control circuit to make the laser output very narrow line-light stability at

different temperatures.Narrow and stable spectral line width is a necessary condition for the practical application of coherent optical communication system. The wider the spectral line of the light source, the greater the phase noise and the worse the sensitivity will be. The spectral line width of the light source itself will determine the lowest bit error rate and the highest receiving sensitivity that the system can achieve[2].

3 Simulation of Coherent Optical Communication System Based on OptiSystem

In this paper, the QPSK coherent optical communication system for local area network (LAN) is simulated.QPSK modulation and demodulation technology, also known as orthographic phase shift keying technology, belongs to a phase modulation technology [9].It is improved by PSK technology, and has many advantages over traditional modulation and demodulation technology. It transmits more information in limited bandwidth, and the interference has little effect on the signal during transmission.

3.1 Design principle

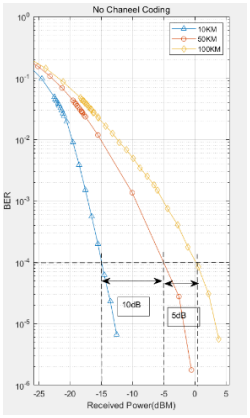


Fig. 4. Design schematic

Figure 4 is the design schematic diagram of QPSK coherent optical communication system.As shown in the figure, the probability of bit error increases with the increase of received power.And the longer the transmission distance, the higher the bit error rate, both of which affect the bit error rate.When bit error rate is 10^{-4}, transmission distance of 50km will reduce 10dB from receiving power of 10km, while transmission distance of 100km will reduce 5dB from receiving power of 50km.

3.2 Circuit design

Figure 5 is the design drawing of the 1Gbit/s QPSK coherent optical communication system.The pseudo-random binary sequence of 1Gbit/s is generated by the pseudo-random sequence generator and input to the QPSK optical transmitter, which sends the modulated signal into the optical transmission link for transmission.Then the QPSK optical receiver demodulates the received signal through coherent detection, then amplifies and filters, and finally recovers the QPSK signal.The performance of the system is evaluated by bit error rate (BER), error vector amplitude percentage, received signal power, and constellation [10].

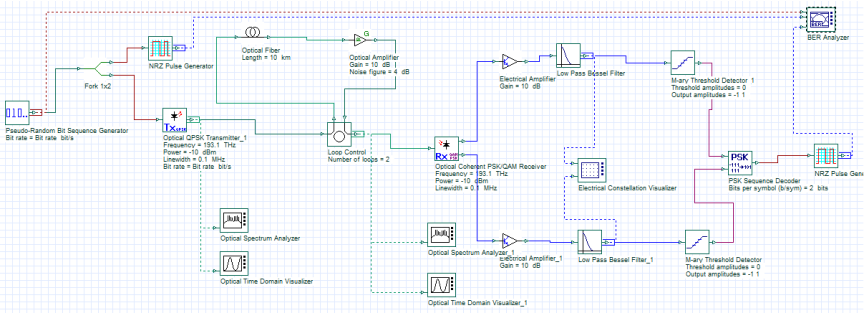


Fig. 5. Circuit diagram

3.3 The results demonstrate

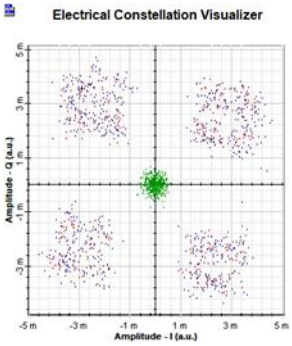


Fig. 6. Result illustration

Figure 6 is the constellation chart of QPSK coherent optical communication system.It can be seen that this constellation chart is not good, appear more obvious constellation blur and deviation.It can be seen from the constellation that the signal is strongly affected by noise and is far away from the ideal vector point with a high error rate.The primary task of communication is to transmit information.Therefore, the main index of communication system quality is the reliability of information transmission.Reliability means the accuracy of information. In digital communication system, the accuracy of information is mainly measured by the frequency in THE BER code.What we need is high reliability and low bit error rate.In order to realize the long distance accurate and stable transmission of optical communication system, we added DSP component to reduce bit error rate. The following is the design of DSP processing unit.

3.4 DSP processing unit core algorithm design

The coherent scheme can not only significantly extend the communication distance, but also significantly improve the information carrying capacity [11].The coherent control method uses digital signal processing (DSP) to balance and compensate PMD polarization mode and electric field nonlinear effect, and improve the circuit's resistance to dispersion and ROADM filtering limit [12].

In coherent QPSK systems, when the symbol phase is$\left\{0,\ \frac{\pi}{2},\ \pi,\ \frac{3\pi}{2}\right\}$ or $\left\{\frac{\pi}{4},\ \frac{3\pi}{4},\ \frac{5\pi}{4},\ \frac{7\pi}{4}\right\}$, the sampling value of the Kth symbol received is assumed to be

$$S(k) = e^{j(\theta_s(k)+\Delta\omega kT+\theta_L(k)+\theta_n(k))}. \quad (1)$$

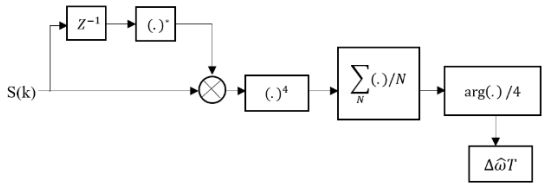


Fig.7. Block diagram of quadrature frequency offset estimation algorithm

The quadriple algorithm's quadratic error assessment process usually consists of four steps: the first step is to eliminate the phase damage θ_L caused by the laser linewidth.In high-speed optical transmission systems, θ_L changes slowly relative to the symbol rate, with θ_L before and after being considered constant.Then, the phase difference between the two symbols can be eliminated by calculating the phase difference

$$S(k)S^*(k-1) = e^{j(\Delta\theta_s+\Delta\omega T+\theta_n)}. \quad (2)$$

The second step is to remove the remaining modulation phase $\Delta\theta_s$ .Whether the modulation phase of the QPSK signal is$\left\{0,\ \frac{\pi}{2},\ \pi,\ \frac{3\pi}{2}\right\}$ or $\left\{\frac{\pi}{4},\ \frac{3\pi}{4},\ \frac{5\pi}{4},\ \frac{7\pi}{4}\right\}$, the QPSK signal has a value of $\left\{0,\ \frac{\pi}{2},\ \pi,\ \frac{3\pi}{2}\right\}$.Through the fourth power operation of the processing result of the first step, the phase component of the removed $\Delta\theta_s$ part is:

$$(S(k)S^*(k-1))^4 = e^{j(4\Delta\omega T+4)\theta_n}. \quad (3)$$

The third step is to remove the noise phase θ_n.In high-speed optical transmission system, the deviation between laser frequency and symbol speed is also very slow.The difference produced by frequency deviation from the adjacent N symbol can be considered as a constant $\Delta\omega T$.The calculated mean value of N results can approximately eliminate the influence of noise phase on the estimated results

$$\sum_N \frac{(S(k)S^*(k-1))^4}{N} = e^{j4\Delta\hat{\omega}T}. \quad (4)$$

The fourth step is to extract carrier frequency offset valuation.

$$\frac{\arg\left\{\sum_N \frac{(S(k)S^*(k-1))^4}{N}\right\}}{4} = \Delta\hat{\omega}T. \quad (5)$$

3.5 Circuit design based on DSP

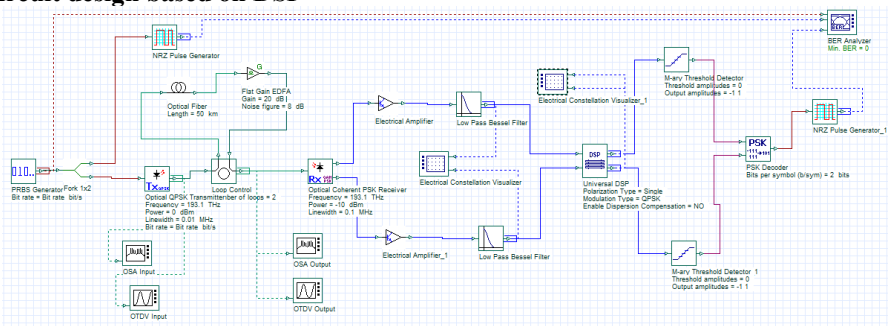
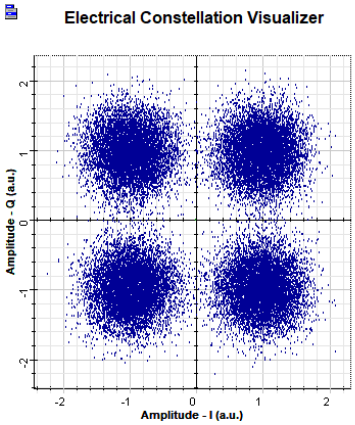


Fig.8. Circuit diagram

Figure 8 is the design drawing of QPSK coherent optical communication system.The parameters of the simulation system are set as follows: the transmitter uses an optical rectilinear phase-shift coded transmitter with a frequency of 193.1THz, and passes through a 50km fiber, which is amplified by an amplifier, with a gain of 20dB and a noise of 8dB.The receiver uses coherent detection, optical coherent PSK receiver, passes through the amplifier and the low pass filter, passes through the DSP module, obtains the constellation chart of this system.Compared with the previous design drawing, in order to realize long distance stable transmission, DSP module was added to carry out digital signal processing and reduce bit error rate.Advanced digital signal processing (DSP) algorithms have been implemented at receivers to compensate for signal damage during propagation [13].

3.6 The results demonstrate



Compared with the constellation chart without DSP module, this chart is more complete than it.As can be seen from the figure, the actual signal distribution in the signal space is very concentrated, and the bit error rate is relatively low at this time.The design of DSP processing unit can reduce the bit error rate under QPSK modulation and improve the transmission efficiency of coherent optical communication.

4 Conclusion

This paper mainly discusses and studies the coherent optical communication system, puts forward the experimental theoretical results, and verifies the theoretical analysis through the simulation experiment.In the introduction, the development trend and advantage of coherent optical communication and its broad development prospect are briefly introduced.The second chapter describes the coherent optical communication system in all aspects.The advantages and disadvantages of coherent optical communication system are introduced and analyzed. The structure principle and key technology of coherent optical communication and its broad application prospect are expounded.The third chapter is the focus of this paper. It is based on the second chapter to carry out experimental simulation and theoretical analysis.In order to reduce bit error rate, DSP processing unit core algorithm design, simulation and analysis.The paper has been basically completed here. Due to the lack of knowledge, there are many shortcomings in this paper.I hope the teacher can point out in time that if I can understand the content of my research after reading this article, this is the goal I want to achieve, and it is also the greatest affirmation and support for my work.

References

[1] A Alphonse;G S Sanyal.Coherent Optical Fibre Communication Systems[J].IETE Technical Review.1986,Vol.3(No.8)：429-433.

[2] Zhihong Zhu. Design and Implementation of Coherent Optical Communication System [J]. the electronic world,2016(09):139-140.

[3] Coherent Optical OFDM for Long-haul Communication System[D]. Beijing University of Posts and Telecommunications,2018.

[4] Xiangdong Lv,Xuerui Liang,Qianchen Yu,Weidong Ma. Research status and development trend of optical communication technology [J]. Telecom science,2019,35(02):70-78.

[5] Jiong zhang. Application and development trend of optical fiber technology in communication engineering [J]. Engineering research,2019,4(15):237-238.

[6] Lihong Niu,Liping Kong. On Coherent Optical Communication [J]. SME management and technology (Under the ten-day),2012(07):219-220.

[7] Jirong Shi,Lin Liu.2.5 Gbit/s Research on Coherent Optical Communication System [J]. Shanghai Electric Technology,2019,12(03):14-16.

[8] Pengxia Li,Xizheng Xi. Experimental study on QPSK modulation and demodulation in coherent optical communication system [J]. Laser technology,2019,43(04):127-132.

[9] Wenbo Guo. Design and Simulation of QPSK Communication System Based on Matlab[J].Information and communication,2019(11):211-212.

[10] Jagana Bihari Padhy,Bijayananda Patnaik. 100 Gbps multiplexed inter-satellite optical wireless communication system[J]. Springer US,2019,51(7). 16.

[11] Ming - Seng Kao;Jingshown Wu.Forward and backward Raman amplifications in coherent optical communication systems[J].Journal of the Chinese Institute of Engineers.1992,Vol.15(No.4)：359-365.

[12] Yaming Wu,Guangcheng Li. Design of 100G coherent optical module [J]. Value engineering,2012,31(18):181-182.

[13] Dhiman Kakati,Subhash C. Arya. A 640-Gbps, 15.2344-b/s/Hz full-duplex optical fiber/wireless single-channel coherent communication system using IQM-based DP-256-QAM and DSP techniques[J]. Springer US,2020,39(1).

[14] Yang Yu. Coherent optical communication and its applications [J]. Physical bulletin,2001(09):43-46.

Performance Analysis of Indoor Visible Light ACO-OFDM Coding

CHEN Peng Yuan[1], LI Ping[1*]

{liping@dlpu.edu.cn}

[1]Research Institute of Photonics, Dalian Polytechnic University, Dalian 116034, China

Abstract: With the research and development of visible light communication technology, more and more researchers have begun to pay attention to more effective and reliable modulation technology and channel coding technology. In order to reduce the bit error rate of the shortwave wireless visible light communication system and increase the signal rate in the transmission process, the high efficiency and reliability of the modulation and demodulation technology must be realized. Therefore, the asymmetric limit orthogonal frequency division multiplexing technology (ACO-OFDM) is due to it's excellent anti-interference ability and power efficiency are widely used in visible light communication systems. This article will introduce in detail the basic principles of ACO-OFDM technology, Hermitian mapping, and use BPSK, 4-PPM,16QAM for bit error rate simulation analysis, and select a modulation method rate with a lower bit error rate; analyze the linear block code, BCH The basic principles of coding and convolutional codes, and ACO-OFDM systems based on these three coding methods are simulated by Simulink, and their bit error rates are compared. Finally, ACO-OFDM based on convolutional codes was selected. This system is more suitable for visible light communication.

Keywords: Visible Light Communication, ACO-OFDM, Channel Coding

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307400

1 Introduction

Visible light communication technology uses fluorescent lights or light-emitting diodes to send out high-speed flickering signals that are invisible to the naked eye to transmit information. It has the advantages of rich spectrum resources, low consumption, low cost, and good confidentiality. With the popularity and rapid development of light-emitting diodes (LED), LED-based lighting and display systems have become more and more integrated into people's lives, and visible light communication technology has become one of the fastest-growing high-tech fields in today's society. In order to realize the effectiveness and reliability of data transmission in the visible light communication system, this article introduces an asymmetric limiting optical orthogonal frequency division multiplexing (ACO-OFDM) technology, and adds channel coding to the visible light communication system to improve. The system's error detection and correction capabilities. First, this article explains the general system framework of the visible light communication system; second, this article explains the working principle of the ACO-OFDM technology and Hermitian mapping; and simulates the ACO-OFDM system under the visible light channel, and compares the ACO-OFDM and 4-PPM , BPSK, QPSK, 16QAM bit error rate simulation comparison; then, this article briefly introduces the linear block code, BCH coding and convolutional code three channel coding techniques, and the three coding techniques for bit error rate comparison simulation. Finally, a modulation and channel coding technology more suitable for visible light communication is selected.

2. Principle of Visible Light Communication

The visible light communication system structure is shown in Figure 2.1. It can be seen that, similar to the traditional radio frequency communication system, the visible light communication system is also divided into three parts: visible light communication transmitter, visible light communication wireless channel, and visible light communication receiver.

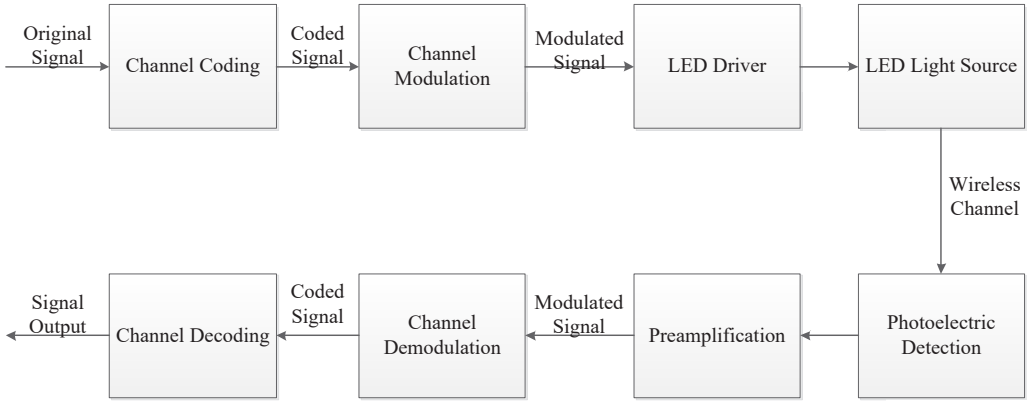


Figure 2.1 Visible light communication system block diagram

The specific functions of each part are as follows:

(1) Visible light communication transmitter

The original signal is a binary bit stream. First, pre-processing coding and modulation is performed, that is, an error correction code is added to the original signal to be transmitted and an appropriate modulation method is selected to improve the error correction capability and reliability of the transmission signal. After the LED is driven, the electrical signal is converted into an optical signal, and after passing through the LED semiconductor device, the electrical energy is converted into light energy and emitted. Among them, the light source used in the optical communication system can be a white LED.

(2) Visible light communication receiver

The main device at the receiving end is a photoelectric detector, which can perform photoelectric conversion of the signal. After the optical signal is converted into an electrical signal, it is demodulated and decoded corresponding to the transmitting end to restore the original signal.

3. ACO-OFDM Technology

3.1 The Technical Principle of OFDM

OFDM technology is one of Multi-Carrier Modulation (MCM) technologies. The so-called OFDM technology is to divide the data stream to be transmitted into several sub-data

streams, each sub-data stream has a much lower transmission bit rate, and the sub-data streams are used to modulate several mutually orthogonal carriers in parallel.

Each OFDM symbol is the addition of multiple modulated sub-carrier signals, and the modulation mode of each sub-carrier can be phase shift keying (PSK) or quadrature amplitude modulation (QAM). If N represents the number of sub-channels and T represents the width of the symbol, $d_i(i=0,1,\cdots,N-1)$ is the data symbol assigned to each sub-channel， f_c is the carrier frequency，The OFDM symbol starting from can be expressed as $t=t_s$ can be expressed as：

$$s(t)=\mathrm{Re}\left\{\sum_{i=N/2}^{N/2-1} d_{i+N/2}\exp\left[j2\pi\left(f_c-\frac{i+0.5}{T}\right)\left(t-t_s\right)\right]\right\},t_s\le t\le t_s+T \quad (1)$$

It is also possible to use the equivalent baseband signal as shown below to indicate the OFDM symbol beginning with $t=t_s$：

$$s(t)=\sum_{i=-N/2}^{N/2-1} d_{i+N/2}\exp\left[j2\pi\frac{i}{T}\left(t-t_s\right)\right],t_s\le t\le t_s+T \quad （2）$$

Among them, the real part and imaginary part in formula (2) respectively correspond to the in-phase component and the orthogonal component of the OFDM symbol. In practice, they can be multiplied by the cos component and sin component of the subcarrier to form the final OFDM signal.

3.2 The Technical Principle of ACO-OFDM

The OFDM signal is a bipolar signal, but for the traditional intensity modulation/direct detection optical communication system, the bipolar signal cannot be directly used because the light intensity cannot appear negative. Therefore, for optical OFDM, the usual solution is to add a DC bias DC bias to the OFDM signal during electrical-optical signal conversion, so that the bipolar signal becomes a unipolar signal, which can be transmitted on light.

However, the use of DC-biased OFDM will increase the average optical power of the transmitted optical signal while reducing the modulation depth, resulting in low power

efficiency of the system. Now a new ACO-OFDM method can be used to solve the problem of bipolar OFDM signals. ACO-OFDM is considered to cut off the part of the original OFDM signal that is less than 0, leaving only positive values.

Figure 3.1 shows that the specific process of the ACO-OFDM system is: at the transmitting end, the original signal is first subjected to channel coding and PSK or QAM modulation, and the mapped signal is serial-parallel converted into a parallel data stream, and the pilot signal is inserted, Hermitian mapping, IFFT, adding a cyclic prefix, zeroing the negative part of the signal, and then converting it into a serial data stream, forming an ACO-OFDM symbol and transmitting it into the wireless channel.

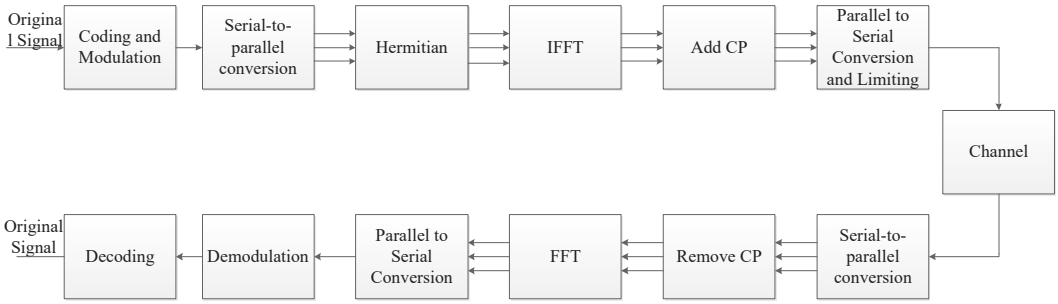


Figure 3.1 Block diagram of visible light communication system based on ACO-OFDM

When a signal is detected at the receiving end, the signal is first converted into a parallel data stream, the cyclic prefix is removed, and the parallel data stream is converted into a serial data stream after FFT. At this time, the PSK or QAM modulated data is obtained. The original data can be obtained by corresponding demodulation.

Hermitian Mapping

Suppose the time domain sampling of an OFDM symbol in the baseband without cyclic prefix is：

$$x(k)=\frac{1}{N}\sum_{m=0}^{N-1}X_m \exp\left(\frac{j2\pi km}{N}\right),\quad 0\le k\le N-1 \quad (3)$$

Where： $x(k)$ represents the sampled value sequence in the time domain； k represents the serial index of time sampling； N represents the number of subcarriers； X_m represent the

frequency domain data symbol modulated onto the m-th subcarrier after mapping.

When X has a Hermitian symmetric structure：

$$\begin{aligned} X &= (X_0, X_1, X_2, \cdots, X_m, \cdots, X_{N-1}) \\ &= (X_0,\ X_1,\ X_2, \cdots,\ X_{N/2-2},\ X_{N/2-1},\ X^*_{N/2-1}, \cdots,\ X^*_1) \end{aligned} \tag{4}$$

Where： X^* represent the conjugate of X ,The OFDM signals obtained after IFFT are all real-valued signals. X in this case, if only odd-numbered subcarriers are selected to transmit modulated data, and at the same time, even-numbered carrier modulated data is set to zero, which is expressed as:

$$X = (0, X_1, 0, \cdots, X_1, 0, X^*_{N/2-1}, \cdots, 0, X^*_1) \quad (5)$$

Then when X satisfies the above structure, the OFDM signal obtained after IFFT is a real value, and the part less than zero is set to zero before transmission, the amplitude of the odd-numbered information transmitted after the debit becomes the original general, and the other The information is not affected. This method of making the signal meet real-valued unipolarity is called asymmetric clipping (ACO).

Suppose the time domain expression of the OFDM signal after IFFT transformation is $x(t)$, $x_{ACO}(t)$ represents the unipolar ACO-OFDM signal after $x(t)$ is limited：

$$x_{ACO}(t) = \begin{cases} x(t), x(t) > 0 \\ 0, x(t) \le 0 \end{cases} \tag{6}$$

The ACO-OFDM system uses the odd-numbered sub-carrier to transmit data in the sub-carrier allocation, and the even-numbered sub-carrier does not transmit information. This method brings two advantages: First. The single-polarization operation of the signal can be realized directly by the method of zeroing the complex signal, which is relatively simple; second, the noise generated during the asymmetric limiting operation only falls on the even carrier that does not transmit information.

Guard Interval and Cyclic Prefix

In the OFDM system, inserting a guard interval between OFDM signals can eliminate inter-symbol interference to the greatest extent. The length of this guard interval is generally greater than the maximum delay extension of the wireless channel, so that the component of a symbol generated by multipath propagation is not will interfere with the next symbol.

In the guard interval, it is not necessary to add a signal, that is, the guard interval is a blank transmission time. However, in the OFDM system, due to multipath propagation, the blank guard interval will cause inter-channel interference (ICI), which will cause the gap between the sub-carriers. The destruction of orthogonality causes interference between different sub-carriers.

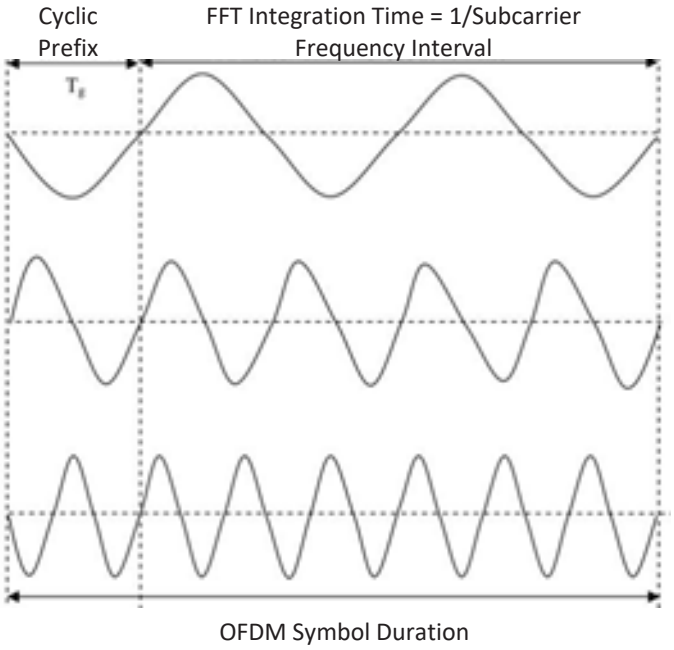


Figure 3.2 OFDM symbols with cyclic prefix

As a result, the original OFDM signal with a width of T can be periodically extended, using the extended signal as a guard interval, as shown in Figure 3.2. The signal in the guard interval is called the cyclic prefix (Cyclic prefix). The signal of the cyclic prefix is the same as the part of the tail time of the OFDM signal. The OFDM symbol transmitted in this way is at the receiving end, and the part with the initial time is first removed, and the remaining part

with a width of T is Fourier transformed, and then demodulated. In the process of time domain convolution, the periodicity of OFDM symbols can be well protected by adding a cyclic prefix, and the ICI of other subcarriers is reduced or eliminated.

4.1 ACO-OFDM System Simulation

This article uses Simulink software to build an ACO-OFDM system. The model is shown in Figure 4.1. The source sampling time is 0.001, 128 sub-data carriers, 512-point FFT, cyclic prefix length 16,1024 OFDM symbols, and the bit error rate statistics module accepts Extension 131076.

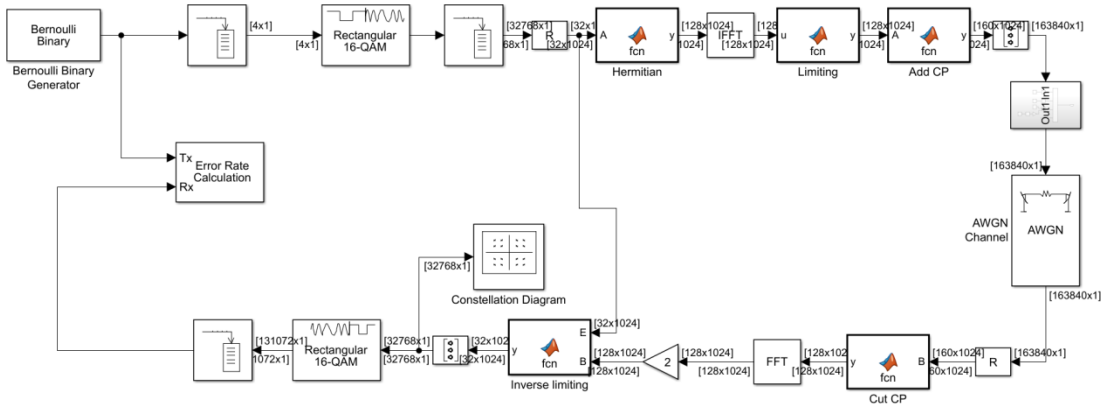


Figure 4.1 ACO-OFDM system model

Set the SNR of the Gaussian white noise module to 40, run the simulation, and get the ACO-OFDM symbol constellation diagram as shown in Figure 4.2.

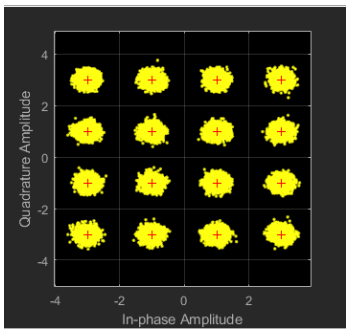


Fig. 4.2 star constellation

It can be observed from Figure 4.2 that the ACO-OFDM technology still decomposes the

data stream into several independent sub-data streams, which greatly improves the spectrum utilization and maintains the orthogonality of each sub-data stream. As shown in Figure 4.2, the ACO-OFDM system mapping in this article uses 16QAM modulation. After the transmitter transmits the signal, the constellation diagram after removing the cyclic prefix, FFT, and inverse clipping after the channel (direct link and Gaussian white noise) is removed. The distortion is small.

4.2 Bit Error Rate Simulation

This article then carried out a comparison simulation of the bit error rate for BPSK, QPSK, 16QAM, ACO-OFDM, and PPM. The sampling time of the Bernoulli binary generator is 0.001, the constellation sequence is Gray, the simulation duration is 100s, and the bit error rate range: 0 —40, the simulation results are shown in Figure 4.3.

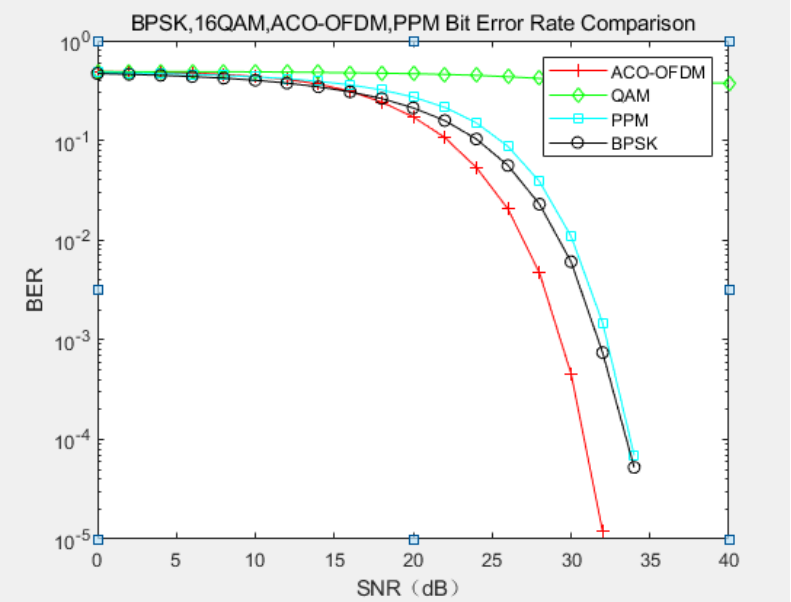


Figure 4.3 BPSK,QPSK,16QAM,ACO-OFDM,PPM bit error rate comparison

As shown in Figure 3.3, the original signal transmitted by the source is transmitted to the visible light communication channel (direct link, Gaussian white noise) after different

modulation methods, and after demodulation, the ACO-OFDM system has the lowest bit error rate, and the PPM system The bit error rate of the BPSK system and the QPSK system is basically the same. The 16QAM system has the highest bit error rate and with the increase of the signal-to-noise ratio, the decrease is not large. So in terms of the bit error rate, the modulation method adopts ACO-OFDM technology. The visible light communication system has the best performance.

5. Channel Coding

Due to the interference and fading in the channel during the transmission of the digital signal, error codes are generated in the transmitted data stream, thereby causing signal distortion at the receiving end. Therefore, through the link of channel coding, the digital signal adopts error and error detection coding technology, and the data stream is processed accordingly, so that the system has a certain error correction ability and anti-interference ability, which can greatly avoid bit errors in the code stream transmission happened.

Channel coding technology has been widely used in mobile communications and wired communications. There are few researches on channel coding in visible light communication systems. Therefore, the research and implementation of channel coding in visible light communication systems are particularly important for improving the performance of visible light communication systems. Below we briefly introduce several main coding methods used in simulation research.

5.1 Linear Block Code

Linear block codes are represented by (n, k, d), n is the code length, d is the mi nimum distance of the code, and k is the length of the information group. Linear bloc k code encoding requires two matrices, one is the generator matrix, represented by G, and the other is the check matrix, represented by H. Find a set of code words in cod e words $g_{k-1}, g_{k-2}g_{k-3}, \cdots, g_1, g_0$, So that all code words can be represented by this set of numbers $C = m_{k-1}g_{k-1} + m_{k-2}g_{k-2} + m_{k-3}g_{k-3} + \cdots + m_1g_1 + m_0g_0$. These codewords form t

he generator matrix G of the linear block code;

The error correction capability of linear block codes is closely related to the minimum distance d. In the linear block code encoding process, the setting of the d value requires special attention:

（1）When the number of random errors detected is e， $d \geq e+1$;

（2）When the number of corrected random errors is t, $d \geq 2t+1$;

（3）When correcting the number of errors t and detecting the number of errors e at the same time（ $e \geq t$ ）, $d \geq t+e+1$;

（4）When correcting the error number t and deleting the error number e at the same time, $d \geq 2t+e+1$.

5.2 BCH Code

The BCH code is a linear block code in a finite field, which has the ability to correct multiple random errors. Given any finite field $GF(q)$ and its extended field $GF(q^m)$, Where q is a prime number or power of a prime number, m is a positive integer.

The most commonly used BCH code is the binary BCH code. All symbols of the binary BCH code are composed of 0 and 1, which is convenient for the realization of the hardware circuit. The binary primitive BCH code has the following important parameters:

1. Yard length: $n = 2\hat{}m-1$;

2. Check bit length: $n-k \leq m*t$;

The generator polynomial of the BCH code is the product of the 2t smallest polynomial least common multiple of $GF(q\hat{}m)$. The generator polynomial of the BCH code with error correction capability t is $g(x) = LCM\left\{m_1(x), m_2(x), \cdots, m_{2t-1}(x), m_{2t}(x)\right\}$, Where LCM represents the least common multiple, m(x) is the smallest polynomial.

Known from the polynomial theory, if the element a^i in the finite field $GF(2\hat{}m)$ is m-th degree, which is approximately the root of the polynomial $m_i(x), (a^i)^2, (a^i)^4, (a^i)^8, \cdots$ is also the root of $m_i(x)$, $(a^i)^2, (a^i)^4, (a^i)^8, \cdots$ is called the conjugate root system. If two

roots are conjugate, they have the same minimum polynomial. Therefore the generator polynomial:

$$\begin{aligned} g(x) &= LCM\left\{m_1(x), m_2(x), \cdots, m_{2t-1}(x), m_{2t}(x)\right\} \\ &= m_1(x) * m_3(x) * \cdots * m_{2t-1}(x) \end{aligned} \quad (7)$$

Through the above steps, the generator polynomial of the BCH code can be obtained. The information can be encoded by obtaining the generator polynomial.

5.3 Convolutional Code

Convolutional code (n,k,N) is a non-blocking code. Although convolutional codes also encode k-bit information segments into n-bit code groups, the supervision symbols are not only related to the current k-bit information segments, but also the same as the previous $(N-1)$ information. Paragraph related. In other words, the supervision symbols in a code group supervise N information segments. Usually N is called the coding constraint degree, which means the number of code segments related to each other during coding. The code rate is defined as: $R_c = k / n$。

Figure 5.1 is the (2,1,2) convolutional code encoder used in this paper with output code length n=2, degree of constraint N=2, input bit k=1. The current input signal, the previous state signal and the previous two state signals are subjected to a modulo two addition operation as the first output, and the current input signal and the previous two state input signals are used as the second output.

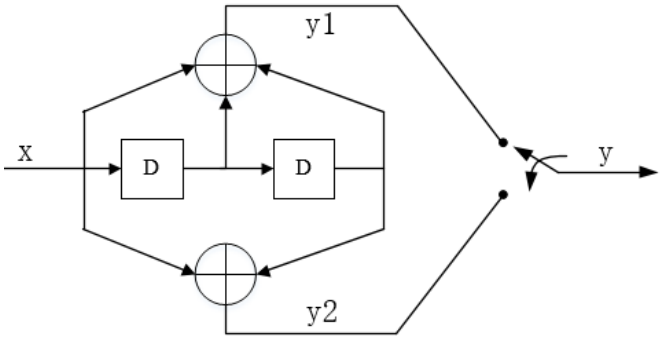


Figure 5.1 (2,1,2)Convolutional code encoder

This article adopts Viterbi algorithm as the decoding of convolutional codes. The basic principle of this algorithm is to compare the received signal sequence with all possible transmission signal sequences, and select the sequence with the smallest Hamming distance as the current transmission signal sequence.

6. ACO-OFDM System Coding Simulation

This article uses simulink to build an ACO-OFDM system using linear block codes, BCH codes and convolutional codes, and compares the bit error rates of different coding methods. Linear block code generation matrix: [[1 1 0; 0 1 1; 1 1 1; 1 0 1] eye(4)]; convolutional code: poly2trellis(7, [171 133]); BCH code word length 15, The message length is 5, and the code rate is 1/3. The simulation results are shown in Figure 6.1.

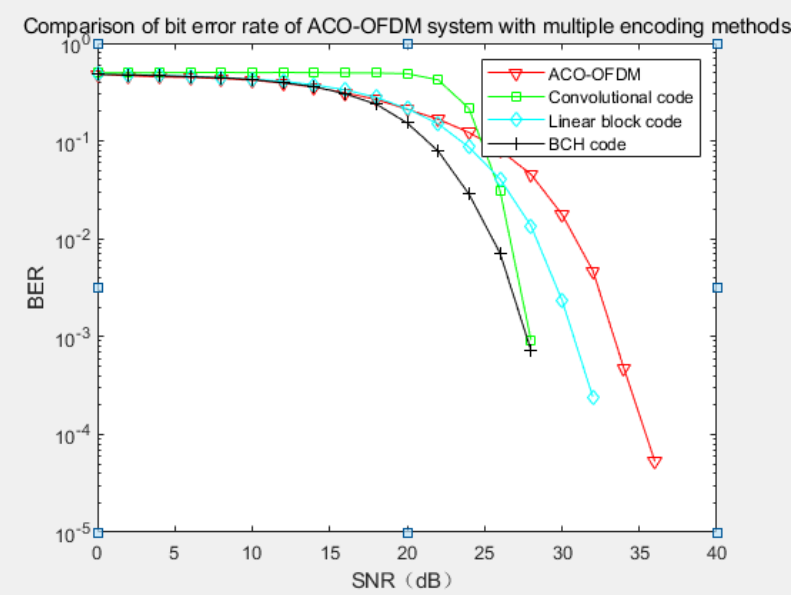


Figure 6.1 Comparison of bit error rate of multiple encoding methods

It can be seen from Figure 6.1 that after the signal generated by the source is subjected to different encoding methods, ACO-OFDM systems, and corresponding decoding, when the channel signal-to-noise ratio is greater than 28dB, the convolutional code has the fastest convergence speed and errors. The bit rate is the lowest; the second is the case of using BCH coding, and the linear block code has the worst improvement ability on the bit error rate.

7. Concluding Remarks

This article briefly introduces the general system framework and channel model of visible light communication, and explains the principles of ACO-OFDM system and PPM system, Hermitian mapping and cyclic prefix. Simulink is used to simulate the visible light communication system based on ACO-OFDM technology and PPM modulation technology, and compares and analyzes the bit error rate with BPSK, QPSK and 16QAM. The experimental results verify the feasibility of ACO-OFDM technology in visible light communication systems and the relatively low bit error rate performance. Combined with ACO-OFDM technology, compared with other modulation methods, it has strong anti-fading ability, high spectrum utilization, suitable for high-speed data transmission, simple modulation

and demodulation, no complex equalization technology at the receiving end, and strong anti-inter-symbol interference (ISI) ability, etc. Advantages, choosing ACO-OFMD technology as the modulation method of the visible light communication system has more advantages. After that, this article introduces the principle and compares and simulates the three coding methods of linear block code, BCH code and convolutional code. It is concluded that the ACO-OFDM system based on convolutional code channel coding is more suitable for the choice of visible light communication system.

References

1. Chen N. Visible light positioning and communication fusion system based on ACO-OFDM [D]. Nanjing University of Posts and Telecommunications.2019.6
2. Yin N M. Research on Dimming Technology of VLC System Based on OFDM [D]. Beijing University of Posts and Telecommunications.2018.3.14
3. Cai C. Performance of asymmetric truncated optical OFDM system based on discrete Hartley transform [D]. Beijing University of Posts and Telecommunications.2016.3.3
4. Jing Y L. Research on optical wireless communication system based on aco-ofdm.[D]. Harbin Engineering University.2016.3
5. Zhou J. Key technologies of optical frequency division multiplexing system [D]. Beijing University of Posts and Telecommunications.2018.5.29
6. Pathak P H,Feng X T,Hu P F,et al. Visible Light Communication,Networking,and Sens-ing A Survey,Potential and Challenges[J].IEEE Communications Surveys and Tutorials,2015,17(4):2047-77
7. Lu Yu,Min Zhang,Dahai Han.FPGA-based design and implementation of an OFDM transmitter for VLC system..Asia Communications and Photonics Conference 2013..
8. Trio Adiono，Yulian Aska，Design of an OFDM System for VLC with a Viterbi Decoder[J].IEIE Transaction on Smart Processing and Computing, 2017, 6(6): 455-465.
9. Komine T, Nakagawa M. Fundamental Analysis for Visible Light Communication System Using LED Light[J]. IEEE Transactions on Consumer Electronics, 2004, 50(1):100-107.

10. Liu H, Darabi H, Banerjee P, et al. Survey of Wireless Indoor Positioning Techniques and Systems [J]. IEEE Transactions on Systems Man&Cybernetica Part C, 2007, 37(6):1067 一 1080.
11. Hou Y, Xiao S, Zheng H, et al. Multiple Access Scheme Based on Block Encoding Time Division Multiplexing in an Indoor Positioning System Using Visible Light [J]. Journal of Optical Communications and Networking, 2015, 7(5): 489-95.
12. Pasha M A, Yuen C, Hassan N U, et al. Indoor Positioning System Designs Using Visible LED Lights:Performance Comparison of TDM and FDM Protocols [J]. Electronics Letters, 2014, 51(1):72-74.
13. Kuo Y S, Pannuto P, Dutta P. Demo: Luxapose: Indoor Positioning with Mobile Phones and Visible Light[C]. Proceedings of the International Conference on Mobile Computing and NETWORKING, F, 2014:55-61.
14. Gu W, Aminikashani M, Deng P, et al. Impact of Multipath Reflections on the Perfor-mance of Indoor Visible Light Positioning Systems [J]. Journal of Lightwave Technology, 2016, 34(10):2578-2587.
15. Thomas Q, Wang Y, Ahmet S, et al. Analysis of an Optical Wireless Receiver Using a Hemispherical Lens with Application in MIMO Visible Light Communications [J]. Journal of Lightwave Technology (S0733-8724), 2013, 31(11): 1744-1754.

About the Author: First Author: Chen Peng far (1997-), male, master, mainly engaged in research communications of visible light.

E-mail: 554855487@qq.com

Contact information：18742059216

Corresponding author: Li Ping (1969-), female, Professor, master, main research directions for optical communications and intelligence systems.

E-mail: liping@dlpu.edu.cn

Room 313, Photonics Institute, Dalian University of Technology, No. 1, Light Industry Garden, Ganjingzi District, Dalian.

High-throughput visible image transmission design based on the X-CT root 3D reconstruction system

Xin Guan[1], Jinpeng Wang[1,*], Yang Zhou[1], Teng Zhang[1], Kemo Jin[2,*], and Nianyu Zou[1]

[1] Information Science and Engineering College,
Dalian Polytechnic University, Dalian 116034, China
wangjp@dlpu.edu.cn

[2] Key Laboratory of Plant-Soil Interactions, Ministry of Education, China Agricultural University, Beijing 100193, China
kemo.jin@cau.edu.cn

Abstract. X-ray computed tomography (X-CT) technology can help us observe plants' root structure in a three-dimensional, non-destructive manner under underground soil conditions. Between the CT imaging terminal and the post-image 3D reconstruction processing terminal, there is a massive amount of image data for transmission, and the process requires a lot of transmission time, which is challenging to meet the characteristics of large-scale and high-throughput research. Today's medical transmission bus Technology does not meet the low cost and high convenience of future plant phenotype research. For large-scale population genetic analysis after the three-dimensional phenotype of plant roots, such as quantitative trait locus mapping (QTL) research, it is necessary to have high the technology of flux image data transmission is added based on the three-dimensional analysis of the root system. In this paper, under the conditions of underground soil, a high-throughput visual light image transmission system for X-CT plant root CT scanning imager is proposed. It uses the characteristics of fast data transmission speed and large throughput of visible light, and image transmission can be useful. Reduce the time of image data transmission between the upper imaging terminal and the lower image processing terminal and reduce the later 3D reconstruction research time cost.

Keywords: X-CT, visible light communication, image transmission.

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2306765

1 Introduction

To further keep pace with global environmental change and world population growth, crop yields need to increase dramatically. Based on modern agricultural technology, real-time monitoring, and improvement of plant root structure changes are expected to improve water and nutrient absorption and utilization efficiency. In this regard, some researchers have proposed the second green revolution and put more plant phenotypic research into improving plants go to the underground root structure [1]. Due to the invisibility and spatial complexity of plant roots under the condition of underground soil, it is a challenge to observe and process high-throughput in-situ soil root images. With the development of the medical imaging field, it is possible to non-invasively study the three-dimensional phenotypic structure of the root system in the soil. X-CT technology and MRI technology gradually used to study the hidden parts of plants [2]. For X-CT technology, it has great potential in the three-dimensional analysis of plant root phenotype. Compared with MRI, the equipment cost of a non-medical CT scanner will be lower [3]. The medical CT image is a two-dimensional image sequence of the target obtained through the imaging terminal, and a three-dimensional model obtained through the calculation of the lower computer and depicted with a specific software [4]. In recent years, by continuously adapting and optimizing the imaging terminal, medical CT has reduced the scanning time compared with the past. Using higher voltage and current in a specific X-CT scanner, the contrast between plant roots and soil has increased, making Later, better results obtained in root-soil segmentation [5]. For CT image segmentation or reconstruction at the processing end [6], it is necessary to collect a large amount of data from the host computer and complete, precise algorithm calculations.

At the beginning of the 21st century, with the rapid development of computer information technology, wireless networks are gradually used in the modern society. Communication-based on wireless visible light technology can bring higher frequency band bandwidth to the system, but for low-frequency systems, the signal attenuation is not obvious [7]. As the devices in the integrated photodetector manufacturing industry become cheaper, the high-speed, low-cost, short-distance, and highly flexible white LED visible light communication system has aroused scholars' in-depth research. Since 2013, China has included visible light communication technology in the 863 project plan, which has promoted the progress of visible-light wireless communication technology, so that white light LEDs are now widely used in lighting, display, signal

emission, and other fields. Visible light communication systems have high application prospects in high-speed communication, smart transportation, indoor positioning, and other fields [8]. Thanks to the high bandwidth and fast transmission rate of white light LED visible light communication, the system can realize some communication application scenarios that meet high throughput.

During this period, the image data transmission needs to consider the conditions of accuracy and high throughput. In recent years, scholars have optimized wireless communication algorithms, continuously optimized the signal-to-noise ratio, and expanded the application field of wireless communication [9-11]. For the data exchange between a specific plant root CT scanner and the lower processing end, the commonly used medical PCI-E bus communication protocol cannot satisfy the low cost and portability characteristics of future plant phenotype detection.

In this paper, under the research conditions of underground soil, a high-throughput method of visible light image transmission for the root system 3D reconstruction system of X-CT technology is proposed, which provides a kind of upper and lower computer for specific high-throughput root CT scanners in the future. The design of visible light image data transmission utilizes the characteristics of fast data transmission speed and large throughput of visible light for image transmission, which can effectively reduce the time of image data transmission between the imaging terminal and the lower computer, and reduce the research time cost of the overall system.

2 Research content

With the rapid development of plant phenotypes, CT detection techniques for roots in specific soil environments have emerged. X-CT technology uses the data to calculate the linear attenuation coefficient of X-rays and plant roots to form a visual image, create cross-sections, stack them, and reconstruct the soil's plant root structure non-destructively. It has the advantages of high spatial resolution and fast scanning speed. The disadvantage is that the imaging end and the processing end have to communicate and exchange many CT image data after scanning. At present, the single technical solution to solve this aspect gradually revealed. In medicine, there is a PCI-E bus communication protocol for the transmission of massive CT data for clinical diagnosis, which can solve the problem of three-dimensional reconstruction of extensive data and high precision in the CT system. Simultaneously, there is still a need for a specific transmission method for underground plant root CT scanners, which aims to realize a

low-cost, high-throughput, and low-loss image transmission system at the imaging and the reconstruction processing end. The position of the visible light image transmission system designed in this paper is shown in Figure 1, which undertakes the high-throughput image transmission work between the upper imaging end and the lower reconstruction end.

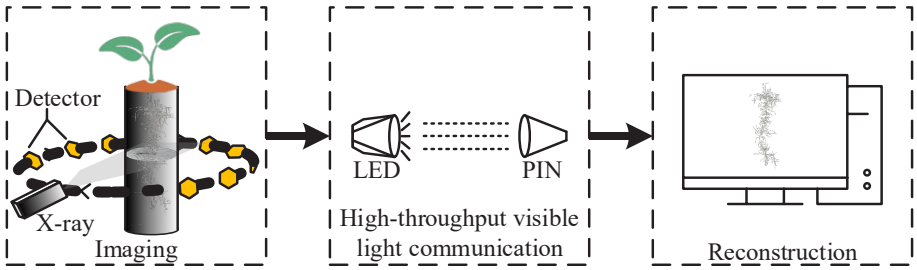


Fig. 1. Location map of the visible light image transmission system

3 Overall design structure

This system based on the white light LED to carry out the visual light communication design between the upper imaging end and the lower reconstruction end of the image information. The CT imaging terminal's image information communicated through the serial port and the signal modulated to the LED white light source. The visible light is transmitted through the optical channel and received by the photodiode. The signal is demodulated and sent to the serial port of the lower computer to recover and display. High-throughput visible light image transmission and communication between the scanner's imaging end and the lower reconstruction end. The overall system structure divided into two parts, the transmitting end of the upper CT imaging and the lower reconstruction. The block diagram is shown in Figure 2.

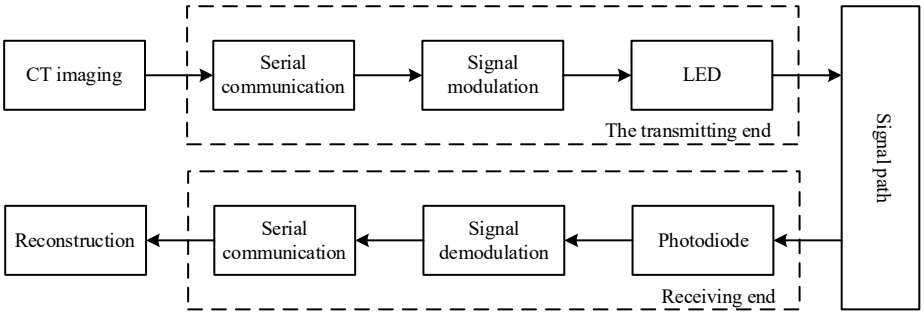


Fig. 2. Schematic diagram of the visible light image transmission system

4 System principle analysis

4.1 The choice of white LED

In the high-throughput visible light image transmission circuit, device selection, and the overall design of the circuit in the future will directly or indirectly affect the performance indicators of communication. This communication system mainly involves the selection of LED and photoelectric detection circuits.

White LED is the core component of the emitter of the entire system. According to the manufacturing method of white LED, it can be divided into two types. One is a single light-emitting chip, and the other is a multi-light-emitting chip based on three primary colours. They emit white light on different principles. Communication performance is also different. The current research shows that the outer surface of a single chip emitting blue LED coated with a yellow phosphor, and the white light emitted by the complementary white light has better luminous efficiency and modulation rate, so the choice of LED will directly affect this high-flux addable light The communication quality of the image transmission system. Figure 3 is a schematic diagram of a single-chip white LED.

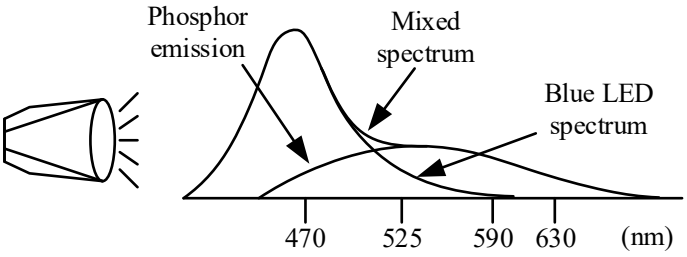


Fig. 3. Schematic diagram of a white LED

4.2 Selection of photodetection diode

In the visible light communication system designed here, the photodetector and circuit's construction directly or indirectly affects the entire communication system's performance. The photodetector diode's function is to convert and restore the optical signal received from the transmitter into an electrical signal. The core component of the receiving end is the photodetection diode. Facing the high-throughput visual light image transmission system based on CT root system, this type of photodetection diode circuit is required:

(1) The response speed is fast, the frequency bandwidth is comprehensive, and it can adapt to the environment of high-throughput transmission of CT images.

(2) The photoelectric conversion circuit is simple, and the photoelectric conversion efficiency is high. It can convert the optical signal emitted by the transmitter to the greatest extent. The photoelectric conversion efficiency determined by the current responsiveness R_I, where I_s is the output current and P is the incident light power.

$$R_I = I_S / P \tag{1}$$

(3) The photoelectric detection circuit has good linearity, and the electrical signal is less interfered with by the internal noise of the circuit.

(4) The photodetector diode's frequency response range should match the white LED's luminous effect at the emitter.

Nowadays, photodetection diodes have been applied to equipment in many fields, including CT scanners. The typical and commonly used silicon photodetection diode has high photoelectric conversion efficiency, a simple design of the front-end amplifying circuit, and a responsivity range of 0.3~0.45A/W. It has high quantum

efficiency in the visible light field and is satisfied with this design. In terms of practical application, because of its low price, it is also widely used. Figure 4 shows the response curve of a typical silicon photodiode.

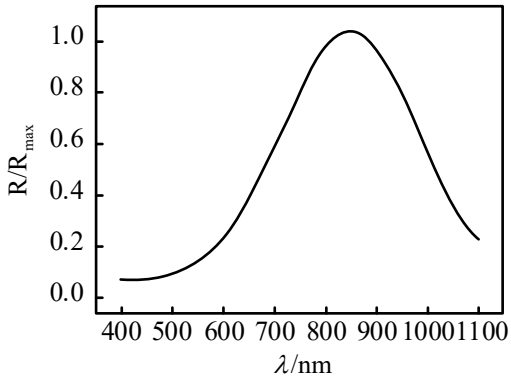


Fig. 4. Silicon photodiode response curve

5 Circuit design

5.1 Transmitter circuit design

The visible-light image transmission transmitter's design based on the X-CT technology root system 3D reconstruction system roughly consists of three parts: USB serial communication module, signal amplifier circuit module, and white light LED drive circuit module. Among them, the white light LED drive circuit is the core part of the transmitting end, and its working principle is to convert the modulated direct current signal into the light signal emitted by the white light LED and amplify the modulated electrical signal. In the visible light communication image transmission system, white light LED is used as the light-emitting signal source. Because of its high-frequency response characteristics, the signal peak value and power consumption heating under long-term DC bias should be considered when designing the white light LED drive circuit. Controllable factors.

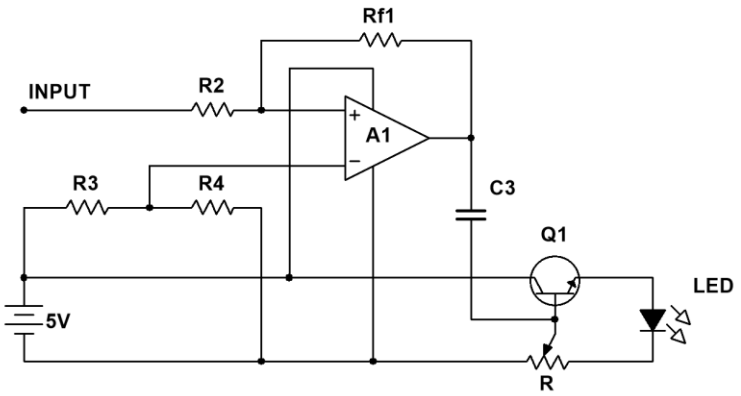


Fig. 5. Schematic diagram of a white LED drive circuit

The transmitting end of this design adopts on-off keying (OOK) signal modulation technology. After the central control chip's digital signal modulated and output, the purpose is to modulate the image binary signal flow at the LED drive circuit's front end to the white light LED to emit high frequency Flashing light signal. OOK modulation is a practical optical communication modulation technology. It controls the transmission frequency according to the amplitude of the signal. When the signal amplitude is high, the carrier frequency is transmitted, which is "on", represented by the digital signal "1"; otherwise, the carrier frequency not transmitted. Frequency, which is off, is represented by a digital signal "0". Figure 6 is the OOK modulation principle diagram. The carrier used to transmit digital information, and the carrier signal used to control the on-off changes.

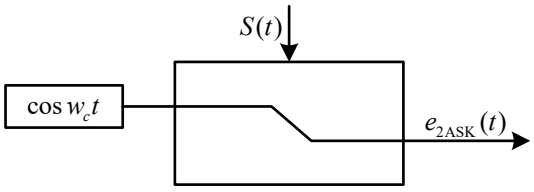


Fig. 6. OOK modulation schematic diagram

The overall system principle diagram of the transmitting end shown in Figure 7. Among them, C1, C2, C3 are filter capacitors. The amplification factor of amplifier A1 is shown in Formula 2.

$$A1 = \frac{R_{f1}}{R_2} \tag{2}$$

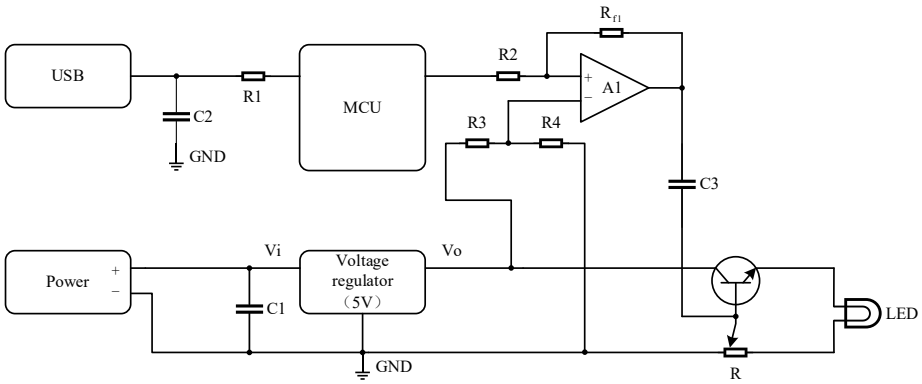


Fig. 7. Schematic diagram of the transmitter circuit

5.2 Receiver circuit design

For the receiving end, it is mainly responsible for receiving optical signals, converting the optical signals into electrical signals through silicon photoelectric detection circuits, and finally returning them to image data streams through operational amplifiers and digital-to-analogue conversion circuits. The receiving end design roughly divided into two modules: the silicon photoelectric detection and receiving amplifier circuits. When the optical signal transmitted in the space channel, it will be interfered with and attenuated by background noise, and the dark current in the photodetection diode must be considered. Therefore, when designing the photoelectric detection circuit, it is necessary to consider the amplification of the electrical signal's amplification and the filtering of the dark current and other noises. Figure 8 is a schematic diagram of a silicon photoelectric detection circuit.

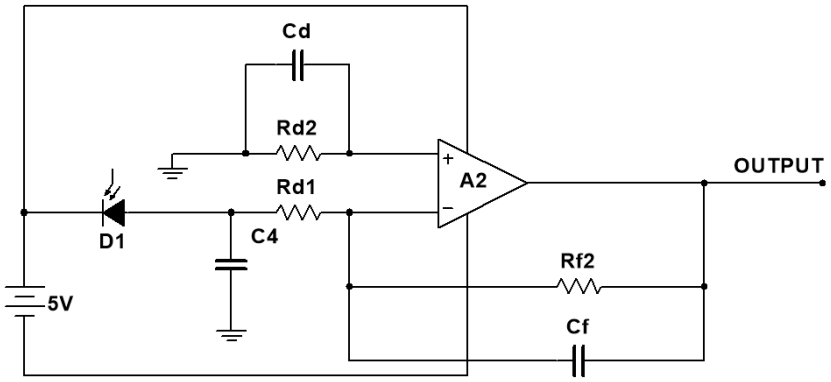


Fig. 8. Schematic diagram of silicon photoelectric detection circuit

R_{f2} is used as the feedback resistor. Because the silicon photodiode's resistance is relatively large, it is easy to oscillate at the feedback resistor and increase noise, so the resistance cannot be selected too large. C_f is used as a feedback capacitor to control the frequency response of the receiving circuit to prevent self-excitation. Generally, this capacitor is several picofarads. The amplification factor of amplifier A2 is shown in Formula 3.

$$A2 = -\frac{R_{f2}}{R_{d1}} \tag{3}$$

The overall system principle diagram of the receiving end shown in Figure 9. Among them, R7 and C5 composed of low-pass filters. Two functions realized in the circuit of this design. One is to amplify the photodetection diode's electrical signal, and the other is to filter. R_{f2} and C_f form a feedback circuit and an operational amplifier form a voltage negative feedback amplifier circuit, making the circuit have the advantages of low noise and frequency bandwidth. The feedback circuit controls the receiving end amplifier circuit's frequency response, and the upper cut-off frequency represented by f. VDD is the working voltage of the chip, in which the resistance value of R7 pull-up resistance is determined by the driving device voltage V_O, current I_O, the sum of leakage current of all the driving gate outputs high current at ordinary times I_{OZ} and the total input current I_{IL} connected to the lower end of the pull-up resistance.

$$f = \frac{1}{2\pi R_{f2} C_f} \tag{4}$$

$$\frac{V_{DD}-V_{O(max)}}{I_{O(max)}-I_{IL}} \le R7 \le \frac{V_{DD}-V_{O(min)}}{I_{OZ}+I_{IL}} \tag{5}$$

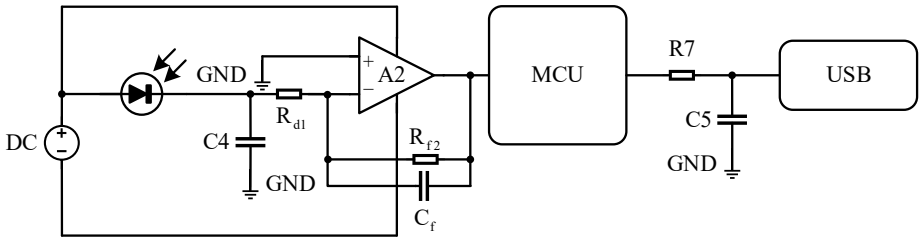


Fig. 9. Schematic diagram of the receiver circuit

6 System transmission test

To simulate the CT imaging terminal's connection at the transmitting end, connect the USB to the upper computer and power on the transmitting end, connect the USB to the lower computer on the receiving end, and power on the receiving end. Under normal lighting conditions in the laboratory, the transmitter and receiver fixed at a certain distance, and the bit error rates at distances of 10cm, 20cm, 30cm, 40cm and 50cm and the high voltage signal at the receiving end counted separately. Figure 10 (a) (b) shown.

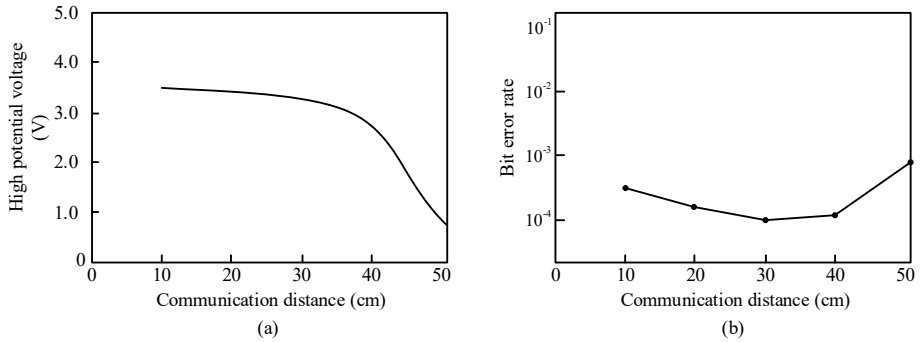
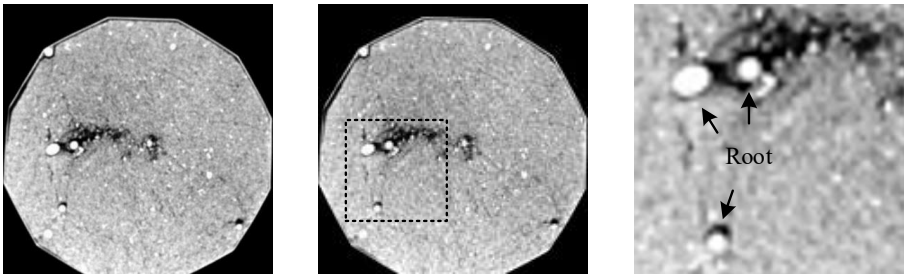


Fig. 10. (a) The relationship between communication distance and high-voltage signal at the receiving end, (b) The relationship between communication distance and bit error rate.

It can be seen from Figure 10 (a) that the high-level output at the receiving end of this system is related to the communication distance. The voltage values of the output

signals at the receiving end at different distances sampled, it can be seen from the change curve that when the communication distance is less than 40cm, the output voltage does not change much with the distance. When the communication distance is greater than 40cm, the output voltage drops smoothly. Figure 10 (a) can be inferred that the best communication distance of this system is within 40cm.

It can be seen from Figure 10 (b) that when the communication distance of this system is less than 30cm, the bit error rate decreases with the increase of the communication distance. When the communication distance is greater than 30cm, the bit error rate increases with the communication distance increase. And increase. In summary of Figure 10 (a) after experimental and theoretical analysis, there is an optimal communication distance with a minimum error rate 10^{-4}, and this system is 30cm. Figure 11 shows the transmission diagram of the transmitting end and the receiving diagram of the receiving end under the optimal communication distance of 30cm.



(a) The original image of the sender

(b) The image received by the receiver

(c) The root characteristics

Fig. 11. Image transfer test diagram

7 Conclusion

In the future, high-throughput three-dimensional analysis of roots in soil will gradually become a hot issue in plant phenotyping. High-throughput technologies will be integrated into the platform for three-dimensional root detection to achieve high-throughput. In the follow-up, large-scale 3D reconstruction quantitative research, the high-throughput visual light image transmission system for root CT scanning imager is a brand new image transmission option for future root phenotype detection. Compared with the traditional medical CT machine's image transmission bus solution, this design considers the high-throughput image information transmission, and the total cost will

significantly be reduced. Future high-throughput plant root phenotype image data research provides a brand new transmission option, which can be used on specific root CT scanners.

References

1. Paya A., Silverberg J., Padgett J.: X-ray computed tomography uncovers root–root interactions: quantifying spatial relationships between interacting root systems in three dimensions. Frontiers in plant science, 6: 274 (2015).
2. Wang, N.: Research on Three-Dimensional Reconstruction and Visualization of Maize Roots Based on Magnetic Resonance Imaging and VTK. Zhejiang University, Hangzhou (2013).
3. Atkinson J., Pound M., Bennett M.: Uncovering the hidden half of plants using new advances in root phenotyping. Current opinion in biotechnology, 55 1-8 (2019).
4. Mooney S., Pridmore T., Helliwell J.: Developing X-ray computed tomography to non-invasively image 3-D root systems architecture in soil. Plant and soil, 352(1) 1-22 (2012).
5. Teramoto S., Takayasu S., Kitomi Y.: High-throughput three-dimensional visualization of root system architecture of rice using X-ray computed tomography. Plant Methods, 16 1-14 (2020).
6. Guan X., Wang J., Zhou Y.: Study on 3D Reconstruction of Plant Root Phenotype Based on X-CT Technique. International Conference on Green Communications and Networking. Springer, Cham, 182-192 (2020).
7. Wang J., Ye Z., Cao F.: Cooperative distributed antenna transmission method based on co-channel interference in 5G mobile communication system. Journal of Jilin University (Engineering and Technology Edition), 48(1) 333-341 (2020).
8. Ren Y.: Research and implementation of image transmission system based on visible light communication. Yanshan University (2017).
9. Wang J., Ye Z., Tarun M.: A Novel Linear Antenna Synthesis for Linear Dispersion Codes Based on an Innovative HYBRID Genetic Algorithm. Symmetry, 11(9),1176 (2019).
10. Wang J., Ye Z., Jeremy G.: A Power Control Algorithm Based on Chicken Game Theory in Multi-Hop Networks. Symmetry, 11(5),718 (2019).
11. Wang J., Cao F., Zou N.: Multi carrier system joint receiving method based on MAI and ICI. Journal of Jilin University (Engineering and Technology Edition), 41(6), 301-305 (2018).

Underwater image transmission system based on visible light communication

Zhou Yang, Wang Jinpeng, Guan Xin, Zhang Teng, Li Bo, Zou Nianyu

Dalian Polytechnic University, 116034, Dalian, China
wangjp@dlpu.edu.cn

Abstract. In order to improve the image transmission quality of underwater visible light communication, this paper designs an optimal scheme of the underwater image transmission system based on blue visible light communication. In the hardware circuit module, the trans-impedance amplifier circuit of the T-feedback network is used to improve the signal receiving performance. Due to the bad feature of the underwater channel, a fast likelihood blind detecting algorithm was adopted to detect, collect and process the data in this article. The method proposed in the paper is able to reduce BER (Bit Error Rate) as well as speed up computation. It can be seen from the experimental results that the scheme of the system optimization effectively decreases DC bias of the photoelectric receiver, as well as decreases the computation complexity and improves the transmitting efficiency under the condition of meeting the requirements for the specific BER of the receiving signal.

Keywords: Underwater visible light communication, Image transmission, Detection algorithm.

1 Introduction

The ocean occupies about 71% of the earth's surface area, and it plays a significant role in the aspects of national security, economy and environment etc. The communication technique, a means for exchanging information, can be a key, which is the basic and supporting power for production activities in different fields of the marine environment. With the rapid and stable development of the modern society, there are more and more marine production activities, consequently, human being has a higher demand to the communication techniques. The transmissions of underwater image data has been widely and deeply applied and researched in many fields, such as exploration of deep-sea resources, monitoring of seabed ecological conditions, underwater rescue, ocean war preparation, and development of intelligent underwater robots [1-4]. At the same time, as the support technology of underwater image transmission, the research on underwater wireless communication technology is also expanding. Currently, the traditional underwater wireless communication technologies mainly include underwater acoustic communication, underwater radio frequency communication and underwater visible light communication. Underwater acoustic communication technology has low data transmission rate and slow

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2306758

propagation speed of sound waves, so there is a high communication delay. In addition, sound waves will also affect the safety of marine life; The communication distance of underwater radio frequency communication technology is extremely short, usually only a few meters, and the communication cost is high. In 2016, the experimental results of the research team of Lloret J et al. showed that radio frequency signals can be transmitted at a rate of up to 100Mbps, but the transmission distance is extremely short, and the effective communication distance is only about 15cm[5]; Compared with the above two methods, the underwater visible light communication technology uses blue-green visible light with a wavelength of 450nm-550nm for data transmission in seawater, which has a very high data transmission rate, low link delay and low implementation cost [6-8]. On November 13, 2020, the "Struggle" manned submersible and the deep-sea video lander "Canghai" and "Lingyun" carried out joint operations in the Mariana Trench, among which the "Struggle" and the "Canghai" realized information exchange through visible light communication technology, and implemented the world's first live TV broadcast on the ocean floor of 10,000 meters [9].

The underwater channels are harsh, and many factors such as the optical properties, absorption and scattering effects of seawater will cause interference to signal transmission and affect communication quality. In order to solve the problem of low efficiency of traditional underwater wireless communication methods and improve the transmission effect of visible light communication, this paper proposes an optimized scheme of underwater image transmission system based on blue visible light communication. Firstly, the paper presents a system model for image transmission using blue visible light communication Secondly, in order to solve the problem of weak light signal the receiving end uses the T-feedback network trans-impedance amplifier circuit to improve system performance. According to the transmission characteristics of the underwater channel, the fast likelihood detection algorithm is used to optimize the detection of the optical signal to improve the accuracy and speed of the received signal. Finally, the feasibility of the system is verified by analyzing the experimental results. The results show that, compared with the traditional underwater communication technology, The scheme of this paper is stable and reliable, and has practical significance for improving the safety and scientific of underwater activities.

2 System design principle and scheme

The underwater image transmission system is mainly composed of three parts: transmitter, underwater channel and receiver, as shown in Figure 1. At the transmitter end part, after the collected image information is encoded and modulated by the MCU, the modulated signal is modulated into the current of the LED drive circuit and emitted into the underwater channel in the form of blue visible light. The signal passing through the underwater channel is received at the signal receiving end, and a photodetector is used to convert the optical signal received from the transmitting end into an electrical signal. The signal is extracted by the fast likelihood blind detection

algorithm, and then the electrical signal is amplified by the operational amplifier circuit. The amplified electrical signal is demodulated and A/D converted to restore the original image data.

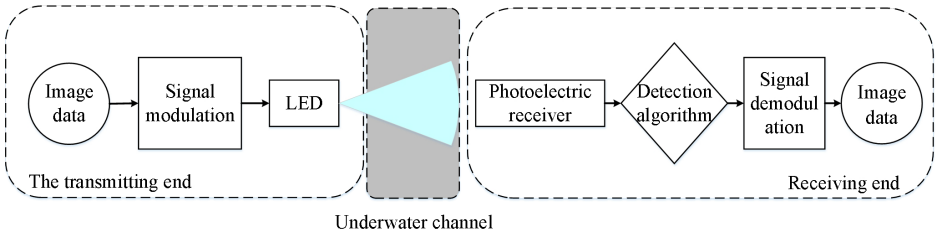


Fig. 1. System block diagram

3 System hardware design

3.1 Transmitter circuit design

After the image data is modulated, the transmitter drives the LED light source to convert electrical signals into optical signals for transmission. The blue LED drive circuit first uses an operational amplifier to form a voltage comparator to solve the problem of insufficient transistor drive capability. The function of the capacitor C1 is to increase the switching rate of the LED drive circuit, so that the high frequency performance of the drive circuit is improved. The effect of R2 is to balance the impedance at the input and the voltage. R3 and R4 protect the circuit and prevent short circuit to the ground. The schematic diagram of the drive circuit is shown in Figure 2.

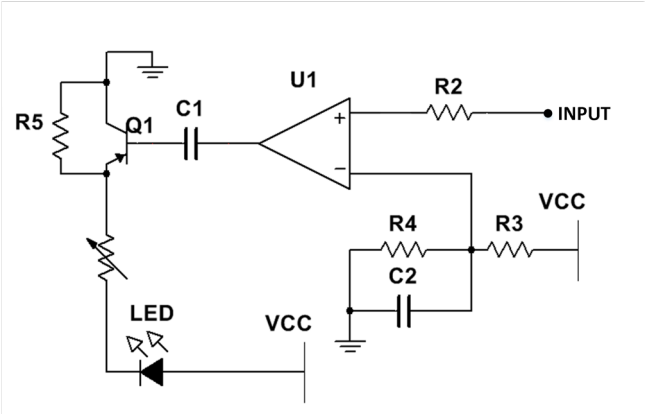


Fig. 2. Schematic diagram of drive circuit

3.2 Receiver circuit design

Since the signal generated by the photodiode is relatively weak, the receiving end generally uses an operational amplifier circuit to amplify the received signal. When the gain of the ordinary operational amplifier circuit is very large, the resistance of the feedback resistor and the matching resistor will also become very large. After the input bias current of the amplifier flows through the matching resistor, a significant voltage is applied to the photodiode. However, due to the sensitive nature of the sensor will produce a larger dark current, which will produce a DC bias at the output, which will affect the circuit performance. In order to ensure the quality of signal transmission and improve the gain level of visible light signals, the trans-impedance amplifier circuit designed by the T-feedback network is introduced at the receiving end. The schematic diagram of the trans-impedance amplifier circuit of the T-feedback network is shown in Figure 3.

In the embodiment, the circuit uses compensation resistor R5 to produce offset components caused by the input current of the operational amplifier, and its resistance value is the same as Rf. C2 is a bypass capacitor matched with the resistor R5 to reduce the noise introduced by the matched resistor R5. The T-feedback network is used to replace a single large-value feedback resistor, where R2 >> R3 >> R1, The equivalent impedance formula of the T-feedback network is:

$$R_{feq} = R_2 + R_3 + \frac{R_2 R_3}{R_1} \approx (1 + \frac{R_3}{R_1}) R_2 \tag{1}$$

The trans-impedance amplifier circuit of the T-feedback network used a smaller resistor R2 to replace the original larger feedback impedance, but achieved the same current gain. The resistance of the compensation resistor decreased with the decrease of the new feedback resistor R2, and the corresponding voltage applied to the two segments of the photodiode decreases, the dark current decreased, and the matching error between the feedback resistor and the compensation resistor is reduced.

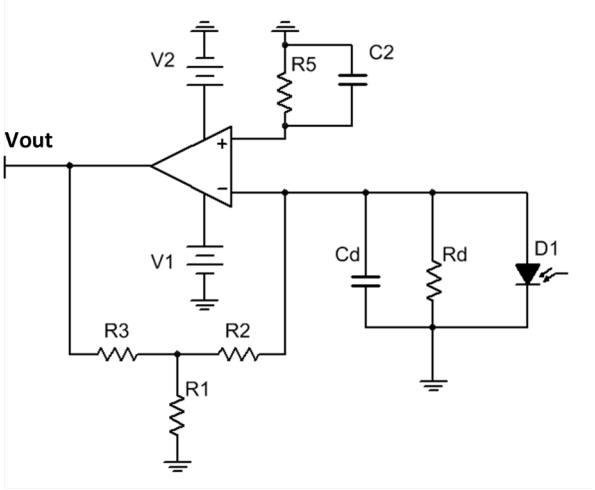


Fig. 3. Receiver circuit schematic diagram

4 System software part design

4.1 Signal modulation method

CAP (Carrier-Free Amplitude / Phase Modulation) modulation technology is based on the development of QAM (Quadrature Amplitude Modulation) technology. Its characteristic is that shaping filters and matched filters are used at the transmitting end and the receiving end to generate and separate the same direction and quadrature signals, so that the generated CAP signal has no carrier in the signal band, so this modulation method is called carrier-free modulation[10]. Since the visible light communication system is very sensitive to system non-linearity, CAP modulation will not cause the system PAPR (Peak to Average Power Ratio) to be too high. This is also the advantage of cap modulation in optical communication system[11]. The CAP modulation principle block diagram is shown in Figure 4.

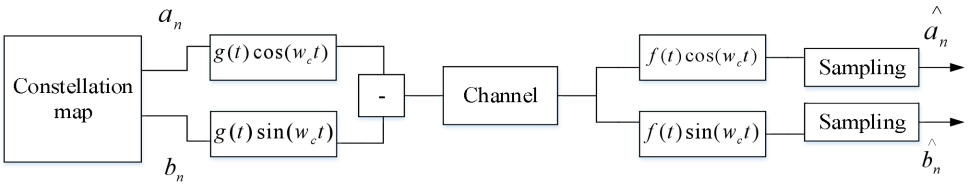


Fig. 4. CAP modulation principle diagram

4.2 Fast-likelihood blind detection algorithm

The indoor visible light channel communication distance is short and the channel is stable. However, the underwater visible light channel is different from the indoor visible light channel. Due to the harsh underwater environment, impurities and particles in the water, turbulence effects, water temperature changes, absorption and scattering effects of seawater and many other factors will interfere with the state of the underwater visible light channel. Therefore, the performance of the underwater visible light communication system is affected. For the transmission characteristics of underwater visible light channel, it is very important to select an appropriate signal detection algorithm.

Blind likelihood detection means that the receiver processes signal under the condition that the mathematical model and statistical parameters of the channel are not known, then it can complete the accurate estimation of the channel and the detection of the signal. Compared with the traditional MMSE (Minimum Mean Square Error) algorithm [12]、 MD (Mean Detection) algorithm[13-14] and blind detection algorithm, the fast likelihood blind detection algorithm has lower complexity and higher accuracy. When the channel conditions such as channel model and parameters are unknown, the fast likelihood blind detection algorithm can complete the signal detection and channel model estimation by processing the received signal.

The visible light channel function obeys the logarithmic normal distribution [15-16], and the probability density function of the received signal r with respect to the transmitted signal s is:

$$f_x(r \mid s, G) = \frac{1}{(\sqrt{2\pi\sigma_n^2})^L} \exp(-\frac{\| r - P_s G s \|^2}{2\sigma_n^2}) \tag{2}$$

Where G is the channel gain, is the transmission power, and L is the data length. According to the maximum likelihood decoding criterion, the sequence that makes the probability density function the maximum can be obtained by formula (2):

$$\hat{s} = \arg\max f_x(r \mid s, G) \tag{3}$$

Where $\hat{s}$ is the transmitted signal sequence, and formula (3) can also be understood as finding the estimate value of the signal that minimizes the Euclidean distance between s and r at the data set:

$$\hat{s} = \arg\min \| r - Gs \|^2 \tag{4}$$

Visible light communication has a high amount of data transmission, and it has a high complexity when searching and calculating according to formula (4). the complexity is 2^L. Now simplify it, expand and merge the above formula to obtain:

$$\| r - P_s G s \|^2 = \sum_{i=1}^{L} r_i^2 - 2P_s G \sum_{i=1}^{L} r_i s_i + P_s^2 G^2 \sum_{i=1}^{L} s_i^2 \tag{5}$$

According to the non-negative characteristics of the visible light channel and the analysis of the signal value only "0" and "1", in order to determine the number of bits "1" in the transmitted signal s, the square sum function of the signal is defined as:

$$S = \sum_{i=1}^{L} s_i^2, 0 \le S \le L \tag{6}$$

At the same time, set the symbol corresponding to bit "1" to be the largest, sort the received signals and find the sum of the first S items:

$$Q_S = r_{(1)} + r_{(2)} + \ldots + r_{(s)}, (r_{(1)} \ge r_{(2)} \ldots \ge r_{(S)}, 1 \le S \le L) \tag{7}$$

Substituting formula (5) and formula (6) into formula (4), simplification can be obtained:

$$\| r - P_s G s \|^2 = \sum_{i=1}^{L} r_i^2 - 2P_s G Q_S + P_s^2 G^2 S \tag{8}$$

Formula (7) can get the decision criterion of the signal sum function **S** through extreme value analysis:

$$\hat{S} = \begin{cases} \arg\max\limits_{1 \le S \le L} \dfrac{Q_S^2}{S}, Q_S > 0 \\ 0, Q_S < 0 \end{cases} \tag{9}$$

According to the judgment criterion of the signal sum function **S**, the received signal $r_{(i)}$ can be judged. The data order judgment condition of the receiving end is:

$$\mathrm{r}_{(i)} = \begin{cases} 1, i \geq \hat{S} \\ 0, i \leq \hat{S} \end{cases} \tag{10}$$

The first relative maximum in the decision function (9) is the absolute maximum of a complete signal transmission. The image of this function is approximately a convex function, so the signal can be demodulated through the first peak of the decision function. This decision method Further reduce the complexity of calculation.

5 Experiment and analysis

5.1 Experimental simulation analysis

The T-feedback network trans-impedance amplifier circuit at the photoelectric receiving end was simulated, and compared with the ordinary trans-impedance amplifier circuit, the input current signal frequency was a square wave signal of 10kHz, and the simulation results are shown in the figure 5. The simulation results show that the bias current generated by the T-feedback network trans-impedance amplifier circuit is significantly reduced. Compared with the ordinary trans-impedance amplifier circuit, the DC bias is reduced by 55.9%. The T-feedback network can effectively reduce the DC bias by the compensation resistor.

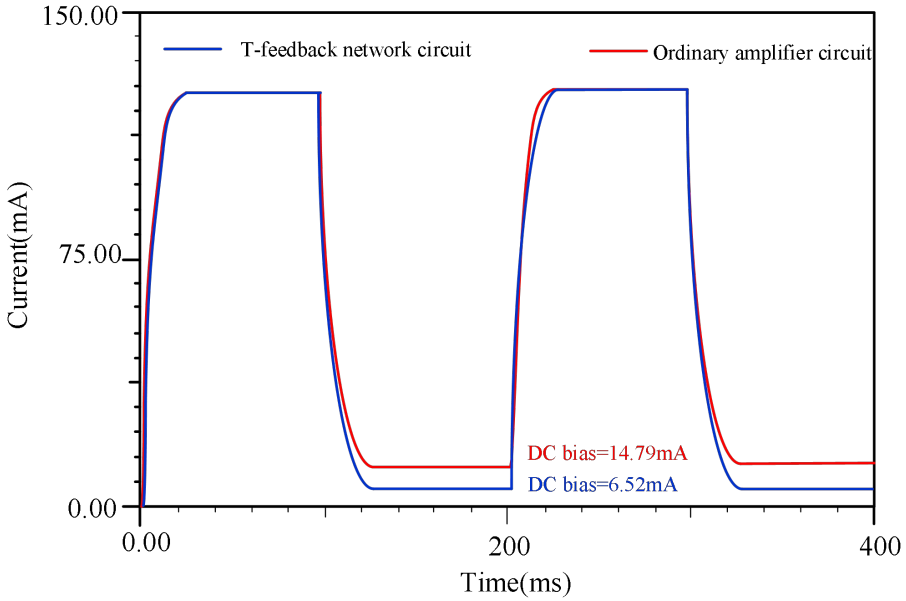


Fig. 5. Simulation diagram of output waveform comparison

For the performance analysis of the algorithm, under the conditions of data length L=20, and variance= $\sigma_n^2 = 0.5$, the simulation results of the BER performance of the Fast-likelihood blind detection algorithm, the MMSE algorithm and the mean detection algorithm are shown in Figure 6.

It can be seen that when the same BER of 10^{-4} is reached, the SNR(Signal-Noise Ratio) of the Fast-likelihood blind detection algorithm is about 28dB, and the SNR of the MMSE algorithm is about 36dB. In contrast, the BER is reduced by 22%. Obviously the fast-likelihood blind detection algorithm is better than other algorithms. In addition, each time the detection is performed, the detection time of the Fast likelihood blind detection algorithm is 3.4s, the Mean detection algorithm is 5.2s, and the MMSE algorithm is 6.8s. It can be seen that the algorithm used in this article not only met the need for BER performance, but also reduced the computational complexity and improves the communication rate. It has advantages in the high-speed signal transmission of the visible light communication system.

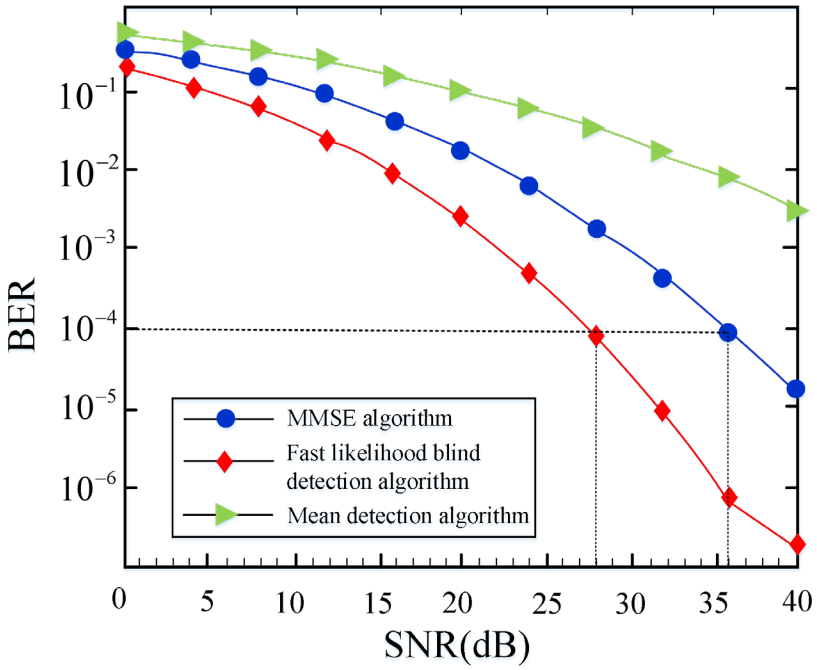


Fig. 6. BER diagram of the three algorithms

5.2 Image transmission test

Figure 7 is the system test diagram built indoors. A water tank with a specification of 100cm*30cm*15cm is used to simulate the underwater channel. During the experiment, adjust the angle of the LED light source at the transmitting end and the angle of the photodiode at the receiving end to align them to achieve the best transmission effect. Figure 8 shows the BER under different transmission distances.

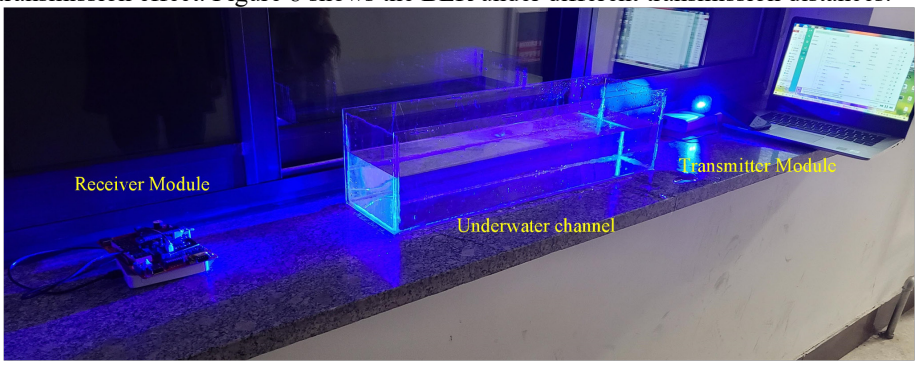


Fig. 7. System experiment diagram

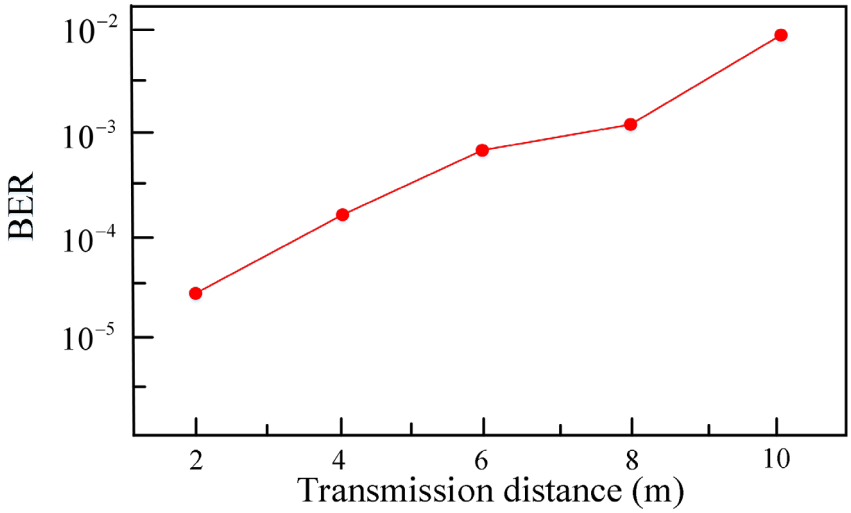


Fig. 8. BER under different transmission distances

In order to better verify the effectiveness of the scheme and simulate underwater image transmission, the sent images are now divided into three types of underwater green, blue and yellow backgrounds. The image size is 380*380 pixels. Figure 9 is the transmitted image, and Figure 10 is the received image. It can be seen from the figure that the three received images are almost the same as the original images, and only some details are blurred.

Fig. 9. The transmitted image

Fig. 10. The received image

5.3 Objective analysis of image quality

In the objective evaluation part, the UCIQE(Underwater Color Image Quality Evaluation) and MSE(Mean-square Error) are used to objectively analyze the received image quality. UCIQE is a linear combination of color concentration, saturation and contrast, which is used to evaluate the color deviation, blurriness and low contrast of underwater images. The higher the value, the higher the image quality. MSE can evaluate the degree of image change. The smaller the value of MSE, the higher the accuracy of the image. The experimental results are shown in Table 1.

Table 1. Comparison of objective evaluation of different methods

	Green image		Blue image		Yellow image	
	UCIQE	MSE	UCIQE	MSE	UCIQE	MSE
Conventional circuit and MD algorithm	0.5328	2563.8	0.5278	3418.2	0.5331	2633.0
T-circuit and MD algorithm	0.6814	1150.0	0.6742	1089.3	0.6625	1201.4
Conventional circuit and MMSE algorithm	0.7056	1245.6	0.7168	1340.5	0.7099	1438.1
T-circuit and MMSE algorithm	0.7341	863.7	0.7584	811.3	0.7528	763.5
This article scheme	0.8129	427.3	0.8236	407.9	0.8086	453.2

It can be seen from Table 1 that the evaluation index of UCIQE of the scheme in this paper is obviously higher than that of other schemes. This is because the dark current of the conventional receiving circuit is large and the signal transmission effect is poor; The MD algorithm algorithm has certain requirements on the data length, and the judgment error is larger when the length is different. In the changing underwater light environment, the MMSE algorithm is not sensitive to the light intensity, which

leads to large errors. There are two reasons for the better image reception of our scheme: 1) The trans-impedance amplifier circuit of the T-feedback network used reduces the dark current of the circuit while ensuring the circuit gain remains unchanged, and at the same time reduces the matching error between the resistors, the circuit performance is better; 2) The Fast-likelihood blind detection algorithm used in this paper filters out edge noise through constrained search, does not rely on channel statistical characteristics, reduces the amount of calculation, and improves the detection accuracy. The scheme in this paper has a higher degree of restoration of the original image's color density, saturation, and contrast.

In addition，the MSE index of the scheme in this paper is also significantly lower than other schemes, indicating that the pixel gap between the received image and the original image is smaller, and the image information fidelity is better.

6 Conclusion

Aiming at the disadvantages of traditional underwater communication methods and the poor underwater visible light communication channels, this paper designed a visible light communication system for underwater image transmission. The trans-impedance amplifier circuit of the T-feedback network was adopted at the receiving end of the system to improve the performance of signal transmission; The use of the Fast-likelihood blind detection algorithm not only reduced the BER of signal transmission, but also reduced the computational complexity. Finally, the effectiveness of the proposed scheme was verified by building a system experiment platform and comparing the image quality under different schemes.

Acknowledgements: This research was financially supported via Project of the National Natural Science Foundation of China (61402069), the 2017 Project of the Natural Science Foundation of Liaoning province (20170540059), the General project of Liaoning education department in 2016(2016J205).

References

1. Lu, H., et al. : CO-Net: A cognitive ocean network. IEEE Wireless Com-mun. 26(3), 90–96 (2019)
2. Wang J P, Ye Z P, Tarun M, Li B and Zou N Y.: A Novel Linear Antenna Synthesis for Linear Dispersion Codes Based on an Innovative HYBRID Genetic Algorithm, Symmetry 2019,11(9),1176.
3. Xiang W D, Yang P, Wang S, et al. : Underwater image enhancement based on red channel weighted compensation and gamma correction model [J]. Opto-Electron Adv 1, 2018，01(10)： 180024.
4. Zeng F J,Yang K J,Yan X,Zhao M M,Yang Ping,Wen L H.: Research progress of underwater laser communication system [J/OL]. Laser and Optoelectronics Progress:1-19[2020-12-21].
5. Wang J P, Ye Z P, Jeremy Gillbanks, Tarun M. Sanders and Zou N Y, A Power Control Algorithm Based on Chicken Game Theory in Multi-Hop Networks, Symmetry 2019,

11(5),718.
6. ZHAO Y H, WANG C, WANG F M.: 1.725Gb/s underwater visible light communication system based on a silicon substrate green LED and equal gain combination receiver [C] // International Conference and Exhibition on Visible Light Communications (Invited Talk 3). Yokohama, Japan: ICEVLC, (2018).
7. Wang J P, Cao F, Zou N Y. Multi carrier system joint receiving method based on MAI and ICI. [J], Jilin Daxue Xuebao (Gongxueban), 2018, 41(6): 301-305.
8. CHI N, WANG C F, LI W P.: Research Progress of Underwater Visible Light Communication Technology Based on Blue/Green LED. Journal of Fudan University(Natural Science), 58(05), 537-548 (2019).
9. Chen H H.: "Struggle": From domestic to "National Innovation". China Science News, 2020-11-30 (1).
10. ZOU P, LIU YF, WANGF M, CHIN, Mitigating nonlinearity characteristics of gray-coding square 8QAM in underwater VLC system [C] / Asia Communications and Photonics Conference. Hangzhou, China: IEEE, 2018: 1-3.
11. Wang J P, Ye Z P, Cao F, and Zou N Y. : Cooperative distributed antenna transmission method based on co-channel interference in 5G mobile communication system, Journal of Jilin University (Engineering and Technology Edition), 2020, 48 (1)：333-341.
12. Ji W Y, Wu G F, Wang C. : Generalized Likelihood Block Detection for Spad-based Underwater Vlc System[J]. Science Letter, 2020.
13. Wang Y, et al. : Fusion-based underwater image enhancement by wavelet decomposition. In: Proceedings of IEEE International Conference on Industrial Technology (ICIT), Toronto, Canada, pp. 1013–1018 (2017)
14. Qi Y, Liu Z, Ni X L. : Research on Maximum Likelihood Detection Algorithm of Random Optical Signal Based on Expectation Maximization[J]. Journal of Changchun University of Science and Technology (Natural Science Edition), 2017, 40(02): 90-93.
15. Asuncion A, Newman D. UCI Machine Learning Repository[EB/OL]. [2015-06-01]. http://archive.ics.uci.edu/ml/datasets.html
16. Yu J J, Chi N. Digital Signal Processing In High-Speed Optical Fiber Communication Principle and Application[M]. Singapore, 2020

Empirical Research on Urban and Rural Education Equality of Inner Mongolia Under the Background of Urban-rural Integration——Analysis Based on Education Gini Coefficient

Linlin Liu[1], Jinpeng Wei[2] and Na Zhang[3]
{liulinlin@imu.edu.cn[1],weijinpeng0202@163.com[2], 2605267865@qq.com[3]}

School of Marxism Inner Mongolia University[1], School of Public Management Inner Mongolia University[1], School of Public Management Inner Mongolia University[23], School of Public Management Inner Mongolia University[23]

Abstract. This paper selects the statistical data of education situation in Inner Mongolia from 2002 to 2018, and analyzes the changes of educational equity in Inner Mongolia by using the decomposition method of education Gini coefficient. The results show that the education level of the population is increasing, the Gini coefficient of education is decreasing, and the overall education equity has achieved remarkable results, which is in a overall fair stage; the gap between urban and rural education years has increased, and the contribution rate of urban and rural education difference has been maintaining at a rather high figure and shows the trend of continuous increase. It can be seen that the difference between urban and rural education is the main cause of restricting the development of educational equity in Inner Mongolia, and will exist for a certain period of time. The restrictive factors that cause the gap between urban and rural education equity are mainly reflected in the imbalance of the economic social structure, the imperfection of government responsibility and public investment system, the dilution of educational resources by educational immigrants, and the low level of protection of rural teachers' professional rights and interests. Promoting the integration of urban-rural compulsory education is the fundamental way to narrow the gap between urban and rural areas and solve the problems of rural education development. Specifically speaking, we should build a multi governance community of urban and

---

[1] Author: Linlin Liu (1981 ~), female, doctoral candidate, School of Marxism, Inner Mongolia University. Jinpeng Wei (1996 ~), male, master of School of public administration, Inner Mongolia University. Zhang Na (1996 ~), female, graduate student, School of public administration, Inner Mongolia University.

Fund Project: scientific research project of colleges and universities in Inner Mongolia Autonomous Region "Research on high quality and balanced development path of compulsory education in Inner Mongolia under the background of urban-rural integration" (NJSY21287).

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307540

rural education led by the government, in order to promote the two-way flow of urban and rural education resources, continue to implement the rural-teacher-support plan, vigorously develop rural preschool education, promote education targeted poverty reduction, and promote education equity through education informatization.

Keywords: Ethnic minority areas; education equality; urban-rural integration; education Gini coefficient.

1 Introduction

The Fifth Plenum of the 19th Central Committee of the Communist Party of China has clearly stated that by the time of 2035, we will achieve the goal of building a education power, reaching a new high level with national qualities and social civilization, achieving new actual developments of the human development in every aspects and common prosperity. Standing at a new historical starting point, developing more fair and quality education is an inevitable choice to adapt to the changes of in the principle social contradictions and the characteristics of high-quality development stage. It's also an inevitable choice to improve the level of education modernization and meet the ever-growing needs for a better life. Achieve these great aims and "Not a target shall be neglected" ,"Not a single region or nation shall be left behind". In recent years, the issue of educational equity has draw the attention of domestic scholars. In empirical studies, standard deviation and Gini coefficient are used to measure the degree of educational inequity, and most of the studies prefer Gini coefficient method. Changzheng Chang, Zhijian Xun, Huaizu Li and Baicai Sun has calculated the education Gini coefficient of China over the years. The results show that the total number of years of education of our population is increasing year by year, the Gini coefficient of education is decreasing year by year, and the degree of education equity has been greatly improved, but there is still a long way to go compared with the world level. [1],[2],[3] Dexin Hu analyzed the education development level and education equity of cities, counties and townships in China.[4] Wanming Zhang used education Gini coefficient, and analyzed the overall education equity and inter provincial education equity from the perspective of education outcome equity. She holds that the overall education equity is improved, while the inter provincial education development is seriously unfair.[5] Zhenhua Wu calculated and analyzed the level and trend of educational equity of rural residents in China between and within provinces. We can tell from his research that Inner Mongolia is an area with strong willingness to receive education and of general public education resources.[6] Based on the trend of education years, education Gini

coefficient of urban and rural population in Inner Mongolia, and the gap between urban and rural education years by gender and by rural or urban areas, this paper expounds the degree of education equity in Inner Mongolia ,the main causes and constraints of education injustice at this stage, and puts forward countermeasures and suggestions.

2 Method AND SUMMARY STATISTICS

2.1 Method

When it comes to the measurement of educational equity, we usually measure the degree of difference and equilibrium through educational development indicators, such as Gini coefficient, difference coefficient, equilibrium index and so on. And education Gini coefficient is widely used by scholars because of its stability and effectiveness.This paper also applied the education Gini coefficient measurement method to decompose the overall equity of Inner Mongolia's education development according to the differences among cities, counties and townships, and thus analyze the contribution of urban-rural gap to the overall education equity, and evaluate the urban-rural equity of Inner Mongolia's education development trend.At the same time, considering the differences of gender equity between urban and rural areas, we also decomposes the gender education equity among cities, counties and townships.

2.1.1 Measurement of educational equity

The measurement of education Gini coefficient is derived from the calculation principle of Gini coefficient. According to the Gini coefficient: when Gini coefficient is lower than 0.2, it means absolute fairness, 0.2 ~ 0.3 means relatively fair, 0.3 ~ 0.4 means relatively reasonable, 0.4 ~ 0.5 means the gap is relatively large, and 0.5 above means a wide gap, and 0.4 is usually used as the "warning line" of Gini coefficient's evaluation.This standard is also applicable to the education Gini coefficient of education, that is, the higher the education Gini coefficient of education, the lower the degree of education equity. The calculation formula is as follows:

$$G = 1 - \sum_{i=1}^{5} (X_i - X_{i-1})(Y_i + Y_{i-1}) \tag{1}$$

As we can see in equation (1), where G is the education Gini coefficient, i is the level of

education, x_i is the cumulative proportion of the educated population at each level in the total population, y_i is the cumulative proportion of the number of years of education of the population at each level in the total number of years of Education,[7] and 5 is the classification of the number of education levels. According to the current school year system and education system of our country, and from the research of other scholars, the education level can be divided into five levels: the first level, no schooling, and its years of education is 0; the second level, primary school, and its years of education is 6; the third level, junior high school, and its years of education is 9; the fourth level, senior high school or technical secondary school, and its years of education is 12; The fifth level is junior college or above, with 16 years of education.

2.1.2 Decomposition of Education Gini coefficient

In order to analyze the contribution of each factor to Gini coefficient, we need to further decompose the composition of Gini coefficient. At present, there are mainly two methods to decompose Gini coefficient : one is the trichotomy method, that is, to decompose Gini coefficient into intra-group, inter-group and residual items, represented by the decomposition formula of Yao (1999), [8] and Hangkong Zhang (2013)[9] and others have conducted relevant research based on this method. The decomposition formula of Gini coefficient of education is as follows:

$$G=G_a+G_b+G_c \qquad (2)$$

As we can see in equation (2), where G_a is the internal components due to the difference of educational level within groups. G_b is the components between groups due to the difference of educational level among groups . G_c refers to the remaining part or the overlapping part of groups to reflect the differences among groups.

The other is dichotomy, which is to divided samples into intra-group and inter-group, represented by the decomposition formula of Junsen Zhang (2002)[10], and Baicai Sun(2014) and others conducted relevant research based on this method. In view of the fact that the remaining items in the first method overlap with intra-group and inter-group, which will weaken the accuracy of inter-group and intra-group's contribution. Thus we chose the second method to decompose the education Gini coefficient. The specific decomposition formula is as follows:

$$G = \sum_{i=1}^{n} P_i^2 \cdot \left(\frac{\mu_i}{\mu} \right)^2 \cdot G_i + G_B \quad (3)$$

As we can see in equation (3), where G is the education Gini coefficient ; P_i is the proportion of group i's population; μ_i and μ represent the average years of education of group i and total population's average years of education respectively; G_i represents the education Gini coefficient of each group, $P_i^2 \cdot \left(\frac{\mu_i}{\mu} \right)^2 \cdot G_i$ is the absolute contribution of intra-group gap to the overall education Gini coefficient, and in the following parts we will refer it as R_i; G_b is the absolute contribution of inter-group gap to the overall education Gini coefficient.[11] The relative contribution of each part and the relationship between them are as follows:

$$\sum_{i=1}^{n} \frac{R_i}{G} \times 100\ \% + \frac{R_B}{G} \times 100\ \% = 1 \quad (4)$$

2.2 Data

"China Population & Employment Statistic Yearbook" has counted the population of cities, counties and townships of all provinces, autonomous regions and cities in China by gender and education level, and the data set is authoritative and available.We selected the statistical data of the education situation of people in Inner Mongolia from 2002 to 2018, divided by city, county and country and by gender.

In 2002, the Ministry of Education issued "the notice on strengthening the management of basic education", which clearly put forward that "To actively promote the balanced development of schools in the stage of compulsory education". Since then, the development of compulsory education has begun to change to the direction of "balanced", and "key schools and key classes" have been gradually canceled. Therefore, 2002 is the key year of China's education reform and development since the reform and opening up.

3 Analysis of research results

3.1 Urban-rural difference of per capita education years in Inner Mongolia

As can be seen from Figure 1, the average number of years of education per capita in

Inner Mongolia has increased from 7.88 years in 2002 to 9.69 years in 2018. The required number of years of compulsory education has been completed. Combined with the proportion of population grouped by education level, we can tell that the overall educational level of Inner Mongolia population is in the transition stage from junior high school to senior high school.

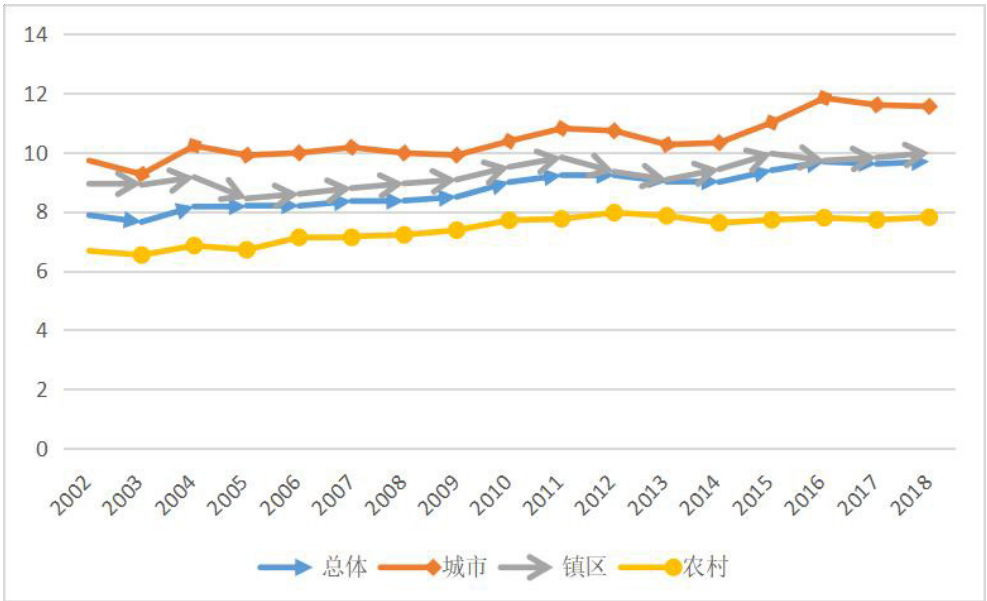


Fig. 1. Architecture of a typical wireless sensor node.

Although the gap of education opportunities between urban and rural population is gradually narrowing, and the per capita education level is steadily improving, the gap between urban and rural residents' per capita education years has expanded, and the gap is mainly reflected in the education opportunities of senior high school and above and also the illiteracy and semi illiteracy rate. In 2018, the number of years of education per capita in cities will reach 11.55 years, and more than 50% of urban residents will have access to high school and above education. Nearly one third of the population is of college education or above, which is way higher than that in towns and rural areas.

The proportion of educational population at all levels in the township is close to the overall level. The number of years of education per capita in rural areas has not yet reached the level of nine-year compulsory education; the education level of the population accounts for the vast majority of primary schools and junior high schools, and the proportion of residents receiving or has received high school education or above is only 15%, but the proportion of

uneducated population is still as high as 9.06%, which is related to the migration of labor force and education derived from urbanization. In combination, the gap between the education level of rural population and the overall is still very wide. This situation will exist for a long time at present and in the future. The improvement of rural education level is still a short link restricting the development of urban-rural education integration.

Table 1. Proportion of urban and rural education population at all levels in Inner Mongolia Autonomous Region in 2018

Education level of population	Proportion			
	Total	City	County	Township
no schooling	5.29%	1.83%	4.38%	9.06%
primary school	22.96%	11.53%	20.16%	35.26%
junior high school	35.64%	28.15%	37.96%	40.73%
senior high school or technical secondary school	16.78%	25.74%	16.12%	9.17%
junior college or above	19.33%	32.75%	21.39%	5.78%

As we can see from Figure 2 , from the perspective of gender differences in the number of years of education per capita, the gap between men and women in the whole and in urban and rural areas is narrowing, and the gap of years of education between men and women in 2018 is about 50% smaller than that in 2002. Among the differences of divided groups of educational years in different genders, the average educational years of men are higher than that of women, and the educational years of urban genders are less than that of rural ones. The average number of years of education for rural women is the lowest. The reason for this trend is not only the unbalanced development of urban and rural education, but also the deep rooted influence of gender discrimination in education concept, especially in rural and pastoral areas.

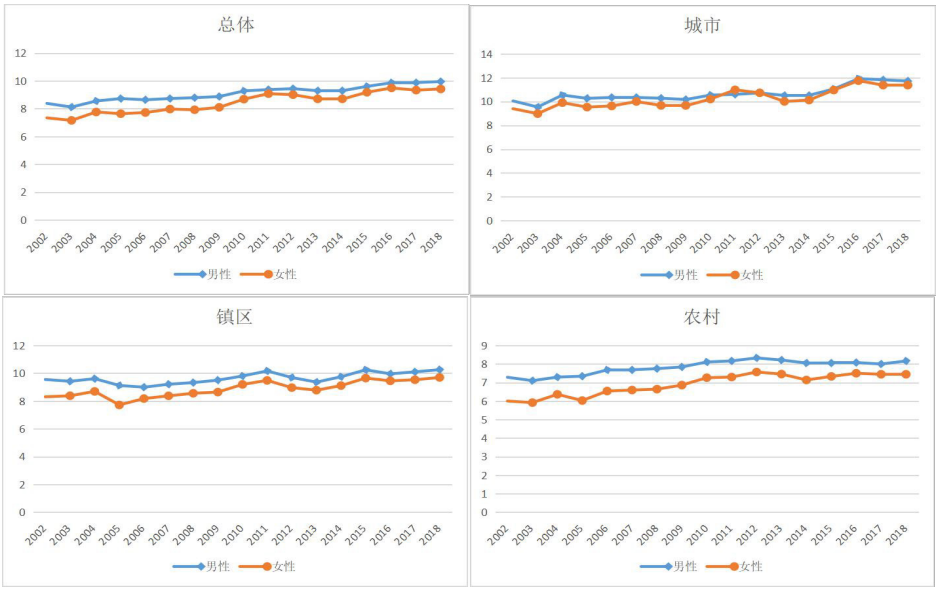


Fig. 2. Years of education for urban and rural residents in Inner Mongolia Autonomous Region (2002-2018)

3.2 Urban-rural differences in education Gini coefficient in Inner Mongolia

In the context of urban-rural dual structure system, the gap between urban and rural education development is obvious, the inequality of education opportunities between urban and rural population has been criticized, and rural education has been accused of being in "poverty". In recent years, the central and local government departments at all levels have attached great importance to the balanced development of urban and rural public education, trying to narrow the education gap by rationally allocating education resources and leaning towards the rural areas, remote poverty-stricken areas and ethnic minority areas.

When it comes to the gap between urban and rural education, the society is more concerned about whether the gap is improving or worsening. As can be seen from Figure 3, the degree of education equity of urban and rural residents in Inner Mongolia has improved rapidly. With the increase of per capita schooling years, the education Gini coefficient of urban and rural shows a downward trend, from 0.262 in 2002 to 0.230 in 2018; and from 0.204, 0.228 and 0.275 in urban, township and rural areas to 0.179, 0.219 and 0.235 respectively.

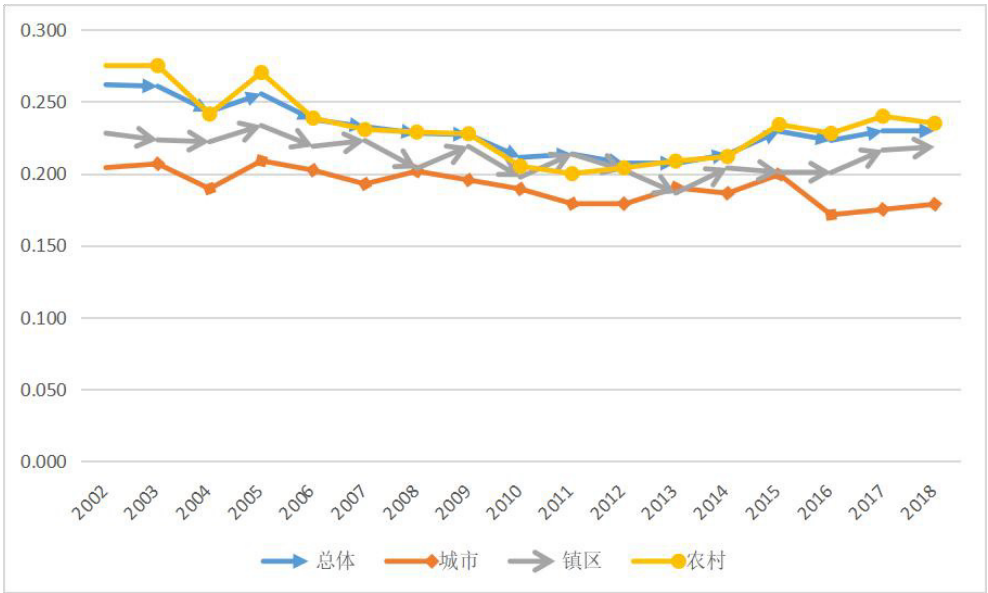


Fig. 3. Education Gini coefficient of urban and rural in Inner Mongolia Autonomous Region (2002-2018)

As we can see in Figure 4, from the perspective of gender between urban and rural areas, the Gini coefficient of the whole and within urban and rural areas shows a downward trend, in which the downward trend of women is higher than that of men, and the downward trend of rural areas is higher than that of urban areas. To some extent, it reflects the improvement of education equity brought by the improvement of rural education in recent years. From the perspective of gender gap, the gap of education Gini coefficient between genders is narrowing, but women is higher than men in all the groups. Among them, the Gini coefficient of rural women is the highest. Combined with the level of per capita education years, it shows that the education situation of rural women is the most unfair in urban-rural areas and gender. This also shows that while vigorously promoting the integration of urban and rural education, we should also protect the right to be educated between different genders in urban and rural areas.

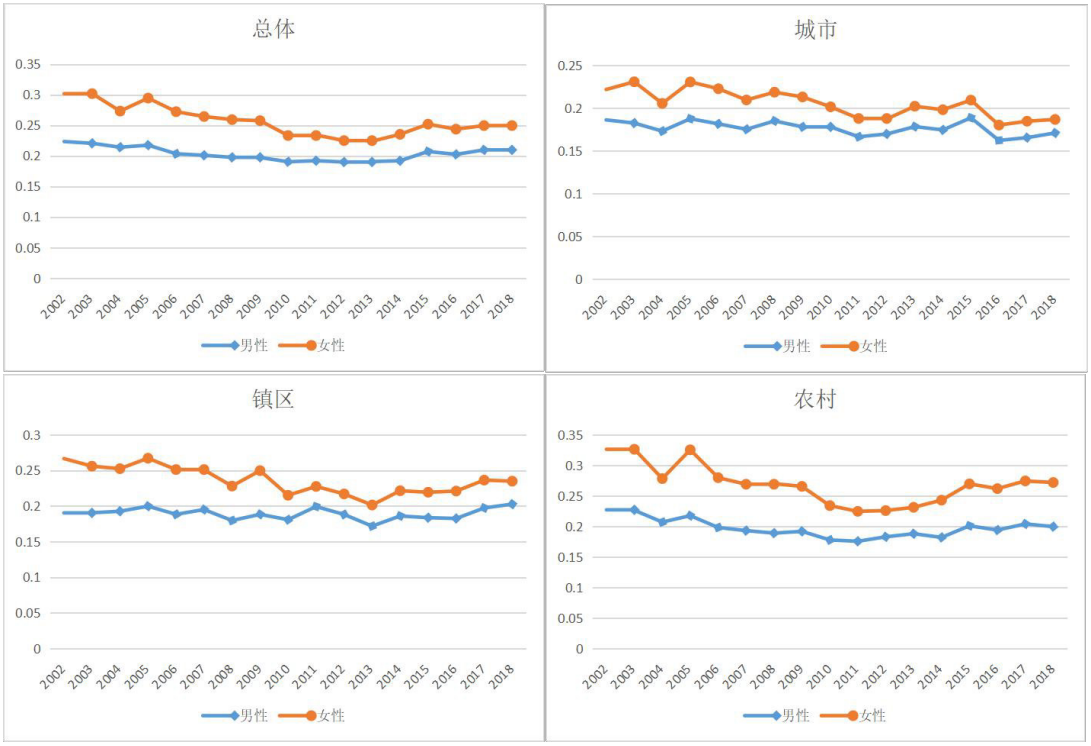


Fig. 4. Gender education Gini coefficient of urban and rural in Inner Mongolia Autonomous Region (2002-2018)

3.3 Analysis of Education Gini coefficient of Inner Mongolia

As we can see in Figure 5, the decomposition results of education Gini coefficient shows the following three characteristics:

First, the contribution rate of cities and towns to the unfair development of education is on the rise. The contribution rate of cities increased from 8.55% in 2002 to 13.14% in 2018; the contribution rate of counties increased from 3.24% to 7.51%. These rising trend is mainly due to the dilution of urban education level after the rural population with lower education level flows into cities and counties, thus expanding the education inequality within cities and counties.

Second, the contribution rate of rural areas to the unfair development of education has decreased significantly. The contribution rate of rural areas decreased from 23.79% in 2002 to 9.47% in 2018. On the one hand, it has benefited from the strong investment and support of the central and local governments to the education of rural pastoral areas; on the other hand,

after the outflow of a large number of rural population, the education level of the remaining population was generally low and the internal difference was small.

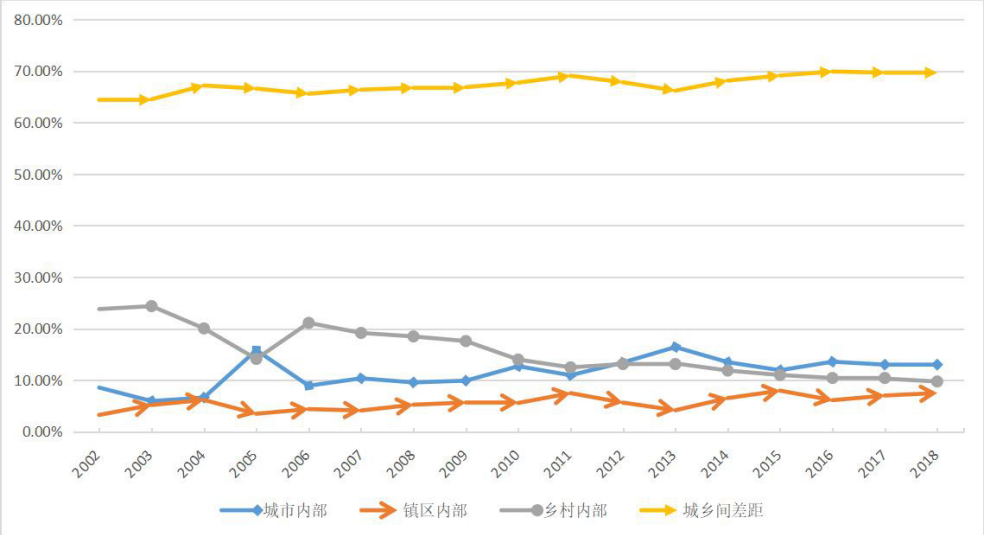


Fig. 5. Contribution of urban and rural areas to the unfair development of education in Inner Mongolia (2002-2018)

Thirdly, the contribution rate of urban-rural differences increased year by year, from 64.41% in 2002 to 69.62% in 2018. Because the contribution rate of urban internal differences is less than that of rural internal differences, the total contribution of urban-rural internal differences to Gini coefficient gradually decreases, and the sum of the relative contribution rate of city, county and township internal differences decreases from 37.87% in 2002 to 30.39% in 2018. This shows that the progress of educational equity mainly comes from the narrowing of the gap between urban and rural areas, especially the improvement of educational environment and the level of educational equity in rural pastoral areas. The difference between urban and rural areas has always been the main cause of the overall education inequity in Inner Mongolia, and the contribution rate to education inequity is increasing year by year.

Table 2. The contribution of education inequality between urban and rural areas

	City contribution rate （%）	County contribution rate （%）	Township contribution rate （%）	Urban-rural difference contribution rate （%）
2002	8.55%	3.24%	23.79%	64.41%
2003	5.94%	5.18%	24.37%	64.52%
2004	6.61%	6.16%	20.04%	67.20%
2005	15.81%	3.47%	14.13%	66.59%

2006	8.91%	4.38%	21.12%	65.59%
2007	10.35%	4.10%	19.16%	66.38%
2008	9.55%	5.23%	18.48%	66.73%
2009	9.89%	5.66%	17.58%	66.87%
2010	12.70%	5.56%	14.00%	67.74%
2011	10.96%	7.48%	12.48%	69.09%
2012	13.42%	5.64%	13.12%	67.81%
2013	16.42%	4.15%	13.25%	66.18%
2014	13.49%	6.51%	11.87%	68.14%
2015	11.92%	7.96%	10.99%	69.12%
2016	13.57%	6.10%	10.42%	69.91%
2017	13.01%	7.01%	10.30%	69.68%
2018	13.14%	7.51%	9.74%	69.62%
2002-2018 average	11.43%	5.61%	15.58%	67.39%

4 Analysis of restrictive factors

4.1 The economic social structure leads to the widening gap between urban and rural education

Since the founding of the People's Republic of China, the Household Register System has been used as a method to restrain rural residents from pouring into the urban, which is closely connected to the allocation of all sorts of resources. Under the effect of the urban-rural-segregated Household Registration System, the unique “urban-rural dual structure” were formed, and the development gap between urban and rural areas exists in various fields. This makes it difficult for cities and rural areas to form a optimum interactive relationship. Among all kinds of unfairness caused by this, the unfairness of urban and rural education development has the greatest and deepest impact on the overall social development of our country, which is mainly reflected in the inequality of urban and rural education opportunities, the inequality of urban and rural education management and investment, and the imbalance of urban and rural education resource allocation. [12] Due to the deep-rooted division and rule of urban and rural education, the unreasonable distribution mode of basic education resources hinders the balanced development of education, and the education in rural and pastoral areas still lags behind the city in general. Even in this new new urbanization state, with tons of rural students entering the urban, the child of the migrant workers still can’t have the same education resources as the urban students because of the Household Registration System and the fact that urban students’ education rights were put first, which makes “education for everyone” a luxury. The economic social development level of Inner Mongolia lags behind

that of the eastern developed areas. The rural and pastoral areas, especially the poor counties in nation level and in autonomous region level, lag behind in its region,county or city, and the education resources it possess is rather little, which restricts the development of the local education. This situation is in urgent need to be improved through the integrated development of urban and rural education.

4.2 With unclear responsibilities of governments at all levels and low level of decentralization of the public investment system, it is difficult for education to develop.

The primary challenge the education investment system is facing is the responsibility of government investment is unclear. And because of that, the financial sharing mechanism is unclear. In the education market, the areas and scope of government independent investment, government leverage investment, guiding market investment and market leading investment are not clear, which leads to the risk and trend of government investment responsibility generalization under the guidance of promoting education equity. [13]China's current education investment system is a local government responsibility system under the unified coordination of the central government, which is characterized by the government leading. This means that the local government bears a more important responsibility in education investment and education resource allocation. At the same time, the level of the main body of cost bearing of compulsory education is too low. The compulsory education financial system of "provincial overall planning, county-based" makes the county-level governments became the main fundraiser, due to the constraints of the economic foundation of the grass-roots government, the current situation of social development and other factors, the financial funds in Inner Mongolia are still relatively concentrated in cities and towns, the resource interaction between urban and rural areas has not yet formed, the radiation belt action of urban education development on rural pastoral areas is insufficient, and the development of rural education is difficult to promote. The inconsistent responsibility and ability of education investment of grass-roots government leads to the imbalance of education resource allocation between urban and rural areas.

4.3 "Education immigration" dilutes the development level of education

Due to the rapid development of urbanization, the decrease of rural school-age population and the slow down of local government's "caring" to rural education, a large number of rural schools have been merged, and the resources of rural schools that have not been merged is

increasingly dropping, which leads to the one-way “urban mobility” of rural students and excellent teachers. The result is the decline and even extinction of rural education. In this background, the educational immigration policy of “two oriented, two included” came into being. From the perspective of implementation effect, although the educational immigration policy integrates educational resources and maximizes the efficiency of educational resources. But in essence, it is through the supply of the system to carry out the institutional redistribution of the original uneven distribution of educational resources. Essentially, it is a remedial system design within the framework of the urban-rural dual education system, which lacks the foresight of the policy implementation effect [14], resulting in a new polarization of education and leads to the decline of rural education; immigrant students have to make a difficult choice between paid high-quality educational resources and declining rural education. The merger or decline of rural schools can only damage the educational interests of the most vulnerable groups in rural areas.

4.4 The lagging behind practical professional interests and low sense of belonging of professional levels of rural teachers weaken the power of education development

The primary reason for the unfair ratio of teachers in urban and rural areas lies in the fact that rural preschool teachers face great difficulties in obtaining their practical professional interests. Due to the location disadvantage, salary supply, development platform and other factors, the construction of rural compulsory education teachers is still relatively backward. [15] As the most dynamic and competitive factor in the development of education, teachers are faced with employment choice, employment environment, job development, promotion prospects and other practical interest factors. There are great differences in educational facilities and resources between urban and rural areas in Inner Mongolia, with backward teaching environment in rural areas, weak educational foundation, lack of incentive mechanism for teachers, it is difficult to attract teachers. High quality education resources are concentrated in the city, which provides teachers with more favorable professional interests and development platform, and can give teachers a good sense of professional belonging. Therefore, high quality teacher resources tend to flow to the city, resulting in the phenomenon that the supplement speed of rural teacher resources is far lower than that of the city, resulting in the imbalance in the number and quality of teachers between urban and rural areas.

5 Countermeasures and suggestions

Fair development and quality improvement of education are two major themes of China's education development in the new era. In the process of promoting China's education modernization, education modernization in ethnic areas is not only the focus of development, but also the short board that needs to be patched urgently. [16] In the background of urban-rural integration, education in ethnic minority areas ushers in rare opportunities for development. Through the supervision and evaluation of Balanced Development of National Compulsory Education in Inner Mongolia Autonomous Region, the comprehensive coefficient of difference of primary school in 103 counties (cities and districts) is less than 0.65, reaching the national standard, with the average value of primary school comprehensive coefficient of difference of 0.451; the comprehensive coefficient of difference of junior middle school in 103 counties (cities and districts) is less than 0.55, reaching the national standard, with the average value of junior middle school comprehensive coefficient of difference of 0.301. [17] In the process of promoting education equity, the development of compulsory education in Inner Mongolia has made a historic leap forward, but there is still a lot of room for development to realize the high-quality balance, especially the unbalanced development of urban and rural education has become the main cause of the unbalanced development of education.The integration of urban and rural compulsory education is conducive to improving the resource sharing of urban and rural areas in the implementation process of compulsory education, and realizing the dynamic balance of distribution. Integration is not non-differentiation or homogenization. Rural education can learn from the concept of urban education, but should not be simply coping and imitating. On the premise of not separating urban education from rural education, according to the development characteristics and current situation of rural education, rural primary and secondary schools should be encouraged to develop their own advantages according to their own characteristics, narrow the gap between urban and rural areas, and integrate urban and rural education resources. The integration of urban and rural areas will promote the overall common development of urban and rural areas.

5.1 Strengthen the top-level design and build a multi-governance community of urban and rural education led by the government

In ethnic minority areas, the awareness of participation of multiple governance subjects is weak, and the government and provincial education management departments are more putting on a “one-man show” in the process of providing public education services, which strengthens the government’s main body function and weakens the joint governance and

collaborative participation of society and market. Government should change the concept of government management while coordinating and making education policies, fully respect the autonomy of rural schools, explore the establishment of education mode to adapt to the sustainable development of rural education, and encourage the linkage of urban and rural basic education collectivization. It is also necessary to activate the principal and teacher's participation consciousness of the school as an educational entity, which is not only limited to mechanical implementation, but also to enable them to jump out of the complicated logistics support and teacher-student safety system affairs, concentrate on running the school and studying, and pay attention to the aspirations of students and parents as the objects of educational service.

In the specific implementation process: first, give the rural school principals full autonomy and academic autonomy , but with strict quality control and process evaluation; second, create a good atmosphere of "respecting teachers and valuing education", give full play to the main role of rural teachers as the "soul of education", encourage rural teachers to innovate teaching methods, give care in life and work, and stimulate rural teachers' enthusiasm and also teachers' sense of professional belonging and lofty mission; third, in the selection and employment of rural school principals, we should strictly assess their moral cultivation and academic philosophy, and encourage them to carry out practical exploration of education reform; fourth, we should establish an assessment and evaluation system. Regularly organize the third-party quality control evaluation and supervision participated by the government, society and the representatives of parents' heads, and regularly feed back education opinions and suggestions, so as to mobilize the enthusiasm of all parties to participate.

5.2 Break the system and mechanism barriers and promote the two-way flow of urban and rural education resources

The concentration of urban high-quality education resources not only attracts high-quality students from rural areas, but also attracts many high-quality teachers. The trend of resource flow is one-way flow from rural to urban areas, which aggravates the impoverishment of rural education resources and the decline of education quality. Thus formed an essential inequality to the "vulnerable groups" who are forced to receive education in rural areas for they don’t have the conditions to study in urban areas . The unfairness of education makes it difficult for a large part of rural groups to enjoy fair and quality education. With the improvement of people's living standards and the popularization and implementation of the compulsory education law, the fairness of educational opportunities has been basically realized, and people

are looking forward to the fairness of high-quality educational opportunities. It is necessary to ensure that every child has the opportunity to enjoy high-quality educational resources at least in the stage of basic education. The key and difficult point lies in the fairness in the process and result of education, that is, to change from basic balance to high-quality balance, from resource allocation as the core to quality improvement as the core, from extension development to connotation development, from passive development relying on the government to independent development of schools. The premise of high quality equilibrium is equilibrium, and the core is high quality. First is to reform the financial sharing mechanism, polish and adjust different education funds output grades, and introduce social capital to alleviate the practical problems of the shortage of educational funds, backward teaching equipment and insufficient teachers in rural schools;[19] second is to focus on the long-term development of Education Driven Rural Revitalization. The central and provincial governments should further strengthen the financial transfer to the poverty-stricken areas of ethnic minorities and hold the bottom line; third is to explore the cooperation between urban and rural schools, and make the education resources in the cities to “Complementary advantages, learn from each other” [19] . Finally and most importantly, we should pay attention to the improvement of education quality, that is, the improvement of "soft power".

5.3 Continue to implement the “rural teacher support plan”, and make the preferential policies to be further extended to the areas with poor teacher resources.

As a result of long-term neglect, rural schools have been in the plight of insufficient teachers, unstable team, unreasonable structure and more substitute teachers, which has become an important factor restricting the development of rural education. There is a big gap between urban and rural teachers in terms of economic income, promotion and living treatment. Compulsory education teachers in the county show a single upward flow into the county schools. The implementation of “rural teachers support plan” has improved the allocation of rural teachers' resources, steadily improved the level of education and teaching ability, guaranteed the reasonable treatment in all aspects according to law, and significantly enhanced the attractiveness of the profession. However, there are still some difficulties, such as lack of finance or insufficient investment in some areas, inadequate implementation of non explicit treatment such as housing provident fund, insurance and turnover dormitory, and ineffective implementation of two-way exchange system in counties. First of all, we should start with the construction of the teaching staff in the areas where the teacher resources are scarce, appropriately increase the proportion of rural teachers, attract excellent talents,

implement the system of county management and school appointment and job rotation within the county, increase the salary of rural teachers, and increase the special allowance for teachers in the areas where the conditions are difficult. The more difficult the local condition is, the higher the special allowance of rural teachers will be, so that the salary of rural teachers will not be lower than or even higher than that of urban teachers. According to the degree of contribution and the spirit of dedication, we should give policy preference to the teachers who work in rural areas and have been engaged in rural education for more than a certain number of years. Secondly, it is necessary to establish a two-way flow mechanism of urban and rural teacher resources. Teachers in county and rural schools flow regularly in terms of quantity and quality, forming a flow state in which the city leads the township, the township helps the city, and urban and rural teachers develop and grow together. [20] Third, rural teachers' further study, joint training and directional training should be normalized, and rural teachers should be selected regularly and proportionally to go to famous universities for further study.

5.4 Vigorously develop rural preschool education to solve the problems of rural children's "difficulty" and "too expensive" to enter the kindergarten.

Preschool education is the starting point of one's education and lifelong learning, which is an important part of the national education system. Due to the influence of historical and practical factors, problems such as "unreasonable layout", "high tuition", "extreme lack of teachers" and "insufficient supporting infrastructure" have emerged in the development process of rural preschool education, which are the direct causes of the problems that are strongly reflected by the rural people—— "difficult and expensive admission" . In Inner Mongolia, after the implementation of the three-phase preschool education action plan, the coverage rate of rural preschool education has greatly improved, but the layout of inclusive kindergartens in remote agricultural and pastoral areas still can not fully meet the needs of local residents, high-quality kindergartens are scarce, and professional preschool teachers are insufficient. To vigorously develop rural preschool education, we should accelerate the formulation of the “preschool education law”, expand the coverage of governance funded preschool, guide social forces to hold more inclusive kindergartens, strengthen the docking and integration of urban and rural preschool education resources, give full play to the advantages of the radiation of urban high-quality preschool education resources and the linkage between urban and rural areas, and promote the construction of integrated rural preschool education supply network of district, county, township and village, which can activate the vitality and endogenous force of rural preschool education development.[21]

Promote preschool education, especially in rural areas, from quantitative development to balanced development, to quality improvement, to characteristic development and innovative development.[22]

5.5 Promote targeted poverty alleviation in education to prevent students from dropping out of school due to family difficulties

President Xi Jinping once pointed out that “the poorer the place is, the poorer the education is, but the poorer the place is, the more education is needed, and the poorer the education is, the poorer the place is.” [23] Education targeted poverty alleviation is the booster of "blocking the intergenerational transmission of poverty" and the catalyst of promoting social classes mobility. Ethnic minority areas are the main battlefields for poverty alleviation. Among the 14 contiguous poverty-stricken areas in China, 11 are distributed in ethnic minority areas. In 2016, the Ministry of education and other departments jointly issued the “13th five year plan for poverty alleviation”, which pointed out: "to realize the full coverage of basic public services for the education of the poor people, such as the poverty card establishment, to guarantee the full funding of all education stages from enrollment to graduation, to ensure that children from poor families can go to school, and not to let a single student drop out of school due to family difficulties." However, in the deep poverty-stricken areas, the concept of education is backward, the self-sufficiency rate of education finance is low, and the strength of endogenous development is weak. We should increase the amount of education funds arranged by the central government, help the poverty-stricken areas fundamentally solve the problem of poverty, “make up the short board”, accelerate the standardization of education public services, and use standardization to drive equalization. We should attach importance to the support for the disadvantaged groups in education, subsidize girls to go to school, increase the number of years of education for rural women, and teach them how to fish, so that they can have the ability to create a better life.

5.6 Promoting education informationization and vigorously developing "Internet +" Education

Educational informatization is a shortcut to give full play to the advantages of technical means and realize the sharing of high-quality educational resources in ethnic areas. It is a "shortcut" to break through the space restrictions and make minority students and Han students learn and grow together. It is an important measure to promote educational equity. We

should vigorously promote the application of “Internet+education” in the compulsory education stage, and promote the public's support for policy implementation in the form of "Internet +". From the student point of view, through the “Internet+education” mode, we can get high-quality educational resources, break the traditional teaching mode of textbooks, desks and blackboards, and enjoy the rich teaching contents provided by the excellent teachers team through appropriate platforms and appropriate educational modes when rural compulsory education schools can’t provide high quality service. “Internet+education” is undoubtedly the best way for rural compulsory education schools to provide quality services. The best way to obtain educational resources. At present, the plan of “Internet+education” is not very difficult to land, but the internet penetration rate in rural areas needs to be improved. The state is in the stage of vigorously supporting education, and “Internet+education” is an inevitable trend of development. We should popularize “Internet+education” and guide all roles involved in education to correctly understand, narrow the gap between urban and rural areas, and promote the integration process of compulsory education in urban and rural areas.[24]

Conclusion

Fairness in education, especially in the stage of compulsory education, is the most easily perceived fairness in every family, and it is also the most important way to improve people's sense of happiness and gain. The unbalanced development of urban and rural education has become an important manifestation of the main contradictions in the field of education in the new era. To implement the integration of urban and rural education in ethnic minority areas, we should adhere to the people-centered principle, vigorously promote educational equity, drive the overall prosperity and development of urban and rural areas, make the achievements of educational development more equitably benefit the people of all ethnic groups, effectively guarantee the basic right to education of the people of all ethnic groups in rural pastoral areas, let every child have a chance to go to school, and let the people of all ethnic groups enlightens the mind through education, so as to realize the value of life and the all-round development of human beings.

References

[1] Changzheng Zhang , Zhijian Jia, Huaizu Li. Empirical research on the degree of education equity in China: 1978-2004 -- Based on the calculation and analysis of Education Gini coefficient[J]. Educational research of Tsinghua University, 2006 (2): 10-14
[2] Baicai Sun. Measuring educational equality in the past 30 years of China's reform and opening up -- An Empirical Analysis Based on Education Gini coefficient [J]. Education

research, 2009 (4): 12-18
[3] [11]Baicai Sun, Yang Zhang, Yunpeng Liu. Educational achievement and educational equity of Chinese ethnic populations: a comparison based on the data of the last three population censuses [J]. Ethnic studies, 2014 (3): 25-36 + 124
[4] [12]Dexin Hu. Regional comparative study on the degree of education equity between urban and rural areas in China [J]. Contemporary education science, 2017 (3): 14-18
[5] Wanming Zhang. Analysis of China's educational resource allocation and Policy Choice -- Based on the calculation of Education Gini coefficient[J]. Journal of Renmin University of China, 2013, (04): 89-97
[6] Zhenhua Wu, Xuemin Zhang. An Empirical Study on education equity of rural residents in China -- Based on the calculation and decomposition of education Gini coefficient from 1988 to 2012 [J]. Education economic review, 2017 (3): 20-37
[7] Yuetang Chen, Zhixiang Lei. Differences and trends in the development of educational equity in China: a regional comparison based on Gini coefficient of Education [J]. JOURNAL OF HUNAN AGRICULTURAL UNIVERSITY (SOCIAL SCIENCE EDITION), 2019,20 (3): 90-96
[8] SHUJIE YAO.On the decomposition of Gini coefficients by population class and income source: a spreadsheet approach and application[J].Applied Economics,1999,31（10）:1249-1267.
[9] Hangkong Zhang, Feixia Ji. Empirical research on education equity in China: 1982-2010 -- Analysis Based on the decomposition of Education Gini coefficient [J]. Educational science, 2013,29 (6): 1-6
[10] Junsen Zhang,Tianyou Li.International Inequality and Convergence in Educational Attainment, 1960–1990[J].Review of Development Economics,2002,6（3）:383-392.
[13] Wanpeng Zhang, Mengqi Li. Moving forward in hesitation -- exploring the way of education investment system with Chinese characteristics [J]. Education development research, 2020,40 (11): 1-7
[14] Hongqi Chu, Qian Zhao, et al. Educational reform in the process of urbanization [M]. Beijing: Educational Science Press, 2016:162-163
[15] Guangfei Wang, LINRONG Fu. The dilemma and Countermeasures of promoting the balanced development of compulsory education through the integration of urban and rural education [J]. Rural economy, 2018 (03): 112-117
[16] Bater Chen. Accelerating the modernization of Ethnic Education [J]. Ethnic education in China, 2020 (1): 15
[17] The data comes from the website of Education Department of Inner Mongolia Autonomous Region.
https://www.nmgov.edu.cn/hygq/yjta/zxta/202010/t20201021_59947.html
[18] Xiaoli Xu. Empirical research on the degree of education equity in China: 2004-2012 -- Based on the calculation and analysis of Gini coefficient of Education [J]. Education and teaching forum, 2018 (10): 211-213
[19] Compilation group. Development report of Chengdu Regional Education Alliance [M]. People's publishing house, 2015:62
[20] Fangting Qiu. Practical problems and Countermeasures of teacher flow in the integration of urban and rural education in County [J]. Education exploration, 2016 (2): 15-18
[21] Gengli Zhang. Practical difficulties and promotion strategies of Rural Preschool Education Collectivization Development from the perspective of urban rural integration [J]. Journal of educational science of Hunan Normal University. 2018 (03): 54-58
[22] Hongzhi Long, Huimin Yuan. Achievements, experiences and Reflections on the

development of preschool education in ethnic minority areas in China in recent ten years [J]. Contemporary education and culture, 2021,13 (01): 109-116
[23] Xi Jinping. How should we do well in education [M]// Xi Jinping. Get rid of poverty. Fuzhou: Fujian people's publishing house, 1992:129.
[24] Lijuan Wang. Differences in educational attainment and new ideas of targeted poverty alleviation in Liangshan area: analysis of Gini coefficient based on the educational years of ethnic and Han students (2000-2015) [J]. Research on ethnic education, 2019,30 (02): 22-30

Algorithm design and simulation of Rectilinear Steiner Tree problems based on concrete class Line

Man feng[1], Mingyang Li[2], Pengyuan Chen[3], Zhenping Lan[4],Ping Li(✉)
{ fengman2020@163.com[1],523155344@qq.com[2],554855487@qq.com[3],lanzp@dlpu.edu.cn[4],liping@dlpu.edu.cn(✉)}

Department of Information Science and Engineering Dalian Polytechnic University Dalian, P. R. China

Abstract. This graduation thesis mainly focuses on the specific RST, namely the Minimum Rectilinear Steiner tree, to study and improve the algorithm, and finally simulation and output results. Steiner tree problem belongs to combinatorial optimization discipline and is NP-hard problem. Over the past three centuries, many mathematicians have extended and extended this problem, making the steiner tree tree problem further developed. This paper mainly deals with a sub-problem of steiner tree, the Minimum Rectilinear Steiner tree problem. In this paper, based on dynamic programming, a classical technique to solve the steiner tree problem, two algorithms, L-MRST and Z-MRST, are proposed to construct the Minimum Rectilinear Steiner tree on a plane. Both algorithms use point data discretization for pre-tree preparation. However, in the same solution mode, the algorithm l-mrst directly iterates on the current subtree to build the final tree. The algorithm zmrst preprocesses the feasible z-layout standby of all points and enumerates all possible layouts when traversing the nodes. This algorithm increases the algorithm complexity, but the results are obviously optimized. Experimental results also bear this out. For the Minimum Rectilinear Steiner tree problem, the two algorithms in this paper have their own advantages and disadvantages, which can help VLSI problem to some extent, and can effectively optimize the large-scale wire network routing problem. The next research direction about the line steiner tree is also proposed.

Keywords: Minimum Rectilinear Steiner tree, Dynamic programming, VLSI, Algorithm Optimization.

1 Overview

1.1 A Sample Introduction

The Steiner tree (ST) problem is a combinatorial optimisation problem. It first sprouted in the early 18th century and appeared in the Fermat problem, which was popularised in the 19th century by the Swiss mathematician Steiner as "the problem of finding a point on a plane such that the sum of the distances from this point to a given number of points on the plane is minimised" [1]. Steiner had no further contribution to make to the Steiner Tree problem, but in his honour the problem was named after him.

✉ Corresponding author: Ping Li
(1969-),female, Professor, master, main research directions for optical communications and intelligence systems. E-mail:liping@dlpu.edu.cn

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307527

Over the centuries, several mathematicians have extended the problem, leading to a deeper development of the Steiner tree problem and a revision of the term "introduction of a point" to "introduction of several points". This extends the scope of its application and significantly increases the difficulty of the problem. The Steiner tree problem has been identified as an NP difficult problem [11]. Initially, researchers were only able to propose an approximation scheme (Polynomial-time Approximation Scheme, PTAS) for the Steiner tree problem, using heuristic algorithms such as the Genetic Algorithm or ant algorithm to find an approximation in the worst-case scenario. The solution is compared with the minimum spanning tree as a measure. It is only in recent times that exact algorithms for the Steiner tree problem have been developed. The first exact algorithms were proposed by Hakimi, a foreign researcher, and were based on an enumerative implementation, but with considerable limitations and high time complexity. Subsequently, new exact algorithms have been proposed and the Steiner tree problem has evolved by expanding its definition, enriching its structure and branching out. The Steiner tree problem has also been shown to be NPcomplete [1].

In recent years, the Steiner tree problem has been defined in detail as a graph consisting of n points on a plane, connecting these points and allowing the addition of up to N-2 points (the additional points are called Steiner points), with the shortest connected lines forming a tree called a Steiner tree. If no new points are added, the Steiner tree problem is equivalent to the classical computational minimum spanning tree problem. For example, it can become an Edge-Weighted Steiner tree (EWST) depending on whether the edges have value or not; similarly, it can become a VertexWeighted Steiner tree (Vertex-Weighted) depending on whether the vertices are assigned value or not. (Steiner tree,VWST). In general, the default weighted Steiner tree problem is the edge-weighted Steiner tree (EWST).

In addition, the Steiner tree problem is subdivided into sub-problems such as directed Steiner trees, undirected Steiner trees, special Steiner trees, bottleneck Steiner trees and parametric Steiner trees. In the case of the European Steiner Tree, which is a Steiner tree on the European plane, the two lines connecting the same point must be at an obtuse angle of 120°, whereas in the case of the Rectirectilinear Steiner Tree (RST), the two connecting lines through the same point must be at 90°, and are therefore also known as right-angled Steiner trees [4].

This thesis studies a sub-case of the rectilinear staner tree (RST) - the minimum rectilinear staner tree!

1.2 Concept and nature of the Steiner tree problem

The concept of a Steiner tree has been defined by many researchers at home and abroad and is basically the same. The shortest tree that connects all these terminal vertices is a Steiner tree (V is the set of vertices and E is the set of edges). The maximum number of related edges over a vertex is three (set). In addition to terminal vertices, there are also Steiner points in V. Steiner points are new points that are added during the connection process in order to have shorter connecting edges and are also known as auxiliary vertices. The Steiner point is the most fundamental feature of the Steiner tree. This is because if there were no Steiner points, i.e. V=R, the Steiner tree would be identical to the minimum spanning tree. There are no more than n-2 Steiner points in a Steiner tree [2].

The Steiner tree basically follows the above definition and is divided according to different requirements.

The Euclidean steiner tree (EST) is a steiner tree on the Euclidean plane, which requires an obtuse angle of 120° on both sides past a point and all steiner points must be within a convex bundle if the area enclosed by the connection of n terminal vertices is a convex bundle.

The line distance between two points a (x_1,y_1) and b (x_2,y_2) is calculated using the European distance d.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

However, in the case of the rectilinear staner tree (RST), it is required that the two sides past a point must be at 90° and that all lines are perpendicular, which is why it is also called the right angle staner tree or vertical static distance staner tree. The connection distance d_1 between two points a1 (x_1,y_1) and b1 (x_2,y_2) is calculated using the absolute distance (also known as the Manhattan distance, where two points can only be connected by a straight or polyline segment).

$$d_1 = |x_1 - x_2| + |y_1 - y_2|$$

I have used figures 1 and 2 to show the difference between a European Steiner tree and a straight Steiner tree.

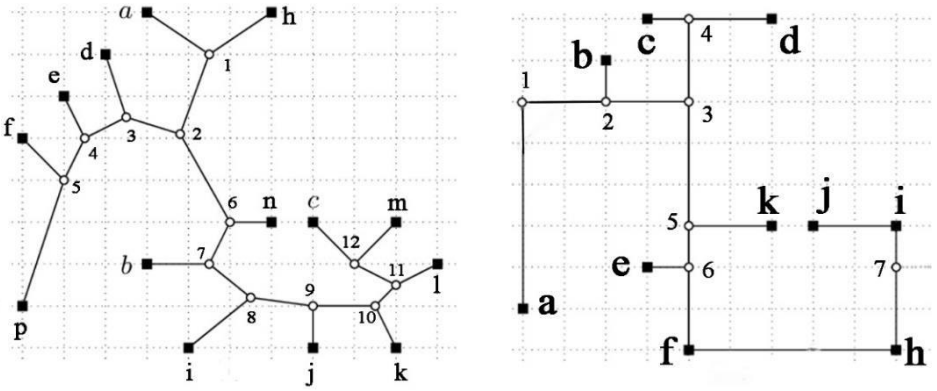


Fig. 1. Euclidean steiner tree and Rectilinear Steiner tree. The solid square dot marked by a letter represents the terminal vertex R; the hollow dot marked by a number represents the Steiner point S. The hollow dot is a hollow dot.

This thesis investigates a sub-case of the rectilinear Steiner tree (RST) - the minimum rectilinear Steiner tree (MRST). As the name implies, the minimum recti-linear Steiner tree is the Steiner tree with the shortest total length and the lowest total cost of connection among many rectilinear Steiner trees.

1.3 Key point analysis of the least rectilinear Steiner tree

The key to solving general Steiner trees is the determination of the Steiner point (S). The rectilinear Steiner tree is no exception, and even the Steiner point of a rectilinear Steiner tree can be better determined. A new concept, the Hanan grid, needs to be introduced here. At each point in a given set of terminal vertices R, a horizontal line and a vertical line lead to a point where the intersection is a Hanan point, and the resulting mesh is a Hanan grid. The mathematician Hanan, who first formulated the minimal rectilinear Steiner tree problem, has a famous Hanan's theorem. Hanan's theorem means that for any rectilinear Steiner tree, when the Steiner points are all located at Hanan's points, it is certain that there must be an optimal rectilinear Steiner tree, i.e. a minimum rectilinear Steiner tree [2].

Thus, the problem of finding a minimum rectilinear Steiner tree on a plane can be completed equivalent to finding a minimum rectilinear Steiner tree on a Hanan grid. The positions of the Steiner points are constrained. The key to the minimum Steiner tree becomes the filtering of

the Hanan grid points to determine the Steiner point positions and finally the rectilinear tree construction to solve the optimal tree [10].

1.4 Various known algorithms and their performance measures

First of all, it is clear that the Steiner tree problem is an NP Hard problem (nondeterministic polynomial (NP), so there may or may not be an infinite number of solutions. Of course, several researchers have proved that the Steiner tree problem is solvable, so there is only an optimal solution and not a single final solution to the problem. This is because as the mathematical body of knowledge and computer technology continues to develop, researchers continue to discover new optimisation algorithms and solutions, and the problem takes on a new lease of life. This is the beauty of the NP problem.

The minimum rectilinear Steiner problem discussed in this thesis and the proposed optimisation algorithm is only one of many methods, and the results are not necessarily optimal, but only partially based on an effective algorithm.

Approximate and exact algorithms. When the rectilinear Steiner tree problem was first studied, researchers were not able to come up with an exact algorithm, so research has focused on the Polynomial-time Approximation Scheme (PTAS) for polynomials. The performance measure of an approximation algorithm is the algorithm approximation, which is the ratio of the approximate solution found in the least ideal case to the optimal solution. The performance of the Steiner tree algorithm in comparison to the minimum spanning tree algorithm is also one of the key research questions. Berman et al. found that

$W_{ST}/W_m \leq 2(1-1/m)$

W_{ST} : Stannard tree weights W_m: the minimum spanning tree weight m: the minimum number of leaf nodes in the spanning tree

From 1980 to the 21st century, the approximation algorithm has been improved by many authors, resulting in a reduction in approximation from 2 to 1.55 and a reduction in time complexity from O(nlogn+m) to n^k (k approaching infinity)[2].However, approximate algorithms can only give approximate solutions and many problems do not have approximate algorithms, so researchers have started to investigate exact algorithms. The earliest exact algorithms were based on enumerative implementations, either based on spanning tree enumeration or on topology enumeration, and I have collated the exact algorithms proposed by some of the researchers in Table 1.

As can be seen from the table, branch definition algorithms have become the key to solving the problem, but the various algorithms have their limitations and deficiencies, are highly complex and are not suitable for situations where the number of vertices is large. Therefore research on the Steiner tree problem is still ongoing.

Table 1. Comparison of the exact Steiner tree algorithm.

Author	Methods	Time complexity	Limitations or deficiencies
Hakimi	Enumeration	$O(n^2\log n+mn+\min\{n^{k-2},2^{n-k}\}k^2)$	Only applicable to R=V
Shore	Branch definition	$O(24^n)$	High time complexity
Aneja,Maculan	Branch definition	$O(1.62^n)$	No time complications
Fomin	Treatment technology	$O(1.59^n)$	Clusters phenomenon

Parameterisation algorithms. Because of the large number of terminal vertices n and the complexity of the diagram in practical use, parametric algorithms have been developed.

Parameterisation algorithms are used to parameterise the Steiner tree problem. The classical parametric algorithm is the Dreyfus-Wagner algorithm, which is based on dynamic planning techniques. The implementation is based on the idea of decomposition, with three points as a small tree, from which a subset of the small trees is ultimately connected to form the entire large optimal Steiner tree, iteratively for each connected subtree. The complexity of the algorithm is O ($^{3rn+2rn2}$). The dynamic planning technique is also one of the most widely used techniques for solving Steiner tree problems. Later researchers applied tree decomposition, also known as path decomposition, to the investigation of the Steiner tree problem, which led to further optimisation of the Steiner tree parametric algorithm problem. Subset convolution and dynamic planning and capacitive theorem and dynamic planning techniques were later applied to optimise both spatial and temporal complexity, and the Steiner tree problem continues to progress.

Heuristic algorithms. A Heuristic Algorithm is an intelligent optimisation algorithm. Heuristic algorithms are usually found in the traditional problem solving experience of seeking a problem oriented strategy, which can then be used to find a relatively good solution in a feasible time frame. At this stage, heuristic algorithms are mainly nature-like algorithms, such as ant colony algorithms, simulated annealing algorithms, genetic algorithms and so on[5].

Simulated annealing algorithms refer to the problem of similar and combinatorial optimisation of the annealing process of crystalline material substances. If the annealing process is modelled, the objective function agrees with the energy function and the control parameter agrees with the important parameter temperature, then the solution space is regarded as the state space and the final simulation of the annealing algorithm's search for the base state (the lowest energy state in which a solid substance is formed) is the optimisation of the minimal (optimal) value of the objective function.

A colony algorithm is the ability of ants to release a pheromone on a path when searching for food. The more ants that travel a certain path, the more pheromone trails are left behind and the higher the probability that the ants will later choose that path. By means of this internal synergy, the ant colony gradually develops an optimum route. This optimal route is the optimal solution.

Genetic algorithms are the most basic type of intelligent optimisation algorithm. It is based on the Darwinian theory of evolution, which is based on the principle of survival of the fittest. "Genetic algorithms simulate the phenomena of mating, reproduction and mutation that occur in the process of natural selection and, according to the principle of survival of the fittest, use genetic operators (there are different genetic operators for different problems, i.e. variable parameters) to carry out calculations and select the best individual from generation to generation. Individual"[6], the optimal solution to be found.

Particle swarm optimisation algorithms are derived from genetic algorithms. It is an intelligent optimisation algorithm based on iterative thinking and can be used to solve most optimisation problems. Particle swarm optimisation is an improvement on the legacy algorithm, which is simpler and less complex than the genetic algorithm.

There are many other effective algorithms for solving the minimum rectilinear Steiner tree problem, but only some of them are listed here as a reference for optimising the algorithms.

2 Algorithm design

2.1 Algorithm design principles and processes

Principle: The algorithm is based on dynamic planning techniques. Dynamic planning is one of the most widely used techniques for solving Steiner tree problems. The algorithm uses decomposition thinking to find the optimal rectilinear Steiner tree for each subset, taking into account all decomposition cases[8].

Dynamic programming is a commonly used mathematical method, developed in the 20th century by mathematicians such as Bellman, which involves taking a large multistage problem and transforming it into a series of small single-stage problems, solving them one by one using the relationships between the stages, and ultimately optimising the solution.

The basic idea of a dynamic planning algorithm is to decompose the problem to be solved into a number of sub-problems, solve the sub-problems first and then derive the solution to the original problem from the solutions to these sub-problems. However, the sub-problems resulting from the decomposition are not normally independent of each other. Therefore, using dynamic planning techniques, it is possible to create an array of variables that hold the answers to the solved sub-problems (e.g. the edges of the spanning tree in this article) and then find the answers to the solved subproblems when needed, i.e. when performing the next recursion, thus avoiding the increased complexity and time costs associated with a large number of repeated calculations.

This is how the algorithm in this thesis works, first finding the root node, then recursively finding all the solutions, finding the optimal solution, moving on to the child nodes and continuing the recursion, but this is not a simple repetition, as each recursion is related to the solution already found, and does not generate a separate subtree, but ultimately a complete tree. Therefore, by using the 'table' idea of dynamic planning techniques, the complexity of the algorithm can be reduced, the algorithm can be run faster and redundancy reduced.

2.2 Design and research of algorithms

The algorithm uses the Strategy design pattern and the display specifies a class, named RST, which can use the changeStrategy function to change the solution mode used (L-MRST or Z-MRST), thus enabling both algorithms to run in the same situation. The ZRST class and the LRST class are used to solve these two different algorithms. At the same time the ZRST and LRST implementations have different interfaces, so the Strategy class also acts as an Adapter.

The RST class is a common solution mode for solving rectilinear Steiner trees. The programme starts with a randomly generated number of terminal vertices in the main function, which are added to the RST class and changed by the changeStrategy function into the L-MRST or Z-MRST solution mode, respectively. Some of the pseudocode is shown below.

The changeStrategy function is first defined to be a member of the RST class and, after a simple if statement and switch selection statement, ultimately returns myStrategyIdx_.

```
1 int RST ::changeStrategy(int s_idx) //define the change function
2 {
3       if (s_idx == myStrategyIdx_) return 0;//judgement
4       if (myStrategyIdx_)
5             delete myStrategy_;
6       myStrategyIdx_ = s_idx;//assign a value to the parameter s_idx
7      switch (s_idx)
8      case 1 :
9            myStrategy_ = new LRSTStrategy;
10                 break;
11           case 2 :
12               myStrategy_ = new ZRSTStrategy;
13                 break;
14      return myStrategyIdx_;//return tree
15 }
```

A new class is then defined, named the MST class, which is used to solve the rectilinear minimum spanning tree in preparation for the subsequent algorithm, which is used to solve for all the edges of the tree after generating the minimum rectilinear Steiner tree (as the minimum spanning tree was originally used when solving the minimum Steiner tree, and although the two trees are different, they have something in common).

The following is part of the code for the MST function. The first step is to traverse all nodes and find all edges using a two-layer for loop.

```
1       void MST::mst() {
2       const size_t n = m_vertexs.size();
3       Dist dist dist_max(std::numeric_limits<int>::max(), 0, 0);
4       vector<Dist> dist(n, dist_max);
5       vector<size_t> parent(n, 0);
6       vector<bool> in_tree(n, false);
7       size_t t = 0; // dist[t] = Dist(m_vertexs[t], m_vertexs[t]);
8       for (size_t k = 0; k < n - 1; k++) {
9             in_tree[t] = true;
10              for (size_t j = 0; j < n; j++)
11                    if (j ! = t) {
12                          Dist tmp(m_vertexs[t], m_vertexs[j]);
13                          if (tmp < dist[j] && !in_tree[j]) {
14                                dist[j] = tmp;
15                                parent[j] = t; // exchange nodes
16                          } }
17              Dist min = dist_max;
18              size_t min_v = 0;
19              for (size_t i = 0; i < n; i++) {
20                    if (dist[i] < min && !in_tree[i]) {
21                          min = dist[i];
22                          min_v = i;
23                    } }
24              t = min_v;
25              m_lines.push_back(Line(parent[t], t));
26      } }
```

A number of constants used in the algorithm are then defined in the Common function, including NUM, i.e. the number of vertices for a given endpoint. By changing the value of NUM in Common.h, a minimum rectilinear Steiner tree can be generated with the corresponding number of vertices.

Finally, the Visualizer class is also specified, which plays a role in the display. The pseudo-code is displayed as follows.

All generated edges are first processed with a for loop, stored and coloured black, and then all points, including terminal vertices and steiner points, are processed with a for loop, also coloured black, for display in the final generated tree.

```
1 void Visualizer::show(const RST* rst)
2 {
3   Mat picture(name);
4   for (const auto& seg : rst   ->v_seg)//processing the generated edge
5   {
6       line(picture,
7                 horizontal axis coordinates, vertical axis coordinates),
8         cv::Point(seg.v.x, seg.v.y),
9         //store as black
10    }
11   for (const auto& p : rst   ->v_op) // Process NUM points
12   {
13       circle(picture,
14             cv::            Point(p.x, p.y),
15                        cv::Scalar(0, 0, 0));
16    }
17 namedWindow("RST");//Naming
18 imshow("RST", picture);//Show
19 }
```

2.3 Implementation and comparison of the two algorithms

The following discussion focuses on the implementation of the LRST and ZRST algorithms. Both algorithms rely on the base class Line and the derived class Line_L. Line_L and Line_Z are L- and Z-shaped Layout. segment is used mainly in the peripheral RST and Visualizer classes to represent each horizontal or vertical line segment (where the polyline has been split). This differs when implementing the construction of a rectilinear Steiner tree in concrete terms.

L-MRST based on the derived class Line_L. LRST is to select an "L-shaped" or "U-shaped" layout for each edge and, out of all the layouts, find the one with the largest overlap. Due to the separable nature, the sub-problems can in fact be independent and the exhaustive search can be translated into a dynamic planning process in the tree.The tree therefore first needs to be created. A node of degree 1 needs to be chosen as the root (this is to ensure that there is initially only one edge). The tree is then built recursively using the list of edges found in the MST function. The tree is represented by two arrays: tree and parent. each element of the tree is a vector, which represents the successor node of a node. parent represents the parent node.Since any resulting edge is only possible on a grid line with dotted x- or y-coordinates (Hanan grid), it is advantageous to discretize the coordinates for subsequent processing. All that is required for discretization is to obtain the coordinates of all points, then sort and de-

weight all x,y coordinates, the resulting ordinate number being the discretized coordinate number. This is done using the map function in STL (the internal implementation of map is an ordered red and black tree). x_coord and y_coord variables record the original coordinates of all the discretized coordinates and discr_points records the discretized coordinates of all the points, respectively.The functions find_layout_L and find_layout_U recursively solve for all cases. Starting from the only child node of the root, each call solves for the optimal solution (the solution is the choice that the children of the current node should adopt) in the case where the edge between the current point and the parent node is chosen L (or U) in the current subtree. The result of the solution for each choice at each point is placed in the layout_l and layout_u arrays, using binary coding, for each bit of code for the choice of a child node (0 for L, 1 for U).An exhaustive search of all possible case points for all children of a node is done recursively using the dfs function. Part of the pseudo-code is shown below.

```
1 dfs(int parent, std::vector<int> &kids,
2         size_t num,
3         int &value, int &result, int choice, int &decision) {
4       if (num == kids.size()) {
5             if (result < value)
6                   { result = value; decision = choice;}
7         // if the shape is L
8         paint(discr_points[parent], discr_points[kids[num]], false, 1, value);
9         value += layout_l[kids[num]];
10          dfs(parent, kids, num + 1, value, result, choice, decision);
11          value -= layout_l[kids[num]];
12          paint(discr_points[parent], discr_points[kids[num]], false, -1, value);
13 // if the shape is U
14        Repeat,false replase true.
```

Z-MRST based on the derived class Line_Z. The first step of the second algorithm requires the discretization of the terminal vertices as well as tree building. These two processes are the same as for L-MRST.

The difference is that for each edge, all possible Z-layouts are pre-processed first. This is achieved with the dfs function.

After the tree has been built with the build_tree function, it is called recursively from the root (with the find_root function) (here with the find_layout function). For each node that is traversed, all of its children are traversed and then all possible layouts for the current point are enumerated (i.e. all combinations of layouts for the children of the current node). The other processes and coordinate points are stored in the same way as the L-MRST algorithm.

The optimal layout for the current point is obtained by the get_result function, which is iterated over and over again to obtain the minimum rectilinear Steiner tree.

```
1 void ZRST ::dfs(int root, int father, size_t stat, layout &lay, vector<size_t>& stack) {
2 if (tree[root].size() == 0)
3 lay.sub_ans = 0;//initialisation
4       int id = stat, son;
5     if (stat == tree[root].size()) {// Traverse from the root node
6 int ans = 0;
7 vector<Line_Z> overLap;
8 for (int i = 0; i < id; i++) {
9          ans += layouts[son = tree[root][i]][stack[i]].sub_ans;
          overLap.push_back(Line_Z(root, layouts[son][stack[i]].mid_point));
10          overLap.push_back(Line_Z(father, root, lay.mid_point));
11          ans += overlap(points(), overLap)
12 point(root).distance(point(father));
13      if (sans < lay.sub_ans) {
14          lay.sub_ans = ans;
15          copy(stack.begin(), stack.begin() + id, lay.best_lay);}
 else {son = tree[root][stat];
16          for (size_t i = 0; (stack[id] = i) < layouts[son].size(); i++) {
17                  dfs(root, father, stat + 1, lay, stack);//continue iterating
18 }
```

2.4 Algorithmic simulation implementation

This algorithm uses command line parameters to control the behaviour of the programme. "LRST" and "ZRST" run the two algorithms respectively. "test.txt" is used to generate the new test data. This means that the input values are the same (same randomly generated number of end vertices), the output form is the same and the spanning tree attribution class is the same, the only difference is the tree generation process.

Output form: Each run displays the resulting graph of the solution and the final total tree length using opencv.

The number of terminal vertices and the constants of the file names can be changed in the code main.cpp and common.h.

3 Analysis of simulation results

3.1 Experimental platforms and use cases

The algorithm experiments are carried out on the same experimental platform in both a hardware and a software environment.

(1) Hardware environment

Mainframe: Intel® Core™ i7-7500U CPU, 8GB RAM

(2) Software environment

Host: Microsoft Windows 10 (64 bit edition), Microsoft Visual Studio 2019, OpenCv.

(3) Software used

Microsoft Visual Studio 2019, OpenCv

The simulation of the two algorithms with different numbers of vertices is based on the same randomly generated Hanan grid structure, where the horizontal and vertical distances of

the lattice points are considered to be of unit length 1. Six test cases with 5, 10, 20, 20, 100 and 500 vertices are selected for the experiment. The resulting graph of the minimum rectilinear Steiner tree is recorded and represented as opencv, the total tree length for each case is output for both algorithms and a txt document is generated to record the coordinates of each point.

3.2 Experimental results

Six groups of experimental cases with a number of terminal vertices of 5, 10, 20, 20, 100 and 500 were selected for the experiment. Below is the total tree length generated for each group in both algorithms.

Table 2. Simulation results.

Case studies	Number of terminal vertices	L-MRST total length	Z-MRST total length
1	5	18	18
2	10	65	65
3	20	159	159
4	50	168	168
5	100	381	379
6	500	8480	8432

The graphical results on the final spanning tree are represented by cases 3, 4 and 6.

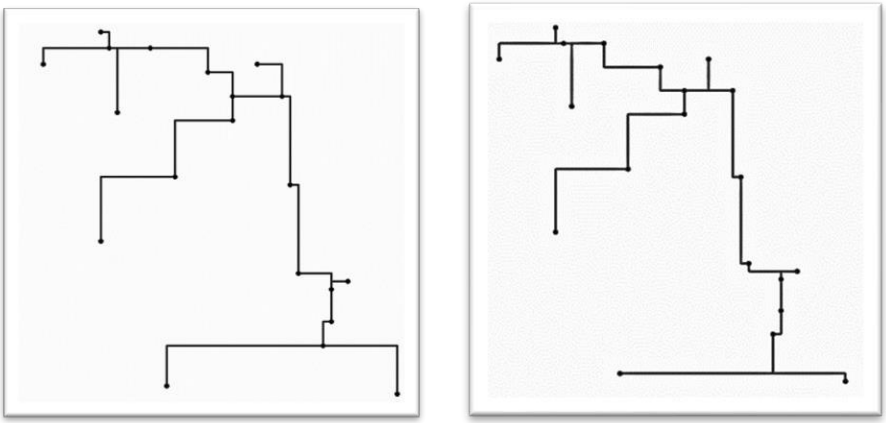


Fig. 2. Case 3.L-MRST & Z-MRST

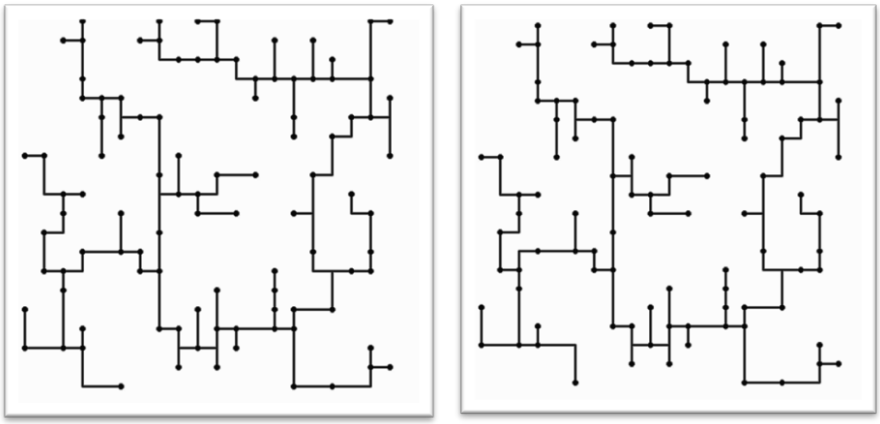


Fig. 3. Case 4.L-MRST & Z-MRST

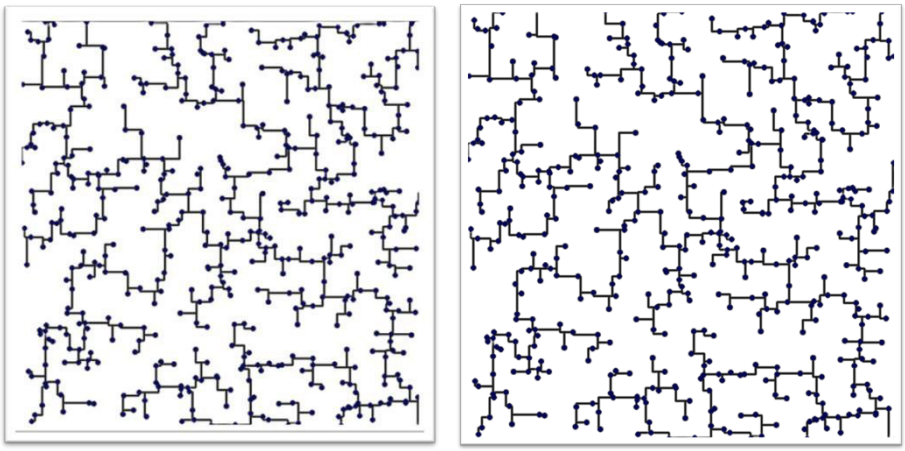


Fig. 4. Case 6.L-MRST & Z-MRST

3.3 Comparative analysis of simulation results

Simulation results: For both the L-MRST and Z-MRST algorithms, the total length of the rectilinear Steiner spanning tree is practically indistinguishable for a small number of vertices (n) (less than 50) and there is a slight difference in the graph of the spanning tree. With a larger number of vertices, Z-MRST has a shorter total rectilinear Steiner spanning tree length and produces better results than the L-MRST algorithm. This advantage is even more pronounced when the number of vertices is higher. Analysis of the results: It can be seen that Z-MRST gives slightly better calculation results than L-MRST. This is due to the fact that Z-MRST takes more cases into account and traverses more leaf node layouts. z-MRST pre-processes all the nodes with possible z-layout spares, because a z-layout can be uniquely identified by its inflection point and so a layout can be represented by saving the coordinates of this inflection point. for each node traversed. All of its children are traversed and then all possible layouts for the current point are enumerated (i.e. all combinations of layouts for the children of the current node). This results in a more optimised rectilinear Steiner tree, but the complexity of the algorithm increases accordingly and it takes longer to get the results from

the running program.

In summary.

When the number of terminal vertices is small, the two algorithms produce similar rectilinear Steiner trees with the same total length.

The Z-MRST calculation outperforms the L-MRST calculation with a larger number of terminal vertices. this advantage becomes more pronounced as the number of terminal vertices increases.

The Z-MRST algorithm is more complex and L-MRST is faster than Z-MRST in terms of algorithm running time, but this advantage is not obvious if the number of terminal vertices is small.

4 Summary and outlook

4.1 Summary

With regard to the rectilinear Steiner tree problem, two algorithms are presented in this thesis based on dynamic planning techniques, MRST for line_L and MRST for line_Z. Both algorithms are based on the base class line derivation. This is because Hanan's theorem states that an optimal rectilinear Steiner tree exists when every Steiner point lies at the intersection of two vertical lines containing terminal vertices. The key to MRST is therefore the determination of the Steiner point, which must lie on the Hanan grid. So the algorithm starts by discretizing and ordering the positions of the given terminal vertices, and segment is used to represent each horizontal or vertical line that passes through these points, which in combination is the Hanan grid. I have put both algorithmic approaches in the same program, the input and output functions are common, differ only in the subtree iteration and are displayed via opencv using a common Visualizer function.

From the analysis of the final results it can be seen that the Z-MRST calculations are slightly better than the L-MRST calculations, an advantage that is more obvious when the number of vertices is higher. This is because Z-MRST takes more scenarios into account during iteration by pre-compiling all possible edges. However, the time cost of harvesting a better solution is increasing, especially with a given number of terminal vertices.

Comparing the two algorithms, each with its own advantages and disadvantages, the time and cost of optimising the results is increasing. Therefore, I believe that the algorithm written in this paper has a lot of room for improvement. It can be further optimised from the point of view of reducing redundancy and unnecessary recursions in order to obtain better results.

4.2 Outlook

This paper has achieved some success in the study of the algorithm for the smallest tree of the rectilinear Steiner tree, but there is still much room for improvement. Because the Steiner tree problem is an NP hard problem, this paper presents only the simplest branch of the problem concerning rectilinear Steiner trees. The Steiner tree problem has a very wide range of applications in many fields. They play an important role in computer networks, in the field of biology, in the humanities and in genetic engineering. The rectilinear Steiner tree problem is a very basic problem in VLSI design and has applications in the design phases of VLSI such as physical synthesis, cabling planning, interconnection planning, global and detailed cabling, etc. It can be used to generate optimal cabling solutions and significantly reduce cabling time. Although some minimum rectilinear Steiner tree algorithms and optimisations are given in this paper, there are many more factors that can interfere in real life use. For example, with real

integrated circuit wiring, there are not only endpoints on the circuit board, but also many other components such as chips, buzzers etc. For these "obstacles" on the circuit board, a minimum rectilinear Steiner tree can be used in the MRST. These 'obstacles' on the circuit board can be further investigated on the basis of the MRST (Minimum Rectilinear Steiner Tree), called the Obstacle-avoiding Rectirectilinear Steiner Minimum Tree (MRST). ORSMT also known as the Restricted Version Bottleneck Rectilinear Steiner Tree[7].

There is still a long way to go with regard to the research on the rectilinear Steiner tree problem for realistic VLSI wiring. How to ensure quality, even if the resulting lines are the shortest and the wiring time is reduced, and how to make the algorithm stable and optimised, are key issues for future research on the rectilinear Steiner tree problem.

It is hoped that future researchers will continue to explore the rectilinear Steiner book problem and optimise the algorithm for application in real production[9].

References

[1] Viet Min-yi. Introduction to combinatorial optimization. Hangzhou: Zhejiang Science and Technology Press(2001).

[2] The Steiner Tree problem[J]. Computer Science and Exploration, 2011,38(10):16-22.

[3] Yusheng Zhang , Zhongbin Wang. Minimum rectilinear Steiner tree and its application in detailed wiring[J]. Journal of Hefei University of Technology (Natural Sciences), 1992(S1):100-107.

[4] Huimin Jin, Liang Ma, Zhoumen Wang. Fast algorithm for Euclidean Steiner's optimal tree[J].Computer Applications Research,2006,23(5):60-62.

[5] Jin Huimin, Ma Liang, Wang Zhoumen. Intelligent optimization algorithm for the Euclidean Steiner minimal tree problem[J]. Computer Engineering,2006,32(10):201-203. DOI:10.3969/j.issn.1000-3428.2006.10.075.

[6] Chen Xiuhua, Zhu Natural. A PSO algorithm for solving the RSMT wiring problem[J]. Journal of Minjiang University, 2014,35(5):39-44.

[7] Zhang Hao, Ye Dongyi, Guo Wenzhong. A multilayer barrier-wound right-angle Steiner minimal tree heuristic[J]. Small Microcomputer Systems, 2016, 37(08):1760-1764.

[8] Zhang Hao, Ye Dongyi, Guo Wenzhong. A multilayer barrier-wound right-angle Steiner minimal tree heuristic[J]. Small Microcomputer Systems, 2016, 37(08):1760-1764.

Yang Changling, Yan Xiaolang. MRST hybrid genetic algorithm based on tree coding and its parallel processing[J]. Microelectronics, 1999,29(2):89-95.

[9] Jun Ma,Bo Yang,Shaohan Ma. A practical algorithm for the minimum rectirectilinear steiner tree[J]. Journal of Computer Science and Technology,2000,15(1).

[10] S. D. Lin and D. H. Kim, "Construction of All Rectirectilinear Steiner Minimum Trees on the Hanan Grid and Its Applications to VLSI Design," in IEEE Transactions on ComputerAided Design of Integrated Circuits and Systems.

[11] Garey M R, Johnson D S. The rectirectilinear Steiner tree problem is NP-complete[J]. SIAM Journal on Applied Mathematics, 1977, 32(4): 826-834.

Intelligent control system of tea garden lamps based on ZigBee

Zhanzheng Xu[1], Yingming Gao[2], Yidi Chen[3], Huiyang Feng [4], Guanying Cao*
{xuzhanzheng@163.com, gycao@dlpu.edu.cn}
Research Institute of Photonics, Dalian Polytechnic University Dalian 116034，China

Abstract. An intelligent control system for tea garden lamps is designed based on ZigBee technology. Image sensors, photosensitive sensors, temperature and humidity sensors located at the ZigBee terminal nodes are applied to monitor the tea garden environment. Real-time information of the tea garden is obtained and transported to the intelligent terminal by wireless communication network as feedback. The intelligent terminal releases instructions to control the working state of the pest phototactic lamps located at the terminal nodes which induces the pests in the tea garden to be killed by the frequency-vibration insecticidal lamps. According to the wavebands that promotes the photosynthesis of tea, a tea growth lamp is designed to solve the lack of light encountered during the growth of tea. The systematic design can not only effectively control pests, improve tea quality and yield of tea, but also reduce the use of chemical reagents and protect the natural environment.

Keywords: ZigBee technology, intelligent control, tea garden lamps, sensors, phototactic lamps

1 Introduction

With the improvement of living standards, tea is gradually indispensable when meeting guests. At the same time, tea gardens have become an important part of the agricultural field and have developed into an eco-tourism resource. However, tea garden pests can cause great harm to tea gardens [1], which has always plagued tea farmers. The traditional pest measures in tea gardens include insecticide board[2-3], insecticidal lamp[4-6] and chemical reagents. However, when the insect board encounters rainy or

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307434

high temperature weather, the effect of the insect board will be greatly reduced. Ordinary insecticidal lamps have a single trapping type of pests, and the effect is not obvious. Using chemical reagents to kill pests will not only affect the quality of tea, but also pollute the ecological environment.

In order to pursue an intelligent and humanized life, ZigBee technology is widely used in various places. Kumar A developed an animal health monitoring system by ZigBee technology, which can acquire animal physiological information in real time[7]. Zhang Q discussed the wireless solution of greenhouse monitoring system based on ZigBee technology, and designed the wireless node, network construction and software system [8]. Khusvinder Gill uses ZigBee design to realize a home automation system, which users can access and control remotely [9].

In order to realize the intelligent management of tea gardens, this paper designs a tea pest control system with controllable lights based on ZigBee technology. The system mainly uses monochrome LED, wireless communication technology and sensor technology to control the work of tea garden lamps to achieve the purpose of attracting different pests in different periods of time in the tea garden.And use frequency-vibration insecticidal lights to trap and kill the attracted pests. The system can not only effectively improve the management of tea gardens, but also improve the quality and yield of tea.

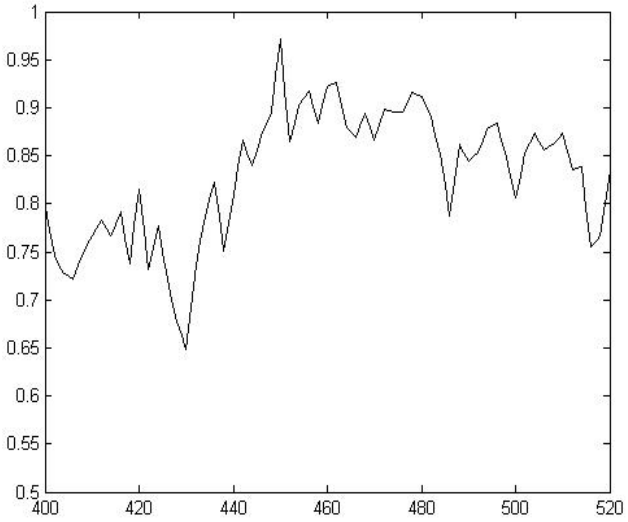
2 Design of tea garden lamps

In the process of tea growth, most tea gardens will face the dilemma of insufficient light. This dilemma will result in slow growth of tea and delayed picking. At the same time, pests such as tea loopers and tea leaf rollers will eat tea leaves, causing the appearance of the tea leaves to be incomplete, affecting the yield of tea and reducing the income of the tea garden. In view of the above problems, the purpose of promoting the growth of tea and controlling pests can be achieved by designing tea garden growth lamps and pest phototactic lamps.

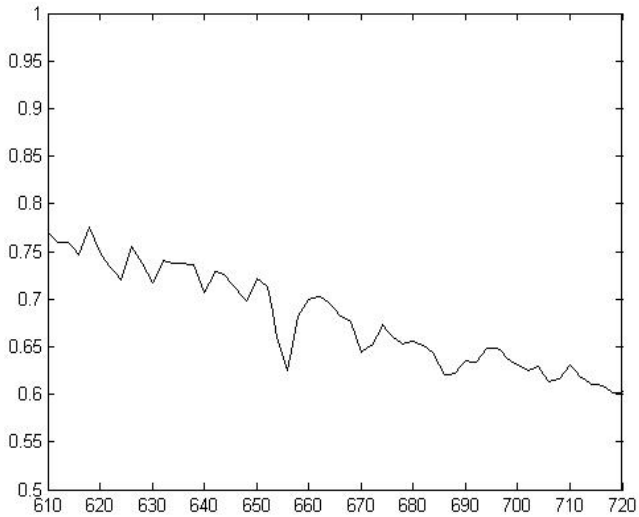
2.1 Tea growth lamps

Wang Jiazhen et al studied that the two wavelengths of 400-520nm and 610-720nm contribute the most to the photosynthesis of plants[10]. Therefore, these two bands can be intercepted on the sunlight spectrum as the ideal lamp spectrum for promoting the growth of tea. And the best LED combination scheme can be obtained by using the least square method. By rationally arranging the LEDs, the light distribution is reasonable

and the illuminance is uniform.This solution is used to supplement the light of the tea, make up for the lack of light and promote the growth of the tea. The ideal spectrum of tea garden growth lamps is shown in Figure 1.



(a) blue violet spectrum



(b) red spectrum

Fig. 1. Ideal growth lamp spectrum of tea garden.

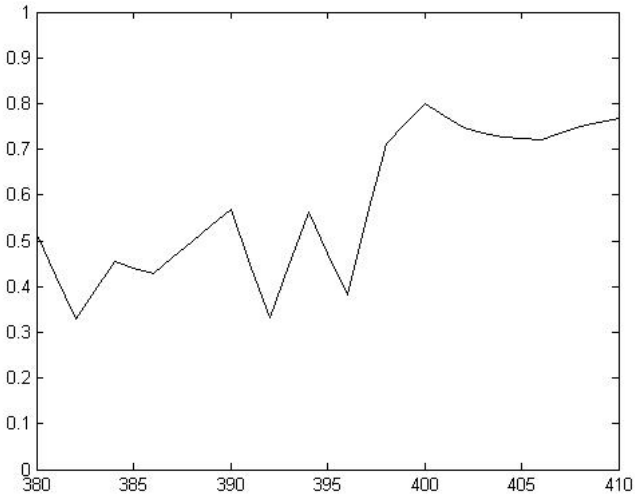
2.2 Phototactic lamps and insecticidal lamps

Phototaxis is one of the most important tropisms formed by insects in the long-term evolution[11-12]. They are sensitive to the waveband of 250-700nm. Different types of pests in tea gardens have different light-taxis bands, as shown in Table 1.

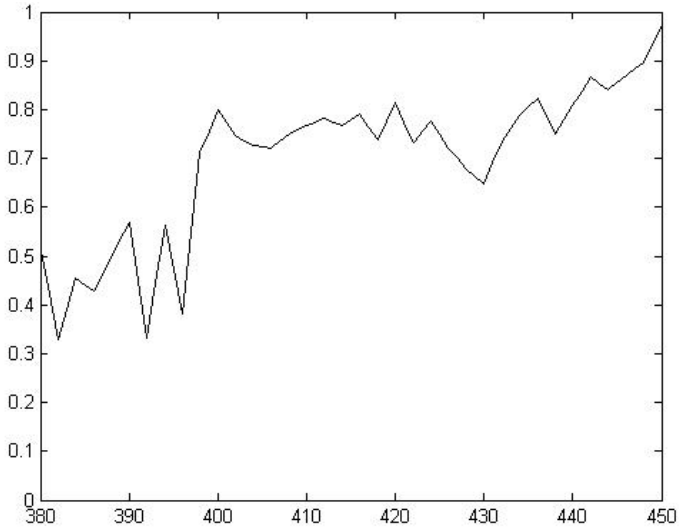
Table 1. Phototactic band and flapping rhythm of main pests in tea garden.

Main pest specie	Phototactic band /nm	Rhythm of lighting
Tea geometrid	380—410	19:00—24:00
Empoasca vitis Gothe	410—420	17:30—19:30
Noctuid	330—410	20:00—24:00
Tea leafroller	330—400	19:00—23:00
Aeolesthes induta Zewman	380—450	18:00—24:00

Taking tea geometrid as an example, tea geometrid is most sensitive to the waveband of 380-410nm. According to the phototactic waveband of tea geometrid, the 330-410nm spectrum can be intercepted on the fitted spectrum as the ideal phototatic spectrum. By using the above method, the ideal phototaxis spectrum of other pests can be obtained, as shown in Figure 2.



(a) phototactic spectrum of tea geometrid



(b) phototactic spectrum of Monochamus alternatus

Fig.2. Phototactic spectrum of tea garden pests.

3 Tea garden lamps control system

In recent years, wireless communication technology has developed rapidly and ZigBee technology has gradually matured.The advantages of low power consumption, short time delay and high security make it more eye-catching when people choose intelligent control methods.

3.1 System framework

According to the difference in phototaxis of different pests at different periods and the light-fluttering rhythm [13-14], the information is obtained and transported to the smart terminal by wireless communication network as feedback and controls the corresponding lights to work, effectively trapping pests and ensuring the normal growth of tea. The framework of the intelligent control system for tea garden lights based on ZigBee technology is shown in Figure 3.

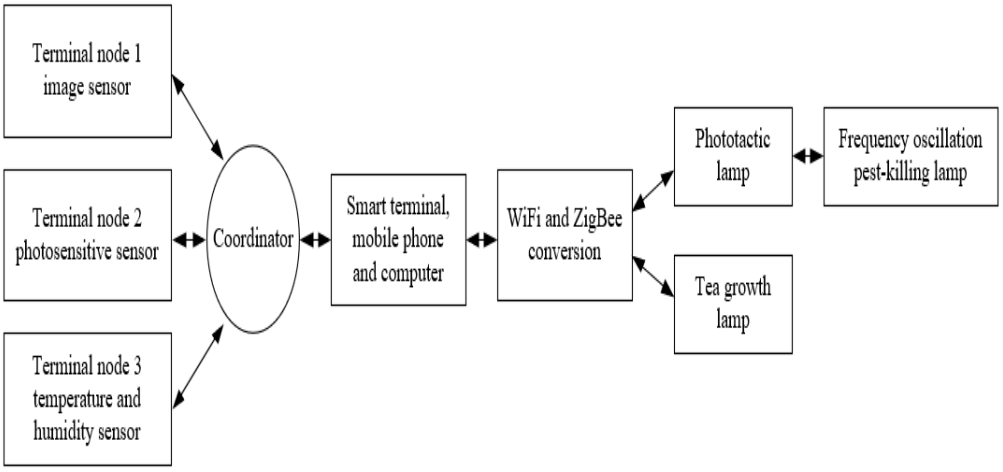


Fig.3. Framework diagram of intelligent control system for tea garden lamps.

The whole system consists of WiFi wireless local area network, ZigBee wireless personal area network, monitoring module and pest control module. The monitoring terminal node will access the preset parameters into the ZigBee LAN, and send the data measured by each sensor at the terminal node location to the coordinator in real time. The coordinator establishes a local area network according to the preset parameters and transmits the information fed back by the monitoring system to the intelligent terminal through the data interface and WiFi network. The intelligent terminal controls all kinds of phototactic lamps to work independently according to the set program instructions, and traps and kills the trapped insects through the frequency-vibrating phototactic lamp.

3.2 Monitoring system

The monitoring system uses sensors to monitor the environment, and sends the information to the host computer through the wireless network after processing. The monitoring module system is shown in Figure 4.

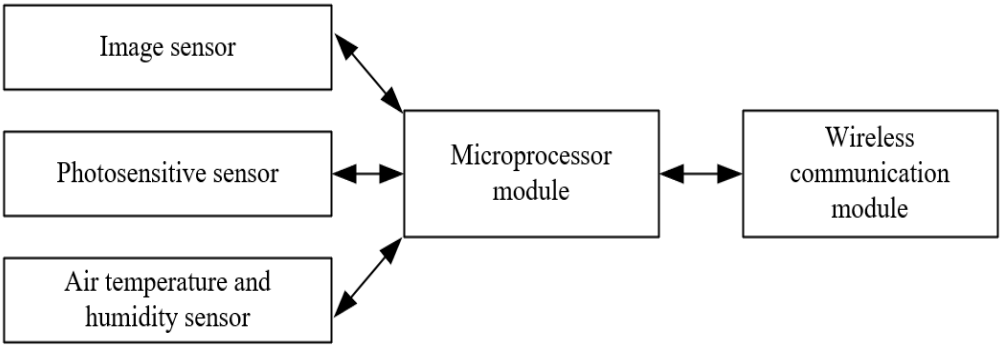


Fig.4. Monitoring module system.

(1) The image sensor is mainly used to monitor the safety of tea gardens. It is made with a 1/4 inch OV2640 million HD CMOS sensor. It has the characteristics of high sensitivity, high flexibility, supporting JPEG output and also contains data interface and control interface, which can well adapt to the different needs of the system.

(2) The photosensitive sensor is mainly used to measure light intensity information in rainy weather. When the light intensity is lower than the light compensation point suitable for the photosynthesis of tea, the managers can control the tea-growth lamp to work through the information fed back by the photosensitive sensors.

(3) The temperature and humidity sensor adopts SM51 series in the sonbest brand. This kind of sensor has low power consumption, wide measurement range and high accuracy. When the temperature and humidity are abnormal, the information can be fed back to the intelligent terminal, which is convenient for management personnel to deal with in time.

(4) The microprocessor module can mainly complete operations such as fetching instructions, executing instructions and exchanging information with external memory and logic components.

(5) The wireless communication module is responsible for establishing the network connection between each module which can realize the transmission function between the data, and has low cost and good applicability.

3.3 WiFi and ZigBee network module

Both WiFi and ZigBee networks are widely used wireless network communication technologies today. However, the two network protocols are not compatible. You can

add a WiFi endpoint at the ZigBee gateway to communicate together to achieve data transmission. The WiFi local area network in this system can realize the network connection between the sensors of the monitoring system and the smart terminal (mobile phone, computer, etc.), so that the administrator can control the operation of the system through mobile phones and other devices. ZigBee network is mainly used in the field of automatic and remote control, which can realize the network connection of various lamp driving and control equipment. Reasonable use of these two wireless network communication technologies can make the management more intelligent and humanized.

3.4 Tea garden lamps module

In order to solve the problems of pests in tea gardens and lack of light in tea gardens, according to the light-tending band of the pests, monochromatic LEDs are used to design all kinds of pest locating lights, and then the pests are killed by frequency-vibrating insecticidal lights. The light intensity information of the tea garden is obtained and transported to the intelligent terminal by wireless communication network as feedback.When the light intensity is less than the light compensation point that promotes the photosynthesis of the tea, the system will control the light that promotes the growth of the tea to work.

In this paper, through the topological structure of the ZigBee star network, the image sensors, photosensitive sensors, temperature and humidity sensors, tea growing lights, and pest phototactic lamps are placed at the terminal nodes of the network. The information is forwarded through the coordinator to realize communication between two terminal nodes (such as photosensitive sensors and tea growth lamps).

3.5 ZigBee network system

In the entire ZigBee network, the node connection includes the connections between various sensors located at the the terminal nodes and the coordinator nodes. The connection between the lamps (the tea growth lamps and the pest phototactic lamps) located at the terminal nodes and the coordinator nodes. The photosensitive sensor node is successfully connected with the coordinator first, and then transmits the real-time monitoring light intensity information to the coordinator. The coordinator stores the information and compares the size with the light compensation point for photosynthesis of tea. If the light intensity measurement value is less than the light intensity value of

the light compensation point, a command will be sent to control the operation of the tea growth lamp.The flow chart of node establishment is shown in Figure 5.

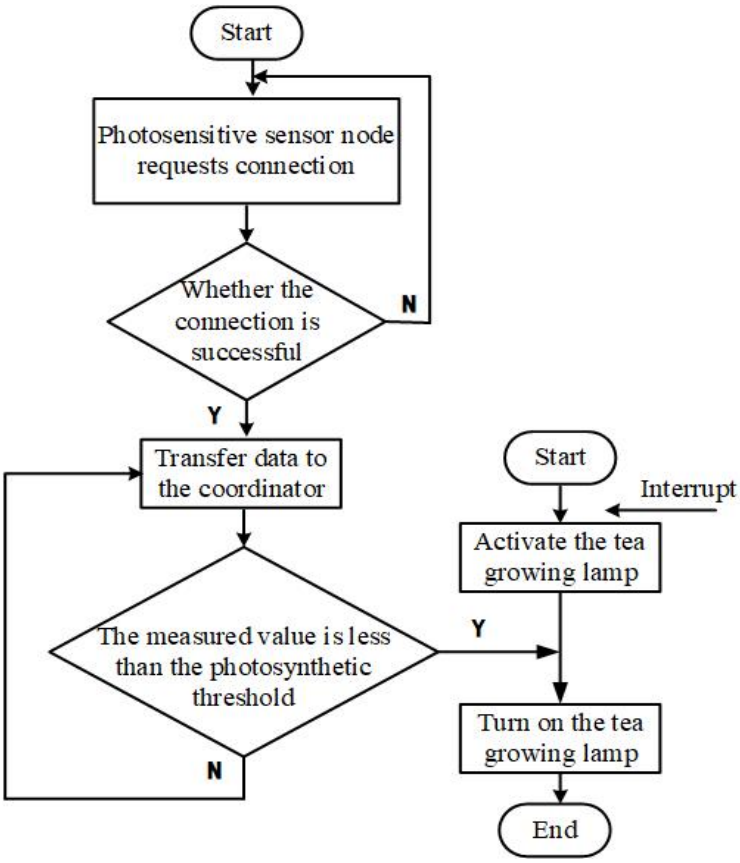


Fig.5. Terminal node establishment process.

4 Conclusion

The ZigBee-based pest control system for tea gardens designed in this paper, with the characteristics of ZigBee's low power consumption and short time delay which enables the system to be installed flexibly at low cost. In addition, this system uses a frequency-vibrating insecticidal lamp to catch and kill the attracted pests, which reduces the environmental pollution caused by chemical reagents. With the increasing demand for pollution-free products, the system will have a good market prospect.

References

[1] Qi L I , Hong-Chun Z , Ji-Cai T , etc.: Occurrence and Damage of Lepidopterous-pest in Tea Gardens of Hunan[J]. Tea Communication, (2011).

[2] Hongtao Wang, Faxiang Jiang, Min Chen, etc.: Trapping effect of different color sticky insect boards on Main Pests in tea garden [J]. China plant protection guide, (2018).

[3] Yuantao Yao, Changjun Wang, Lili Tian, etc.: Study on the control effect of insect proof board on tea aphid in northern asexual tea garden [J]. Northern horticulture, (1): 112-114(2014).

[4] Li, Tianhua, Pan, et al.: Research on the LED Pest Control Light for Tea Plantations[J] Journal of Computational and Theoretical Nanoence, 13(4):2296-2299(2016).

[5] Zhaoming W, Jingfeng W, Youbing P.: The Effect of Jiaduo Multi-frequency Vibration-killing Lamps Controlling Pest in Organic Tea Garden [J]. Chinese Agricultural Science Bulletin, (2009).

[6] Xiang Xu, Li MA, Zelin Yue and other.:LED single wavelength lamp insecticidal lamps on tea garden pests [J]. China plant protection guide, 37 (12): 53-56(2017).

[7] Kumar A, Hancke G P.: A ZigBee-based animal health monitoring system[J]. IEEE sensors Journal, 15(1): 610-617(2014).

[8] Zhang Q, Yang X, Zhou Y, etc.: A wireless solution for greenhouse monitoring and control system based on ZigBee technology[J]. Journal of Zhejiang University-Science A, 8(10): 1584-1587(2007).

[9] Gill K , Yang S H , Yao F , etc.: A ZigBee-based home automation system[J]. IEEE Transactions on Consumer Electronics, 55(2):422-430(2009).

[10] Jiazhen Wang, Xinyu Zhang, Xing Jin, etc.: Effect of night light supplement on Yield and quality of spring tea in red and blue lamp tea garden [J]. Journal of Fujian agriculture, 34 (1): 46-52(2019).

[11] Lei Bian, Xiaoling Sun, Yu Gao, etc.: Phototaxis mechanism and application progress of insects [J]. Acta Entomologica Sinica, 49 (6): 1677-1686(2012).

[12] XiangFeng J, Chaoliang L.: Advances in research on photoaxis of insects and the mechanism[J]. Kun Chong zhi shi= Kunchong Zhishi, 41(3): 198-202(2004).

[13] Haihua Tu, Xianfei Kuang, Yu Mao, etc.: Control effect of dual band led solar insecticidal lamp on tea pests [J]. China plant protection guide, 38 (10): 53-57(2018).

[14] Haihua Tu, Xiuxia Hu, Yu Mao, etc.: Development and application of dual spectrum led insecticidal lamp for tea garden [J]. Anhui Agricultural Sciences(3): 36(2019).

Light Environment Design to Enhance the Uniformity of Strawberry Canopy Lighting in Greenhouse

Xiaoyang He [1,2,3*] , Yan Liu [3] , Qipeng He [4], Chenyu Zhao [3], Haoyu Hu [3], Jinshuai Wang [3], Xiao Li [3], Ling Yu [3*]

{hexy@dlpu.edu.cn,1258570169@qq.com,hqpdyx@163.com}

University of Chinese Academy of Sciences,Beijing,100049,China [1], Shenyang Institute of Computing Technology Chinese academy of sciences, Shenyang, Liaoning Province,110168,China[2],Dalian Polytechnic University, Dalian, Liaoning Province,116034,China[3],Guangchuang service center of Guizhou Zhifu Optical Valley Investment Management Co., Ltd.[4]

Abstract. Under different installation methods, the illumination uniformity of strawberry canopy was tested, and the light environment was simulated by dialux evo. Through the simulation test of different arrangement, installation height and luminous angle, the data of three different working faces are observed. Through analysis and calculation, a kind of light environment design which can improve the illumination uniformity of strawberry canopy in greenhouse is obtained.

Keywords: illumination uniformity; light environment design; plant lighting; plant factory;

1 Introduction

With the development of lighting technology, people are no longer satisfied with the growth of plants under the sun, but supplement the light of plants. In some areas with bad climate, The efficient plant lighting technology can shorten plant growth cycle and increase crop quality. A LED lamp containing a spectrum suitable for plant growth acts on the growing plant to make up for the lack of light caused by environmental reasons, which can not only promote its growth, but also improve the quality of the plant.

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307391

Because the leaf distribution of plants is not uniform, under certain light conditions, the top leaf reaches saturation, but the bottom leaf does not reach light saturation at this time, the illumination will be uneven, which will affect the growth of plants. As a herbaceous plant, the plant height range in the growth cycle is generally 10 cm-40cm, and the number of leaves with different heights varies greatly. Under the condition of natural light, it is easy to produce the above uneven illumination, which is suitable for artificial light supplement experiment[1]. Therefore,strawberry was selected as the experimental object. LED light source has the advantages of long service life, using DC, low heating and so on. At the same time, the LED light source system can also regulate the light quality, and can realize close irradiation with crops, which is suitable for this simulation experiment. In the process of experiment, the illumination and evenness of strawberry in vertical direction were studied by adjusting the installation position and luminous angle of lamps and lanterns. Dialux evo is used to simulate the light environment, three working faces with different heights are set up, and the illumination uniformity of the three working faces is calculated. By reducing the difference of illumination in the vertical direction of plant leaves, a suitable light environment for strawberry growth is obtained.

2 Design principles

In order to solve the problem that the canopy leaves are affected by uneven light, the illumination difference of strawberry plants in vertical direction is reduced by adjusting the lamp arrangement, installation height and luminous angle, so that the crops can receive light evenly in vertical direction. As a whole, the light distribution mode mainly starts from four aspects :1.illumination 2.illumination uniformity 3.installation capacity reduced to minimum 4.distance to high ratio (s/h)[2]. With different installation height, the lower the installation safety, the greater the illumination uniformity of the working face, so the appropriate intermediate value is found by testing.

The luminous angle of the light source is increased from 20° to 50° in 2° . Since the vertical height of mature strawberry plants is not more than 30 cm, three height working faces (1.3 m、1.5m、1.7m) are set up in the simulation process to compare the data, and the suitable luminous angle is selected, and the optimal light environment design scheme is obtained.

3 Location Design of Light Source

3.1 Lamp installation

The height of plant lighting installation should be determined by the species of vegetables cultivated in solar greenhouse and their botanical and biological characteristics. According to environmental factors, the installation height of the light-filling lamp in greenhouse is 1.9 m-3.3m[1]. By setting this height as the simulation height, it can be seen from the parameters that the lower or higher the installation height is, and the uniformity decreases obviously.`

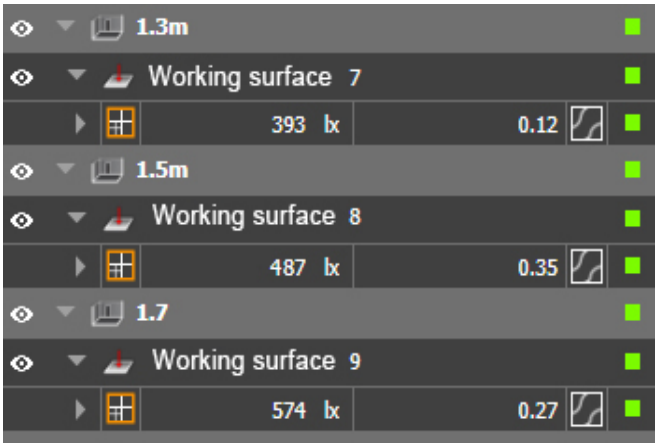


Fig .1. Illumination and illuminance uniformity of the lamp at 2 meters

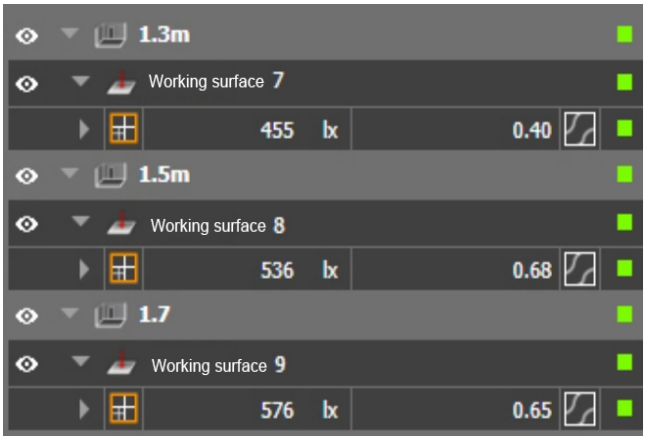


Figure .2. Illumination and illuminance uniformity of the lamp at 2.3 meters

Whether the arrangement of lamps and lanterns is reasonable or not depends on the high ratio (s/h). When the distance between s(lamps and lanterns) has been determined h(the higher the height of the lamp distance from the working face), the better the illumination uniformity and the poor economy. Therefore, in the early stage of the test, it is found that when the ratio of distance to height is smaller, the difference of parameters between the three working faces is smaller. Therefore, in the similar experimental environment, summarized the existing plant lighting test installation. The lowest installation height and the lowest installation height are obtained, and the simulation test is carried out on the installation of 1.9 m-3.3m height lamps. The results are as follows: when the ratio of distance to height is 0.48, the uniformity of the three working faces is the smallest and the highest. The installation height is 2.5 m, so the installation height 2.5m.

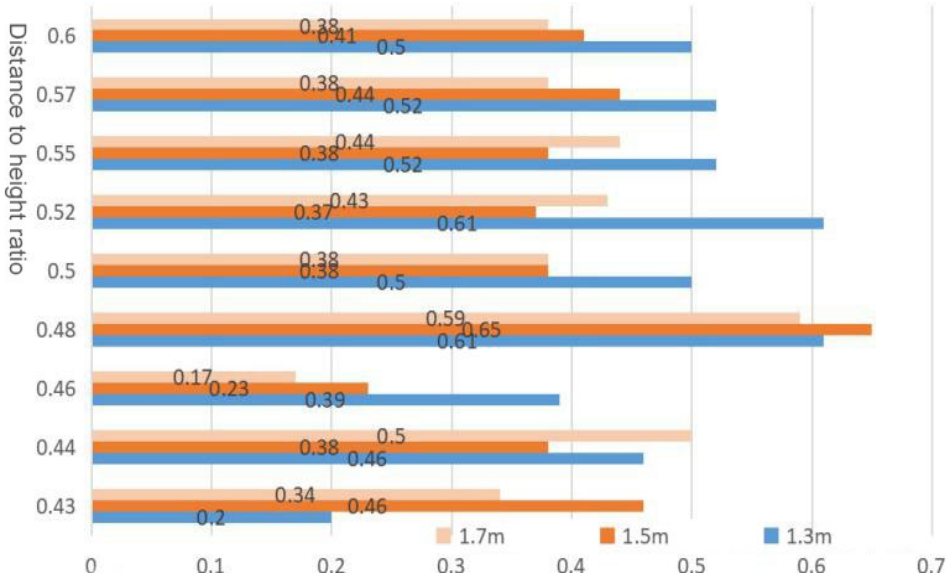


Figure.3. Effect of different distance and height ratios on uniformity

3.2 Lighting arrangement

A total of two ways of lighting layout to carry out simulation tests, mainly meet the requirements :1.working face illumination is not less than 500 lx, select the three working face on the minimum difference between the test data 2. the three working face illumination uniformity difference is the minimum 3. the appropriate direction of light 4.light source installation capacity reduced to the minimum 5.layout neat and beautiful[1].

Two kinds of arrangement schemes are used to compare, uniform arrangement and cross arrangement, as shown in the following figure, the number of lamps used is 48 by using the spatial coefficient method. The advantages of this scheme are neat and beautiful layout, high utilization rate of lamps and small capacity. However, the disadvantage is that the light source irradiation on the vertical surface is not uniform, the illumination uniformity and illumination uniformity of the three working faces are quite different, and the plants will appear the phenomenon of unsaturated at the top of the canopy. The biggest advantage of cross arrangement is that it can solve the problem of uniform arrangement, which can significantly improve the situation of large difference in working face[4]. The disadvantage of this scheme is that the installation capacity of lamps is large and the installation is not convenient.

The dialux evo simulation lamp is uniform, the arrangement mode is 6×8, the rotation angle of the lamp is 0°, and the installation position is directly above the strawberry plant (2.5 m from the ground). According to the simulation results, the illumination of the three working faces is lx、475lx、537lx.:306 The illumination uniformity is 0.5,0.47,0.24, respectively. The higher the illumination is, the greater the ratio of distance to height is, so the low illumination uniformity can not guarantee the significantly reduced illumination difference in the three working faces. The use of cross-cloth lamp increases the number of lamps, improves the illumination and illumination uniformity of the working face, so in the same case, the luminous angle of the lamp is 0 degrees, and the difference between the three working faces is obviously improved. So choose cross-cloth lamp.

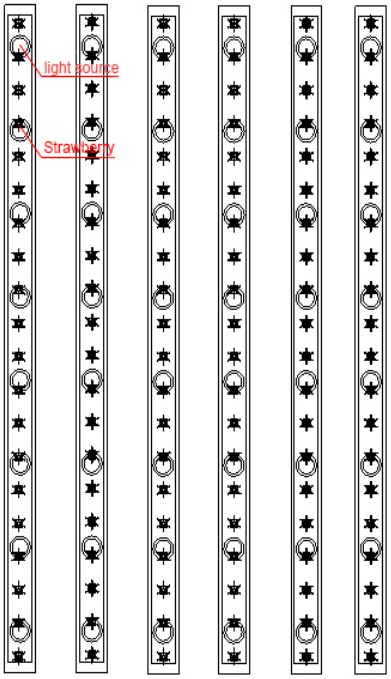


Figure.4. Uniform layout of lamps

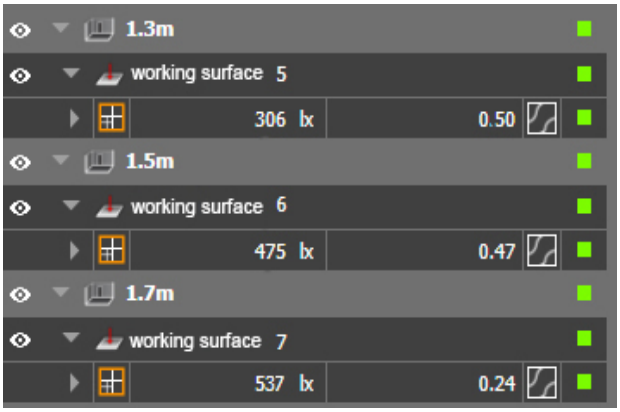


Figure.5. Illumination and illumination uniformity of the working surface with uniform lighting

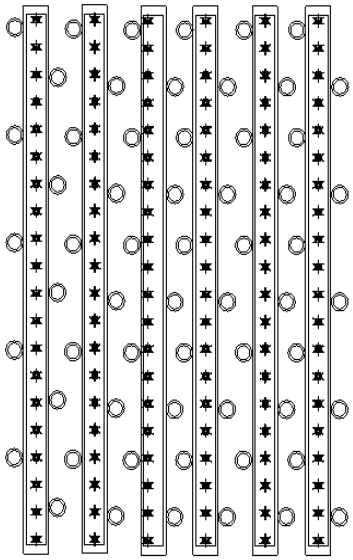


Figure.6. Cross-cutting arrangements

Fig.7. Illumination and illuminance uniformity of working surface with cross-lights

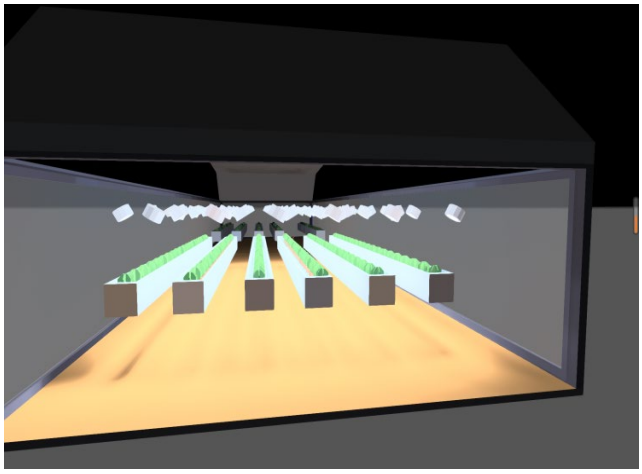


Figure.8. Lamp installation diagram

3.3 Luminescence angle

It is found that the shading problem between different layers of fixed cultivation frame is improved to some extent.The direction of light irradiation is parallel to the direction of cultivation frame and the direct light is used to the maximum extent. The main problem of this method is that the upper layer of the same planting frame shading the lower layer, which will lead to different layers of strawberry lighting environment inconsistent, affecting its yield and quality so do not change the way of plant cultivation[2]. Whether changing the light source installation angle is also a solution. Therefore, we carry out simulation experiments, because in vertical cultivation.Experimental Environment Reference [5], the tilt angle range of the plant is 20-50°, which can guarantee the reasonable installation capacity of lamps and lanterns, so the same range is selected at the luminous angle of the light source. A total of 16 groups of experimental data 20°-50° were increased in 2°[3] At the same time, in order to increase the uniformity of the working face, avoid the adjustment of the left and right sides of the plant due to the luminous angle.

4 Simulation results and analysis

Using dialux evo simulation, the whole length of greenhouse space is 28 m、wide 8 m, the whole is divided into left and right parts, and symmetrical, only one side is tested in the test. Strawberries evenly placed in six rows at intervals of 0.6 m. The lamp installation height is 2.5 musing the cross cloth lamp way, carries on the simulation test from 20° to 50°, the test result arranges the data as follows:

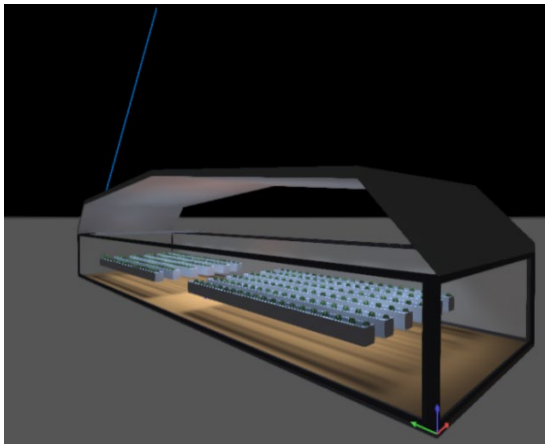


Fig .9. Simulation results of greenhouse

Table .1. Illumination and uniformity of lamps and lanterns in three working faces at different angles

		20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	50
Illumination lx	1.3 m	599	590	585	579	566	**564**	555	554	502	491	474	430	430	401	359	368
	1.5 m	682	682	669	659	643	**646**	606	588	520	520	477	476	434	429	392	371
	1.7 m	748	741	727	704	691	**686**	637	612	598	597	569	566	519	474	427	399
Uniformity	1.3 m	0.69	0.67	0.60	0.55	0.55	**0.65**	0.65	0.65	0.64	0.57	0.59	0.61	0.61	0.50	0.52	0.52
	1.5 m	0.46	0.38	0.54	0.54	0.61	**0.67**	0.60	0.59	0.59	0.51	0.43	0.37	0.41	0.38	0.39	0.39
	1.7 m	0.24	0.37	0.37	0.49	0.48	**0.66**	0.55	0.59	0.46	0.37	0.22	0.43	0.44	0.38	0.28	0.20

When the luminous angle is less than 30°, the difference between the three working faces

decreases linearly with the increase of luminous angle, but the change is significant but the luminous angle is more than 30°. The gap between the three working faces also increases, can be said that 30°is a critical value. Therefore, the most suitable luminous angle at the same installation height should be 30°. At an angle of 30°, the illuminance of its three working faces is :564、646、686. In terms of illuminance and uniformity, the illuminance of the three working faces is the smallest in 16 groups of experiments, and the maximum difference of illuminance uniformity is only 0.02. To sum up, this scheme is the optimal scheme at present.

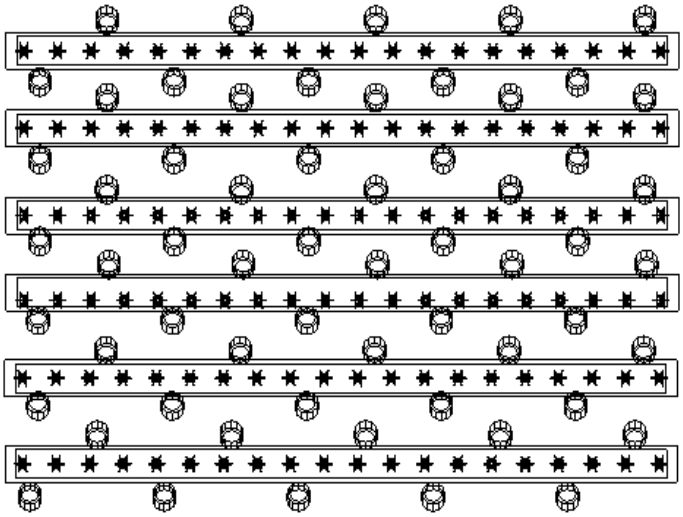


Figure.10. Cross arrangement of lamps and lanterns

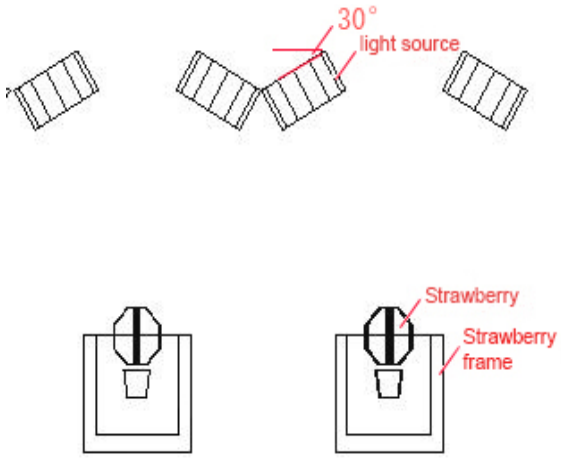


Figure .11. Left view of lighting method

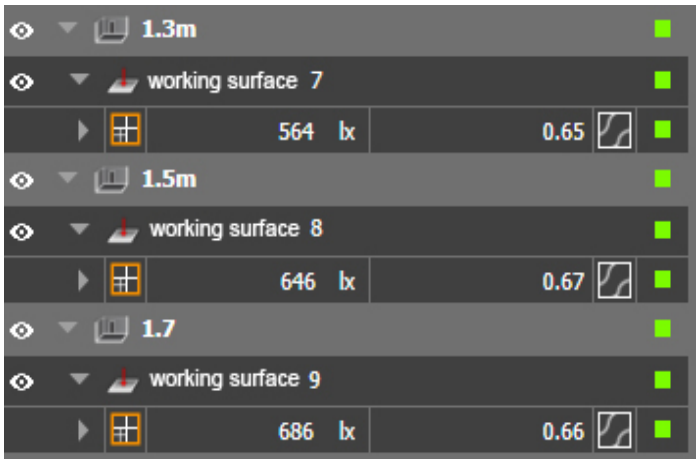


Figure.12. Simulation results at 30° angle

5 bundles

When the installation height and arrangement mode are determined, when the luminous angle is less than 30°, the difference between the three working faces decreases linearly with the

increase of luminous angle, but when the luminous angle is more than 30°, the gap between the three working faces increases with the increase of luminous angle 30°and the installation height is 2.5m. The cross-arrangement method used as the light environment design scheme of strawberry shed, and the optimized simulation parameters are obtained. The illumination of three high working faces is 564 lx、646lx、686lx. respectively The uniformity of illumination was 0.65、0.67 and 0.66 respectively.

A systematic study has been carried out in this experiment on LED light supplement, and the most suitable installation method for strawberry plant light supplement has been obtained, which provides a theoretical basis for the production and application of sunlight greenhouse light supplement in practical strawberry planting. Generally speaking, the LED and light supplement in the solar greenhouse not only ensures the yield of production, but also improves the quality of strawberry fruit and is suitable for production. Now in agricultural production, the solar greenhouse in our country has been greatly popularized and applied, and the LED light-filling technology will become more and more important[6]. LED light-filling can accurately regulate the growth and development of plants and their fruit, and can actively regulate plant growth if properly applied. In agricultural production, we can obtain high value and high return agricultural products according to different plant demand for light. At the same time, it also meets the requirements of high efficiency and high yield of scientific and technological agriculture in modern society, and has a very good prospect of application and popularization.

References

[1] Lingling Wei, Qichang Yang,Shuili Liu . Research Status and Application Prospect of LED in Plant Factories [J]. China Agronomy Bulletin ,2007,23(011):408-411.

[2] Chunling Wang,Weitang Song,Shumei Zhao,etc. Improving the Light Temperature Environment and Improving Strawberry Yield [J].with Japanese Strawberry Stereo Cultivation Frame Journal of Agricultural Engineering ,2017,33(011):187-194.

[3] Weitang Song .Shumei Zhao . A three-dimensional strawberry cultivation frame for improving light temperature environment and increasing strawberry yield in Quming Mountain [Journal of J] Agricultural Engineering（2017)

[4] Zhengmin Di, Dongxiang Yu , Jianfeng Zhang, etc. [J]. on Vegetable Production and Management Technology for Haze Weather Facilities Chinese melon and vegetable ,2016,29(02):44-45.

[5] Jun Liu,Conghua Xie . The invention relates to an efficient method for producing potato trial tube potato and a culture box thereof [P],china:CN02131125.0,2004-04-14.

[6] Xueyi Zhu,Juan Tang,Muqing Liu, etc. A LED Remote Lighting System for Promoting the Growth of Dendrobium candidum[P],china:CN201520265016.5,2015-10-07.

Experiment teaching design of high-power smart LED driver based on Arduino

Yingming Gao[1], Weidong Kuang[2]

{ gyming@dlpu.edu.cn[1],781894587@qq.com[2]}

School of information science and engineering, Dalian Polytechnic University, Dalian 116034, China

Abstract: For the comprehensive experimental course of light source and lighting specialty, the teaching system of high-power intelligent LED driver is developed, which is composed of Arduino board and high-power LED extended circuit board. The expansion board includes an indicator unit, an input unit and a high-power LED drive unit. The process of determining the key components of the drive unit is given in detail. The expansion board can drive 1~2 1W high-power LED within the system working voltage range of 8~12V, with stable output current and 88% peak efficiency. The relation curve between voltage difference and efficiency is given. For different loads, the efficiency is different even in the coincidence interval of pressure difference, which indicates that efficiency is not only related to pressure difference, But also needs to consider the load the voltage difference caused by debugging to achieve high efficiency. Experimental design is beneficial to enhance students' perceptual knowledge of high-power ledsdriver and enhance their design ability.

Keywords: Arduino,HV9910B,smart LED driver,Experimental teaching system

1 The introduction

Arduino, as a popular open source hardware platform, has been paid attention to by many universities. Arduino is very popular in European and American universities as a basic technology for teaching and research[1]. Many colleges and universities in China have carried out teaching reform and curriculum construction based on Arduino, so as to achieve teaching goals and improve students' innovation ability.Based on Arduino platform, Dalian University of Technology designed the basic experimental course of The Internet of Things, which

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307232

enables students to quickly understand the hierarchical structure of the Internet of Things and cultivate their development and innovation abilities[2]. The teaching reform of Arduino-based electronics courses in USTC has stimulated students' interest in learning and promoted the cultivation of students' comprehensive quality[3]. Based on the existing teaching experience and practical teaching, Shanghai University proposed a new electronic practice teaching scheme based on Arduino open source hardware platform[4]. Shanghai Normal University offers Arduino as a professional elective course for first-year students, which cultivates students' professional interest and innovation ability[5]. By introducing Arduino into the teaching reform of electronic practical training and innovative practice, Shanghai Institute of Power inspires and improves the innovation ability of students, which is of reference significance for colleges and universities to carry out electronic practical training and innovative practical teaching reform[6].

Arduino's original goal was to give art students an easy to use microcontroller to help them design[7]. Its INTEGRATED development environment is simple and friendly. The initial state of the hardware is set by functions rather than by various registers. It is not necessary to understand the working principle of the single chip microcomputer.With objects defined by reusable class libraries, a few lines of code can achieve the required functionality, greatly improving programming efficiency.Arduino's ease of use makes it suitable for cultivating students' innovation ability. Arduino can be used to quickly build prototype systems like "building blocks", so as to avoid spending too much time and energy on basic functions and focus on innovative work of method verification or functional experience.This experimental teaching system combines Arduino with LED drive technology to quickly build an intelligent lighting prototype system with sensing and network functions.Through this system, students' innovation ability of intelligent lighting is cultivated, and their verification and experience of intelligent lighting effect and even research on basic methods of lighting control are carried out, laying a foundation for their future career in intelligent lighting industry.

2 Smart drive design

The chip-level intelligent LED driver has dimming function and can communicate with the microcontroller through SPI, I2C, single-bus and other serial communication interfaces[8]. This driver has no computing power and can only get the dimming information from the single chip microcomputer through serial interface. Since the Arduino itself provides a variety of serial interfaces, this intelligent driver can be simulated by combining the Arduino and LED

driver, as shown in Figure 1 in the form of "Arduino+ high-power LED Drive extension board".It has a stronger ability to expand, the serial interface if connected to the network communication module and sensor module, easy to be extended to the intelligent lighting system.

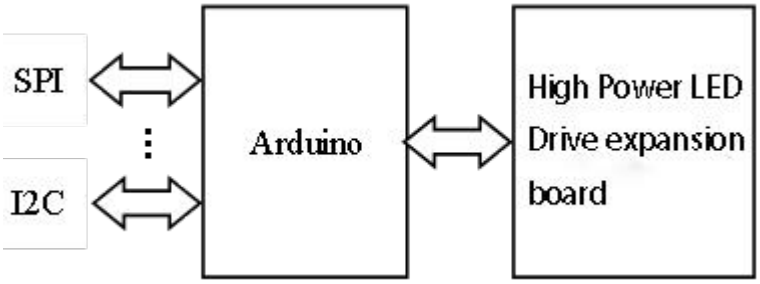


Fig. 1. Structure of intelligent LED driver

2.1 Expansion board design of high-power LED driver

We used Arduino UNO[9]As a processor, a high-power LED drive extension board is developed to realize the structure of the intelligent driver.The expansion board includes key input unit, LED indicator unit and high-power LED drive unit based on constant current driver HV9910B.The key is used to adjust the duty ratio of PWM dimmer signal.The LED indicator unit consists of 3 straw hats encapsulated LEDS, and displays the duty ratio of the current PWM signal with a combination of "light/dark" state.

Table 1. Pins of Arduino UNO for the extension board connection

Arduino pin types	The name of the pin	Directions for use
The power supply pins	VIN	Power supply 7~12V for Arduino
The power supply pins	GND	The reference to
Digital pins	Pin 3	PWM signal output for dimming
Digital pins	Pins 5, 6	Input pins are used to connect keys
Digital pins	Pins 8, 9, 10	Output pins are used for low power leds

The Arduino UNO pins used in the extension board are shown in Table 1. They insert the pins into the slot of the Arduino board, so that the extension board can be placed above the Arduino in the installation form shown in Figure 2.Arduino UNO provides three types of pins, namely, the power pin, the analog input pin, and the digital pin. Among them, the digital pin is multiplexed with the PWM output pin and the serial communication pin.Dc power supply is adopted in the system. The positive pole is connected to the Arduino VIN slot through an extension board, and the negative pole is connected to the Arduino GND slot through another inserting pin.The input voltage range of Arduino UNO is 7~12V, and the operating voltage range of the drive unit HV9910B is 8~450V[10][11]. Therefore, the input voltage range of smart drivers is 8~12V.Digital pins 8~10 are used to drive three low-power leds in the indicating unit. "bright" and "dark" respectively represent binary numbers 1 and 0. Their "light/dark" combination can indicate eight PWM signals with different duty ratio.

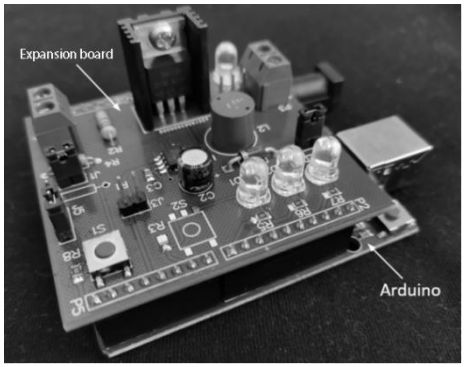


Figure. 2. The connection form between the extension board and Arduino

Fig. 3 is the high-power LED drive unit designed by BUCK.HV9910B can output two kinds of PWM switch signals with constant cycle and constant low level time to drive the step-down converter circuit.As a buck converter and operating in continuous current mode, when the output voltage U_{OUT}Greater than the input voltage U_{IN}When 50%, that is, when the duty ratio of the switching signal is greater than 0.5, it is advisable to use the switching signal with constant low level time to avoid sub-harmonic oscillation[12].The switch signal may be transmitted by oscillating resistance R_TSettings, R_TOne end is connected to pin 8, the other end is earthed and the other end is connected to pin 4, which is a switch signal with constant low-level time[7]. By introducing jumper wire into the circuit (P3 in FIG. 3), the system can switch between these two switch signals, ensuring the stability of driving multiple high-power leds.

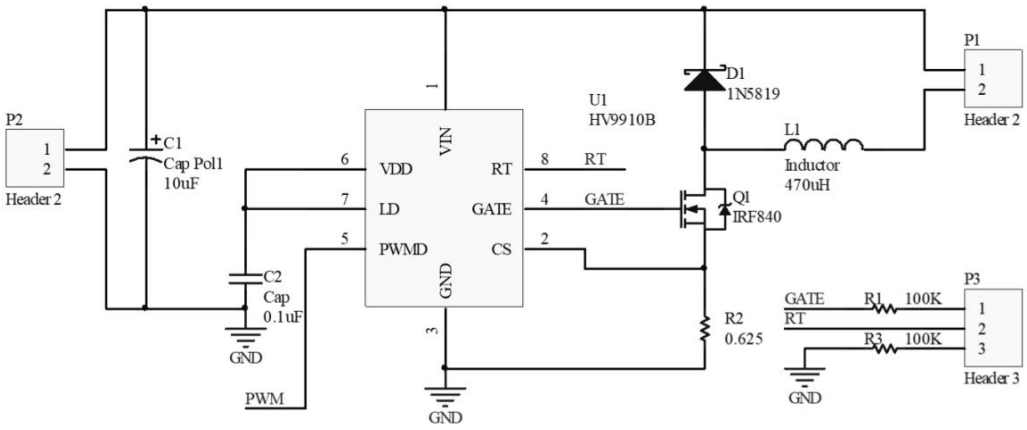


Fig. 3. High power LED drive circuit design

2.2 Determination of key components

The design indicator of the expansion board is: input voltage U_{IN}When the range is 8~12V, the output voltage U_{OUT}Range is 3.5~7V, output current I_O=350mA (i.e., drive 1~2 1W white leds), ripple current is 30%.Since the ratio between the maximum output voltage 7V and the minimum input voltage 8V is 0.9 and greater than 0.5, the low-level time-constant PWM signal is adopted.The components to be determined are current detector resistor R2, oscillation resistor R1 or R3 (determined by jumper), power inductor L1 in the switch converter circuit, field effect transistor Q1 and diode D1, as shown in Fig. 3.

$$R_{CS} = \frac{0.25V}{1.15 \cdot I_{LED}(A)} \quad (1)$$

$$R_T(K\Omega) = 25 \cdot t_{OSC}(\mu s) - 22 \quad (2)$$

Equation (1) calculates the galvanometer resistance.HV9910B belongs to the peak current controller, and the current detection resistance is determined by the peak current.As the ripple current is 30%, it is the peak current according to the relationship between average current and ripple current. $I_O = I_{LED}1.15 \cdot I_{LED}$Substitute 0.35a into equation (1), and you get RI_{LED}CSR2 = = 0.62 Ω.

Equation (2) is used to calculate the oscillation resistance.If THE PWM signal is periodic, then in

the formula is the period of the signal.t_{OSC}If PWM signal is low level time constant, then in the formula is low level time.t_{OSC}For low-voltage input, a low level time of 5 s, corresponding to R, is recommended when using a PWM signal with a constant low level time$_{T}\Omega$ = R1 = 100 k.

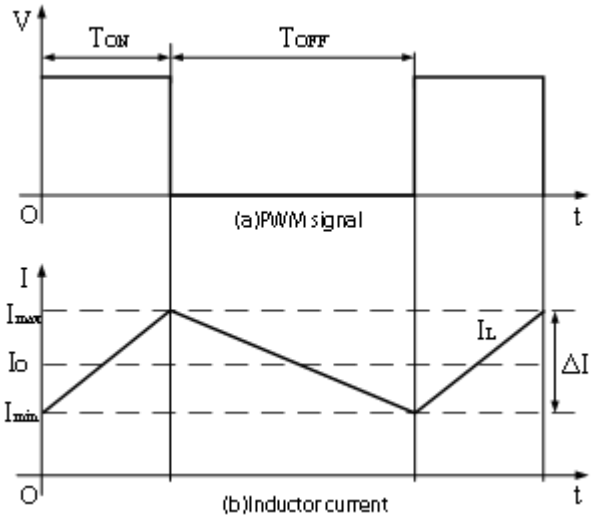


Fig. 5. Relationship between step-down circuit PWM signal and inductive current

$$L=\frac{U_L \cdot \Delta t}{\Delta I} \tag{3}$$

Formula (3) calculates the power inductance, where is the inductance voltage, is the increment of the inductance current, and is the corresponding time increment.$U_L \Delta I \Delta t \Delta I$In the PWM signal T_{OFF}It is more convenient to calculate inductance when period Q is cut off. At this time, field effect tube Q is cut off and diode D is switched on. Fig.6 Circuit (a) is equivalent to circuit (c) and inductance voltage U_LIs equal to the output voltage U_{OUT}.It can be seen from Fig. 5 that the decrease of the inductance current during the period is the ripple current.$\Delta t = T_{OFF} \Delta I$To drive two 1W leds, set U_L=U_{OUT}If A is substituted into equation (3), the power inductance L=333 H can be considered as the minimum theoretical inductance that meets the requirements, and the ripple current can be reduced by using A larger inductance than it.$\Delta t = T_{OFF} = 0.5\mu s \Delta I = 0.35 \times 0.3 = 0.1$

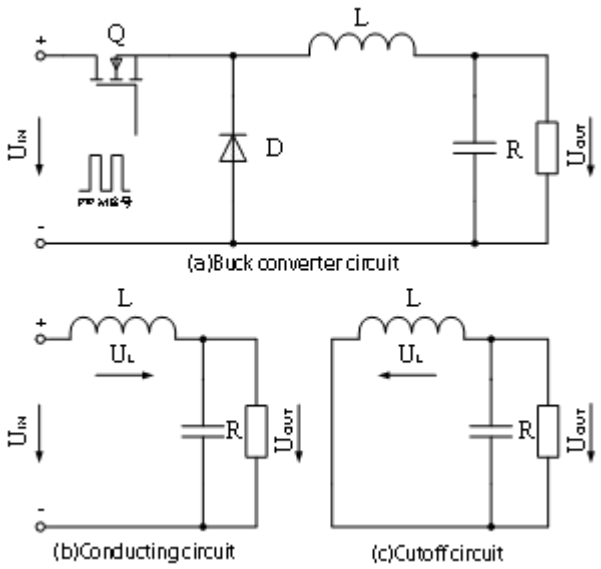


Fig. 6. Topological structure and state transition of step-down circuit

The system uses field effect tube IRF840 and Schottky diode 1N5819 as switching elements.In addition to the voltage, current bearing capacity to consider, their switching speed and PWM signal frequency match, otherwise the circuit can not work normally. As shown in Fig. 6, PWM signal makes field effect tube Q in step-down circuit (A) switch between on and off, and diode D switch between off and on, thus making circuit (a) cycle switch between circuit (b) and circuit (c) to realize voltage transformation function.Ideally, the switch is instantaneous and fully synchronized with THE PWM signal, as shown in Fig.7 (b) Curve 1.However, under the influence of junction capacitance, the voltage at both ends of PN junction cannot mutate, and the switching state conversion of devices always lags behind the change of PWM signal[13]. Only when the device's switching time T_{on}And t_{off}The state conversion can be completed only when the holding time of PWM signal is less than that of PWM signal, as shown in Fig.7 (b) Curve 2.If the PWM signal frequency is too fast, its holding time is shorter than the switching time of the device, as shown in Fig. 7(b) curve 3, resulting in the device never being able to transfer state and the step-down switch circuit is unable to perform voltage transformation function.

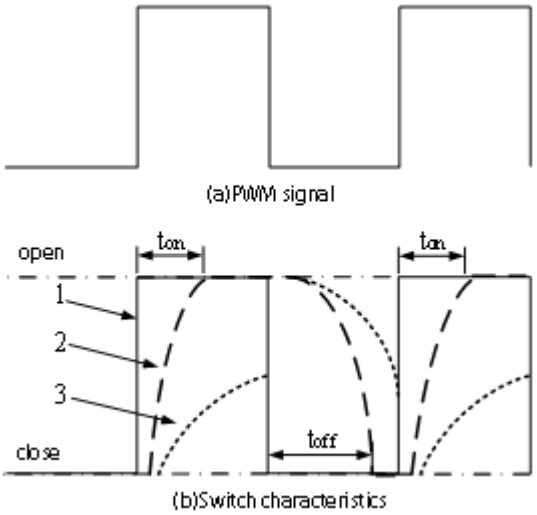


Fig. 7. Relationship between PWM signal and device switching time

3 System perfor test

3.1 Index Test

Efficiency is equal to the ratio of output power to input power. For dc power supply, power is equal to the product of voltage and current. The input voltage and current are measured respectively. After the output voltage and current, the input power and output power are calculated to obtain the efficiency.

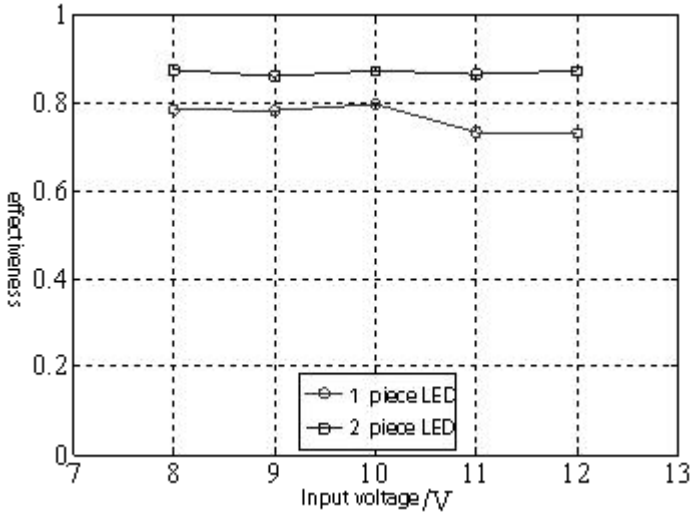


Fig. 8. Relation curve between input voltage and efficiency

As shown in Fig. 8: The input voltage of the expansion plate is within the range of 8~12V, and the maximum efficiency of driving 1 1W high-power LED and 2 1W high-power LED is measured as 80% and 88% respectively.The efficiency of 2 1W high-power leds driven by expansion board is higher than that of 1W high-power leds driven by expansion board.

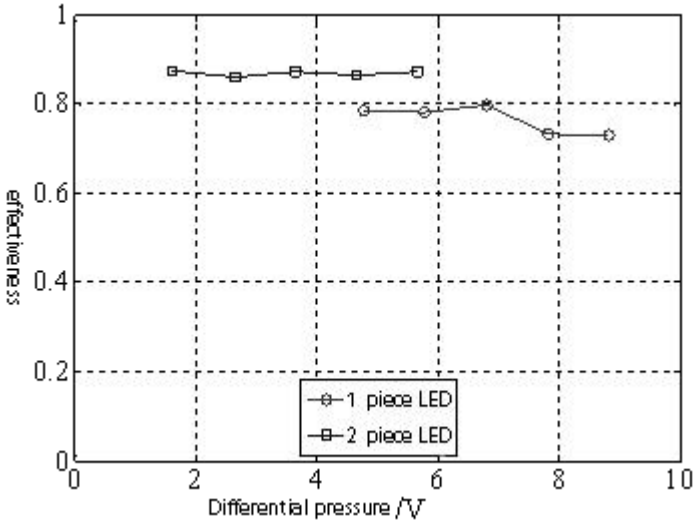


Fig. 9. Relation curve between pressure difference and efficiency

It can be seen from Figure 9 that when the input voltage is 8-12V, the efficiency is relatively stable when driving 2 leds with the increase of the difference between input voltage and output voltage, while when driving 1 LED, the efficiency tends to decline with the increase of the pressure difference.But the pressure difference and efficiency are not one to one correspondence relation, for the above two kinds of loads, in the overlap area of the pressure difference, the efficiency has the more obvious difference. It shows that the efficiency is related to the pressure difference and the load, and the load and the pressure difference caused by the load should be considered comprehensively to obtain the high efficiency.

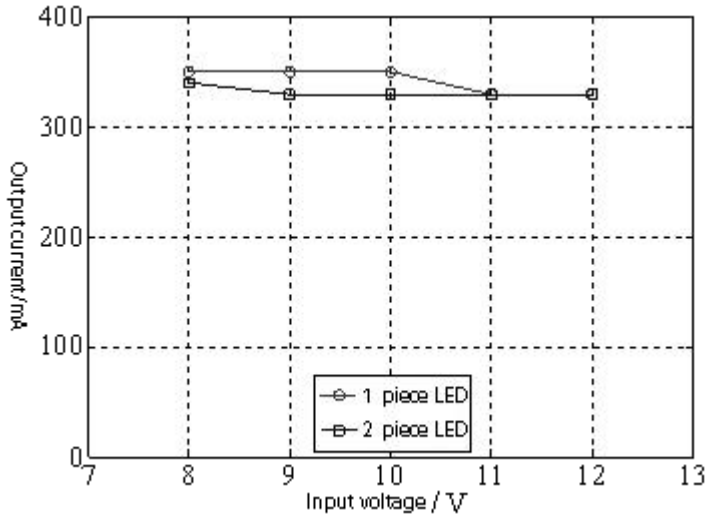


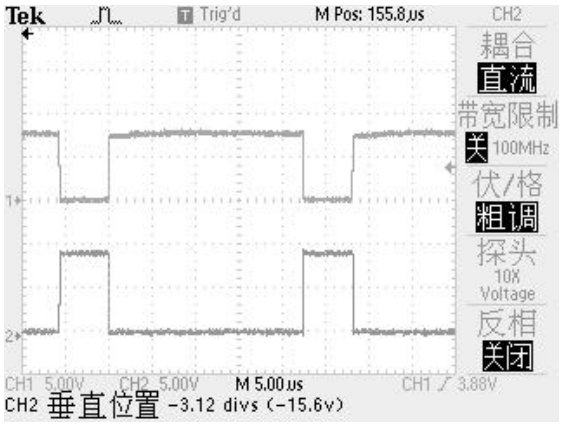
Fig. 10. Relationship between input voltage and output current

Fig. 10 shows the corresponding relationship between the output current and the input voltage when the input voltage of the expansion plate is 8~12V and 1 LED and 2 LED are respectively driven.As the output voltage changes, the output current is relatively stable, with a maximum of 345mA and a minimum of 327mA, slightly smaller than the target output current of 350mA.

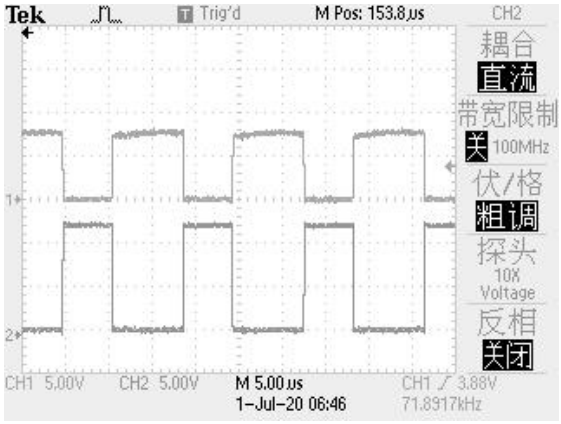
3.2 Signal testing

Fig. 11 shows the gate voltage waveform of the field effect tube (the switching signal output by HV9910B) and the corresponding diode positive voltage waveform at the input voltage of 9V and 12V. Channel 1 is the PWM signal waveform and channel 2 is the diode positive voltage waveform.According

to the measurement results, for low level PWM signal with constant time, its period decreases with the increase of input voltage under the condition of certain load. This is consistent with the effect of the output voltage/input voltage on duty cycle of the step-down switch converter circuit. The field-effect tube grid voltage waveform is complementary to the diode positive voltage waveform. The maximum voltage of the diode positive electrode varies with the input voltage and is the same as the input voltage. The output switching signal of HV9910B does not change with the input voltage.



(a) Waveform of input voltage 9V

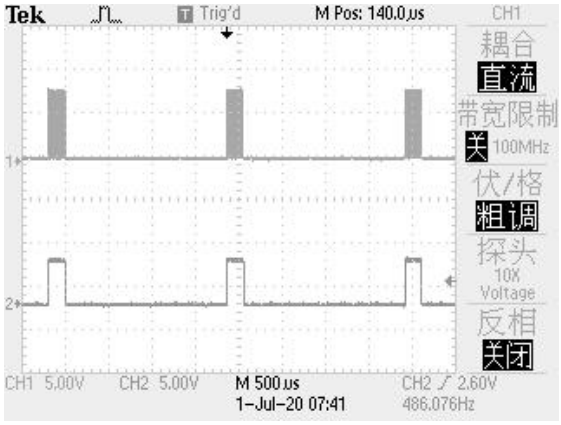


(b) Waveform with input voltage of 12V

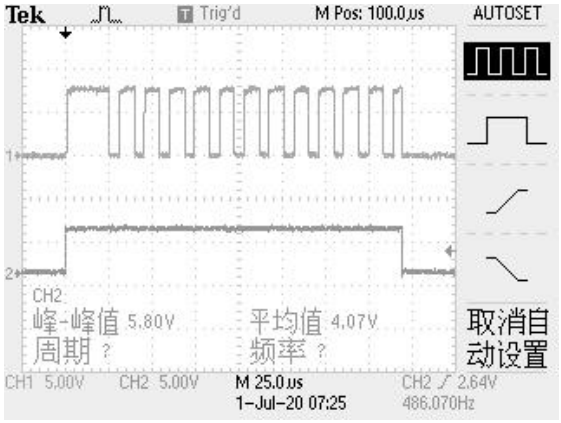
Fig. 11. Voltage waveform of switching signal and diode positive pole

Fig. 12 shows the corresponding relationship between the switching signal output from HV9910B and the dimming signal input from the pin of PWMD (duty ratio is 10%). Channel 1 is the switching signal and channel 2 is the dimming signal.In dimming, the switching signal output from HV9910B is

not output until the dimming signal is at high voltage, but not at low voltage. The actual switching signal output from HV9910B is the result of logic "and" operation between the dimming signal and the continuous switching signal.



(a) The overall characteristics of the two signals



(b) Local characteristics of two signals

Fig. 12 .Waveform of switching signal and dimming signal

4 Experimental teaching content design

This experimental teaching system can meet the needs of comprehensive experiment of light source and lighting specialty. The Arduino application can complete PWM signal output and serial communication functions by calling relevant functions. Compared with the traditional register operation mode, it saves a lot of repetitive basic code writing, and quickly builds intelligent lighting system to

verify and study the lighting control strategy.The contents that can be offered in experimental teaching include:

(1) High-power LED dimming experiment.The program changes the duty ratio of dimming signal output by Arduino digital pin 3 by pressing buttons to achieve dimming of high-power LED.The duty ratio of dimmer signal starts from 0. As the key is pressed down each time, the cycle increases the duty ratio to 100% in accordance with Table 2 and then returns to 0. At the same time, the "light/dark" combination of three low-power leds on the expansion board is changed accordingly.Measure the working current of high-power LED under different duty ratio, draw the working current curve of duty ratio and working current, and intuitively understand the relationship between LED working current and dimming duty ratio.

Table 2. Is the corresponding relationship between duty ratio and LED on-off combination

LED2	LED1	LED0	The serial number	Duty ratio
0 / dark	0 / dark	0 / dark	0	0
0 / dark	0 / dark	1 / bright	1	5%
0 / dark	1 / bright	0 / dark	2	10%
0 / dark	1 / bright	1 / bright	3	20%
1 / bright	0 / dark	0 / dark	4	40%
1 / bright	0 / dark	1 / bright	5	60%
1 / bright	1 / bright	0 / dark	6	80%
1 / bright	1 / bright	1 / bright	7	100%

(2) When not dimming, use oscilloscope to observe switch signals output by GATE pins of HV9910, record signal cycle, holding time of high level and low level, and compare with theoretical values.During dimming, use oscilloscope to observe the switch signal output by GATE pin, and record the change of output signal with dimming signal. Explain the logic operation relation between switching signal in dimming, switching signal in non-dimming and dimming signal.

(3) Illumination control experiment. The GY-30 illumination sensor module of I2C interface and the intelligent LED driver constitute the intelligent lighting control system. With the ON/OFF control strategy, the program realizes the illumination control of the working surface, that is, when the illumination of the day is lower than the threshold, the light is turned ON for illumination, and when the daylight is higher than the threshold, the light is turned OFF for energy saving.

5 Conclusion

The experimental teaching system of high-power intelligent LED drive developed in this paper can be used in the teaching of comprehensive experimental courses of light source and lighting specialty, to help students deeply understand the working principle of LED drive, enhance perceptual knowledge, and lay a foundation for future related work. The experimental design combines Arduino technology with lighting control technology, which reduces the difficulty of programming and enables students to focus more on experimental content related to system performance test and illumination control.

References

[1] Shi Zhixiong. Application exploration of Arduino in Engineering Teaching in University [J].Journal of Xichang University, 2012, 26(4):50-54.

[2] Liu Wenjie, Zhu Ming, Tan Zhenquan. Arduino-based Temperature Control Experiment Design for the Internet of Things [J].Experimental Technology and Management, 2017, 34(1), 150-152.

[3] Wang Fan, Zhou Wenhui, Liu Baojun, et al. Teaching Reform of Electronic Design Practice Based on Arduino [J].Modern Computers, 2018, (35), 81-85.

[4] Yuan Li. Preliminary Exploration of Electronic Internship Education Reform based on open source hardware [J]. Education & Teaching Forum, 2013, (22), 39-42.

[5] Zhang Chongming, Ye Hong, Ni Jifeng, et al. Teaching Practice of Arduino Elective course in the first year of undergraduate [J].Experimental Technology and Management, 2018, 35(3), 169-171,194.

[6] Tang Min, Jin Yiming. Preliminary Exploration on the Introduction of open source hardware into Electronic Training and Innovation Practice [J].Guide to Knowledge, 2016, 43(2), 75-76.

[7] Cui Yiwen. How did gizmodo invented by several art teachers 10 years ago give birth to today's intelligent hardware industry?[EB/OL]. [2015-08-03]. http://www.qdaily.com/articles/12926.html?source=feed.

[8] Sha Zhanyou, Yu Guoqing. Typical Application of Intelligent LED Driver [J].Power Supply Technology Applications, 2011, 14(5): 56-60.

[9] Chen Luzhou. Arduino Programming Foundation (Version 2) [M].Beijing: Beijing University of Aeronautics and Astronautics Press, 2015.

Huang Jianping, SHEN Hanxin. Design and Research of LED Step-down Drive Circuit based on HV9910B [J].Modern Electronic Technology, 2014, (19): 139-142

[11] Microchip.HV9910B chip data [EB/OL]. [2014/03-09]. Http//www.supertex.com.

[12] Steve Winder. Power Supplies for LED Driving[M]. United States of America: Elsevier Inc., 2008.

[13] Yan Shi, WANG Hong. Foundation of Digital Electronic Technology (the sixth Edition) [M].Beijing: Higher Education Press, 2016.

Research and design of a dimming control system based on STM32

Xiaoyang He[1,*], Chenyu Zhao[1], Wanglong Zhao[2], Yingming Gao[1], Ling yu[1], Hongbin Sun[1], Yutao Fu[1], Anqi Wang[1], Qipeng He [3,*]

{hexy@dlpu.edu.cn,2054662703@qq.com, hqpdyx@163.com }

Dalian Polytechnic University, Dalian, Liaoning Province,116034,China[1]

Hainan University, Haikou, Hainan Province, 570228,China[2]

Guizhou Zhifu Optics Valley Investment Management Co., Ltd, Bijie, Guizhou Province, 551700,China[3]

Abstract. The dimming control system with high efficiency and energy saving, simple and convenient operation is of great significance to daily life and industrial production. This paper we uses STM32 as the control core, designs a dimming control system based on PWM dimming mode and Bluetooth to send control commands. The experimental results show that the system can realize the smooth adjustment of the brightness of the four LEDs remotely.

Key words: STM32;Light-emitting-diodes(LEDs);Pulse width modulation(PWM);

1 Introduction

With the rapid development of LED lighting technology, people's demand for professional lighting is also increasing[1]. Compared with traditional lighting sources with single function, high energy consumption, nonlinear dimming and other shortcomings, LED has the advantages of high photoelectric conversion efficiency, low energy consumption, long life, fast response, etc. And LED is easy to combine with loT technology, which can better meet people's lighting needs[2]. At present, the common LED dimming technologies include analog dimming, PWM dimming and SCR dimming[3].Combined with the rapid response characteristics of LED, this paper uses PWM dimming technology to design a LED dimming control system based on STM32, and uses Bluetooth communication to realize the remote brightness control of LED.

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307129

2 Principle of PWM dimming

Pulse width modulation dimming(PWM dimming) is a method to control analog circuit by digital output of microprocessor. A series of pulse waves with different period and relative amplitude are used to replace the required waveform[4]. The luminous intensity of the LED is proportional to the current flowing through it, so the brightness of the LED can be adjusted by controlling the current. Combined with the fast response characteristics of LED, PWM dimming is used to make the switch circuit switch LED quickly at a certain frequency, that is to change the average current flowing through the LED to realize the switch and brightness adjustment of LED. In order to avoid the influence of light flicker on the human eye, the switch frequency should be set high enough that the human eye can't recognize(generally greater than 100Hz)[5-6].

The basic principle of PWM dimming is shown in Figure 1. The pulse counter TIMx_CNT is set to count up, and the value of the auto reload register TIMx_ARR is ARR.As the clock source keeps counting, the value CNT of the pulse counter TIMx_CNT will be compared with the present value CCRx of the compare register TIMx_CCRx. When CNT>CCRx, output high electrical level(or low electrical level), otherwise output electrical low level(or high electrical level). Changing the value of CCRx can change the duty cycle of output PWM, and change the value of ARR can change the period of output PWM. The Timer period is ARR, and the duty cycle of the output PWM wave is $\frac{CCRx}{ARR+1}$.

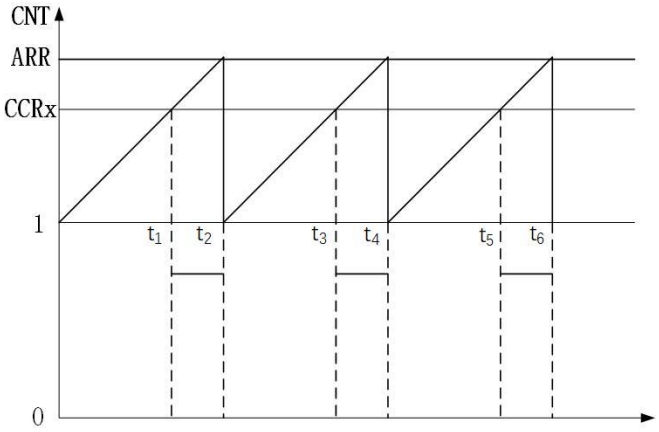


Fig.1. The operating principle of PWM

3 Overall design of dimming system

This paper selects STM32F103VET6 as the main control chip, which is mainly composed of Bluetooth module, LED drive module, power module and LED load, etc. The LED dimming control system is shown in Figure 2.Send control commands to the Bluetooth module through the mobile phone. The Bluetooth module communicates with STM32 microcontroller through serial port, control STM32 to output PWM waves with different duty ratio, and then outputs PWM waves to LED driver circuit to realize adjustment of LEDs' brightness remotely.

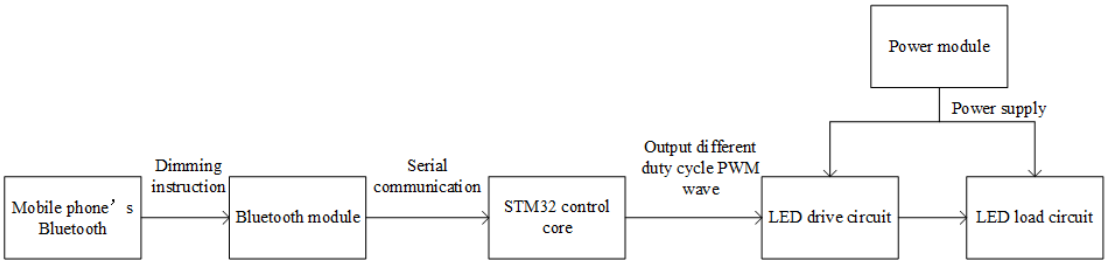


Fig.2. System overall design block diagram

4 The hardware design of dimming system

4.1 Main control module and peripheral circuit

In this paper, STM32F103VET6 single chip microcomputer is selected as control core. The peripheral circuit mainly includes reset circuit, crystal oscillator circuit, power supply circuit, etc. The reset circuit makes the whole system return to the initial state, the crystal oscillator circuit provides accurate clock signal for the dimming system, and the power supply circuit supplies power for the whole system.

The reset circuit is made up of 10K resistance and $0.1\mu F$ electric capacity, the circuit is shown in Figure 3. At the moment the system is powered on, the voltage across the capacitor cannot change suddenly. At this time, the input voltage to the NRST pin is low and the system is reset. Then the power-on voltage V_{cc} charge the capacitor through the resistor R, and the voltage across the capacitor continues to rise. At this time, the high voltage level is input to the NRST pin, and the system works normally.

After pressing the reset button RESET, the power-on voltage V_{cc}, the resistance R, and the reset button RESET form a complete loop. The capacitor is short-circuited. At this time, the capacitor begins to discharge, and the voltage across the capacitor continues to decrease. The

input voltage to the NRST pin is low and system reset.

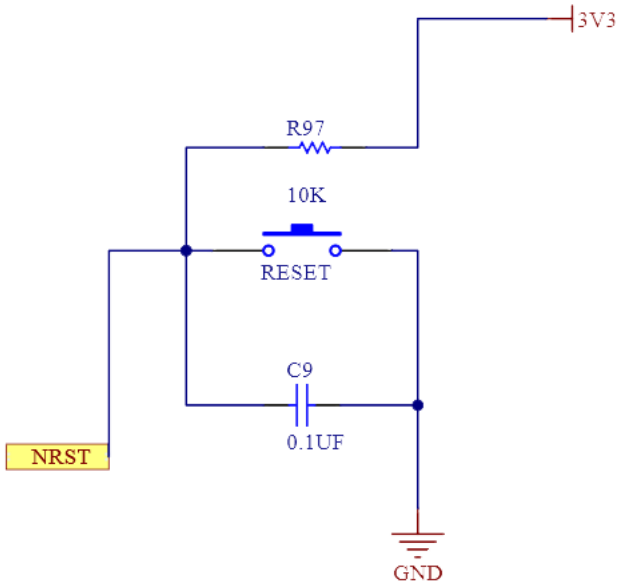


Fig.3. Reset circuit

STM32 mainly has two oscillation modes, namely HIS(internal high-speed clock) and HSE(external high-speed clock). In order to improve the clock accuracy, HSE is selected as the main clock source. Figure 4 shows the crystal oscillator circuit. The crystal oscillator selects an 8MHz HC-48SMD. The selection of C_1 and C_2 must be compatible with the circuit. The numerical relationship between C_1 and C_2 is shown in the following expression.

$$C_L - C_S = \frac{C_1 - C_2}{C_1 \cdot C_2} \tag{1}$$

C_L represents the load capacitance, C_S represents the parasitic capacitance between the two pins of the transistor[7], the input capacitance of the STM32 crystal oscillator is 5 pF, and the parasitic capacitance of the PCB trace is between 3~5 pF, so $C_L = 20pF$, $C_S = 9pF$. Normally, take C_1=C_2 and substitute the above formula to get C_1=$C_2 = 22pF$. To ensure normal start-up, the transconductance of the microcontroller g_m is required to be much large than transconductance of the crystal oscillator g_{mcrit}. Generally, it's required to be more than five times, and g_{mcrit} is calculated by the following expression.

$$g_{mcrit} = 4 \cdot ESR \cdot (2\pi F)^2 \cdot (C_0 + C_L)^2 \quad (2)$$

ESR is the equivalent series resistance of the crystal oscillator, F is the oscillation frequency of the crystal oscillator, C_0 is the parasitic capacitance between the two pins, and C_L is the load capacitance[8]. According to the actual choice of single-chip microcomputer and crystal oscillator, substitute relevant data into the above formula to calculated g_m=25mA/V, $g_{mcrit} = 0.5894\text{mA/V}$, and $\frac{g_m}{g_{mcrit}} \doteq 42 \gg 5$. Therefore, choosing $C_1 = C_2 = 22pF$ can meet the starting vibration requirements.

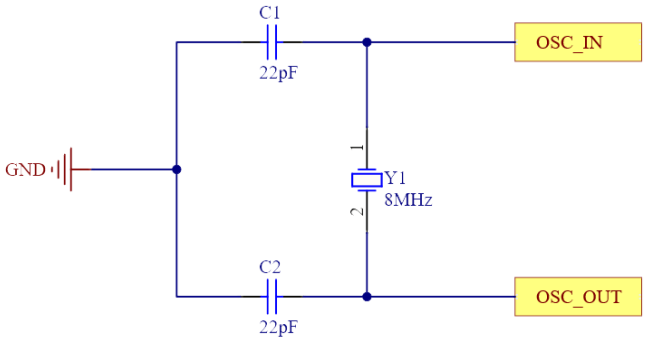


Fig.4. Crystal oscillator circuit

4.2 Bluetooth module and LED driver circuit

The Bluetooth module chooses ATK-HC05, this chip is mainstream Bluetooth chip which support Bluetooth V2.0 protocol standard CSR, the data transmission speed is within the range of 2M/s, and it's connected to the outside through 6 pin headers with a pitch of 2.54mm and the serial port module power supply voltage 3.3v, working current is about 40mA[9]. The module comes with a blue LED indicator. After it is successfully powered on, the LED flashes quickly to indicate that it is in the pairable state. After successful pairing, the LED double flashes. The HC-05 Bluetooth module can enter the AT state and communicate with the PC through the serial port. The initial properties of the Bluetooth can be set by inputting AT commands, including the device name, connection password, baud rate, master-slave mode, etc. The HC-05 Bluetooth module is connected to the microcontroller through four pins. RXD and TXD are respectively connected to the TXD and RXD pins on the microcontroller, namely the PA2 and PA3 pins. V_{CC} and GND are connected to the power and ground respectively. The connection circuit is shown

in Figure 5.

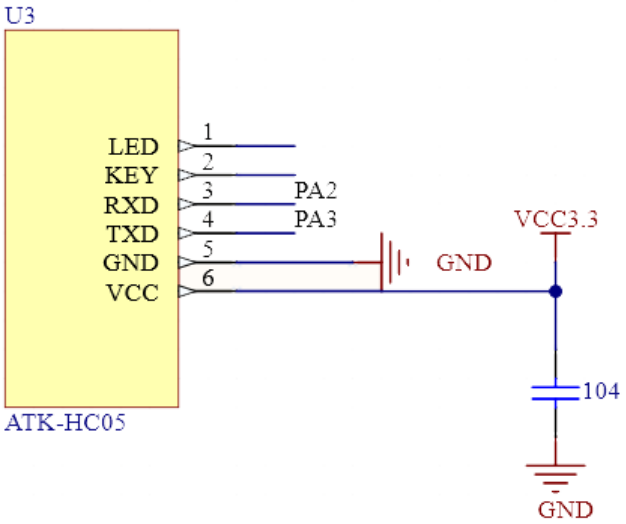


Fig.5. Bluetooth module connection circuit

The LED drive circuit uses the XL4001 module, which is a 150KHz fixed frequency step-down pulse width modulation DC/DC converter with a wide input voltage range of 4.5V~40V and a load driving capacity of 2A. It has good load capacity and only a small amount of external components is needed.

5 The software design of dimming system

This design program is written by Keil MDK programming software, and the computer language used for programming is C language. The whole process of program design is to initialize the system clock, GPIO port and serial port, etc. Then run to the main function, which is the concrete realization of the Bluetooth control function. First judge whether the Bluetooth data transmission is completed according to the status receiving flag n, then judge the transmitted data, increase or decrease the PWM duty cycle of the corresponding output channel according to the preset setting and return the value at the same time. This program realizes the regulation of the duty cycle of the four-way output PWM, and the main program flowchart is shown in Figure 6.

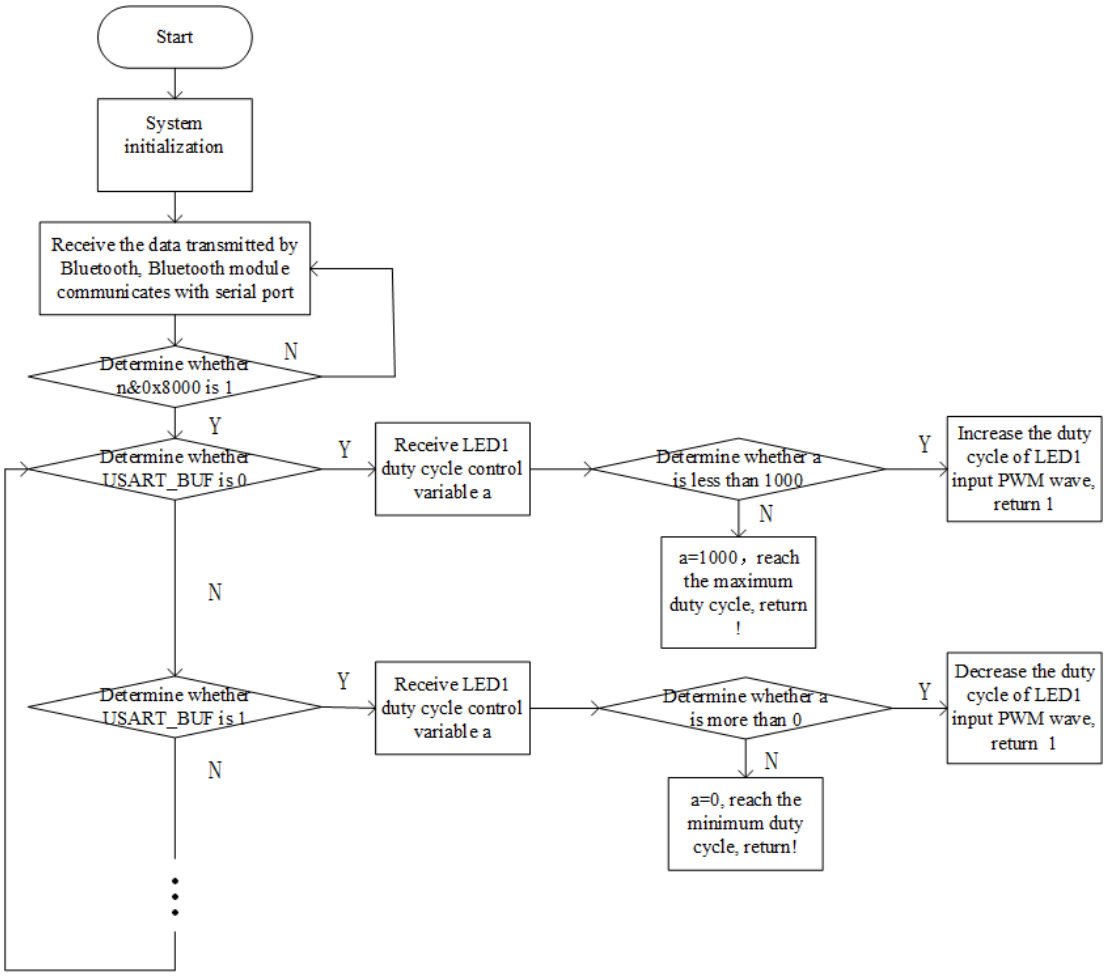


Fig.6. Main program flow chart

Choose Timer 3 and Timer 4 to output PWM wave, and set the frequency to 1KHz. The output ports are respectively set to PA6, PA7, PB6, PB7, respectively connected to the PWM input of the LED driver module, the initial PWM duty cycle of each output port is set to 0, and the frequency is 2KHz. Enable serial port 2 and complete the connection between the Bluetooth module and the microcontroller. PA2 and PA3 are the transmit-receive pins connected to the Bluetooth module, the baud rate of serial port 2 is set to 9600bps, and the interrupt processing function of the serial port is used to store the data transmitted by Bluetooth, Configuration functions are as follow:

TIM_Config();//Output PWM wave timer configuration, output frequency is set to 2KHz

PWM_Config();//Output PWM wave configuration, set its output mode, output comparison polarity, etc.

NVIC_PWM2_Config();//Timer 4 interrupt configuration function

LED_Config();//Configure four PWM wave output pins, set their output mode, pin rate, etc.

NVIC_Config();//Serial port 2 interrupt configuration function

USART_Config();//Serial port 2 configuration function, set its transceiver pin, transceiver mode, baud rate, etc.

void USART2_IRQHandler(void)// Serial port 2 interrupt service function, store data transmitted by Bluetooth

void TIM3_IRQHandler(void)// Timer 3 interrupt service function

void TIM3_IRQHandler(void)// Timer 4 interrupt service function

The main program judges the data transmitted by Bluetooth through the while(1) loop, and uses the status receiving flag n to judge whether the data transmission is completed. For the output PWM wave of channel 1, send a value of 0, after the Bluetooth module successfully receives the data through the serial port, the data is stored in the array USART_BUF[], the output PWM wave duty cycle increases by 1%, and the value 1 will be return. When increased to100%, continue to send the value 0, its duty cycle remains unchanged and returns !, at this time, the duty cycle has been adjusted to the maximum. Control program are as follow:

```
if(USART_BUF[0]== '0')
{
  if(a<1000)
  {
    a = a+10;
    TIM3->CCR1=a;
  USART_SendData(USART2,0x31);// Successful execution of the operation, return value 1
}
```

```
        if(a>=1000)
        {
          a=1000;
          TIM3->CCR1=a;
        USART_SendData(USART2,0x21);// Increase the duty cycle to the maximum, return !
        }
    }
```

Send a value of 1, after the Bluetooth module successfully receives the data through the serial port, the data is stored in the array USART_BUF[], the output PWM wave duty cycle will decrease by 1% and the value 1 will be return. When decreased to 0%, continue to send the value 1, its duty cycle remains unchanged and return !, at this time, the duty cycle has been adjusted to the minimum. Control program are as follow:

```
        if(USART_BUF[0]== '1')
        {
          if(a>0)
          {
           a=a-10;
           TIM3->CCR1=a;
           USART_SendData(USART2,0x31);// Successful execution of the operation, return value 1
          }
          if(a<=0)
          {
           a=0;
           TIM3->CCR1=a;
           USART_SendData(USART2,0x21);//Decrease the duty cycle to the minimum, return !
```

```
    }
}
```

The other three PWM output procedures are similar. Writing the program in this way is easy to control the duty cycle of each output PWM wave, and we can change the step length of each increase or decrease of the duty cycle to achieve more convenient and precise brightness control of each LED, and the return value can be very intuitive shows whether the data transmitted by Bluetooth is successfully received or whether the duty cycle has been adjusted to the upper and lower limits.

6 Experiment and result analysis

According to the above software and hardware design, the dimming control system is completed as shown in Figure 7.

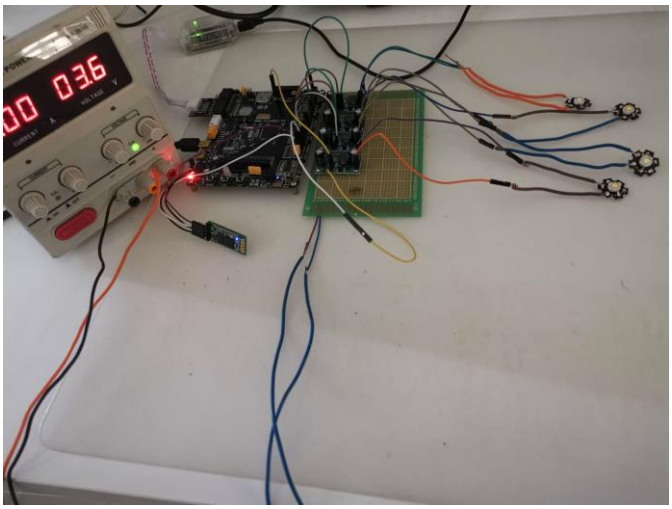


Fig.7. Dimming control system

Perform dimming experiments with different duty cycles on the dimming control system. In order to ensure the stability of the system, the power supply outputs 3.6V and the LED brightness can be kept stable. The mobile phone is connected to the Bluetooth module, and the blue indicator light of the Bluetooth module flashes double after successful pairing. Through the Bluetooth serial port assistant of the mobile phone, set the duty cycle of the four output PWM waves to 10%, 40%, 70% and 100% respectively, and observe with an oscilloscope. The

output waveforms are shown in Figure 8 and Figure 9. The dimming effect of each LED at different duty cycles is shown in Figure 10. The four LEDs input PWM wave duty cycles from top to bottom are 10%, 40%, 70% and 100%.

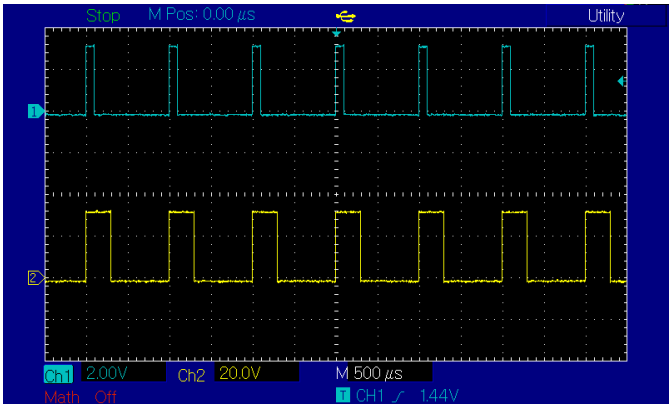


Fig.8. Output PWM waveforms with duty ratios of 10% and 40%

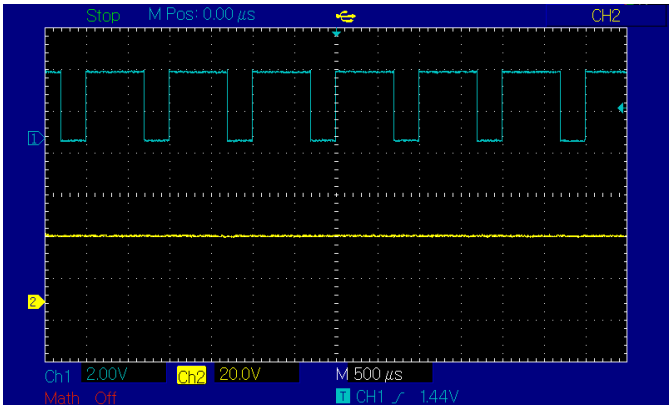


Fig.9. Output PWM waveforms with duty ratios of 70% and 100%

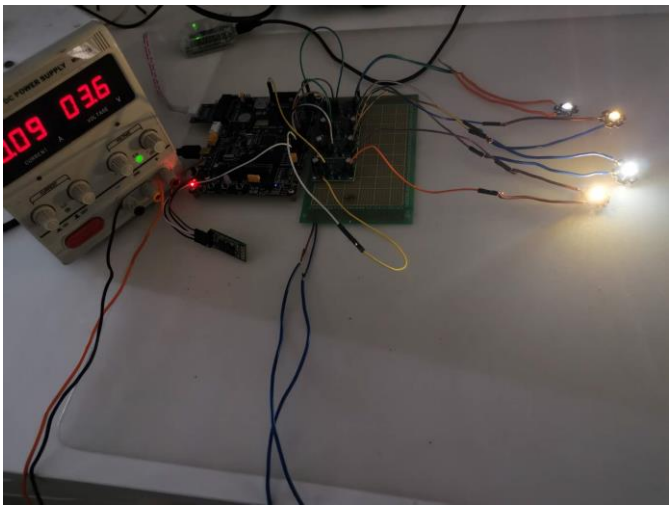


Fig.10. The dimming effect of each LED under PWM wave input with different duty cycles

From the above experiments, the LED dimming system controls the output of a stable voltage and current, and smooth brightness adjustment of the four LEDs can be achieved through the Bluetooth control of the mobile phone, which meets the design requirements of the dimming system.

7 Conclusion

With the deep integration of LED lighting technology and Internet of Things technology, its applications in the lighting field are becoming wider and wider, and at the same time it can better meet people's lighting requirements. This paper selects STM32F103VET6 as the main control chip, combined with Bluetooth wireless communication and achieve brightness control of the four LEDs remotely. It can be seen from the experimental results that the system can smoothly adjust the brightness of the four LEDs to meet the expected requirements. The system has simple structure and convenient control, which provides a reference for the research and design of LED wireless control system.

References

[1] Renjie Liang.: Development status and future prospect of intelligent lighting control technology. Vol. 25, pp. 15-26+31. Illuminating Engineering Journal, China(2014)

[2] Weian Luo, Han Wang, Xindu Chen, Yaobing Zeng, etc.: Three-channel PWM digital control dimming and color matching method. Vol. 39, pp. 414-421. Journal of Luminescence, China(2018)

[3] Chunyuan Zhang.: LED dimming technology and its application. pp. 104-106+42. Fujian Building Materials, China(2019)

[4] Junfeng Hong, Wenqiang Bu, Rongxin Zhang, En Cheng, etc.: Design of Sine Signal Generator Based on STM32 PWM. Vol. 37, pp. 188-191+224. Journal of Hefei University of Technology(Natural Science), China(2014)

[5] Congrong Guan, Weicheng Zhao, Tianyue Zhu, etc.: Design of high-power LED dimming controller based on STM32. Vol. 30,pp. 94-99+104. Illuminating Engineering Journal, China(2019)

[6] Shih-Chang H , Jyun-Jia C , Po-Yu K.: A single chip for 4-channel LED dimming driver of 240W true color display with SPI control. Vol. 78, pp. 16-25. Microelectronics journal, US(2018)

[7] Anan Xu, Chengcheng Ren, Shuping He, etc.: Design of Intelligent LED Desk Lamp Dimming System Based on STC Series Single Chip Microcomputer. Vol. 27, pp. 88-93+116. Illuminating Engineering Journal, China(2016)

[8] Jinbao Yang, Wei Xiang, Chaohua Chen.: Precision capacitance tester based on single chip microcomputer. pp. 15-18. Instrumentation Technology, China(2016)

[9] Qun Zhang, Xu Yang, Zhengyan Zhang, etc.: Design and Implementation of Serial Communication of Bluetooth Module. Vol. 31, pp. 79-82. Research and Exploration in Laboratory, China(2012)

The Design and Verification of a New Light Distribution Curve Applied in Classroom

Gao Wen[1], Zou Nianyu[1,2], Cao Fan[1], He Xiaoyang[1],

Yang Miaomiao[2], Li Jian[2], He Qipeng[2*]

{1621161796@qq.com[1], n_y_zou@dlpu.edu.cn[1,2],*Corresponding author: hqpdyx@163.com[2]}

[1]Research Institute of Photonics, Dalian Polytechnic University, Dalian 116034, China

[2]Innovation & Service Center For Lighting of Guizhou Zhifu Optical Valley Investment Management Co., Ltd., Bijie 551700, China

Abstract: The visual health of teenagers has become the focus of social attention. Good classroom light environment plays an important role in the growth process of teenagers. Classroom lamps are the key factor to build a good classroom light environment. This article is based on the existing common lambert body light distribution curve based on the classroom light building light environment, to further improve the classroom space illumination uniformity, visual comfort, and improve the utilization efficiency of the light, put forward a new type of classroom lamp light distribution curve, and space environment simulation, light type and lambert with comparing the classroom light building light environment, from data analysis, the classroom space uniformity, visual comfort and a considerable improvement of utilization rate of light and verify the feasibility of this kind of light distribution curve, for the construction of high-quality classroom light environment has the reality significance.

Keywords: Classroom light environment, Visual comfort and health, Uniformity of illumination, Light efficiency

1 Introduction

The data show that the overall myopia rate of children and adolescents in China is 53.6% and showing an increasing trend.The prevalence of myopia generally presents a trend of younger age [1-2]. Studies have shown that a good classroom light environment is conducive to the protection of myopia among adolescents [3] and can improve their learning efficiency [4-6]. Classroom light plays an important role in classroom lighting. Proper light distribution curve can build a good light environment. Today's classroom lamps are basically Lambert shaped light distribution curves, and there is a certain distance between lamps and lanterns, resulting in low illumination evenness. The overall visual comfort of the classroom needs to be

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2306914

improved, and the light utilization is not high. In order to achieve the required illumination evenness, the number of lamps has to be increased and energy consumption has to be increased. Therefore, according to the characteristics of the classroom space environment, a new classroom light distribution curve is designed. Using DIALUX EVO built a typical spatial layout of the classroom, import IES of the original file, the designed light distribution curve and lambert light distribution curve for the actual data analysis and contrast of the lighting effect, discovered in this paper, design of light distribution curve, enhances the intensity of illumination uniformity of the classroom, reduces glare, so as to improve the overall visual comfort of the classroom, and improve the utilization rate of light energy, reduce the overall energy consumption of the space, proves the feasibility of the use of the light distribution curve, have the practical significance for the construction of high quality light environment of classroom.

2 Design idea and realization of new light distribution curve

Appropriate light distribution curve plays an important role in building a high-quality light environment, and can improve the utilization of light. In this paper, according to the particularity of classroom desk lighting, according to the scope of irradiation target, target illuminance, illuminance uniformity, the luminescence curve of luminaires in space should be derived in reverse. According to national standards, the average horizontal illumination of classroom desks should be above 300Lx, and the uniformity of illumination should be above 0.7. The height of lamp height from desk surface should be no less than 1.7m [8].

Therefore, this design idea is planned according to the above standards. The simulation classroom area is 62.1m^2 (6.9m × 9m), the test range (the area where the desk surface is located) is 36m^2(6m × 6m), and the test height (the high standard of the desk surface) is 0.75m. The luminescent points are arranged 2 m away from the target irradiation surface, and 4 classroom lights are used to illuminate the desk surface. Therefore, the irradiation range of the target surface of a single lamp is 9 m^2(3 m*3 m) square. In order to achieve better illuminance uniformity, the illuminance is the same everywhere in the range of the designed irradiation surface. In an ideal state, the uniformity is 1.

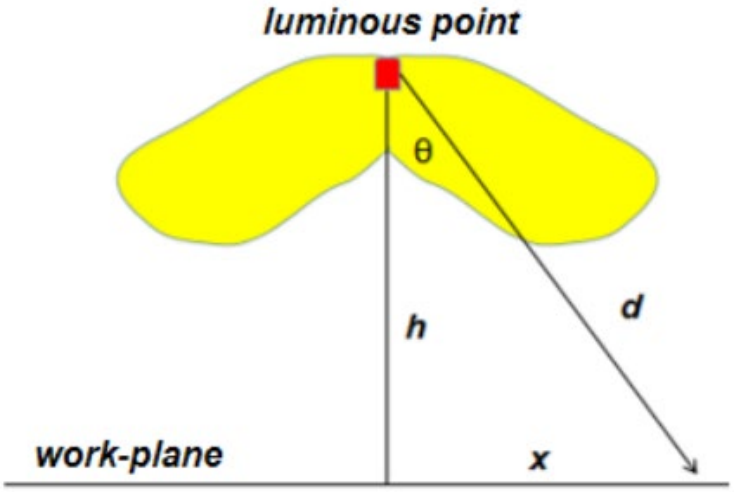


Fig. 1. Calculation diagram

$$I = I(\theta) \tag{1}$$

$$E = \frac{I(\theta)COS(\theta)}{d^2} = \frac{I(\theta)COS^3\theta}{h^2} \tag{2}$$

$$I(\theta) = \frac{h^2 E(\theta)}{COS^3\theta} \tag{3}$$

Where h represents the vertical height of the luminous point from the illuminated surface, and D represents the spatial distance from the luminous point to each irradiated point. From the above equation, we get the computational relationship between luminescence intensity I and illumination, the vertical height of the luminescence point from the irradiation surface h, and the Angle between the luminescence point and the irradiated point. In this calculation, in order to ensure the illumination is equal everywhere, E is determined as a constant value, and H has also been determined as 2m. Therefore, the relationship between light intensity and Angle can be obtained.

Through this formula, the distribution curve of the same circular area with illumination can be calculated, but the design needs to realize the illumination distribution within the

rectangular area. Therefore, on this basis, horizontal Angle C is introduced to construct a cartesian coordinate system along with vertical Angle, as shown in the figure below:

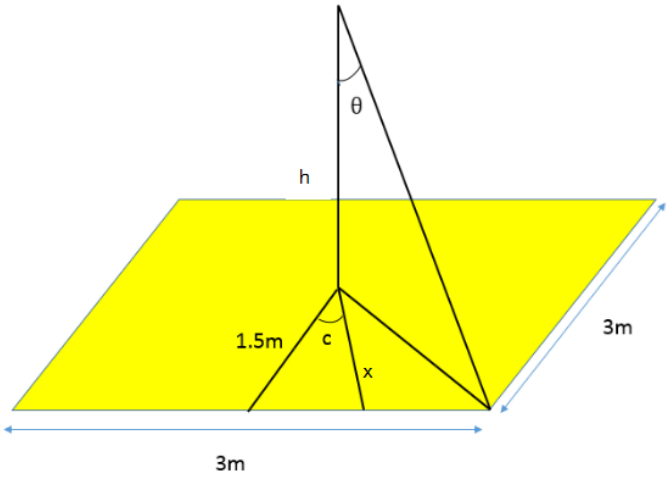


Fig. 2. Schematic diagram of exposure condition

The irradiation area in this design is symmetrical in accordance with quadrant, so only 1/4 area needs to be calculated and designed. Therefore, c=0,3,6... At 90 degrees, according to different c angles, the maximum amount of can be cut off at different c angles. In order to ensure the accuracy of calculation, a smaller interval is taken when calculating, =0,0.5,1.0..... The delta theta is computed as follows:

$$\frac{1.5}{x} = \cos c \tag{4}$$

$$\tan\theta = \frac{\mathrm{x}}{\mathrm{h}} \tag{5}$$

$$h = 2 \tag{6}$$

From (4), (5) and (6), get (7) $$x = 2\tan\theta = \frac{1.5}{\cos c} \tag{7}$$

$$\frac{0.75}{\cos c} = \tan\theta \tag{8}$$

$$\theta \text{ truncatio n} = arc\frac{0.75}{\cos c} \tag{9}$$

Then the relation (3) of spatial light intensity I is calculated from above. The light intensity of each small element is calculated, and the light intensity distribution data of the whole space is obtained. As shown in the following table, the spatial strength distribution corresponding to some micro-elements calculated by software is obtained. In the figure, horizontal Angle C and vertical Angle jointly determine a micro element, form a matrix, calculate the light intensity of each micro element, and jointly constitute the spatial light intensity distribution.

The horizontal Angle distribution C

The vertical Angle distribution is θ

	0	3	6	9	12	15	18	21	24	27	30
0	2400	2400	2400	2400	2400	2400	2400	2400	2400	2400	2400
0.5	2400.274	2400.274	2400.274	2400.274	2400.274	2400.274	2400.274	2400.274	2400.274	2400.274	2400.274
1	2401.097	2401.097	2401.097	2401.097	2401.097	2401.097	2401.097	2401.097	2401.097	2401.097	2401.097
1.5	2402.469	2402.469	2402.469	2402.469	2402.469	2402.469	2402.469	2402.469	2402.469	2402.469	2402.469
2	2404.391	2404.391	2404.391	2404.391	2404.391	2404.391	2404.391	2404.391	2404.391	2404.391	2404.391
2.5	2406.866	2406.866	2406.866	2406.866	2406.866	2406.866	2406.866	2406.866	2406.866	2406.866	2406.866
3	2409.894	2409.894	2409.894	2409.894	2409.894	2409.894	2409.894	2409.894	2409.894	2409.894	2409.894
3.5	2413.48	2413.48	2413.48	2413.48	2413.48	2413.48	2413.48	2413.48	2413.48	2413.48	2413.48
4	2417.625	2417.625	2417.625	2417.625	2417.625	2417.625	2417.625	2417.625	2417.625	2417.625	2417.625
4.5	2422.333	2422.333	2422.333	2422.333	2422.333	2422.333	2422.333	2422.333	2422.333	2422.333	2422.333
5	2427.608	2427.608	2427.608	2427.608	2427.608	2427.608	2427.608	2427.608	2427.608	2427.608	2427.608
5.5	2433.455	2433.455	2433.455	2433.455	2433.455	2433.455	2433.455	2433.455	2433.455	2433.455	2433.455
6	2439.878	2439.878	2439.878	2439.878	2439.878	2439.878	2439.878	2439.878	2439.878	2439.878	2439.878
6.5	2446.884	2446.884	2446.884	2446.884	2446.884	2446.884	2446.884	2446.884	2446.884	2446.884	2446.884
7	2454.478	2454.478	2454.478	2454.478	2454.478	2454.478	2454.478	2454.478	2454.478	2454.478	2454.478
7.5	2462.666	2462.666	2462.666	2462.666	2462.666	2462.666	2462.666	2462.666	2462.666	2462.666	2462.666
8	2471.456	2471.456	2471.456	2471.456	2471.456	2471.456	2471.456	2471.456	2471.456	2471.456	2471.456
8.5	2480.856	2480.856	2480.856	2480.856	2480.856	2480.856	2480.856	2480.856	2480.856	2480.856	2480.856
9	2490.872	2490.872	2490.872	2490.872	2490.872	2490.872	2490.872	2490.872	2490.872	2490.872	2490.872
9.5	2501.515	2501.515	2501.515	2501.515	2501.515	2501.515	2501.515	2501.515	2501.515	2501.515	2501.515
10	2512.794	2512.794	2512.794	2512.794	2512.794	2512.794	2512.794	2512.794	2512.794	2512.794	2512.794
10.5	2524.718	2524.718	2524.718	2524.718	2524.718	2524.718	2524.718	2524.718	2524.718	2524.718	2524.718
11	2537.298	2537.298	2537.298	2537.298	2537.298	2537.298	2537.298	2537.298	2537.298	2537.298	2537.298
11.5	2550.546	2550.546	2550.546	2550.546	2550.546	2550.546	2550.546	2550.546	2550.546	2550.546	2550.546
12	2564.473	2564.473	2564.473	2564.473	2564.473	2564.473	2564.473	2564.473	2564.473	2564.473	2564.473

Fig. 3. Light intensity distribution corresponding to partial micro-elements

The above light intensity distribution data will be compiled in the format of IES file. Classroom lights are designed in this paper, so the data will be integrated in the format of class C indoor lamps. [9]Take a 0.5 degree interval between the angles on the vertical axis θ ,The range of points is 0-90 degrees, and the number of points is 181 times. The horizontal axis Angle is taken at intervals of 3 degrees. The range of points is 0-90 degrees, and the number of points is 31 times. The two forms a matrix, divides the irradiated surface into 5611 small elements, and the light intensity of each part forms the light intensity distribution curve.

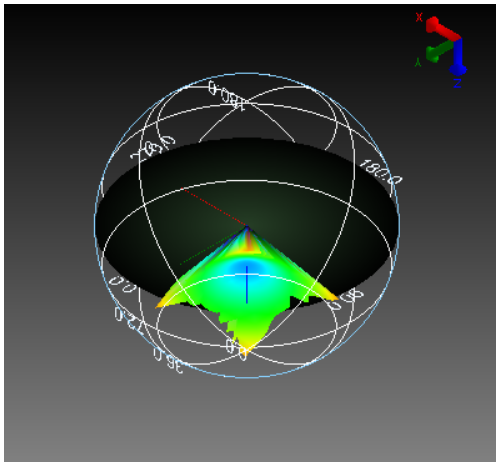


Fig. 4. 3D distribution of the designed light distribution curve

3 Simulate the effect of classroom environment

After get the light distribution curve, according to the current common classroom layout model, import the design files of the IES of light distribution curve and the market is more popular nowadays lambert shape light distribution curve of the classroom light, to build two kinds of light distribution curve of the indicators of the classroom light environment carries on the analysis comparison, in addition to energy consumption are compared.

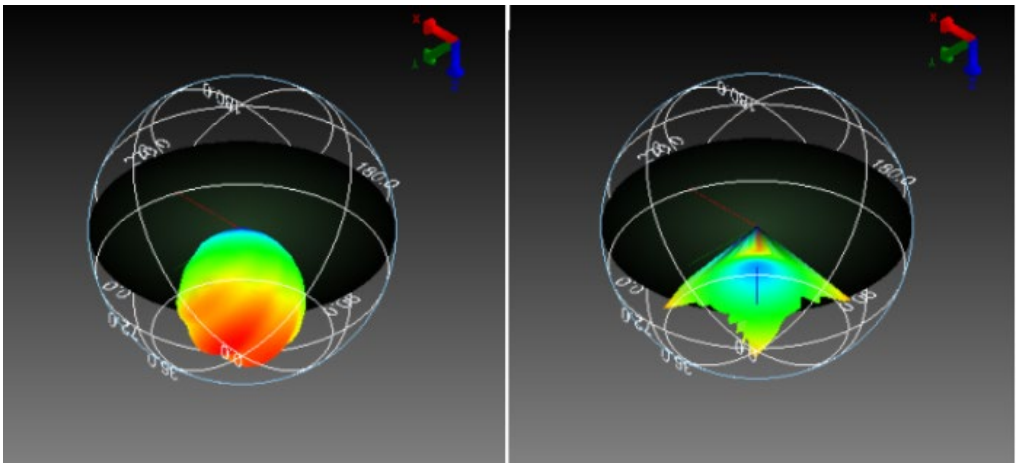


Fig. 5. Lambert type distribution for comparison (left) and design light distribution curve (right)

The simulation classroom covers an area of 62.1m^2 (6.9m×9m). According to national standards, the reflection ratio of each area is set. The reflection ratio of side wall and back wall is 0.77, the reflection ratio of front wall is 0.6, the ground is 0.35, and the desktop is 0.26. The test area (desk surface area) is 36m^2 (6m×6m), and the test height (desk surface height) is 0.75m. UGR calculation point is set in the middle chair of the last row, with a height of 1.2m, where the glare of the whole classroom is the most severe, simulating the overall glare received by the students in the last row. Since there is no clear height regulation for the installation of classroom lights, it is only required to be no less than 1.7m from the desk surface. In combination with the design requirements, the height of both lamps is set to 2m to illuminate the desk surface where the students are. The layout is perpendicular to the blackboard surface and uniformly distributed in the space. The specific parameters and lighting effect parameters are shown in Table 1.

Table 1. Comparison of classroom parameters and effects of two kinds of light distribution

Classroom parameters and lighting effects of two types of light distribution		
	Lambert's light distribution	This design distributes light
Overall classroom area	6.9m*9m=62.1m^2	
Target test area	6m*6m=36m^2	
Number of classroom lights	9 lamps	4 lamps
Luminance flux (lm)/power (w)	2000lm/30w	4000lm/60w
horizontal illumination（lx）	307lx	383lx
Uniformity of horizontal illumination	0.73	0.86
UGR	19	15

Fig. 6. Renderings of simulated classroom space(reading area)

By analyzing the data obtained from the simulation, it is found that the light distribution curve is precisely designed according to the environment used in this paper, so that the light environment of the whole classroom is improved. Compared with the traditional Lambert-type curving classroom lamps, the illuminance and uniformity of the desks in the classroom are improved under the premise of the same height of the lamps. The average illuminance of the desks in the classroom is 383lx and the uniformity is 0.86. UGR glare has also been reduced from 19 to 15, with some improvement in glare. In addition, the design of light distribution curve, improve the utilization of light energy, on the premise of reducing the space total light flux, improve the target surface illuminance. The luminous flux of the whole classroom was saved by 2000LM and energy consumption was nearly 11%, but the illumination of the target surface and the uniformity of illumination were increased by 24% and 18%, respectively. The data analysis shows that the light distribution curve of this design can be well applied to classroom lighting and play a role in improving the quality of light environment.

Fig. 7. Classroom lighting environment rendering using the light distribution curve lamps

4 Summary

According to the present classroom space environment and actual lighting requirements, this paper designs a new light distribution curve suitable for classroom lamps. This light distribution curve was introduced into a typical classroom space, DILUX EVO was used for simulation, and the average illumination and illuminance uniformity of the target exposed surface (desk surface) were considered. It was found that the illuminance uniformity of the target surface was increased by 18% compared with the traditional Lambert-type light distribution classroom lamp under the premise of the same light distribution height. The overall UGR of the space is also controlled, and the classroom glare value using this light distribution curve is 15. In addition, the utilization rate of light energy was improved, the overall luminous flux consumption was reduced by nearly 11%, and the illumination was increased by 24%, which verified the feasibility of using the distribution curve. It is of practical significance to construct high quality classroom lighting environment.

References

[1] The National Health Commission: in 2018, the overall myopia rate of children and adolescents was 53.6% [J]. Modern Health (second Half),2019,(5):1-2.

[2] Kang P． Optical and pharmacological strategies of myopia control［J］． Clin Exp Optom，2018，101(3):321-332.DOI:10．1111/cxo．12666．

[3] Jiang Sibin, Wang Zhenghe, Yu Hong, Chen Manman, Xu Jianxing, Ma Jun. Influence of classroom light transformation on vision and poor vision of Primary and middle school students [J]. Journal of lighting engineering,2019,30(03):15-18.

[4] Govén T, Raynham P, Laike T, et al. The influence of ambient lighting on pupils in classrooms-considering visual, biological and emotional aspects as well as use of energy[J]. Adjunct proceedings experiencing light 2009, 2009: 13.

[5] MAO Peng, Wang Wenhan, Li Jie, et al. Effects of classroom illumination on learning efficiency [J]. Journal of Southeast University (Medical Edition), 2016,37(4):662-666. DOI:10.3969/ J.issn.1671-6264.2018.04.023.

[6] Guo Ya, Zeng Shanshan, HAO Wentao, et al. Effect of LED classroom lighting on human eye physiological characteristics [J]. Journal of lighting engineering,2019,30(3):7-14. DOI:10.3969/ j.issn.1004-440x.2019.03.002.

[7] GB 7793-2010, Hygienic Standards for daylighting and Lighting in classrooms of primary and secondary schools [S].

[8] Zou Jiping. Light Distribution Curve of Lamps and its Standard format of electronic Document [J]. Building Electric,2008(01):37-41.

Parking Space Recognition Based on Deep Convolutional Neural Network

Zhuowen Chen, Zijun Gao, Jiaqi Li, Junjie Zhang, Yanan Dai, Wenbo Hu, Changmao Li
{1176655862@qq.com , gaozj@dlpu.edu.cn , 1499842219@qq.com}

School of Information Science and Engineering, Dalian Polytechnic University, Dalian, China

Abstract. Automatic real-time recognition for lots of spots is a crucial and challenging task, which can significantly improve the intelligence level of a city and make it that much more convenient for people to travel. Deep learning models have been applied in various fields, and significant development has also been achieved. However, taking into account interference in recognition images and occlusions of lane lines, deep learning still lacks accuracy and real-time.Therefore, the paper argues that combining OpenCV processing and deep learning should be utilized. This method can improve recognition for identifying parking spaces and increase the intellectualization of a city. In this paper, a network model based on VGG19 has been proposed, and the SENet network module has been added to its output along with a performing secondary enhancement on training images. Advanced pre-processing to obtain accurate point coordinates has finally obtained a satisfactory effect. It also[1] predicts images with severe interference and images with line occlusion. These two experimental results show that the network can effectively identify even with severe interference. At the same time, a large number of experimental evaluations also show that this method can also be applied to small target recognitions in many fields and can also be used as a basic product for small target detections and recognitions in the future.

Key Words: Pretreatment, Deep Learning, Point Transformation, SENet Network, Feature Enhancement

1 Introduction

Parking space recognition refers to capturing images through a fixed camera, using the captured images to determine the changes in parking spaces and the total number of vehicles in the parking lot in real time. This is a key task in the field of target recognition. Because of its wide application from video surveillance to target tracking, it has attracted widespread attention in the industry.

Although true [1,2] proposed a new method of detecting parking spaces in 2007, in which empty parking spaces in static images are used as markers to identify parking spaces frame by frame. This method uses the feature point distribution and color histogram of each car, and uses a combination method to identify and predict parking spaces. However, due to its large amount of calculation and error, it couldn't be accurately identified in real time, which couldn't meet the requirements of intelligent parking space identification. In 2009, Ichihashi et al. [3] proposed a new camera system (ParkLotD), which is based on C-means fuzzy clustering and

Project: Scientific Research Project of Education Department of Liaoning Province (J202012)

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307757

particle swarm optimization to adjust the hyperparameters, improving the accuracy of indoor vehicle detection and outdoor space, and the success rate of parking space detection. In 2018, Karakaya et al. [4] suggested another method of parking sp。ace detection based on deep learning, this method obtains a model by constructing a deep convolutional neural network to process parking lot vehicles, simply collect useful information and make predictions. However, The prediction accuracy and real-time results of this model are poor, and it is unable to predict the parking situation of the parking lot, and it has not conducted in-depth research on the network.

Real-time detection of parking spaces is essential for the development of an intelligent society. but there are still deficiencies in this part of the research. This paper proposes a real-time detection and recognition method for parking spaces based on deep convolutional neural network VGG19[5,6,7] combined with SENet[7,8] module. The premise of this method is pre-processing by OpenCV. Compared with other methods, this method does not have interference-free images for initialization, but uses single-frame images extracted from the video for processing. Through the prediction of the model, interference is effectively eliminated, and the prediction accuracy reaches 95%. Since it is based on single-frame image processing, our method has good generalization and can be extended to more fields.

2 Related Work

Completing the intelligent recognition of parking spaces is one of the key steps to promote urban modernization, and there are several representative studies on it. The first is the shortest search method based on genetic algorithm by intercepting video frames as a management medium. The second method is to recognize the license plate and extend it to the vehicle. However, due to the pollution and corrosion of the license plate, and the camera installation specifications, the license plate recognition system is prone to failures, recognition errors and high costs, which is not universal [9]. The third method is to conduct research on the basis of wireless communication technology and image processing technology. The above methods still have disadvantages such as low degree and slow speed.

The core technology of finding vacant parking spaces in parking lots through video frames captured by surveillance cameras is to accurately identify vacant parking spaces in parking spaces. In this paper, through a lot of analysis of the existing domestic intelligent parking lot management system parking space detection technology, comprehensive advantages and disadvantages, found that vehicles detection algorithm based on the image features of vehicles and empty parking spaces can be well applied to large area of outdoor parking lots. An image processing system based on video frame objects is proposed. Use the deep convolutional neural network VGG19-SENet to enhance vehicle target features, improve accuracy, and reduce loss. The information collection function can help drivers find free parking spaces quickly and accurately, save time to look for vacant parking spaces and occupy as little traffic resources as possible.

Through a large amount of research and analysis on the state and features of vehicle images, after the necessary noise filtering and other morphological preprocessing of the overall image of the parking lot [12], the key areas and irrelevant areas of the parking lot are specifically identified. Carry out image segmentation [13], so as to obtain the area of interest with less interference in the parking space, and then obtain the specific parking space image, achieve

the purpose of obtaining the spatial structure information of the parking space, and facilitate the analysis and integration of image information.

The image is input into the VGG19-SENet network structure[10], the important advantage is that the simple structure of VGGNet is adopted and the SENet image enhancement network is introduced. At the same time, it strengthens the learning ability of the network structure for parking features, and the training model can classify images more accurately. After all the fully connected layers are removed, computing resources are saved, and system performance is not significantly reduced.

3 OpenCV Pretreatment

3.1 Introduction to preprocessing

In order to improve the accuracy of parking space detection results, we need to perform image preprocessing on the parking lot video frames collected by the camera, so as to achieve the purpose of removing redundant information in the image and improving the reliability of the parking space image input to the network. In this paper, the preprocessing of the parking lot video image includes morphological operations to remove noise and to obtain the ROI area of each detection feature by fitting a straight line.

3.2 Image Denoising

Image denoising [14] is a process of reducing noise in digital images, sometimes called image denoising, to improve the image signal-to-noise ratio, so that the image quality in the input network is better. In this paper, due to the influence of the optical characteristics of the camera and the lighting changes of the captured video, there is a certain amount of noise in the collected parking map. In order to prevent noise from negatively affecting the final parking detection during the training model, the key is to eliminate noise. Noise removal is divided into two parts: coarse filtering and fine filtering. First, the dark spots and the extraneous color information in the top view of the parking lot of this paper are binarized according to the threshold, performing bitwise_and operation on the processed image and the original image. For the photos of the parking lot, median filtering is generally used for the first processing to filter out most irrelevant factors.

Fig. 1. Comparison of original image and coarse filtered image.

Fine filtering mainly uses the Canny operator in the OpenCV library to perform edge detection on the grayscale image of the parking lot. The working mode is mainly to detect the horizontal and vertical of the image and the diagonal pixels of the operator to obtain the horizontal direction G_x and the vertical direction G_y respectively. The purpose of the first-order derivative value is to obtain the edge detection binary image that can express the pixel gradient intensity G and the gradient direction θ, thereby obtaining the edge detection binary image of the overall range of the parking lot. The Canny algorithm is calculated as formulas (1) and (2):

$$G=\sqrt{G_x^2+G_y^2} \tag{1}$$

$$\theta=\arctan\left(\frac{G_y}{G_x}\right) \tag{2}$$

The above work can complete the conversion of RGB images of the complex parking lot to the image denoising tasks that can clearly distinguish the edge of the parking space and part of the car outline, which is more suitable for machine processing [16].

Fig. 2. Edge binary image obtained by fine filtering.

3.3 ROI Area Selection

After obtaining a clearer parking space edge and contour image in formula (3), the main goal of this section is to select the fixed range of the parking lot by the method of straight line fitting, and use appropriate for the edge contour of the binarized image. The geometric feature extraction algorithm [17] draws the exact position of the edge of the parking space [18], then divides each train space into a cluster, and finally divides each parking space.

Firstly, the boundary corner points of the parking lot are selected to determine the general range of the parking lot, multiple boundary line equations are obtained by linear fitting through each corner point, the mask image of the parking lot is drawn, and the image after edge detection is matched through the mask operation to obtain the required ROI image.

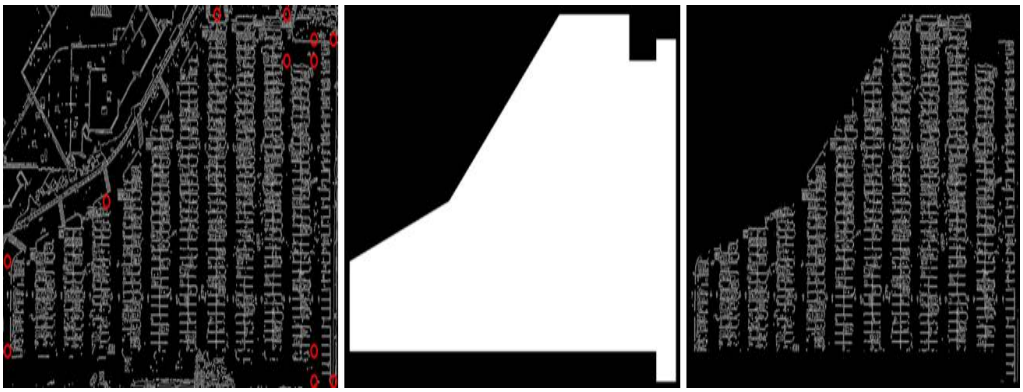


Fig. 3. ROI area edge image.

According to the feature difference between the straight line feature of the parking space and the irregular shape of the vehicle contour, the Houph transform that can filter the curve to determine the straight line can be used to obtain the boundary of each parking space [19]. The main principle of Houph straight line detection is to detect each edge, a straight line that may contain pixels in the Cartesian coordinate system is mapped to the polar coordinate system. The straight line is extracted by voting, the mapping relationship is as follows:

$$\rho = x\cos\theta + y\sin\theta \tag{3}$$

However, the shortcoming of Houph transform that it is easy to lose the line segment length information. As a result, there is still some redundant information in the parking space boundary image, and no clear and distinct parking space segmentation map is obtained. In view of this situation, the bubbling sorting of each Houph line with the boundary conditions of the parking space can filter out these redundant parking space boundary lines, and get a clear parking space distribution boundary [20]. By calculating the starting parking point according to the space line spacing, the boundaries of all parking spaces can be determined, and the coordinates of each parking point can be obtained, then the initial state of each parking space can be obtained, which is convenient for subsequent segmentation and sampling of the parking space image.

4 Network model structure

This paper designed a deep convolutional neural network. The purpose is to accurately predict the situation of parking spaces in the parking lot and mark them in real time in the original image. Because the actual situation in the later stage is affected by complicated and special circumstances such as vehicle blurring and parking space occlusion, the final judgment result shows that the VGG19-SENet [21,22,23] network must have strong suppression ability against interference and noise, and have strong robustness. The network can still get better results in various complex situations.

4.1 SENet

Figure 4 shows the overall structure of the SENet network, which mainly contains the Squeeze module and the Excitation module. The Squeeze module performs convolution operations in a local space, and encodes all spatial features on a channel into a global feature. This paper uses global average pooling as the first step, complex aggregation methods can also be used. The Excitation module is an operation that integrates the relationship between channels after obtaining the global description characteristics. It has two characteristics. First, it is very flexible and can obtain the nonlinear relationship between each channel; the second is the learning relationship can be fused, allowing multi-channel features to be learned. Therefore, it can reduce the complexity of the model and improve anti-interference ability [25,26]. The basic mapping relationship is as follows:

$$Z_c = F_{sq}\left(u_c\right) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} U_c\left(i,j\right) \tag{4}$$

$$s = F_{ex}\left(z,w\right) = \sigma\left(g\left(z,w\right)\right) = \sigma\left(W_2 \operatorname{Re} LU\left(w,z\right)\right) \tag{5}$$

$$X_c = F_{scale}\left(u,S_c\right) = S \cdot U_c \tag{6}$$

The F_{sq} and F_{ex} in formulas (1) and (2) respectively represent the feature compression and extraction of the input feature maps [27,28] to achieve the purpose of reducing the dimensionality of the data. The flexibility of the SENet network is that it can easily connected with the existing network to achieve the effect of enhancing feature information and speeding up training. In this paper, the network that VGG19 and SENet are connected together is used.

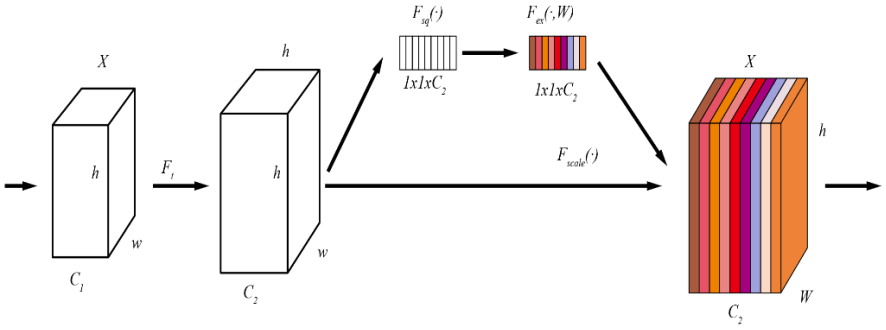


Fig. 4. SENet network structure.

4.2 VGG19-SENet

Based on Karen Simonyan's VGG network, this paper constructs the VGG19-SENet network model. VGGNet explores the relationship between the depth of the convolutional neural network and its performance, with multiple 3×3 small convolution kernels and 2×2 maximum pooling layer, VGGNet successfully constructed a 16-19 layer deep convolution. The entire network uses the same size as the convolution kernel (3×3) and the maximum pooling size (2×2). Convolutional concatenation uses a larger convolution kernel alone, has fewer parameters, and has more nonlinear transformations than a single convolution layer. However, because VGG has three fully connected layers, it uses more parameters, consumes more computing resources, and reduces training speed. In order to optimize the network structure and reduce the number of parameters, the three full connection layers of VGG are removed, and the SENet network is added before OUTPUT to improve the training speed of the model and improve the accuracy of prediction. Finally, after the feature enhancement of SENet, the softmax function is used for output, and better results can be obtained.

Network structure, the VGG19-SENet network structure is shown in Figure 5. After the input image [48,48,3] is entered, first two layers of convolution are performed to obtain [48,48,46], and one layer of pooling output is performed [48,48,128]. The output image obtained after five rounds of convolution and pooling operations in the figure is [1,1,512]. Then enter the SENet network for feature enhancement, finally add a full connection layer, and use the Softmax function for output.

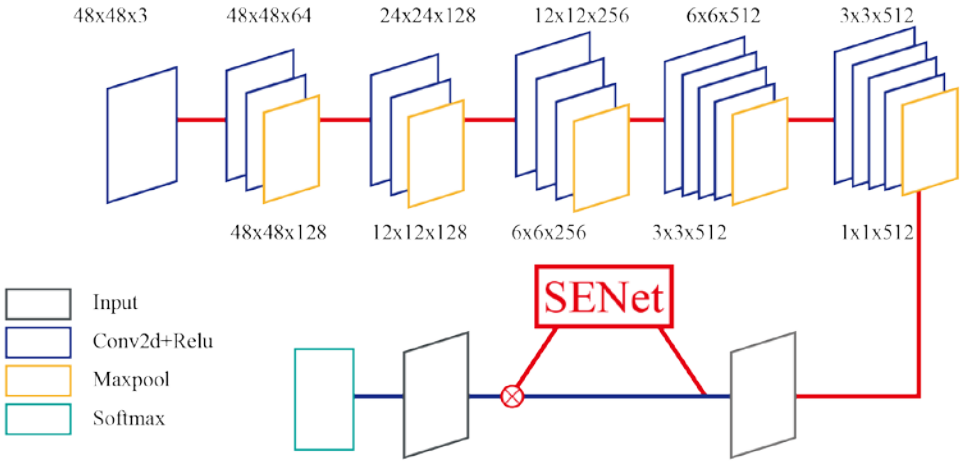


Fig. 5. Final network structure flow chart.

5 Experiment

In this paper, the VGG19-SENet network model is designed using the Keras framework. The optimizer selects SGD. Set Lr = 0.0001, momentum = 0.9, and learning rate 0.002. Using categorical_crossentropy as the loss function, and set other hyperparameters to default values. When generating a model, ImageDataGenerator is used to generate images in batches and enhance image features. As an image generator, it can iteratively loop and feed images into the network in batches.

5.1 Datasets and Expeerimental Settings

Data set: The data set used in this paper is composed of two parts, one uses the image obtained after OpenCV preprocessing, then put the coordinates of each parking space together. As shown in Figure 6, save the image of each car, the image size is 48*48. The other part uses Python crawlers to get top views of different types of cars on the Internet, which is also saved as 48*48. In the end, training set is 6000 images, and test set is 2000 images. This paper mainly sets the ROI area of the predicted images, removes the severely disturbed area, and reconstructs the place where the lane line is blocked, to show the results of OpenCV preprocessing and the excellent characteristics of the network.

Evaluation scheme: In the experiment, in this paper, each frame of parking lot image is predicted, and the accuracy of the prediction is used as an index to evaluate the performance of the VGG19-SENet model. At the same time, it can still perform well for videos with low definition and strong interference. In the case of severe interference, predicting objects with lower definition is a great challenge for our model and data set. We also evaluate the performance of the model through this experimental result.

5.2 Implementation Details

The prediction of parking spaces in this paper is based on two parts. The first part is to use the OpenCV tool to preprocess to obtain the point coordinates of each parking space, input the point coordinates into the model, and predict each parking space through the model in each frame of the image. The second part is to design a VGG19-SENet network structure, through SENet network and ImageDataGenerator class twice to enhance the image to improve accuracy. After basic preprocessing, the key is to classify the point coordinates into clusters, divide the parking lot by column, divide the parking lot line equidistantly, and accurately index each parking space. In the VGG19-SENet network, this paper uses different network models to make predictions and compares them with our network. The results are shown in Table 1. For our VGG19-SENet network, after experiments, we adjusted the convolution kernel to 3*3 and the step size to 1, and the predicted experimental results were the best, as shown in Figure 6.

Table 1. Accuracy of different networks and time used.

NET	Accuracy	Time
VGGNet	84.3%	12.4s
ResNet-50	78.2%	9.2s
ResNet-101	91.3%	15.1s
ResNet-152	90.3%	12.8s
BN-Inception	86.5%	14.2s
VGG19-SENet	95.6%	8.3s

Fig. 6. Predict experimental results

6 Conclusion

This paper proposes a method of VGG19-SENet network to improve the accuracy of parking space recognition. This method first uses OpenCV preprocessing to cut the image, obtains coordinates through point transformation, and then uses a new network model to predict the parking spaces. The key point is to use SENet network and ImageDataGenerator class to enhance the target feature twice to improve the success rate of small target detection. This is a method that has never been explored in the current technology. For this reason, we have added the SENet module to the VGG19 network. The function is to classify and strengthen the features after the output of the full connection layer of VGG19 to make the experimental results better. By comparing the experimental results of different networks and the prediction results under severe interference, the effectiveness of this method is verified, which is better than some current methods. In addition, experimental verification shows that our proposed network has good generalization and can be applied to different scenarios.

References

[1] Xu Lexian, Chen Xijiang, Ban Ya, Huang Dan.: Intelligent parking space detection method based on deep learning. China Laser. Vol. 46, pp. 230-241 (2019)
[2] True N. Vacant parking space detection in static images. pp. 17 (2007). San Diego:University of California
[3] Ichihashi H, Notsu A, Honda K, et al. Vacant parking space detector for outdoor parking lot by using surveillance camera and FCM classifier. IEEE International Conference on Systems, Man and Cybernetics. pp. 3603-3608 (2013)
[4] Karakaya M, Akıncı F C. Parking space occupancy detection using deep learning methods.2018 26th Signal Processing and Communications Applications Conference. pp. 1-4 (2018)
[5] Soft Computing; Findings from Xinxiang Medical University Provides New Data about Soft Computing (A Novel Multi-focus Image Fusion By Combining Simplified Very Deep Convolutional Networks and Patch-based Sequential Reconstruction Strategy). pp. 1275-1279 (2020). Computer Weekly News
[6] Saha S. Pagnozzi Alex, Bourgeat Pierrick, et al. Predicting motor outcome in preterm infants from very early brain diffusion MRI using a deep learning convolutional neural network (CNN) model. Neuroimage. pp. 215:116807 (2020)
[7] Hua Cui, Gege Yuan, Ni Liu, et al. Convolutional neural network for recognizing highway traffic congestion. Journal of Intelligent Transportation Systems. Vol. 24, pp. 279-289 (2020)
[8] Ze.L. Design of automatic parking space visual inspection and positioning system based on deep learning and OpenCV. pp. 15-85 (2019). Jiangsu University
[9] Fei.X. Research on Automatic Recognition Technology of Parking Space Based on Visual Inspection. pp. 1-72 (2015). Shenyang Polytechnic University
[10] Wujie.D.: Research and Application of Parking Space Recognition Method Based on Digital Image Technology. pp. 3 (2018). Guangxi Normal University
[11]Wenwen.W, Hongcai.T.: Research on Anomaly Detection Model based on Optimized VGG19 Convolutional Neural Network. pp. 253-258 (2020). Journal of Chengdu University of Information Technology
[12]Liangchao.J, Chuanyou.L, Fanqing.Y.: Lane Detection Based on OpenCV. pp. 41-44 (2018). Chang'an University
[13] Zhang Wan, Jin Fengxiang, Zhao Xiangwei, Ji Min, Li Ting.: Research on close-range image enhancement and segmentation algorithm based on OpenCV. pp. 881-886 (2018). Beijing Surveying and Mapping
[14] Hong.Y, Tong.Z.: Research on Classical Image Denoising Algorithms. Information and Computer. pp. 66-67. (2020)
[15] Lei.L.: Parking space detection and recognition method based on panoramic vision automatic parking. pp. 1-80 (2018). Xi'an Electronic TechnologyUniversity
[16] Yunong.W.: Research on Recognition Method of Parking Space Based on Image Similarity. pp. 1-68 (2018). Ji'lin University
[17] Guoxu.D.: Research and Practice of OpenCV Contour Recognition. Equipment manufacturing technology. pp. 101-103 (2020)
[18] Xiaochen.Z.: Research and Implementation of Parking Space Detection and Reversing Assistant Algorithm Based on Computer Vision. pp. 1-77 (2011). Northeastern University
[19] Yusong.H.: Research and Implementation of Parking Space Detection in Garage Based on Binocular Vision. pp. 1-90 (2019). Chong'qing University
[20] Xiaohang.L, Jia.G, Fulun.P, Jianjun.M, Cheng.W, Bo.S.: Improved algorithm of road detection based on Hough transform. Applied Optics. pp. 229-234 (2016)
[21] Daniel Octavian Melinte, Luige Vladareanu.: Facial Expressions Recognition for Human–Robot Interaction Using Deep Convolutional Neural Networks with Rectified Adam Optimizer. Sensors. pp. 1-18 (2020)
[22] Yang Ke, Wang Weibin, Xu Weiqing.: Parking space detection method based on small deep learning network. Automobile Practical Technology. pp. 155-158 (2020)

[23] Yongfei.X, Zhongming.Y, Yucheng.Y.: Outdoor parking lot empty space detection based on Faster R-CNN and ResNet. Fujian Computer. pp. 32-34 (2018)
[24] Zhifeng.Z, Jinqing.L, Wenzao.S.: Intelligent parking space query system based on deep learning. Computer System Applications. pp. 107-114 (2019)
[25] Sven Koitka, Lennard Kroll, Eugen Malamutmann, et al.:Correction to: Fully automated body composition analysis in routine CT imaging using 3D semantic segmentation convolutional neural networks. pp. 1-2 (2020)
[26] Zhang, C, Liu, S, Xu, X, Zhu, C.: C3ae: Exploring the limits of compact model for age estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 12587-12596 (2019)
[27] Xie, J.C, Pun, C.M.: Deep and ordinal ensemble learning for human age estimation from facial images. IEEE Transactions on Information Forensics and Security 15. pp. 2361-2374 (2020)
[28] Shen, W, Guo, Y, Wang, Y, Zhao, K, Wang, B, Yuille, A.L.: Deep regression forests for age estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2304-2313 (2018)

Defect Detection and Recognition of Mobile Phone Membrane Based on Convolutional Neural Network

Changmao Li[1], Enbo Zhang[2],Li Liu(✉)
{2538001101@qq.com[1],248290248@qq.com[2],link_liuli@hotmail.com (✉)}

Department of Information Science and Engineering Dalian Polytechnic University Dalian, P. R. China

Abstract. With the upgrading of mobile phone equipment, automatic detection of mobile phone film defects has been paid more and more attention in industrial production quality. Mobile phone film defect detection is a huge workload and challenging problem. Traditional methods can also detect some industrial identification defects, but these methods can only detect defects under specific conditions, such as obvious defect outline, strong contrast, low noise conditions. The defect detection method of mobile phone film proposed in this paper is to locate the target area with input images obtained from the industrial environment, remove the background, and then classify them into their designated classes through convolutional neural network. Experimental results show that this method can meet the robustness and accuracy of mobile phone film defect detection.

Keywords: Mobile phone film; Target location; Convolutional Neural Network; Defect detection.

1 Introduction

Mobile phone film defect detection is an extremely important part of the operation for manufacturers, of course, manufacturers also make a lot of efforts on defect detection and quality control. In the photos of mobile phone film surface, problems such as strong reflection, background noise and difficult observation of defects are likely to occur, which increase the difficulty of detection. **Figure 1** shows an image of mobile phone film defect.

It can be seen from **Figure 1** that although there are friction trace defects, the background noise is complex and the reflection is strong, which poses a great challenge to the detection of mobile phone film defects. In recent years, machine vision-based methods can overcome many man-made shortcomings, such as low detection accuracy and poor performance, and have become a new research method for trend surface defect detection. Real time, high subjectivity and high work intensity. These inspection systems based on machine vision appear in many industrial applications, such as steel strip inspection [1.2].

✉ Corresponding author: Li Liu
Fund Project:Danlian Science and Technology Innovation Fund (2020JJ26GX029)

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307711

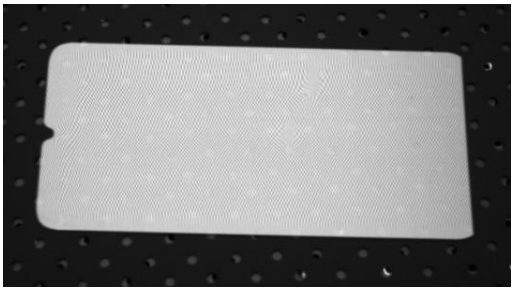


Fig. 1. Mobile phone film defect image.

Liquid crystal display (LCD) [3], organization inspection [4.5], aluminum profile [6], railway track inspection [7], food inspection [8], optical component inspection [9]. The application of machine vision technology on surface defects has been extensively studied and tested. The traditional image processing method is based on the detection of local abnormal image pixels to solve the problem of prediction defects, which can be divided into structural methods, threshold methods, spectral methods and model-based methods [10]. Structural methods include model matching [11] and morphological processing [12]. Threshold methods include iterative optimization method threshold [13], OTSU method [14], contrast adjustment threshold [15], watershed method [16] and so on. Spectrum methods usually include Fourier transform [17] and wavelet transform [18]. Model-based methods include the Gaussian mixture entropy model [19]. A method based on machine learning usually includes two stages: feature extraction and model classification. Feature extraction includes local binary model (LBP) function [1] and gradient direction histogram (HOG) function [20]. **Figure 1** is the friction defect of the mobile phone film. As can be seen from the figure, the background is mostly black with white spots. These methods usually target specific conditions and lack adaptability and robustness to the aforementioned detection environment. Mobile phone film defect detection is a huge workload and challenging problem. Traditional methods can also detect some industrial identification defects. Although these detection algorithms have obtained good detection results in all aspects of surface defect detection, they cannot be directly applied to the above-mentioned mobile phone film surface detection. In the past decade, many researches have been devoted to the application of machine vision technology in surface defect detection, and The neural network algorithm is classifying the target, classification of natural scenes got good results [21]. A flexible multi-layer deep feature extraction framework based on CNN is proposed to detect anomalies in anomalous data sets [22].Lin et al. Established a convolutional neural network (CNN) for fault checking of LED chips [23]. Defective areas are located using activation-like mapping technology without the need for annotation at the human area level. Liu et al. A detection system based on neural network detection level (DCNN) has been proposed.

2 Image preprocessing

As shown in **Figure 1**, the defect image of mobile phone film was collected in the industrial environment by using an industrial camera. Gaussian smoothing is done to the input image, and the density function of one-dimensional Gaussian normal distribution is used：

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad (1)$$

Where μ is the mean of x and σ is the standard deviation of x. Since each calculation takes the current calculation point as the origin, μ is equal to 0. Morphological processing is performed on the image, and interference factors in the background of the image are removed through corrosion and expansion operations. Corrosion is to replace the pixel value of the image under the anchor point overlap with the minimum value, and the expansion is to replace it with the maximum value on the contrary. Finally, through binarization processing, the gray histogram of the image is first calculated. Assuming that the mean value of the gray value is 130, the mean value is called M. Now arbitrarily select A gray value t, then the histogram can be divided into two parts, called A and B, respectively, and the average value of these two parts is MA and MB. The proportion of the number of pixels in part A to the total number of pixels is PA. Same thing with PB. The inter-class variance given by binarization is defined as:

$$ICA = PA * (MA - M)^2 + PB * (MB - M)^2 \quad (2)$$

Since the mobile phone film is easy to distinguish from the image background, through the above method, this article first uses Gaussian filtering to smooth the image, and then calculates the histogram of the picture to get the pixel threshold of the background and the mobile phone film. Use this critical value to apply binarization processing to the mobile phone. The membrane and the background can be separated, and then the mobile phone membrane can be cut out according to the image. The final picture of the mobile phone membrane pretreatment is as follows:

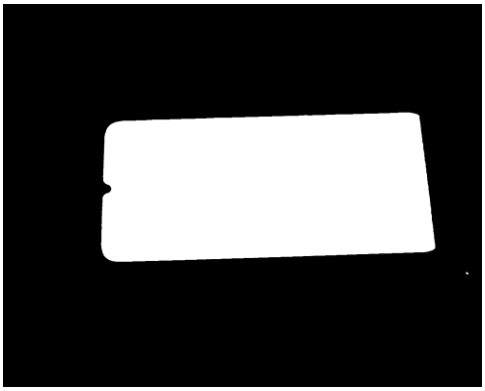


Fig. 2. Binarized mobile phone film image.

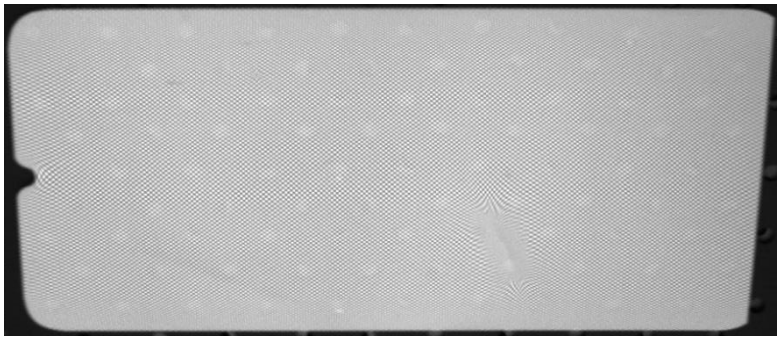


Fig. 3. Mobile phone membrane area.

3 Establish convolutional neural network

Feature pyramid is added to enhance image features while convolutional neural network is constructed. The overall neural network structure in this paper is divided into CSPDarknet53, SPP and PANet structures. These three parts are combined to finally get the defect characteristics of the mobile phone film image. The flow chart of the convolutional neural network is shown in **Figure 4**. The activation function used in this paper is MISH function, and the formula of MISH function is as follows:

$$Mish = x * tanh(ln(1 + e^{x})) \tag{3}$$

The MISH function diagram is shown in **Figure 5**:

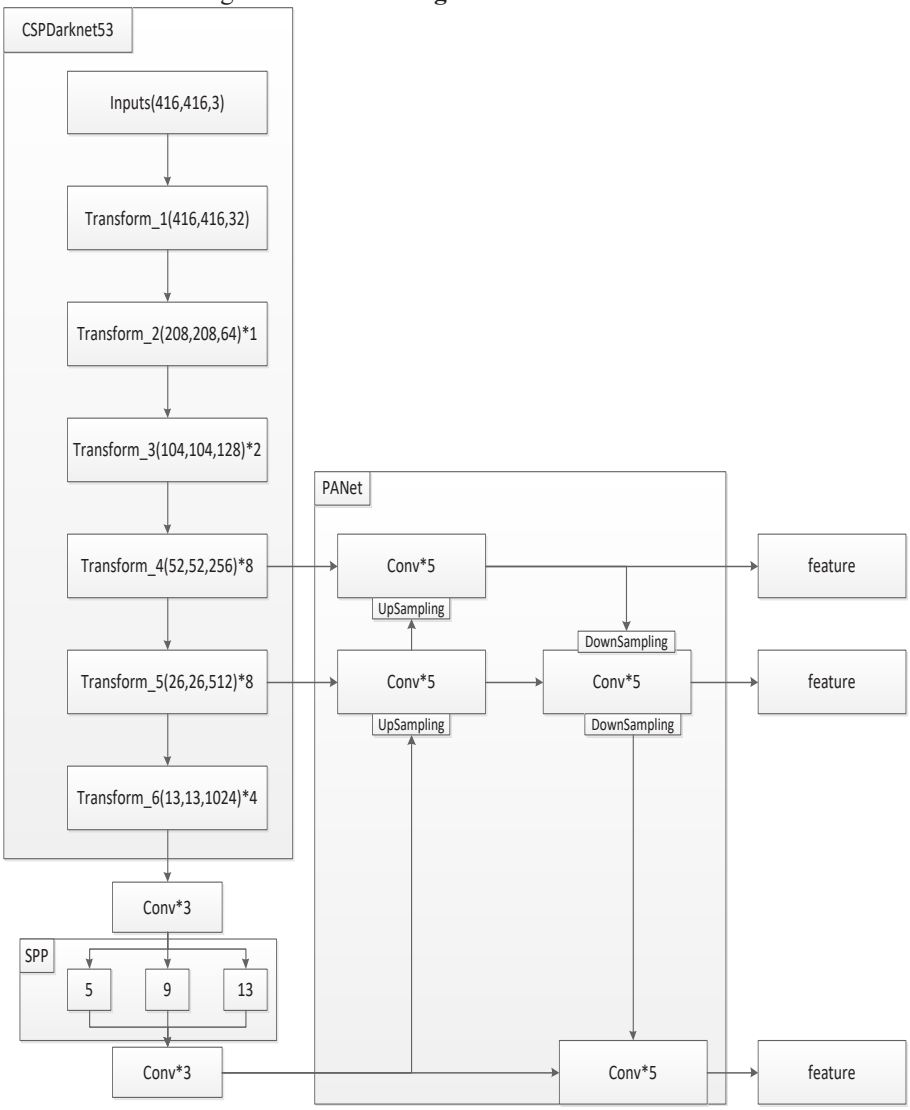


Fig. 4. Convolutional Neural Network flow chart.

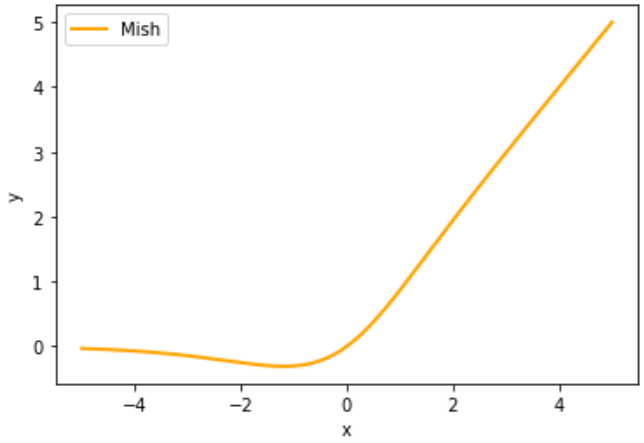


Fig. 5. Coordinate diagram of the Mish function.

In the CSPDarknet53 structure, the convolution kernel is 3*3 with a step size of 1, and the convolution is carried out successively to obtain the transformation diagram. In the SPP structure, the pooling kernel of 5*5,9*9,13*13 is used to carry out the maximum pooling operation, which can greatly increase the receiver field and separate the most significant context features. PANet structure has the meaning of repeatedly improving features. Upsampling and Downsampling are respectively adopted to realize repeated feature extraction to reduce semantic loss. Finally, the acquired feature is detected and recognized. In terms of data enhancement, this article uses four images to be spliced in a certain direction, through the image flip, color gamut change, and zoom. The great advantage of this is to enrich the background of the detection target and also speed up the calculation. And this article also uses the learning rate cosine annealing decay algorithm. The learning rate cosine annealing decay is shown as follows：

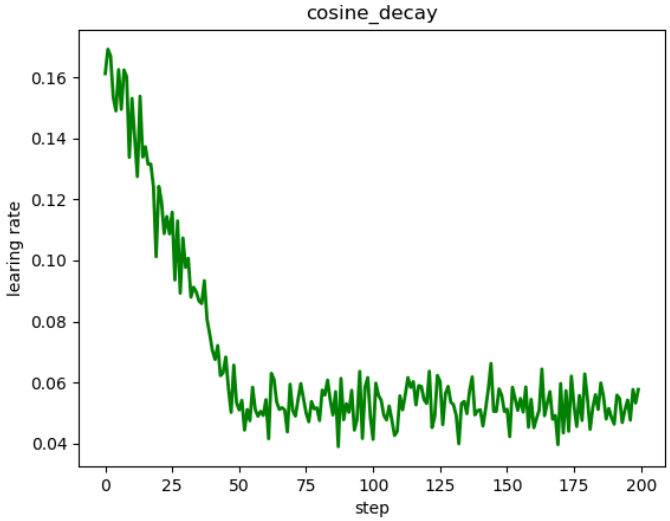


Fig. 6. Learn the cosine function of rate decay diagram.

Table 1. Convolutional Neurl Network Structure Configuration Table.

Network structure	kernel size	Stride
CSPDarkNet53	3*3	1
PANet	1*1	1
SPP	5*5or9*9or13*13	1

The learning rate will first rise and then fall. When it rises, it uses a linear rise. When it falls, the simulated cosine function decreases. Repeatedly execute it many times until the desired effect is achieved.

Table 1 can show the parameters of the network in detail.

Composition of loss Calculate parameters required by loss When calculating loss, it is actually a comparison between Y_pre and Y_true: Y_pre is the output of an image after passing through the network, which contains contents of three feature layers inside; Y_true is the offset position, length, width, and type of the grid (19, 19), (38, 38), and (76, 76) corresponding to each of its real boxes in a real image.The final output content of the network is the prediction box and its type corresponding to each grid point of the three feature layers.

Y_pre is the three feature layers respectively correspond to the position, confidence degree and type corresponding to the three prior boxes on each grid point after the picture is divided into grids of different sizes. For y1, y2, and y3 of the output, [..., : 2] refers to the offset relative to each grid point, [..., 2: 4] refers to the width and height, [..., 4: 5] refers to the confidence of the box, and [..., 5:] refers to the prediction probability of each category. For now, Y_pre is still undecoded. After decoding, it will only look like the real image. Y_true is the offset position, length, width, and type of the grid (19, 19), (38, 38), and (76, 76) corresponding to each real box in a real image. It still needs to be coded to match the structure of Y_pre Loss values need to be processed on three feature layers, taking the smallest feature layer as an example.

The calculation steps of loss are as follows:

(1). Use Y_true to extract the position of the point (m, 19, 19, 3, 1) and its corresponding category (m, 19, 19, 3, 80) of the real target in the feature layer.

(2). After the prediction value of prediction is output and processed, the predicted value of reshape is Y_pre and shape is (m, 19, 19, 3, 85). And x y, w h decoded.

(3). For each picture, calculate the IOU of all the real boxes and prediction boxes. If the coincidence degree of some prediction boxes and real boxes is greater than 0.5, it will be ignored.

(4). Calculate CIOU as regression loss. Here, only regression loss of positive samples is calculated.

(5). The loss for calculating confidence is composed of two parts. The first part is that there is actually a target, and the confidence value in the prediction result is compared with 1; In the second part, there is actually no target. In the fourth step, the value of its maximum IOU is compared with 0.

(6). Calculate the loss of the predicted category, which calculates the gap between the

predicted class and the real class where there is a target in fact.

The actual existing box, the value of the confidence in the prediction result is compared with 1; For boxes that do not actually exist, the confidence value in the prediction result is compared with 0. In this text, the boxes that do not contain the target that are ignored are removed. The actual existing boxes, and the comparison between the predicted results and the actual results.

IOU is the concept of ratio, which is insensitive to the scale of the target object. However, the regression loss optimization of commonly used BBOX and IOU optimization are not completely equivalent, and ordinary IOU cannot directly optimize the non-overlapping part. The CIOU will get the loss value.

Therefore, some people propose to directly use CIOU as regression optimization loss, and CIOU is a very excellent idea. CIOU takes into account the distance between the target and anchor, overlap rate, scale and penalty items, which makes the target box regression more stable and avoids problems such as divergence in the training process like IOU and CIOU. The penalty factor takes into account the comparison between the length and width of the predicted box and the aspect ratio of the target box. Formula is as follows:

$$\mathrm{CIOU} = \mathrm{IOU} - \frac{\rho^2(b,b^{gt})}{c^2} - \alpha v \tag{4}$$

$\rho^2(b, b^{gt})$ respectively represent the Euclidean distance of the center point of the prediction box and the real box. C represents the diagonal distance of the smallest closure region that can contain both the prediction box and the real box.

4 Experiment

In actual industrial production line, the images are fewer defects, and the image acquisition and tags require manual operation, so we choose 400 images as the data set, through the image preprocessing operations intercept membrane partial area to keep the mobile phone, then the image according to the ratio of 9:1, 360 images as the training set, the rest for the prediction set. This article uses labelimg image annotation software to mark the training set of mobile phone film separately, mark the friction marks in the mobile phone film defect as 01, frame it with a box, and save it in xml format. These images were then input into the established convolutional neural network for training. In order to evaluate the detection results, we used IOU and precision to quantitatively evaluate the performance of the two subtasks.

For the split task, IOU is defined as:

$$IOU(FM, SM) = \frac{Area((FM \cap SM)}{Area(FM \cup SM)} \tag{5}$$

Where FM represents Friction Marks, and SM represents Scratch Marks. The accuracy is used to quantitatively evaluate the performance of classification tasks, and the calculation formula is as follows:

$$Accuracy = \frac{Tp}{Tp+Fp} \tag{6}$$

From the second step, we can obtain the prediction results of the three feature layers, whose shapes are (N, 19, 19, 255), (N, 38, 38, 255), and (N, 76, 76, 255) respectively, corresponding to the positions of three prediction boxes on the grid of 19x19, 38x38, and 76x76 for each graph. However, this prediction result does not correspond to the position of the final prediction box on the picture, which can only be completed by decoding. The three feature layers divide the whole map into grids of 19x19, 38x38 and 76x76 respectively, and

each network point is responsible for the detection of a region. We know that the prediction results of feature layers correspond to the positions of three prediction boxes. We first reshape them, and the results are (N, 19, 19, 3, 85), (N, 38, 38, 3, 85), (N, 76, 76, 3, 85).The 85 in the last dimension contains 4+1+80, representing X_offset, Y_offset, H and W, confidence, and classification results, respectively.

The decoding process is to add each grid point with its corresponding X_offset and Y_offset, and the result is the center of the prediction box. Then, the length and width of the prediction box are calculated by combining the prior box with H and W. So that gives you the position of the entire prediction box.

Where TP and FP represent the number of defect areas correctly and incorrectly classified into their own categories. Of course, score sorting and non-maximum inhibition screening are also needed to pick out the score of each category that is greater than the set score. Then, non-maximum inhibition is carried out by using the position and score of the box, and the position of the defect can be framed finally.

Both of which are accurately framed, and 01 represents the friction mark defect. Through the overall measurement of the accuracy of defect location and defect recognition, it is found that the accuracy of defect prediction is 98%. And there is only one defect of friction mark in the picture, and there are many interference factors, which have a greater impact, but In general, this method has a good effect in predicting defects.

The result of the test is shown in **Figure 7**:

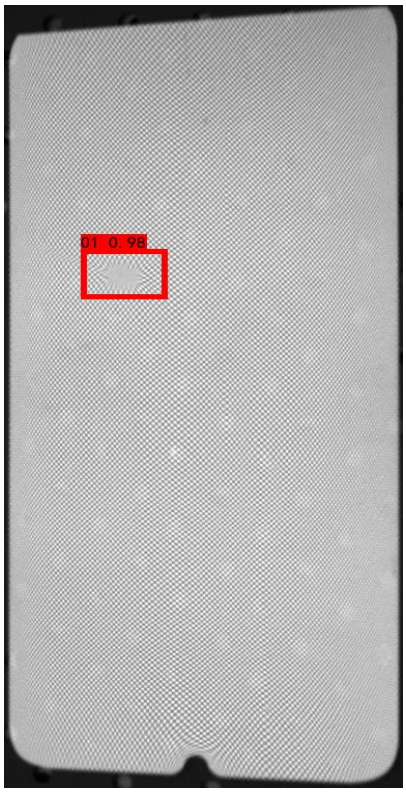


Fig. 7. Detection effect diagram.

5 Summary

In this paper, a new convolutional neural network (convolutional neural network) structure, combined with image preprocessing operation, is proposed for defect detection of mobile phone film in industry. By improving the convolutional neural network structure can exactly detect the defects on the membranes of the mobile phone and recognition, using industrial acquisition data sets, forecast data, the test results show that this method can effectively detect the defects, visualization and quantitative experiments have shown that the method can satisfy the requirement of complicated industrial environment. However, one limitation of this method is that it requires a large number of people to manually mark, which will take a lot of time and expense. In the future, we need to improve industrial equipment to automatically mark the defects of mobile phone film images and reduce a lot of unnecessary costs.

Acknowledgment

The authors acknowledge financial support from the Natural Science Foundation of Educational Department of Liaoning Province (Grant: J2020053), and Technology Innovation Fund (Grant: 2020JJ26GX029) and would like to express many thanks to the support of Dalian Key Laboratory of Smart Micro-grid and Green Recycling Industry.

References

[1] Song, K.: noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. Appl. Surf. Sci. pp. 858–864 (2013)
[2] Wu, Y.: A UAV-based visual inspection method for rail surface defects. Appl. Sci. pp. 245-248 (2018)
[3] Cen, Y. G.: Defect inspection for TFT-LCD images basedon the low-rank matrix reconstruction. Neurocomputing. pp. 1206–1215 (2015)
[4] Lei, J.: Scale insensitive and focus driven mobile screen defect detectionin industry. Neurocomputing. pp. 72–81 (2018)
[5] Li, Y.: Deformable patterned fabric defect detection with Fisher criterion-based deeplearning. IEEE Trans. Autom. Sci. Eng. pp. 1256–1264 (2017)
[6] Chondronasios, A.: Feature selection for surface defect classification of extruded aluminum profiles. Int. J. Adv. Manuf. Technol. pp. 33–41 (2016)
[7] Gibert, X.: Deep multitask learning for railway track inspection. IEEE Trans. Intell.Transp. Syst. pp. 153–164 (2017)
[8] De Araújo, S.A.: Beans quality inspection using correlation-based granulometry.Eng. Appl. Artif. Intell. pp. 84–94 (2015)
[9] Tao, X.: Weak scratch detection and defect classification methods for a large-aperture optical element. Opt. Commun. pp. 390–400 (2017)
[10] Ren, R.: A generic deep-learning-based approach for automated surface inspection.IEEE Trans. Cybern. pp. 929–940 (2018)
[11] Jian, C.: Automatic surface defect detection for mobile phone screen glass based on machine vision. Appl. Soft Comput. pp.348–358 (2017)
[12] Mak, K. L.: Fabric defect detection using morphological filters. Image Vis. Comput. pp. 1585–1592 (2009)
[13] Li, X.: Quantitative surface crack evaluation based on eddy current pulsed thermography. IEEE Sens. J. pp. 412–421 (2017)

[14] Yuan, X.: An improved Otsu method using the weighted object variance for defect detection.Appl. Surf. Sci. pp. 472–484 (2015)
[15] Win, M.: A contrast adjustment threshold ingmethod for surface defect detection based on mesoscopy. IEEE Trans. Ind. Inform. pp. 642–649 (2015)
[16] Wang, L.: Calculation of flexible printed circuit boards (FPC) global and local defect detection based on computer vision. Circ. pp. 49–54 (2016)
[17] Bai, X.: Saliency-based defect detection in industrial images by using phase spectrum. IEEE Trans. Ind. Inform. pp. 2135–2145 (2014)
[18] Borwankar, R.: An Optical Surface Inspection and Automatic Classification Technique Using the Rotated Wavelet Transform. IEEE Trans. Instrum. Meas. pp. 690–697 (2018)
[19] Susan, S.: Automatic texture defect detection using Gaussian mixture entropy modeling. Neurocomputing. pp. 232–237 (2017)
[20] Shumin, D.: Adaboost learning for fabric defect detection based on hog and SVM. In Proceedings of the International Conference on Multimedia Technology, Hangzhou, China, pp. 26–28 (2011)
[21] Ganovska, B.: Design of the model for the on-line control of the AWJ technology based on neural networks. Indian J. Eng. Mater. Sci. pp. 279–287 (2016)
[22] Natarajan, V.: Convolutional networks for voting-based anomaly classification in metal surface inspection. In Proceedings of the IEEE International Conference on Industrial Technology, Toronto, ON, Canada. Pp. 22–25 (2017)
[23] Lin, H.: Automated defect inspection of LED chip using deep convolutional neural network. J. Intell. Manuf. pp. 1–10. (2017)

Keyword extraction and ranking based on crawler and natural language processing

Enbo Zhang[1], Changmao Li[2], Li Liu(✉)
{ 248290248@qq.com[1], 2538001101@qq.com[2], link_liuli@hotmail.com (✉) }

Department of Information Science and Engineering Dalian Polytechnic University Dalian, P. R. China

Abstract. This paper adopts crawler, Hidden Markov Model, Viterbi algorithm to make a segmentation of text data on Internet, and adopt TF-IDF algorithm to extract and sort the keywords. Secondly, an experiment was carried out to extract and sort keywords from analyzing online recruitment text data. Through the experience the authors come to the conclusions: The method described in this paper can analyze the keywords of the online text and apply to various situations.

Keywords: crawler, Hidden Markov Model, Viterbi algorithm, natural language processing

1 Introduction

With the rapid development of the Internet, due to the rapid growth of the amount of data on the Internet and the reduction of the difficulty of data acquisition, people are often confused and induced by the text information on the Internet in work, study and life, and it is difficult to quickly and efficiently obtain the information they really need. In addition, due to the huge amount of data, only a small part of the data can be obtained and read by human beings, which results in a very one-sided and subjective view of things. If can analyze large amounts of data, to extract the keywords and returned to the user,this problem will be solved. Therefore, it is important to extract and sort keywords, In browsing the recruitment information. For example, when browsing the recruitment information, extracting keywords can make people quickly understand the job requirements, so as to make overall judgment.

Data acquisition is the first step,if a large amount of data is analyzed. Crawler is a common technology to get generous data. For example, there are studies[1] based on crawlers to obtain the behavior data of online forums. There are also researches[2] on price index calculation based on crawler. Processing language text data belongs to the field of natural language processing.

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307707

Table 1. Comparison between crawler and traditional data collection methods.

Comparision target	Crawler	Traditional method
Acquisition efficiency	High	Low
Energy expended	Low	High
Amount of data obtained	High	Low

For example, some studies[3] use natural language processing to identify enterprise names, and some studies[4] identify professional vocabulary.

Therefore, based on crawler and natural language processing, this paper attempts to use HMM, Viterbi algorithm and TF-IDF algorithm to extract keywords from text and sort them according to weight. By extracting keywords and weights, people can not only quickly understand the main content of the text, but also understand the problem from a macro perspective.

2 Crawler

2.1 Definition of crawlers

Crawler is a program or script that automatically grabs the World Wide Web information according to certain rules. The crawler used in this paper is focused crawler, that is, according to a certain web page analysis algorithm to filter links irrelevant to the topic, keep useful links and put them into the URL queue waiting to be crawled until a certain condition of the system is reached[5].

2.2 Advantages of crawlers

With the explosive growth of data, it is more and more difficult to obtain a large number of online target information quickly and accurately. In the past, if people want to obtain data, the usual way is to use human, artificial to search all kinds of websites on the Internet, looking for the information they want. This method usually takes a lot of time and energy. As people step into the era of big data, the amount of data people need is more and more huge, which highlights the advantages of crawler.This method usually cost a lot of time and energy. As people step into the era of big data, the amount of data people need is more and more huge, which highlights the advantages of crawler.

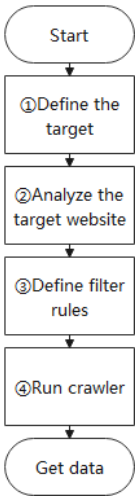


Fig. 1. Flow chart of a Crawler program

Compared with the traditional method, the main advantages of crawler are shown in Table 1, high efficiency, low energy consumption, and large amount of data acquisition. In terms of data acquisition efficiency, crawlers can obtain a large amount of information in a limited time in seconds (the specific efficiency depends on the type of target data, network speed and response speed of the target website). However, traditional methods need to search, judge and screen manually, which takes a long time and has low acquisition efficiency. In terms of energy consumption, after finishing the crawler program, let the computer run the program without supervision. The traditional method is to use manpower all the time. In terms of the amount of information, crawler can obtain a huge amount of data, but traditional methods are difficult to achieve this.

Crawler can obtain huge amount of data in a short time, which is efficient and reliable.

2.3 Crawler programming steps

Definition of regular expression: regular expression is a logic formula for string operation, that is, some specific characters defined in advance and the combination of these specific characters are used to form a "regular string", which is used to express a filtering logic for string[6].

Definition of XPath expression: XPath uses path expressions to select nodes or node sets in XML documents. Nodes are selected by following paths or steps. Reference[7] uses XPath expression to extract agent information of web pages.

As shown in Figure 1, there are four steps in using crawlers to collect data:

① Describe or define the crawling target: it is to determine the URL of the website to be crawled.

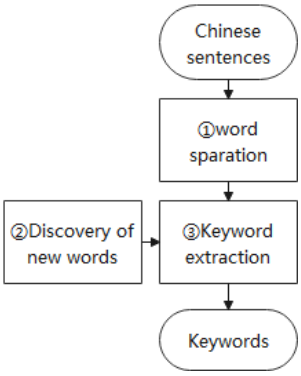


Fig. 2. Flow chart of a keyword extraction

② Write out the filter rules: generally, the filter rules are described by programming Regular Expressions or XPath Expressions.

③ Run: run the crawler program, get the data on the target website and save it according to the written filtering rules.

3 natural language processing

3.1 Definition of natural language processing

Natural language processing is an important direction in the field of computer science and artificial intelligence. It studies various theories and methods that can realize effective communication between human and computer with natural language. Basic natural language processing technologies include stop word removal, word segmentation, root extraction and part of speech tagging[8].

3.2 Keyword extraction algorithm flow

The field of natural language processing involved in this paper is syntactic semantic analysis. The main algorithm is shown in Figure 2. Firstly, the Chinese text crawled down by the crawler is segmented; secondly, new words are found to prevent word segmentation errors when new words do not exist in the dictionary appear in the text; finally, TF-IDF algorithm is used to calculate the keywords.

3.3 Process and principle of word segmentation

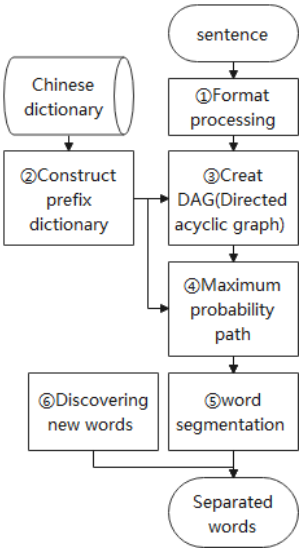


Fig. 3. Flow chart of a word segmentation

（1）Word segmentation process：

As shown in Figure 3, firstly, format the Chinese text, that is, use non Chinese characters to segment the Chinese text; secondly, construct a prefix dictionary based on the Chinese dictionary to prepare for the subsequent construction of directed acyclic graph; thirdly, construct a directed acyclic graph of sentences according to the prefix dictionary; fourthly, find the maximum probability path according to the probability in the prefix dictionary; fifthly, find the maximum probability path according to the maximum probability Probability path is used to segment sentences;

finally, new words are found, and the words that do not appear in the dictionary are recorded; the new words and the divided words are regarded as the final segmentation results, and the segmentation process ends. The Chinese dictionary and non Chinese character set files are obtained from GitHub.

（2）Principle of word segmentation

①Using the non Chinese character set obtained from GitHub, the Chinese text to be segmented can be segmented by Regular Expression or string operation.

②The prefix dictionary is constructed based on Chinese dictionary, and a trie tree is constructed to store the prefix dictionary. A prefix dictionary is a dictionary where words with the same prefix are stored together. For example: The prefixes of "Gong Ye Da Xue" are "Gong", "Gong Ye" and "Gong Ye Da". Prefix dictionaries store words with the same prefix together. The construction of prefix dictionary is to facilitate the subsequent construction of directed acyclic graph.

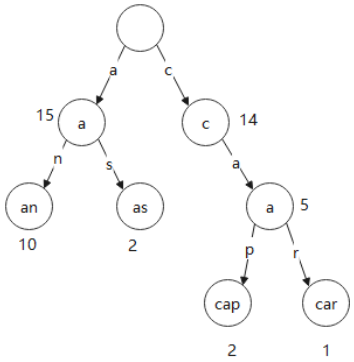


Fig. 4. Trie tree

Trie tree, also known as prefix tree[9], is an ordered tree, which stores words with the same prefix together and has the advantage of fast search speed. All descendants of a node have the same prefix, that is, the string corresponding to the node, while the root node corresponds to an empty string. The above feature is just used to store the prefix dictionary. As shown in Figure 4, key is labeled in the node and value is labeled outside the node. Each complete word corresponds to a specific integer. In this paper, the value of the node is the word frequency of the word. An empty string is saved in
the root node. The left child node is "a", and the saved value is 15, which means that the word frequency of "a" is 15 times; "the left child node of" a "is" an ", and the value is 10, which means that the word frequency of" an "is 10 times;" the right child node of "a" is "as", and the value is 2, which means that the word frequency of "as" is 2; and "an" and "as" have the same prefix "a" .

③According to the prefix dictionary, the sentence is divided into words, and the directed acyclic graph is constructed. In this paper, we use the method of calculating the optimal path through directed acyclic graph, as described in reference[10], which is one of the common applications of DAG.

④Calculate the maximum probability path: in the directed acyclic graph, each vertex is weighted, and the corresponding weight is the word frequency of the word.

Let the *route* = (w_1,w_2,w_3,...,w_n) we want to demand make $\sum weight(w_i)$ the largest

As shown in formula (1), the weight $\{R_{i\to j}\}$ of any path from W_i to W_j is equal to the sum of the weight of the path from wi to W_j.

$$\{R_{i\to j}\} = \{R_I + weight(j)\} \tag{1}$$

As shown in formula (2), the weight $\{R_{i\to k}\}$ of any path from W_i to W_k is equal to the sum of the weight of the path from W_i to W_k.

$$\{R_{i\to k}\} = \{R_I + weight(k)\} \tag{2}$$

Therefore, for nodes W_j and W_k with common precursor node W_i, it is necessary to repeatedly calculate the probability of the path to W_i, which is a repetitive subproblem.

The optimal path R_{max} of the whole graph and a terminal node W_x, for its possible existence of multiple precursors W_i, W_j, $W_{k....}$ Let the maximum probability paths to W_i, W_j and W_k be R_{maxi}, R_{maxj} and R_{maxk} respectively, then:

$$R_{max} = \max\left(R_{maxi}, R_{maxJ,}\ R_{maxK} \ldots\right) + weight\ (W_x) \tag{3}$$

Therefore, the problem can be transformed into solving R_{maxi}, R_{maxj} and R_{maxk}... The optimal solution in the substructure is a part of the global optimal solution, forming the optimal substructure problem.

It satisfies the repeated subproblem and the optimal substructure problem, so it turns into a dynamic programming problem. Similar to the method described in reference[11], the optimal path (maximum probability path) in the directed acyclic graph is calculated by dynamic programming. Finally, Chinese text segmentation is realized according to the maximum probability path.

⑤Find new words based on HMM.

HMM: Hidden Markov model, is a statistical model based on Markov hypothesis[12]. Hidden Markov model (HMM) can be described by five elements, including two state sets and three probability matrices[13]:

（1）Implicit state S

（2）Observable state O

（3）Initial state probability matrix π

（4）Implicit state transition probability matrix A

（5）Observation state transition probability matrix B

As described in reference[14], the process of Chinese word segmentation can be transformed into marking the position of each word in a string of Chinese characters sequence, which is brought in as a parameter of the model. In this paper, the position B (begin), M (middle), E (end) and S (single) of the word in the word are regarded as the hidden state, and the word is the observed state. and the remaining parameters in the dictionary file are used, which is a standard decoding problem. According to the probability, Viterbi algorithm is used to solve the maximum possible hidden state. We can find new words. In this paper, the trained HMM is used.

In fact, the essence of Viterbi algorithm is to use dynamic programming to solve HMM prediction problem[15], that is, to use dynamic programming to find the

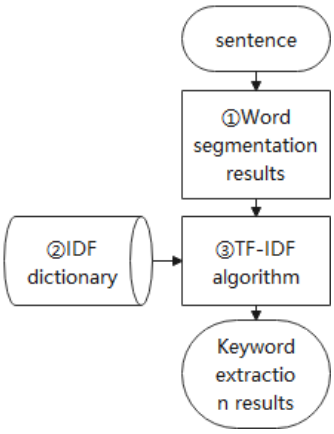


Fig. 5. Flow chart of a keyword extraction (TF-IDF part)

maximum probability path (optimal path). In this case, a path corresponds to a sequence of states.

3.4 Keyword extraction and its principle

Keywords are mainly obtained by TF-IDF algorithm. As shown in Figure 5, after the sentence is divided into words, each divided word is calculated based on the IDF dictionary by using TF-IDF algorithm, and the result is calculated to get the $TF\text{-}IDF_{ij}$ The larger the value of ij, the more important the word is relative to the text.

Definition of TF-IDF algorithm: TF-IDF (term frequency – inverse document frequency) is a common weighting technology for information retrieval and data mining. TF is term frequency and IDF is inverse document frequency. TF-IDF is a statistical method to evaluate the importance of a word to a file set or one of the files in a corpus. The importance of a word increases with the frequency of its appearance in the document, but decreases inversely with the frequency of its appearance in the corpus[16].

The principle of TF-IDF algorithm is as follows

$$\mathrm{TF} - IDF_{ij} = tf_{ij} \times idf_i \tag{4}$$

$$tf_{ij} = \frac{n_{ij}}{\sum_k n_{kj}} \tag{5}$$

As shown in formula 4, the importance of a word to a file set $TF\text{-}IDF_{ij}$ is equal to the product of word frequency *TFij* and inverse text frequency index idf_i.

As shown in formula 5, word frequency tf_{ij} is equal to the number of times that *i* word appears in *j* document nij divided by the sum of the number of times that all words appear in *j* document $\sum_k n_{kj}$. The inverse text frequency index *idf* in this paper is directly read into the dictionary.

4 Keywrd sorting

4.1 All Chinese Texts

When the text is all Chinese, according to the weight of each word calculated by TF-IDF algorithm after word segmentation, all the words are arranged in descending order, and a threshold *U* is specified. The words whose weight is greater than the threshold *U* are the keywords of the text.

4.2 The main body is Chinese text mixed with English

When the text is a mixture of Chinese and English (such as the recruitment information on the recruitment website), most English words are separated by spaces except some special words (such as New York), so the difficulty of word segmentation is relatively simple. First of all, we need to screen out and save the English text separately with regular expression; second, we need to get the information like "new" on the Internet The third step is to filter the mixed English text, screen out and save the special English words separately; the fourth step is to divide the English words by spaces; finally, the selected special English word set and the separated ordinary English word set are combined to complete the English text segmentation.

After word segmentation in the English part, to filter the second time, we need to delete "an", "and" and other meaningless function words from the English word set, and finally calculate the frequency of each English word, that is, the frequency of each English word.

In the Chinese text part, the word frequency is calculated by the method described in this paper.

According to word frequency, Chinese and English words are sorted in descending order. Take the first *N* words as the keywords of the mixed language text.

5 Experiment

Taking the information text of Chinese recruitment website as an example, based on Python and MySQL, this paper uses the method described in this paper to analyze the keywords of "Python development engineer" on a certain website.

5.1 Crawler section

Fig. 6. The composition rule of URL

First, determine the URL of the website to be crawled and its naming rules. The data to be crawled in this paper is the detailed recruitment data of "Python development engineer" in Dalian area on a recruitment website. After manual login, the naming rules of URL are analyzed, and the specific format of URL naming is shown in Figure 6, which is the combination of "main URL", "region number", "position number" and "number of pages". Through manual search and comparison, you can get the region number of Dalian and the position number of Python development engineer, so you can make your own URL to crawl.

The second step is to manually log on the target web page, watch the front-end code of the web page, and formulate the crawling strategy according to the target data. For example, observe the format of data in the front-end code, storage rules, etc. For example, the recruitment data to be captured in this paper is stored in the P tag under div with multiple category attributes of position number.

The third step is to start to write the crawler program and write the filtering rules. In order to improve the efficiency of filtering, this experiment uses Regular Expression and XPath Expression according to the situation. In this step, we use XPath Expression and Regular Expression to filter out the text data, and then use regular expression to clean the filtered text data again. This paper mainly crawls the text data for each recruitment data.

Finally, we start to run the crawler program and save the text data in the database to prepare for the natural language processing steps.

5.2 Natural language processing

First of all, the crawler program crawled down the text format processing, to prepare for word segmentation. In this paper, the target site in Dalian area accurate search "Python development engineer" recruitment data, a total of 28, this article put 28 data into a text, start word segmentation.

The recruitment data is a mixture of Chinese and English text, so the flow chart is shown in Figure 7. First, program regular Expressions to separate the Chinese and English in the text. Second, segment words separately.

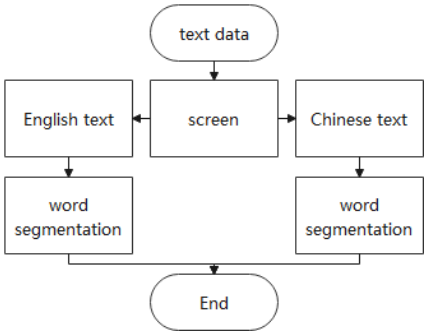


Fig. 7. Flow chart of Chinese English mixed text

Key	Value	
0:	[0,2,5]	“有”，“有经验”，“有经验者”
1:	[1,2]	“经”，“经验”
2:	[2]	“验”
3:	[3]	“者”
4:	[4,5]	“优”，“优先”
5:	[5]	“先”

Fig. 8. Key-Value diagram

In the process of English text segmentation, the special English words are separated based on the special English word dictionary, and then the space and punctuation are used as the segmentation position. After the word segmentation, it is combined with the separated Special English words to form a new array. And calculate the frequency of each English word.In the process of Chinese text segmentation, we first download the non Chinese character set from GitHub and use regular expressions to segment Chinese text.

Second, we need to build a trie tree to process the downloaded Chinese dictionary into prefix dictionary.

In the third step, according to the prefix dictionary, the result of the first step is further divided into words, and the directed acyclic graph is constructed.

When constructing a directed acyclic graph, we need to create a python list or dictionary to store the mapping relationship between vertices. For example, The sentence “有经验者优先。” is segmented according to the prefix dictionary, it can be divided into eight words: “有” “经” “验” “者” “优” “先” “有经验” “经验” “优先” As shown in Figure 8, this paper marks each word with the position in the text, the position of the first word is set to 0, and the position of the last word is set to 5, which is used as the key; the list formed by the end position of the divided words is used as the value. As shown in Figure 8, there are six words “有经验者优先”, so the key label is 0-5; Because there are three words related to “有” after word segmentation: “有”, “有经验” and “有经验者”, the key "0" Representing “有”, "the key" 2 "Representing “经” and the key " 3 "Representing “者” are written into the value of key0, and the value is [0,2,5]. The rest of the key value pairs are the same.

Construct a directed acyclic graph with the above words and words, as shown in Figure 9. The directed acyclic graph has six vertices, representing six words in the text respectively. The serial number above the vertex represents key, and the vertices are connected according to the order of appearance in the text. After word segmentation, , “有经验”、“经验” “有经验者” and “优先” are divided together, so that “有” points to “验” and “者”, “经” points to “验” and “优” points to “先”, and the representative is divided into one word.

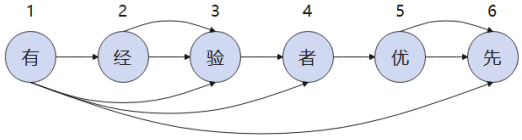


Fig. 9. DAG（Directed acyclic graph）

In the fourth step, by writing a bottom-up dynamic programming algorithm, the maximum probability path of the directed acyclic graph is calculated, and the Chinese text segmentation is realized according to the maximum probability path.

In the fifth step, new words are recognized based on HMM and Viterbi algorithm, and the results are added to the word segmentation results in the fourth step.

Finally, the TF-IDF algorithm is used to calculate the Chinese keywords, and the Chinese and English keywords and the word frequency of each word are saved in the database.

5.3 Keyword sorting and display

The Chinese and English keywords in the database are sorted in descending order according to word frequency.

In order to better show the conclusions (keywords and weight), data visualization technology is used to display the results in the word cloud and figures.

The final result is shown in Figure 10, keywords are displayed in the figure of word cloud. All the words in the figure are the key words analyzed by the system. The larger the font of the word, the greater the weight of the word. That is to say, the more important the word is to the position of "Python development engineer". For example, the biggest word is "Python" and the second is "MySQL". It is proved that the two most important skills for the position of "Python development engineer" are "Python" and "MySQL".

Font color is generated randomly, just to distinguish keywords, does not have special meaning.

Further data are shown in figures 11, 12 and 13. As shown in Figure 11, the x-axis represents all the analyzed keywords, and the y-axis represents the word frequency of the keyword, that is, the number of times the keyword is required to be mastered by the recruiter.

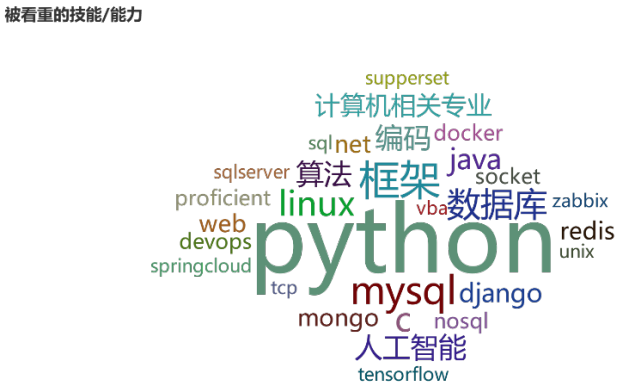


Fig. 10. The word cloud of keywords

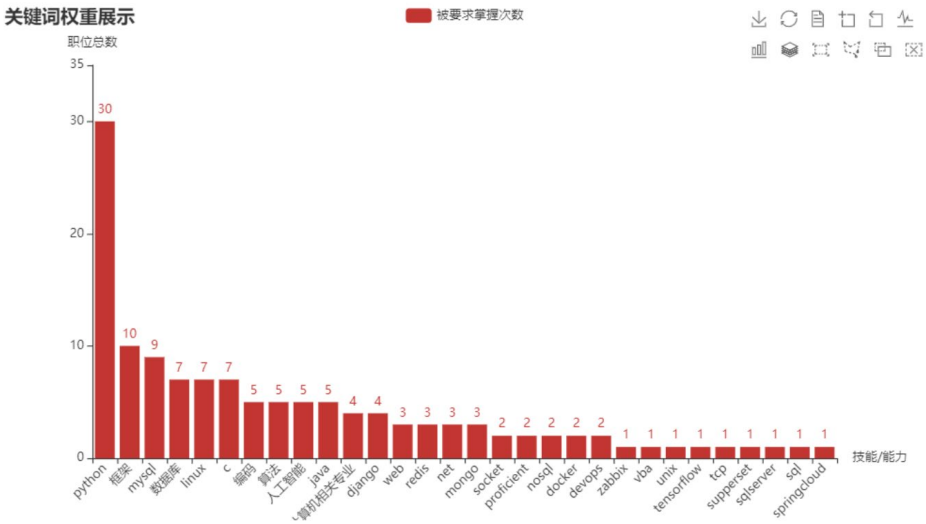


Fig. 11. Keywords overview

As shown in Figure 12, each pie chart represents a keyword, and the percentage at the center of the pie chart represents the probability of the keyword appearing in the recruitment requirements.

Figure 13 shows the degree to which keywords are required to be mastered by the recruiter. The degree is divided into four categories: "expert","Familiar", "know" and "master". As shown in Figure 13, the x-axis represents the keyword, and the y-axis represents the probability that the keyword is required by the recruiter in the four standards.

Fig. 12. Keywords pie chart

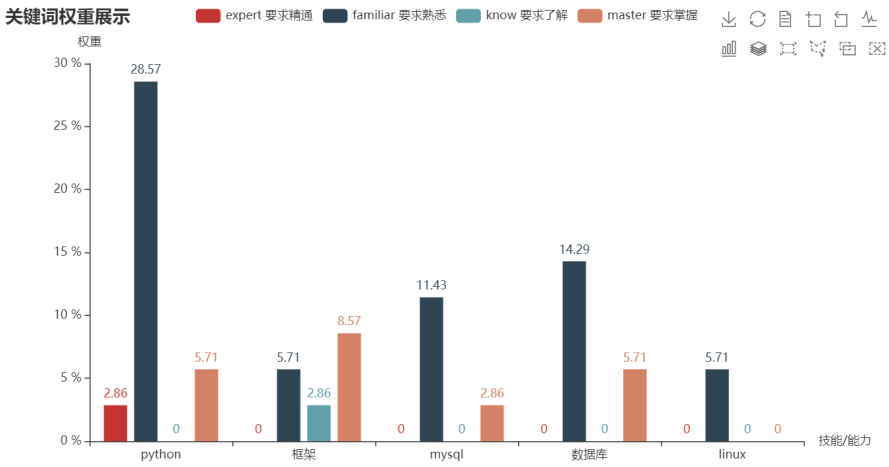


Fig. 13. Figure showing keywords in different degrees

6 Conclusion

In the era of big data, with the explosive growth of data volume and the reduction of data acquisition difficulty, people are often induced and confused by the information on the Internet, and it is difficult to obtain the real valuable information. If we use the keyword capture and sorting method described in this paper, we can produce different effects according to the size of the data and the actual application scenarios. When the amount of data is small, the key words can help users understand the central idea of multiple texts, identify the main content of the text, and improve the efficiency of learning and work; when the amount of data is large, taking the experiment in this paper as an example, we can analyze the key words through recruitment data, and summarize the requirements of most companies for a certain position or major General requirements and specific skills expectations can make users more aware of the current enterprise recruitment trends, industry hot spots and other information, so that users will not be limited to limited recruitment information, but based on the huge amount of data, have a macro understanding of the specific skills and job requirements of recruitment, so that users can understand what skills are more valued by recruiters, so as to help them Users make choices on the way to study and work. And by modifying the crawler program and analyzing the recruitment data of different regions, we can get the industry characteristics of a region, which is more region specific.

The method described in this paper is not only applicable to the above examples, but also applicable to other situations, and can be widely used in a variety of situations.

Acknowledgment

The authors acknowledge financia support from College Student Innovation and Entrepreneurship Project.

References

[1] Zheng JJ, CHENG Y, MA G, et al.: Research on evolution of public environmental protection behavior based on data crawler and dynamic social network. System Engineering-Theory ＆Practice. Vol. 2, pp. 219-229 (2020)

[2] LIU YB, ZHAO ZD, LIU HB.: A model of compiling price index based on the “Web Scraping” technology. Statistical Research. Vol. 31, pp. 74-80 (2014)

[3] HUANG GX, ZHU SX, WANG XH, et al.: Natural language processing and machine learning-based suspected soil. Chinese Journal of Environmental Engineering. Vol.14, pp. 3234-3242 (2020)

[4] ZHU TT, DU YF, LI RF, et al.: An unsupervised approach to recognizing new words in power domain. Electric Power Engineering Technology. Vol. 39, pp. 159-165 (2020)

[5] ZHOU LZ, LIN L.: Survey on the research of focused craw ling technique. ComputerApplications. Vol. 25, pp. 1965-1969 (2005)

[6] JeffreyE FF.: Introduction to regular expressions. Vol. 1, pp. 4-5. Mastering Regular Expressions, CN(2009)

[7] Antonov E, Lopatina E, Ionkina K, et al.: Agent data merging. Procedia Computer Science. Vol. 169, pp. 473-478 (2020)

[8] WANG CH, ZHANG M, MA SP.: A survey of natural language processing in information retrieval. JOURNAL OF CHINESE INFORMATION PROCESSING. Vol. 21, pp. 35-45 (2007)

[9] ZHANG M , CHEN W , WANG Q.: Data structure and algorithm. China Machine Press, CN(2010)

[10] Borenstein D.: A directed acyclic graph representation of routing manufacturing flexibility. European Journal of Operational Research. pp. 78-93 (2000)

[11] LV YJ, ZHAO TJ, YANG MY, et al.: Leveled unknown Chinese words Resolution by dynamic programming. JOURNAL OF CHINESE INFORMATION PROCESSING. Vol. 15 (2001)

[12] HU CJ, HAN ZQ.: Application study of Hidden Markov model based part-of-speech tagging. Computer Engineering and Applications. Vol. 6, pp. 62-64 (2002)

[13] Rose R C.: A hidden Markov model based keyword recognition system. Proc of Icassp Albuquerque Nm Usa. Vol. 1, pp. 129-132 (1990)

[14] WANG WF, XU HJ, YANG WZ, WU XL.: Review of Chinese word segmentation algorithms. Group Technology ＆ Production Modernization. Vol. 35, pp.1-8 (2018)

[15] CHEN K T.: Veterbi decoding method for convolutionally encoded signal. (2008)

[16] SHI CY, XU CJ, YANG XJ.: Study of TFIDF algorithm. Journal of ComputerApplications. Vol. 29, pp. 167-180 (2009)

Learning Deep Representation of The Emotion Speech Signal

Junyi Duan[1], Zheng Song[2], Jianfeng Zhao[3*]
{duanjy@chinatowercom.cn[1], 1031761899@qq.com[2], nmgzjf@buaa.edu.cn[3*]}

Inner Mongolia Branch of China Tower Corporation Limited, 010000, Hohhot, Inner Mongolia, China [1], School of Electronic Information Engineering, Inner Mongolia University, 010021, Hohhot, Inner Mongolia, China[2], Hangzhou Innovation Institute, Beihang University, 310000, Hangzhou, China[3*]

Abstract. This paper aims at learning deep representation of emotion speech signal directly from raw audio clip using a 1D convolutional encoder, and reconstructing the audio signal using a 1D deconvolutional decoder. The learned deep features which contain the essential information of the signal, should be robust enough to reconstruct the speech signal. The location of the maximal value in the pooled receptive field of the max pooling layer is passed to the corresponding unpooling layer for reconstructing the audio clip. Residual learning is adopted to ease the training process. A dual training mechanism was developed to enable the decoder to reconstruct the speech signal from the deep representation more accurate. After completing the training of the convolutional-deconvolutional encoder-decoder as a whole, the decoder with transferred features was trained again. Experiments conducted on Berlin EmoDB and SAVEE database achieve excellent performances.

Keywords: Deep representation, emotion speech signal, 1D deconvolutional decoder, convolutional-deconvolutional encoder-decoder

1 Introduction

With the development of deep learning, there emerge a trend towards deriving a deep representation of the input signal directly from raw data. The learned deep representation of the unprocessed signal can lead to better performance in many practical tasks. The basic approach of learning deep representation is to construct a deep architecture to extract the high-level features, which can represent the input signal. So, a series of new neural network architectures have been introduced to learn deep representation, such as autoencoder network, convolutional neural network (CNN), or recurrent neural network (RNN) [1], [2], [3].

---

[1] Junyi Duan(1976~), male, engineer, Inner Mongolia Branch of China Tower Corporation Limited.

[2] Zheng Song(1995~), female, Master candidate, School of Electronic Information Engineering, Inner Mongolia University.

[3] Jianfeng Zhao(1978~), corresponding author, male, associate professor, Hangzhou Innovation Institute, Beihang University.

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307539

Due to the high learning capacity, deep representation learning has made a great improvement in several fields, especially in learning hieratical image representation. Several very deep autoencoders have been proposed to learn the deep representation of the image, and used for many tasks [4], [5], [6], [7]. The deep convolutional encoder-decoders with deep residual learning blocks (RLBs) [8], [9], [10], which allow skip connections, have also achieved great success and obtained better performance in learning deep representation from images.

As regards the research fields relevant to speech, there are a lot of researchers have devoted to complete the recognition or prediction task using different deep networks. The autoencoder, CNN, and RNN have reformed the speaker, speech, and speech emotion recognition by learning deep features from different hand-crafted features, especially from mel-frequency cepstral coefficients (MFCCs) and short-time Fourier transform (STFT) [11], [12], [13], [14]. The deep features learned by the deep neural network (DNN) can form a hierarchical representation of raw data, and perform better in some experiments, such as emotion prediction [15]. The learning of the deep representation, which composes of multiple linear and non-linear transformations, from raw audio clip also has been brought into focus.

Inspired by the success of image restoration using deep residual learning, we designed a 1D deep convolutional-deconvolutional encoder-decoder with RLBs to learn the deep representation from raw audio clip using convolution and pooling, and reconstruct the audio signal from deep representation using deconvolution and unpooling. To unpooling properly, the exact location of the maximal value [16], which is lost during the max pool, is transferred from the layer of encoder to the corresponding layer of decoder. Residual learning block (RLB) is used to ease the training process. A dual training mechanism was developed to enable the decoder to reconstruct the speech signal from the deep representation more accurate. After the completion of the training of encoder-decoder as a whole, the decoder with transferred features was trained again. The experimental results show that the dual training mechanism can make the deconvolutional decoder reconstruct the emotion speech signal from the deep representation effectively with the help of the dual training mechanism.

Our key contributions are as follows: 1) we complete the design of 1D deep convolutional-deconvolutional encoder-decoder to learn deep representation from speech signal using convolution and pooling, and restore the signal from the learned deep representation using deconvolution, and unpooling; 2) we adopt a different RLBs in encoder and decoder to ease the training of the very deep encoder-decoder; 3) we develop a dual training mechanism, which enable the encoder-decoder perform better. The dual training mechanism can enable the decoder to restore the speech signal from the deep representation more accurate. To our knowledge, it is the first framework that applies such a convolutional-deconvolutional encoder-decoder to learn deep representation from speech signal, and the results show that the designed architecture can learn the deep representation and reconstruct the speech signal with high fidelity Successfully.

2 Related Work

With the development of deep learning, deep representation learning has made great strides over the last several years, especially in image processing, such as image restoration. Hinton and Krizhevsky designed a very deep autoencoder to learn the deep representation of the image, which was used to implement the application of content-based image retrieval [4]. Pascal Vincent *et al*. proposed a stacked denoising autoencoder to learn useful representations

with a local denoising criterion, which could learn Gabor-like edge detectors from natural image patches and larger stroke detectors from digit images contrary to ordinary autoencoders [5]. Junbo Zhao *et al*. presented a stacked what-where auto-encoder, which produced the reconstruction of the image by integrating discriminative and generative pathways and providing a unified approach without relying on sampling during training [6]. Yiyi Liao *et al*. proposed a graph regularized auto-encoder, which preserved the local connectivity from the original image space to the representation space and provided explicit encoding model for fast inference and powerful expressive capacity for complex modelling [7]. When residual learning framework was present by Kaiming He *et al*. [8], [9] the deep convolutional encoder-decoder designed by Xiao-Jiao Mao *et al*. [10] with symmetric skip connections, which was composed of multiple layers of convolution and deconvolution operators, achieved better performance in recovering the original image.

Learning deep features from raw speech waveform to implement speech-related tasks has also attracted increasing attention of many researchers. In addition to the above deep representation learning approaches on image-related tasks, there were many other encouraging achievements [17], [18], [19], [20], [21] that also motivated us to achieve better performance in learning deep representation from the raw speech signal. Jaitly and Hinton proposed learning an intermediate representation by training a restricted Boltzmann machine (RBN) directly on the speech time signal for automatic speech recognition (ASR). Experiments on the TIMIT phoneme recognition task achieved state-of-the-art results at the time [22]. Bhargava and Rose presented a stacked bottleneck deep neural network, which was trained on windowed speech waveforms. The experimental results were only worse than corresponding MFCCs on the same architecture [23]. Palaz *et al*. proposed a CNN architecture, which was directly trained on the speech signal to estimate phoneme class conditional probabilities. The experimental results showed that the architecture had a more robust performance in noisy conditions [24], [25]. Sander Dieleman and Benjamin Schrauwen designed a CNN architecture to learn features directly from raw audio signals, which was successfully applied to an automatic tagging task [26].

3 The Proposed Approach

To learn compact and robust features, a deep convolutional-deconvolutional encoder-decoder was designed according to the requirements of the experiments and the hardware resources. Convolution and pooling are exactly proved to outperform on learning features, which are inspired from biological visual cortex and designed to emulate the behaviour of the visual system [29]. With the distinguishing excellences, spatially-local connectivity and shared weights [30], convolution and pooling perform the function of the learning filters in the encoder. In order to reconstruct the speech signal, deconvolution and unpooling were adopted to build the decoder for symmetry.

3.1 1D Convolution and Deconvolution

The 1D convolution layer plays the role of feature extractors and learns the local features which restrict the receptive fields of the hidden layers to be local (see Figure 1a). The convolution kernel, which has a small receptive field, is convolved across the input. The convolution will produce a features map by computing the dot product between the entries of the filter and the input. It is the core building block of a CNN.

If 1D convolution layer takes as input a speech signal $x(n)$, the feature map $z_k(n)$ can be obtained by convolving the signal $x(n)$ with the k-th convolution kernel $c_k(n)$ of size l.

$$z_k(n) = x(n) * c_k(n) = \sum_{m=-l}^{l} x(m) \cdot c_k(n-m) \tag{1}$$

To restore the speech signal x, the deconvolution, which reverses the effects of convolution, needs to be applied. The restored speech signal $\hat{x}(n)$ can be represent as a linear sum of feature map $z_k(n)$ convolved with filter $c_k(n)$ [37].

$$\hat{x}(n) = \sum_{k=1}^{K} z_k(n) * c_k(n) \tag{2}$$

Where K is the number of the filters.

The 1D deconvolution performs a transformation going in the opposite direction of a normal 1D convolution. To complete the transposed convolutions, the connectivity pattern of the transposed convolution operation needs to be compatible with the convolution. The input of the deconvolution is the output of the convolution, while the output of the deconvolution is the input of the convolution.

3.2 1D pooling and Unpooling

The 1D pooling layer, which makes the features robust against noise and distortion [31], performs the non-linear downsampling function and reduces the resolution of the features. The most common used non-linear functions to implement pooling is max pooling. Max pooling will partition the input into a set of non-overlapping regions and output the maximal value of each sub-region. But the max pooling will lose the exact location of the maximal value in the pooled receptive field, which is very useful for the reconstruction of the input. The features produced by max-pooling layer can be expressed as:

$$z_k^o(n) = \max_{\forall p \in \Omega_k} z_p^{o-1}(n) \tag{3}$$

Where $z_k^o(n)$ and $z_p^{o-1}(n)$ represent the maximum feature of the k-th pooling region at the o-th layer and the input features at the $(o-1)$-th layer, Ω_k represents the pooling region with index k.

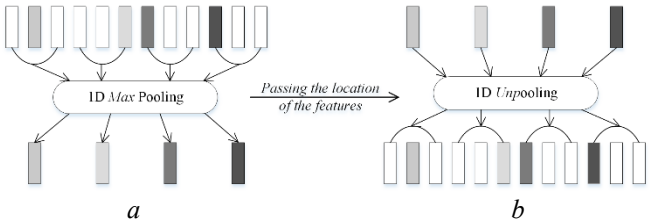


Fig. 1. Illustration of 1D max pooling and unpooling

The most common used inverse operation of 1D pooling is upsampling, which will repeat the data along the time axis (see Figure 1b). However, the upsampling function cannot correctly place the features in decoder, which will become noise and affect the quality of the signal reconstruction.

To reconstruct the signal with much higher fidelity, Goroshin *et al.* introduced a soft version of *max* and *argmax* pooling operators to hand the exact location of the maximal value from a pooling layer to the corresponding unpooling layer [38]. The *max*-pooling operator will indicate the position of the maximum value in each sub-region. Whereas the *argmax*-pooling operator will return the position of the maximum value defined over the sub-region.

Inspired by the idea of Goroshin *et al.*, we achieve this goal in our architecture using a simple method. We use hard version pooling to output the maximal value of each pooling region in encoder, and upsampling to repeat the maximal value in each unpooling region in decoder. The location of the maximal value in each pooling region in encoder will be handed to the corresponding unpooling region in decoder by a mask.

$$\underset{\Omega_k}{mask} = \partial \sum z_k^o(n) / \partial z_k^o(n) \tag{4}$$

Where $z_k^o(n)$ denotes the features of the k-th pooling region at the o-th layer, n stands for temporal location, Ω_k represents for the k-th pooling region.

The mask contains switches indicating. The value of the maximum value location of the pooling region is 1 in the mask, and the value of other location is 0. Therefore, the reconstructed signal in a unpooling region can be represented as:

$$\hat{\hat{z}}_k^o(n) = \underset{\Omega_k}{mask} \cdot \hat{z}_k^o(n) \tag{5}$$

Where $\hat{z}_k^o(n)$ denotes the upsampled feature of the k-th unpooling region at the o-th layer, $\hat{\hat{z}}_k^o(n)$ denotes the unpooled feature of the k-th unpooling region at the o-th layer, n stands for temporal location, Ω_k represents for the k-th unpooling region.

3.3 Residual Learning Block

Deep residual learning allows skip connections, which enable the network to be as linear or non-linear as the data sees fit. It was first introduced by Kaiming He *et al.* [8], [9] for image recognition. Residual learning block always consists of several layers, and can ease the training of deep neural networks (see Figure 2). The skip connections in a RLB are connections which skips one or several layers and connects to the next available layer.

Different RLBs were used in encoder and decoder to ease the training of the designed very deep encoder-decoder. The RLB adopted in encoder contains two convolution layers (see Figure 2a), and the RLB used in decoder includes two deconvolution layers (see Figure 2b). RLB can be expressed in a form:

$$x_{l+1} = F(x_l) + x_l \tag{6}$$

Where x_l and x_{l+1} are the inputted features and outputted features, and F is the residual function which is determined by the building layers. The activation function (ELU) acts as a pre-activation of the convolution and deconvolution layers, in contrast with the conventional method of post-activation.

3.4 Architecture

3.4.1 Architecture Design

The designed architecture, which is illustrated in Figure 3, consists of two parts: the encoder and decoder. The encoder is formed by a stack of convolution and pooling layers,

because of the symmetry requirement between the encoder and decoder, the decoder is built by the same number of deconvolution and unpooling layers.

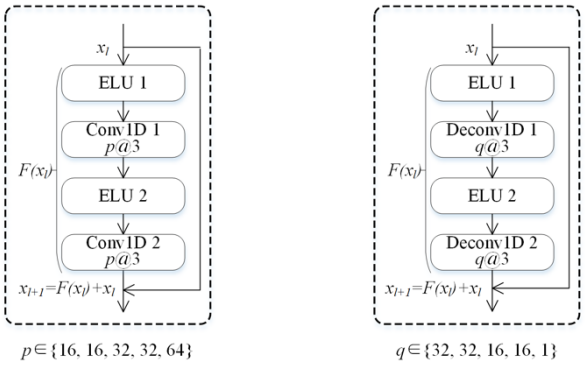


Fig. 2. Illustration of residual learning blocks

The encoder consists of five RLBs, five convolution and pooling layers. The RLB in the encoder is constructed by stacking two ELU and convolution layers with a kernel size of 3 (see Figure 2a). In the same RLB, the convolution will produce the same number of feature maps. The number of the feature maps in each RLB is 16, 16, 32, 32 and 64. The 1D convolution is applied with a kernel size of 3 on the audio data. The number output of filters in each convolution layer is 16, 16, 32, 32 and 64. That is to say, there are fifteen convolution layers in the designed encoder in total. In each pooling layer, the 2 downsampling is applied on each feature map in each RLB. This leads to the same number of feature maps with reduced temporal resolution. When the audio clip, a 64000-dimensional vector, is entered into the encoder, the dimension will be reduced to 2000.

The corresponding decoder of the designed encoder consists of five RLBs, five deconvolution and unpooling layers. The RLB in the decoder is built by linking two ELU and deconvolution layers with a kernel size of 3 (see Figure 2b). In the same RLB, the deconvolution will produce the same number of feature maps. The number of the feature maps in each RLB is 32, 32, 16, 16 and 1. The 1D deconvolution is applied with a kernel size of 3 on the audio data, and will output 32, 32, 16, 16 and 1 feature maps in each deconvolution layer. So, there are fifteen deconvolution layers in the decoder in total. In each unpooling layer, the 2 upsampling is applied on each feature map of each RLB. This also will produce the same number of feature maps with increased temporal resolution. When the deep representation, a 2000-dimensional vector, is entered into the decoder, the audio clip will be rebuilt.

3.4.2 Optimization Algorithm

The aim of hyperparameter optimization is choosing a set of hyperparameters for a deep architecture, usually with the goal of optimizing the deep network's performance on an independent data set. An optimization algorithm will find the “best fitting” model by solving a convex minimization or concave maximization problem [32].

Adaptive moment estimation (Adam) optimizer is adopted to train the network for faster backpropagation and smoother convergence [33]. We use the Adam with the learning rate of $\gamma = 0.001$ and exponential decay rates of $\beta_1 = 0.9, \beta_2 = 0.999$.

The objective function is mean squared error (MSE), which will minimize the loss between the recovered speech signal and the ground truth during the training of the network.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2 \tag{7}$$

Where $\hat{Y}_i$ is the recovered signal of the inputs, and Y_i is the speech signal corresponding to the inputs. After the network has been trained with optimized parameters, the best model will be returned.

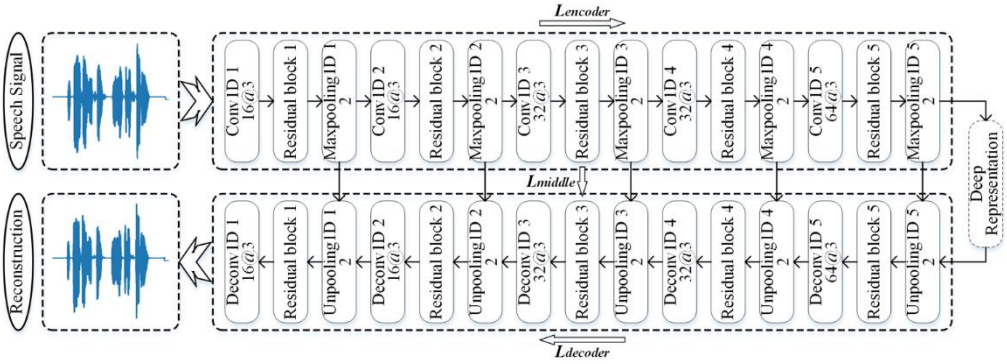


Fig. 3. The overall architecture of the designed 1D deep convolutional-deconvolutional encoder-decoder. The upper half of the architecture is an encoder, the bottom half of the architecture is a decoder. The exact location of the maximal value in the pooled receptive field is handed from the layer of encoder to the corresponding layer of decoder.

3.4.3 Regularization Methods

Overfitting is a common problem which arises with the deep architectures. Overfitting always occurs when the deep architecture is very complex, especially when the deep network has too many parameters relative to the number of observations. If a deep network has been overfit, it will have poor predictive performance. So there are many regularization methods have been proposed to overcome overfitting, such as weight decay, dropout, early stopping [34], [35], [36]. This paper used weight decay and early stopping to combat overfitting.

4. Experiments and Results

The raw data contain the whole information, so abundant features can be obtained from the audio clips. So two emotion speech databases, which contain clear features, were selected to accomplish the experiments. The audio clips were preprocessed to meet the requirements of the experiments and balance the hardware limitation against time-consuming.

4.1 Data Preprocessing

We select two emotional databases, the Berlin emotional database (Berlin EmoDB) and the Surrey Audio-Visual Expressed Emotion (SAVEE) database, both of which possess a high degree of naturalness to test and verify the designed deep convolutional-deconvolutional encoder-decoder. The two selected databases are all acted speech database, in which pre-determined sentences with the required emotions are expressed by the invited actors.

4.1.1 Databases Introduction

Berlin EmoDB - The Berlin EmoDB which was recorded in 2005 provides labeled audio clips and some analysis results. It contains seven emotions, and each emotion comprises nearly the same number of utterances to properly evaluate the classification accuracy. These acted emotional utterances were spoken by ten professional actors in an angry, boredom, disgust, fear, happy, neutral and sadness way. Five actresses and five actors feigned five short and five longer German utterances with the simulated emotions. There are 535 sentences of the utterances which come from everyday communication and can be interpretable in all applied emotions. The utterances were recorded in an anechoic chamber and were evaluated in a perception test regarding the recognisability of emotions and their naturalness [27].

SAVEE Database - The SAVEE database was released in April 2011 and provides the audio, visual, audio-visual data and other analysis data. It consists of 480 British English utterances which expressed by four native English male actors in 7 different emotions. These emotions have been described psychologically in discrete categories: anger, disgust, fear, happiness, sadness, surprise and neutral. The sentences of the emotional utterances were picked from the standard TIMIT corpus and phonetically-balanced for every emotion. The utterances were recorded, processed and labeled in a visual media lab. And ten subjects evaluated the recordings for checking the quality of data [28].

The two acted databases, which are recorded by the professional institutions, have the same number of the emotion categories. They share the following emotions: angry, disgust, fear, happy, neutral and sadness following the palette theory. But there is the boredom but without the surprise in Berlin EmoDB when compared with SAVEE database, and vice versa.

Evaluation protocol - In experiments each dataset was split into two sets randomly, one is training set which takes 80% of the clips, and the other one is testing set which takes the remaining 20% of the clips.

4.1.2 Data Preprocessing

The audio clips are processed simply to meet the requirements of learning deep representation. The audio clips of each database vary in sample rate and length. The audio files of the Berlin EmoDB consist of audio WAV files sampled at 16 kHz, and the audio clips of the SAVEE database sampled at 44.1 kHz. The sample rate of audio clips used in the experiments is 16 kHz. The audio clips of the SAVEE database are all resampled to 16 kHz. All the raw audio clips in the experiments are segmented to 4 s long. At 16 kHz sampling rate, they can be represented as a 64000-dimensional vector. The audio clips of the two selected databases are truncated or padded up to the length of 64000. If the length of the audio clips is longer than 4 s, they are truncated to the length of 64000. Otherwise, they will be padded to 4 s long. Then all the audio data are pre-processed to have zero mean and unit variance. To use the hard softmax, which is applied as a bias, all the audio data are mapped to [0, 1].

4.2 Network Pre-training

The designed 1D deep convolutional-deconvolutional encoder-decoder as a whole, which consists of the encoder and decoder, was trained firstly. During the training, the experimental data was split into two sets randomly, one is training set which takes 80% of the data from the whole dataset, and the other one is testing set which takes the remaining 20% of the data for testing the network.

Early stopping is adopted to combat the overfitting, which leads the model to lost predictive performance. When the monitored quantity stops improving after the patience, the training will stop. The training accuracy, training loss, validation accuracy or validation loss is always the quantity to be monitored in training progress. The patience is the number of epochs with no improvement of the monitored quantity. If the patience is too small, the training of the model will stop when a monitored quantity does not improve; on the contrary, the training will last for a long time. In the pre-training, the monitored quantity is validation accuracy, the patience is set to 8. This patience not only prevents the model from prematurely cancelling training, but also avoids a long training.

4.3 Network Fine-tuning

After the pre-training, the decoder of the designed 1D deep convolutional-deconvolutional encoder-decoder was fine-tuned separately to obtain better performance. The main steps of the fine-tuning are as follows: 1) save the weights of the 1D deep convolutional-deconvolutional encoder-decoder after the pre-training, 2) transfer the learned features of the designed 1D deep convolutional-deconvolutional encoder-decoder to the encoder, 3) save the prediction of the experimental data using the encoder, 4) transfer the learned features to the decoder, and fine-tune the decoder using the prediction and the experimental data. The processing is illustrated in Figure 4.

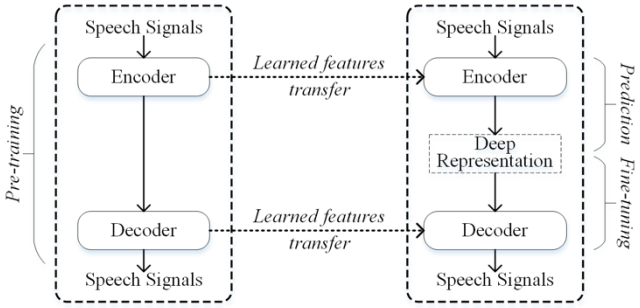


Fig. 4. Illustration of the pre-training and the fine-tuning

Early stopping is also used to combat the overfitting during the fine-tuning. The monitored quantity is validation accuracy, and the patience is also set to 8. In the experiments, the fine-tuning lasted much longer than the training.

4.4 Experimental Results

4.4.1 Experimental Results on Berlin EmoDB

The experimental results on Berlin EmoDB are shown in Table 1. The results were obtained on all the 535 audio clips, which were split into two sets randomly in the ratio of 8:2. From the table, we can see that the training loss and the testing loss in fine-tuning are much smaller than that in pre-training.

Table 1. Berlin EmoDB results for prediction of audio clips. The loss is mean squared error.

Loss	Pre-training	Fine-tuning
Training loss	0.00136	0.00077
Testing loss	0.00105	0.00080

The training and validation losses in pre-training and fine-tuning over epochs on Berlin EmoDB are illustrated in Figure 5. From the figures, we can see that the training and validation losses in fine-tuning are significantly smaller than that in pre-training.

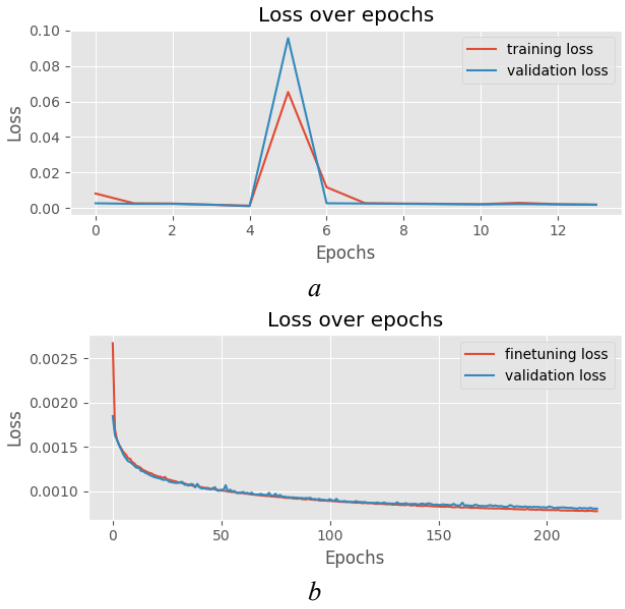


Fig. 5. The training and validation losses in pre-training and fine-tuning over epochs on Berlin EmoDB

The waveforms of the original signal, and the reconstructions produced by the pre-trained model and the fine-tuned model are illustrated in Figure 6. The first waveforms in Figure 6 are the graphical representation of the original signal. The second waveforms are restored by the pre-trained model, and the third waveforms are reconstructed by the fine-tuned model. From the figure, we can see that the waveforms of the reconstruction produced by the fine-tuned model have much higher fidelity when compared to that of the signal restored by the pre-trained model.

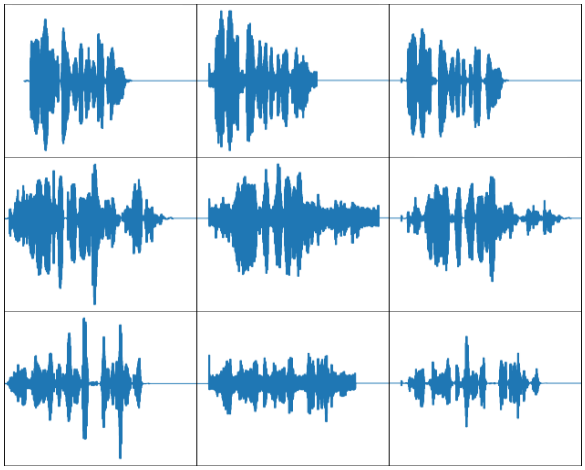


Fig. 6. The waveforms of the original signal, and the reconstructions produced by the pre-trained model and the fine-tuned model

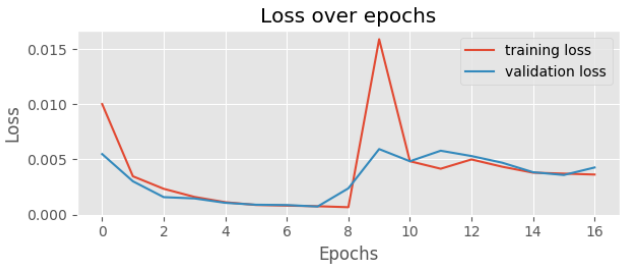
4.4.2 Experimental Results on SAVEE Database

The experimental results on SAVEE Database are shown in Table 2. The results were gained on all the 480 audio clips, which were split into two sets randomly, one is training set which takes 80% of the clips, and the other one is testing set which takes the remaining 20% of the clips. We can see that the training loss and the testing loss of the fine-tuning are much smaller than that of the pre-training.

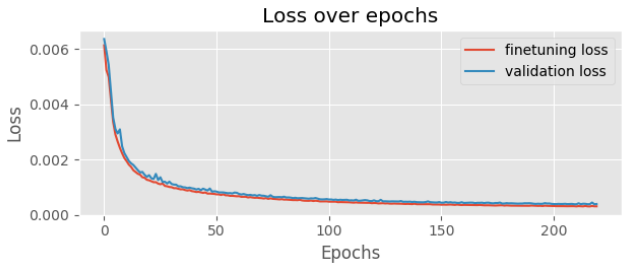
Table 2. SAVEE Database results for prediction of audio clips. The loss is mean squared error.

Loss	Pre-training	Fine-tuning
Training loss	0.00075	0.00031
Testing loss	0.00072	0.00039

The training and validation losses in pre-training and fine-tuning over epochs on SAVEE database are showed in Figure 7. These figures also show that the training and validation losses in fine-tuning are also smaller than that in pre-training. Therefore, in order to learn the more accurate representation of the signal, it is import to fine-tune the model after pre-training.



a



b

Fig. 7. The training and validation losses in pre-training and fine-tuning over epochs on SAVEE database

The waveforms of the original signal, and the reconstructions restored by the pre-trained model and the fine-tuned model are illustrated in Figure 8. The layout of the waveforms in Figure 8 is the same as that in Figure 6 We can also see that the waveforms of the signal produced by the fine-tuned model have much higher fidelity that of the reconstruction produced by the pre-trained model.

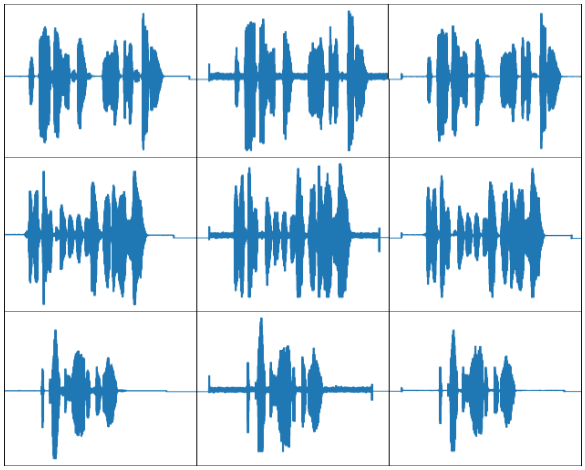


Fig. 8. The waveforms of the original signal, and the reconstructions produced by the pre-trained model and the fine-tuned model

4.5 Results Analysis

The loss of the designed 1D deep convolutional-deconvolutional encoder-decoder as a whole contains three parts:

$$L = L_{encoder} + L_{decoder} + L_{middle} \tag{8}$$

Where $L_{encoder}$ and $L_{decoder}$ are the losses of the encoder and decoder, L_{middle} is intermediate reconstruction loss.

In pre-training, the 1D convolutional-deconvolutional encoder-decoder will be trained as a whole, so the loss is composed of $L_{encoder}$, $L_{decoder}$ and L_{middle}. But in fine-tuning, the decoder is trained separately, so the loss consists only of $L_{decoder}$. The exact location of the maximal value will not be transferred from the encoder to the decoder, but the location information relating to the maximal value will be handed to the decoder by transferred features. The upsampling will not affect the quality of the signal reconstruction significantly. Therefore, the training and validation losses in fine-tuning are significantly smaller than that in pre-training in our experiments.

5. Future Work

The designed 1D deep convolutional-deconvolutional encoder-decoder has obtained better results in the experiments conducted on the Berlin EmoDB and the SAVEE database, and the waveform of the speech restoration has high fidelity. But there are many aspects worthy of further discussion and study:

1. In our experiments, only the maximal value of a pooling region was restored correctly by unpooling. So how to restore all the values of a pooling region by unpooling deserves deeply study.
2. To restore a speech signal from convoluted features accurately, the same filters adopted in convolution should be applied in deconvolution. So how to hand the information of convolutional kernels to deconvolution layer also need for further research.
3. Comparing to the distortion of the restored image, the alteration of the reconstructed speech signal is more likely to be perceived. So how to detect the distortion of the restored speech signal is worth delving into.
4. As we know, the toleration level of a person to an image is higher than that to the speech. So how to measure the distortion level also need for further research.

6. Conclusion

In this paper, we design one 1D deep convolutional-deconvolutional encoder-decoders to learn high-level representation from the raw audio clips. Further, we adopt a dual training mechanism to train the designed architecture, which can leverage the training in a more effective way. By using our proposed mechanism, the loss dropped dramatically, and the reconstruction has much higher fidelity. The results show that the designed architecture can learn deep representation from raw speech signal, and model high-level abstractions of the speech signal from the raw audio clips.

Acknowledgments. Zheng Song and Jianfeng Zhao's work in this paper was supported in part by the National Natural Science Foundation of Inner Mongolia under Grant No. 2020MS06024.

References

[1] Hinton, Geoffrey E., and Richard S. Zemel. "Autoencoders, minimum description length and Helmholtz free energy." *Neural Information Processing Systems* (1993): 3-10.

[2] Lecun, Y., et al. "Backpropagation applied to handwritten zip code recognition." *Neural Computation* 1.4 (1989): 541-551.

[3] Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors." *Nature* 323.6088 (1986): 533-536.

[4] Krizhevsky, Alex, and G. E. Hinton. "Using very deep autoencoders for content-based image retrieval." *Esann 2011, European Symposium on Artificial Neural Networks, Bruges, Belgium, April 27-29, 2011, Proceedings* DBLP, 2011.

[5] Vincent, Pascal, et al. "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion." *Journal of Machine Learning Research* 11.12(2010):3371-3408.

[6] Zhao, Junbo, et al. "Stacked What-Where Auto-encoders." *Computer Science* (2016).

[7] Liao, Y., Y. Wang, and Y. Liu. "Graph Regularized Auto-Encoders for Image Representation. " *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society* PP.99(2016):1-1.

[8] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

[9] He, Kaiming, et al. "Identity mappings in deep residual networks." *European Conference on Computer Vision*. Springer International Publishing, 2016.

[10] Mao, Xiaojiao, Chunhua Shen, and Yu-Bin Yang. "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections." *Advances in Neural Information Processing Systems*. 2016.

[11] Deng, Li, et al. "Binary coding of speech spectrograms using a deep auto-encoder." *Interspeech*. 2010.

[12] Graves, A., N. Jaitly, and A. R. Mohamed. "Hybrid speech recognition with Deep Bidirectional LSTM." *Automatic Speech Recognition and Understanding* 2013:273-278.

[13] Han, Kun, D. Yu, and I. Tashev. "Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine." *Interspeech* 2014.

[14] Huang, Zhengwei, et al. "Speech emotion recognition using CNN." *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014.

[15] Trigeorgis, George, et al. "Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network." *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 2016.

[16] Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." *European conference on computer vision*. Springer International Publishing, 2014.

[17] Shen, F., et al. "Hashing on nonlinear manifolds." *IEEE Transactions on Image Processing a Publication of the IEEE Signal Processing Society* 24.6(2015):1839-51.

[18] Yao, Yazhou, et al. "Exploiting Web Images for Dataset Construction: A Domain Robust Approach." *IEEE Transactions on Multimedia* PP.99(2016):1-1.

[19] Huang, Chen, et al. "Learning Deep Representation for Imbalanced Classification." *IEEE Conference on Computer Vision and Pattern Recognition* IEEE, 2016:5375-5384.

[20] Shen, Fumin, et al. "A Fast Optimization Method for General Binary Code Learning." *IEEE Transactions on Image Processing* 25.12(2016):5610-5621.

[21] Yao, Yazhou, et al. "A Domain Robust Approach for Image Dataset Construction." *ACM on Multimedia Conference* ACM, 2016:212-216.

[22] Jaitly, Navdeep, and Geoffrey Hinton. "Learning a better representation of speech soundwaves using restricted boltzmann machines." *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011.

[23] Bhargava, Mayank, and Richard Rose. "Architectures for deep neural network based acoustic models defined over windowed speech waveforms." *Interspeech*. 2015.

[24] Palaz, Dimitri, Ronan Collobert, and Mathew Magimai Doss. "Estimating phoneme class conditional probabilities from raw speech signal using convolutional neural networks." *arXiv preprint arXiv:1304.1018* (2013).

[25] Palaz, Dimitri, Mathew Magimai Doss, and Ronan Collobert. "Analysis of CNN-based Speech Recognition System using Raw Speech as Input." *conference of the international speech communication association* (2015).

[26] Dieleman, Sander, and B. Schrauwen. "End-to-end learning for music audio." *IEEE International Conference on Acoustics, Speech and Signal Processing* IEEE, 2014:6964-6968.

[27] Burkhardt, Felix, et al. "A database of German emotional speech." *Interspeech 2005 - Eurospeech, European Conference on Speech Communication and Technology*, Lisbon, Portugal, September 2005:1517-1520.

[28] Jackson, Philip J. B., and S. U. Haq. "Surrey Audio-Visual Expressed Emotion (SAVEE) database." (2011).

[29] Boser, Bernhard E., et al. "An analog neural network processor and its application to high-speed character recognition." *Neural Networks, 1991., IJCNN-91-Seattle International Joint Conference on*. Vol. 1. IEEE, 1991.

[30] Behnke, Sven. "Discovering hierarchical speech features using convolutional non-negative matrix factorization." *Neural Networks, 2003. Proceedings of the International Joint Conference on*. Vol. 4. IEEE, 2003.

[31] Shen, Fumin, et al. "Face image classification by pooling raw features." *Pattern Recognition* 54(2016):94-103.

[32] Shen, F., et al. "Approximate least trimmed sum of squares fitting and applications in image analysis." *IEEE Transactions on Image Processing* 22.5(2013):1836-1847.

[33] Kingma, Diederik, and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980* (2014).

[34] Krogh, A. "A Simple Weight Decay Can Improve Generalization." *Advances In Neural Information Processing Systems* 4 1992:950--957.

[35] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." *Journal of Machine Learning Research* 15.1(2014):1929-1958.

[36] Loughrey, John, and P. Cunningham. "Using Early Stopping to Reduce Overfltting in Wrapper-Based Feature Weighting." *Trinity College Dublin Department of Computer Science* (2005).

[37] Zeiler, Matthew D., et al. "Deconvolutional networks." *Computer vision and pattern recognition* (2010).

[38] Goroshin, Ross, Michael Mathieu, and Yann Lecun. "Learning to linearize under uncertainty." *Neural information processing systems* (2015): 1234-1242.

The chaotic system of simplify-memristor simulator in jerk circuit

Junjie Wen, Miao Miao*, Huizhen Yan

{miaomiao@dlpu.com}

School of Information Science and Engineering, Dalian polytechnic University, Dalian, 116034, China

Abstract: The paper puts forward a simplified jerk circuit simulator, based on a simplified have resistance device emulator jerk circuit with unstable saddle points. According to the circuit structure, circuit of the dimensionless model is set up, in the simulation At the same time, many basic dynamic properties, such as stability and equilibrium, are analyzed. The dynamic characteristics in this system are analyzed by the commonly used dynamic analysis methods such as Lyapunov index and bifurcation diagram, and several special dynamic phenomena are observed. For instance: chaos, quasi-periodic, periodic, and some complex behavior. The research results, which will offer experimental basis and theoretical for the application of the chaotic system pseudorandom sequence in the field of public information security, for instance, digital communication security.

Keywords: Jerk circuit, Simplified-memristor, Chaos

1 Introduction

Over the past few years, this inherent nonlinearity of memristors [1] has often been used to design new nonlinear chaotic circuit and system is formed by using memristor to replace the non-linear or linear elements in the existing chaotic circuit system, resulting in a rich variety of non-linear dynamic behaviors [2-15] including quasi-period, period, chaos. Because so far, it has the required structure relationship of today also can't buy in the market, so many have been proposed by the ready-made discrete component implementation of today the emulator, they mainly composed of operational amplifier and analog circuit based on multiplier today with different nonlinear device simulator and related to the second-order and first-order filter network have resistance model. The memristor simulators realized by these equivalent circuits are suitable for the research of hardware experiments. In addition, the jerk system [16-18] is a

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2308042

simple form in the system's algebraic structure, achieving compact circuits with fewer circuit elements. For this reason, this paper introduces a simplified memristor emulator into the existing jerk circuit. Six discrete elements constitute a simplified memristor [19]: integrator U_1, two resistors R, a capacitor C, an analog multiplier M_1 and a resistor R_a. Integrating capacitance C and resistor R are used in parallel in order to avoid DC voltage integration drift. Compared with the proposed memristor, the simplified memristor has the advantages of simple mathematical model and easy circuit system realization.

The paper is laid out as follows. In the second part, a simplified memristor simulator model is introduced, and the shrinkage hysteresis loops are validated by calculation and experiment. We put the dynamic analysis of chaotic system in the third part, and this conclusion of this paper is obtained in the last chapter.

2 Simplified memristor simulator

For convenience study the chaos dynamics of jerk circuits based on memristor simulators, a simplified memristor [19] simulator is proposed based on the references. The simplified memristor [19] simulator consists of six discrete elements: an integrator U_1, a capacitor C_0 and two resistors R_1, a resistor R_e and an analog multiplier A_1. Integrating capacitance C and resistor R are used in parallel in order to avoid DC voltage integration drift. Compared with the proposed memristor, the simplified memristor has the advantages of simple mathematical model and easy circuit system realization..

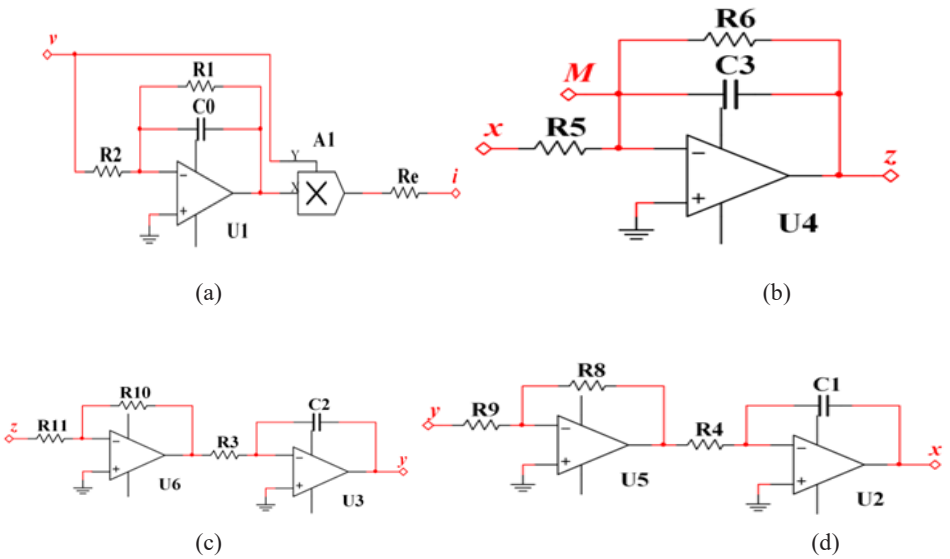


Fig.1 The analog circuit corresponding to the system, (b) (c) (d), memristor (a)

2.1 Memristor model description

The HNN is described by the nonlinear differential equation of the group corresponding to n neurons [21]. In this paper, we consider the connection topology of a non-autonomous mHNN based on two neurons, which is shown in Fig.4. Here, I_m and F are amplitude and frequency parameters associated with the stimulus , by using the external stimulus input $I_{(s)}$ with a sine function $I_{(s)} = I_m \sin(2pF_s)$.

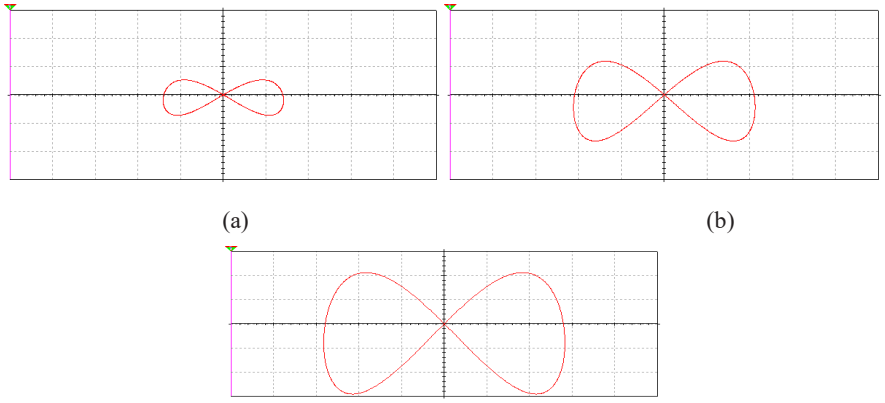


Fig.2 Numerical simulation of simplified pinch hysteresis loop of a memristor in the *v*–*i* plane, (a) *f* = 1 kHz with *v*=2V (b) *f*=1kHz with *v*=3V (c) *f*=1kHz with *v*=4V

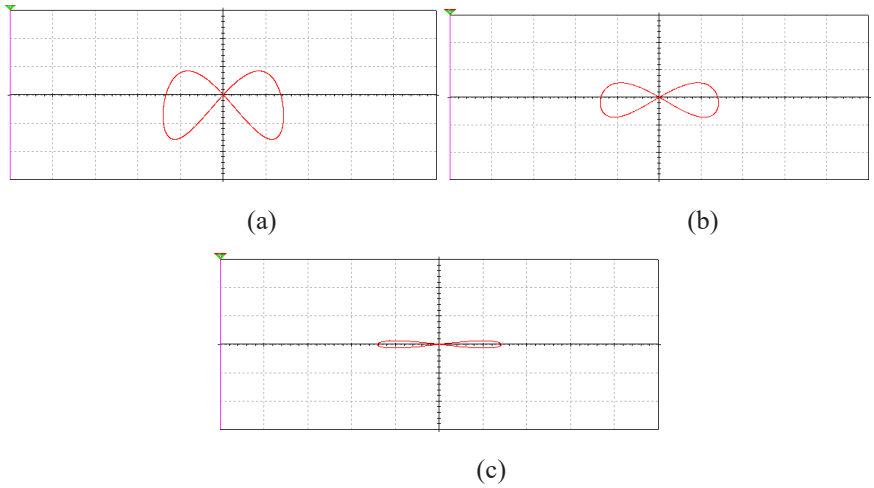


Fig.3 Numerical simulation of simplified pinch hysteresis loop of a memristor in the *v*–*i* plane: (a) *f* = 500 Hz with *v*=2V (b) *f*=1 kHz with *v*=3V (c) *f*=5 kHz with *v*=4V

The model of memristor is obtained by Kirchhoff's law, and the basic relationship of discrete elements is expressed as follows

$$\begin{cases} i = W(v_0)v = \dfrac{1}{R}v_0 v \\ \dfrac{dv_0}{dt} = -\dfrac{1}{R_e}(v_0 + v) \end{cases}, \tag{1}$$

The current and voltage are represented by I and V separately, V_0 is the internal state variable and g is the gain of the multiplier M_1.

2.2 Simplified memristor emulator-based jerk circuit

In particular, resistors R_a and R_b are used as tunable circuit parameters are R = 10kΩ, R_a = 80, R_b = 5kΩ, C_1 = C_2 = C_3 = 10 nF. In Fig. 1, the voltage at both ends of the second order memristor diode bridge and the current flowing through the bridge are represented by V and I, respectively. According to [13], the voltage-current relationship of the simplified memristor is expressed as

$$\begin{cases} C_1\dfrac{dv_1}{dt} = \dfrac{1}{R_a}v_2 \\ C_2\dfrac{dv_2}{dt} = \dfrac{1}{R_b}v_3 \\ C_3\dfrac{dv_3}{dt} = -\dfrac{1}{R_c}v_1 - \dfrac{1}{R_a}v_3 - \dfrac{g}{R}v_0 v_2 \\ C\dfrac{dv_0}{dt} = -\dfrac{1}{R_e}(v_0 + v_2) \end{cases}, \tag{2}$$

Where V_0 is the voltage of capacitor C for the second-order memristor in Fig.1, the voltages V_0, V_1, V_2, and V_3 at both ends of capacitors C_1, C_2, C_3 and C are treated as variables. Since the output voltage of U_4 is equal to V_2, the input voltage of the second-order memristor satisfies the relation $V_2 = V$.

By introducing parameters and new variables in dimensionless form

$$\begin{cases} x_1 = \rho v_1 \\ x_2 = \rho v_2 \\ x_3 = \rho v_3 \\ x_4 = \rho v_4 \\ \tau = \dfrac{t}{R_c}, C = C_1 = C_2 = C_3 \\ a = \dfrac{1}{R_a}, b = \dfrac{1}{R_b}, c = \dfrac{1}{R_c}, g = \dfrac{g}{R}, e = \dfrac{1}{R_e} \end{cases} \tag{3}$$

The circuit equation in can be converted into a normalized system equation

$$\begin{cases} x_1 = ax_2 \\ x_2 = bx_3 \\ x_3 = -cx_1 - ax_3 - gx_4x_2 \\ x_4 = -e(x_4 + x_2) \end{cases} \tag{4}$$

Set the following parameters: a=12.3, b=2.15, c=4.4, e=4.0, g=1.2.

2.3 Eigenvalues and equilibrium point.

That means x_1=x_2= x_3= x_4= 0. It is easy to solve that a primary equilibrium point occurs at (4) which can be express as: E=(0, 0, 0, 0).

The Jacobian [20] at E is derived as

$$J = \begin{bmatrix} 0 & a & 0 & 0 \\ 0 & 0 & b & 0 \\ -c & -gw & -a & gy \\ 0 & -e & 0 & -e \end{bmatrix} \tag{5}$$

This characteristic root equation is:

$$\det(\lambda E - J) = 0 \tag{6}$$

When this parameters are setting as a=12.3, b=2.15, c=4.4, g=1.2, e=4. And this eigenvalues corresponding to the equilibrium point are λ_1=0.1547, λ_2=-0.9030, λ_3=-0.9030, λ_4=0.9030i, There are three positive numbers λ_1, λ_2 and λ_3, and a conjugate complex numbers λ_4. So this system has a saddle-focus equilibrium point (0, 0, 0, 0).

3 Numerical diagram of the dynamical behaviors

Two variable system parameters are determined by formula (3) as b and c. The parameter region of interest is specified as $b \in [1.9, 3]$, $c \in [4, 6]$.

3.1 System phase diagram

When this system parameters a=12.3, b =2.15, c = 4.4, e=4, g=1.2, the time step is 0.01s, and the initial conditions are (1, 1, 1, 1), the four Lyapunov indices L_1=9.7061, L_2=-0.0217, L_3=-3.3504, and L_4=-12.9289,by calculation respectively under this parameter condition. Fig.4 show the phase diagram of chaotic attractor obtained through simulation.

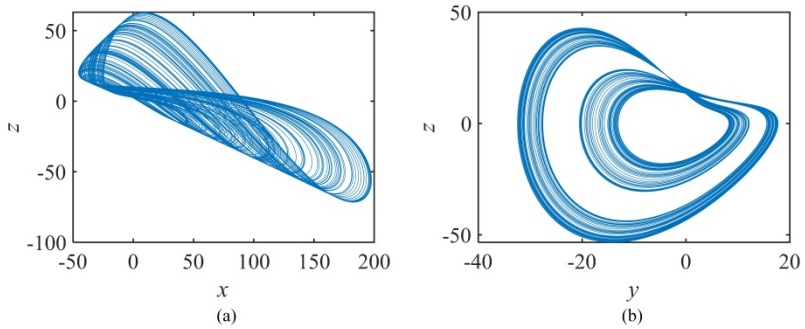


Fig.4 When a =-12.3, b =-2.15, c = 4.4, e=4, g=1.2 chaotic attractor phase diagram, (a) x-y plane (b) y-z plane

3.2 Lyapunov exponential spectrum and bifurcation diagram of the system

This new system is a 4-dimensional, and its states under different parameters can be obtained by combining bifurcation diagram with Lyapunov exponential spectrum. In the following, parameters b and c are taken as variables, initial values are set as (1, 1, 1, 1), step size h=0.01s, remaining parameters of the equation are fixed, and different states of the chaotic system are observed by changing parameters b and c.

Take the parameter $b \in$ [1.9, 3], and let a=12.3, c=4.4, e=4.0, g=1.2. Fig. 5 shows the Lyapunov exponential spectrum and bifurcation diagram of a chaotic system. According to Fig.5, when b=2.3, By finding the maximum Lyapunov exponent L_1 is equal to 0, we know that the system is in a limit cycle state. When $b \in$ [2.6, 2.75], When the maximum Lyapunov index L_1 is less than 0, the system is in a periodic state.It can be seen from the figure that at other times, the maximum Lyapunov index L_1 is always greater than 0 is less than 0, and the system presents a periodic state. At other times, the maximum Lyapunov exponent L_1 is

always above the zero coordinate line , L_2=0, and other Lyapunov exponents are all less than 0, so the system presents a chaotic state.

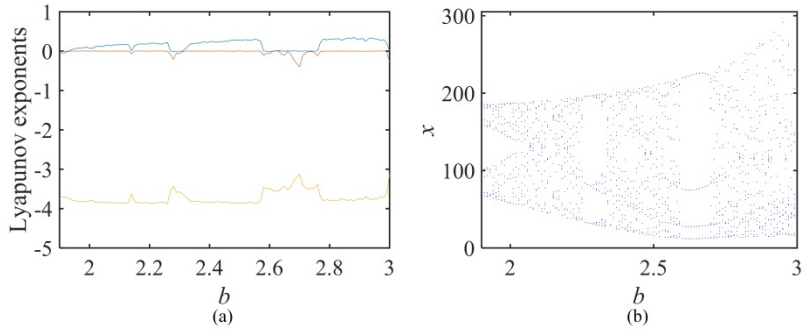


Fig.5 Lyapunov exponential spectrum and bifurcation diagram of this system when a change

Fig.6 shows this phase diagram of *x*-*z* when the parameters *a*=12.3, *c*=4.4, *e*=4, *g*=1.2 are fixed, and *b*=2.3. When *b*=2.3, the phase diagram shows the limit cycle state. When *b*=2.9, the phase diagram is in a chaotic condition. In the meantime the Lyapunov exponent of this system is both less than 0, one is equal to zero and the other is greater than 0, which is in a chaotic condition. Through analysis, it can be easy seen that this result is integrally corresponding to Lyapunov exponential spectrum and the bifurcation diagram shown above.

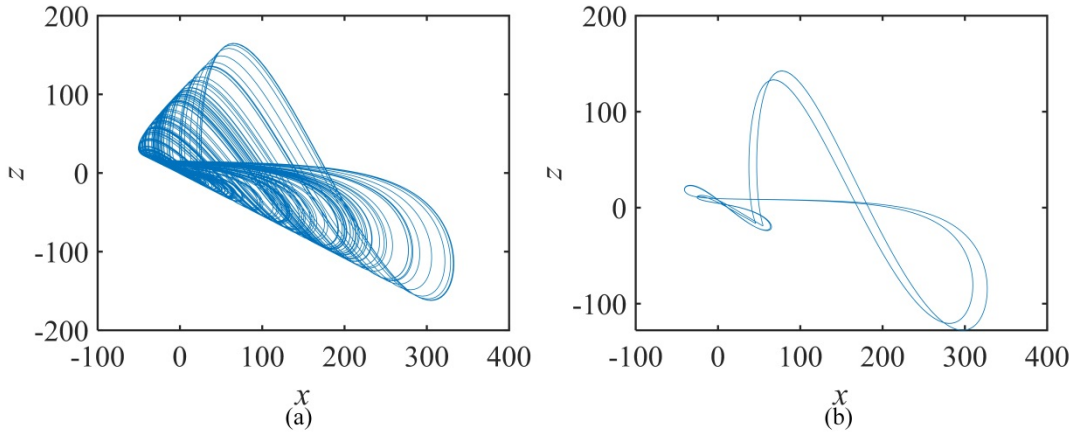


Fig.6 The phase diagram of the system is *b* (a) *b*=2.9 (b) *b*=2.3

With the parameter *c*∈ [4, 6], *b*=2.15, *a*=12.3, *e*=4, *g*=1.2, Fig. 8 shows the Lyapunov exponential spectrum and the bifurcation diagram. When *c*∈ [4.65, 4.73], the maximum Lyapunov exponent L_1 is equal to 0, and the system appears as a limit cycle. When *c*∈ [4.74, 5.26], the maximum Lyapunov exponent L_1 is greater than 0, and no other exponent is greater than 0, then the system shows a chaotic state. When *c*∈ [5.67, 6], When the maximum

Lyapunov index L_1 is greater than 0, and the other indices are not greater than 0, the system can be known to be in a chaotic state is greater than 0, and no other exponent is greater than 0, the system shows a chaotic state. When $c\in$ [5.29, 5.68], When the maximum Lyapunov exponent L_1 is equal to 0, then we know that the system is represented as a limit cycle.

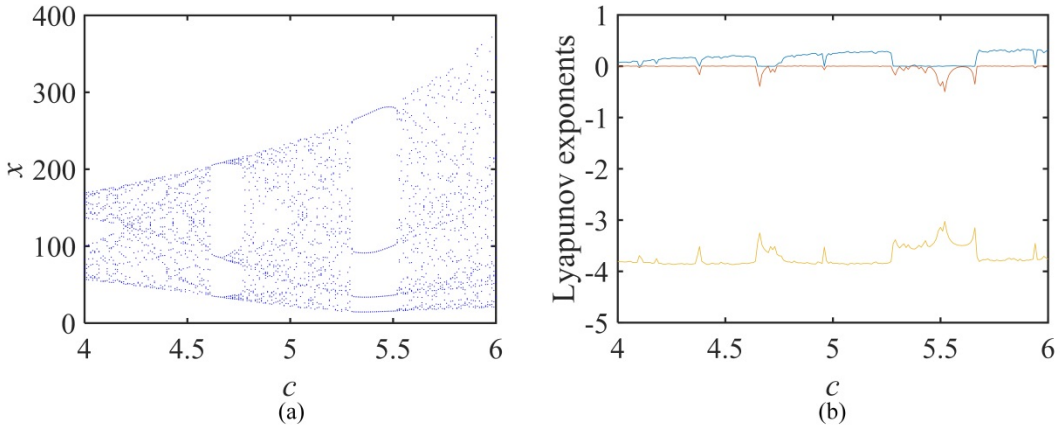


Fig.7 Bifurcation diagram of the system and Lyapunov exponential spectrum when a change (a) Bifurcation diagram (b) Lyapunov exponential spectrum

Fig.8 shows the phase diagram of x-y when the parameters b=2.15, a=12.3, e=4, g=1.2 are fixed, and c is 4.7, 5, 5.4, and 5.9, respectively. When c=4.7 and c=5.4, the phase diagram is limit cycle. When c=5 and c=5.9, the phase diagram is in a chaotic condition. When the Lyapunov exponent of the system is less than 0, one is equal to 0, and the other is always above 0, so we know that the system is in a chaotic state. Through the above analysis, it can be seen that this result is completely corresponding to the bifurcation diagram, and the Lyapunov exponential spectrum of the above system is in a chaotic state. Through analysis, it is easy to see that this result completely corresponds to the bifurcation diagram and Lyapunov exponential spectrum mentioned above.

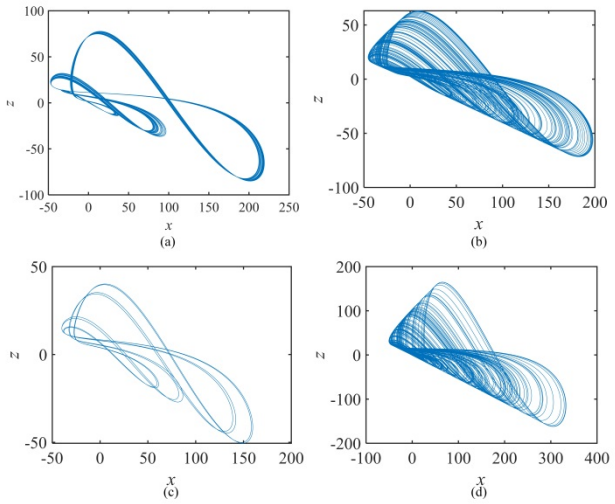


Fig.8 The phase diagram of the system is c (a) c=4.7 (b) c=5 (c) c=5.4 (d) c=5.9

3.3 The analysis of System complexity

The complexity research involves scholars, and various fields in different fields have imparity understandings of complexity. Up to now, there is no qualitative concept of complexity. The complexity of the system can be known by the proximity between the chaotic sequence and the random sequence measured by the correlation algorithm. The higher the complexity value is, the higher the random sequence is and the higher the corresponding security is.

The analysis of system complexity includes complexity of behavior and structure. Behavioral complexity refers to the possibility of generating new models within a short time window measured by some method from the chaotic sequence itself. The more likely a new model is to be generated, the more complex the sequence becomes. Now, there are some algorithms to calculate the behavior complexity of the chaotic pseudo-random sequences, most of which are based on Shannon entropy and Kolmogorov method. These algorithms are quick in calculation and relatively accurate in results. However, if the dimension of pseudo-random sequences is too high or the symbol space of pseudo-random sequences is too large, the calculation results will overflow or even get no results. Structure complexity refers to the frequency characteristics through the transformation domain. The more balanced the energy spectrum distribution in the sequence transformation domain is, when the original sequence is closer to the random signal, the complexity of the sequence is greater. The corresponding spectral entropy value can be calculated by combining the concept of Shannon entropy. The

energy characteristics of the transformation domain are analyzed by structural complexity, which is aimed at whole sequences, not just local sequences, so compared with the behavior complexity algorithm, the results have global statistical significance. In this paper, SE and C0 algorithm are used to analyze the structural complexity.

In this part, parameter c is used as a variable to analyze the complexity of the system. When parameters b=2.15, a=12.3, e=4, g=1.2 are fixed, and $c \in$ [4, 6], Figure 9 shows the simulation results.

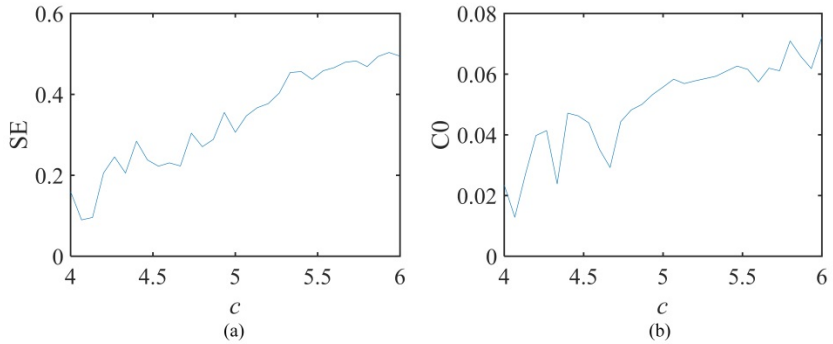


Fig.9 shows that SE algorithm and C0 algorithm have a high degree of synchronization

4 Conclusion

In the paper, we designed this new chaotic circuit and introduce its dimensionless equation. A new chaotic attractor is found in this chaotic system. By analyzing the Lyapunov exponential spectrum, bifurcation diagram and complexity, The dynamic nature was discovered , which show a high degree of complexity and sensitivity with the change of system parameters. Through numerical simulation, we observe the special phenomenon of chaotic. At the same time, some reasonable explanations are given for these phenomena. Because of these rich dynamic behaviors, this new chaotic system has a good application prospect in information encryption and secure communication.

References

[1] K. Eshraghian, O. Kavehei, K.R. Cho, J.M. Chappell, A. Iqbal, S.F. Al-Sarawi, D. Abbott, Proc. IEEE 100, 1991 (2012).

[2] B.C. Bao, T. Jiang, G.Y. Wang, P.P. Jin, H. Bao, M. Chen, Nonlinear Dyn. 89, 1157 (2017).

[3]. M. Chen, B.C. Bao, T. Jiang, H. Bao, Q. Xu, H.G. Wu, J. Wang, Int. J. Bifurc. Chaos 28, 1850120 (2018).

[4]. M. Chen, M.X. Sun, H. Bao, Y.H. Hu, B.C. Bao, IEEE Trans. Ind. Electron. 67, 2197 (2019)
[5]. A.L. Fitch, D.S. Yu, H.H.C. Iu, V. Sreeram, Int. J. Bifurc. Chaos 22, 1250133 (2012)
1058 The European Physical Journal Special Topics
[6]. I. Petras, IEEE Trans. Circuits Syst. II: Express Briefs 57, 975 (2010)
[7]. S.P. Wen, Z.G. Zeng, T.W. Huang, Phys. Lett. A 376, 2775 (2012)
[8]. Q. Xu, Y. Lin, B.C. Bao, M. Chen, Chaos Soliton Fractals 83, 186 (2016)
[9]. H. Bao, W.B. Liu, M. Chen, Nonlinear Dyn. 96, 1879 (2019)
[10]. B.C. Bao, L. Xu, N. Wang, H. Bao, Q. Xu, M. Chen, AEÜ Int. J. Electron. Commun. 94, 26 (2018)
[11]. J. Kengne, Z.N. Tabekoueng, V.K. Tamba, A.N. Negou, Chaos 25, 103126 (2015)
[12]. Z.N. Tabekoueng, J. Kengne, H.B. Fotsin, A.N. Negou, D. Tchiotsop, Chaos Solitons Fractals 91, 180 (2016)
[13]. H.G. Wu, B.C. Bao, Z. Liu, Q. Xu, P. Jiang, Nonlinear Dyn. 83, 893 (2016)
[14]. J.H. Zhang, X.F. Liao, AEÜ Int. J. Electron. Commun. 75, 82 (2017)
[15]. V.T. Pham, S. Vaidyanathan, C.K. Volos, S. Jafari, N.V. Kuznetsov, T.M. Hoang, Eur. Phys. J. Special Topics 225, 127 (2016)
[16]. J.C. Sprott, IEEE Trans. Circuits Syst. II: Express Briefs 58, 240 (2011)
[17]. C.B. Li, J.C. Sprott, H.Y. Xing, Phys. Lett. A 380, 1172 (2016)
[18]. H. Bao, N. Wang, B.C. Bao, M. Chen, P.P. Jin, G.Y. Wang, Commun. Nonlinear Sci. Numer. Simul. 57, 264 (2018)
[19].Xu Q, Lin Y, Bao BC, Chen M. Multiple attractors in a non-ideal active voltage-controlled memristor based Chua's circuit. Chaos Solit Fractals 2016;83:186–200.
[20].A. Wolf, J.B. Swift, H.L. Swinney, J.A. Vastano, Physica D 16, 285 (1985).
[21].Bao BC, Qian H, Xu Q, Chen M, Wang J, Yu YJ. Coexisting behaviors of asymmetric attractors in hyperbolic-type memristor based Hopfield neural network. Front Comput Neurosci 2017; 11:81.
[22].Chen, Chen; Sun, Kehui; He, Shaobo An improved image encryption algorithm with finite computing precision.107340 (2020).

A new fractional-order chaotic system with plane equilibrium: Bifurcation analysis, Multi-stability and DSP implementation

Tianming Liu[1], Xiaoyu Zhang[2], Peng Li[1,*], and Huizhen Yan[1]

{lipeng@dlpu.edu.cn}

[1]School of Information Science and Engineering, Dalian polytechnic University, Dalian, 116034, China
[2]Agronomy College, ShenYang Agricultural University, Shenyang, 110161, China

Abstract: In this paper, based on the definition of conformable differential, a 4D chaotic system is discretized by Adomian decomposition method (ADM) and its numerical solution is obtained. The stability of the system at the equilibrium point is analyzed. From the coexisting phase diagram, bifurcation model and Lyapunov exponential spectrum, the multi-stability of fractional chaotic systems is analyzed. Spectral entropy complexity is used to study the complex performance of fractional order systems. Then, according to the measurement result of spectrum entropy complexity, three sets of parameters with higher complexity are selected to design a pseudo-random number generator (PRNG). The designed PRNG passed all randomness tests, indicating that the system has high practical application value. Meanwhile, the related design of the circuit implementation of the chaotic system was completed and verified on the DSP platform. Research shows that fractional-order chaotic systems have special dynamic behaviors and multiple stability phenomena that evolve with q. The pseudo-random sequence generator and digital circuit designed in this research provide guidance for the application of fractional-order chaotic systems.

Keywords: Dynamics, CADM algorithm, Coexisting attractors, PRNG.

1 Introduction

Chaos has received widespread attention from scholars since its discovery, and it has now become an important part of the nonlinear field. Lorentz accidentally noticed the existence of chaos in the atmosphere and constructed the Lorentz system[1-4]. For a long time, people

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307991

have been committed to building systems with stronger chaotic characteristics. The theory of fractional calculus is used to model and explain the phenomena of nature, which can reveal the dynamic behavior of the actual system more objectively. The fractional order system has great advantages in describing the memory and heredity of various systems[5]. Therefore, some scholars try to introduce the definition of fractional differential into chaotic systems. The definition of fractional differential was introduced into some typical chaotic systems, not only the phenomenon of chaos exists, but the chaotic characteristics of the system become more complicated[6-8]. For example, fractional-order Lorentz system[9], fractional-order Liu system[10], etc. These systems are now widely used in various applications, such as [11, 12], encryption applications[13, 14], biomedical engineering[15], etc.

In recent years, a peculiar phenomenon of coexistence attractors has been discovered in chaotic systems[16]. The typical characteristic of a chaotic system is its high sensitivity to initial conditions. However, some systems have some special phenomena, which need to ensure that a certain parameter remains unchanged. Different initial conditions make the system have multiple numerical solutions[17]. The existing research results show that some low-dimensional nonlinear differential equations will not only present chaotic phenomena, but also produce multiple coexisting attractors[18-21]. Ma et al. studied the multi-stability phenomenon in the chaotic system of Hopfield neural network[22]. Yu et al. proposed a new memristive hyperchaotic system with multiple stability and designed a circuit and chaotic synchronization control in Ref[23]. Liu et al. analyzed the chaotic system with infinite attractors and implemented DSP in Ref[24]. Compared with the research on coexistence attractors of integer order systems, the research literature on coexistence attractors of fractional order systems is relatively scarce. Compared with general fractional-order chaotic systems, fractional-order chaotic systems with coexistence phenomena can show better performance in the application field[25-28]. Nowadays, the multiple stability of fractional-order chaotic systems attracts scholars to further study[29, 30].

The algorithms for solving fractional chaotic systems are mainly frequency domain method (FDM), predictive correction method (ABM) and Adomian decomposition method (ADM)[31]. These algorithms usually use the Riemann-Liouville definition and Caputo definition[32]. Khalil et al. proposed the definition of consistent fractional differential[33], which combined with the ADM algorithm (CADM) to solve the fractional differential equations can overcome some of the shortcomings of the existing algorithms[34]. Compared with ADM algorithm, CADM algorithm has the advantages of fast convergence speed, fast calculation speed and less resource consumption [35]. At present, there are few studies on the use of improved ADM algorithm (CADM) to solve fractional chaotic systems. So far, many

studies have been conducted on the application of fractional chaotic systems, such as nuclear magnetic resonance[36], cellular neural networks and secure communication[37]. The application of chaos relies on digital signal processing technology and the design of hardware circuits. Digital signal processor (DSP) has been widely used in engineering due to its excellent performance and convenient processing[38-40]. Therefore, this paper considers the realization of the chaotic system solving process on the DSP platform. C language is used for programming to obtain an approximate chaotic attractor phase diagram on digital circuits.

The rest of this paper is organized as follows: the definition of consistent fractional-order and the numerical solution of fractional chaotic systems are introduced in Sect 2. The dynamics of the system and its multiple stability are studied In Sect 3. In Sect 4, a pseudo-random number generator was designed and tested for randomness. In Sect 5, Realization of Fractional Chaos System on DSP. Finally, the results of this paper are summarized.

2 Definition of Fractional Calculus and solution of Chaotic System

2.1 Conformable differential definition

Conformable differential definition can be better combined with ADM decomposition algorithm for solving fractional differential equations. Based on this definition, the differential equation can be written as

$$\begin{cases} T_{\mathrm{t}}^{q} x(t) = Lx(t) + Nx(t) + g \\ x^{(k)}(t_0^{+}) = b_k, k = 0, \cdots, m-1 \\ m \in N, m-1 < q \le m \end{cases} \tag{1}$$

here $*T_t^q$ is the derivative operator. The state variables describe as $x(t)=[x_1(t), x_2(t), \ldots, x(t)]^{\mathrm{T}}$. The initial value of the differential equation is expressed as b_k. L and N respectively represent the linear and nonlinear terms of the differential equation. After applying T_t^q to both sides of Eq. (1), it can be described as

$$x(t) = I_{t_0}^{q} Lx(t) + I_{t_0}^{q} Nx(t) + I_{t_0}^{q} g + x(t_0) \tag{2}$$

where $*I_{t_0}^{q}$ is the integral operator, Here, some properties of the integral operator need to be listed as follows

$$I_{t_0}^{q} C = \frac{(t-t_0)^q}{q} C \tag{3}$$

$$I_{t_0}^{q_n} ... I_{t_0}^{q_2} I_{t_0}^{q_1} C = C \frac{(t-t_0)^{q_1+q_2+....+q_n}}{q_1(q_2+q_2)...(q_{1}+...+q_n)} \tag{4}$$

The discrete formula of the nonlinear term is

$$\begin{cases} A_j^i = \frac{1}{i!}[\frac{d^i}{d\lambda^i} N(v_j^i(\lambda))]_{\lambda=0} \\ v_j^i(\lambda) = \sum_{k=0}^{i} (\lambda)^k x_j^k \end{cases} \tag{5}$$

here, i = 0, 1, …, ∞ and j = 1, 2, …, n. According to Equation 5, the nonlinear term can be rewritten as

$$Nx = \sum_{i=0}^{\infty} A^i(x^0, x^1, \cdots, x^i) \tag{6}$$

The solution of the differential equation is represented by x, and the solution formula is as follows

$$x = \sum_{i=0}^{\infty} x^i = I_{t_0}^q L \sum_{i=0}^{\infty} x^i + I_{t_0}^q \sum_{i=0}^{\infty} A^i + I_{t_0}^q \, \mathrm{g} + x(t_0) \tag{7}$$

Among them, x_i can be obtained by the following formula for further solving

$$\begin{cases} x^0 = I_{t_0}^q g + x(t_0) \\ x^1 = I_{t_0}^q L x^0 + I_{t_0}^q A^0(x^0) \\ x^2 = I_{t_0}^q L x^1 + I_{t_0}^q A^1(x^0, x^1) \\ \cdots \\ x^i = I_{t_0}^q L x^{i-1} + I_{t_0}^q A^{i-1}(x^0, x^1, x^2, \cdots, x^{i-1}) \\ \cdots \end{cases} \tag{8}$$

2.2 Fractional-order chaotic system and its solution

A chaotic flow with a plane of equilibria system (CFPES) is proposed in Ref[41].The four-dimensional chaotic system equation is

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 \\ \dot{x}_3 = x_3 + ayw - x_3 x_4 \\ \dot{x}_4 = x_1 x_2 + b x_2 x_3 \end{cases} \tag{9}$$

The parameters of the system are a and b. Based on conformable derivative, FCFPES can express as

$$\begin{cases} T_{t_0}^q x_1(t) = x_2 \\ T_{t_0}^q x_2(t) = x_3 \\ T_{t_0}^q x_3(t) = x_3 + ayw - x_3 x_4 \\ T_{t_0}^q x_4(t) = x_1 x_2 + b x_2 x_3 \end{cases} \tag{10}$$

where q represents the order of the system.

Decompose FCFPES according to formula (3) as follows

$$\begin{bmatrix} L_{x_1} \\ L_{x_2} \\ L_{x_3} \\ L_{x_4} \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \\ x_3 \\ 0 \end{bmatrix}, \begin{bmatrix} N_{x_1} \\ N_{x_2} \\ N_{x_3} \\ N_{x_4} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ a x_2 x_4 - x_3 x_4 \\ x_1 x_2 + b x_2 x_3 \end{bmatrix}, \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{11}$$

Among them, the nonlinearity of the system is formed by the multiplication of two state variables. FCFPES contains similar nonlinear terms. According to Eq (5), one of the nonlinear terms $x_1 x_2$ is decomposed as follows

$$\begin{cases} A_1^0 = x_1^0 x_2^0 \\ A_1^1 = x_1^1 x_2^0 + x_1^0 x_2^1 \\ A_1^2 = x_1^2 x_2^0 + x_1^1 x_2^1 + x_1^0 x_2^2 \\ A_1^3 = x_1^3 x_2^0 + x_1^2 x_2^1 + x_1^1 x_2^2 + x_1^0 x_2^3 \\ A_1^4 = x_1^4 x_2^0 + x_1^3 x_2^1 + x_1^2 x_2^2 + x_1^1 x_2^3 + x_1^0 x_2^4 \end{cases} \tag{12}$$

here, letting the initial value $x_0 = [x_1(t_0^+), x_2(t_0^+), x_3(t_0^+), x_4(t_0^+)]$. Then, the initial value can be described as

$$\begin{cases} x_1^0 = x_1(t_0^+) \\ x_2^0 = x_2(t_0^+) \\ x_3^0 = x_3(t_0^+) \\ x_4^0 = x_4(t_0^+) \end{cases}. \tag{13}$$

According to the two properties of the integrated integral operator, the iterative relation of the system is

$$\begin{cases} c_1^1 = c_2^0 \\ c_2^1 = c_3^0 \\ c_3^4 = c_3^0 + ac_2^0c_4^0 - c_3^0c_4^0 \\ c_4^1 = c_1^0c_2^0 + bc_2^0c_3^0 \end{cases} \begin{cases} c_1^2 = c_2^1 \\ c_2^2 = c_3^1 \\ c_3^2 = c_3^1 + a(c_2^1c_4^0 + c_2^0c_4^1) - (c_3^1c_4^0 + c_3^0c_4^1) \\ c_4^2 = c_1^1c_2^0 + c_1^0c_2^1 + b(c_2^1c_3^0 + c_2^0c_3^1) \end{cases} \tag{14}$$

$$\begin{cases} c_1^3 = c_2^2 \\ c_2^3 = c_3^2 \\ c_3^3 = c_3^2 + a(c_2^2c_4^0 + 2c_2^1c_4^1 + c_2^0c_4^2) - (c_3^2c_4^0 + 2c_3^1c_4^1 + c_3^0c_4^2) \\ c_4^3 = c_1^2c_2^0 + 2c_1^1c_2^1 + c_1^0c_2^2 + b(c_2^2c_3^0 + 2c_2^1c_3^1 + c_2^0c_3^2) \end{cases} \tag{15}$$

$$\begin{cases} c_1^4 14 = c_2^3 23; \\ c_2^4 24 = c_3^3 33; \\ c_3^4 34 = c_3^3 33 + a(c_2^3c_4^0 + 3c_2^2c_4^1 + 3c_2^1c_4^2 + c_2^0c_4^3) - (c_3^3c_4^0 + 3c_3^2c_4^1 + 3c_3^1c_4^2 + c_3^0c_4^3); \\ c_4^4 44 = c_1^3c_2^0 + 3c_1^2c_2^1 + 3c_1^1c_2^2 + c_1^0c_2^3 + b(c_2^3c_3^0 + 3c_2^2c_3^1 + 3c_2^1c_3^2 + c_2^0c_3^3); \end{cases} \tag{16}$$

$$\begin{cases} c_1^5 = c_2^4 \\ c_2^5 = c_3^4 \\ c_3^5 = c_3^4 + a(c_2^4c_4^0 + 4c_2^3c_4^1 + 6c_2^2c_4^2 + 4c_2^1c_4^3 + c_2^0c_4^4) - (c_3^4c_4^0 + 4c_3^3c_4^1 + 6c_3^2c_4^2 + 4c_3^1c_4^3 + c_3^0c_4^4); \\ c_4^5 = c_1^4c_2^0 + 4c_1^3c_2^1 + 6c_1^2c_2^2 + 4c_1^1c_2^3 + c_1^0c_2^4 + b(c_2^4c_3^0 + 4c_2^3c_3^1 + 6c_2^2c_3^2 + 4c_2^1c_3^3 + c_2^0c_3^4); \end{cases}$$

(17)

So, the numerical solution of FCFPES obtained by CADM algorithm is

$$x_j(t) = \sum_{i=0}^{5} c_j^i \frac{(t-t_0)^{iq}}{i!q^i} \tag{18}$$

here j=1,2,3,4. Ensure that the calculation accuracy after iteration is high enough, and the numerical solution can be calculated after five iterations. Setting a=-1.1, b=0.92, q=0.9, h=0.01 and initial value x_0= [2.5, 0, -1, -0.4]. The Lyapunov exponents are L_1=0.2567, L_2=0, L_3=0, and L_4=-1.528, and the Lyapunov dimension is D_L= 3.1680. L_1 is a positive Lyapunov exponent which means that this system is chaotic in this set of parameters. The phase diagrams on different planes of the system are shown in Fig. 1.

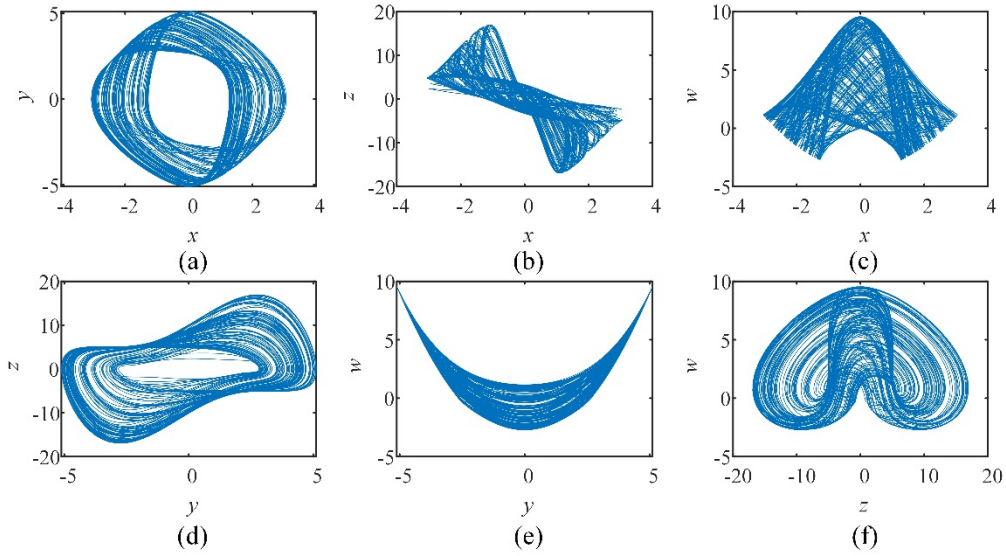


Fig. 1 Phase portraits of FCFPES (q=0.9) (a) x-y plane (b) x-z plane (c)x-w plane (d)y-z plane(e) y-w plane (f)z-w plane

3 Dynamical of different parameters of FCFPES

3.1 Stability analysis of fractional-order chaotic system

Analyzing the stability at the equilibrium point is an indispensable research content of chaos theory in fractional chaotic systems. As we all know, the stability of the fractional order judgment balance point is inconsistent with the integer order. The necessary conditions for the asymptotic stability of the fractional order system using eigenvalues are given in Ref[42].

Lemma 1. The characteristic value satisfies the following conditions

$$\left|\arg(\lambda)\right| \geq \frac{\pi\alpha}{2}, \alpha = \max(q_i, i = 1,2,3) \tag{19}$$

where q is the order of the fractional order, and λ is the eigenvalue. The eigenvalues of the system satisfy Theorem 1, which means the equilibrium point is stable. The stable area of the fractional chaotic system exceeds the imaginary axis and presents a fan-shaped area with an origin, as shown in Fig. 2. Obviously, if the eigenvalue is outside the fan-shaped area, the equilibrium point is unstable. The unstable equilibrium point is an important condition for the

chaotic phenomenon of the system.

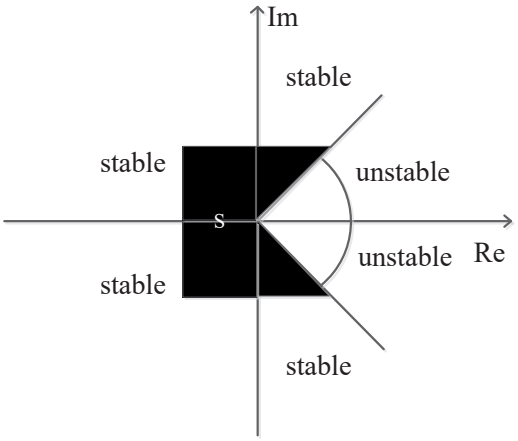


Fig. 2 Stability region of fractional-order system

Setting the left side of Eq. (1) to zero, FCFPES has a flat set of equilibrium points expressed as $E=\{(x, 0, 0, w)|x, w\in R\}$. The Jacobian matrix J of FCFPES can be described as

$$J_E = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & aw & 1-w & 0 \\ 0 & x & 0 & 0 \end{bmatrix} \tag{20}$$

The corresponding characteristic equation is

$$\lambda^2(\lambda^2 + (w-1)\lambda - aw) = 0 \tag{21}$$

FCFPE always has the value of $\lambda_{1,2}$ as zero, and $\lambda_{3,4}$ are related to the initial value of w and the system parameter a. Some cases and the corresponding characteristic values and their stability are list in Table1.

Table 1 Different values, balance point, eigenvalue and stability (q=0.9)

a	q	w	λ	stability
-1	0.9	1	(0, 0, i, -i)	Hopf branching point
		-0.4	(1.64, 0, 0, -0.24)	Unstable focus
		1.1	(0, 0, -0.05+1.0476i, -0.05-1.0476i)	Stable saddle-focus
		0.25	(0.38+0.3356i, 0.38+0.3356i, 0,0)	Unstable saddle-focus

3.2 The analysis of dynamic characteristics evolves with parameters *a* and *b*

In this section, a and b are set as bifurcation parameters, and the evolution process of a

chaotic system with order q= 0.9 is analyzed through the bifurcation model and LEs. The fixed parameter b=0.92, and the LEs for $a\in$(-1.7, -0.942) are shown in Fig. 3(a). Fig. 3(b) shows the bifurcation model, which corresponds well to the LES. Among them, FCFPES has a positive cycle doubling bifurcation at a=-1.411. After several period-doubling bifurcations, it enters the chaotic state at a=-1.382, and then enters the periodic window through some boundary crisis points and then enters the chaotic state again through the slicing and bifurcation. Finally, at a=-0.991, it branches from the reverse period doubling into the periodic state. The bifurcation behavior of parameter a is a typical chaotic bubble. Figs. 3(c) and 3(d) are the LEs and bifurcation models that vary with the parameter b. For a=-1.1, the positive period doubling bifurcation, edge crisis and split bifurcation can also be seen in Fig. 3(d). Table 2 gives detailed status to parameters a and b.

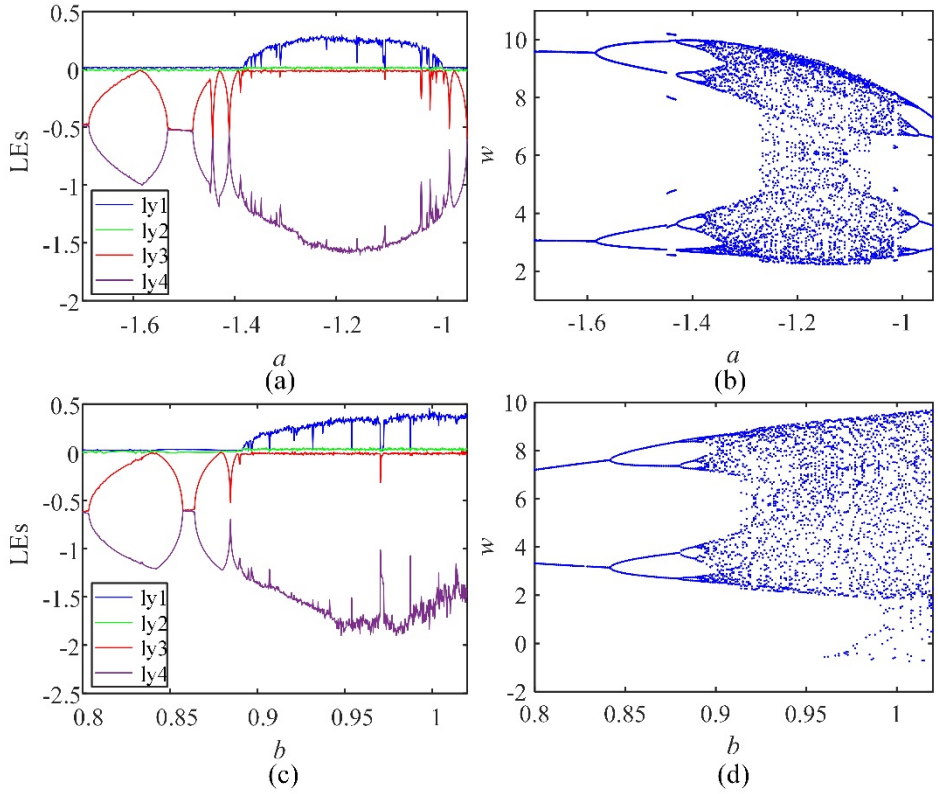


Fig. 3 Lyapunov exponent spectrum (LEs) and bifurcation diagram for different parameters. (a)-(b) The dynamics that varies with a when b=0.92; (c)-(d) The dynamics that varies with b when a=-1.1.

3.3 The analysis of dynamic characteristics evolves with order q

FCFPES has a specific parameter q, which is used as a bifurcation parameter to study the evolution behavior of chaotic systems. The definition of conformable derivative satisfies the definition of integer order when the order q is 1. Therefore, this section can compare the integer-order and the fractional-order. Letting a=-1.1 and b=0.92, and the initial value is also x_0= (2.5, 0, -1, -0.4). In Fig. 4, the bifurcation diagram and Les are obtained. Obviously, FCFPES are all chaotic in $q \in (0.7, 1)$. In addition, when q gradually decreases, the largest Lyapunov exponent gradually increases from 0.26 to 1.15, and the number of bifurcation points becomes denser. Obviously, FCFPES has a larger positive Ly value, which proves that the reconstructed FCFPES has a greater ratio of complexity and dynamic characteristics.

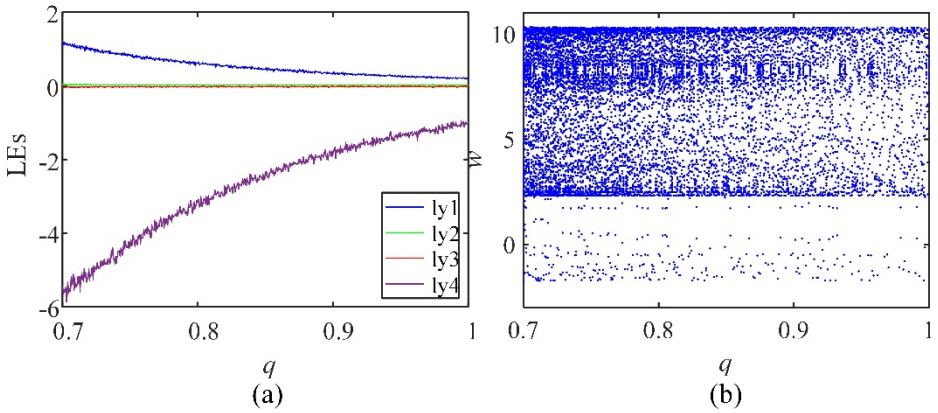


Fig. 4 Dynamical behaviors when a=-1.1, b=0.92, $q \in (0.7, 1)$. (a) LEs; (b) bifurcation diagram

3.4 Multi-stability analysis

Multi-stability phenomena exist in some chaotic systems, and this special phenomenon plays a vital role in the field of information security. It means that the system has multiple sets of solutions when the parameter values remain unchanged. The bifurcation graph that vary with the initial value x_0 verifies the multi-stability in FCFPES. The fixed parameters a=-1.1, b=0.92 and q=0.9. Figure 5 is a bifurcation model, where (a), (b) and (c) correspond to x_1= (k_1, 0, -1, -0.4), x_2= (2.5, k_2, -1, -0.4) and x_3= (2.5, 0, -1, k_3). In Fig. 5(a), an obvious positive period-doubling bifurcation is observed, and the system state enters chaos from period through period-doubling bifurcation. In Fig. 5(b), there are periodic states and chaotic states, and they

are symmetrical and discontinuous bifurcation models. Fig. 5(c) Chaos enters the periodic state through reverse period-doubling bifurcation behavior. With k as the bifurcation parameter, each state is a different solution of the system, that is, the phenomenon of coexistence attractor. The coexistence phase diagram can explain this phenomenon more intuitively. The coexistence phenomena of different types of attractors including period-period, period-chaos and chaos-chaos coexistence phenomena are shown in Fig. 6. In particular, Fig. 6(b) is a heterogeneous chaotic coexistence attractor, and Fig. 6(d) is a heterogeneous periodic coexistence attractor. The above research shows that FCFPES has a relatively rich coexistence phenomenon, which will be more conducive to its application in confidential communication and other fields.

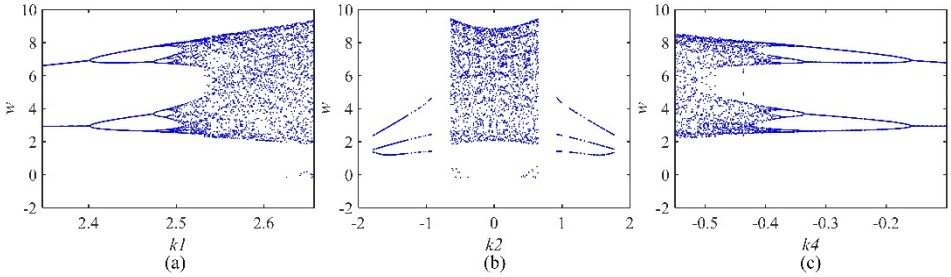


Fig. 5 Bifurcation model of w that changes with the initial value. (a) x_1= (k_1, 0, -1, -0.4); (b) x_2= (2.5, k_2, -1, -0.4); (c) x_3= (2.5, 0, -1, k_3).

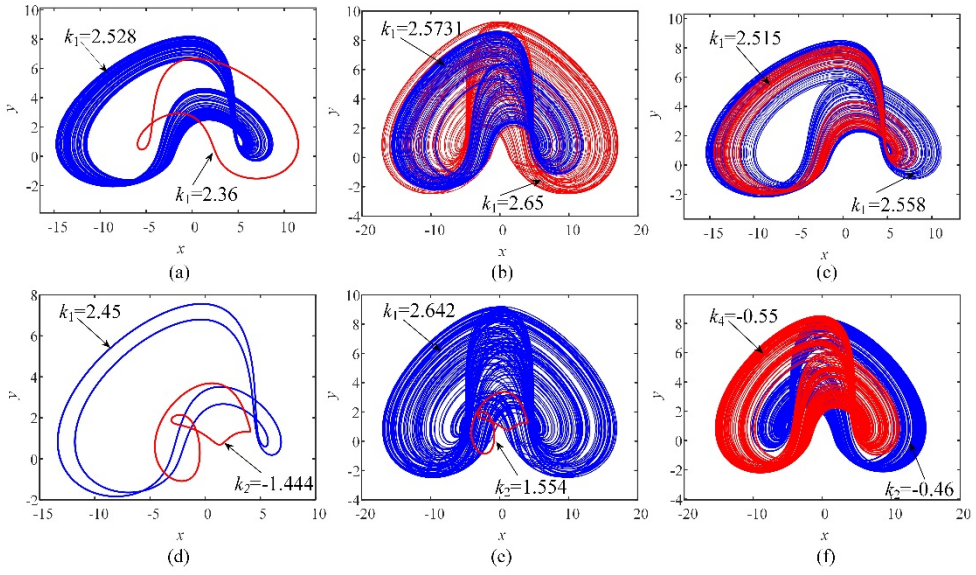


Fig. 6 Coexistence attractor on x-y plane with different k. (a) k_1=2.528, k_1=2.36; (b) k_1=2.5731,

k_1=2.65; (c) k_1=2.515, k_1=2.558; (d) k_1=2.45, k_2=-1.444; (e) k_1=2.642, k_2=1.554; (f) k_2=-0.46, k_4=-0.55.

3.5 Attractor with different parameter plane

The basin of attraction drawn based on the Lyapunov exponent can effectively reflect the state of the system. The system state distribution can be obtained from the attractor basin, and at the same time it also provides a selection instruction for its application. It should be noted that the parameter selection should avoid taking values at the boundary as much as possible, because the different calculation accuracy may cause the degradation and divergence of the system state. If it is used in image encryption and other fields, the value should be set within the orange range as much as possible for better encryption effect. Fig. 8 shows the attraction basins of two planes *a*-*q* and *b*-*q*. The different states of the system are marked with different colors. In order to distinguish periodic, chaotic and divergent are represented by yellow, orange and blue respectively.

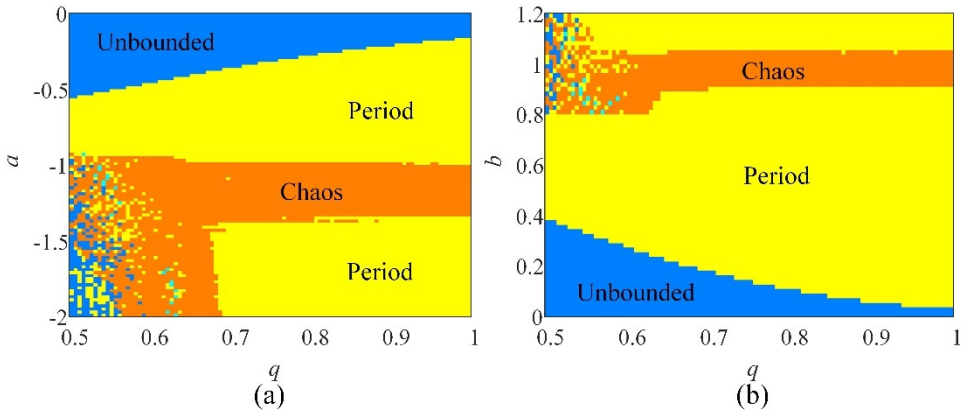


Fig. 7 Two-dimensional attractor basin (a) b=0.92, $a\in$(-2, 0), $q\in$(0.5,1); (b) a=-1.1; $b\in$(0,1.2), $q\in$(0.5,1).

3.6 Complexity test of chaotic sequence

The sequence generated by the chaotic system has special pseudo-randomness, which is an important performance of the application of the chaotic system. Therefore, it is necessary to measure the randomness of the sequence. The spectral entropy (SE) algorithm is improved from the Shannon entropy algorithm. The algorithm not only has a fast calculation speed and

can better reflect the related structure of the sequence, but also can measure the complexity of the system more effectively. Especially when calculating continuous and stable time series, the advantage of the SE algorithm is more obvious. It should be pointed out that when calculating SE complexity, the first 20,000 data needs to be discarded to eliminate the influence of errors caused by system startup.

Fig. 8 is a 3D SE complexity diagram of different parameter planes with x_0= (2.5, 0, -1, -0.4). The complexity graph on the *a*-*b* plane with q = 0.9 as shown in Fig. 8(a). In this region, the spectral entropy almost reaches 0.7, which means that the higher the SE value, the better the pseudo-randomness of the sequence. Therefore, the selection parameter should try to take the value in the red area to ensure high complexity. Figs. 8(b) and 8(c) are the complexity diagrams on the *a*-*q* plane and the *b*-*q* plane. A similar trend is found from the two figures, that is, the complexity will increase correspondingly as q decreases within a certain range. In addition, when a or b is unchanged, SE can take the lowest value at q=1. In other words, FCFPES have higher complexity and better random performance than integer-order chaotic systems. Meanwhile, the complexity graph provides some basis for the selection of parameters.

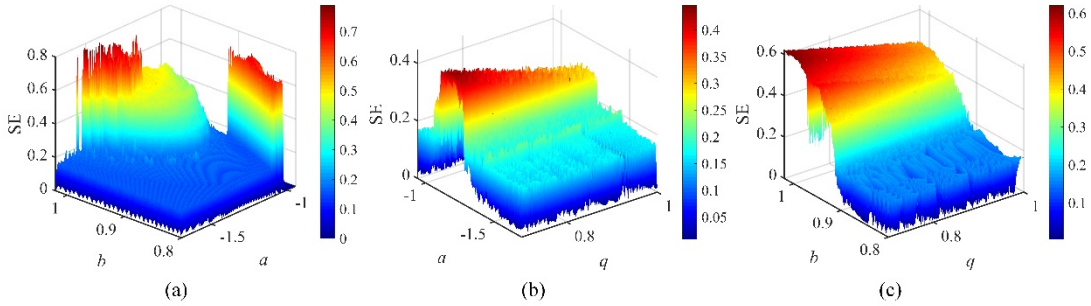


Fig. 8 Spectral Entropy (a)on *a*-*b* plane; (b)on *a*-*q* plane; (c)on *b*-*q* plane.

4 Chaos-based PRNG

Chaos theory has applications in many fields, among which it can be designed to be a pseudo-random number generator (PRNG) with pseudo-randomness in its generated sequence. The requirements of PRNGs appear in many encryption applications. The chaotic system with multiple stability can be used to design a PRNG with good performance. In the previous section, we observed the high sensitivity of CFPES by changing its initial value by generating various coexistence attractors. Therefore, three sets of parameters are selected in the higher

numerical area in the complexity graph to construct the PRNG respectively. Here, a new PRNG generation algorithm is proposed based on FCFPES. Considering the time consumed by iterative calculations and the amount of data that needs to be generated to achieve fast calculations on hardware, our experiment is based on MATLAB to generate 16-bit significant digital PRNG process as shown in Fig9.

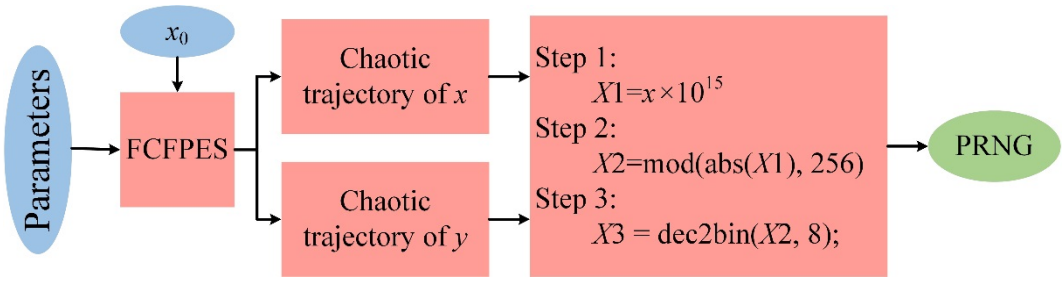


Fig. 9 The flowchart of generating PRNG

Our experiment uses statistical software NIST-800-22, which is used for statistical testing to test the randomness of PRNG. NIST-800-22 contains multiple empirical statistical tests, which can provide a true assessment of the randomness of PRNG. Generally speaking, when p>0.01 obtained by testing a binary sequence, we think PRNG has passed the NIST test. According to the algorithm, when the parameters are regarded as input, we can obtain three PRNGs, corresponding to the parameters (a, b, q) = (-1.189, 0.512, 0.9), (-1.134, 0.92, 0.8) and (-1.1, 1.019, 0.7). The data used for measurement is divided into 100 groups, each of which is a 1,000,000-bit binary sequence. The p-values of PRNG1, PRNG2 and PRNG3 tested by NIST-800-22 are listed in Table 3. Obviously, all p-values are greater than 0.01, that is, three PRNGs pass all statistical tests.

Table 3. NIST tests results of FCFPES.

Each group contains 1,000,000 dates					
	NIST tests	p-value			Result
		PRNG1	PRNG2	PRNG3	
1.	Block frequency	0.3041	0.0210	0.0909	Random
2.	Frequency	0.6787	0.7198	0.3505	Random
3.	FFT	0.0909	0.4373	0.7598	Random
4.	Approximate entropy	0.3505	0.1154	0.2023	Random
5.	Cumulative sums	0.9569	0.0762	0.8248	Random
6.	Serial-1	0.8311	0.4197	0.4343	Random
7.	Runs	0.7981	0.1626	0.6163	Random
8.	Longest run of ones	0.7792	0.3041	0.4190	Random

9.	Overlapping template	0.4012	0.4373	0.8343	Random
10.	Non-overlapping template	0.4930	0.4996	0.5164	Random
11.	Linear complexity	0.9717	0.6993	0.7197	Random
12.	Binary matrix rank	0.8514	0.7792	0.0519	Random
13.	Universal statistical	0.3838	0.9717	0.9114	Random
14.	Random excursions	0.4621	0.3578	0.3232	Random
15.	Random Excursions Variant	0.4217	0.3534	0.3808	Random

5 DSP implementation

Compared with analog circuits, digital circuits have faster processing speed, high precision and reliability. Therefore, the realization of the digital circuit of the chaotic system is more reliable. In this section, we use DSP to implement CFPES. We choose the DSP chip model as TMS320F28335. The chip's operating speed reaches 150MHz and 16-bit floating point operations can also meet the required accuracy requirements. In addition, a DA converter is needed to convert the digital signal output by the DSP into an analog signal and import it into the oscilloscope. The model of the D/A converter is DAC8552. Fig. 10 shows a schematic diagram of each hardware connection. C language is used to design the software design of the chaotic system. Fig. 11 shows the 1DSP platform designed and implemented.

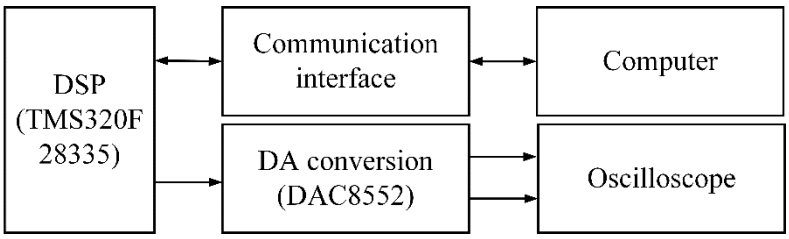


Fig. 10 DSP connection diagram.

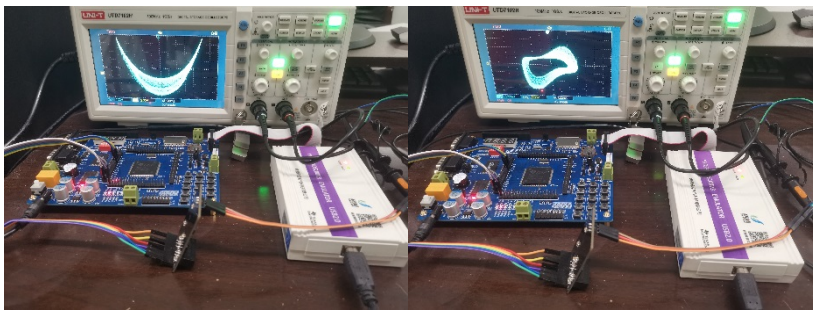
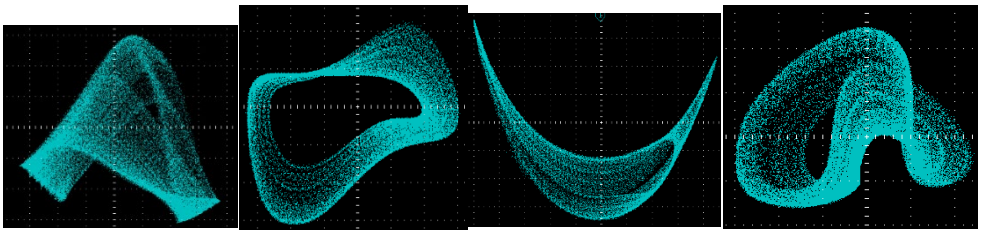


Fig. 11 Digital circuit experimental platform

The DSP implementation parameters are set as, h = 0.01, x_0 = (2.5, 0, -1, -0.4), a = -1.1, b = 0.92, and q = 0.9. It should be pointed out that this set of parameters is the same as the attractor parameters in Fig. 1. The oscilloscope shows the attractor of the signal after D/A conversion on a plane. Adjust the oscilloscope to take a screenshot, and the result is shown in Figure 12. The experimental results obtained the approximate chaotic attractors.



(a) (b) (c) (d)

Fig. 12 Chaotic attractors on DSP platform

6 Conclusions

In this paper, a new four-dimensional fractional-order chaotic system is constructed and its numerical solution are obtained by CADM. The balance point set of the system is distributed on a plane. It is interesting that the stability at the equilibrium point is affected by the initial value. When the system changes with parameters, some typical bifurcations appear, such as positive and negative cycle doubling bifurcations, slicing bifurcations, and boundary crises, showing abundant dynamics. When changing with the order q, the value of LEs reaches the maximum value at q=0.7. The dynamic behavior changing with the parameters indicates that FCFPES has a higher complexity chaotic sequence and richer dynamic evolution behavior. In addition, the multi-stable phenomenon of FCFPES is studied through the coexistence phase diagram and bifurcation model, which includes the asymmetric coexistence periodic attractor and coexistence attractor, and the symmetric coexistence chaotic attractor phenomenon. The basin of attraction drawn based on LE provides a basis for the selection of the system parameters. The reconstructed fractional-order chaotic system still has multiple stability and different coexistence types. Then, the three PRNGs were tested using NIST-800-22, and the results showed that they all passed all statistical tests. Finally, the realization of the DSP platform demonstrates the feasibility of the system. This paper provides a certain research foundation for the research of fractional chaotic systems, and also provides new ideas for the

application of fractional chaotic systems.

Acknowledgements

This work is supported by Scientific Research Funding Project of the Educational Department of Liaoning Province, China (J2020051).

References

[1] S. Mukhopadhyay and S. Banerjee, "Learning dynamical systems in noise using convolutional neural networks," Chaos, vol. 30, no. 10, p. 103125, Oct 2020, doi: 10.1063/5.0009326.

[2] L. Zhang, K. Sun, S. He, H. Wang, and Y. Xu, "Solution and dynamics of a fractional-order 5-D hyperchaotic system with four wings," The European Physical Journal Plus, vol. 132, no. 1, 2017, doi: 10.1140/epjp/i2017-11310-7.

[3] M. Wang, Y. Deng, X. Liao, Z. Li, M. Ma, and Y. Zeng, "Dynamics and circuit implementation of a four-wing memristive chaotic system with attractor rotation," International Journal of Non-Linear Mechanics, vol. 111, pp. 149-159, 2019, doi: 10.1016/j.ijnonlinmec.2019.02.009.

[4] K. Sun, X. I. A. Wang, and J. C. Sprott, "Bifurcations and Chaos in Fractional-Order Simplified Lorenz System," International Journal of Bifurcation and Chaos, vol. 20, no. 04, pp. 1209-1219, 2010, doi: 10.1142/s0218127410026411.

[5] Chenguang et al., "Coexistence of multiple attractors for an incommensurate fractional-order chaotic system," The European Physical Journal Plus, vol. 135, no. 1, pp. 1-21, 2020.

[6] C. Ma, J. Mou, Y. Cao, T. Liu, and J. Wang, "Multistability analysis of a conformable fractional-order chaotic system," Physica Scripta, vol. 95, no. 7, 2020.

[7] S. Eshaghi, "Synchronize and control chaos of chaotic fractional Laser system," in The Third Conference on Chaotic Dynamical Systems, 2020.

[8] M. Farman, A. Akgül, M. U. Saleem, S. Imtiaz, and A. Ahmad, "Dynamical behaviour of fractional-order finance system," Pramana, vol. 94, no. 1, pp. 1-10, 2020.

[9] S. He, K. Sun, and S. Banerjee, "Dynamical properties and complexity in fractional-order diffusionless Lorenz system," The European Physical Journal Plus, vol. 131, no. 8, 2016, doi: 10.1140/epjp/i2016-16254-8.

[10] C. Liu, X. Cai, Y. Wang, L. Liu, Y. Wang, and G. Zheng, "Fractional-Order Hidden Attractor Based on the Extended Liu System," Mathematical Problems in Engineering, vol.

2020, 2020.

[11] A. Khan, L. S. Jahanzaib, and P. Trikha, "Changing Dynamics of the First, Second and Third Approximations of the Exponential Chaotic System and Their Application in Secure Communication Using Synchronization," International Journal of Applied & Computational Mathematics, vol. 7, no. 1, 2021.

[12] N. A. A. Fataf, S. K. Palit, S. Mukherjee, M. R. M. Said, D. H. Son, and S. Banerjee, "Communication scheme using a hyperchaotic semiconductor laser model: Chaos shift key revisited," European Physical Journal Plus, vol. 132, no. 11, p. 492, 2017.

[13] F. Yang, J. Mou, K. Sun, and R. Chu, "Lossless image compression-encryption algorithm based on BP neural network and chaotic system," Multimedia Tools & Applications, vol. 79, no. 1–2, 2020.

[14] F. Yang, J. Mou, C. Luo, and Y. Cao, "An improved color image encryption scheme and cryptanalysis based on hyperchaotic sequence," Physica Scripta, 2019.

[15] A. N. Kengnou Telem, D. Tchiotsop, T. Kanaa, H. B. Fotsin, and D. Wolf, "A robust chaotic fast Walsh transform encryption for gray scale biomedical image transmission," Signal & Image Processing An International Journal, vol. 6, no. 3, pp. 81-102, 2015.

[16] F. T. Arecchi, R. Meucci, P. GP, and T. JR, "Experimental Evidence of Subharmonic Bifurcations, Multistability, and Turbulence in a Q-Switched Gas Laser," Physical Review Letters, vol. 49, p. 1217, 1982.

[17] H. Natiq, M. R. M. Said, M. R. K. Ariffin, S. He, L. Rondoni, and S. Banerjee, "Self-excited and hidden attractors in a novel chaotic system with complicated multistability," The European Physical Journal Plus, vol. 133, no. 12, 2018, doi: 10.1140/epjp/i2018-12360-y.

[18] F. Yu et al., "Multistability Analysis, Coexisting Multiple Attractors, and FPGA Implementation of Yu–Wang Four-Wing Chaotic System," Mathematical Problems in Engineering, vol. 2020, p. 7530976, 2020/08/29 2020, doi: 10.1155/2020/7530976.

[19] J.-J. He and B.-C. Lai, "A novel 4D chaotic system with multistability: Dynamical analysis, circuit implementation, control design," Modern Physics Letters B, vol. 33, no. 21, p. 1950240, 2019, doi: 10.1142/s0217984919502403.

[20] F. Li, C. Tai, H. Bao, J. Luo, and B. Bao, "Hyperchaos, quasi-period and coexisting behaviors in second-order-memristor-based jerk circuit," The European Physical Journal Special Topics, vol. 229, no. 6-7, pp. 1045-1058, 2020, doi: 10.1140/epjst/e2020-900123-5.

[21] B. Bao et al., "Chaotic Bursting Dynamics and Coexisting Multistable Firing Patterns in 3D Autonomous Morris–Lecar Model and Microcontroller-Based Validations," International Journal of Bifurcation and Chaos, vol. 29, no. 10, 2019, doi: 10.1142/s0218127419501347.

[22] C. Ma, J. Mou, F. Yang, and H. Yan, "A fractional-order hopfield neural network chaotic system and its circuit realization," The European Physical Journal Plus, vol. 135, no. 1, p. 100, 2020/01/23 2020, doi: 10.1140/epjp/s13360-019-00076-1.

[23] F. Yu et al., "Dynamic Analysis, Circuit Design, and Synchronization of a Novel 6D Memristive Four-Wing Hyperchaotic System with Multiple Coexisting Attractors," Complexity, vol. 2020, p. 5904607, 2020/05/19 2020, doi: 10.1155/2020/5904607.

[24] T. Liu, J. Yu, H. Yan, and J. Mou, "A Fractional-Order Chaotic System With Infinite Attractor Coexistence and its DSP Implementation," IEEE Access, vol. 8, pp. 199852-199863, 2020, doi: 10.1109/access.2020.3035368.

[25] K. Rajagopal, A. J. M. Khalaf, Z. Wei, V.-T. Pham, A. Alsaedi, and T. Hayat, "Hyperchaos and Coexisting Attractors in a Modified van der Pol–Duffing Oscillator," International Journal of Bifurcation and Chaos, vol. 29, no. 05, 2019, doi: 10.1142/s0218127419500676.

[26] C. Li, W. Joo-Chen Thio, J. C. Sprott, H. H.-C. Iu, and Y. Xu, "Constructing Infinitely Many Attractors in a Programmable Chaotic Circuit," IEEE Access, vol. 6, pp. 29003-29012, 2018, doi: 10.1109/access.2018.2824984.

[27] S. Qureshi, A. Atangana, and A. A. Shaikh, "Strange chaotic attractors under fractal-fractional operators using newly proposed numerical methods," The European Physical Journal Plus, vol. 134, no. 10, 2019, doi: 10.1140/epjp/i2019-13003-7.

[28] C. Li, F. Min, and C. Li, "Multiple coexisting attractors of the serial–parallel memristor-based chaotic system and its adaptive generalized synchronization," Nonlinear Dynamics, 2018.

[29] F. Yu et al., "A New 4D Four-Wing Memristive Hyperchaotic System: Dynamical Analysis, Electronic Circuit Design, Shape Synchronization and Secure Communication," International Journal of Bifurcation and Chaos, vol. 30, no. 10, p. 2050147, 2020.

[30] A. B. A, K. R. B, A. J. M. K. C, S. J. A, G. D. L. D. E, and J. K. D, "Dynamical analysis of a new multistable chaotic system with hidden attractor: Antimonotonicity, coexisting multiple attractors, and offset boosting," Physics Letters A, vol. 383, no. 13, pp. 1450-1456, 2019.

[31] D. Peng, K. H. Sun, and A. O. A. Alamodi, "Dynamics analysis of fractional-order permanent magnet synchronous motor and its DSP implementation," International Journal of Modern Physics B, vol. 33, no. 06, p. 1950031, 2019, doi: 10.1142/s0217979219500310.

[32] S. He, K. Sun, X. Mei, B. Yan, and S. Xu, "Numerical analysis of a fractional-order chaotic system based on conformable fractional-order derivative," The European Physical Journal Plus, vol. 132, no. 1, 2017, doi: 10.1140/epjp/i2017-11306-3.

[33] R. Khalil, M. A. Horani, A. Yousef, and M. Sababheh, "A new definition of fractional derivative," Journal of Computational and Applied Mathematics, vol. 264, 2014.

[34] J. Ruan, K. Sun, J. Mou, S. He, and L. Zhang, "Fractional-order simplest memristor-based chaotic circuit with new derivative," European Physical Journal Plus, vol. 133, no. 1, p. 3, 2018.

[35] O. Acan and D. Baleanu, "A new numerical technique for solving fractional partial differential equations," Miskolc Mathematical Notes, vol. 19, no. 1, 2018, doi: 10.18514/mmn.2018.2291.

[36] S. Bhalekar, V. Daftardar-Gejji, D. Baleanu, and R. Magin, "Transient chaos in fractional Bloch equations," Computers & Mathematics with Applications, vol. 64, no. 10, pp. 3367-3376, 2013.

[37] Y. Li, Z. Li, M. Ma, and M. Wang, "Generation of grid multi-wing chaotic attractors and its application in video secure communication system," Multimedia Tools & Applications, vol. 79, no. 1, 2020.

[38] H. Wang, K. Sun, and S. He, "Characteristic Analysis and DSP Realization of Fractional-Order Simplified Lorenz System Based on Adomian Decomposition Method," International Journal of Bifurcation and Chaos, vol. 25, no. 06, p. 1550085, 2015, doi: 10.1142/s0218127415500856.

[39] S. Dai, K. Sun, W. Ai, and Y. Peng, "Novel discrete chaotic system via fractal transformation and its DSP implementation," Modern Physics Letters B, p. 2050429, 2020.

[40] C. L. Li, C. Q. Liu, Y. G. He, and X. B. Yang, "Analysis, Stabilization, and DSP-Based Implementation of a Chaotic System with Nonhyperbolic Equilibrium," Complexity, vol. 2020, 2020.

[41] A. Bayani, K. Rajagopal, A. J. M. Khalaf, S. Jafari, G. D. Leutcho, and J. Kengne, "Dynamical analysis of a new multistable chaotic system with hidden attractor: Antimonotonicity, coexisting multiple attractors, and offset boosting," Physics Letters A, vol. 383, no. 13, pp. 1450-1456, 2019, doi: 10.1016/j.physleta.2019.02.005.

[42] E. et al., "Equilibrium points, stability and numerical solutions of fractional-order predator–prey and rabies models - ScienceDirect," Journal of Mathematical Analysis & Applications, vol. 325, no. 1, pp. 542-553, 2007.

A new Lorentz chaotic system and its circuit implementation

Xingce Liu[1], Xiaoyu Zhang[2], Peng Li[1,*], Huizhen Yan[1]

{lipeng@dlpu.edu.cn}

[1]School of Information Science and Engineering, Dalian polytechnic University, Dalian, 116034, China
[2]Agronomy College, ShenYang Agricultural University, Shenyang, 110161, China

Abstract： In this paper, a newfashioned Lorentz chaotic system are found. According to the circuit structure, we set up the dimensionless model of the circuit. In the simulation, three different shapes of chaotic attractors are found. The equilibrium point and stability of chaotic system are analyzed by using the traditional dynamical analysis methods, and the dynamical behaviors with two varying parameters of this circuit are analyzed in detail. In addition, The specific phenomenon of chaotic attractor coexistence is discovered. Finally, the circuit simulation is carried out using Multisim, and the results are compared with the numerical simulation, which proves the validity of the theoretical analysis. Combined with Theoretical analysis and simulation consequence, it is shown that the new chaotic system has abundant dynamic characteristics.

Keywords: Memristor, Lorentzian chaotic system, Attractor coexistence, Multisim simulation

1 Introduction

Chaos is a kind of motion state that exists universally in nature. It has many characteristics worth studying, such as sensitivity to initial conditions, randomness of change and unpredictability of long term behavior. Macroscopically, although it presents an irregular state, its essence is an orderly movement[1,2].If there is to be chaos in a circuit, there must be nonlinear elements. Memristor can be said to be the most representative element of nonlinear elements. In 1971, Professor Chua projected the existence of memristors are based on the completeness principle of variable combinations[3].In 1976, Professor Chua further elaborated the composition principle and application characteristics of memristors, and explained that memristors can be divided into magneto controlled memristors and charger-controlled memristors[4]. Due to no memory element has been found, memristor research has not received much attention from the scientific and engineering circles. In 2008, HP Labs completed the implementation of memristor, and great progress has been made in the research

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307990

of memristor[5-8].

Lorentz system is familiar to most chaos scholars[9-11]..Lorentzian equation was determined by American meteorologist Lorentz when he simulated the weather, which is an aperiodic phenomenon[12]. The three variables of this equation simulate temperature, humidity and pressure respectively[13]. Based on his results, Lorenz concluded that small initial differences would become larger over time, and based on this, he found that long-term weather forecasting was impossible[14]. This is also regarded as the beginning of the study of nonlinear chaotic theory, so Lorentz system performs an increasingly important role in the research of nonlinear science [15-19]. A newfashioned fifth-order chaotic system is established by combining Lorentz system with a new memristor. The dynamic behavior of Lorentz system combined with the new memristor is studied emphatically.

In this system, a new memristor is introduced on the basis of Lorentz system, and several chaotic attractors are found[20].As is known to all, the dynamical behavior of chaotic circuits containing memristors is closely relevant to the initial state of memristors, which may lead to multi-stability or extremely multi-stability of the system[21-23]. Multistable or extremely multi-stable phenomenon refers to the coexistence of multiple attractors or infinitely many attractors, which is a nonlinear dynamics phenomenon with great research significance [24]. It is a kind of dynamic behavior related to initial conditions, which is often encountered in some nonlinear dynamic systems[25]. Since the system parameters aren’t changed and only the initial values are changed, the trajectory of the system can asymptotically approach to a variety of stable states, such as periodic, quasi-periodic, chaotic or hyperchaotic[26]. Multistable or extremely multi-stable nonlinear dynamic systems not only provide great selectivity for their potential engineering applications, but also pose new challenges to control multi-stable states[27].The coexistence behavior of chaotic attractor is found in this system, and the system is discussed deeply[28].

The rest of this article is outlined as follows. In the second section, the model of memristor is introduced. In the third section, the simple chaotic circuit is introduced, the equivalent differential equation of the circuit is introduced, and the dynamic characteristics of the system is analyzed. In the fourth section, the consequence of theoretical analysis are verified on Multisim platform. The fifth section is a summary of the conclusions of this research.

2 The model of memristor

Professor Bao has proposed a new type of magneto controlled memristor. This memristor can describe a smooth quadratic nonlinear characteristic curve, and its memristor’s model is

shown as:

$$\begin{cases} i = W(\varphi)v \\ \dfrac{d\varphi}{dt} = v \\ W(\varphi) = -a + b|\varphi| \end{cases}, \tag{1}$$

Thus, the *v–i* curve can be expressed as:

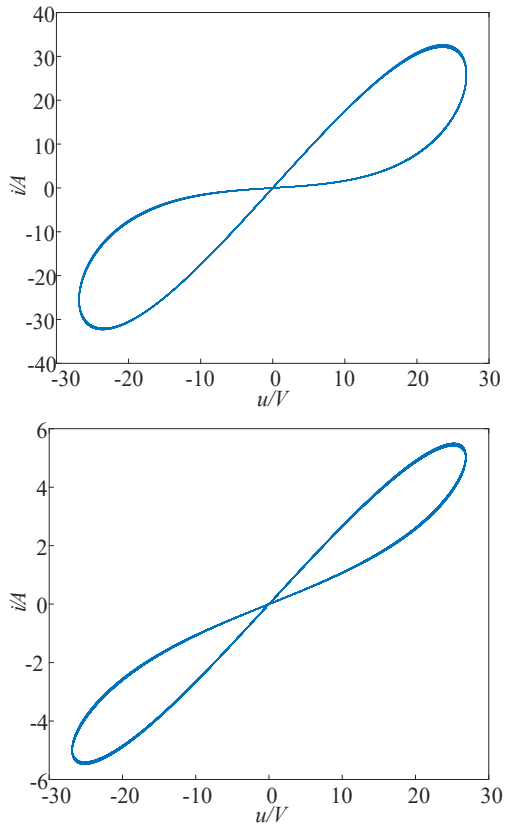
$$\begin{cases} i_m = W(\varphi)v_s \\ \dot{\varphi} = v_s \end{cases}, \tag{2}$$

W(φ) in Equation (1) is the relation of magnetic flux *φ*. The graph of the function is the characteristic curve that goes through the origin. The slope of the curve is called memorialization, which represents the ratio between charge and magnetic flux. The units of memorialization and conductance are Siemens.Memorific can be expressed as:

$$\mathrm{W}(\varphi) = \alpha + \beta|\varphi|, \tag{3}$$

The *α* and *β* in the equation are the coefficients of the memristor. The magnetic flux of *Φ* represents today.

V_S is AC power supply in Equation (2), It can be expressed as: $V_S=Asin(2\pi ft)$, Where A is the swing of the AC power supply, and *f* is the frequency of the AC power supply. The sinusoidal AC supply V_S is the input of the memristor, α=0, β=1 in Equation (3). When A=19 and the frequency *f* is 5Hz and 50Hz respectively, the volt-ampere characteristics of the memristor are shown in Figure 1. When sinusoidal AC excitation is used, the volt-ampere diagram of the memristor is a closed "8" curve through the origin.With the increase of frequency *f*, the area of the side lobe of the curve decreases gradually. This is consistent with the basic properties of memristors.



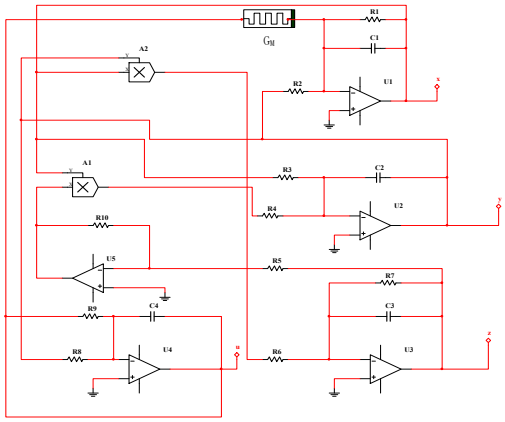
(a) f=5Hz (b) f=50Hz

Fig. 1 Hysteresis curve

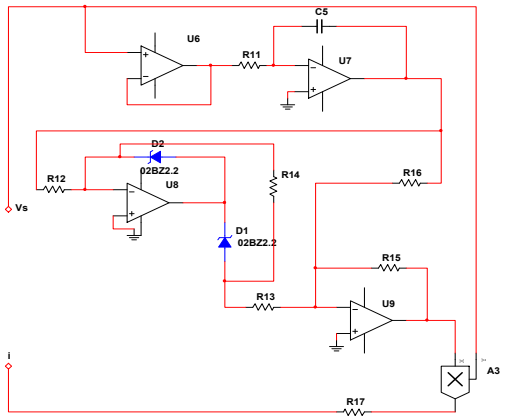
3 A memristor–based novel chaotic circuit

3.1 A novel chaotic circuit

The schematic diagram of the novel Lorentz type chaotic circuit is shown in Figure 2 (a), the G_M in the figure is a new memristor, and its circuit schematic diagram is shown in Figure 2 (b):



(a)



(b)

Fig. 2 (a) the main circuit diagram (b) the memristor circuit

The values of components shown in Fig. 2 circuit are shown in Table 1:

Tab 1. Circuit element value list

name	category	numerical	unit
$R1$，$R2$，	resistance	5	KΩ
$R3$	resistance	1.2	KΩ
$R4$	resistance	0.51	KΩ
$R5$，$R7$，	resistance	10	KΩ
$R8$，$R10$，	resistance	5	KΩ
$R11$	resistance	50	KΩ
$R12$，$R14$，	resistance	10	KΩ
$R16$	resistance	10	KΩ
$R17$	resistance	69	KΩ
$C1$，$C2$，	capacitanc	100	nF
$C4$，$C5$	capacitanc	100	nF

$A1$	multiplier	0.1	\
$A2$，$A3$	multiplier	1	\

The ideal magnetron memristor in the figure is composed of voltage follower, reverse integral circuit, absolute value circuit, multiplier and resistance R. its mathematical model can be easily obtained by analyzing the relationship between input voltage and output current. The relationship between output i and input v_s can be expressed as follows:

$$\begin{cases} i = W\left(v_w\right)v_s = \dfrac{g_1}{R}\left|v_w\right|v_s = \dfrac{k}{R}\left|v_w\right|v_s \\ \dot{v}_w = \dfrac{v_s}{RC_0} \end{cases}, \tag{4}$$

Where V_w is the internal variable of the ideal magnetron memristor, and $g1$ is the gain of the multiplier A_0, where $k=g_1R_1/R_2$, and the memristor $W(v_w) = \frac{kv_w}{R_1}$.

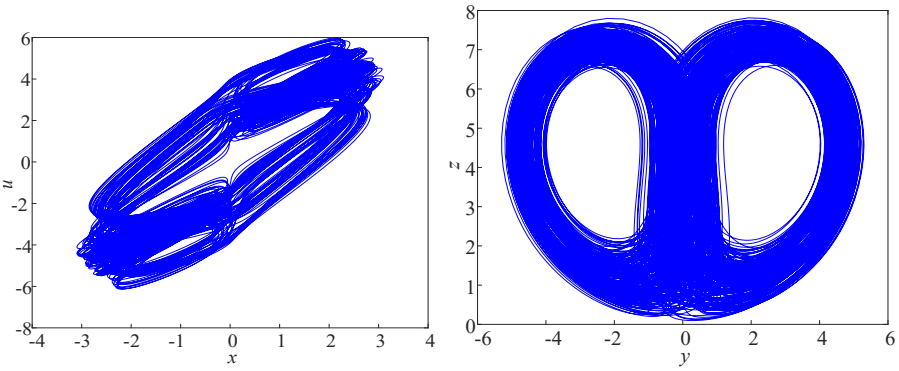
According to Kirchhoff's law of voltage and current and the basic characteristics of each element, the system equation can be obtained for the chaotic circuit shown in Figure 2 (a), which is expressed as:

$$\begin{cases} RC\dfrac{dv_x}{dt} = -\dfrac{Rv_x}{R_1} - \dfrac{Rv_y}{R_2} - \dfrac{g_1}{R_{11}}R\left|v_w\right|v_u \\ RC\dfrac{dv_y}{dt} = \dfrac{g_2R_fRv_zv_x}{R_3R_4} - \dfrac{Rv_x}{R_5} \\ RC\dfrac{dv_z}{dt} = -\dfrac{g_3Rv_xv_y}{R_6} - \dfrac{R}{R_7}v_z \\ RC\dfrac{dv_u}{dt} = -\dfrac{R}{R_8}v_y - \dfrac{R}{R_9}v_u \\ RC\dfrac{dv_w}{dt} = \dfrac{R}{R_{11}}v_u \end{cases}, \tag{5}$$

Where V_x, V_y, V_z, V_u are four circuit variables, V_w is the internal variable of ideal magnetron memristor,$g1,g2,g3$ are the gains of multipliers A1,A2,A3,by dimensionless treatment of the equation, the equation can be expressed as:

$$\begin{cases} \dot{x} = a(x+y) - |w|u \\ \dot{y} = 9.8xz - 45x \\ \dot{z} = -10xy - cz \\ \dot{u} = by - 5u \\ \dot{w} = u \end{cases}, \quad (6)$$

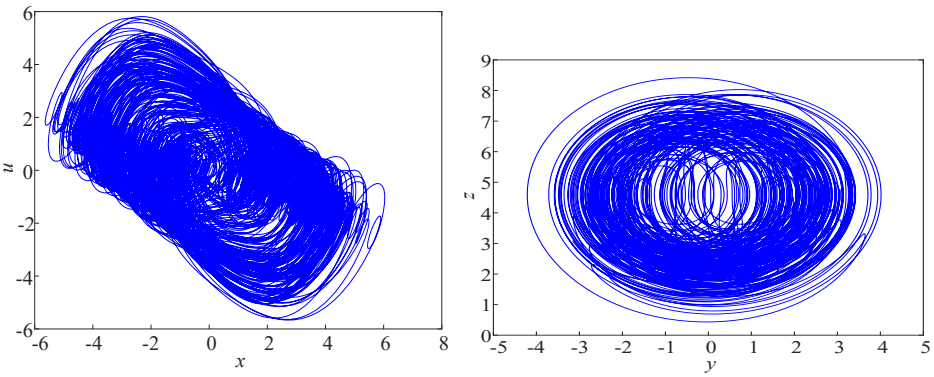
When the parameters of system are chosen to be a=-10, b=-10, c=8, the initial value are chosen to be (1, 1, 0, 0, 0) and the time step is selected 0.01 seconds, the five Lyapunov exponents are expressed as L1=0.6966, L2=0, L3= -0.1671, L4=-2.6321, L5=-20.8969. The three kinds of different chaotic attractors obtained by simulation can be shown in Figure 3.



(a) x-u (b) y-z

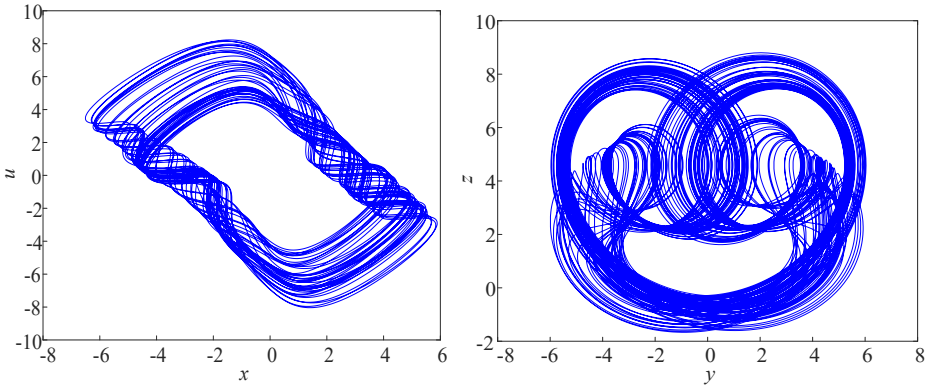
Fig. 3 When a =-10，b =-10，c = 8，chaotic attractor phase diagram

Since the parameters a, b, c are changed and the initial condition and step size are kept unchanged, the following two kinds of chaotic attractor phase diagrams are obtained:



(a) *x*-*u* (b) *y*-*z*

Fig. 4 When a =-5，b =-15，c = 3，chaotic attractor phase diagram



(a) *x*-*u* (b) *y*-*z*

Fig. 5 When a =-10，b =-10，c = 15，chaotic attractor phase diagram

3.2 Equilibrium point and stability

According to the dynamical equations (6) of system, the divergence can be expressed as:

$$\nabla V=\frac{\partial \dot{x}}{\partial x}+\frac{\partial \dot{y}}{\partial y}+\frac{\partial \dot{z}}{\partial z}+\frac{\partial \dot{u}}{\partial u}+\frac{\partial \dot{w}}{\partial w}=a-c-5-1\,, \tag{7}$$

Let the system parameters a=-10, b=-10, c=8, the initial conditions are selected (1, 1, 0, 0, 0) ,the divergence of the system is less than zero, so the system is dissipative, and the system might have chaotic attractors. Let the equations (6) of system be equal to zero, then the equilibrium point E (0, 0, 0, 0, 0) of the system can be obtained,according to the equilibrium point, the Jacobian matrix can be expressed as follows:

$$J=\begin{bmatrix} a & a & 0 & 0 & 0 \\ -45 & 0 & 0 & 0 & 0 \\ 0 & 0 & -c & 0 & 0 \\ 0 & b & 0 & -5 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \tag{8}$$

Its characteristic polynomial can be expressed as follows:

$$P(\lambda)=\lambda^5+13\lambda^4-460\lambda^3+393.5593\lambda^2+18000.0092\lambda\,, \tag{9}$$

According to the characteristic equation, the eigenvalues are λ_1 = 0, λ_2 = 5, λ_3 = - 8, λ_4 = -

26.7945, and $\lambda_5 = 16.7945$. According to the Routh–Hurwitz criterion, the equilibrium point E (0, 0, 0, 0) is unstable.

3.3 The impacts of parameters

This system is a five-dimensional chaotic system. By combining the bifurcation diagram with Lyapunov exponent spectrum, the states of the system with different parameters can be obtained. In the following, the parameters a and c are put as variables, the initial conditions are selected (1, 1, 0, 0, 0) the step size is $h = 0.01$, the remaining parameters of the equation are fixed, and the different states of the system are observed by changing the parameters a and c.

Take parameter $a \in [-10,1]$, let b=-10, c=8, Lyapunov exponent spectrum and bifurcation diagram can be expressed in Figure 6.It can be realized from Fig.6 that when a=-8.25, -8.15 and 0, the maximum Lyapunov exponent L1 is equal to 0, and the system behaves as a limit cycle state. At other times, the L1 is always greater than zreo,L2=0, and other Lyapunov exponents are less than 0, so the system is a chaotic system.

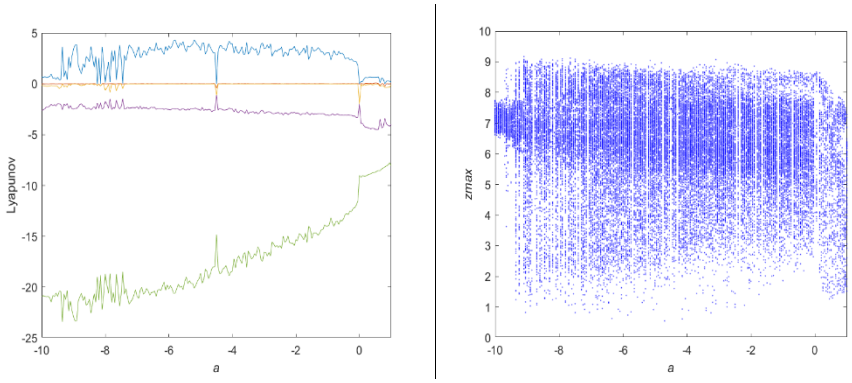
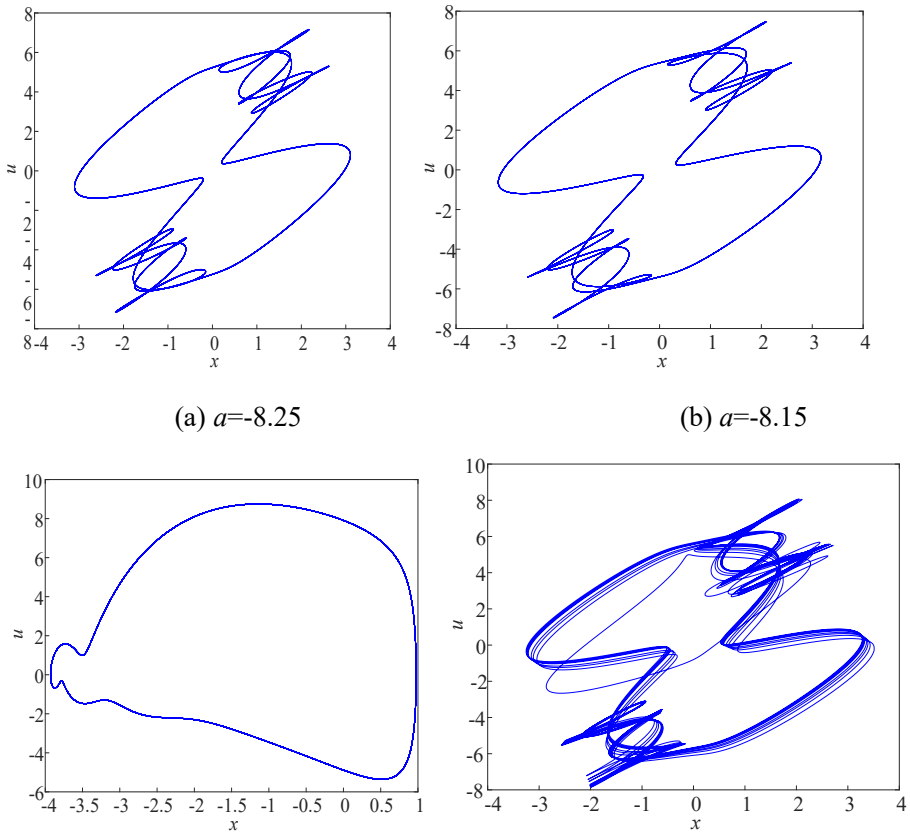


Fig. 6 Lyapunov exponential spectrum and bifurcation diagram of the system when a changes

This result shows the phase diagram of x-u when c=8, b= -10, fixed, a = -8.25, -8.15, 0 and -8, respectively. When a =-8.25, a =-8.15 and a=0, the phase diagram shows a limit cycle state. According to the Lyapunov exponent spectrum, there are four exponents less than 0 and one equal to 0, and the system is in a periodic state; When a =-8, the phase diagram is chaotic. At this time, the other Lyapunov exponents of the system are less than 0, L1>0, and L2=0, which are in chaotic state. The analysis shows that this result is completely corresponding to the bifurcation diagram and Lyapunov exponent spectrum shown above.



(a) a=-8.25 (b) a=-8.15

(c) a=0 (d) a=-8

Fig. 7 The phase diagram of the system with different a

Take the c∈[0, 18], a = -10, b = -10, and the Lyapunov exponent spectrum and bifurcation diagram under this condition are expressed in Figure 8.When c∈[0, 0.55], the L1=0, and the system behaves as a limit cycle. When c=0.55, chaos appears. When c∈ [0.55, 8.85], the L1> 0, and the other exponents is less than 0, then the system shows chaotic state. Since c∈[8.85, 8.9], the maximum Lyapunov exponent L1 is equal to 0, and the system behaves as a limit cycle. When c ∈ [8.9, 16.9], the L1> 0, and the other exponents is less than 0, and the system is in chaotic state. Since c∈[16.9, 18], the L1 is equal to zero and the other items are less than 0, and the system behaves as a limit cycle state.

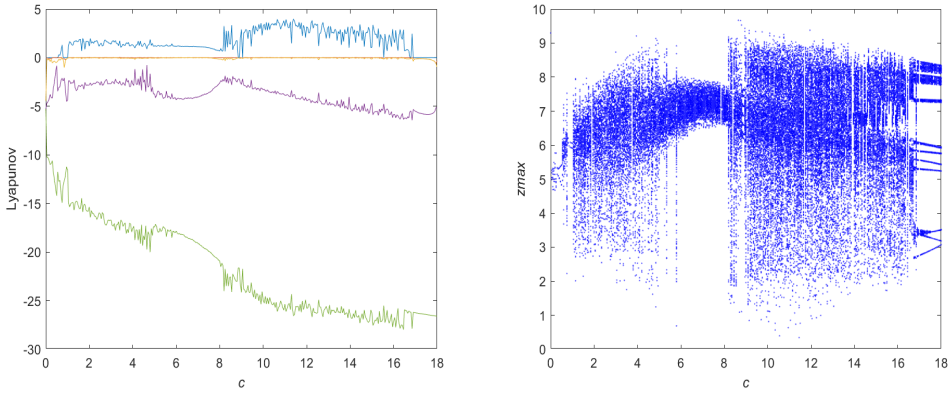
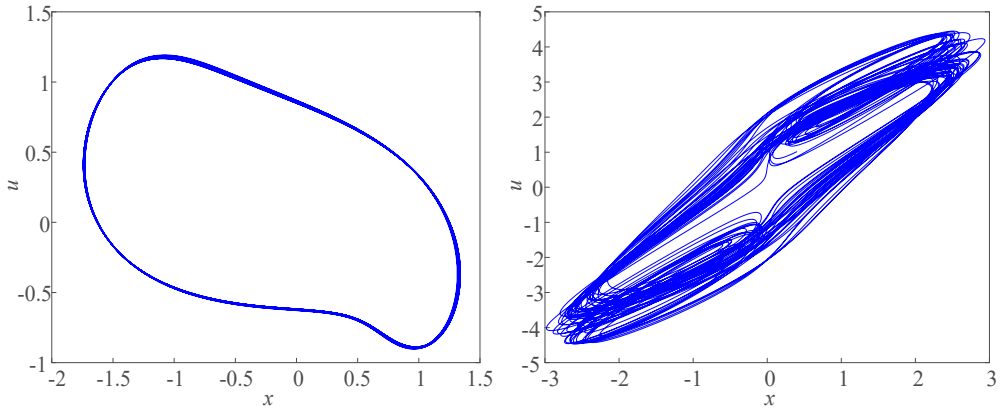
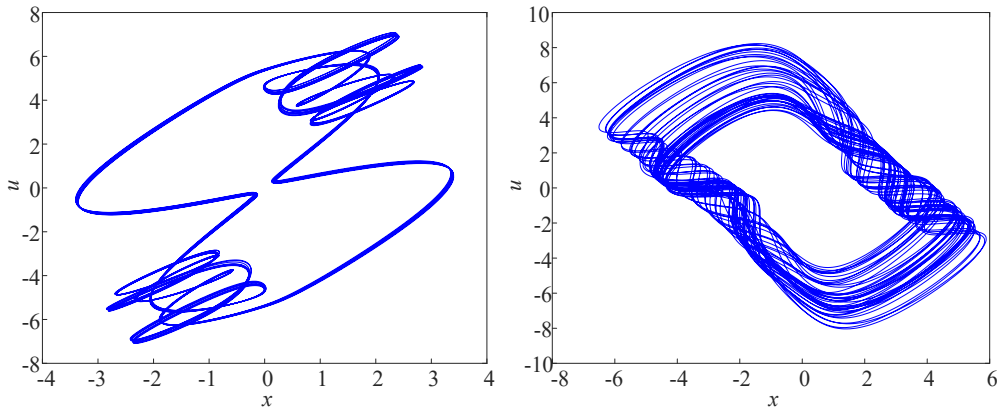


Fig. 8 Lyapunov exponential spectrum and bifurcation diagram of the system when *c* changes

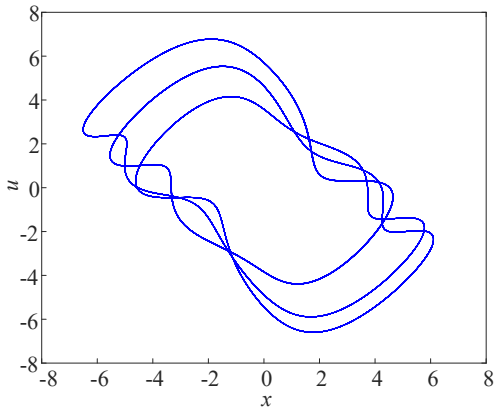
This result shows the phase diagram of *x*-*u* when the parameters *a*=-10, *b*=-10, fixed and c are 0.45, 5, 8.85, 15 and 17.4 respectively. When *c* =0.45, *c*=8.85 and *c*=17.4, the phase diagram is a limit cycle. According to the Lyapunov exponent spectrum, there are four exponents of the system that are less than 0, another is equal to 0, and the system is in cycle state; When *c*=5 and *c*=15, the phase diagram is chaotic. At this time, three Lyapunov exponents of the system are less than 0, one equal to zero and another greater than 0, it means that it is chaotic system. The analysis shows that this result is completely corresponding to the bifurcation diagram and Lyapunov exponent spectrum shown above.



(a) *c*=0.45 (b) *c*=6



(b) c=8.85 (d) c=15



(e) c=17.4

Fig.9 The phase diagram of the system with different c

3.4 Complexity analysis

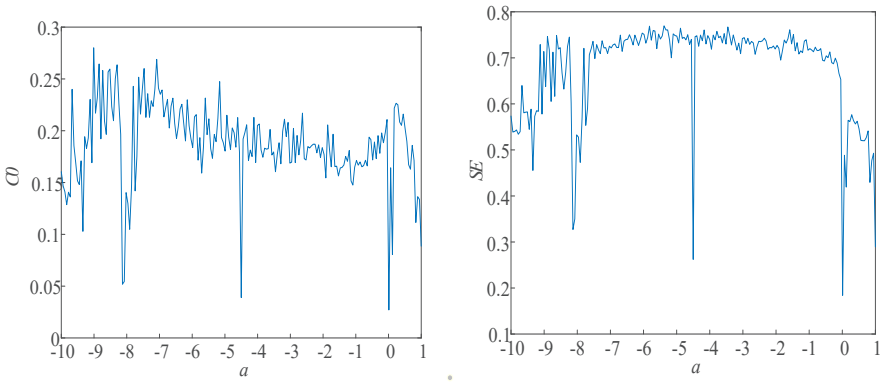
The study of complexity involves various fields, and scholars in different fields have different understandings of complexity. So far, there is no unified concept of complexity. The complexity of chaotic system is to measure the degree of chaotic sequence approaching random sequence by using correlation algorithm. The greater the complexity, the closer the sequence is to the random sequence, the higher the security. In essence, the complexity of chaotic systems belongs to the study of chaotic dynamics.

Analysis of complexity includes behavioral complexityand structural complexity.Behavioral complexity refers to using some method to measure the probability of generating new patterns in a short time window from the chaotic sequence itself.The more likely a new pattern is to be generated, the more complex the sequence. At present, there are

many algorithms to calculate the behavior complexity of chaotic pseudo-random sequences, and they are all based on Kolmogorov method and Shannon entropy. This kind of algorithm has fast calculation speed and accurate results. However, if the selected dimension is too high or the symbol space of pseudo-random sequences is too large, the calculation results will overflow or even get no results. The more balanced the energy spectrum distribution in the transform domain, the closer the original sequence is to the random signal, that is, the greater the complexity of the sequence. Combined with the concept of Shannon entropy, the corresponding spectral entropy can be calculated. The energy characteristics of this region are analyzed by using structural complexity, which is for the whole sequence rather than the local sequence. Therefore, compared with the behavior complexity algorithm, the results obtained are more global. In this paper, the complexity of the system are analyzed by using SE algorithm and CO algorithm.

SE algorithm mainly uses the Fourier transform, through the energy distribution in the Fourier transform domain, combined with Shannon entropy to obtain the spectral entropy value. The gist of C0 algorithm is to decompose the sequence into two parts, regular and irregular, and calculate the irregular part to get the result.

In this part, the complexity of the system is analyzed with parameter c as a variable. When the parameters c =8 , b=-10, fixed, and $a\in$[-10,1], the simulation results are shown in Figure 10.



(a) C0 Complexity (b) SE Complexity

Fig. 10 Structural complexity curve with parameter a

By analyzing the complexity, it can be seen that SE algorithm and C0 algorithm have a high degree of synchronization. According to the image, when a =-8.25, -8.15 and 0, the system shows a limit cycle state. On other scales, the system appears chaotic. By analyzing the SE complexity diagram and the C0 complexity diagram, it can be found that the complexity is

at a low point when the system is in the periodic state, while the complexity increases significantly when the system enters the chaotic state. The results presented by the complexity are consistent with the Lyapunov exponential spectrum and bifurcation diagram as shown in Figure 8.

3.5 Coexistence of attractors

Attractor coexistence is a peculiar phenomenon, which primarily emerges in some peculiar chaotic systems. It is very important in the study of nonlinear systems. Since the system parameters remain unchanged and the original values change, the orbits of the system tend to different states, such as point, quasi periodic, periodic or chaotic. In order to explore this special phenomenon, let a =-10, b =-10, c =5, with the change of initial value, we can observe the change of system state. The results of numerical simulation are shown in Figure 11. The blue part in the figure indicates that the initial value is (1, 2, 3, 4, 10), and the red part indicates that the initial value is (1, 0, 3, 4, 1).

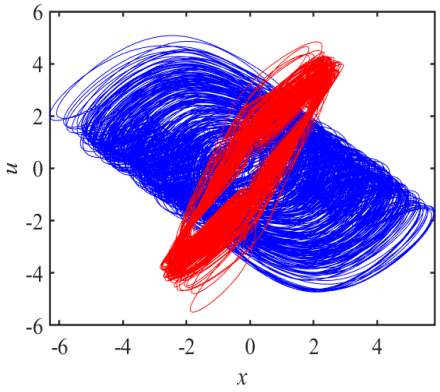
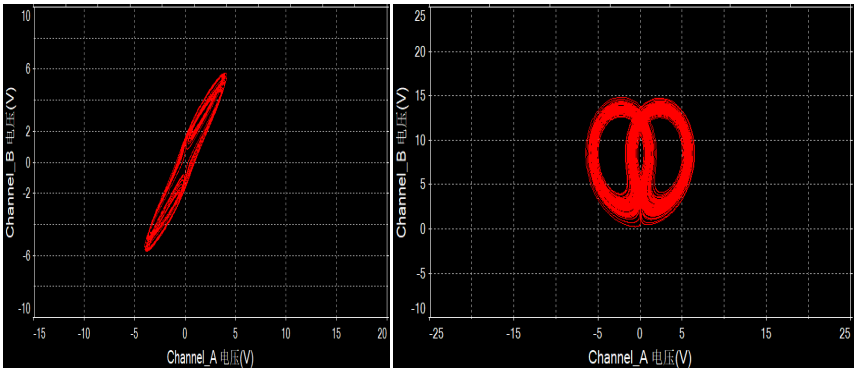


Figure 11. The diagram of attractor coexistence.

4 Multisim simulation

Because chaotic system is easily affected by external disturbance when it is implemented by analog circuit, it is difficult to control the relevant characteristic conditions accurately in the actual circuit. In this section, the system is simulated by using Multisim platform,and the realization of the system is proved.

The simulated electric circuit is illustrated in Figure 2, and the arguments of each component in the circuit are illustrated in Table 1. The corresponding parameters are a=-10, b=-10, and c=8, and the image shown in Figure 12 is obtained.



(a)*x-u* (b)*y-z*

Fig. 12. The phase diagram of Multisim simulation

5 Conclusion

In this article, we devise a newfangled Lorentz type chaotic circuit and introduce its dimensionless equation. Three different kinds of chaotic attractors are found in this chaotic system. By analyzing the Lyapunov exponential spectrum, bifurcation diagram and complexity of the system, it is found that the dynamic characteristics of the new system show a high degree of complexity and sensitivity with the change of system parameters. Through numerical simulation, we observe the special phenomenon of chaotic attractor coexistence. At the same time, some reasonable explanations are given for these phenomena. At last, the circuit simulation is obtained on Multisim platform to prove the consistency between the circuit simulation and the numerical simulation. Due to these rich dynamic behaviors, the novel Lorentz chaotic system has a good application prospect in information encryption and secure communication. Next, we are going to go further, and we will try to go further and explore the applications of new circuits.

Acknowledgements

This work is supported by Scientific Research Funding Project of the Educational Department of Liaoning Province, China (J2020051).

References

[1] Taher A A , Christos V , Gerodimos N A , et al. A Novel Chaotic System without Equilibrium: Dynamics, Synchronization, and Circuit Realization[J]. Complexity, 2017, 2017:1-11.

[2] Ma X , Mou J , Liu J , et al. A novel simple chaotic circuit based on memristor–memcapacitor[J].

Nonlinear Dynamics, 2020, 100(5).

[3]Chua, L.: Memristor-the missing circuit element. IEEE Trans. Circuit Theory 18(5), 507–519 (1971). https://doi. org/10.1109/TCT.1971.1083337

[4] Chua, L.O., Kang, S.M.: Memristive devices and systems. Proc. IEEE 64(2), 209–223 (1976)

[5] Ventra, M.D., Pershin, Y.V., Chua, L.O.: Circuit elements with memory: memristors, memcapacitors, and meminductors. Proc. IEEE 97(10), 1717–1724 (2009)

[6] Biolek, D., Biolkova, V.: Mutator for transforming memristor into memcapacitor. Electron. Lett. 46(21), 1428 (2010). https://doi.org/10.1049/el.2010.2309

[7] Volos, Christos, K, et al. Antimonotonicity, Crisis and Multiple Attractors in a Simple Memristive Circuit[J]. Journal of Circuits Systems & Computers, 2018.

[8] Xia, Q., Robinett, W., Cumbie, M.W., Banerjee, N., Cardinali, T.J., Yang, J.J., Wu, W., Li, X., Tong, W.M., Strukov, D.B.: Memristor-CMOS hybrid integrated circuits for reconfigurable logic. Nano Lett. 9(10), 3640 (2009)

[9] Benkouider K , Bouden T , Halimi M . Dynamical Analysis, Synchronization and Circuit Implementation of a New Hyperchaotic System with Line Equilibrium[C]// 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT). IEEE, 2019.

[10] Chua L O, Komuro M, Matsumoto T. The double scroll family[J]. 1986, 33(11):1072-1118.

[11] Lorenz, Edward N. Deterministic Nonperiodic Flow[J]. Journal of Atmospheric Sciences, 20(2):130-141.

[12] Pehlivan I , Uyaroglu Y . Simplified chaotic diffusionless Lorentz attractor and its application to secure communication systems[J]. Iet Communications, 2007, 1(5):1015-1022.

[13] Kurylev Y , Lassas M , Uhlmann G . Inverse problems for Lorentzian manifolds and non-linear hyperbolic equations[J]. Inventiones Mathematicae, 2018.

[14] Sundarapandian V , Karthikeyan R . Anti-Synchronization of Hyperchaotic Lorenz andHyperchaotic Chen Systems by Adaptive Control[J]. International Journal of Engineering Science & Technology, 2012, 3(5):41-50.

[15] Hu K , Wang Y K , Dai H L , et al. Nonlinear and chaotic vibrations of cantilevered micropipes conveying fluid based on modified couple stress theory[J]. International Journal of Engineering Science, 2016, 105(Aug.):93-107.

[16] Tlidi M , Clerc M G . [Springer Proceedings in Physics] Nonlinear Dynamics: Materials, Theory and Experiments Volume 173 || Hyper-Chaotic and Chaotic Synchronisation of Two Interacting Dipoles[J]. 2016, 10.1007/978-3-319-24871-4(Chapter 20):261-272.

[17] Mahmoud E E , Alá A. Eshmawi. Secure communication and synchronizations in light of the stability theory of the hyperchaotic complex nonlinear systems[J]. Journal of Intelligent and Fuzzy Systems, 2019, 38(3):1-15.

[18] Liu C S . A novel Lie-group theory and complexity of nonlinear dynamical systems[J]. Communications in Nonlinear Science & Numerical Simulation, 2015, 20(1):39-58.

[19] He X , Li C , Huang J , et al. Generalized synchronization of arbitrary-dimensional chaotic systems[J]. Optik - International Journal for Light and Electron Optics, 2015, 126(4):454-459.

[20] Bao H , Hu A , Liu W . Bipolar Pulse-Induced Coexisting Firing Patterns in Two-Dimensional Hindmarsh–Rose Neuron Model[J]. International Journal of Bifurcation and Chaos, 2019, 29(01).

[21] Fang, Yuan, Guangyi, et al. Extreme multistability in a memristor-based multi-scroll hyper-chaotic system.[J]. Chaos, 2016.

[22] Chang H , Li Y , Chen G , et al. Extreme Multistability and Complex Dynamics of a Memristor-Based Chaotic System[J]. International Journal of Bifurcation and Chaos, 2020, 30(08):434-445.

[23] Bao B , Jiang T , Wang G , et al. Two-memristor-based Chua's hyperchaotic circuit with plane equilibrium and its extreme multistability[J]. Nonlinear Dynamics, 2017.

[24] Kengne J . Coexistence of Chaos with Hyperchaos, Period-3 Doubling Bifurcation, and Transient Chaos in the Hyperchaotic Oscillator with Gyrators[J]. International Journal of Bifurcation & Chaos, 2015, 25(4):4285-762.

[25] Ling G , Ge M F , Hu F Y , et al. Coexistence of multi-attractor in small genetic regulatory networks[C]// 2020 39th Chinese Control Conference (CCC). 2020.

[26] Bao B , Xu L , Wu Z , et al. Coexistence of multiple bifurcation modes in memristive diode-bridge-based canonical Chua's circuit[J]. International Journal of Electronics, 2018:00207217.2018.1426122.

[27] Schaller G , Kiesslich G , Brandes T . Counting Statistics in Multi-stable Systems[J]. Physical review. B, Condensed matter, 2010, 81(20):2498-2502.

[28] Chen C , Chen J , Bao H , et al. Coexisting multi-stable patterns in memristor synapse-coupled Hopfield neural network with two neurons[J]. Nonlinear Dynamics, 2019, 95(10).

Analysis of the dynamical characteristics of fractional-order chaotic system and its DSP realization

Xintong Han, Junqiao Liu*, Huizhen Yan
{dalianliujunqiao@126.com}
School of Information Science and Engineering, Dalian polytechnic University, Dalian, 116034, China

Abstract. In this paper, the 4D fractional-order chaotic system is devised, and the accurate numerical solution of the system is obtained based on the CADM algorithm. The dynamical characteristics of the fractional-order chaotic system are analyzed by employing bifurcation diagram, Lyapunov exponential spectrum, 0-1 test and the poincaré section. Meanwhile, the ranges of parameters of the fractional-order chaotic system are given in practical application through the SE and C0 complexity algorithm. The results demonstrate that the fractional-order chaotic system with the complex dynamical behaviors. Finally, digital circuit of the chaotic system is implemented on the DSP platform. This study offers a practical and theoretical basis for the application of fractional-order chaotic system in the safe communication and other fields.

Keywords: fractional-order system, dynamical characteristics, the CADM algorithm, DSP implementation

1 Introduction

In 1695, Leibniz and Lopida proposed the concept of fractional calculus, but for a long time, its practical application background has not been discovered, leading to slow progress in the related research of fractional calculus. However, recent studies have shown that, compared to integer-order calculus, the fractional-order calculus could better describe the objective physical world [1]. Therefore, the research on fractional calculus has become a hot spot. Research on the Chen system [2], Liu system [3] and simplified Lorenz system [4] and other nonlinear chaotic systems found that, compared to integer-order chaotic systems, fractional-order chaotic systems have more complex dynamical characteristics [5-7].

Currently, the solving algorithms of the fractional-order chaotic systems mainly include frequency domain method (FDM) [8], predictive correction method (ABM) [9] and Adomian

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307951

decomposition method (ADM) [10]. These algorithms generally use the Riemann-Liuville definition and Caputo definition. But they all have some problems. First, they cannot satisfy some important properties of the fractional calculus [11], such as the product rule and chain rule. Second, the calculation process is very complicated [12]. Therefore, Khalil et al. proposed the new definition of the conformable fractional calculus [13]. The combination of the definition of the fractional calculus and the ADM algorithm can make up for the shortcomings of the existing algorithms for solving fractional differential equations [12, 14]. The CADM algorithm improves the ADM algorithm, reduces the complexity of calculation, and has the advantages of faster convergence speed, calculation speed, and less resource consumption [15]. At present, the use of this CADM algorithm to resolve the fractional-order chaotic systems is gradually becoming a research hotspot [16].

At present, many studies have been carried out on the application of fractional-order chaotic systems, such as nuclear magnetic resonance [17], cellular neural networks [18], and secure communications [19-21]. But it really applies to hardware and software technologies that rely on signal processing. Digital signal processor (DSP) [22, 23] has been widely used in engineering due to its superior performance and convenient processing. Based on this, this paper uses DSP technology to implement the hardware implementation of the proposed four-dimensional fractional-order chaotic system.

Based on the a four-dimensional Sprott-B chaotic system, the 4-D fractional-order chaotic system is constructed through using the conformable fractional calculus definition. The remaining structure of the paper is as follows. In sect.2, the numerical solution of the 4-D fractional-order chaotic system is solved by the CADM algorithm. The dynamical behaviors of the 4-D fractional-order chaotic system are analysed in sect.3. The 4-D fractional-order chaotic system is achieved on the DSP platform In sect.4. Finally, the experimental research results are summarized and a future direction is proposed.

2 Numerical solution of the fractional-order chaotic system

2.1 CADM algorithm of the fractional-order chaotic system

Assuming that the conformable fractional differential equation is expressed as

$$\begin{cases} T_t^q y(t) = Ly(t) + Ny(t) + g \\ y^{(k)}(t_0^+) = b_k, k = 0, \cdots, m-1, \\ m \in N, m-1 < q \le m \end{cases} \tag{1}$$

where T_t^q is differential operator, $x(t)=[y_1(t)\ y_2(t)\ \cdots\ y_n(t)]^T$ are the state variables. $Lx(t)$ is the linear part of the system, $Ny(t)$ is the non-linear part of the system function. g represents the constant term, and b_k represents the initial value of the chaotic system. According to the conformable fractional calculus definition, one can get

$$y(t) = I_{t_0}^q Ly(t) + I_{t_0}^q Ny(t) + I_{t_0}^q g + y(t_0)\,, \tag{2}$$

here $I_{t_0}^q$ is the integral operator, $y(t_0)$ is the initial value condition of the chaotic system. The nonlinear term can be expressed as

$$Ny = \sum_{i=0}^{\infty} A^i(y^0, y^1, \cdots, y^i)\,, \tag{3}$$

then the nonlinear term can be decomposed by

$$\begin{cases} A_j^i = \dfrac{1}{i!}[\dfrac{d^i}{d\lambda^i} N(v_j^i(\lambda))]_{\lambda=0} \\ v_j^i(\lambda) = \sum\limits_{k=0}^{i} (\lambda)^k y_j^k \end{cases}, \tag{4}$$

where i=0,1, ..., j=1,2, …, n. So, the numerical solution of the equation is

$$y = \sum_{i=0}^{\infty} y^i = I_{t0}^q L \sum_{i=0}^{\infty} y^i + I_{t0}^q \sum_{i=0}^{\infty} A^i + I_{t0}^q \,\mathrm{g} + y(t_0)\,, \tag{5}$$

here y_i is calculated by

$$\begin{cases} y^0 = I_{t_0}^q g + y(t_0) \\ y^1 = I_{t_0}^q Ly^0 + I_{t_0}^q A^0(y^0) \\ y^2 = I_{t_0}^q Ly^1 + I_{t_0}^q A^1(y^0, y^1) \\ \cdots \\ y^i = I_{t_0}^q Ly^{i-1} + I_{t_0}^q A^{i-1}(y^0, y^1, y^2, \cdots, y^{i-1}) \\ \cdots \end{cases}. \tag{6}$$

2.2 Solution of the fractional-order chaotic system

A new four-dimensional autonomous chaotic system on the basis of the proposed Sprott-B system. The new system equation can be expressed as

$$\begin{cases} \dot{x}_1 = a(x_2 - x_1) + x_4 \\ \dot{x}_2 = x_1 x_3 \\ \dot{x}_3 = b - x_1 x_2 \\ \dot{x}_4 = -cx_4 + x_1 x_3 \end{cases}, \tag{7}$$

here x_1, x_2, x_3, x_4 indicates the state variables of the system, a, b and c are positive parameters of the system. Based on conformable derivative, the fractional form of the chaotic system equation is

$$\begin{cases} T_t^q x_1 = a(x_2 - x_1) - x_4 \\ T_t^q x_2 = x_1 x_3 \\ T_t^q x_3 = b - x_1 x_2 \\ T_t^q x_4 = -c x_4 + x_1 x_3 \end{cases}, \tag{8}$$

here q is the order of the system. According to the above formula (2), the linear, non-linear and constant terms are

$$\begin{bmatrix} L_{x_1} \\ L_{x_2} \\ L_{x_3} \\ L_{x_4} \end{bmatrix} = \begin{bmatrix} a(x_2 - x_1) + x_4 \\ 0 \\ 0 \\ -c x_4 \end{bmatrix}, \begin{bmatrix} N_{x_1} \\ N_{x_2} \\ N_{x_3} \\ N_{x_4} \end{bmatrix} = \begin{bmatrix} 0 \\ x_1 x_3 \\ -x_1 x_2 \\ x_1 x_3 \end{bmatrix}, \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \mathrm{b} \\ 0 \end{bmatrix}. \tag{9}$$

Decomposing the Eq. (6) the non-linear terms decompositions are

$$\begin{cases} A_2^0 = x_1^0 x_3^0 \\ A_2^1 = x_1^1 x_3^0 + x_1^0 x_3^1 \\ A_2^2 = x_1^2 x_3^0 + x_1^1 x_3^1 + x_1^0 x_3^2 \\ A_2^3 = x_1^3 x_3^0 + x_1^2 x_3^1 + x_1^1 x_3^2 + x_1^0 x_3^3 \\ A_2^4 = x_1^4 x_3^0 + x_1^3 x_3^1 + x_1^2 x_3^2 + x_1^1 x_3^3 + x_1^0 x_3^4 \\ A_2^5 = x_1^5 x_3^0 + x_1^4 x_3^1 + x_1^3 x_3^2 + x_1^2 x_3^3 + x_1^1 x_3^4 + x_1^0 x_3^5 \end{cases}, \tag{10}$$

$$\begin{cases} A_3^0 = -x_1^0 x_2^0 \\ A_3^1 = -x_1^1 x_2^0 - x_1^0 x_2^1 \\ A_3^2 = -x_1^2 x_2^0 - x_1^1 x_2^1 - x_1^0 x_2^2 \\ A_3^3 = -x_1^3 x_2^0 - x_1^2 x_2^1 - x_1^1 x_2^2 - x_1^0 x_2^3 \\ A_3^4 = -x_1^4 x_2^0 - x_1^3 x_2^1 - x_1^2 x_2^2 - x_1^1 x_2^3 - x_1^0 x_2^4 \\ A_3^5 = -x_1^5 x_2^0 - x_1^4 x_2^1 - x_1^3 x_2^2 - x_1^2 x_2^3 - x_1^1 x_2^4 - x_1^0 x_2^5 \end{cases}, \tag{11}$$

$$\begin{cases} A_4^0 = x_1^0 x_3^0 \\ A_4^1 = x_1^1 x_3^0 + x_1^0 x_3^1 \\ A_4^2 = x_1^2 x_3^0 + x_1^1 x_3^1 + x_1^0 x_3^2 \\ A_4^3 = x_1^3 x_3^0 + x_1^2 x_3^1 + x_1^1 x_3^2 + x_1^0 x_3^3 \\ A_4^4 = x_1^4 x_3^0 + x_1^3 x_3^1 + x_1^2 x_3^2 + x_1^1 x_3^3 + x_1^0 x_3^4 \\ A_4^5 = x_1^5 x_3^0 + x_1^4 x_3^1 + x_1^3 x_3^2 + x_1^2 x_3^3 + x_1^1 x_3^4 + x_1^0 x_3^5 \end{cases} . \tag{12}$$

here, assuming the initial value $x_0=[x_1(0), x_2(0), x_3(0), x_4(0)]$, the first term can be described as

$$\begin{cases} x_1^0 = x_1(t_0) \\ x_2^0 = x_2(t_0) \\ x_3^0 = x_3(t_0) + b\dfrac{h^q}{q} \\ x_4^0 = x_4(t_0) \end{cases} . \tag{13}$$

Let

$$\begin{cases} c_1^0 = x_1^0 \\ c_2^0 = x_2^0 \\ c_3^0 = x_3^0 \\ c_4^0 = x_4^0 \end{cases}, \tag{14}$$

according to the Eq. (6), the second expansion is

$$\begin{cases} x_1^1 = \left[a(c_2^0 - c_1^0) + c_4^0 \right] \dfrac{(t - t_0)^q}{q} \\ x_2^1 = \left[c_1^0 c_3^0 \right] \dfrac{(t - t_0)^q}{q} \\ x_3^1 = \left[-c_1^0 c_2^0 \right] \dfrac{(t - t_0)^q}{q} \\ x_4^1 = \left[-cc_4^0 + c_1^0 c_3^0 \right] \dfrac{(t - t_0)^q}{q} \end{cases} . \tag{15}$$

Let $h=t-t_0$, by the same method, the coefficients of the other five terms are

$$\begin{cases} c_1^2 = a(c_2^1 - c_1^1) + c_4^1 \\ c_2^2 = c_1^1 c_3^0 + c_1^0 c_3^1 \\ c_3^2 = -c_1^1 c_2^0 - c_1^0 c_2^1 \\ c_4^2 = -cc_4^1 + c_1^1 c_3^0 + c_1^0 c_3^1 \end{cases}, \tag{16}$$

$$\begin{cases} c_1^3 = a(c_2^2 - c_1^2) + c_4^2 \\ c_2^3 = c_1^2 c_3^0 + 2c_1^1 c_3^1 + c_1^0 c_3^2 \\ c_3^3 = -c_1^2 c_2^0 - 2c_1^1 c_2^1 - c_1^0 c_2^2 \\ c_4^3 = -cc_4^2 + c_1^2 c_3^0 + 2c_1^1 c_3^1 + c_1^0 c_3^2 \end{cases}, \tag{17}$$

$$\begin{cases} c_1^4 = a(c_2^3 - c_1^3) + c_4^3 \\ c_2^4 = c_1^3 c_3^0 + 3c_1^2 c_3^1 + 3c_1^2 c_3^1 + c_1^0 c_3^3 \\ c_3^4 = -c_1^3 c_2^0 - 3c_1^2 c_2^1 - 3c_1^1 c_2^2 - c_1^0 c_2^3 \\ c_4^4 = -cc_4^3 + c_1^3 c_3^0 + 3c_1^2 c_3^1 + 3c_1^2 c_3^1 + c_1^0 c_3^3 \end{cases}, \tag{18}$$

$$\begin{cases} c_1^5 = a(c_2^4 - c_1^4) + c_4^4 \\ c_2^5 = c_1^4 c_3^0 + 4c_1^3 c_3^1 + 6c_1^2 c_3^2 + 4c_1^1 c_3^3 + c_1^0 c_3^4 \\ c_3^5 = -c_1^4 c_2^0 - 4c_1^3 c_2^1 - 6c_1^2 c_2^2 - 4c_1^1 c_2^3 - c_1^0 c_2^4 \\ c_4^5 = -cc_4^4 + c_1^4 c_3^0 + 4c_1^3 c_3^1 + 6c_1^2 c_3^2 + 4c_1^1 c_3^3 + c_1^0 c_3^4 \end{cases}, \tag{19}$$

$$\begin{cases} c_1^6 = a(c_2^5 - c_1^5) + c_4^5 \\ c_2^6 = c_1^5 c_3^0 + 5c_1^4 c_3^1 + 10c_1^3 c_3^2 + 10c_1^2 c_3^3 + 5c_1^1 c_3^4 + c_1^0 c_3^5 \\ c_3^6 = -c_1^5 c_2^0 - 5c_1^4 c_2^1 - 10c_1^3 c_2^2 - 10c_1^2 c_2^3 - 5c_1^1 c_2^4 - c_1^0 c_2^5 \\ c_4^6 = -cc_4^5 + c_1^5 c_3^0 + 5c_1^4 c_3^1 + 10c_1^3 c_3^2 + 10c_1^2 c_3^3 + 5c_1^1 c_3^4 + c_1^0 c_3^5 \end{cases}, \tag{20}$$

so, the CADM numerical results of the fractional-order chaotic system with six terms are

$$x_j(t) = \sum_{i=0}^{6} c_j^i \frac{(t-t_0)^{iq}}{i!q^i}. \tag{21}$$

3 Dynamical analysis of the fractional-order chaotic system

3.1 Simulation analysis of the system

Setting a=4, b=9, c=5, h=0.01, the system order q=0.8, and the initial values x_0 = [1, 1, 1, 1]. The Lyapunov exponents of the system are L_1=1.0170, L_2=0, L_3=-12.5479, and L_4=-16.7206, and the Lyapunov dimension is D_L= 2.09. At this time, the system has a positive Lyapunov exponent, indicating that the system is chaotic. Fig. 1 shows the phase diagrams in different planes of the system. To further verify the dynamical characteristics of the system under the parameters, the Poincaré section of the system at this time are shown in Fig. 2. The Poincaré section in the figure is neither a finite point set nor a closed curve. It is a piece of dense points with a fractal structure. This structure has the typical characteristics of a chaotic system.

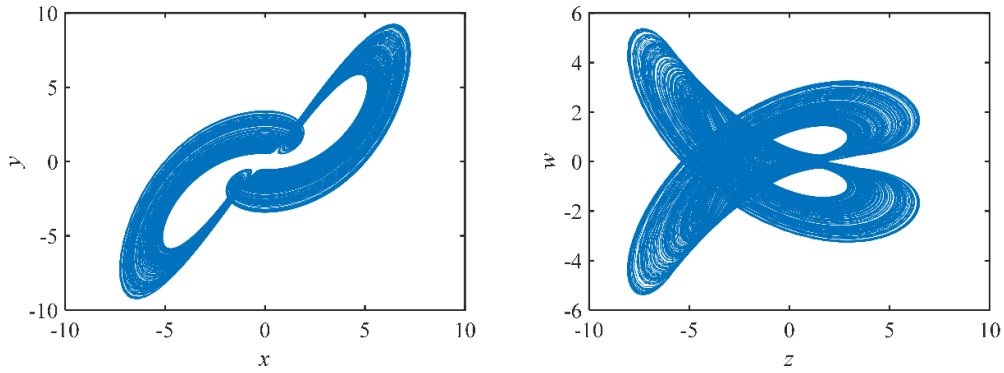


Fig. 1 The phase diagrams of system with q=0.8

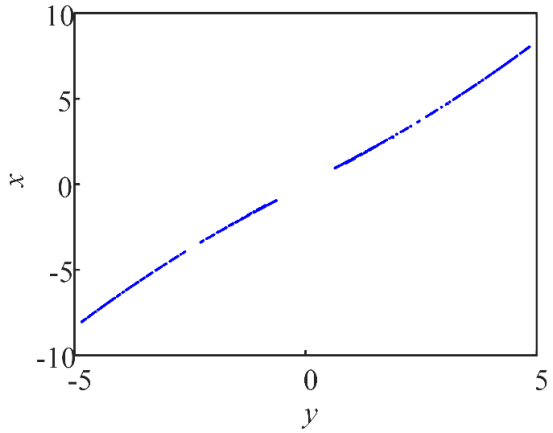


Fig. 2 The Poincaré section of system (8) on the y-x plane when q = 0.8

3.2 Stability analysis of the system

For Eq. (7), let $\dot{x}_1=\dot{x}_2=\dot{x}_3=\dot{x}_4=0$, one can get the equilibrium point of the system $S_{1,2}$= (± √ b, ± √ b, 0, 0). The Jacobian matrix J for the system (Eq. (7)) at the plane equilibrium $S_{1,2}$ is obtained as

$$J=\begin{pmatrix} -a & a & 0 & 1 \\ 0 & 0 & \pm\sqrt{b} & 0 \\ \mp\sqrt{b} & \mp\sqrt{b} & 0 & 0 \\ 0 & 0 & \pm\sqrt{b} & -c \end{pmatrix}, \tag{22}$$

when a=4, b=9, and c=5, the equilibrium point of the system is $S_{1,2}$= (±3, ±3, 0, 0). The

characteristic polynomial equation is yielded as

$$\lambda^{32}+9\lambda^{24}+29\lambda^{16}+126\lambda^{8}+360=0\,. \tag{23}$$

If the eigenvalue satisfies

$$\left|\arg(\lambda)\right|\geq\frac{\pi}{2M}\,, \tag{24}$$

the system is gradually stable at the equilibrium point. Substitute eigenvalues into the Eq. (24),

$$0.05\pi-\min_{i}\left\{\left|\arg(\lambda_i)\right|\right\}=-0.0229<0\,, \tag{25}$$

thus, at the equilibrium point $S_{1,2}$, the system is stable.

3.3 Dynamical characteristics with the parameter varying

3.3.1 Dynamical evolution with parameter *a*

Setting the parameters b=9, c=5, the order q=0.8, h=0.01, and x_0= [1 1 1 1]. Fig. 3 indicates the bifurcation diagrams and the Lyapunov exponents spectrum when a changes from1 to 5. When a belong to [2.98,3.03], [3.46,3.5], [4.27,4.33], [4.34,5] and so on, the largest Lyapunov exponent of the system is zero, so when the parameter a is within these ranges, the system is in the periodic state. However, in other ranges, the largest Lyapunov exponent is positive, and the system is in the chaotic state. Complex dynamical characteristics of the system with the parameter a varying are summarized in the Table 1.

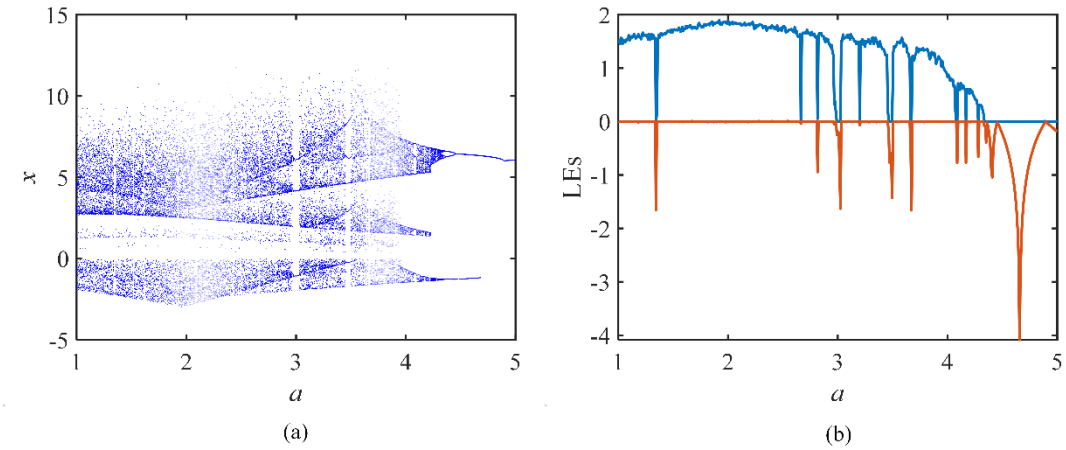


Fig. 3 The system varies with the parameter a (a)Lyapunov exponent spectrum; (b) Bifurcation diagram.

Table 1. System status with the parameter *a*

a	dynamical characteristic	*a*	dynamical characteristic
1~2.97	Chaotic attractor	4.18~4.20	Chaotic attractor
2.98~3.03	Period-2 state	4.21~4.22	Period-7 state
3.04~3.45	Chaotic attractor	4.23~4.27	Chaotic attractor
3.46~3.5	Period-3 state	4.28~4.33	Period-3 state
3.51~3.65	Chaotic attractor	4.33~4.34	Chaotic attractor
3.66~3.67	Period-2 state	4.34~4.36	Period-4 state
3.68~4.15	Chaotic attractor	4.36~4.45	Period-2 state
4.16~4.17	Period-5 state	4.46~5	Limit cycle

From the Table 1, when *a* from 1 to 5, a typical chaotic attractor and six different types of periodic states appear in the system. In the Fig.4, it can be found that when the chaotic system is in a periodic state, there is a bounded regular motion on the *p*-*s* plane; when the system is in the chaotic state, the *p*-*s* plane is an unbounded motion similar to Brownian motion.

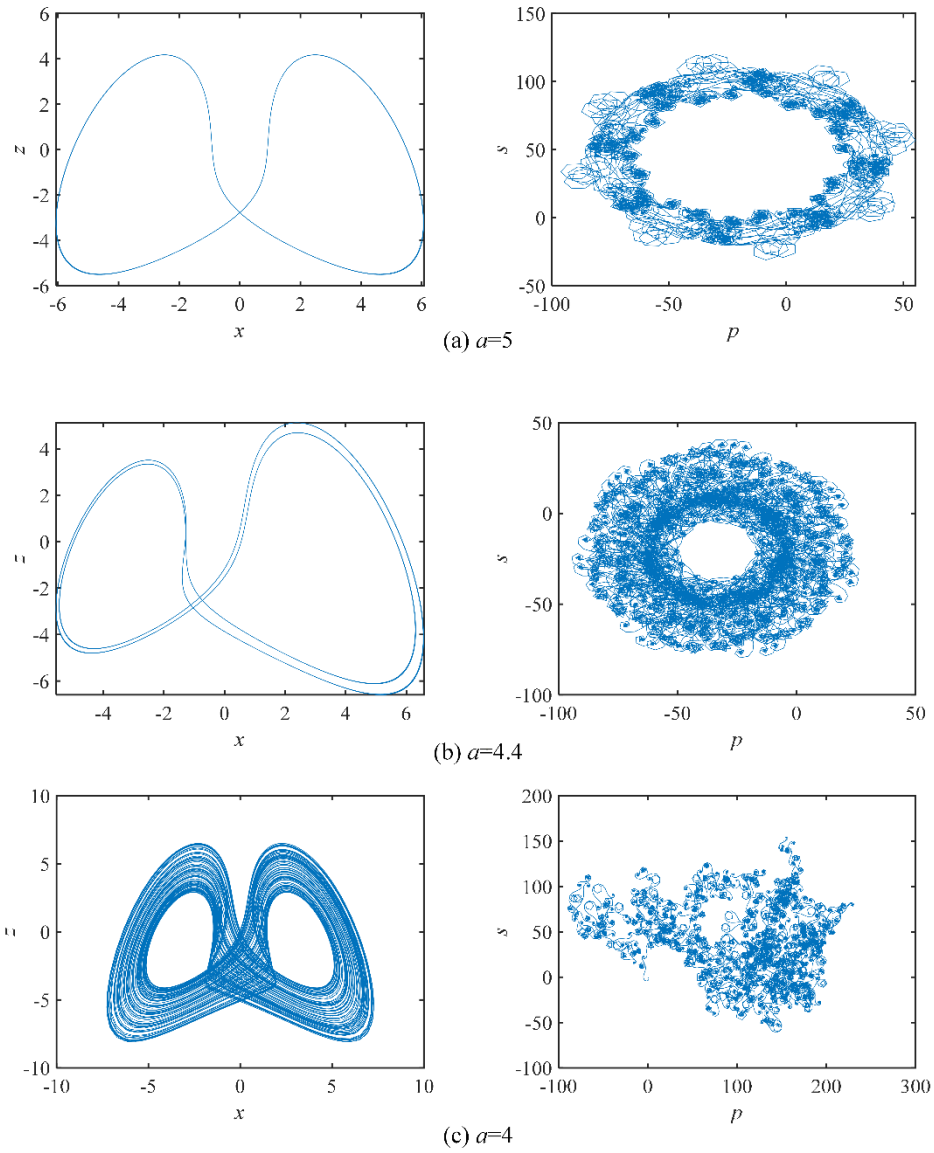


Fig. 4 Phase diagram and *p*-*s* plane with different values of *a*.

3.3.2 Dynamical evolution with the parameter *b*

Setting a=4, c=5, h=0.01, the order q=0.8, and x_0= [1 1 1 1]. Fig.5 indicates the Lyapunov exponents spectrum and the bifurcation diagrams when the parameter b varies in the range of (7, 11). The results can be found from the bifurcation diagram that as the parameter b

increases, the system enters the chaotic state from a periodic state through the period-doubling bifurcation. At the same time, when the parameter b is at [8.53,8.59] and [10.21,10.25], there are two relatively obvious period window, which corresponds to the Lyapunov exponential spectrum. When the parameter b varying, dynamical characteristics of the system are shown in the Table 2.

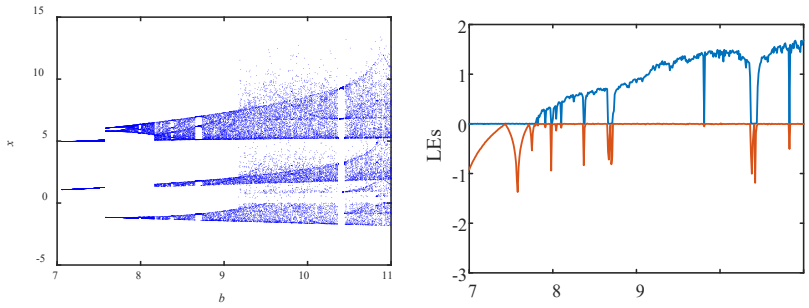


Fig. 5 The system with the parameter b varies. (a) Lyapunov exponent spectrum (b) Bifurcation diagram

Table 2. System status with the parameter b

b	dynamical characteristic	b	dynamical characteristic
7~7.33	Limit cycle	8.27~8.52	Chaotic attractor
7.34~7.72	Period-2 state	8.53~8.59	Period-3 state
7.73~7.8	Period-4 state	8.60~9.77	Chaotic attractor
7.81~7.97	Chaotic attractor	9.78~9.80	Period-3 state
7.8~8.01	Period-3 state	9.81~10.20	Chaotic attractor
8.02~8.24	Chaotic attractor	10.21~10.25	Period-2 state
8.25~8.26	Period-5 state	10.26~11	Chaotic attractor

3.3.3 Dynamical evolution with the order q

Let a=4, b=9, c=5, h=0.01, the x_0= [1 1 1 1]. When $q\in$[0.5,1], the bifurcation diagram and the Lyapunov exponents spectrum are shown in the Fig.6. From Fig.6(a), when $q\in$[0.53,1], the system is in a chaotic state. When q<0.52, there is no value in the Lyapunov exponent spectrum and bifurcation diagram, and the chaotic system is in the divergent state.

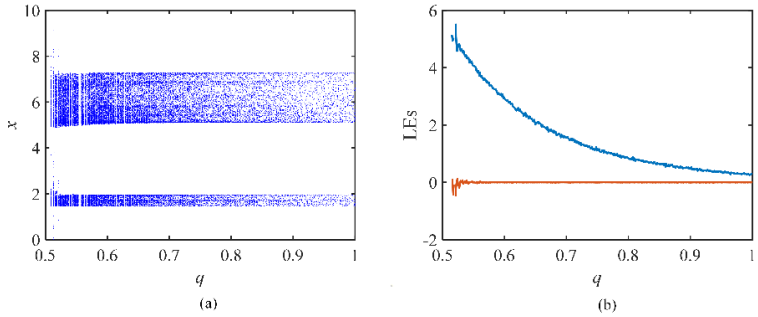


Fig. 6 The system with the parameter q varies. (a) Lyapunov exponent spectrum (b)Bifurcation diagram

To further analyze the system is applicable to the field of secure communication parameter selection range, observing better dynamics chaotic system, where SE and C0 algorithm [24-26] of the system are analyzed. Let the order $q \in [0.5,1]$, sequence length N=50000, SE and C0 complexity of the system are shown in the Fig.7. When $q > 0.52$, the change trend of SE complexity and C0 complexity is gradually reduced. When q=1, the value of complexity reaches the minimum. Therefore, the fractional-order chaotic system is more complicated than the integer-order chaotic system. As the order decreases, the system complexity will increase correspondingly, which indicates that the fractional-order state of the system has higher application value than the integer-order state.

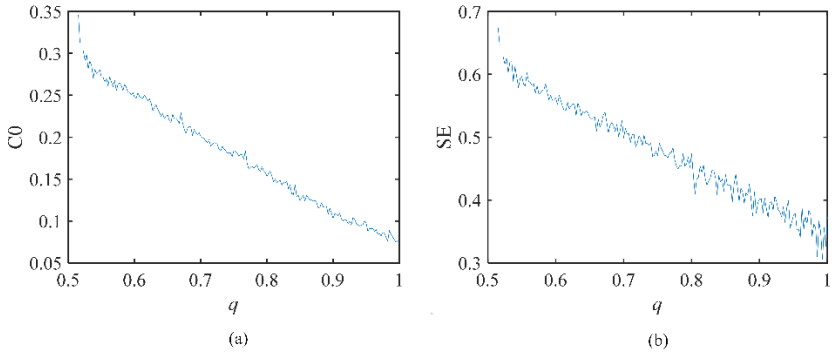


Fig. 7 Complexity varies with the parameter q (a) C0 complexity (b)SE complexity.

4 Hardware circuit based on DSP

In this section, the digital circuit fractional chaotic systems are constructed and implemented on a DSP platform. DSP processor is widely used in engineering practice, due to its powerful performance and low price. Where, the processing speed of the DSP chip TMS320F28335 is accelerated, the precision is higher, and the reliability is higher. TMS320F28335 16-bit DSP is

a kind of floating-point arithmetic operation at 150MHz. The DSP board is connected to 16 pairs of interface channels by DAC DAC8552 SPI (Serial Peripheral Interface). The program flow for DSP implementation and the DSP experimental platform of the fractional-order chaotic system are shown in the Fig. 8 and Fig. 9. Setting a=4, b=9, c=5, h=0.01, q=0.8, and x_0= [1 1 1 1]. The experimental results of the DSP platform are shown in the Fig. 10. Experimental results with the corresponding numerical results agree well in Fig. 1. That is to say, the 4-D fractional-order chaotic system has been perfectly tested on the DSP platform.

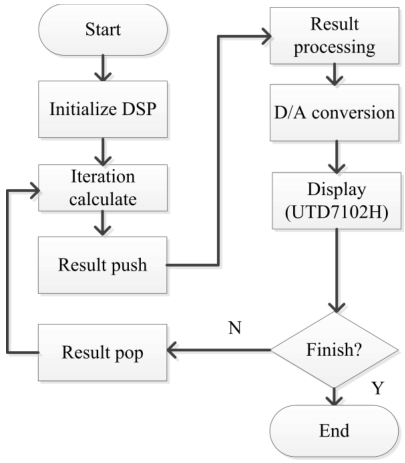


Fig. 8 The program flow for DSP implementation

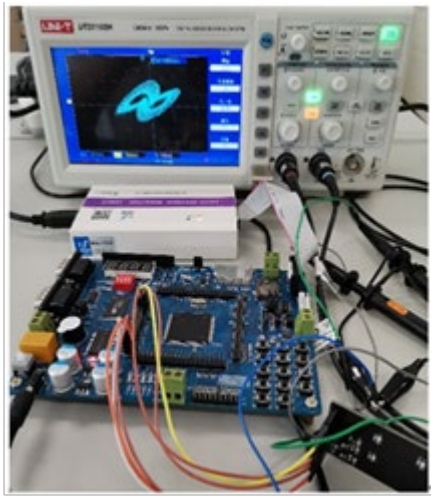


Fig. 9 The DSP experimental platform

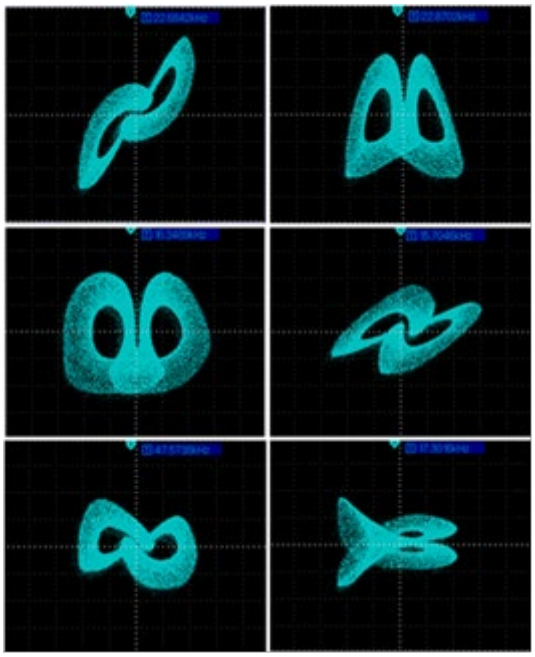


Fig. 10 The DSP implementation of the system

5 Conclusion

In this paper, the 4-D fractional-order chaotic system is designed. The numerical solutions of the fractional-order chaotic system are solved by using the CADM algorithm. Meanwhile, the dynamical behaviors of the system were analyzed through phase diagrams, bifurcation diagrams and Lyapunov exponents. The simulation results show that the fractional-order chaotic system has complex dynamics. The minimum order of chaos in the fractional-order chaotic system is 2.12. In addition, the analysis of complexity shows that, when q=0.53, system complexity is the highest, the randomness of the chaotic sequence is the best and the safety performance is the highest. Finally, the system is implemented by using the DSP platform. The experimental results reflect the correctness of the CADM algorithm and the physical feasibility of the fractional-order chaotic system. The results for the system described herein provides a theoretical basis in the application of engineering.

References

1. Tavazoei, M.S., et al., *Some Applications of Fractional Calculus in Suppression of Chaotic Oscillations.* IEEE Transactions on Industrial Electronics, 2008. **55**(11): p. 4094-4101.
2. A, C.L. and G.C. B, *Chaos in the fractional order Chen system and its control.* Chaos, Solitons

& Fractals, 2004. **22**(3): p. 549-554.

3. Lu, J.J. and C.X. Liu, *Realization of fractional-order Liu chaotic system by circuit.* Chinese Physics B, 2007. **16**(6): p. 1586.
4. Sun, K., X. Wang, and J.C. Sprott, *Bifurcations and Chaos in fractional-Order Simplified Lorenz System.* International Journal of Bifurcation & Chaos, 2010. **20**(04): p. 1209-1219.
5. Shaobo, et al., *Solution and dynamics analysis of a fractional-order hyperchaotic system.* Mathematical Methods in the Applied Sciences, 2016.
6. Zhang, L., et al., *Solution and dynamics of a fractional-order 5-D hyperchaotic system with four wings.* European Physical Journal Plus, 2017. **132**(1): p. 31.
7. Yu, F., et al., *Dynamic Analysis, Circuit Design, and Synchronization of a Novel 6D Memristive Four-Wing Hyperchaotic System with Multiple Coexisting Attractors.* Complexity, 2020. **2020**.
8. Charef, A., et al., *Fractal system as represented by singularity function.* IEEE Transactions on Automatic Control, 2002. **37**(9): p. 1465-1470.
9. Sun, H.H., A.A. Abdelwahab, and B. Onaral, *Linear approximation of transfer function with a pole of fractional power.* Automatic Control IEEE Transactions on, 1984. **29**(5): p. 441-444.
10. Adomian, G., *A review of the decomposition method and some recent results for nonlinear equations.* Mathematical & Computer Modelling, 1991. **13**(7): p. 17-43.
11. He, S., et al., *Numerical analysis of a fractional-order chaotic system based on conformable fractional-order derivative.* European Physical Journal Plus, 2017. **132**(1): p. 36.
12. Abdeljawad, T., *On conformable fractional calculus*. 2015: Elsevier Science Publishers B. V.
13. Khalil, R., et al., *A new definition of fractional derivative.* Journal of Computational and Applied Mathematics, 2014. **264**(5): p. 65–70.
14. Gzütok, N.Y. and U. Gzütok, *Multivariable Conformable Fractional Calculus.* Filomat, 2018. **32**(2): p. 45-53.
15. Allahviranloo, T., N. Ahmady, and E. Ahmady, *Numerical solution of fuzzy differential equations by predictor–corrector method.* Information Sciences, 2007. **177**(7): p. 1633-1647.
16. Nadia, et al., *A conformable fractional calculus on arbitrary time scales.* Journal of King Saud University Science, 2016.
17. Bhalekar, S., et al., *Transient chaos in fractional Bloch equations.* Computers & Mathematics with Applications, 2013. **64**(10): p. 3367-3376.
18. elik, V. and Y. Demir, *Chaotic Fractional Order Delayed Cellular Neural Network.* 2010.
19. Kiani-B, A., et al., *A chaotic secure communication scheme using fractional chaotic systems based on an extended fractional Kalman filter.* Communications in Nonlinear Science & Numerical Simulation, 2009. **14**(3): p. 863-879.
20. Xu, C., J. Sun, and C. Wang, *An Image Encryption Algorithm Based on Random Walk and*

Hyperchaotic Systems. International Journal of Bifurcation and Chaos, 2020.

21. Lan, R., et al., *Integrated Chaotic Systems for Image Encryption.* Signal Processing, 2018. **147**(JUN.): p. 133-145.
22. Penaud, S., et al., *DSP implementation of self-synchronised chaotic encoder-decoder.* Electronics Letters, 2002. **36**(4): p. 365-366.
23. Kharel, R., et al. *Implementation of a secure digital chaotic communication scheme on a DSP board*. in *International Symposium on Communication Systems*. 2010.
24. Ye, X., et al., *Complexity Analysis of a Mixed Memristive Chaotic Circuit.* Complexity, 2018. **2018**: p. 1-9.
25. Ye, X., et al., *Dynamic Characteristic Analysis for Complexity of Continuous Chaotic Systems Based on the Algorithms of SE Complexity and C0 Complexity.* 2017.
26. Xiaolin Ye, et al., *Analysis of continuous chaotic complexity based on SE and C0 algorithm.* Journal of Dalian Polytechnic University, 2018. **037**(001): p. 67-72.

Characteristic analysis of fractional-order simplest chaotic circuit based on Adomian decomposition method

Haiying Hu[1], Yinghong Cao[2], Huizhen Yan[3]

{cncaoyinghong@dlpu.edu.cn}

School of Information Science and Engineering, Dalian polytechnic University, Dalian, 116034, China

Abstract: In this paper, a fractional order chaotic system is reconstructed in the simplest memristive chaotic circuit. The solution of this system is obtained through ADM. Phase diagrams, LEs, bifurcation diagrams and complexity are used in the dynamics study of this fractional chaotic system. At the same time, the stability of the chaotic circuit is concluded and its stability region is given. Meanwhile, the digital circuit of the system was designed and verified on the DSP board. Research results indicate that the ADM algorithm can accurately analyze and calculate effective numerical solutions of fractional-order chaotic systems. According to dynamic analysis, we can get that the system has complex dynamic behavior. This article provides a new direction for the study of the simplest fractional-order memristive chaotic circuits, and provides guidance for the application of proposed systems in the actual field.

Key words: ADM decomposition algorithm; fractional-order memristive chaotic circuit; dynamical characteristic; DSP implementation

1. Introduction

In 1971, Professor Cai reported the memristor for the first time and analyzed the relevant content of the memristor [1], and elaborated on the physical characteristics, basic principles and applications of memristive in 1976 [2]. For a long time, since the actual components that meet the characteristics of the memristive have not been discovered, the research on the memristive

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307855

has not yet attracted scholars' attention.. It was not until 2008 that HP Labs reported the hardware implementation of the memristor for the first time [3], which made the research of memristive aroused widespread concern all over the world.

Memristive usually includes two types of typical nonlinear elements: charge-controlled memristive and magnetron memristive. Because memristive have memory characteristics, and non-linear circuits based on memristor elements are more likely to produce chaotic signals. Scholars have been focusing on the design and implementation of memristive chaotic circuits and have made great progress [4-11]. Fernando et al. analyzed the dynamical behaviors of one kind of memristor oscillators [4]. A simplest chaotic fractional memristor was proposed by Lin et al. [5]. Bao et al. constructed some new circuits based on the smooth memristive model and found that their dynamic behavior is closely related to the initial conditions [6]. The six synchronization algorithms for chaotic systems based on memristor is presented in Ref. [7]. Ye et al. [8, 9] reached the complexity of the hybrid memristive chaotic circuit and analyzed a novel 5-D hyperchaotic memristive system. Most of the existing researches on memristive chaotic circuits are aimed at chaotic systems of integer order.

However, many references indicate that fractional-order chaotic systems have better performance chaotic sequences, higher complexity and richer dynamics, which are more suitable for the field of secure communication [12-19]. Therefore, the relevant research on the dynamics of fractional-order chaotic circuits with memristor is curious. It is very meaningful to research the dynamics of fractional-order chaotic circuits based on a simplest memristive chaotic circuit.

For discretizing fractional-order chaotic systems, the existing references mostly use frequency domain approximation methods[20], predictor-corrector method, and Adomian decomposition method (ADM) [21]. Among them, ADM has the advantages of better accuracy of calculation results and faster calculation speed, and is more suitable for solving fractional differential equations than other decomposition methods [24, 25]. Therefore, we choose ADM algorithm to solve the chaotic system.

In this paper, a fractional-order chaotic equation is constructed based on the circuit containing the simplest memristor and its dynamic behavior is analyzed. Its organizational structure: In Section 2, the memristor is introduced and the simplest memristive circuit model is proposed. In Section 3, the ADM algorithm discretizes the simplest fractional order chaotic circuit system equation and obtains its numerical solution. In Section 4, the dynamics of the system are studied. Design and implement the digital circuit of the system on the DSP platform. Finally, some experimental conclusions are summarized.

2. Simplest Memristive Chaotic Circuit

2.1 Introduction to the memristive model

The memristor model is used to illustrate the connection between magnetic charge q and flux φ. The q and φ passing through the memristive can be represented by a curve $f(\varphi, q)=0$ in the rectangular coordinate system. If the magnetic flux φ plays a dominant role, then the memristor is called a magnetron memristor, and $W(\varphi)=dq(\varphi)d\varphi$, here $W(\varphi)$ is called its meminductor value, and the volt-ampere characteristic between the current flowing through the element and the voltage at both ends is $i(t)=w(\varphi)u(t)$; If the amount of charge in the memristor plays a leading role, then the memristor is a charge-controlled memristor, and $M(q)=d\varphi(q)/dq$, here $M(q)$ is called its memristive value, the volt-ampere characteristic between the current flowing pass the memristive and the voltage across it is $u(t)= M(q)i(t)$. Memristive usually has basic criteria such as passivity, closure, existence, uniqueness, and complexity. At present, the real realization of the memristor is still at the experimental level, so the research on the memristor at this stage is mainly aimed at its equivalent circuit model or mathematical model.

2.2 The memristive circuit model□

Figure 1 plotted the simplest memristive circuit model.

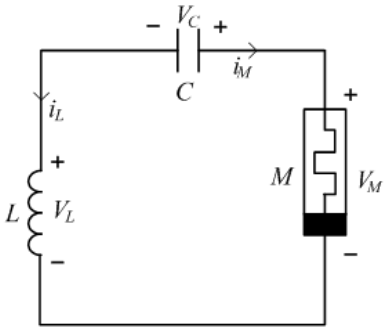


Fig. 1 Simplest memristive circuit

where M, V_M, V_L, V_C, i_M and i_L is the state variables. The memristor model can be expressed as

$$\begin{cases} V_M = \beta(z^2-1)i_M \\ \dot{z} = i_M - \alpha z - z i_M \end{cases} \tag{1}$$

The circuit model can be obtained by equations (2).

$$\begin{cases} \mathrm{C}\dfrac{dV_C(t)}{dt} = i_L(t) \\ L\dfrac{di_L(t)}{dt} = -(V_C(t) + \beta(z^2(t)-1)i_L(t)) \\ \dfrac{dz(t)}{dt} = i_L(t) - \alpha z(t) + i_L(t)z(t) \end{cases} \tag{2}$$

Let $1/C=c$, $1/L=d$, $V_C(t)=x$, $i_L(t)=y$, $z(t)=z$, the normalization operation of Eq. (2) is:

$$\begin{cases} \dot{x} = cy \\ \dot{y} = -d(x + \beta z^2 y - \beta y), \\ \dot{z} = -y(-\alpha + y)z \end{cases} \tag{3}$$

here the system parameters are c, α, β, d, set $\alpha=0.6$, $c=1$, $\beta=1.5$, $d=1/3$, and the initial values are [0.1, 0, 0]. The Lyapunov exponent can be calculated as follows: $L_1=0.045$, $L_2=0$, $L_3=-0.603$. A Lyapunov exponent is greater than 0 proves that the system is chaotic with this parameter.

Combining the definition of Caputo's fractional differential with the Eq. (3), a fractional system equation is reconstructed. The expression equation is as follows

$$\begin{cases} {}^*D_{t_0}^q x = cy \\ {}^*D_{t_0}^q y = -d(x + \beta(z^2 - 1)y), \\ {}^*D_{t_0}^q z = -y - \alpha z + yz \end{cases} \tag{4}$$

Among them t_0 is the initial value, ${}^*D_{t_0}^q$ is the Caputo operator, q is the order ($0<q\leq 1$).

3. The solution of chaotic system

3.1 ADM decomposition algorithm

A fractional differential equation ${}^*D_{t_0}^q(t)=f(x(t))$, $x(t)=[x_1(t), x_2(t), x_3(t), \ldots, x_n(t)]^T$ is a function variable, and ${}^*D_{t_0}^q$ denotes q-order Caputo differential operator, then the system can expressed by

$$\begin{cases} {}^{*}D_{t_0}^{q}x(t) = Nx(t) + Lx(t) + g(t) \\ x^{(k)}(t_0^{+}) = b_k, k = 0,1,...,m-1 \\ g(t) = [g_1(t), g_2(t), \text{K}, g_n(t)]^T \end{cases}, \tag{5}$$

here, N represents the nonlinear term of the differential equation. L is the linear part of the Eq. (1), $g(t)$ is constant of the autonomous system, and b_k is given initial value. After performing operations of Eq. (4), we can get

$$x = J_{t_0}^{q}Nx + J_{t_0}^{q}Lx + J_{t_0}^{q}g + \sum_{k=0}^{m-1} b_k \frac{(t-t_0)^k}{k!} , \tag{6}$$

here, $J_{t_0}^{q}$ is the R-L fractional integral operator of order q (q⩾0). The basic features of this operator are

$$J_{t_0}^{q}(t-t_0)^{\gamma} = \frac{\Gamma(l+1)}{\Gamma(l+1+q)}(t-t_0)^{l+q} , \tag{7}$$

$$J_{t_0}^{q}C = \frac{C}{\Gamma(q+1)}(t-t_0)^{q} , \tag{8}$$

$$J_{t_0}^{q}J_{t_0}^{r}x(t) = J_{t_0}^{q+r}x(t) , \tag{9}$$

where C is a constant, l≥-1, r≥0, t∈$[t_0,t_1]$,. As a nonlinear term, N needs to be further decomposed. The decomposition rules are as follows

$$\begin{cases} A_j^i = \frac{1}{i!}[\frac{d^i}{d\lambda^i}N(v_j^i(\lambda))]_{\lambda=0} \\ v_j^i(\lambda) = \sum_{k=0}^{i}(\lambda)^k x_j^k \end{cases}, \tag{10}$$

Among them j=1, 2, 3, …, n, i=0, 1, 2, 3, … , ∞. The nonlinear term is

$$Nx = \sum_{i=0}^{\infty} A^i(x^0, x^1, ..., x^i) . \tag{11}$$

From equation (11), the solution of equation (5) is

$$\begin{cases} x^0 = J_{t_0}^q g + \sum_{k=0}^{m-1} b_k \dfrac{(t-t_0)^k}{k!} \\ x^1 = J_{t_0}^q L x^0 + J_{t_0}^q A^0(x^0) \\ \cdots \\ x^i = J_{t_0}^q L x^{i\text{-}1} + J_{t_0}^q A^{i\text{-}1}(x^0, x^1, \ldots, x^{i\text{-}1}) \\ \cdots \end{cases} . \tag{12}$$

3.2 Solution of Fractional Memristive Chaotic System

The iterative algorithm of the chaotic system is:

$$\begin{cases} x_{m+1} = x_m + c y_m \dfrac{h^q}{\Gamma(q+1)} + cd(x_m - \beta y_m + \beta y_m z_m^2) \\ \quad \dfrac{h^{2q}}{\Gamma(2q+1)} + \ldots \\ y_{m+1} = y_m + c(x_m - \beta y_m + \beta y_m z_m^2) \dfrac{h^q}{\Gamma(q+1)} \\ \quad + (cdy_m - d^2 \beta x_m + \ldots) \dfrac{h^{2q}}{\Gamma(2q+1)} + \ldots \\ z_{m+1} = z_m + (y_m z_m - y_m - \alpha z_m) \dfrac{h^q}{\Gamma(q+1)} + (z_m b x_m \\ +\ldots) \quad \dfrac{h^{2q}}{\Gamma(2q+1)} + \ldots \end{cases} . \tag{13}$$

Here, the step size is set *h*. the iteration algorithm are

$$\begin{cases} C_{10} = x_m \\ C_{20} = y_m \\ C_{30} = z_m \end{cases} . \tag{14}$$

$$\begin{cases} C_{11} = cC_{20} \\ C_{21} = dC_{10} - d\beta C_{20} + d\beta C_{20}C_{30}^2 \\ C_{30} = -C_{20} - \alpha C_{30} + C_{20}C_{30} \end{cases}. \tag{15}$$

$$\begin{cases} C_{12} = cC_{21} \\ C_{22} = dC_{11} - d\beta C_{21} + d\beta(C_{21}C_{30}^2 + 2C_{20}C_{30}C_{31}) \\ C_{32} = -C_{21} - \alpha C_{31} + C_{21}C_{30} + C_{20}C_{31} \end{cases}. \tag{16}$$

$$\begin{cases} C_{13} = cC_{22} \\ C_{23} = dC_{12} - d\beta C_{22} + d\beta(C_{22}C_{30}^2 + 2C_{20}C_{30}C_{32}) \\ \quad + d\beta(2C_{21}C_{30}C_{21} + C_{20}C_{31}^2)\dfrac{\Gamma(2q+1)}{\Gamma^2(q+1)} \\ C_{33} = -C_{22} - \alpha C_{32} + C_{22}C_{30} + C_{20}C_{32} + C_{21}C_{31} \\ \quad \dfrac{\Gamma(2q+1)}{\Gamma^2(q+1)} \end{cases}. \tag{17}$$

$$\begin{cases} C_{14} = cC_{23} \\ C_{24} = dC_{13} - d\beta C_{23} + d\beta(C_{23}C_{30}^2 + 2C_{20}C_{30}C_{33} + \\ \quad (2C_{22}C_{30}C_{31} + 2C_{21}C_{30}C_{32} + 2C_{20}C_{31}C_{32}) \\ \quad \dfrac{\Gamma(3q+1)}{\Gamma(q+1)\Gamma(2q+1)} + C_{21}C_{31}^2\dfrac{\Gamma(3q+1)}{\Gamma^3(q+1)}) \\ C_{34} = -C_{23} - \alpha C_{33} + C_{23}C_{30} + C_{20}C_{33} + (C_{21}C_{32} \\ \quad + C_{22}C_{31})\dfrac{\Gamma(3q+1)}{\Gamma(q+1)\Gamma(2q+1)} \end{cases}. \tag{18}$$

$$\begin{cases} C_{15} = cC_{24} \\ C_{25} = dC_{14} - d\beta C_{24} + d\beta(C_{24}C_{30}^2 + 2C_{20}C_{30}C_{34} \\ \quad +(2C_{23}C_{30}C_{31} + 2C_{21}C_{30}C_{33} + 2C_{20}C_{31}C_{33}) \\ \quad \dfrac{\Gamma(4q+1)}{\Gamma(q+1)\Gamma(3q+1)} + (2C_{22}C_{30} + C_{20}C_{32}^2)\dfrac{\Gamma(4q+1)}{\Gamma^2(2q+1)} \cdot \\ \quad +(C_{22}C_{31}^2 + 2C_{21}C_{31}C_{32})\dfrac{\Gamma(4q+1)}{\Gamma(2q+1)\Gamma^2(q+1)}) \\ C_{35} = -C_{24} - \alpha C_{34} + C_{24}C_{30} + C_{20}C_{34} + (C_{23}C_{31} \\ \quad +C_{21}C_{33})\dfrac{\Gamma(4q+1)}{\Gamma(q+1)\Gamma(3q+1)} + C_{22}C_{32}\dfrac{\Gamma(4q+1)}{\Gamma^2(2q+1)} \end{cases} \tag{19}$$

3.3 Solution of the Lyapunov exponents spectra

Based on the QR decomposition and iterative formula, the algorithm of Lyapunov exponential spectrum calculation of chaotic system is designed. And QR decomposition is

$$\begin{aligned} & qr[J_h ... J_1] \\ & = qr[J_h J_{h-1} ... J_2 (J_1 Q_0)] \\ & = qr[J_h J_{h-1} ... J_3 (J_2 Q_1)][R_1] \\ & = qr[J_h J_{h-1} ... (J_i Q_{i-1})][R_{i-1} R_{i-2} ... R_2 R_1] . \\ & = ... \\ & = Q_h [R_h ... R_1] \\ & = Q_h R \end{aligned} \tag{20}$$

where QR decomposition process is represented by $qr[\cdot]$. Jacobian matrix is expressed by J, and h represents the number of iteration. Then we get the LE spectrum

$$\lambda_k = \frac{1}{mh}\sum_{i=1}^{m} \ln\left|R_i(k,k)\right|, \tag{21}$$

here k is the system dimension.

4. Analysis of dynamic characteristics of chaotic system and DSP realization

4.1 Stability analysis

Let $-b(x+\beta(z^2-1)y)=0$, $ay=0$, $y-\alpha z+yz=0$. Obviously, there is an equilibrium point, which is $x=y=z=0$. The Jacobian matrix is

$$J_{(P_0)}=\begin{bmatrix} 0 & a & 0 \\ -b & b\beta & 0 \\ 0 & 1 & -\alpha \end{bmatrix}. \tag{22}$$

In the equilibrium point, $(\lambda+\alpha)(\lambda^2-b\beta\lambda+ab)=0$ is the characteristic polynomial of the Eq. (22), and the characteristic root of the characteristic polynomial is

$$\begin{cases} \lambda_1=-\alpha \\ \lambda_2=(b\beta+j\sqrt{4ab-b^2\beta^2})/2 \\ \lambda_3=(b\beta-j\sqrt{4ab-b^2\beta^2})/2 \end{cases} \tag{23}$$

According to the chaotic system equilibrium point analysis theorem[16], the system stability can be analyzed in detail as follows.

Theorem 1: If all the characteristic roots of a chaotic system satisfy $|\arg(\mathrm{eig}(J))|=|\arg(\lambda_i)|>q\pi/2$ ($i=1, 2, 3, \ldots, n$), it indicates the system is gradually stable.

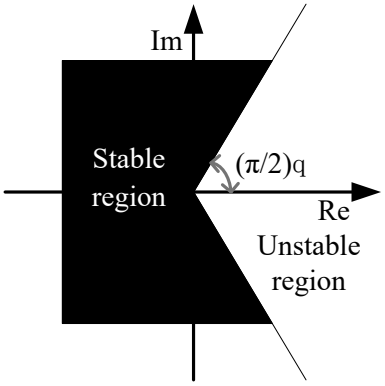


Fig.2 Unstable and stable regions in f chaotic system

It can be seen from Fig. 2 that all the characteristic roots in the black area satisfy $q\pi/2<|\arg(\lambda_i)|$, which means that system is stable. In the white area, all the characteristic roots

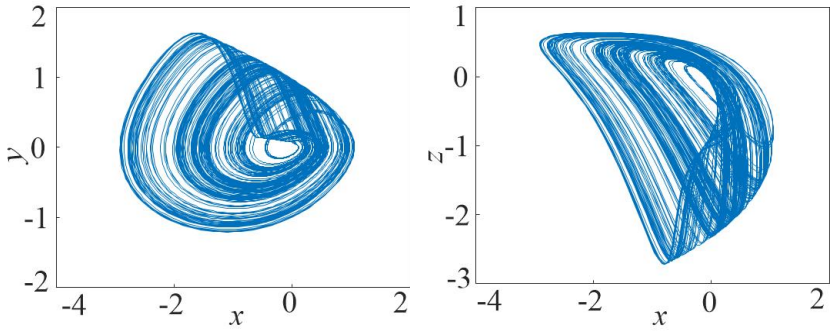
satisfy $|\arg(\lambda_i)|\leq q\pi/2$, and the system is unstable at this time.

For q=1, a system with integer order is obtained. If only consider the case of $q_1=q_2=q_3=q$, that is, all variables in the system have the same order, then the following theorem can be used to analyze the system state [16].

Theorem 2: Assume that the unstable characteristic root of the rolling saddle point is $\lambda_{2.3}=r_{2.3}\pm j\omega_{2.3}$. If and only if the characteristic root $\lambda_{2.3}$ is located in the unstable region. This is a chaotic system. The equal order in the fractional-order system satisfies $q > (\pi/2)a\tan(|\omega_{2.3}|/r_{2.3})$.

4.2 Phase diagram of the system

Set $[x_0, y_0, z_0]$ = [0.1, 0, 0], d=1/3, c=1, β=1.5, α=0.6, h=0.01, and q=0.545. Fig. 3 plots the phase diagram of the system. The L_1=0.346, L_2=0, L_3=-8.166. We can see that the largest Lyapunov exponent is greatly improved compared to the integer order state.

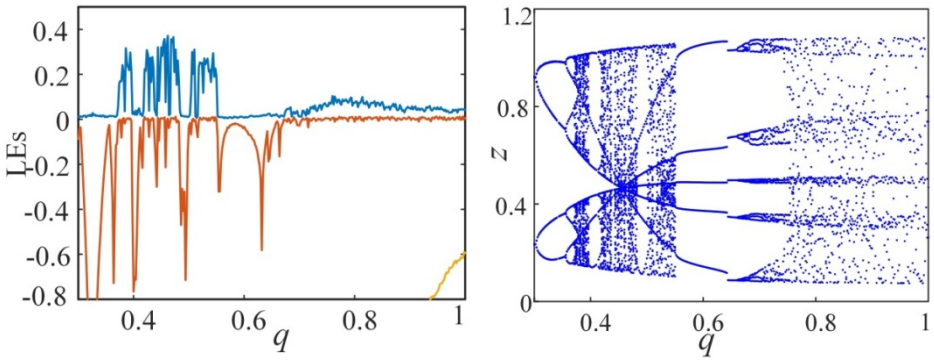


(a) (b)

Fig.3 Phase diagram of the system (a) x-y plane (b) x-z plane

4.3 LEs and bifurcation diagram of chaotic system

Taking the initial value is [0.1, 0, 0] and c=1, d=1/3, α=0.6, β=1.5, h=0.01. When $q\in$[0.3, 1], Fig. 4 depicts the LEs and bifurcation diagrams of the chaotic system.



(a) (b)

Fig.4 LEs and bifurcation diagram of q (a) LEs (b) Bifurcation diagram

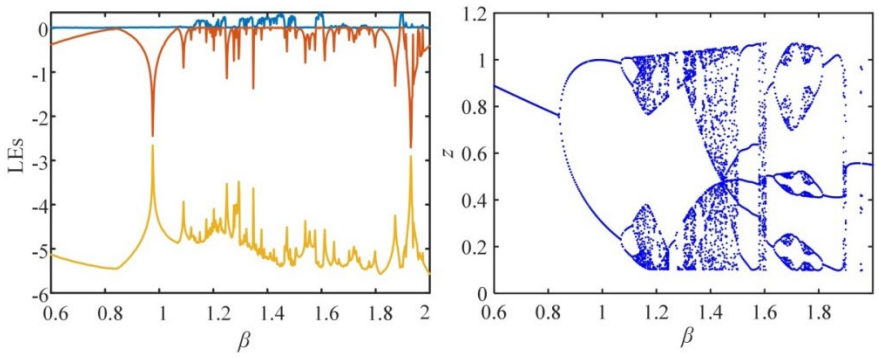
The analysis found that the performance of the fractional-order system has been greatly improved. In addition, the smallest order of chaos in this system is q=0.38×3=1.14. Table 1 lists the state of the system under different q.

Table 1. Dynamic characteristics when the order q changes

q	Dynamical characteristics	q	Dynamical characteristics
0~0.3	divergence	0.45~0.48	chaotic state
0.3~0.34	4 cycle	0.49~0.5	7 cycle
0.35~0.37	8 cycle	0.51~0.55	chaotic state
0.38~0.39	chaotic state	0.56~0.57	5 cycle
0.4~0.41	6 cycle	0.68~0.69	chaotic state
0.42~0.43	chaotic state	0.7	15 cycle
0.44	8 cycle	0.71~1	chaotic state

When q is less than 0.3, the system is divergent. When q is between 0.3 and 1, the system has typical chaotic attractors and many different periodic states.

Let q=0.545 and fix other parameters unchanged. By changing the parameter β, the LEs and bifurcation diagram are plotted in Fig. 5. From Fig. 5, when β takes different values, the LEs obtained correspond to the state of the bifurcation diagram.



(a) (b)

Fig.5 LEs and bifurcation diagram of β (a) LEs (b) Bifurcation diagram

Table 2 shows the state of the system when β changes. It can be seen from Table 2 that when $\beta \in [0.6, 2]$, the system has a classic chaotic attractor and seven different periodic states, and typical periodic doubling branches and anti-periodic branches appear in the process.

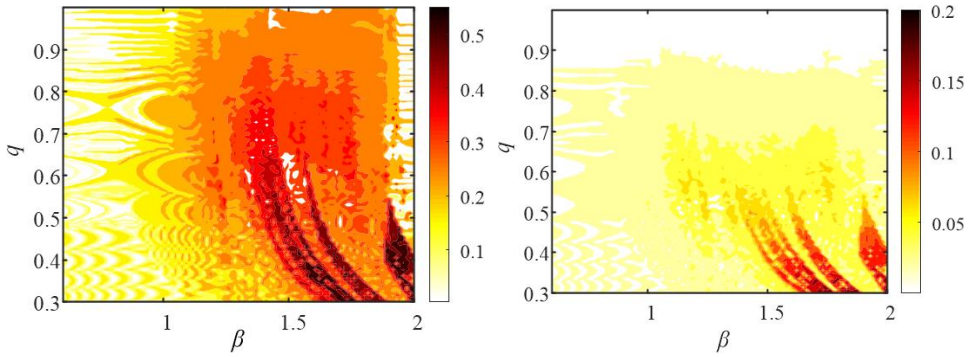
Table 2. Dynamic characteristics when the parameter β changes

β	Dynamical behaviors	β	Dynamical behaviors
0.6~0.9	1 cycle	1.45~1.47	6 cycle
0.91~1.14	2 cycle	1.48	chaotic attractor
1.15~1.28	4 cycle	1.49	8 cycle
1.29~1.32	2 cycle	1.5~1.56	chaotic attractor
1.32~1.33	4 cycle	1.57~1.61	5 cycle
1.33~1.36	8 cycle	1.62~1.67	chaotic attractor
1.36	chaotic attractor	1.68~1.71	3 cycle
1.37	8 cycle	1.72~1.82	6 cycle
1.38	chaotic attractor	1.83~1.88	3 cycle
1.39~1.42	4 cycle	1.89~1.97	chaotic attractor
1.43~1.44	chaotic attractor	1.98~2	Limit cycle

4.4 Complexity of chaotic systems

Complexity is a significant indicator for measuring chaotic pseudo-random sequences. Random performance is better for sequences with high complexity. When applying chaotic sequences to these fields, it is necessary to select sequences with higher randomness and stronger complexity. At present, the commonly used methods for dynamic analysis of chaotic systems include LEs and bifurcation diagram. The two methods can qualitatively evaluate the system's feature, but cannot quantitatively reflect the complexity and randomness of the chaotic sequence. Therefore, spectral entropy (SE) and C_0 complexity are selected to evaluate the randomness of the chaotic system, so that chaotic sequences with higher complexity are selected when the system is applied to chaotic image encryption.

Set c=1, d=1/3, α=0.6, step h=0.01, and [x_0, y_0, z_0] = [0.1, 0, 0]. Based on complexity algorithm, when q and β change at the same time, the complexity of the chaotic system is displayed in Fig. 6.



(a) (b)

Fig.6 Complexity of the chaotic system when q and β change simultaneously (a) SE complexity (b) C_0 complexity

From Fig. 6, the darker colors represent higher levels of complexity. From the Fig.6, when SE is 0.55 the system is the most complex, in this case order q=0.41, the parameter β=1.9, the maximum C_0 complexity is 0.21, the order q=0.33, and the parameter β=1.78. The two complexity changes in Fig. 6 (a) and (b) have the same trend, but the specific values are quite different. Because the spectral entropy (SE) complexity is calculated according to the energy distribution of the system; and the C_0 complexity is Remove the regular part in the sequence, and then analyze and calculate the irregular part; the two algorithms are different, which leads to deviations in specific values. However, both of these two complexity algorithms can express the dynamics of the system and the process of entering chaos, and the trends of the two are consistent. In summary, when β=1.94, 0.4<q<0.5, the complexity and randomness of the system are better. Therefore, when we apply the system, the chaotic sequence in this area should be selected as much as possible.

4.5 Randomness of chaotic system

One of the methods to effectively check the sequence's close to random sequence is the NIST test, which contains 15 performance indicators to check the randomness of the chaotic sequence. Generally speaking, the result meets two criteria, that is, the sequence passed the NIST test.

The first criterion is that the test result obtains the P value, which intuitively reflects whether the sequence is uniformly distributed, and is calculated according to the following

formula:

$$\begin{cases} \chi^2 = \sum_{i=1}^{10} \frac{(F_i - 0.1m)^2}{0.1m} \\ P - value = \text{igmac}(4.5, \frac{\chi^2}{2}) \end{cases}, \tag{24}$$

here F_i is the number of P-value, m represents how many groups. If the P-value>0.01, the randomness of the reconstructed fractional chaotic system has passed the randomness test and has good performance.

Another criterion is the pass rate of passing the test, that is, the proportion of the sequence that passes the test in the entire sequence. We believe that when the following formula is

$$1 - \alpha \pm \sqrt{\frac{\alpha(1-\alpha)}{m}}, \tag{25}$$

where $m \geqslant 1000$ and the value of α is 0.01. The chaotic sequence detection report is compared with the reference [26] in Table 3.

Table 3. NIST test

Test items	Ours		Ref [26]	
	P-value	Pass rate	*P*-value	Pass rate
Frequency	0.275709	0.98	0.867692	0.99
Block Frequency	0.224821	0.98	0.77918	1
Cumulative Sums	0.275709	0.98	0.739918	0.99
Runs	0.249284	0.99	0.779188	0.98
Longest Run	0.350485	0.99	0.055361	1
Rank	0.055361	0.98	0.474986	0.99
FFT	0.851383	1	0.062821	1
Non Overlapping Template	0.162606	0.99	0.071177	0.99
Overlapping Template	0.637119	0.98	0.013569	0.99
Universal	0.262249	1	0.108791	0.99
Approximate Entropy	0.304126	0.99	0.759756	1
Random Excursions	0.249284	0.98	0.249284	0.98
Random Excursions Variant	0.062821	0.98	0.025193	0.98
Serial	0.445368	0.99	0.137282	0.98
Linear Complexity	0.071177	0.99	0.227821	0.97

4.6 DSP realization of chaotic system

Use DSP platform to achieve the chaotic system. The hardware implementation platform is displayed in Fig. 7. Among them, the DSP chip with model TMS320F28335 is used, and the D/A converter is DAC8552, and the oscilloscope is UTD7102H.

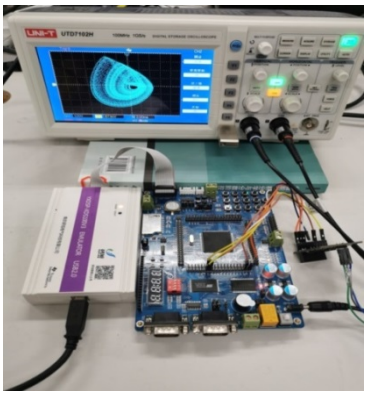
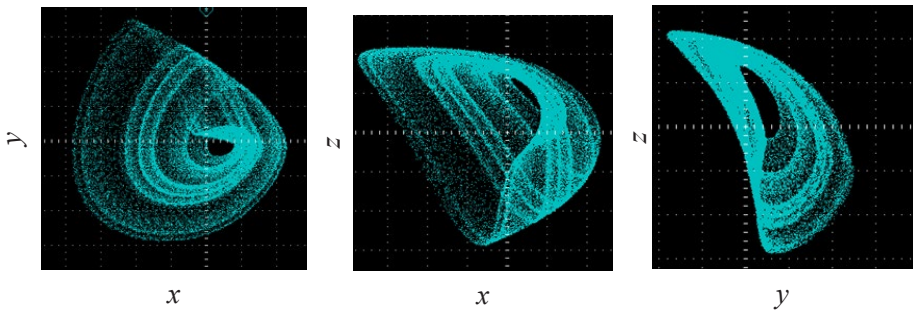


Fig. 7 DSP hardware platform

Setting parameters are shown in section 4.2, and the phase diagram implemented by the DSP platform is plotted in Fig. 8(a)-(c), it is consistent with the computer simulation.



(a) (b) (c)

Fig. 8 Phase diagram (a) *x*-*y* plane (b) *x*-*z* plane (c) *y*-*z* plane

5. Conclusion

According to the definition of fractional differential and the memristive chaotic circuit, we construct the fractional form, and the solution of the chaotic system is calculated using ADM algorithm. On this basis, the system's dynamic features are studied using phase diagram, LEs, bifurcation diagram, and complexity. The analysis of dynamic characteristics shows that the application of ADM decomposition algorithm can effectively solve the fractional-order chaotic system, and the proposed chaotic system has complex performance. When the parameter β=1.94 and the order $q \in (0.4,0.5)$, the randomness and complexity of the system are the best. Under this condition, the chaotic sequence obtained is more suitable for chaotic applications. The process of solving the fractional order chaotic system is realized on the DSP

platform, and the approximate attractor is obtained on the oscilloscope. The conclusions obtained from this research provide great help for the application of memristive chaotic circuits based on fractional order.

Reference

[1] L. Chua, "Memristor-The missing circuit element," IEEE Transactions on Circuit Theory, vol. 18, no. 5, pp. 507-519, 1971.

[2] L. O. Chua and K. Sung Mo, "Memristive devices and systems," Proceedings of the IEEE, vol. 64, no. 2, pp. 209-223, 1976.

[3] L. O. Chua, "The Fourth Element," Proceedings of the IEEE, vol. 100, no. 6, pp. p.1920-1927, 2012.

[4] F. Corinto, A. Ascoli, and M. Gilli, "Nonlinear Dynamics of Memristor Oscillators," IEEE Transactions on Circuits & Systems I Regular Papers, vol. 58, no. 6, pp. 1323-1336, 2011.

[5] L. Teng, H. H. C. Iu, X. Wang, and X. Wang, "Chaotic behavior in fractional-order memristor-based simplest chaotic circuit using fourth degree polynomial," Nonlinear Dynamics, vol. 77, no. 1-2, pp. 231-241, 2014.

[6] B. C. Bao, Z. Liu, and J. P. Xu, "Steady periodic memristor oscillator with transient chaotic behaviours," Electronics Letters, vol. 46, no. 3, pp. 237-238, 2010.

[7] B. Zhang and F. Deng, "Double-compound synchronization of six memristor-based Lorenz systems," Nonlinear Dynamics, vol. 77, no. 4, pp. 1519-1530, 2014.

[8] X. Ye, J. Mou, C. Luo, F. Yang, and Y. Cao, "Complexity Analysis of a Mixed Memristive Chaotic Circuit," Complexity, vol. 2018, pp. 1-9, 2018.

[9] X. Ye, J. Mou, C. Luo, and Z. Wang, "Dynamics analysis of Wien-bridge hyperchaotic memristive circuit system," Nonlinear Dynamics, 2018.

[10] F. Yu, L. Liu, H. Shen, Z. Zhang, and Q. Wan, "Multistability Analysis, Coexisting Multiple Attractors, and FPGA Implementation of Yu–Wang Four-Wing Chaotic System," Mathematical Problems in Engineering, vol. 2020, pp. 1-16, 2020.

[11] F. Yu et al., "Dynamic Analysis, Circuit Design, and Synchronization of a Novel 6D Memristive Four-Wing Hyperchaotic System with Multiple Coexisting Attractors," Complexity, vol. 2020, 2020.

[12] M. Jun, S. Kehui, W. Huihai, and R. Jingya, "Characteristic Analysis of Fractional-Order 4D Hyperchaotic Memristive Circuit," Mathematical Problems in Engineering, vol. 2017, pp. 1-13, 2017.

[13] Y. Xu, K. Sun, S. He, and L. Zhang, "Dynamics of a fractional-order simplified unified system based on the Adomian decomposition method," European Physical Journal Plus, vol. 131, no. 6, pp. 1-12, 2016.

[14] S. He, K. Sun, X. Mei, B. Yan, and S. Xu, "Numerical analysis of a fractional-order chaotic system

based on conformable fractional-order derivative," European Physical Journal Plus, vol. 132, no. 1, p. 36, 2017.

[15] L. Zhang, K. Sun, S. He, H. Wang, and Y. Xu, "Solution and dynamics of a fractional-order 5-D hyperchaotic system with four wings," European Physical Journal Plus, vol. 132, no. 1, p. 31, 2017.

[16] I. Petras, "Fractional-Order Memristor-Based Chua's Circuit," IEEE Transactions on Circuits & Systems II Express Briefs, vol. 57, no. 12, pp. 975-979, 2010.

[17] Ivopetr, "CHAOS IN FRACTIONAL-ORDER POPULATION MODEL," International Journal of Bifurcation & Chaos, vol. 22, no. 4, 2012.

[18] T. T. Hartley, C. F. Lorenzo, and H. K. Qammer, "Chaos in a fractional order Chua's system," Circuits & Systems I Fundamental Theory & Applications IEEE Transactions on, vol. 42, no. 8, pp. 485-490, 1995.

[19] D. Cafagna and G. Grassi, "AN EFFECTIVE METHOD FOR DETECTING CHAOS IN FRACTIONAL-ORDER SYSTEMS," International Journal of Bifurcation & Chaos, vol. 20, no. 03, pp. 1002595-, 2010.

[20] H. Y. Jia, Z. Q. Chen, and G. Y. Qi, "Chaotic Characteristics Analysis and Circuit Implementation for a Fractional-Order System," IEEE Transactions on Circuits & Systems I Regular Papers, vol. 61, no. 3, pp. 845-853, 2017.

[21] W. Gu, Y. Yu, and W. Hu, "Artificial Bee Colony Algorithm-based Parameter Estimation of Fractional-order Chaotic System with Time Delay," IEEE/CAA Journal of Automatica Sinica, vol. 4, no. 1, pp. 107-113, 2017.

[22] M. S. Tavazoei and M. Haeri, "Unreliability of frequency-domain approximation in recognising chaos in fractional-order systems," Iet Signal Processing, vol. 1, no. 4, pp. 171-181, 2007.

[23] M. S. Tavazoei and M. Haeri, "Limitations of frequency domain approximation for detecting chaos in fractional order systems," Nonlinear Analysis Theory Methods & Applications, vol. 69, no. 4, pp. 1299-1320, 2008.

[24] S. B. He, K. H. Sun, and H. H. Wang, "Solution of the fractional-order chaotic system based on Adomian decomposition algorithm and its complexity analysis," Acta Physica Sinica, vol. 63, no. 3, pp. 030502-72, 2014.

[25] D. Cafagna and G. Grassi, "BIFURCATION AND CHAOS IN THE FRACTIONAL-ORDER CHEN SYSTEM VIA A TIME-DOMAIN APPROACH," International Journal of Bifurcation & Chaos, vol. 18, no. 07, pp. 1845-1863, 2008.

[26] F. Yang, J. Mou, K. Sun, Y. Cao, and J. Jin, "Color Image Compression-Encryption Algorithm Based on Fractional-Order Memristor Chaotic Circuit," IEEE Access, vol. 7, pp. 58751-58763, 2019.

A novel chaotic system with 11 piecewise linear functions and its analog circuit implementation

Jieyang Wang[1], Jiawu Yu[1], Huizhen Yan[1]

{yujiawu_dlpu@sina.com}

[1]School of Information Science and Engineering, Dalian polytechnic University, Dalian, 116034, China

Abstract. In this paper, a new four-dimensional chaotic system with 11 piecewise linear functions is proposed in this paper. The dynamic characteristics of the four-dimensional chaotic system are analyzed through the Lyapunov exponential spectrum and bifurcation diagram. The state of the system under different bifurcation parameter changes is observed and recorded, and the exact parameters of the system entering the chaotic state are determined by the 10 test method. Numerical simulation results show that the system has entered chaos from the intermittent chaotic road, and the system has hidden attractors. Finally, the corresponding analog circuit is designed through system dynamics equations and circuit theory, which provides a complete theoretical basis for the system.

Keywords: 11 piecewise linear function; intermittent chaos road; hidden attractor; analog circuit implementation

1 Introduction

Chaos is a non-linear phenomenon widely existing in nature, and it is a basic form of motion [1-2]. The motion model of a dynamic system depends on two factors, one of which is the dynamic equation of the system, and the other is the initial conditions of the system [3-5]. For a deterministic system, when the system equation and initial conditions are determined, the result is determined, that is, the deterministic equation gets a definite solution, but in a chaotic system, even if the equation and initial conditions are determined, an uncertain solution will still be obtained [4-6]. In the chaotic system, three main roads are leading to chaos, namely the period-doubling bifurcation road, the PM-type intermittent road, and the pseudo-period road [7-8]. The way to enter chaos through PM intermittent roads is relatively rare compared to the other two, and it has a higher research value [9-11].

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307842

The PM type intermittent road was proposed by French scientists Y. Pomeau and P. Manneville, and it is also called the intermittent chaos road. The generation mechanism of PM-type intermittent roads is closely related to the tangential bifurcation [13-15]. This way occurs before the starting point of the tangential bifurcation, and the performance is that the periodic state and the chaotic state alternately appear with parameter changes [16-20]. When one of the parameters of the system is lower than a certain value, the system exhibits regular periodic motion [20-22]. When the parameters gradually change in a certain direction along the number axis, the system exhibits an obvious approximate periodic motion for a long time [23-24]. In the process of gradual change, the system has a brief chaotic behavior, and then it becomes a periodic motion [25]. This method is repeated continuously [26-28]. With the change of parameters, this sudden phenomenon becomes more and more frequent, and the approximate periodic motion becomes shorter and shorter [29-30]. Finally, the system enters a completely chaotic state [31].

Piecewise linear function is a kind of nonlinear function, which is usually used to make multi-scroll chaotic systems and has important application value [32-34]. According to the different values of the slope of the function, piecewise linear functions can be divided into four categories [35-36]. By changing a parameter in the function, the number of wings of the multi-scroll chaotic system can be flexibly changed [37-38]. In this paper, by adding 11 piecewise linear functions to the chaotic system to increase the complexity, a new four-dimensional chaotic system with complex dynamics is obtained. The stability of the four-dimensional chaotic system is analyzed, and the results show that it is in a stable state, and the attractor of the system is hidden [39]. The complex dynamics and balance characteristics of the new system make it have a high application potential in military information transmission and confidential communication [40-42].

In this paper, the structure of the new 11 piecewise linear function and the attractor state of the system are introduced in the first part, and the equilibrium point and stability analysis of the new chaotic system are described in the second part. By using a bifurcation diagram, Lyapunov exponential spectrum, and a test method, the dynamic characteristics of this system are analyzed in the third part. The analog circuit corresponding to the new four-dimensional chaotic system is designed through Kirchhoff's law and circuit principle in the fourth part. The summary of the research results of this paper are presented in the fifth section.

2 Novel four-dimensional chaotic system

A new four-dimensional chaotic system with 11 piecewise linear functions can be expressed as

$$\begin{cases} \dot{x} = -a(x+by) + cxv \\ \dot{y} = -d(y-bx) + eyz \\ \dot{z} = -x^2 z + x + H(x) \\ \dot{v} = -(1+h+100y^2)v/100 + (h-1)/10 \end{cases}, \tag{1}$$

among them, H(x) represents an 11 piecewise linear function, and piecewise function H(x) is an odd function, which is usually used to generate an even number of multi-scroll functions. In this paper, it is used to construct a new chaotic system. The construction method greatly increases the complexity of the chaotic system. The corresponding mathematical expression is shown in Formula 2, and the corresponding function image is shown in Figure 1.

$$H(x) = m_N x + 0.5\sum_{i=1}^{N}(m_{i-1} - m_i)(|x+E_i| - |x-E_i|). \tag{2}$$

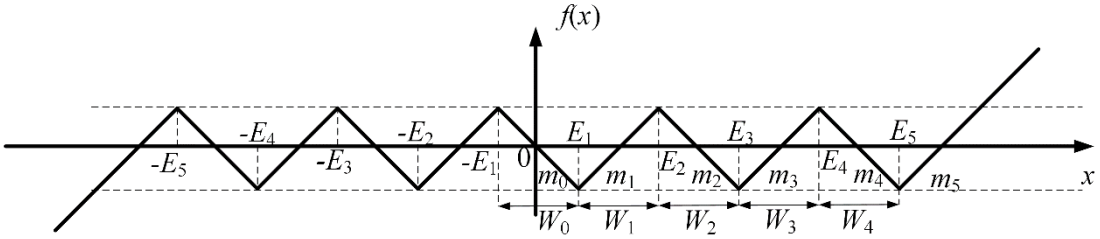


Fig 1 Function graph of 11 piecewise linear functions in the system

In Figure 1, the turning point value is E_i, the width of the linear interval is W_i, the slope of each interval is m_i, and the height of each turning point is equal. From this, the relationship between the width W_i of each interval and the first turning point E_1 can be expressed as

$$\begin{aligned} W_0 &= 2\frac{|m_0|}{|m_0|}E_1 \\ &\vdots \\ W_1 &= 2\frac{|m_0|}{|m_1|}E_1 \\ W_i &= 2\frac{|m_0|}{|m_i|}E_1 \end{aligned}, \tag{3}$$

in formula 3, i=0,1,2, the slope m_i of each interval and the first turning point value E_1 are

incorporated into the formula, and the calculation formula for turning point E_n can be obtained as

$$E_n = \left(1+2\sum_{i=1}^{n-1}\frac{|m_0|}{|m_i|}\right)E_1, \tag{4}$$

expand it to get formula 5

$$\begin{cases} E_2 = \left(1+2\dfrac{|m_0|}{|m_1|}\right)E_1 \\ E_3 = \left(1+2\dfrac{|m_0|}{|m_1|}+2\dfrac{|m_0|}{|m_2|}\right)E_1 \\ E_4 = \left(1+2\dfrac{|m_0|}{|m_1|}+2\dfrac{|m_0|}{|m_2|}+2\dfrac{|m_0|}{|m_3|}\right)E_1 \\ E_5 = \left(1+2\dfrac{|m_0|}{|m_1|}+2\dfrac{|m_0|}{|m_2|}+2\dfrac{|m_0|}{|m_3|}+2\dfrac{|m_0|}{|m_4|}\right)E_1 \end{cases}, \tag{5}$$

the specific expression of the 11 piecewise linear functions can be expressed as

$$\begin{aligned} H(x) &= m_5 x + 0.5(m_0 - m_1)(|x+E_1| - |x-E_1|) \\ &\quad + 0.5(m_1 - m_2)(|x+E_2| - |x-E_2|) + 0.5(m_2 - m_3)(|x+E_3| - |x-E_3|) \\ &\quad + 0.5(m_3 - m_4)(|x+E_4| - |x-E_4|) + 0.5(m_4 - m_5)(|x+E_5| - |x-E_5|) \\ &= f_0(x) + f_1(x) + f_2(x) + f_3(x) + f_4(x) + f_5(x) \end{aligned}. \tag{6}$$

When the system parameters take a=4.5, b=1.6, c=19.5, d=9, e=6.7, h=8.1. m_0=m_2=m_4=-0.5, m_1=m_3=m_5=0.3, the first turning point value E_1=0.4. The initial value is (0.1 0.1 0.1 0.1), when the step length is 0.01s, the Lyapunov exponent of the system is (0.3588, 0.0063, -1.1288, -4.9655), and the system phase diagram is shown in Figure 2, the time domain diagram of the system is shown in Figure 3.

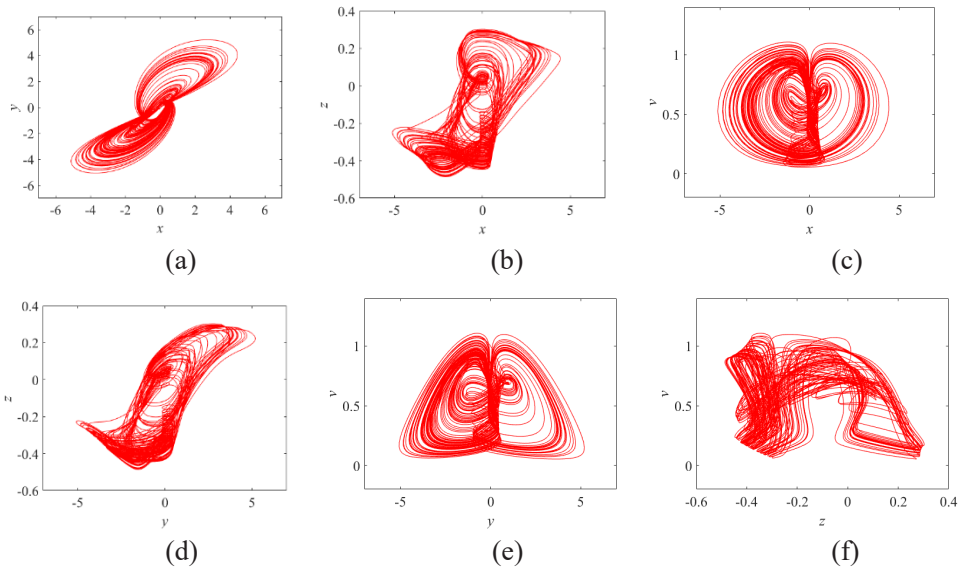


Fig 2 The phase diagram of the system, (a) *x*-*y* plane, (b) *x*-*z* plane, (c) *x*-*v* plane, (d) *y*-*z* plane, (e) *y*-*v* plane, (f) *z*-*v* plane

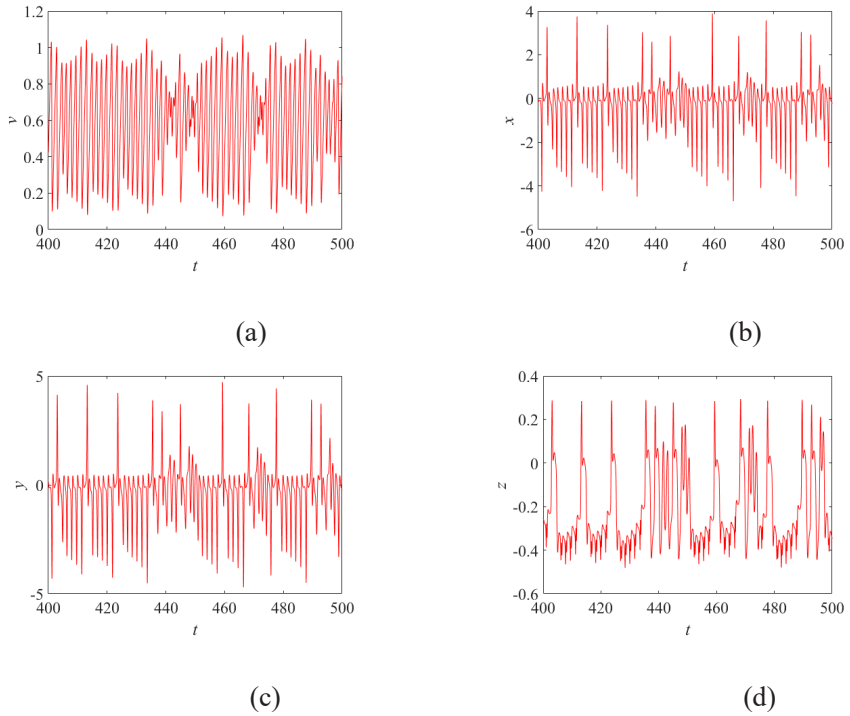


Fig 3 Time domain diagram of the system, (a) *t*-*x* plane, (b) *t*-*y* plane, (c) *t*-*z* plane, (d) *t*-*v* plane

3 Balance point analysis of system

The divergence of the four-dimensional chaotic system is

$$\begin{aligned}\nabla V &= \frac{\partial \dot{x}}{\partial x}+\frac{\partial \dot{y}}{\partial y}+\frac{\partial \dot{z}}{\partial z}+\frac{\partial \dot{v}}{\partial v} \\ &= -a+cv-d+ez-x^2-(1+h+100y^2)/100\end{aligned}, \tag{7}$$

when $-a+cv-d+ez-x^2-(1+h+100y^2)/100<0$, the system is dissipative, and the volume of system is 0, when t approaches infinity. Let the parameters a=4.5, b=1.6, c=19.5, d=9, e=6.7, h=8.1. $m_0=m_2=m_4$=-0.5, $m_1=m_3=m_5$=0.3, E_1=0.4. When the initial value is (0.1 0.1 0.1 0.1), the divergence is less than 0, the system at this time is dissipative, and has the conditions to produce chaos.

Let the differential equations of the system be equal to zero, that is, when the system equation takes $\dot{x}=\dot{y}=\dot{z}=\dot{v}=0$, the equilibrium point of the system can be obtained.

$$\begin{cases}-a(x+by)+cxv=0\\ -d(y-bx)+eyz=0\\ -x^2z/100+x+H(x)=0\\ -(1+h+100y^2)v/100+(h-1)/10=0\end{cases}, \tag{8}$$

it can be easily found that the equilibrium point of the system is P(0, 0, 1.3303, 7.8022). Linearize the system at point P, and the Jacobian matrix of the system is

$$\begin{bmatrix} cv-a & -ab & 0 & cx \\ bd & ez-d & ey & 0 \\ P(x) & 0 & -x^2 & 0 \\ 0 & -2vy & 0 & -y^2-h/100-0.01 \end{bmatrix}, \tag{9}$$

among them, expand P(x) to get formula 10

$$\begin{aligned}P(x) = &+(sgn(E_1+x)+0.5sgn(E_1-x))(m_0-m_1)-2xz\\ &+(sgn(E_2+x)+0.5sgn(E_2-x))(m_1-m_2)+m_5\\ &+(sgn(E_3+x)+0.5sgn(E_3-x))(m_2-m_3)+1\\ &+(sgn(E_4+x)+0.5sgn(E_4-x))(m_3-m_4)\\ &+(sgn(E_5+x)+0.5sgn(E_5-x))(m_4-m_5)\end{aligned}. \tag{10}$$

By analyzing the Jacobian matrix, the characteristic polynomial of the system can be

expressed as

$$f(\lambda) = A_0\lambda^4 + A_1\lambda^3 + A_2\lambda^2 + A_3\lambda^1 + A_4 \tag{11}$$

according to the Rouss-Herwitz stability criterion, the system is stable under the following conditions.

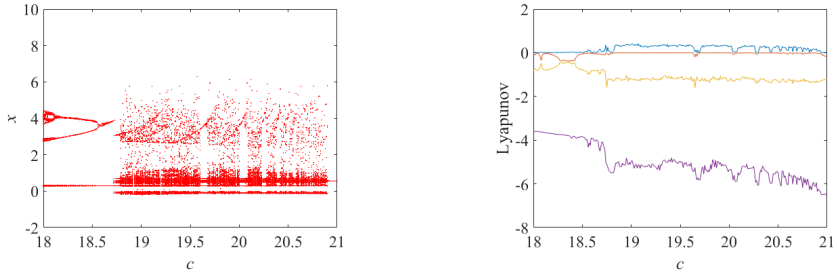
$$\begin{aligned} &A_n > 0 \\ &A_1A_2 > A_0A_3 \\ &A_1A_2A_3 > A_0A_3^2 + A_1^2A_4 \end{aligned} , \tag{12}$$

putting A_n into formula 12 to calculate in turn shows that the system is stable according to the stability judgment. Therefore, all attractors in this system are hidden attractors.

4 Dynamics analysis of system

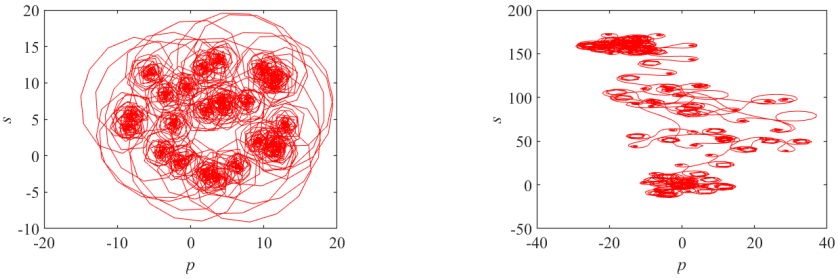
The bifurcation graph and Lyapunov exponent spectrum are important indicators to determine the state of the system. To obtain more accurate parameter values when the system enters chaos, the 10 test method is used for system analysis. This test method can accurately determine the parameter values when the system enters chaos.

Let the system parameters a=4.5, b=1.6, d=9, e=6.7, h=8.1. m_0=m_2=m_4=-0.5, m_1=m_3=m_5=0.3, E_1=0.4. the initial value is (0.1 0.1 0.1 0.1), when the step length is 0.01s, the bifurcation diagram and Lyapunov index spectrum obtained according to the change of $c \in (18, 21)$ are shown in Figure 4. According to Figure 4, as the parameter c decreases from 21, the system begins to produce a transient state of chaos. When the parameter value is c=20.909, the system enters chaos for the first time, the Lyapunov exponent is (0.0996, -0.0165, -1.3122, -6.2662), and the duration gradually becomes longer, and finally, it enters a continuous chaotic state. When c=18.468, the system begins to show obvious period-doubling bifurcation. At this time, the Lyapunov exponent is (0.0133, -0.0801, -0.7587, -3.8666), and the system gradually enters a new chaotic state. The 10 test chart when the system enters chaos and periodic state for the first time is shown in Fig. 5.



(a) (b)

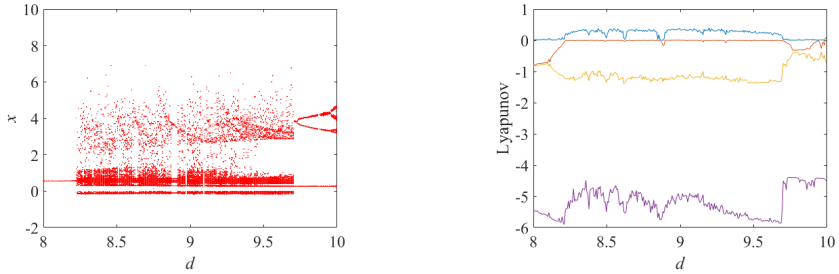
Fig 4 The Lyapunov index spectrum and bifurcation diagram of the system when c changes, (a) the bifurcation diagram of the system, (b) the Lyapunov index diagram of the system



(a) (b)

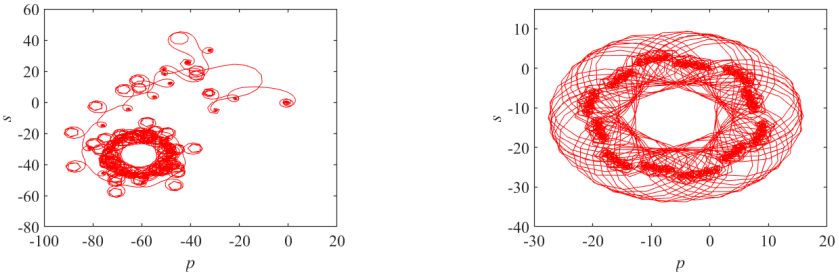
Fig 5 The 10 test chart of the system when c changes, (a) c=20.90, (b) c=18.468

Let the system parameters a=4.5, b=1.6, c=19.5, e=6.7, h=8.1. m_0=m_2=m_4=-0.5, m_1=m_3=m_5=0.3, E_1=0.4, the initial value is (0.1 0.1 0.1 0.1). When the step size is 0.01s and $d \in (8, 10)$, the bifurcation graph and Lyapunov index spectrum of the system are shown in Figure 6. And, Figure 7 is the 10 test chart when the system enters the chaotic state and the periodic state for the first time under different values of the parameter d. With the increase of d, when d=8.224, the system enters a chaotic state, at this time, the Lyapunov index of the system is (0.2243, -0.0034, -1.2404, -5.3852). When $d \in (8.861,8.889)$, the system behaves as a periodic state. When d=8.623, the system is in the periodic window, and the Lyapunov index of the system at this time is (0.02565, -0.03074, -1.371, -5.538). When d>9.719, the system enters a periodic state. At this time, the Lyapunov exponent is (0.0171, -0.0763, -0.6976, -4.4898).



(a) (b)

Fig 6 The Lyapunov index spectrum and bifurcation diagram of the system when d changes, (a) the bifurcation diagram of the system, (b) the Lyapunov index diagram of the system



(a) (b)

Fig 7 The 10 test chart of the system when d changes, (a) d=20.90, (b) d=18.468

5 Analog circuit design of system

In order to perfect the theory of the system, an analog circuit corresponding to the new system shown in formula 13 is designed. The analog circuit is composed of the main arithmetic circuit and an 11 piecewise linear function circuit. The complete circuit consists of resistors, capacitors, multipliers, and operational amplifiers.

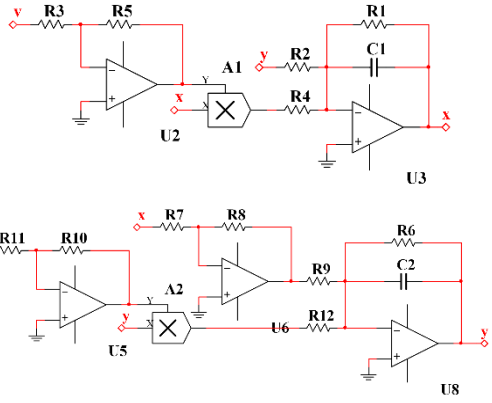
$$\begin{cases} \dot{x} = -a(x+by)+cxv \\ \dot{y} = -d(y-bx)+eyz \\ \dot{z} = -x^2z+x+H(x) \\ \dot{v} = -(1+h+100y^2)v/100+(h-1)/10 \end{cases} . \tag{13}$$

Dimension formula 13 to obtain the equation system shown in formula 14. The first and second differential equations in the equation system are coupled with each other, corresponding to the two sets of circuits in Figure 8 (a) and (b), the piecewise linear function is added in the third

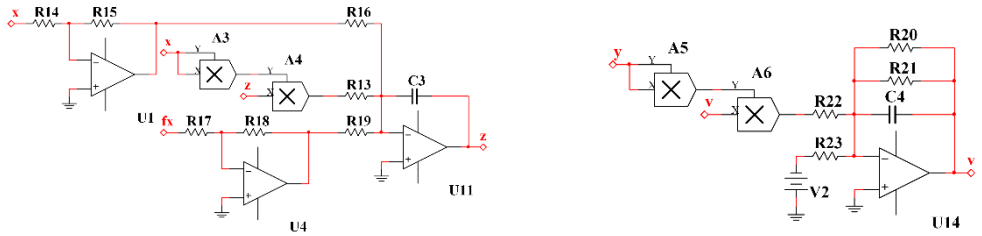
differential equation.

$$\begin{cases} C_1V_x = \dfrac{V_x}{R_1} + \dfrac{V_y}{R_2} + \dfrac{R_4 g_1 V_x V_v}{R_3 R_5} \\ C_2V_y = \dfrac{V_y}{R_6} + \dfrac{R_7 V_x}{R_8 R_9} + \dfrac{R_{10} g_2 V_y V_z}{R_{11} R_{12}} \\ C_3V_z = \dfrac{g_3 g_4 V_x{}^2 V_z}{R_{13}} + \dfrac{R_{14} V_x}{R_{15} R_{16}} + \dfrac{R_{17} H(V_x)}{R_{18} R_{19}} \\ C_4V_v = \dfrac{V_v}{R_{20}} + \dfrac{V_v}{R_{21}} + \dfrac{g_5 g_6 V_v V_y{}^2}{R_{22}} + \dfrac{V}{R_{23}} \end{cases}, \tag{14}$$

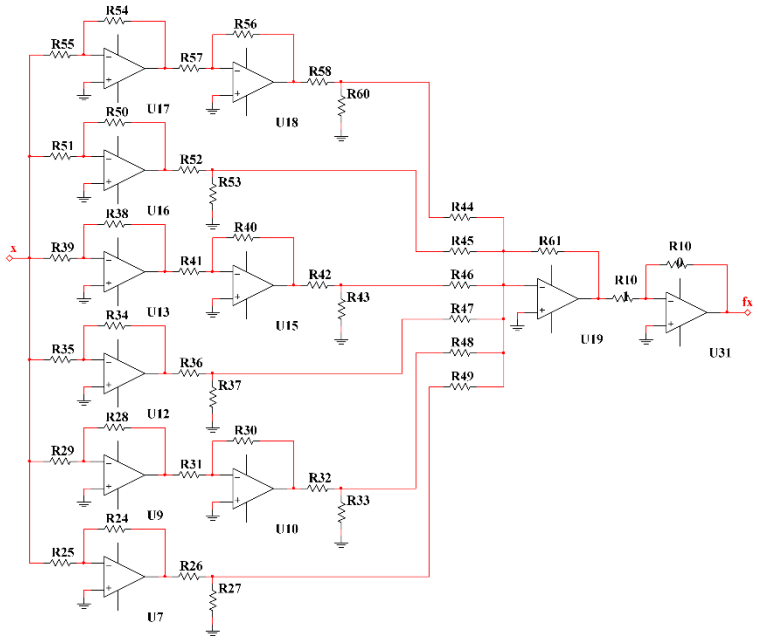
among them, v_x, v_y, v_z, v_v are the four circuit variables of the simulation circuit, g_1, g_2, g_3, g_4, g_5, g_6 are the amplification coefficients of the multiplier, and V represents the constant term. According to the logical relationship shown in formula 14, combined with Kirchhoff's law and circuit principle, the analog circuit corresponding to system 13 can be obtained.



(a) (b)



(c) (d)



(e)

Fig 8 The simulation circuit corresponding to the four-dimensional chaotic system, (a) the first differential equation (b) the second differential equation (c) the third differential equation (d) the fourth differential equation (e) 11 piecewise linear function

6 Conclusion

In this paper, by introducing 11 piecewise linear functions to the chaotic system, a new four-dimensional chaotic system with complex dynamic behavior is proposed. Methods such as bifurcation diagram, Lyapunov exponent spectrum, phase diagram, and 10 test are used to analyze the system. Numerical simulation results show that due to the introduction of 11 piecewise linear functions, the system has hidden attractors and complex dynamics. For example, when some parameters are fixed, when the parameter c becomes smaller, the system shows obvious intermittent chaos. The chaotic state of the system appears more and more frequently and lasts longer and longer until it finally enters the chaotic zone completely. Finally, the analog circuit diagram corresponding to the system is designed using the circuit principle, which provides a complete theoretical basis for scholars to study the system. The complex dynamics of the system provide a good prospect for its practical application.

Acknowledgments

This work was supported by Provincial Natural Science Foundation of Liaoning (Grant Nos.

2020-MS-274); National Natural Science Foundation of China (Grant Nos. 62061014).

Author contributions

Jieyang Wang designed and carried out experiments, data analyzed and manuscript wrote. Jiawu Yu made the theoretical guidance for this paper. Huizhen Yan improved the algorithm. All authors reviewed the manuscript.

Conflicts of Interest

No conflicts of interests about the publication by all authors.

References

[1] Tigan G , Opri D . Analysis of a 3D chaotic system[J]. Chaos Solitons & Fractals, 2008, 36(5):1315-1319.

[2] Qi G , Du S , Chen G , et al. On a four-dimensional chaotic system[J]. Chaos Solitons & Fractals, 2005, 23(5):1671-1682.

[3] Zhang F D . The Synchronization of a Fractional-Order Chaotic System[J]. Advanced Materials Research, 2013, 655-657:1488-1491.

[4] Lu J , Wu X , Jinhu Lü. Synchronization of a unified chaotic system and the application in secure communication[J]. Physics Letters A, 2002, 305(6):365-370.

[5] Chen, Shi-Hua, Zhao, et al. Parameter identification and synchronization of an uncertain Chen chaotic system via adaptive control[J]. Chinese Physics, 2002, 11(6):543-543.

[6] Li Y , Yang B J , Badal J , et al. Chaotic system detection of weak seismic signals[J]. Geophysical Journal International(3):1493-1522.

[7] Sun, H, Cao. Chaos control and synchronization of a modified chaotic system[J]. CHAOS SOLITONS AND FRACTALS, 2008.

[8] Muthukumar P , Balasubramaniam P , Ratnavelu K . Synchronization and an application of a novel fractional order King Cobra chaotic system[J]. Chaos, 2014, 24(3):10000-968.

[9] Kodama H , Sato S , Honda K . Reconsideration of the renormalization-group theory on intermittent chaos[J]. Physics Letters A, 1991, 157(6-7):354-356.

[10] Ditto W L , Rauseo S N , Spano M L . Experimental control of chaos[J]. Controlling Chaos, 1996, 65(26):105-107.

[11] Cosenza M G , Alvarez-Llamoza O , Ponce G A . Lyapunov exponent for type-III intermittent chaos[J]. Communications in Nonlinear Science & Numerical Simulation, 2010, 15(9):2431-2435.

[12] Abarbanel H D I , Rulkov N F , Sushchik M M . Generalized synchronization of chaos: The auxiliary system approach[J]. Physical Review E Statistical Physics Plasmas Fluids & Related

Interdisciplinary Topics, 1996, 53(5):4528-4535.

[13] Tse C K , Zhou Y , Lau F C M , et al. Intermittent chaos and subharmonics in switching power supplies[C]// Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on. IEEE, 2003.

[14] Gabriel Pérez, Cerdeira H A . Extracting Messages Masked by Chaos[J]. Physical Review Letters, 1995, 74(11):1970-1973.

[15] Hisao O , Hazime M , Shoichi K . Global Spectral Structures of Type III Intermittent Chaos[J]. Progress of Theoretical Physics(3):784-801.

[16] N, Mori, T, et al. Scaling Structures and Statistical Mechanics of Type I Intermittent Chaos[J]. Progress of Theoretical Physics, 1989, 81(1).

[17] Njougouo T , Simo G R , Louodop P , et al. Effects of intermittent coupling on synchronization[J]. Chaos Solitons & Fractals, 2020, 139:110082.

[18] Koronovskii A A , Moskalenko O I , Pivovarov A A , et al. Intermittent route to generalized synchronization in bidirectionally coupled chaotic oscillators[J]. Chaos, 2020, 30(8):083133.

[19] Pierre Bergé, Pomeau Y , Vidal C . Order within chaos: towards a deterministic approach to turbulence. Transl. from the French[J]. 1987.

[20] Jahanshahi H , Onofre Orozco-López, Munoz-Pacheco J M , et al. Simulation and experimental validation of a non-equilibrium chaotic system[J]. Chaos Solitons & Fractals, 2021, 143:110539.

[21] Yang Y , Wang L , Duan S , et al. Dynamical analysis and image encryption application of a novel memristive hyperchaotic system[J]. Optics & Laser Technology, 2021, 133:106553.

[22] Khan A , Jahanzaib L S , Trikha P . Changing Dynamics of the First, Second and Third Approximations of the Exponential Chaotic System and Their Application in Secure Communication Using Synchronization[J]. International Journal of Applied and Computational Mathematics, 2021, 7(1).

[23] Qi G , Gou T , Hu J , et al. Breaking of integrability and conservation leading to Hamiltonian chaotic system and its energy-based coexistence analysis[J]. Chaos, 2021, 31(1):013101.

[24] Kubba Z M J , Hoomod H K . Modified PRESENT Encryption algorithm based on new 5D Chaotic system[J]. IOP Conference Series: Materials Science and Engineering, 2020, 928(3):032023 (8pp).

[25] Li H , Deng L , Gu Z . A Robust Image Encryption Algorithm Based on a 32-bit Chaotic System[J]. IEEE Access, 2020, 8:30127-30151.

[26] Niu Y , Sun X , Zhang C , et al. Anticontrol of a Fractional-Order Chaotic System and Its Application in Color Image Encryption[J]. Mathematical Problems in Engineering, 2020, 2020.

[27] J. C , Antoranz, and, et al. Intermittent incommensurate chaos[J]. Physica D: Nonlinear Phenomena, 1985, 16(2):184-202.

[28] Goldhirsch I , Noskowicz S H , Schuss Z . Band-edge localization as intermittent chaos[J]. Physical Review B Condensed Matter, 1993, 47(4):1918.

[29] Ohtaki Y , Takara K , Onodera F , et al. A SIMPLE MODEL OF INTERMITTENT CHAOS[J]. 2015.

[30] Takeshi K , Yoshiro K . Intermittent Chaos Generated by Logarithmic Map[J]. Progress of Theoretical Physics(3):581-586.

[31] MORI, Hazime, OKAMOTO, et al. Global Spectral Structures of Intermittent Chaos[J]. Ippj Rev, 1986.

[32] Shaobo H , Kehui S , Huihai W . Modified multiscale permutation entropy algorithm and its application for multiscroll chaotic systems[J]. Complexity, 2016:52-58.

[33] Liu S , Wei Y , Liu J , et al. Multi-Scroll Chaotic System Model and Its Cryptographic Application[J]. International Journal of Bifurcation and Chaos, 2020.

[34] Shao X , Chen F . The Projective Synchronization of Fractional-order Multi-scroll Chaotic Systems[C]// 2015 4th International Conference on Sensors, Measurement and Intelligent Materials. 2016.

[35] Gui Z , Wu X , Lin Q . Finite Time Synchronization of Nonautonomous Multiscroll Chaotic Systems[C]// International Workshop on Chaos-fractals Theories & Applications. 0. Senouci, Abdelkader, Boukabou, et al. Fuzzy modeling, stabilization and synchronization of multi-scroll chaotic systems.[J]. Optik - International Journal for Light & Electron Optics, 2016.

[36] A A K M , B S D , A V K Y . Finite-time synchronization of multi-scroll chaotic systems with sigmoid non-linearity and uncertain terms - ScienceDirect[J]. Chinese Journal of Physics, 2020.

[37] Khelifa M A , Boukabou A . Design of an intelligent prediction-based neural network controller for multi-scroll chaotic systems[J]. Applied Intelligence, 2016, 45(3):1-15.

[38] Yuan F , Wang G , Wang X . Extreme multistability in a memristor-based multi-scroll hyper-chaotic system[J]. Chaos, 2016, 26(7):507-519. Xiulan, Zhang, Xin, et al. Generalized synchronization for multidirectional saturated multi-scroll chaotic systems[J]. International Journal of Applied Mathematics & Statistics, 2013.

[39] Khelifa, Mohammed, Amin, et al. Design of an intelligent prediction-based neural network controller for multi-scroll chaotic systems[J]. Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 2016.

[40] Cong W , Hong-Li Z , University X . Parameter identification for fractional-order multi-scroll chaotic systems based on original dual-state transition algorithm[J]. Acta Physica Sinica, 2016, 65(6).

Appendix

The coefficient A_n of the characteristic polynomial involved in the third part of this paper is shown in here.

$$A_0 = 1$$

$$A_1 = x^2 + y^2 + a + d + h/100 - cw - ez + 1/100$$

$$\begin{aligned} A_2 = {} & a/100 + d/100 + x^2y^2 + ad + ah/100 + dh/100 - cw/100 \\ & - ez/100 + ax^2 + ay^2 + dx^2 + dy^2 + hx^2/100 + x^2/100 - cwx^2 \\ & - cwy^2 - ex^2z - ey^2z - cdw - aez - chw/100 - ehz/100 + ab^2d + cewz \end{aligned}$$

$$\begin{aligned} A_3 = {} & (ad)/100 + (ax^2)/100 + (dx^2)/100 + (dhx^2)/100 \\ & - (cwx^2)/100 - (ex^2z)/100 + ax^2y^2 + dx^2y^2 + (adh)/100 \\ & - (cdw)/100 - (aez)/100 + (ab^2d)/100 + adx^2 + ady^2 \\ & + (ahx^2)/100 + abey - (cdhw)/100 - (aehz)/100 + (cewz)/100 \\ \\ & + (ab^2dh)/100 - cdwx^2 - cdwy^2 - aex^2z - aey^2z - (chwx^2)/100 \\ & - (ehx^2z)/100 + ab^2dx^2 + ab^2dy^2 - cwx^2y^2 - ex^2y^2z + abem5y \\ & + (cehwz)/100 + cewx^2z + cewy^2z + (abem0ysign(E1-x))/2 \\ & - (abem1ysign(E1-x))/2 + (abem1ysign(E2-x))/2 \\ \\ & - (abem2ysign(E2-x))/2 + (abem2ysign(E3-x))/2 \\ & - (abem3ysign(E3-x))/2 + (abem3ysign(E4-x))/2 \\ & - (abem4ysign(E4-x))/2 + (abem4ysign(E5-x))/2 \\ & - (abem5ysign(E5-x))/2 + 2bcdwxy - 2abexyz \\ \\ & + (abem0ysign(E1+x))/2 - (abem1ysign(E1+x))/2 \\ & + (abem1ysign(E2+x))/2 - (abem2ysign(E2+x))/2 \\ & + (abem2ysign(E3+x))/2 - (abem3ysign(E3+x))/2 \\ & + (abem3ysign(E4+x))/2 - (abem4ysign(E4+x))/2 \\ & + (abem4ysign(E5+x))/2 - (abem5ysign(E5+x))/2 \end{aligned}$$

$$
\begin{aligned}
A_4 = {} & (adx^2)/100 + (abey)/100 + abey^3 + (adhx^2)/100 - (cdwx^2)/100 \\
& - (aex^2z)/100 + (ab^2dx^2)/100 + adx^2y^2 + (ab^2dhx^2)/100 - cdwx^2y^2 \\
& - aex^2y^2z + (abehy)/100 + (abem5y)/100 + ab^2dx^2y^2 + abem5y^3 \\
& - (cdhwx^2)/100 - (aehx^2z)/100 + 2cewxy^2 + (cewx^2z)/100 + 2bcdwx^3y \\
& - 2abexy^3z + (cehwx^2z)/100 + 2cem5wxy^2 + (abem0ysign(E1-x))/200 \\
& - (abem1ysign(E1-x))/200 + (abem1ysign(E2-x))/200 \\
& - (abem2ysign(E2-x))/200 + (abem2ysign(E3-x))/200 \\
& - (abem3ysign(E3-x))/200 + (abem3ysign(E4-x))/200 \\
& - (abem4ysign(E4-x))/200 + (abem4ysign(E5-x))/200 \\
& - (abem5ysign(E5-x))/200 + (abem0y^3sign(E1+x))/2 \\
& - (abem1y^3sign(E1+x))/2 + (abem1y^3sign(E2+x))/2 \\
& - (abem2y^3sign(E2+x))/2 + (abem2y^3sign(E3+x))/2 \\
& - (abem3y^3sign(E3+x))/2 + (abem3y^3sign(E4+x))/2 \\
& - (abem4y^3sign(E4+x))/2 + (abem4y^3sign(E5+x))/2 \\
& - (abem5y^3sign(E5+x))/2 + (abem0y^3sign(E1-x))/2 \\
& - (abem1y^3sign(E1-x))/2 + (abem1y^3sign(E2-x))/2 \\
& - (abem2y^3sign(E2-x))/2 + (abem2y^3sign(E3-x))/2 \\
& - (abem3y^3sign(E3-x))/2 + (abem3y^3sign(E4-x))/2 \\
& - (abem4y^3sign(E4-x))/2 + (abem4y^3sign(E5-x))/2 \\
& - (abem5y^3sign(E5-x))/2 - 3cewx^2y^2z + (abehm5y)/100 \\
& - (abexyz)/50 + (abem0ysign(E1+x))/200 \\
& - (abem1ysign(E1+x))/200 + (abem1ysign(E2+x))/200 \\
& - (abem2ysign(E2+x))/200 + (abem2ysign(E3+x))/200 \\
& - (abem3ysign(E3+x))/200 + (abem3ysign(E4+x))/200 \\
& - (abem4ysign(E4+x))/200 + (abem4ysign(E5+x))/200
\end{aligned}
$$

$$-(abem5ysign(E5+x))/200+cem0wxy^2sign(E1-x)$$
$$-cem1wxy^2sign(E1-x)+cem1wxy^2sign(E2-x)$$
$$-cem2wxy^2sign(E2-x)+cem2wxy^2sign(E3-x)$$
$$-cem3wxy^2sign(E3-x)+cem3wxy^2sign(E4-x)$$
$$-cem4wxy^2sign(E4-x)+cem4wxy^2sign(E5-x)$$
$$-cem5wxy^2sign(E5-x)-(abehxyz)/50$$
$$+(abehm0ysign(E1+x))/200-(abehm1ysign(E1+x))/200$$
$$+(abehm1ysign(E2+x))/200-(abehm2ysign(E2+x))/200$$
$$+(abehm2ysign(E3+x))/200-(abehm3ysign(E3+x))/200$$
$$+(abehm3ysign(E4+x))/200-(abehm4ysign(E4+x))/200$$
$$+(abehm4ysign(E5+x))/200-(abehm5ysign(E5+x))/200$$
$$+(abehm0ysign(E1-x))/200-(abehm1ysign(E1-x))/200$$
$$+(abehm1ysign(E2-x))/200-(abehm2ysign(E2-x))/200$$
$$+(abehm2ysign(E3-x))/200-(abehm3ysign(E3-x))/200$$
$$+(abehm3ysign(E4-x))/200-(abehm4ysign(E4-x))/200$$
$$+(abehm4ysign(E5-x))/200-(abehm5ysign(E5-x))/200$$
$$+cem0wxy^2sign(E1+x)-cem1wxy^2sign(E1+x)$$
$$+cem1wxy^2sign(E2+x)-cem2wxy^2sign(E2+x)$$
$$+cem2wxy^2sign(E3+x)-cem3wxy^2sign(E3+x)$$
$$+cem3wxy^2sign(E4+x)-cem4wxy^2sign(E4+x)$$
$$+cem4wxy^2sign(E5+x)-cem5wxy^2sign(E5+x)$$

A Novel Color Image Encryption Scheme Based on Chaotic Sequence and DNA Mutation Principle

Jin Hao, Miao Miao, HuiZhen Yan

{miaomiao@dlpu.edu.cn}

School of Information Science and Engineering, Dalian polytechnic University, Dalian, 116034, China

ABSTRACT: Based on the simplified three-dimensional fractional unified chaotic system and the principle of DNA mutation, this paper proposes a color image encryption method to further improve the security of image information transmission. On the other hand, the pseudorandom sequence generated by the 3D score simplified unified chaotic system uses the DNA mutation principle algorithm to scramble the color image and expand the pixel value. Finally, the security of the proposed cryptographic algorithm is analyzed using key space, correlation, information entropy, etc. Numerical simulation results show that the algorithm has good security. The results of this paper provide a theoretical basis for the unification of the chaotic system applying the principle of fractional simplification and DNA mutation to image security.

Keywords: Chaotic System; Image Encryption Scheme;DNA Mutation Principle

1.Introduction

With the rapid development of networks and communication technologies, data transmission methods have undergone tremendous changes, and the demand for data transmission security is also increasing. However, Some existing encryption designs are no longer sufficient to support the encryption operation of huge information pictures, let alone achieve good results. Therefore, the research on digital image encryption algorithms has become one of the research topics. Because chaotic systems are random or uncertain motions in a specific system, they have many unique characteristics, such as inherent randomness and sensitivity to initial values, making them suitable for designing image encryption

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2308045

algorithms.[1-6] .

People are actively researching fractional chaotic systems that reflect natural phenomena more accurately than integer chaotic systems[7,8]. Compared with traditional fractional-order system solving algorithms, the Adomian decomposition method has the advantages of fast convergence, It consumes fewer resources, and the calculation speed is faster[8-15]. For example, Wang et al. study the connection between image encryption and chaotic system with memristor, and put it into practice [8]. Zhang et al. analyzed the synchronization and adaptation problems of fractional-order chaotic systems [10]. INnubushi et al. studied the dynamics of fractional feedback control[16]. Natiq et al. study the connection between image encryption and fractional chaotic system, and put it into practice [17]. This paper proposes a three-dimensional fractional-order simplified unified system, and analyzes the dynamic characteristics of its fractional-order system.

At present, a variety of image encryption algorithms have been proposed based on chaotic systems[18-22]. For example, literature [19,23-25] proposed an image encryption algorithm based on discrete chaotic system, and literature [26-30] used hyperchaotic system image encryption algorithm. Literature [7,31-45] proposed a new image encryption algorithm based on chaotic system and DNA manipulation. Improve the security performance of image encryption algorithms this paper introduces the DNA mutation theory into the encryption algorithm.

The structure of the text is as follows. The second part introduces the operation rules of DNA sequence and the simplified 3D score calculation unified system. The third part gives the concrete steps of the image plus decoding scheme. The fourth part gives the simulation results and analyzes the security characteristics of the improved image encryption algorithm. Finally, Section 5 summarizes the research.

2. Basic principle

2.1 DNA mutation principle

Table 1 shows that only 8 of the 24 coding rules meet the Watson click to complete rule. In addition to the regular double addition/subtraction, the addition/subtraction of DNA is also obtained. So on the basis of the existing, we designed eight addition and subtraction rules based on the principle of subdivision. Table 2 shows DNA addition rule 1 and subtraction rule 1 taking DNA coding rule 1 as an example.

Table 1 The law of encoding

rule	1	2	3	4	5	6	7	8
00	A	A	T	T	G	G	C	C
01	C	G	C	G	T	A	T	A
10	G	C	G	C	A	T	A	T
11	T	T	A	A	C	C	G	G

Table 2 Addition and Subtraction rules

+	A	C	G	T	-	A	C	G	T
A	A	C	G	T	A	A	T	G	C
C	C	G	T	A	C	C	A	T	G
G	G	T	A	C	G	G	C	A	T
T	T	A	C	G	T	T	G	C	A

Therefore, according to the selection and matching of the actual DNA double-strand complementary, we designed the pairing principle. As shown in the formula, in the encryption process, the existing information is paired according to the pairing principle.

$$\begin{cases} a_i \neq S(a_i) \neq S\left(S(a_i)\right) \neq S\left(S\left(S(a_i)\right)\right) \\ a_i = S\left(S\left(S\left(S(a_i)\right)\right)\right) \end{cases} \tag{1}$$

From the equation, we can find the appropriate combination of six complementary base pairs, as shown in (2).

$$\begin{aligned} &(1)\,L1\,(A)=T,\ L1\,(T)=T,\ L1\,(C)=G,\ L1\,(G)=A; \\ &(1)\,L2\,(A)=T,\ L2\,(T)=T,\ L2\,(G)=C,\ L2\,(C)=A; \\ &(1)\,L3\,(A)=C,\ L3\,(C)=T,\ L3\,(T)=G,\ L3\,(G)=A; \\ &(1)\,L4\,(A)=C,\ L4\,(C)=T,\ L4\,(G)=T,\ L4\,(T)=A; \\ &(1)\,L5\,(A)=G,\ L5\,(G)=T,\ L5\,(T)=C,\ L5\,(C)=A; \\ &(1)\,L6\,(A)=G,\ L6\,(G)=T,\ L6\,(C)=T,\ L6\,(T)=A; \end{aligned} \tag{2}$$

In the actual operation and matching process of DNA double-strands, special changes

will occur, that is, they will not happen according to the predetermined design. One form of this is that mismatched bases are matched and individual mutations occur. Therefore, mutations in this will appear randomly, individuals with specific mutations, and specific mutations in genes, all are the results of random mutations. This kind of sudden application can meet the requirements of high randomness and high change rate of image information encryption.

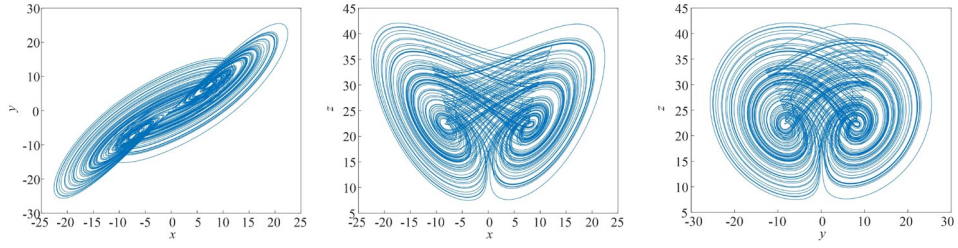
2.2 Dynamic analysis

This paper proposes a fractional three-dimensional simplified unified system based on Lu system and Chen system. Here, x_1, x_2, x_3 are the chaotic state variables q ($q \leq 1$) is the rank of the fractional equation, c is the system parameter

$$\begin{cases} T_t^q x_1 = (25c+10)(x_2 - x_1) \\ T_t^q x_2 = (27-6c)x_2 - x_1 x_3 \\ T_t^q x_3 = x_1 x_2 - (8+c)x_3 / 3 \end{cases} \tag{3}$$

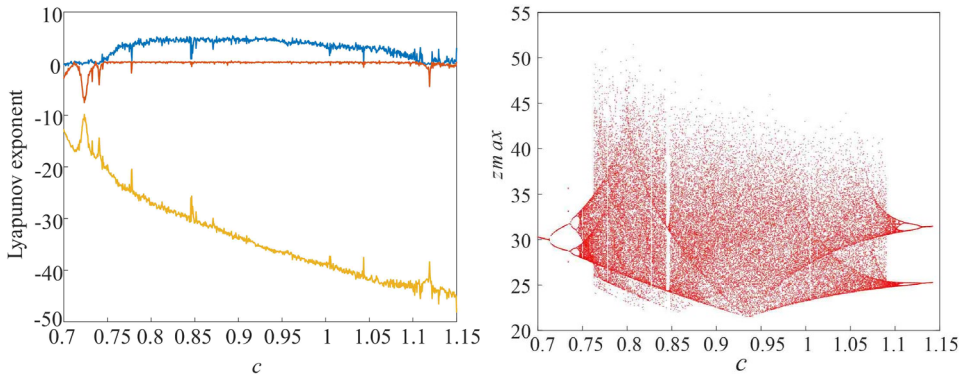
Let the system parameters c=0.9, q=0.9, the time step is t=0.001s, and the system initial value $[x_0, y_0, z_0]$ = [0.1,0.2,0.3]. At this time, the chaotic attractor phase diagram of the system is shown in Figure 1. At the same time, the Lyapunov index (L_1, L_2, L_3) = (5.0824, 0, -3.9031) can be calculated. Since the system has only one positive Lyapunov index value, and the sum of all Lyapunov indexes is negative, so the system is in a chaotic state under current conditions. When the system parameters q=0.9 and $c \in$[0.7:1.15], The period window is only close to c= 0.84. Figure 3 shows the Lyapunov exponent spectrum and bifurcation diagram of the system parameter c= 0.9, $c \in$[4.4: 1.15]. It can be seen from Figure 3 that the system is mostly chaotic in the range of $q \in$[0.4, 1.15], and there is only a periodic window near q= 0.84.

It can be seen from the above that the chaotic regions of the unified system with simplified scores are distributed in a larger region.



(a)x-y plane (b)x-z plane (c)y-z plane

Fig.1. Chaotic attractor phase diagram



(a) (b)

Fig. 2. Dynamical properties analysis of quantum logistic system ($c \in [0.7,1.15]$)

2.3 Test the chaos of the chaotic system

The international standard NIST test is used to verify the randomness of the chaotic sequence generated by the laser chaotic system. Pre-set the initial value of the system and each parameter, We extract samples from the values generated by the chaotic system, perform corresponding tests on the series, and compare the results with international standards.

The NIST test contains two standards to measure the randomness of the sequence. The first method is to test the pass rate of random sequence samples and count the test results.

$$\hat{s} = \sqrt{\frac{\hat{s}(1-\hat{s})}{i}}, \tag{4}$$

where P=1-α, i represents the number of samples in this series of numbers. If the calculation result falls within the critical value of s, it means that the chaotic system is very chaotic. Another equation (5)

$$x^2 = \sum_{i=1}^{10} \frac{(Fi - 0.1s)^2}{0.1s}, \tag{5}$$

This is an international test, a chaotic test of the random number sequence generated by the chaotic system. The evaluation results are given below. After comparing the results with international standards, it can be seen that the chaos of the system is very suitable for practical applications that require this feature.

Table 3. The randomness test result

Test name		Test sequence x		
		$p\text{-value}_{mean}$	$p\text{-value}_T$	Pass Rate
Frequency		0.4845	0.5913	0.98
Block Frequency		0.5251	0.5433	0.97
Runs		0.5254	0.2489	1
Longest Run		0.6738	0.3578	1
FFT		0.4701	0.3186	1
Universal		0.4772	0.1456	0.99
Approximate Entropy		0.5197	0.4586	0.99
Linear Complexity		0.5123	0.3986	1
Non Overlapping Template		0.4934	0.3895	0.98
Overlapping Template		0.5186	0.3188	0.99
Cumulative Sums	Forward	0.5071	0.9558	1
	Reverse	0.4627	0.9458	0.99
Serial	p-value[1]	0.4903	0.4980	0.99
	p-value[2]	0.5265	0.4286	1
Random Excursions	x=-4	0.3844	0.8728	0.9776
	x=-3	0.2798	0.5964	0.9776
	x=-2	0.2751	0.2279	0.9776
	x=-1	0.2721	0.6584	0.9833
	x=1	0.2639	0.4862	0.9811
	x=2	0.2946	0.1058	0.9777
	x=3	0.2778	0.3368	0.9814
	x=4	0.3301	0.0587	1

Table 4. Random excursions variant

Test name		Test sequence *x*		
		p-value$_{mean}$	p-value$_T$	Pass Rate
Random excursions variant	*x*=-9	0.1987	0.2015	1
	x=-8	0.3548	0.1950	1
	x=-7	0.3578	0.4859	0.9667
	x=-6	0.3484	0.1056	1
	x=-5	0.3587	0.1056	1
	x=-4	0.3015	0.4584	1
	x=-3	0.2954	0.6584	1
	x=-2	0.3150	0.2189	0.98
	x=-1	0.3018	0.4150	0.9867
	x=1	0.2979	0.4587	0.99
	x=2	0.2916	0.1950	0.99
	x=3	0.3254	0.1958	0.99
	x=4	0.3589	0.1245	1
	x=5	0.3152	0.4680	1
	x=6	0.3114	0.3842	1
	x=7	0.2986	0.6890	0.9667
	x=8	0.2956	0.7654	0.9866
	x=9	0.3015	0.4958	0.9866

3. Image encryption and decryption scheme

3.1 Encryption method

The first part uses a three-dimensional fractional chaotic graph to generate a random chaotic sequence, which will interfere with channels of the image. The second part uses three-dimensional fractional chaos coupled with DNA coding operations to obtain a series of very chaotic numbers. The scrambled channel perform diffusion operations, that is, complement, addition, and mutation operations. Figure 3 shows the main encryption process.

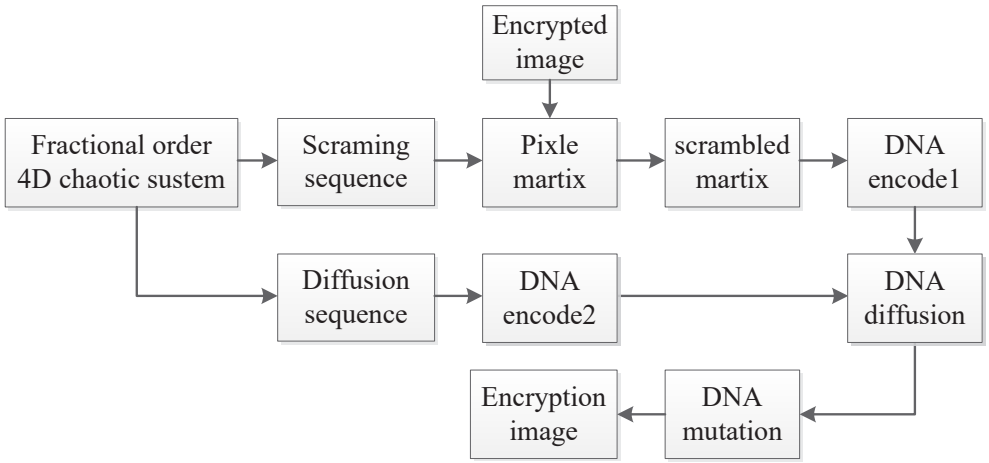


Fig.3. The architecture of encryption algorithm

The specific procedures we designed are listed here:

Step 1: The picture is decomposed into three primary colors and three pixel matrices. Set the key value and use the expression algorithm to calculate

The new initial value conditions of the fractional three-dimensional chaotic system:

$$s=\frac{\sum_{i=1}^{H} I(i,j)}{10^{10}} \tag{6}$$

$$\begin{cases} x_0^{'} = x_0 + s \\ y_0^{'} = y_0 + s \\ z_0^{'} = z_0 + s \end{cases} \tag{7}$$

Step 2: Set *L*. Subsequently, we stipulated that the setting of the disturbance step should be in accordance with the following standards:

$$\begin{cases} Cx=\mathrm{x}\times 10^{16},\mathrm{i}\times\mathrm{z} \\ Cy=\mathrm{y}\times 10^{16},\mathrm{i}\times\mathrm{z} \\ Cz=\mathrm{z}\times 10^{16},\mathrm{i}\times\mathrm{z} \end{cases} \tag{8}$$

$$k=\mathrm{mod}\left(\begin{bmatrix} 1 & Ax(a,b) \\ Ay(a,b) & Ax(a,b)\times Ay(a,b)+1 \end{bmatrix}\times[a;b],[H;W]+[1;1]\right) \quad (9)$$

Where: C_x,, C_y, C_z are matrix replacement coefficients, $k(1)$, $k(2)$ are the coordinates of the obtained random exchange values.

Step 3: We stipulate that any point in the matrix must be moved, but the movement of each point is prescribed. Calculate the corresponding exchange points according to the equation, and exchange them one by one. The matrix is reconstructed to obtain the scrambled image matrix tk. Among them, the rules we designed for pixel disturbance are evolved from genetic principles. Including a series of numerical operations, double-chain collocation operations, and burst collocation operations. The specific arrangements are as follows

Step 4: Replace the scrambled matrix with binary, so that the resulting matrix is enlarged. Using the encryption rules specified in the previous article, the binary matrix is compiled from a matrix defined by genetic letters.

Step 5: Set the initial values of the chaotic system x_0, y_0, z_0, to obtain the chaotic sequence, Then, by repeating (1) $n + h\times w$ times, the first n values are discarded. equation (10)

$$\begin{cases} a_1 = \mathrm{mod}\left(\left(\left\lfloor |x_i| \right\rfloor\right)\times 10^{16}, 256\right) \\ a_2 = \mathrm{mod}\left(\left(\left\lfloor |y_i| \right\rfloor\right)\times 10^{16}, 256\right) \\ a_3 = \mathrm{mod}\left(\left(\left\lfloor |z_i| \right\rfloor\right)\times 10^{16}, 256\right) \end{cases} \quad (10)$$

Step 6: Use the base substitution mutation rule in gene mutation to increase the randomness of image information encryption. The specific operation is that after the DNA diffusion operation, the value of each pixel has been expressed as an encrypted combination of four bases (such as ATCG), and the bases in each group are randomly exchanged. Get a new encrypted image pixel matrix C_1.

$$N(a,b)=\begin{cases} TGA & N(a,b)=(A), N(a,b+1)=(C), N(a,b+2)=(G) \\ ACG & N(a,b)=(T), N(a,b+1)=(G), N(a,b+2)=(A) \end{cases}$$

(11)

Step 7: Convert gene encoding matrix back to digital matrix

Step 8: Output the final finished image result to the finished process.

3.2 Decryption method

The decoding algorithm: First, reverse encoding, we use the calculation principle and gene double-strand operation principle specified in the previous article to change the encrypted information back to the gene letter sequence, and then use the encryption design to change the letter sequence back to the lion sequence ,use the three-dimensional system to generate chaotic sequences and transform them into pseudo-random sequences to restore the scrambling steps of the encrypted image, obtain the original decrypted pixel matrix, and encode it into a digital matrix, and restore the original image through these three matrices.

Step 1: The encrypted image is still divided into three matrices according to the previous operation, and genetic coding is carried out according to the pre-set coding rules.

Step 2: Use gene mutation base substitution mutation to carry out reverse mutation change. Generate the encrypted matrix of the original DNA.

Step 3: Convert the chaotic sequence. Use subtraction to recover the DNA sequence that encrypts the image to obtain the encrypted diffusion sequence. Then, according to the predetermined base pair principle, the chaotic sequence generated by the chaotic system is also coded into a gene matrix using the coding rules.

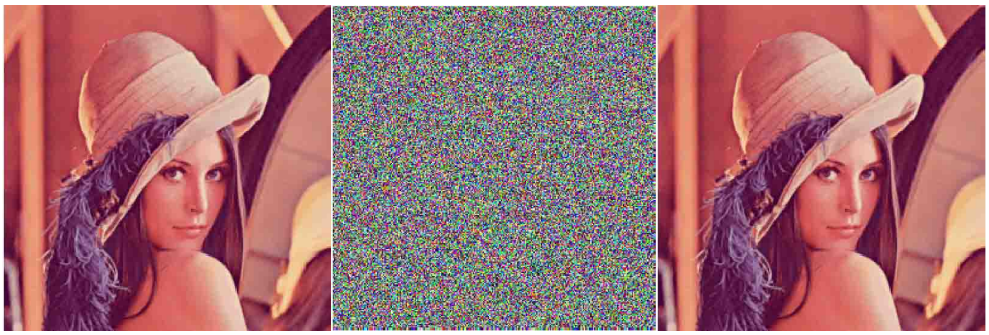
Step 5: Reverse scrambling for reverse operation is performed, and the specific operation process is shown in the encryption process.

Step 6: The three primary color pixel matrixes restore the digital matrix, and the solved image is obtained through the three matrices.

4. Simulation and performance analysis

4.1 Algorithm simulation results

Verify the validity of the proposed cryptographic algorithm. 256x256 Lena images are used for algorithm testing. Set the parameters, initial values and levels of the fractional chaotic system. Based on the MATLAB platform, the corresponding experimental test results are shown in Figure 3.



(a) (b) (c)

Fig.4. Encryption and decryption test (a)original image (b)encrypted image (c)decrypt the image

4.2 Key space analysis

A good image encryption algorithm needs enough key space to resist brute force cracking. In our cryptography, the key space size corresponding to the key will be given. c, q, x_0, y_0, z_0 is about 2^{249}; for the other parameter, because DNA has four kinds of acids and bases, the key space is $2^2\times2^6\times2^{20}=2^{28}$. The key space of the proposed algorithm is 2^{277}, therefore, the calculation method has enough space to resist brute force guessing attacks.

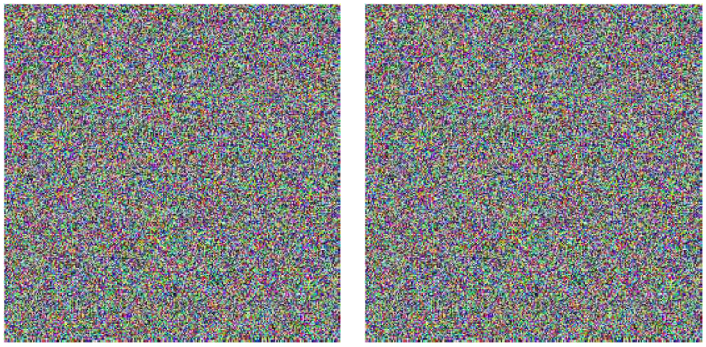
Table 5. Table of key space

	Proposed algorithm	Chen's [23]	Luo's [31]	Li's [33]	Shu's [41]	Wang's [46]
Key space	2^{277}	2^{266}	2^{213}	2^{267}	2^{248}	2^{215}

4.3 Key sensitivity analysis

In this experiment, the keys (x_0+10^{-15}), (y_0+10^{-15}), (z_0+10^{-15}), $(c+10^{-15})$ of the decryption algorithm were changed, and then the encryption Table 6 shows the comparison results. We compare, and the comparison result clearly proves that the correctly decrypted image is very different from the decrypted result with slightly modified initial value, which proves that our design is very sensitive to the initial setting, and the similarity rate is less than 0.1%.

(a) (b)



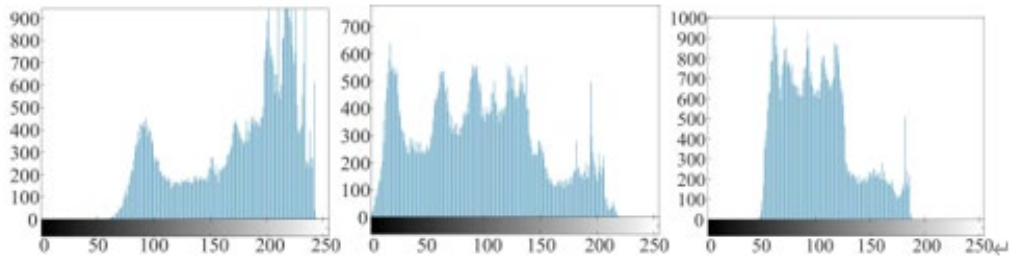
(b) (d)

Fig.5. Sensitivity to initial value (a) Decrypt the complete image (b) x_0+10^{-15} (c) y_0+10^{-15} (d) z_0+10^{-15}

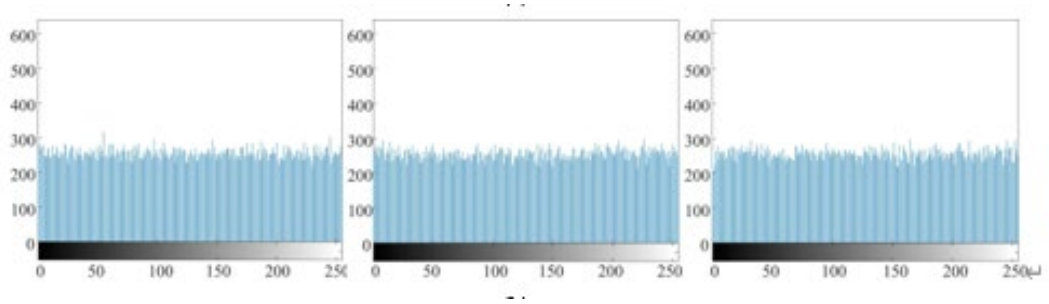
4.4 Statistical performance analysis

4.4.1 Histogram analysis

The histogram test is shown in the figure, and the experiment proves that the performance is good.



(a)



(b)

Fig.6. Histogram analysis results (a)plain image (b)permuted image

4.4.2 Image correlation coefficient

There is a strong correlation between adjacent pixels of the information it stores. The purpose of the image encryption algorithm to break the correlation between adjacent pixels. The calculation formula for judging the pixel correlation coefficient is:

$$r_{xy} = \frac{\operatorname{cov}(a,b)}{\sqrt{D(A)D(B)}}, \tag{12}$$

$$\operatorname{cov}(x,y) = E\left\{\left[\mathrm{x} - d(x)\right]\left[y - D(x)\right]\right\}, \tag{13}$$

$$E(x) = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{14}$$

$$D(a) = \frac{1}{N}\sum_{i=1}^{N}\left[x_i - E(a)\right]^2. \tag{15}$$

The table data shows that the original image has significant correlation, while the small correlation of the encrypted image indicates that the effect of the encryption algorithm meets the requirements.

In order to see more directly whether there is a correlation. **Figure** 7 shows the correlation distribution between the level of the Lena image and the adjacent pixels. The result proves that the surroundings of each pixel of the image before encryption are very related, which makes the encryption process full of difficulties. And the arrangement of the pixels is on the diagonal line. However, as shown in the second half of **Figure** 7, the pixel distribution of the encrypted image is very uniform and will not be counted, which greatly reduces the

relationship between different encrypted images in the encrypted image.

Table 6. Correlation coefficients in *R, G, B* channels

Channel	Direction	Plain image	Cipher image	Ref [7]
R Channel	Horizontal	0.9556	0.0024	0.0095
	Vertical	0.9780	6.3593×10^{-4}	-0.0026
	Diagonal	0.9434	-0.0010	0.0078
G Channel	Horizontal	0.9443	0.0067	0.0183
	Vertical	0.9711	2.8685×10^{-4}	0.0001
	Diagonal	0.9301	-9.3236×10^{-4}	-0.0039
B Channel	Horizontal	0.9280	-8.7964×10^{-4}	0.0034
	Vertical	0.9575	3.9657×10^{-4}	-0.0035
	Diagonal	0.9093	0.0044	0.0052

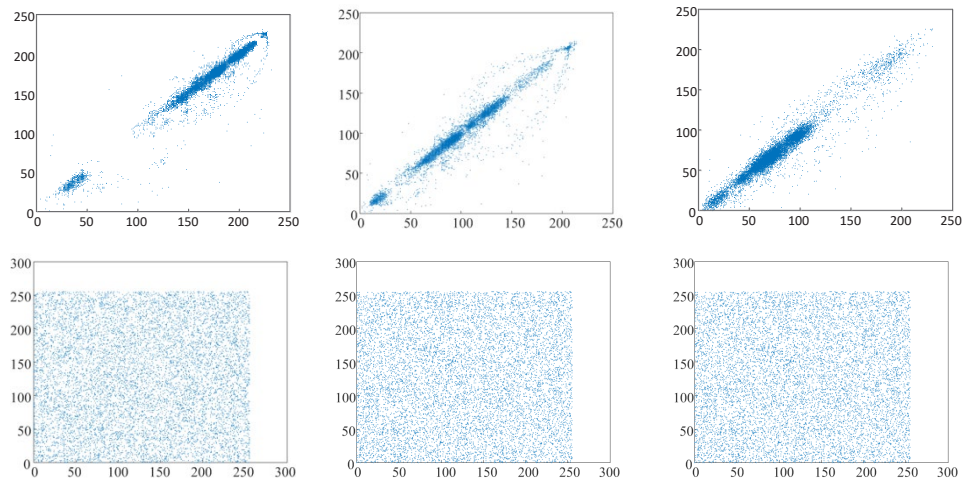


Fig.7. The correlation coefficient analysis: (a) “Lenna” image (b) “Fruits” image (c) “Flowers” image (d) “Pepper” image

4.5 Information entropy

We often use this characteristic value to judge whether the information is chaotic. We take the average of the average chaos of the input information. As shown in Eq. (16):

$$H(m)=\sum_{i=1}^{L-1} I(m_i)\times\log\frac{1}{I(m_i)}. \tag{16}$$

where, $I(m_i)$ is the probability of m_i appearing, L is the number of statistical feature points m_i.

Table 10 shows the average information entropy of each image, and the test value of the new algorithm meets the theoretical requirements. This result shows that the obtained image has completely lost the possibility of discrimination, and his information entropy is very close to the theoretical value. The results show that the encrypted image can be approximately regarded as a random image. This proves the superiority of the algorithm.

Table 8. information entropy

Image name	Information entropy
Lena	7.9972

Table 9. Another scheme

Encryption scheme	IE
Proposed scheme	7.9975
Liu's [5]	7.9964
Li's [19]	7.9973
Sun's [30]	7.9967
Liu's [47]	7.9971
Yang's [48]	7.9981

4.6 Differential attack analysis

In order to verify the sensitivity of the ciphertext to the plaintext, we use differential attacks to test. Usually, two standards, pixel change rate (NPCR) and uniform average change rate (UACI), are commonly used to judge the performance of encryption algorithms.

The values of NPCR and UACI values are calculated by:

$$NPCR=\frac{\sum_{j=1}^{L} E(i,j)}{L}\times 100\%, \tag{17}$$

$$UCAI = \frac{1}{L}\sum_{\substack{i=1\\j=1}}^{L}\frac{I(i,j)-I_1(i,j)}{255}\times 100\%. \tag{18}$$

$$D(\mathrm{x},y)=\begin{cases}1 & C((\mathrm{x},y)\neq C(\mathrm{x},y)\\ 0 & C(\mathrm{x},y)=C(\mathrm{x},y)\end{cases}. \tag{19}$$

We use the following two reference values to distinguish the performance of the system. One of the criteria for judging is as follows:

$$N_{\alpha}^{*}=\frac{F-\Phi^{-1}(\alpha)\sqrt{F/L}}{F+1}. \tag{20}$$

The UCAI thresholds (U_{α}^{*-}, U_{α}^{*+}) are obtained as

$$\begin{cases}D_{\alpha}^{*-}=\mu_u-c^{-1}(a/2)b_u\\ D_{\alpha}^{*+}=\mu_u+c^{-1}(a/2)b_u\end{cases}, \tag{21}$$

We analyze this algorithm from the perspective of three-color pixels, and the results obtained are averaged to show the advantages of the system. The results are obtained by comparing with other algorithms. A detailed comparison is given in Table 10.11.12, which proves that the algorithm has classic information theory security and good performance beyond classic image encryption algorithms.

Table 10. NPCR Values

Image name	NPCR (mean)	Eligibility boundary $D_{0.05}^{*-}$=99.5963%	$D_{0.01}^{*-}$=99.5527%	$D_{0.001}^{*-}$=99.5341%
Lena	99.67%	effective	effective	effective

Table 11.UACI Values

Image name	UCAI (mean)	Critical Value		
Lena	33.45%	effective	effective	effective

4.7 Anti-noise attack performance analysis

The intensity of 0.05, 0.06, 0.07 Gaussian noise was added in the encryption process. And decrypt the contaminated image, the experimental results are shown in **Figure** 8. The test results prove that the algorithm has the ability to resist noise pollution.

Table 12. compare with another scheme (Take Lena image as an example)

Encryption scheme	NPCR	UACI
Proposed scheme	99.67%	33.45%
Liu's [5]	99.62%	33.38%
Khwvana's [18]	99.64%	33.46%
Luo's [31]	99.63%	33.42%
Sun's [30]	99.61%	33.45%
Yang's [48]	99.61%	33.44%

Fig.8. Anti-pollution attack test.

5. Conclusion

We analyzed the dynamic characteristics of the simplified unified system based on the 3D fractional scale of the ADM algorithm. The data showed that the chaotic system has very good advantages, and we used his principle in the practical application of encrypted images. which indicates that they are suitable for chaotic cipher systems. In this paper, we propose a new image encryption method using fractional chaotic system This paper analyzes the security of the encryption algorithm, and applies the fractional simplified unified system and the principle of DNA mutation to the image encryption algorithm. Need to be improv. It provides a theoretical and practical basis for cryptography confidential communication and information security.

References

[1] Lai Q, Kuate P D K, Liu F, et al. An Extremely Simple Chaotic System with Infinitely Many Coexisting Attractors[J]. Circuits and Systems II: Express Briefs, IEEE Transactions on, 2019, PP(99):

1-1.

[2] Lai Q, Wan Z, Kuate P D K, et al. Coexisting attractors, circuit implementation and synchronization control of a new chaotic system evolved from the simplest memristor chaotic circuit[J]. Communications in Nonlinear ence and Numerical Simulation, 2020: 105341.

[3] Murillo-Escobar M A, Cruz-Hernández C, Abundiz-Pérez F, et al. A RGB image encryption algorithm based on total plain image characteristics and chaos[J]. Signal Processing, 2015, 109: 119-131.

[4] Hikal N A, Eid M M. A new approach for palmprint image encryption based on hybrid chaotic maps[J]. Journal of King Saud University - Computer and Information Sciences, 2018.

[5] Liu H, Nan H. Color image security system using chaos-based cyclic shift and multiple-order discrete fractional cosine transform[J]. Optics & Laser Technology, 2013, 50: 1-7.

[6] Chen X, Hu C J. Adaptive medical image encryption algorithm based on multiple chaotic mapping[J]. Saudi J Biol Sci, 2017, 24(8): 1821-1827.

[7] Yang F, Mou J, Luo C, et al. An improved color image encryption scheme and cryptanalysis based on a hyperchaotic sequence[J]. Physica Scripta, 2019, 94(8): 085206.

[8] Wang Y, Chang Y, Wu Q, et al. Fuzzy model of hyperchaotic laser systems with parameters perturbation[J]. High Power Laser & Particle Beams, 2012, 24(09): 2063-2067.

[9] Glushkov A V, Buyadzhi V V, Kvasikova A S, et al. Non-Linear Chaotic Dynamics of Quantum Systems: Molecules in an Electromagnetic Field and Laser Systems[J], 2017.

[10] Zhang L W, Shao M. Chaotic Synchronization with Single-Ring Erbium-Doped Fiber Laser Systems[C]. International Conference on Computational Intelligence & Security, 2010.

[11] Louis, M., Pecora, et al. Synchronization of chaotic systems[J]. Chaos An Interdisciplinary Journal of Nonlinea, 2015.

[12] Rosenbluh M, Aviad Y, Cohen E, et al. Spiking Optical Patterns and Synchronization[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2007, 76(4 Pt 2).

[13] Stoffels E, Kieft I E, Sladek R E J, et al. Plasma needle for in vivo medical treatment: recent developments and perspectives[J]. Plasma Sources Ence & Technology, 2006, 15(4): S169.

[14] Chen H K. Global chaos synchronization of new chaotic systems via nonlinear control[J]. Chaos Solitons & Fractals, 2005, 23(4): 1245-1251.

[15] Yassen M T. Chaos synchronization between two different chaotic systems using active control[J]. Chaos Solitons & Fractals, 2005, 23(1): 131-140.

[16] Inubushi M, Yoshimura K, Arai K, et al. Physical random bit generators and their reliability: focusing on chaotic laser systems[J]. Nonlinear Theory & Its Applications Ieice, 2015, 6(2): 133-143.

[17] Natiq H, Said M R M, Al-Saidi N M G, et al. Dynamics and Complexity of a New 4D Chaotic Laser System[J]. Entropy, 2019, 21(1): 34.

[18] Khurana M, Singh H. Two level phase retrieval in fractional Hartley domain for secure image

encryption and authentication using digital signatures[J]. Multimedia Tools and Applications, 2019, (3).

[19] Li P, Xu J, Mou J, et al. Fractional-order 4D hyperchaotic memristive system and application in color image encryption[J]. EURASIP Journal on Image and Video Processing, 2019, 2019(1).

[20] Vilardy J M, Torres C O, Jimenez C. Fractional convolution and nonlinear operations applied to the image encryption[J]. Journal of Physics Conference, 2019.

[21] Li G-D, Wang L-L. Double chaotic image encryption algorithm based on optimal sequence solution and fractional transform[J]. Visual Computer, 2019, 35(9): 1267-1277.

[22] Chen L, Hao Y, Huang T, et al. Chaos in fractional-order discrete neural networks with application to image encryption[J]. Neural Networks, 2020, 125: 174-184.

[23] Chen G, Mao Y, Chui C K. A symmetric image encryption scheme based on 3D chaotic cat maps[J]. Chaos, Solitons & Fractals, 2004, 21(3): 749-761.

[24] Zhou N, Wang Y, Gong L. A Novel Scheme of Image Encryption Based on Synchronization of Fractional Order Chaotic Systems[J]. Optics Communications, 2011, 284(13): 3234-3242.

[25] Sajasi S, Eftekhari Moghadam A-M. An adaptive image steganographic scheme based on Noise Visibility Function and an optimal chaotic based encryption method[J]. Applied Soft Computing, 2015, 30: 375-389.

[26] Wang X-Y, Zhang Y-Q, Zhao Y-Y. A novel image encryption scheme based on 2-D logistic map and DNA sequence operations[J]. Nonlinear Dynamics, 2015, 82(3): 1269-1280.

[27] Chai X, Gan Z, Lu Y, et al. A novel image encryption algorithm based on the chaotic system and DNA computing[J]. International Journal of Modern Physics C, 2017, 28(05): 1750069.

[28] Zhang L-M, Sun K-H, Liu W-H, et al. A novel color image encryption scheme using fractional-order hyperchaotic system and DNA sequence operations[J]. Chinese Physics B, 2017, 26(10): 100504.

[29] Xu B, Wang G, Iu H H-C, et al. A memristor–meminductor-based chaotic system with abundant dynamical behaviors[J]. Nonlinear Dynamics, 2019, 96(1): 765-788.

[30] Suryadi M T, Satria Y, Fauzi M. Implementation of digital image encryption algorithm using logistic function and DNA encoding[J]. Journal of Physics: Conference Series, 2018, 974: 012028.

[31] Luo Y, Du M, Liu J. A symmetrical image encryption scheme in wavelet and time domain[J]. Communications in Nonlinear Science and Numerical Simulation, 2015, 20(2): 447-460.

[32] Silva-García V M, Flores-Carapia R, Rentería-Márquez C, et al. Substitution box generation using Chaos: An image encryption application[J]. Applied Mathematics and Computation, 2018, 332: 123-135.

[33] Li X, Zhou C, Xu N. A secure and efficient Image encryption algorithm based on DNA coding and spatiotemporal chaos[J]. International Journal of Network Security, 2018, 20: 110-120.

[34] Bukharmetov M, Nyrkov A, Sokolov S, et al. Robust Method for Protecting Electronic

Document on Waterway Transport with Steganographic Means by Embedding Digital Watermarks into Images[J]. Procedia Engineering, 2017, 178: 507-514.

[35] Qin Y, Wang Z, Wang H, et al. Robust information encryption diffractive-imaging-based scheme with special phase retrieval algorithm for a customized data container[J]. Optics and Lasers in Engineering, 2018, 105: 118-124.

[36] Zhang S, Liu H, Li S. Robust adaptive control for fractional-order chaotic systems with system uncertainties and external disturbances[J]. Advances in Difference Equations, 2018, 2018(1).

[37] Farwa S, Muhammad N, Bibi N, et al. RETRACTED: Fresnelet approach for image encryption in the algebraic frame[J]. Applied Mathematics and Computation, 2018, 334: 343-355.

[38] Khan A, Bhat M A. Projective synchronization via feedback controller of fractional-order chaotic systems[J]. International Journal of Modelling and Simulation, 2019, (4): 1-9.

[39] Peng Y, Sun K, He S, et al. Parameter Identification of Fractional-Order Discrete Chaotic Systems[J]. Entropy, 2019, 21(1): 27.

[40] Khennaoui A-A, Ouannas A, Bendoukha S, et al. On fractional–order discrete–time systems: Chaos, stabilization and synchronization[J]. Chaos, Solitons & Fractals, 2019, 119: 150-162.

[41] Shukla M K, Siva D, Mahajan A, et al. A Novel Scheme of Image Encryption Based on Synchronization of Fractional Order Chaotic Systems[C]. 2018 International Conference on Intelligent Circuits and Systems (ICICS), 2018.

[42] Sun S. A Novel Hyperchaotic Image Encryption Scheme Based on DNA Encoding, Pixel-Level Scrambling and Bit-Level Scrambling[J]. IEEE Photonics Journal, 2018, 10(2): 1-14.

[43] Wang X, Li P, Zhang Y. A novel color image encryption scheme using DNA permutation based on the Lorenz system[J]. CrossMark, 2018, 77: 6243-6265.

[44] Ye X, Wang X, Gao S, et al. A new chaotic circuit with multiple memristors and its application in image encryption[J]. Nonlinear Dynamics, 2019.

[45] Chen C, Sun K, He S. An improved image encryption algorithm with finite computing precision[J]. Signal Processing, 2020, 168: 107340.

[46] Wang X, Zhang Y, Zhao Y. A novel image encryption scheme based on 2-D logistic map and DNA sequence operations[J]. CrossMark, 2015, 82: 1269-1280.

[47] Liu W, Sun K, He Y, et al. Color Image Encryption Using Three-Dimensional Sine ICMIC Modulation Map and DNA Sequence Operations[J]. International Journal of Bifurcation and Chaos, 2017, 27(11): 1750171.

[48] Yang F, Mou J, Ma C, et al. Dynamic analysis of an improper fractional-order laser chaotic system and its image encryption application[J]. Optics and Lasers in Engineering, 2020, 129: 106031.

An image encryption scheme based on swapping operation and block scrambling

Xinyu Gao, Junqiao Liu, Huizhen Yan

{ email: dalianliujunqiao@126.com}

School of Information Science and Engineering, Dalian polytechnic University, Dalian 116034, China

Abstract. For purpose of meeting the requirement of encryption security, this paper brings an image encryption scheme based on chaotic system, block scrambling, swapping of rows and columns. The chaotic matrixes and index sequences are obtained to realize the rows and columns exchange and block scrambling, and the diffusion method is used to make the pixels fully diffuse so that the security can be improve. The image gray distribution, the correlation, the key space and the robustness against cropping attack are analyzed and simulated. The results show that the proposed scheme has high security and provides theoretical guidance.

Keywords: Swapping operation; Block scrambling; Chaotic system.

1 Introduction

Image information has the characteristics of intuitive and vivid, which is widely used in the digital age, but it is vulnerable to various attacks by hackers [1]. In recent years, due to the sensitivity of initial value, aperiodic of chaos, many scholars have used chaotic systems to generate random data streams to encrypt images [2-4]. There are many kinds of chaos system from one-dimensional to five-dimensional or even seven-dimensional [5-8]. One-dimensional map is simple, and the generation time is short, but the distribution of blank windows is not uniform. The chaotic system with higher dimensions has stronger chaotic dynamics, but it takes a long time. The encryption process usually has two steps: pixels scrambling in which the pixel positions are changed and diffusion, and in diffusion the pixel values are changed [9]. There are many methods of scrambling, such as Arnold scrambling, two-dimensional image spreading into one-dimensional image scrambling, two-dimensional image directly scrambling [10]. The common methods of diffusion are addition and modulo operation, XOR operation and cyclic shift operation.

Large number of encryption algorithms about chaos have been suggested [11-14]. Wang et al. presented a plaintext-related image encryption scheme according to Josephus traversing and pixel permutation [15]. Zhu et al. [16] used 3-D cat map 3-D DNA level permutation scheme for encryption scheme. Farah et al. [17] introduced a hybrid chaotic map and used it to generate chaotic sequence to encrypt a gray image. Sun et al. encrypted color images by using

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307948

a new memristive chaotic system [18]. Similarly, Hua et al. introduced chaotic maps with good dynamic performance and used one of their to encrypt images [19]. The combination of chaotic system and image encryption has good performance and gradually becomes a research hotspot [20].

This paper proposed an encryption scheme which uses block operation and swapping operation. The scrambling process swaps the rows and columns of the image, then breaks the image into small pieces for further scrambling. The diffusion process enhances the security of image encryption.

This paper is made up of six parts. In Section 2, the chaotic system, swapping operation and block scrambling are introduced. The image encryption scheme and simulation results are shown in Section 3 and Section 4. Section 5 and Section 6 give security analysis and conclusions.

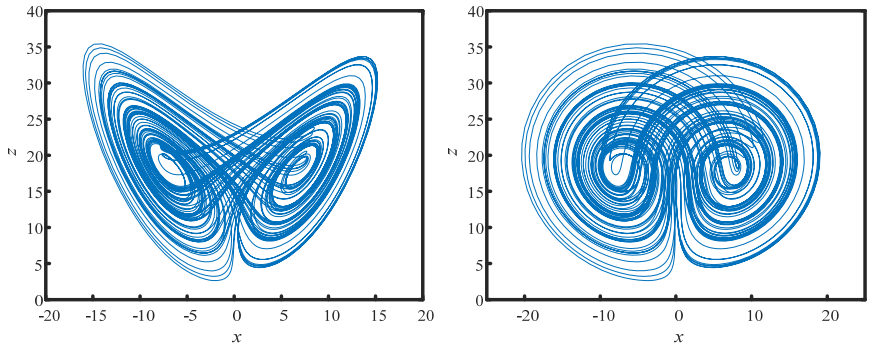
2 Preliminaries

2.1 Lorenz system

In the encryption scheme, simple Lorenz system is defined by

$$\begin{cases} \dot{x}=10(y-x) \\ \dot{y}=-xz+(24-4c)x+cy \\ \dot{z}=xy-\dfrac{8}{3}z \end{cases}, \tag{1}$$

Where c∈[-1.59,7.75]. In this paper, we set control parameter c=2, initial condition $[x_0, y_0, z_0]$ = [1, 2, 3], iteration time step h=0.001. The attractor phases of simple Lorenz system are given in Figure 1.



(a) (b)

Fig.1 Attractors phase: (a) *x*-*z* plane, (b) *y*-*z* plane

2.2 Swapping operation

In this section, the swapping operation is introduced. Firstly, the index sequences q_1 and q_2 are as same as chaotic sequences x and y. Secondly, sequence $q1$ and sequence $q2$ are used for row swapping and column swapping. The details are described as follows:

Step 1: By using the control parameter and initial conditions, three chaotic sequences x, y, z can be produced by equation (1).

Step 2: The sequences are dealt with Eq. (2) and Eq. (3)

$$\begin{cases} X = x\left(end - 512\times512 + 1 : end\right) \\ Y = y\left(end - 512\times512 + 1 : end\right) \\ Z = z\left(end - 512\times512 + 1 : end\right) \end{cases}, \tag{2}$$

$$\begin{cases} X(i) = \left(\left(X(i) + 100\right)\times10^{10}\right) \backslash 256 + 1 \\ Y(i) = \left(\left(Y(i) + 100\right)\times10^{10}\right) \backslash 256 + 1 \\ Z(i) = \left(\left(Z(i) + 100\right)\times10^{10}\right) \backslash 256 + 1 \end{cases}, \tag{3}$$

where equation (2) and equation (3) include the operation of remainder after division. Then, three sequences are changed to three matrixes with size of 512×512.

Step 3: Diagonal elements from the matrixes X and Y are picked out and arranged as the index sequences q_1 and q_2.

Step 4: The rows and columns are processed by equation (4) and equation (5)

$$\begin{cases} t_1 = C(i,:) \\ C(i,:) = C\left(q_1(i),:\right) \\ C\left(q_1(i),:\right) = t_1 \end{cases}, \tag{4}$$

$$\begin{cases} t_2 = C(:,j) \\ C(:,j) = C\left(:,q_2(j)\right) \\ C\left(:,q_2(j)\right) = t_2 \end{cases}, \tag{5}$$

we can get image C which has swapped the rows and columns.

2.3 Block scrambling

Block scrambling is to divide the original image into $m1\times n1$ small pieces and scramble those small pieces. The scrambling step consists of the following four steps:

Step 1: The chaotic matrixes are obtained as step 1 and step 2 in section 2.2. Then, three sequences a, b and q are processed by equation (6)

$$\begin{cases} a = X(:) \\ b = Y(:) \\ q = (a+b) \setminus 256 + 1 \end{cases}. \tag{6}$$

Where $(a+b)\backslash 256$ is the operation of taking the remainder.

Step 2: The original image is divided into $m1 \times n1$ small blocks with size of $m \times n$.

Step 3: All image blocks are arranged into block vectors. Then, the block vectors are processed by equation (7)

$$\begin{cases} t_3 = C(i) \\ C(i) = C(q(i)), \\ C(q(i)) = t_3 \end{cases} \tag{7}$$

we can obtain image C which has scrambled by block scrambling.

Step 4: Block vectors are pieced into $(m1 \times m) \times (n1 \times n)$ image.

3 Encryption scheme

The whole encryption procedures are drawn in Figure 2.

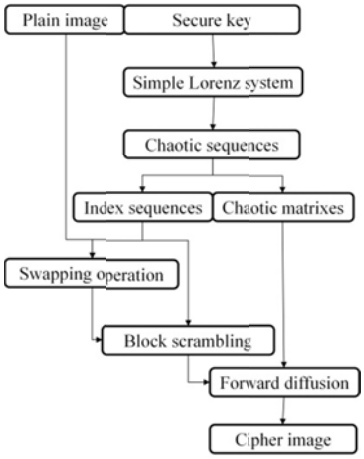


Fig.2 The diagram of encryption algorithm

Step 1: According to the parameter x_0, y_0, z_0 and k, the chaotic matrixes X, Y, Z and index sequences q_1, q_2 can be obtained as the step 1-3 in section 2.2.

Step 2: Each row of the image is put into the swap space t_1, then the corresponding swap sequence is found by using the pseudo-random index sequence and swapped.

Step 3: Each column of the image is put into the swap space t_2, then the corresponding swap sequence is found by using the pseudo-random index sequence and swapped.

Step 4: The image is divided into $m1 \times n1$ blocks, then the image is scrambled in blocks, as presented in section 2.3.

Step 5: The first pixel is following the operation by equation (8)

$$C(1,1) = I(1,1) \oplus X(1,1) \oplus Z(1,1), \tag{8}$$

where $I(1,1) \oplus X(1,1)$ means xor operation.

Step 6: The first row is processed by equation (9)

$$C(1,j) = I(1,j) \oplus X(1,j) \oplus C(1,j-1), \tag{9}$$

where j from 2 to end.

Step 7: The first row is processed by equation (10)

$$C(i,1) = I(i,1) \oplus X(i,1) \oplus C(i-1,1), \tag{10}$$

where i from 2 to end.

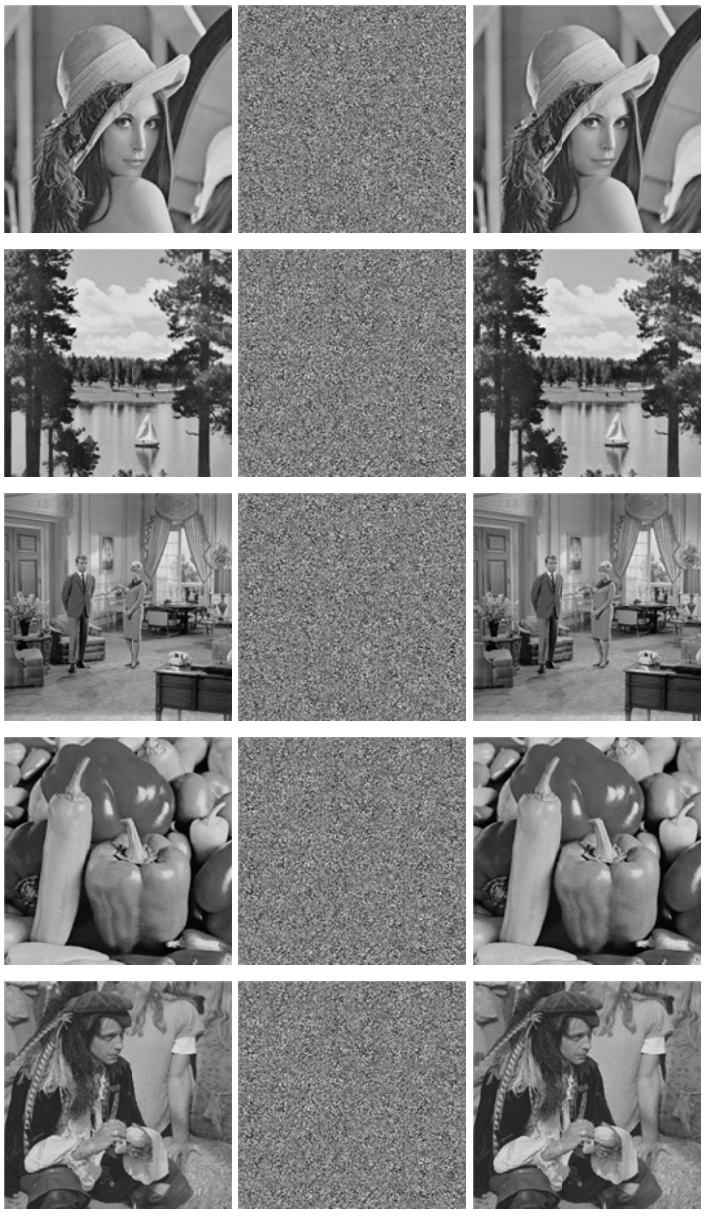
Step 8: The rest of the image pixels are processed by equation (11)

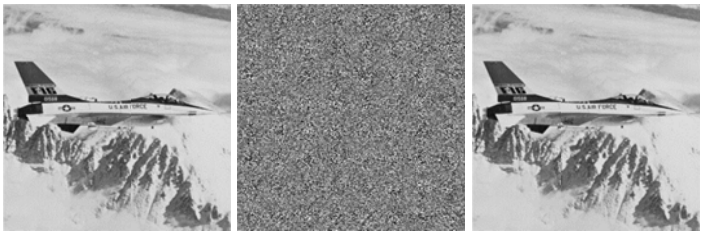
$$C(i,j) = I(i,j) \oplus X(i,j) \oplus C(i-1,j) \oplus C(i,j-1). \tag{11}$$

4 Simulation results

The experiments are simulated on MATLAB (version R2018a) to testify the security of the image algorithm. In experiment, the test images “Lena”, “Lake”, “Livingroom”, “Peppers”, “Man”, and “Jet” with size of 512 ×512 are tested. Simplified Lorenz system initial keys x_0, y_0, z_0 and c are fixed as 1, 2, 3 and 2. Figure 3 gives simulation results: test images mentioned above is arranged in the first column of Figure 3 (a), the corresponding encrypted images and the decrypted images after applying the decryption strategy to it are shown in Figure 3(c) and (d), respectively. First of all, through the observation of the image, the encrypted image obtained after encryption cannot be seen visually with any valuable information. Secondly, the

image obtained by the decryption experiment is compared with the test image, and the experimental results show that the encryption algorithm mentioned in this paper is reversible. In summary, this encryption algorithm meets the two most basic requirements of image encryption, namely, the hiding of information and the recoverability of encrypted images.





(a) (b) (c)

Fig.3 Experimental results of encryption and decryption

5 Security analysis

5.1 Key space test

For an attacker, the easiest way to crack the encryption system may be a brute force attack. In other words, when the key space of the encryption system is small, the attacker can crack the encryption system after a limited number of tests whether the key is true or not. Therefore, after research by cryptographers, the key space of a good encryption system should be at least 2^{100}.

The keys in our algorithm are composed of control parameter c and initial values x_0, y_0, z_0 and the specific value of the key space is approximately 2^{212} after giving the accuracy of the computer 10^{-16}. This shows that brute force attack [21] is not feasible for the encryption system with such a large key space.

5.2 Information entropy

Image information entropy is a statistical form of features, which reflects the average amount of information in the image. It is generally believed that in an encryption system, when the image is encrypted, the higher the information entropy of the image, the less visible information of the image. This entropy is defined by equation (12)

$$H(S)=\sum_{S} P(S_i)\log_2\frac{1}{P(S_i)}bit\,, \tag{12}$$

where S_i represents source that value belongs [0 255], $P(S_i)$ represents probability of occurrence of S_i. When the gray value distribution of the image is consistent, the information entropy value of the image is the ideal value 8. That is to say, the closer the information entropy of the encryption algorithm is to 8, the better its performance.

Table 1 lists that specific values of information entropy of plain images and the encrypted images, the test results are close to the expectation value. The test results found that the information entropy of the encrypted image has been greatly improved compared to the information entropy of the flat image, and its information entropy value is around 7.9993, which indicates that the entropy-related attack [22] is invalid for the encryption algorithm proposed in this paper [23].

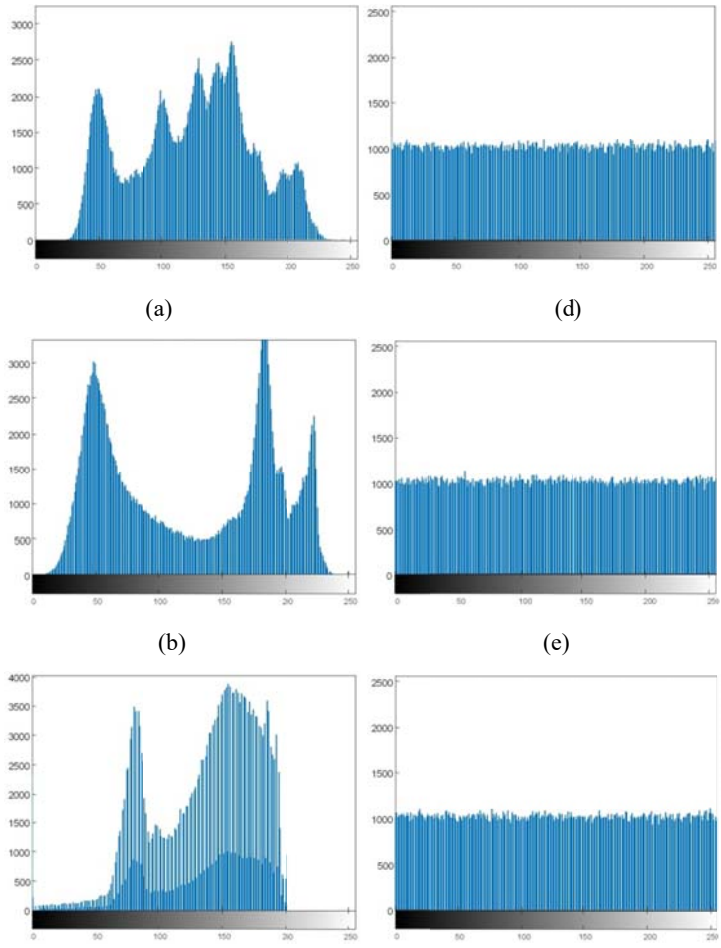
Table 1 Information entropies of six grayscale images

Image	Lena	Lake	Living room	Peppers	Man	Jet
Plain image	7.4451	7.4826	7.2925	6.7624	7.2367	6.7135
Cipher image	7.9992	7.9992	7.9993	7.9993	7.9993	7.9993

5.3 Statistic analysis

5.3.1 Histogram

Histogram is the most intuitive way to view the distribution of image pixel information. Histogram distribution of encrypted image from a good encryption scheme should be uniform. Figure 4 presents the histograms of the test images and its encrypted images in this experiment. From the Figure 4, the histograms of encrypted images distribution are fairly uniform while their corresponding original images are undulating. It can get a conclusion that statistical attacks [16] alone are not capable of deciphering the encryption scheme proposed in this paper.



(c) (f)

Fig.4 Histograms of the original and encrypted images, (a) (d) original Lena and its encrypted image; (b) (e) original Lake and its encrypted image; (c) (f) original Woman and its encrypted image.

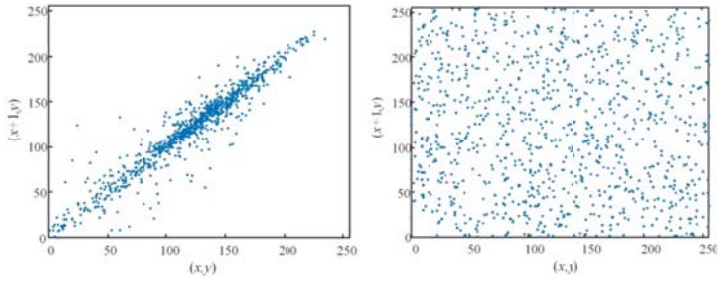
5.3.2 Correlation coefficient

Correlation test refers to the degree of connection between adjacent pixels where the correlation is measured by correlation coefficient. For an original image, absolute value of correlation coefficient is about 1 [24]. On the contrary, correlation will be destroyed after image encryption while the closer the absolute value of the correlation is to 0, the better the encryption effect of the image. The correlation coefficient can be derived from Equation (13)

$$\begin{cases} r_{xy} = \dfrac{\operatorname{cov}(x,y)}{\sqrt{D(x)}\sqrt{D(y)}} \\ \operatorname{cov}(x,y) = \dfrac{1}{N}\sum_{i=1}^{N}\left(x_i - E(x)\right)\left(y_i - E(y)\right) \\ D(x) = \dfrac{1}{N}\sum_{i=1}^{N}\left(x_i - E(x)\right)^2 \\ E(x) = \dfrac{1}{N}\sum_{i=1}^{N} x_i \end{cases}, \tag{13}$$

where x and y are gray values of two adjacent pixels x, which can be the horizontal (H), vertical (V) and diagonal (D) directions.

The correlation analysis is shown in Table 2 and Figure 5 in a quantitative and qualitative manner by using the results calculated by Equation 15. From the Figure 5, it can be seen that the plain image has strong correlation close to 1 while the cipher image does not close to 0. Table 2 is a further verification of the conclusions obtained in Figure 5 by numerical methods. This just proves that our encryption algorithm can not only change image correlation, but also it can resist statistical attacks [25].



(a) (d)

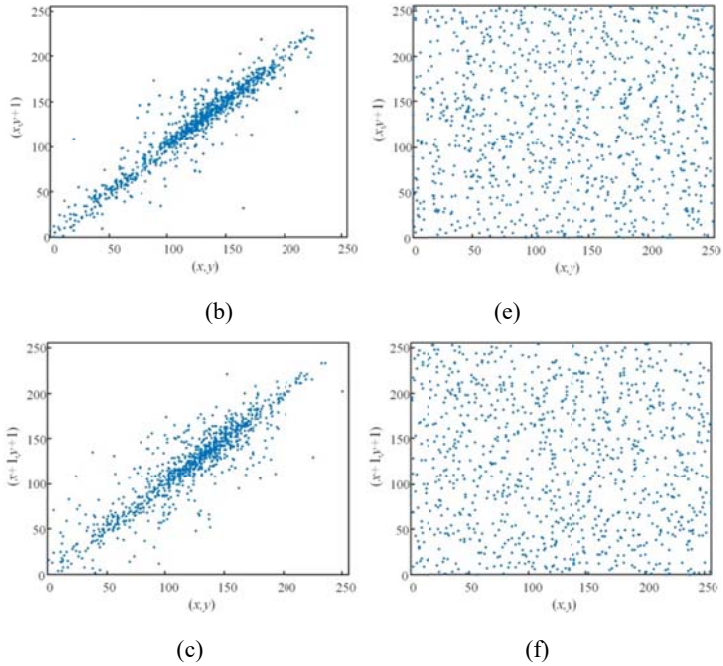


Fig.5 Test results of correlation distribution: (a) (b) (c) correlation of "Living room" in H, V, D direction; (d) (e) (f) correlation of cipher image in H, V, D direction.

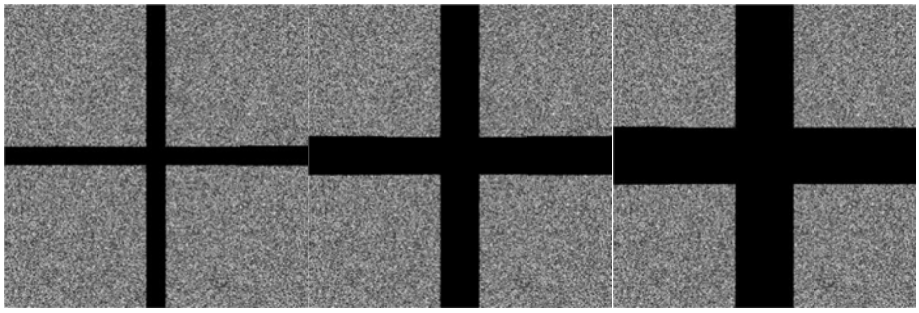
Table 2 Correlation of different images in H, V, D directions

Image	Direction	Plain image	Cipher image
Lena	H	0.9849	0.0020
	V	0.9718	-00017
	D	0.9590	-0.0031
Lake	H	0.9764	-0.0001
	V	0.9767	0.0009
	D	0.9622	0.0001
Living room	H	0.9524	-0.0008
	V	0.9459	0.0011
	D	0.9116	-0.0009
Peppers	H	0.9804	0.0003
	V	0.9802	0.0003
	D	0.9705	-0.0008
Man	H	0.9694	0.0021
	V	0.9602	0.0029
	D	0.9421	0.0026
	H	0.9702	-0.0005

Jet	V	0.9730	-0.0012
	D	0.9491	0.0005
	H	-	0.0010
Average	V	-	0.0014
(absolute value)	D	-	0.0013

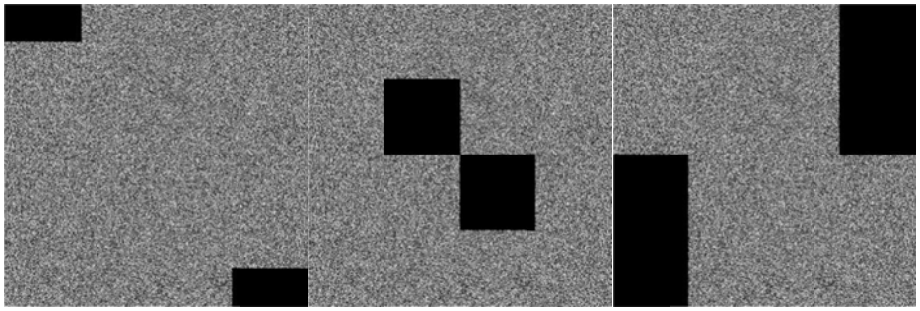
5.4 Cropping attack analysis

The cipher image could be cropped during the transmission process, which leads to the loss of some essential date in the decrypted image. Therefore, whether the key information of the image can be smoothly restored after a part of the information is lost becomes an important indicator for evaluating the encryption system. That is to say, even if some information of cipher image is lost, the cipher image can also be decrypted successfully, so the encryption scheme performs well. Figure 6 shows the experiment results after the information lost: the plain image "Lena" as test image is cropped according to different size and different direction. Since the image can be restored with a higher quality when a small amount of information is lost, the main outline information of the image can still be seen when a large amount of information is lost, which indicates that the loss of information will not become the weak point of this encryption scheme.



(a) data loss of 12.1% (b) data loss of 17.9% (c) data loss of 23.4%

(d) decrypted image, 12.1% (e) decrypted image, 17.9% (f) decrypted image, 23.4%



(g) data loss of 6.25% (h) data loss of 12.5% (i) data loss of 25%

(j) decrypted image, 6.25% (k) decrypted image, 12.5% (l) decrypted image, 25%

Fig.6 Cropping attack results

6 Conclusion

The algorithm in this paper uses swapping operation and block operation to encrypt the gray image. Firstly, the pseudo-random sequences are constructed by simple Lorenz system, and then the image is scrambled by exchanging rows and columns and scrambling blocks, so as to realize pixel scrambling and improve the encryption effect. Secondly, image encryption can be achieved by replacing pixel values. Theoretical analysis and simulation results indicate that encryption system used in this paper can pass most of the performance tests smoothly like key space, statistical and entropy tests, and has good robustness against cropping. Therefore, this algorithm provides another new encryption strategy for the secure transmission of image information on the Internet, and has certain application value.

References

[1] M. Roy, S. Chakraborty, K. Mali, D. Roy, and S. Chatterjee, "A robust image encryption framework based on DNA computing and chaotic environment," Microsystem Technologies, pp. 1-11, (2021)

[2] Y. Xian and X. Wang, "Fractal sorting matrix and its application on chaotic image encryption," Information Sciences, vol. 547, pp. 1154-1169, (2021)

[3] Y. Yang, L. Wang, S. Duan, and L. Luo, "Dynamical analysis and image encryption application of a novel memristive hyperchaotic system," Optics & Laser Technology, vol. 133, pp. 106553-106567, (2021)
[4] X. Yan, X. Wang, and Y. Xian, "Chaotic image encryption algorithm based on arithmetic sequence scrambling model and DNA encoding operation," Multimedia Tools and Applications, no. 1, pp. 1-35, (2021)
[5] Z. Hua, Y. Zhou, C. M. Pun, and C. L. P. Chen, "2D Sine Logistic modulation map for image encryption," Information Sciences, vol. 297, pp. 80-94, (2015)
[6] Liu, Hongjun, Kadir, and Abdurahman, "Asymmetric color image encryption scheme using 2D discrete-time map," Signal Processing the Official Publication of the European Association for Signal Processing, pp. 104-112, (2015)
[7] C. Li, K. Qian, S. He, H. Li, and W. Feng, "Dynamics and Optimization Control of a Robust Chaotic Map," IEEE Access, vol. PP, no. 99, pp. 1-1, (2019)
[8] Z. H. A, Z. Z. A, S. Y. B, Z. Z. C, and H. H. A, "Cross-plane colour image encryption using a two-dimensional logistic tent modular map - ScienceDirect," Information Sciences, vol. 546, pp. 1063-1083, (2021)
[9] H. Li, Y. Wang, and Z. Zuo, "Chaos-based image encryption algorithm with orbit perturbation and dynamic state variable selection mechanisms," Optics & Lasers in Engineering, vol. 115, no. APR., pp. 197-207, (2019)
[10] M. Kaur and V. Kumar, "A Comprehensive Review on Image Encryption Techniques," Archives of Computational Methods in Engineering, pp. 1-29, (2018)
[11] R. Lan, J. He, S. Wang, T. Gu, and X. Luo, "Integrated Chaotic Systems for Image Encryption," Signal Processing, vol. 147, no. JUN., pp. 133-145, (2018)
[12] R. Lan, Y. Zhou, Z. Liu, and X. Luo, "Prior Knowledge-Based Probabilistic Collaborative Representation for Visual Recognition," IEEE Transactions on Cybernetics, pp. 1-11, (2018)
[13] S. Wang, C. Wang, and C. Xu, "An image encryption algorithm based on a hidden attractor chaos system and the Knuth-Durstenfeld algorithm," Optics and Lasers in Engineering, vol. 128,pp. 105995-106008, (2020)
[14] C. L. Li, Z. Y. Li, W. Feng, Y. N. Tong, and D. Q. Wei, "Dynamical behavior and image encryption application of a memristor-based circuit system," AEU - International Journal of Electronics and Communications, vol. 110, pp. 152861-152881, (2019)
[15] Y. Niu and X. Zhang, "A Novel Plaintext-Related Image Encryption Scheme Based on Chaotic System and Pixel Permutation," IEEE Access, vol. PP, no. 99, pp. 22082-22093, (2020)
[16] C. Zhu, Z. Gan, Y. Lu, and X. Chai, "An image encryption algorithm based on 3-D DNA level permutation and substitution scheme," Multimedia Tools and Applications, vol. 79, no. 4, (2020)
[17] M. A. B. Farah, A. Farah, and T. Farah, "An image encryption scheme based on a new hybrid chaotic map and optimized substitution box," Nonlinear Dynamics, vol. 99, no. 1, pp. 1-24, (2020)
[18] J. Sun, C. Li, T. Lu, A. Akgul, and F. Min, "A memristive chaotic system with hypermultistability and its application in image encryption," IEEE Access, pp. 1-10, (2020)
[19] Z. Hua, Y. Zhou, and H. Huang, "Cosine-transform-based chaotic system for image encryption," Information Sciences, vol. 480, pp. 403-419, (2019)
[20] C. Xu, J. Sun, and C. Wang, "An Image Encryption Algorithm Based on Random Walk and Hyperchaotic Systems," International Journal of Bifurcation and Chaos, vol. 30, no. 04, pp. 2050060-2050075, (2020)

[21] Chai et al., "A novel image encryption algorithm based on the chaotic system and DNA computing," International Journal of Modern Physics C Physics & Computers, vol. 28, no. 5, pp. 1750069-1750092, (2017)
[22] X. Chai, Z. Gan, K. Yang, Y. Chen, and X. Liu, "An image encryption algorithm based on the memristive hyperchaotic system, cellular automata and DNA sequence operations," Signal Processing Image Communication, vol. 52, pp. 6-19, (2017)
[23] X. Jin, S. Yin, N. Liu, X. Li, G. Zhao, and S. Ge, "Color image encryption in non-RGB color spaces," Multimedia Tools and Applications,vol. 77, pp. 15851-15873, (2017)
[24] S. Som, S. Dutta, R. Singha, A. Kotal, and S. Palit, "Confusion and diffusion of color images with multiple chaotic maps and chaos-based pseudorandom binary number generator," Nonlinear Dynamics, vol. 80, no. 1-2, pp. 615-627, (2015)
[25] Zheng et al., "Encryption method based on a new secret key algorithm for color images," Aeu Archiv Fur Elektronik Und Ubertragungstechnik Electronic & Communication, vol. 70, no. 1, pp. 1-7, (2016)

A gray image encryption algorithm based on 3D chaotic map and DNA operations

Yuwen Sha[1] , Yinghong Cao[2] , Huizhen Yan[3]

{dlpusyw@sina.com[1], caoyinghong@dlpu.edu.cn[2], yanhz@dlpu.edu.cn[3]}

School of Information Science and Engineering, Dalian polytechnic University,

Dalian, 116034, China

Abstract. In this paper, a new gray image encryption system based on 3D chaotic map and deoxyribonucleic acid (DNA) sequence operations is presented. SHA is designed to give the initial value of the encryption system. The whole algorithm is designed based on permutation and diffusion framework. First, modify the chaotic sequence generated by the chaotic system, and the Arnold function is permuted the plane image. Secondly, DNA level diffusion is introduced, we start by performing a DNA XOR on the DNA coding image. Lastly, decoding is performed on the diffused DNA matrix. The results of experiment and analyses indicate that our encryption algorithm can pass the attack tests and has a certain application prospect.

Keywords: Image encryption, 3D chaotic map, DNA XOR.

1 Introduction

With the advent of the electronic information age, the ways of information exchange have been constantly enriched. In the process of communication, a large amount of information comes along, and the security of transmission and storage of the information on the Internet is threatened. Compared with the encryption of digital images, text images already have classical encryption algorithms, such as RSA, DES, IDEA [1]. Due to the digital image information not only has the characteristics of huge amount of data, but also has strong correlation and redundancy. This makes the encryption technology based on text image difficult to meet the needs of image encryption. Therefore, how to ensure that these private images are not attacked and stolen in the transmission process has become a research hotspot [2, 3]. This makes more and more cryptographers keen on designing a secure and effective image encryption method. Recently, after the study of scholars in the image encryption field of many novel technologies, including optical transform, chaotic systems, DNA computing, compressive sensing Fourier transform, cellular automata, wavelet transform.

Encryption and decryption of digital images require a large number of passwords that are the same size as the plaintext. How to produce pseudo random numbers with good statistical characteristics has become an urgent demand for digital image encryption. Chaotics complex structure, difficult to analyze and predict, can produce a large number of passwords, so it is used in the field of digital data encryption. In 1978, R.Matthews put forward a generalized

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307857

Logistic mapping based on the study of Logistic, and applied this permutation [4]. In 1998, J. Friedrich applied the state values of chaotic systems directly to the pixel method of permutation images. He found that permutation, which simply changed the position of pixels in an image, was not resistant to statistical analysis. This makes what appears to be a successful encryption operation still insecure. Permutation can flatten the encrypted histogram and change the statistical characteristics of the original histogram by changing the value of pixels. However, the visual effect after encryption is not good, so he proposed the classic permutation and diffusion frame structure. This is also the structure adopted by the vast majority of encryption schemes [5-12]. But encryption algorithms of security level are too low. Pareeket al. [5] used pseudo-random sequences produced chaotic logistic map for an image encryption algorithm. However, due to the one-dimensional mapping, the small key space once leads to poor security performance. Later, some high-dimensional chaotic maps were proposed for image encryption. Wang al. [6] use the mixed chaotic sequence to design an encryption algorithm. The pseudo-random sequences are generated by mixing Logistic map, Henon map and Lorenz system. Although the mixing of multiple systems can increase the size of the key space, the security of the encryption strategy is insufficient. Wei al. [13] introduced hyper-chaotic system for RGB image encryption algorithm. Although it can generate more complex chaotic sequences than low-dimensional chaotic systems, but the encryption and decryption speeds are not ideal because the system is too complex. One-time password is currently the most secure encryption algorithm [14-17], but the one-time password has become a new problem, so DNA encryption technology has been incorporated into the encryption system.

In 1994, Adleman published his famous DNA computing research article [18]. With the deepening of DNA computing research, many advantages of DNA computing have been gradually discovered, such as large amounts of parallelism, low power consumption, and large storage. Therefore, a new area of DNA image encryption has emerged. With the application of DNA addition, subtraction, XOR, and complementation in the field of image encryption, researchers found that an encryption system based on chaotic systems and DNA computing is easier to implement while ensuring image security [19-26]. For example, Chai al. [25] presented encryption algorithm in which SHA 256 hash function, a new there-dimensional chaotic system and DNA XOR operation are to enforce system security. NIU al. [26] presented Josephus traversing and pixel permutation for image encryption scheme. However some algorithms also have some security drawbacks [27-29], for example, Zhang al. [28] introduced DNA addition and two Logistic chaotic maps to encrypt image encryption. Liu al. [27] presented combining DNA complementary rule and chaotic maps to encrypt mage. But these two systems have a common disadvantage of small key space. Zhang al. [29] used encoding and chaos map to improve an RGB image encryption algorithm. After Liu al. [30] analyzed its security, it was found that this encryption system has two security risks. First, through experimental analysis of known plane images and ciphertext images, the key is not difficult to obtain process, the encryption system is less sensitive to the plane. In order to solve the shortcomings of chaotic encryption and DNA encryption, such as small key space, strong correlation, weak anti-attack ability, and poor sensitivity to encrypted images. In this paper, the gray-scale image design method improves the above a new chaotic system.

The following content will be divided into 5 sections and introduced in turn. Firstly, section 2 gives Preliminary works. Section 3 gives our encryption scheme by the architecture of permutation and diffusion. In Section 4, input experimental parameters to test performance. In Section 5, we give Security analyses. Finally, Section 6 puts conclusion at last.

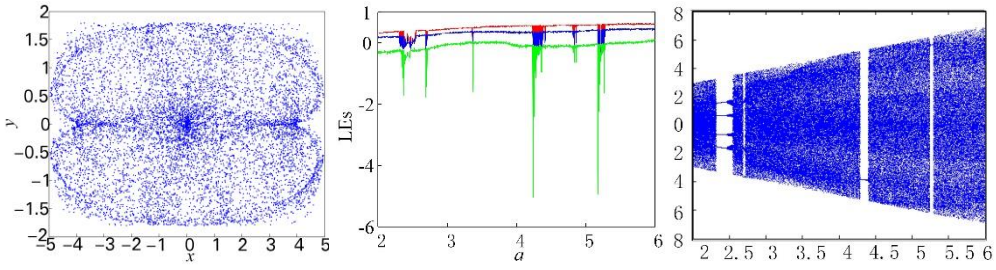
2 Preliminary Works

2.1 The model of system

The three-dimensional discrete chaotic map will be used in the encryption system of this paper and is obtained in Ref [31]. Its mathematical formula is given below:

$$\begin{cases} x_{n+1} = \sin x_n \sin y_n - a \sin z_n \\ y_{n+1} = b \sin x_n \cos y_n \\ z_{n+1} = c y_n \end{cases}, \quad (1)$$

where a, b, c are control parameters, when $(a, b, c) = (4, 4, 2)$ and the initial value $(x_0, y_0, z_0) = (0.5,0.2,0.1)$. The phase diagram is given in Fig.1 (a). Fixed the system parameter $b \in [1,5]$, the LEs (Lyapunov exponent spectrum) is shown in Fig.1 (b). The BDs (bifurcation diagram) is plotted in the Fig.1 (c).



(a) (b) (c)

Fig.1 Phase diagram, LES and BDs of 3D discrete chaotic map: (a) Phase diagram in x-y plane, (b) LES with b=4, c=2, $a \in [2,6]$, (c) BD with b=4, c=2 , $a \in [2,6]$.

2.2 DNA information

2.2.1 Convert image to DNA code

In DNA computing, the bases A, T, C, and G are used to represent information. Therefore, the image pixels are converted to 8-bit binary and use 00, 01, 10, 11 to represent A, T, C, and G respectively, corresponding to a total of 24 encoding rules, but Binary encoding and DNA encoding to produce structural and characteristics of the link. So, you can get eight rules for DNA coding from Table.1. This will also be used as the decoding rule in the decryption process.

In this paper, pixel is the basic unit of image. The gray value of pixel point is expressed as 8-bit binary sequence in computer. According to the rule of binary pair and DNA coding, it can be expressed as 4 nucleic acid bases. For example, A, T, C, and G are represented as 00, 11, 10, and 01 respectively. For pixel points is 121 using encoding rule 1 in Table.1, the binary sequence is represented as [01111001], and the corresponding DNA sequence is represented as [C T G C]. Similarly, if the DNA sequence given is ATCG, follow coding rule 1 in Table.1. We can get a decoded binary sequence of 00110110, the decimal number is “54”. This is how the DNA sequence is decoded.

Table.1 DNA encoding rules

Rule	1	2	3	4	5	6	7	8
00	A	A	T	T	G	G	C	C
01	C	G	C	G	T	A	T	A
10	G	C	G	C	A	T	A	T
11	T	T	A	A	C	C	G	G

2.2.2 DNA operations

When binary pixels are encoded as DNA, the corresponding binary operation rules are inherited by DNA operations as shown in Table.2 and Table.3.

Table.2 XOR operation

XOR	A	G	C	T
A	A	G	C	T
G	G	A	T	C
C	C	T	A	G
T	T	C	G	A

Table.3 Addition and Subtraction operation

+	A	C	G	T	-	A	C	G	T
A	A	C	G	T	A	A	T	G	C
C	C	G	T	A	C	C	A	T	G
G	G	T	A	C	G	G	C	A	T
T	T	A	C	G	T	T	G	C	A

2.3 Optimized Arnold transformation

The following is the optimization of Arnold matrix transformation method: assume the plaintext image is P, expand P into A one-dimensional row vector, and denote it as A. Arnold transformation is performed on the position $(1, j)$ of coordinates of any point of vector A, and the new coordinate position (p, q) is obtained through the following formula.

$$\begin{cases} p = 1 + aj \\ q = b + (ab+1)j \end{cases}. \tag{2}$$

Considering only the above equation, the exchange between pixel points $(1, j)$ and $(1, q)$ can be realized through pseudo-random variables a and b. At the same time, ab+1 is regarded as a new random number denoted by a, then Equation (2) becomes as shown in Equation (3).

$$q = b + aj\ . \tag{3}$$

3 Encryption Scheme

3.1 The Secret Sequence Generator

SHA 256 is a hash function. For messages of any length, SHA 256 will generate a 256-bit hash value, called a message digest. Therefore, using the SHA 256 function to encrypt the image can generate a 256-bit hash value, which will be used to calculate the initial value of the chaotic system.

Firstly, the plane image is entered as a parameter to SHA 256 function, then the key K with 256 bits is generated, and K is divided into 32 groups, each consists of 8-bital numbers. K can be derived as follows:

$$K = k_1, k_2, \ldots, k_{32}, \quad \text{subject to: } k_i = \left\{ k_{i,0}, k_{i,1} \cdots, k_{i,7} \right\}, \tag{4}$$

where $1 \leq i \leq 32$ and $0 \leq j \leq 7$.

The new stream key can be obtained through the following formula:

$$\begin{cases} k_a = \mathrm{mod}(10^{-15}\, mean(K(1:8) \oplus K(9:16), 2^6) \\ k_b = \mathrm{mod}(10^{-15}\, mean(K(17:24) \oplus K(25:32), 2^6) \\ k_c = \mathrm{mod}(10^{-15}\, sum(K(1:8) \oplus K(9:16) \oplus K(17:24), 2^6) \end{cases}, \tag{5}$$

where k_a, k_b, k_c and k_d are the disturbance parameters of the chaotic system, mean (□) is to find the mean operator, mod denotes the modular operator and $K_{2i} \oplus K_{2i+1}$ is the XOR operation.

In this encryption algorithm, the initial value is given as the key, and the initial value of the new chaotic system can be calculated from the following formula:

$$\begin{cases} x_1 = x_0 + k_a \\ y_1 = y_0 + k_b \\ z_1 = z_0 + k_c \end{cases}. \tag{6}$$

3.2 Image encryption flows

First, the encryption flow chart is given in Fig.2, specific details are described in the following two chapters.

3.2.1 Permutation steps

Arnold can generate a lot of pseudo-random sequences, so it can be used in scrambling with the help of sequences generated by 3D discrete chaotic map. The permutation process can be represented in follows.

***Step* 1.** Input grayscale plain image P of size $M \times N$ where M and N are lengths and widths.

***Step* 2.** Use the initial values x_1, y_1, z_1 and u_1 produced by Sec. 3.1, then initial values are taken into 3D discrete chaotic map and iterate the chaotic map for $r+l$ ($r=M \times N$, $l \geqq 500$) times. discarding former l times values of iteration result, key sequence X, Y and Z can be generated by Eq. (1), the four integer sequences $X1$, $Y1$ and $Z1$ are obtained from the following formulas:

$$\begin{cases} X = \{x_1, x_2, \cdots, x_r\} \\ Y = \{y_1, y_2, \cdots, y_r\} \\ Z = \{z_1, z_2, \cdots, z_r\} \end{cases}, \tag{7}$$

$$\begin{cases} X1 = \mathrm{mod}(floor((X+100)\times 10^{10}), 10\times \max(M,N)) + 1 \\ Y1 = \mathrm{mod}(floor((Y+100)\times 10^{10}), 10\times \max(M,N)) + 1 \\ Z1 = \mathrm{mod}(floor((Z+100)\times 10^{10}), 10\times \max(M,N)) + 1 \end{cases}, \tag{8}$$

where floor function is a downward rounding of the results, max (M, N) is the maximum number of M, N.

***Step* 3.** Divide the key sequence $X1$, $Y1$ and $Z1$ into the three groups that can be expressed as: $G(1)=[X1, Y1]$, $G(2)=[X1, Z1]$, $G(3)=[Y1, Z1]$. The first value of the plane pixel will participate in the selection of the chaotic sequence. when $((P(1)\times 1000) \bmod 6)+1=x$, we get key sequence $G(x)$ that is used to permute gray image.

***Step* 4.** In this step, permutation method used is Arnold. Firstly, the image P is expanded into a row vector, assume chaotic sequence array $G(1)$ is obtained in Step 3 then get the random variable $a = X1$, $b = Y1$ as setting in Eq. (3). After the permutation operation can be completed, the image P' is obtained.

3.2.2 Diffusion steps

In the diffusion process, we use DNA XOR operation, which may need to be help of a key matrix produced by 3D discrete chaotic system, to diffuse the image P'. The detailed diffusion steps are given in follows.

***Step* 1.** Transform P' into a binary matrix D ($M\times N\times 8$). Then, DNA matrix W ($M\times N\times 4$) can be obtained after performing matrix D with DNA encoding l_1 rule ($1\le l_1\le 8$) operation.

***Step* 2.** Further manipulation of the chaotic sequences X, Y and Z obtained from Eq. (7), then $X1$, $Y1$ and $Z1$ can be calculated in follow:

$$\begin{cases} X1 = \dfrac{sum(abs(X) + floor(Y))}{r} \\ YI = \dfrac{sum(abs(Y) + floor(Z))}{r} \\ Z1 = \dfrac{sum(abs(X) + floor(Z))}{r} \end{cases}. \tag{9}$$

***Step* 3.** The final key matrix can be obtained by the following formula:

$$K2 = \mathrm{mod}(round(\frac{3\times 10^8 \times T}{F}), 256), \tag{10}$$

$$\begin{cases} T(i,j) = (X1+i)^2 + (Y1+i)^2 + Z1^2 \\ F(i,j) = (X1+i)^2 + j^3 + Y1\times i\times j \end{cases}, \tag{11}$$

where T and F represent modified chaotic sequence, which used to calculate the key matrix $K2$.

***Step* 4.** Convert $K2$ to a DNA matrix $K3$ with encoding l_1 rule ($1\le l_1\le 8$), the encrypted S matrix is shown in follow:

$$S(i) = D(i) \oplus K3(i) \oplus D(i-1), \tag{12}$$

where $\oplus$is DNA XOR operator, $K3(i)$ is the ith (i = 1, 2, …., MN) element of matrix $K1$, $S(i)$ is the output DNA data. When i=1, $S(0) = D$(end).

***Step* 5.** Input the decoding rule l_2, then decode DNA matrix S after diffusion and the ciphertext image S' is generated.

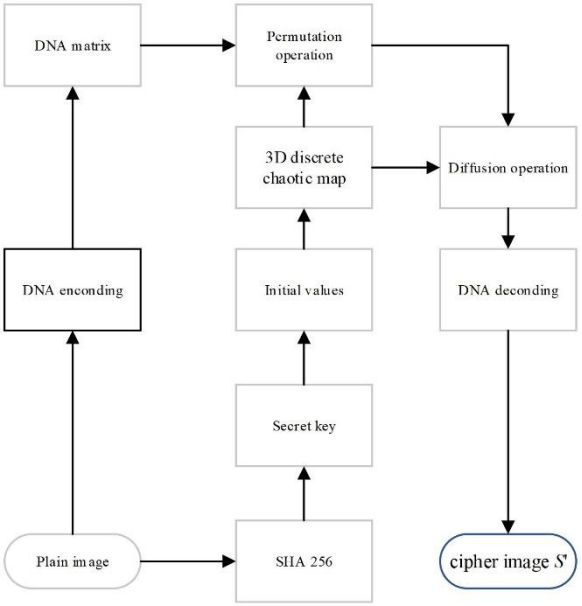
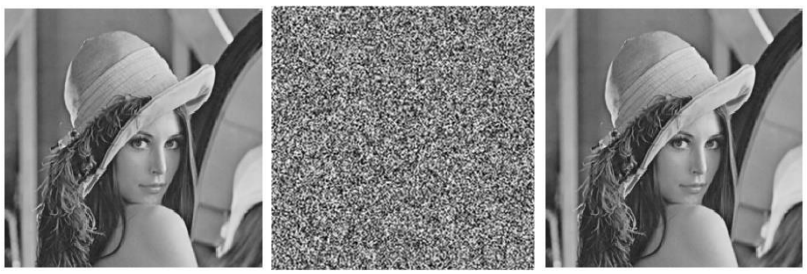


Fig.2 Flow chart of encryption scheme.

4 Simulation tests

Firstly, the initial values and parameters of 3D discrete chaotic system are set as: (x_0, y_0, z_0) = (0.5, 0.2, 0.1), (a, b, c) = (4, 4, 2). Secondly, DNA encoding rules l_1=1 and l_2=3 and iteration numbers l=500; Lena (256 256) as a grayscale image to be encrypted. Lastly, the experimental results are presented in Fig.3 (a)–(f).



(a) (b) (c)

Fig.3 Encryption and decryption test: (a) Original image Lena, (b) encrypted image, (c) decrypted image.

It can be seen from Fig.3 (a)–(b) that, the plane image is encrypted by the encryption algorithm in this paper, it becomes a picture similar to noise, the plaintext information has been successfully hidden. Then after observe the Lena image and decryption as show in Fig.3 (a)–(c), the decrypted image is found to be the same as the original image, it shows that the encryption algorithm with this paper is feasible.

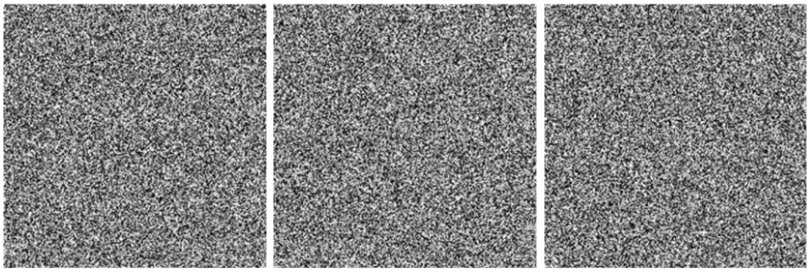
5 Security analyses

5.1. Analysis of key space

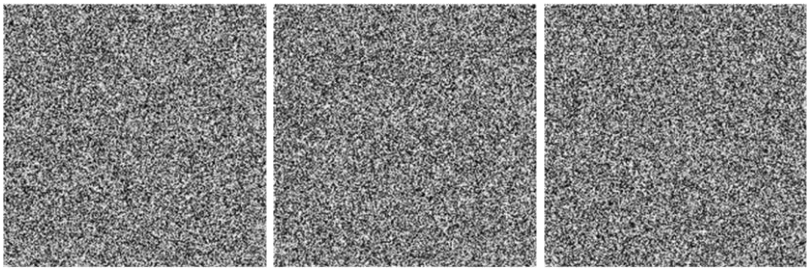
Key space is a collection of all legal keys. The cryptosystem's key space should be large enough so that can effectively combat exhaustive attacks, especially encryption and decryption of very fast cryptographic systems, the length of the password should be at least 100 bit [32]. In the proposed algorithm, key space collection consist of:
(1) The 256-bit external secret key which generated by the plan image are used to give initial values encryptin system.
(2) The chaotic system parameter (x_0, y_0, z_0, a, b, c).
(3) Iterating parameter l in order to obtain chaotic sequences.
(4) The first pixel of the original image ($P(1)$).
(5) Encoding rule l_1 and decoding rule l_2.
Since the calculation of the key space is related to the 10^{-15} of the computer, it is assumed here that accuracy is 10^{-15}, only the key of 3D discrete chaotic system x_0, y_0, z_0, a, b, c and the external key generated by SHA 256 hash function space will be $10^{208} \approx 2^{691}$. When the encryption system has such a large encryption space, only brute force attacks are invalid for the system.

5.2. Analysis of key sensitivity

The analysis of key sensitivity is to make a key change slightly while the other keys remain unchanged, and then decrypt the encrypted image with it. In experiments, the test of key sensitivity scheme uses (x_0, y_0, z_0, a, b, c), which belongs 3D discrete chaotic map, as secret key. Fig.4 shows the decrypted image of the encrypted image after slightly changing the value of parameter, x_0+10^{-16}, y_0+10^{-16}, z_0+10^{-16}, $a+10^{-15}$, $b+10^{-15}$, $c+10^{-15}$ and keeping remaining parameters same. Key sensitivity experiment shows in Fig.5 that the small change of the key will seriously affect the decryption effect of the image, nothing information about the original image to be found due to the image becomes chaotic. Therefore, the algorithm has good security.



(a) (b) (c)

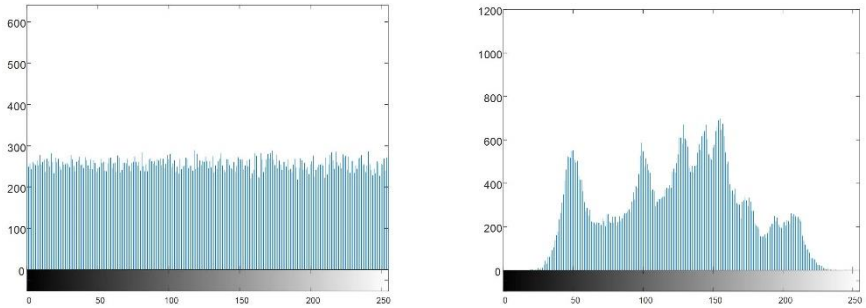


(d) (e) (f)

Fig.4 Key sensitivity tests: (a) Changed key with x_0+10^{-16}, (b) Changed key with y_0+10^{-16}, (c) Changed key with z_0+10^{-16}, (d) Changed key with $a+10^{-15}$, (e) Changed key with $b+10^{-15}$, (f) $c+10^{-15}$.

5.3 Histogram analysis

The histogram of the image is often used to analyze image information, which can be considered as an approximation of the gray density function, Although the histogram cannot directly reflect the image content, it can analyze the performance of an encryption scheme. For example, the histogram of an ideal encrypted image should look like a flat image, if it is not enough, image information can obtained by information statistical analysis. The gray histograms of Lena image before and after encryption are showed in Fig. 5.



(a) (b)

Fig.5 Histogram experimental analysis: (a) Histogram of the original image Lena, (b) histogram of encrypted image.

It can be seen from Fig.5 (b) that the images have almost the same gray value and the statistical Therefore, the encryption algorithm has better performance against statistical analysis.

5.4 Correlation analysis

The correlation coefficient r_{xy} is calculated as follows:

$$r_{xy} = \frac{\operatorname{cov}(x, y)}{\sqrt{D(x)D(y)}}, \tag{13}$$

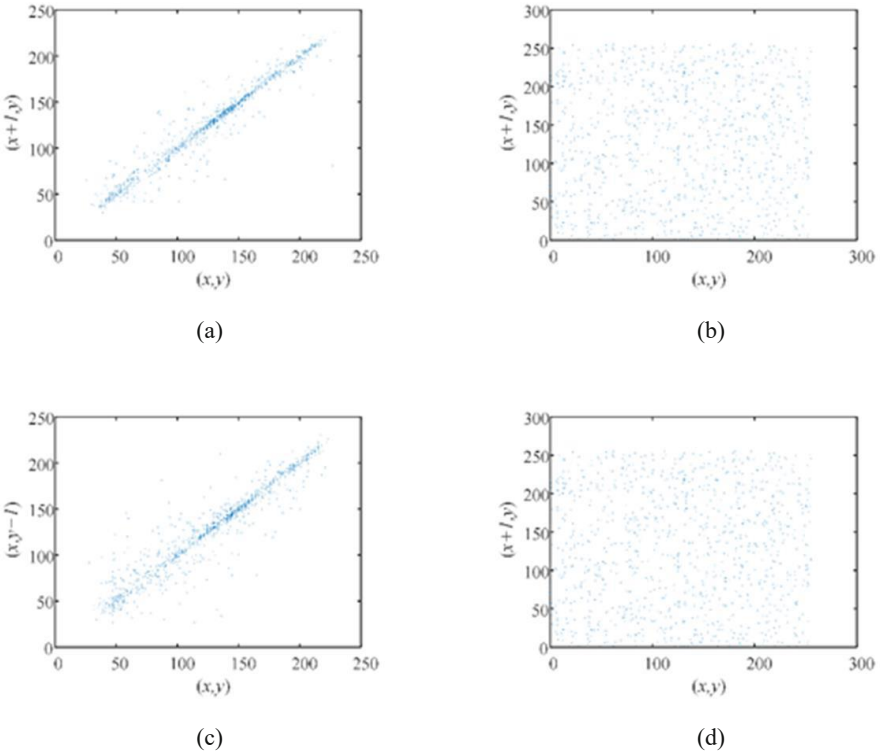
$$\operatorname{cov}(x,y) = E\left\{\left[x - E(x)\right]\left[y - E(y)\right]\right\}, \tag{14}$$

$$E(x) = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{15}$$

$$D(x) = \frac{1}{N}\sum_{i=1}^{N}\left[x_i - E(x)\right]^2, \tag{16}$$

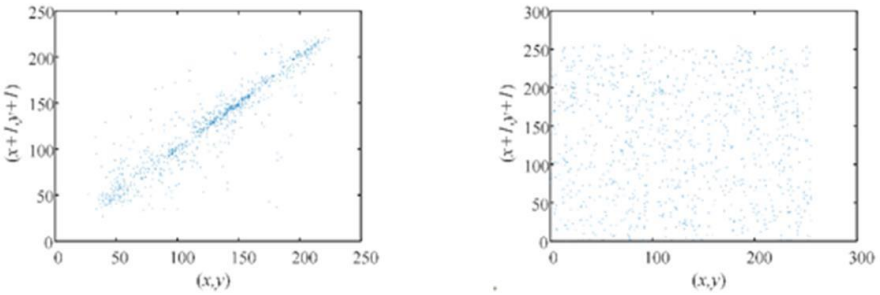
where x and y are the gray value of adjacent pixels, N is the total number of pixels, cov(x, y) is the covariance, $E(x)$ is the average value of pixels and $D(x)$ is the variance. The greater the absolute value of r_{xy}, the stronger the correlation.

In order to visually observe the correlation changes of images before and after encryption, Fig.6 give the 2000 pairs correlation adjacent pixels of the plain image Lena (256 256) and its cipher image in three directions, and the results are compared with the image encryption literatures in recent years, as shown in Table.4.



(a) (b)

(c) (d)



(e) (f)

Fig.6 Correlation experimental analysis (a) Horizontal correlation of the original image Lena (b) Horizontal correlation of the encrypted image Lena (c) Vertical correlation of the original image Lena (d) Vertical correlation of the encrypted image Lena (e) Diagonal correlation of the original image Lena (f) Vertical correlation of the encrypted image Lena

Table.4 Correlation value table

Direction	Lena image	Our algorithm	Ref [33]	Ref [34]	Ref [35]
Horizontal	0.9692	0.0018	-0.0065	0.0058	0.0198
Vertical	0.9423	0.0016	0.0672	0.0094	0.0141
Diagonal	0.9160	-0.0008	0.0047	0.0214	0.0025

It can be seen from Table.4 that the correlation coefficients of plaintext are all above 0.9, but the correlation coefficients of adjacent elements of ciphertext are all approximately 0. The correlation index of this algorithm is better than that of other algorithms.

5.5 Information entropy analysis

Information entropy is an important indicator reflecting the randomness of information. The more random the pixel value distribution, the higher the information entropy of the image. When the occurrence probability of each gray value is equal, the information entropy of the image is the Ideal value. the information entropy can be derived from the following formula:

$$H(m)=\sum_{i=0}^{L-1} p(m_i)\log\frac{1}{p(m_i)}, \tag{17}$$

where $P(m_i)$ represents the probability of the standard state m_i and L total number of state variables m_i. Suppose there are 28 state values for information m and the probability of their occurrence. According to Eq. (17), if the information entropy of ciphertext close to the ideal value of $H(m)=8$ which represents encryption algorithm can fight against entropy attacks. It can be concluded from Table.5 that the information leakage of ciphertext is very small and the algorithm is safe.

Table.5 Information entropy comparison with other algorithms.

Images	Our algorithm	Ref [36]	Ref [37]	Ref [38]
Camera	7.9972	7.9953	7.4101	7.9937
Couple	7.9972	7.9948	7.4740	7.9938

6.Conclusion

Through a series of experiment and safety analysis, this encryption system in this paper has the following two advantages. First, the hybrid system used for encryption is a new 3D chaotic system. It has a fast iteration speed, which can speed up the running speed of the system. At the same time, due to the large number of parameters, which solves the key space security problem to a certain extent. and by analyzing phase diagram, LEs and BDs illustration, this proves that this encryption system can be applied to chaotic image encryption. Secondly, the plaintext image is closely related to the chaotic sequence by using the SHA 256 hash function, which improved the sensitivity of the encryption system to flat images. in other words, once the plaintext changes slightly, the encryption system will be extremely sensitive to this change. Simulation results show that the algorithm has large key space, high key sensitivity, reduced correlation, enhanced pseudo-randomness, higher security, and stronger ability to resist various attacks.

References

[1] G. Ye, "A block image encryption algorithm based on wave transmission and chaotic systems," *Nonlinear Dynamics,* vol. 75, no. 3, pp. 417-427, 2014.

[2] S. Wang, C. Wang, and C. Xu, "An image encryption algorithm based on a hidden attractor chaos system and the Knuth-Durstenfeld algorithm," *Optics & Lasers in Engineering,* vol. 128, 2020.

[3] H. Dong, E. Bai, X. Q. Jiang, and Y. Wu, "Color Image Compression-Encryption Using Fractional-Order Hyperchaotic System and DNA Coding," *IEEE Access,* vol. 8, pp. 163524-163540, 2020.

[4] R. A. J. Matthews, "On the derivation of a Chaotic encryption algorithm," *Cryptologia,* 1984.

[5] N. K. Pareek, V. Patidar, and K. K. Sud, "Image encryption using chaotic logistic map," *Image and Vision Computing,* vol. 24, no. 9, pp. 926-934, 2006.

[6] W. Mei-Lin, L. Qing, and L. I. Ya, "An image encryption algorithm based on the mixed chaotic sequence," *Optoelectronics Letters,* vol. 6, no. 004, pp. 310-313, 2010.

[7] L. Huang, S. Cai, X. Xiong, and M. Xiao, "On symmetric color image encryption system with permutation-diffusion simultaneous operation," *Optics & Lasers in Engineering,* vol. 115, no. APR., pp. 7-20, 2018.

[8] Y. He, F. Wang, S. Wang, and B. Chen, "Diffusion Adaptation Framework for Compressive Sensing Reconstruction," 2017.

[9] E. Borowski, Y. Chen, and H. Mahmassani, "Social media effects on sustainable mobility opinion diffusion: Model framework and implications for behavior change," *Travel Behaviour & Society,* vol. 19, 2020.

[10] A. Kulsoom, D. Xiao, Aqeel-Ur-Rehman, and S. A. Abbas, "An efficient and noise resistive selective image encryption scheme for gray images based on chaotic maps and DNA complementary rules," *Multimedia Tools & Applications,* vol. 75, no. 1, pp. 1-23, 2016.

[11] Wang *et al.*, "Concerted Ion-Exchange Mechanism for Sodium Diffusion and Its Promotion in Na3V2(PO4)(3) Framework," *Journal of Physical Chemistry C Nanomaterials & Interfaces,* 2018.

[12] J. Chang, S. Karra, and K. B. Nakshatrala, "Large-Scale Optimization-Based Non-negative Computational Framework for Diffusion Equations: Parallel Implementation and Performance Studies," 2017.

[13] X. Wei, L. Guo, Q. Zhang, J. Zhang, and S. Lian, "A novel color image encryption algorithm based on DNA sequence operation and hyper-chaotic system," *Journal of Systems & Software,* vol. 85, no. 2, pp. 290-299, 2014.

[14] A. Argyris, E. Pikasis, and D. Syvridis, "Gb/s One-Time-Pad Data Encryption With Synchronized Chaos-Based True Random Bit Generators," *Journal of Lightwave Technology,* vol. 34, no. 22, pp. 5325-5331, 2016.

[15] S. M. H. Alwahbani and E. B. M. Bashier, *Speech scrambling based on chaotic maps and one time pad*. 2013.

[16] A. Shakiba, "A Randomized CPA-Secure Asymmetric-Key Chaotic Color Image Encryption Scheme based on the Chebyshev Mappings and One-Time Pad," *Journal of King Saud University Computer & Information Sciences,* 2019.

[17] W. X. a, W. S. a, Z. Y. b, and L. C. c, "A one-time pad color image cryptosystem based on SHA-3 and multiple chaotic systems - ScienceDirect," *Optics & Lasers in Engineering,* vol. 103, pp. 1-8, 2018.

[18] Adleman and L., "Molecular Computation Of Solutions To Combinatorial Problems," *Science,* vol. 266, no. 5187, pp. 1021-1024, 1994.

[19] X. L. a, L. W. b, Y. Y. b, and P. L. b. a. c, "An improvement color image encryption algorithm based on DNA operations and real and complex chaotic systems," *Optik,* vol. 127, no. 5, pp. 2558-2565, 2016.

[20] X. Wei, G. Ling, Z. Qiang, J. Zhang, and S. Lian, "A novel color image encryption algorithm based on DNA sequence operation and hyper-chaotic system," *Journal of Systems & Software,* vol. 85, no. 2, pp. 290-299, 2014.

[21] X. Y. Wang, H. L. Zhang, and X. M. Bao, "Color image encryption scheme using CML and DNA sequence operations," *Bio Systems,* pp. 18-26, 2016.

[22] X. Zhang and R. Ye, "A novel RGB image encryption algorithm based on DNA sequences and chaos," *Multimedia Tools & Applications,* pp. 1-25, 2020.

[23] H. R. Shakir, "A Color-Image Encryption Scheme Using a 2D Chaotic System and DNA Coding," *Advances in Multimedia,* vol. 2019, pp. 1-13, 2019.

[24] Xiaopeng *et al.*, "A novel color image encryption algorithm based on DNA sequence operation and hyper-chaotic system," *Journal of Systems & Software,* 2012.

[25] Chai *et al.*, "A novel image encryption algorithm based on the chaotic system and DNA computing," *International Journal of Modern Physics C Physics & Computers,* 2017.

[26] Y. Niu and X. Zhang, "A Novel Plaintext-Related Image Encryption Scheme Based on Chaotic System and Pixel Permutation," *IEEE Access,* vol. PP, no. 99, pp. 1-1, 2020.

[27] Q. Wang, Q. Zhang, X. Wei, X. Xue, and L. Guo, "Image Encryption Based on Chaotic Map and DNA Coding," *Journal of Computational and Theoretical Nanoscience,* vol. 7, no. 2, pp. 388-393, 2010.

[28] Q. Zhang, L. Guo, and X. Wei, "Image encryption using DNA addition combining with chaotic maps," *Mathematical & Computer Modelling,* vol. 52, no. 11-12, pp. 2028-2035, 2010.

[29] L. Liu, Z. Qiang, and X. Wei, *A RGB image encryption algorithm based on DNA encoding and chaos map*. Pergamon Press, Inc., 2012.

[30] Y. Zhang, Y. Li, W. Wen, Y. Wu, and J. X. Chen, "Deciphering an image cipher based on 3-cell chaotic map and biological operations," *Nonlinear Dynamics,* 2015.

[31] F. Yang, J. Mou, C. Luo, and Y. Cao, "An improved color image encryption scheme and cryptanalysis based on hyperchaotic sequence," *Physica Scripta,* 2019.

[32] S. M. Seyedzadeh, B. Norouzi, M. R. Mosavi, and S. Mirzakuchaki, "A novel color image encryption algorithm based on spatial permutation and quantum chaotic map," *Nonlinear Dynamics,* vol. 81, no. 1-2, pp. 511-529, 2015.

[33] X. Wang, Q. Wang, and Y. Zhang, "A fast image algorithm based on rows and columns switch," *Nonlinear Dynamics,* vol. 79, no. 2, pp. 1141-1149, 2015.

[34] N. Zhou, A. Zhang, J. Wu, D. Pei, and Y. Yang, "Novel hybrid image compression–encryption algorithm based on compressive sensing," *Optik - International Journal for Light and Electron Optics,* vol. 125, no. 18, pp. 5075-5080, 2014.

[35] Y. Zhang, B. Xu, and N. Zhou, "A novel image compression–encryption hybrid algorithm based on the analysis sparse representation," *Optics Communications,* 2017.

[36] Zhou *et al.*, "Image compression-encryption scheme based on hyper-chaotic system and 2D

compressive sensing," *Optics & Laser Technology,* 2016.

[37] X. Liu, Y. Cao, P. Lu, X. Lu, and Y. Li, "Optical image encryption technique based on compressed sensing and Arnold transformation," *Optik - International Journal for Light and Electron Optics,* vol. 124, no. 24, pp. 6590-6593, 2013.

[38] Ponuma, R., and Amutha, "Compressive sensing based image compression-encryption using Novel 1D-Chaotic map," *Multimedia Tools & Applications,* 2018.

An image encryption algorithm based on neural network

Xuejun Li, Jiawu Yu*, Huizhen Yan
{yujiawu_dlpu@sina.com}

School of Information Science and Engineering, Dalian Polytechnic University, Dalian 116034, China

Abstract: Chaotic encryption provides a practical method for the confidentiality protection of today's digital images. Aiming at the problem of the secure transmission of image information in the network, an image encryption algorithm based on laser chaotic system and neural network is designed in this work. The phase diagram, Lyapunov exponential spectrum and bifurcation diagram analysis confirmed the chaotic characteristics of the single-mode laser power system. On this basis, the plaintext image is first compressed by the BP neural network, and then the compressed image is chaotically generated and diffused using the chaotic sequence obtained by the single-mode laser chaotic system to obtain the ciphertext image, and finally the histogram, Adjacent pixel correlation, anti-differential attack, information entropy and other aspects to test the security of the proposed encryption scheme. Simulation results show that the image encryption algorithm proposed in this work can not only save the channel bandwidth used in ciphertext transmission, but also have good encryption effect and high security performance.

Keywords: Laser chaotic system, Neural network, Arnold transform, Image encryption

1 Introduction

With the popularization of the digital images have been widely applied in military, medicine, industrial engineering, public security, trademark protection [1], and other fields. Images are stored and transmitted through multiple platforms or channels [2]. Because many images contain private or sensitive information, image security issues have received more and more attention in recent years [3].

Traditional image encryption methods usually treat digital images as binary data sequences, and then use traditional data encryption techniques (such as AES [4]) to encrypt the data sequences. For these reasons, traditional encryption schemes are actually not suitable for image encryption. Chaos has the characteristics of initial value sensitivity, unpredictability and ergodicity. These characteristics make the pseudo-random sequence very appropriate to image

EAI GreeNets 2021, June 06-07, Dalian, People's Republic of China

DOI 10.4108/eai.6-6-2021.2307843

encryption [6]. At present, a variety of image encryption algorithms have been designed by using memristor system [7], chaos theory [8-10], quantum theory [11,12], compressed sensing [13-15], DNA coding [16,17], and other technologies. For example, an adaptive medical image encryption algorithm based on improved chaotic mapping was proposed by Chen et al. [18], which used logistics chaotic mapping to scramble plane images, and at the same time uses a hyperchaotic system to adaptively encrypt sub-blocks. In Literature [19], a new encryption and decryption algorithm was proposed, which used projection transformation to improve the diffusion characteristics of the algorithm. Chai et al. proposed a highly adaptive medical image encryption algorithm, which is mainly based on the replacement of pure images and the Latin square matrix and two-way adaptive diffusion [20]. A new image encryption scheme based on the hidden attractor chaotic system was proposed in the literature [21]. Literature [22] proposed a new image encryption algorithm using chaotic cross-mapping to replace and diffuse the matrix of each image color channel. Niu et al. proposed a new color image encryption algorithm based on the anti-control of fractional-order chaotic system [23]. Literature [24] proposed a chaotic-based parallel encryption scheme, which makes full use of modern computer processors to encrypt images. Literature [25] proposed a three-dimensional fractional discrete Hopfield neural network based on left Caputo discrete increments, and at the same time studied the dynamic characteristics and synchronization characteristics of the neural network, and applied it to image encryption. Zhang et al. designed an image encryption scheme based on an integer domain-like perceptron network. It takes the perceptron network as the core to realize the information storage and dissemination of ordinary images [26]. Liu et al. studied an optical image encryption algorithm based on the hyperchaotic system and public-key cryptography theory by combining double random phase encoding of the Fresnel domain [27].

Compared with traditional chaotic systems, the dynamic characteristics of laser chaotic systems are more complex and more sensitive to system parameters. Therefore, the application of laser chaotic system in secure communication has a good application prospect. Therefore, in this work, a single-mode laser chaotic system is used to generate the pseudo-random sequence of pixel scrambling and diffusion. At the same time, a new digital image encryption scheme is designed by combining the neural network method, Arnold scrambling method, and the diffusion method of adding mode operation.

The rest of the paper is arranged as follows: The single-mode laser chaotic system is discussed in section 2. Section 3 introduces the proposed encryption algorithm and decryption algorithm in detail. The encryption and decryption test results and security performance are

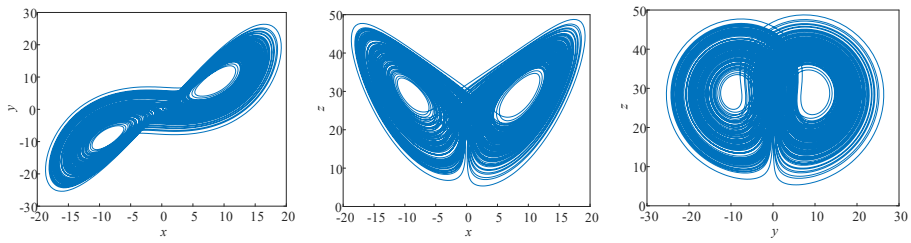
analyzed in section 4. The 5 section summarizes the work of this article and draws important conclusions.

2 Single-mode laser chaotic system

The equation of the single-mode laser Lorentz system can be described as follows:

$$\begin{cases} \dfrac{dx}{dt} = a(y-x) \\ \dfrac{dy}{dt} = (c-z)x - y \\ \dfrac{dz}{dt} = xy - bz \end{cases}, \tag{1}$$

where a, b and c are the system parameters. Let the system initial x_0=1, y_0=2, z_0=3, the system parameter a=10, b=8/3, c=30, and step length h=0.01. The calculated Lyapunov exponent of the system is (1, 0, -14.79). Among them, there is a positive Lyapunov exponent, and the sum of all values is negative. Therefore, the single-mode laser Lorenz system is a chaotic system. Fig.1 shows the system phase diagrams.

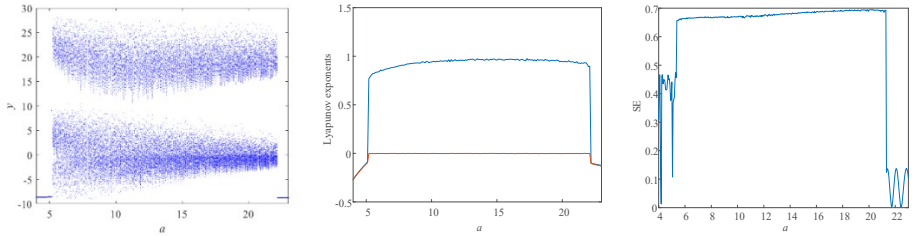


(a) x–y plane (b) x–z plane (c) y–z plane

Fig.1 Phase diagram of single-mode laser Lorenz system

2.1 Influence of different system parameters on dynamic performance

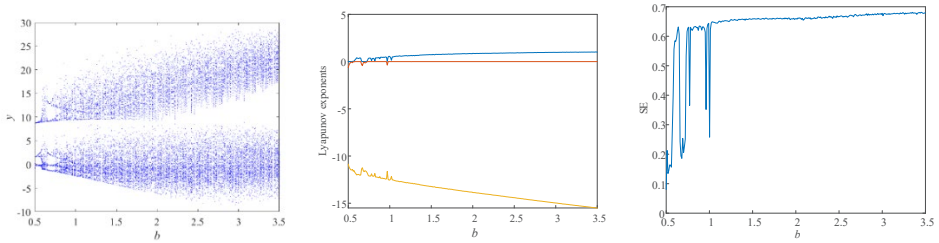
Let parameter b=8/3, c=30, step length h=0.01, initial value x_0=1, y_0=2, z_0=3. When $a \in [4, 23]$, the bifurcation graph, Lyapunov exponent and SE complexity of the system are shown in Fig.2. As can be seen from Fig.2(a, b), when $a \in (4, 5.2) \cup (22.18, 23)$, the system Lyapunov exponent is less than 0, so the system is in a stable state. When $a \in (5.2, 22.17)$, there is a positive Lyapunov exponent and the system is in a chaotic state. Fig.2(c) shows that the complexity of the system is large when it is in a chaotic state. When the system is in a chaotic state, the SE complexity of the system is very high. When the system is in equilibrium, the SE complexity of the system is very low.



(a) Bifurcation diagram (b) Lyapunov exponents (c) SE complexity

Fig.2 Dynamic characteristics of the system at parameter $a \in [4, 23]$

According to the parameters a=10,c=30, step size h=0.01, initial value x_0=1,y_0=2,z_0=3, and $b \in [0.5, 3.5]$, the dynamic characteristics of the system are shown in Fig.3. In Fig.3(a, b), when $b \in [0.5, 0.58) \cup (0.65, 0.73) \cup (0.96, 0.98)$, the maximum Lyapunov exponent is 0, and the system is in the form of periodic motion. The system has a period-doubling bifurcation at b=0.54 and 0.73. When $b \in (0.58, 0.65) \cup (0.73, 0.96) \cup (0.98, 3.5]$, there is a positive Lyapunov exponent, and the system is in a chaotic state. Fig.3(c) shows that the complexity is large in the range of chaotic states, which is in good agreement with the states of bifurcation diagram and Lyapunov exponential spectrum.



(a) Bifurcation diagram (b) Lyapunov exponents (c) SE complexity

Fig.3 Dynamic characteristics of the system at parameter $b \in [0.5, 3.5]$

3 Encryption and decryption algorithm design

3.1 Arnold transform

Arnold transform is also called “Cat transform”. It is a chaotic mapping method that performs repeated folding and stretching transformations in a limited area [28]. The digital image can be regarded as a two-dimensional matrix, and the pixel position in the image can be regarded as the position of the corresponding coordinate in the two-dimensional matrix. After Arnold transform, the pixel positions of the image will be rearranged to achieve the effect of scrambling and encrypting the image. Arnold transformation algorithm is realized by formula (2) (positive transformation):

$$\begin{bmatrix} \alpha_{n+1} \\ \beta_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & b \\ a & ab+1 \end{bmatrix} \begin{bmatrix} \alpha_n \\ \beta_n \end{bmatrix} \bmod(N), \tag{2}$$

where α_n and β_n represent the positions of the image pixels before the transformation, α_{n+1} and β_{n+1} are the positions of the pixels after the transformation, a and b mean the number of current transformations, N is the length or width of the image.

The inverse transformation is shown in formula (3):

$$\begin{bmatrix} \alpha_{n+1} \\ \beta_{n+1} \end{bmatrix} = \begin{bmatrix} ab+1 & -b \\ -a & 1 \end{bmatrix} \begin{bmatrix} \alpha_n \\ \beta_n \end{bmatrix} \bmod(N), \tag{3}$$

The two transformation matrices are reciprocal matrices, and the inverse matrix is still an integer matrix.

3.2 BP neural network

The BP neural network model has three main components, which are layers (input layer, hidden layer, and output layer), neurons, and weights between neurons. The input layer neuron receives the input information and transmits it to each neuron in the hidden layer. The hidden layer neurons are responsible for processing and transforming the received information, and the output layer outputs the processing results. Neurons in each layer are only fully connected to neurons in the adjacent layer, and neurons in the same layer are not connected. The calculation process of each neuron in the BP neural network method is showed in Fig.4. The calculation process of the BP neural network model can be expressed as

$$\hat{Y} = f_{output} \sum_{j=1}^{H} w_{kj} \left(f_{hidden} \left(\sum_{i=1}^{I} w_{ji} X_i + b_j \right) + b_k \right), \tag{4}$$

where I and H mean the number of neurons in the input layer and the hidden layer, X_i is the input information, b_k and b_j mean output layer deviation and hidden layer deviation respectively, f_{output} and f_{hidden} represent the transfer functions of the neurons in the hidden layer and the output neuron respectively, w_{ji} is the weight that connects the input layer and the hidden layer, w_{kj} is the weight between the hidden layer and the output layer.

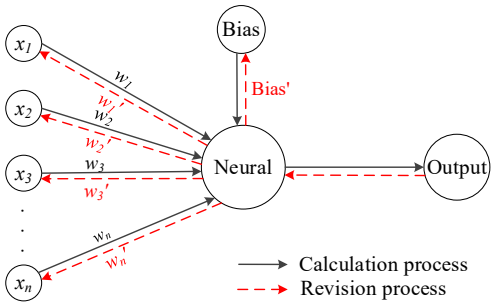


Fig.4 Principle of neuron calculation

The error between the output value and the actual value is measured by Equation (5). If the error exceeds the tolerance, the weight and deviation are corrected by the gradient descent method. The output value is retrained through the modified weight and deviation. The above process is repeated until the output is within the tolerance, which is expressed as

$$E=\frac{1}{N}\sum_{n=1}^{N}\left(Y'_n-Y_n\right)^2 \quad n=1,2,3,\cdots,N\ , \tag{5}$$

where Y'_n and Y_n represent the predicted output and actual output of the training vector respectively, and N means the number of training samples.

3.3 Encryption algorithm

Assuming the input image size is $W\times H$. Fig.5 shows the image encryption process proposed in this paper, and the specific encryption steps can be described as follows:

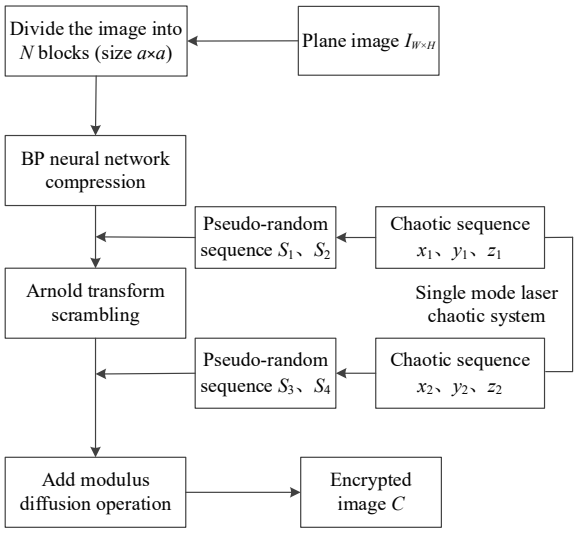


Fig.5 Encryption process

Step 1: Input image $I_{W\times H}$ and divide the image into N sub-image blocks of size $a \times a$.

Step 2: The mean distribution preprocessing is used to normalize the generated sub-image block matrix. The grayscale range of the image to be processed is $[x_{\min}, x_{\max}]$, and the transformation domain is $[y_{\min}, y_{\max}]$. Assuming that the pixel to be processed is x_{value}, then y_{value} can be obtained by the following formula:

$$y_{value} = \frac{(y_{\max} - y_{\min})(x_{value} - x_{\min})}{x_{\max} - x_{\min}} + x_{\min}. \tag{6}$$

Step 3: Use the newff function for training to get compressed data. The transfer function is as follows:

$$\log sig(n) = \frac{1}{1+e^{-n}}. \tag{7}$$

The characteristic of log$sig(n)$ function is that the data in the range of $(-\infty, +\infty)$ is mapped to the interval (0, 1), and n is the input of the number of neuron nodes.

$$\tan sig(n) = \frac{2}{1+e^{-2n}} - 1. \tag{8}$$

In the tan$sig(n)$ function, the output is limited to the interval (-1, 1).

Step 4: Setting the parameters and initial values of the chaotic system, and iterate the chaotic system (t+ $W\times H$) times. The first t values are discarded to avoid interference to the chaotic system. Combine three chaotic sequences x_1, x_2, x_3 to obtain a pseudo-random sequence S of floating-point number type.

Step 5: From the pseudo-random sequence S of floating-point number type, we get the pseudo-random number vector X of integer type length $2WH$, $X_i \in \{1, 2, ..., 10WH\}$. Then two pseudo-random sequences S_1 and S_2 are obtained from X.

Step 6: Convert the image matrix Q compressed by the neural network into a one-dimensional vector T. Pseudo random sequences S_1 and S_2 are used to scramble T. Then the one-dimensional vector is reduced to a $W\times H$ pixel matrix.

Step 7: Similarly, pseudo-random sequences S_3 and S_4 are obtained according to the fifth and sixth steps.

Step 8: pseudo-random sequences S3 and S4 are combined with diffusion algorithm $C_i=(C_{i-1}+S_i+P_i)\text{mod}(256)$ to diffuse pixel values. Among them, the scrambled image is expanded into P, S represents the cipher vector, and the corresponding ciphertext is C, i=1, 2, 3, ..., $W\times H$.

$C_i=(C_{i-1}+S_i+P_i)\text{mod}(256)$ expands as:

$$C_n = (C_0 + S_1 + S_2 + \cdots + S_n + P_1 + P_2 + \cdots + P_n) \bmod (256). \tag{9}$$

The positive diffusion can be expressed as:

$$C_i = \left(C_{i-1} + S_i + P_i\right) \bmod \left(256\right). \tag{10}$$

The reverse diffusion can be expressed as:

$$P_i = \left(2 \times 256 + C_i - C_{i-1} - S_i\right) \bmod \left(256\right). \tag{11}$$

Step 9: The ciphertext vector C can be obtained according to the above formula, and then it is restored to pixel moment. Then the encrypted image C is obtained.

3.4 Decryption algorithm

The image decryption process is the reverse process of image encryption. Fig.6 shows the decryption process, and the detailed decryption steps can be described as:

Step 1: Similarly, by the step 6 and 8 of the encryption processes, the pseudo-random sequences S_1, S_2, S_3 and S_4 of the reverse process are obtained.

Step 2: Restore the pixel value by formulas (12) and (13) to obtain the pixel matrix E.

The positive diffusion is:

$$C_i = \left(C_{i+1} + S_i + P_i\right) \bmod \left(256\right). \tag{12}$$

The reverse diffusion is:

$$P_i = \left(2 \times 256 + C_i - C_{i+1} - S_i\right) \bmod \left(256\right). \tag{13}$$

Step 3: Combined with the Arnold inverse transform algorithm, the pixel position of the pixel matrix after inverse diffusion is restored to obtain a one-dimensional vector F.

Step 4: Restore the pixel value of the one-dimensional vector F from [0, 1] to [0, 255], and the vector is restored to $a \times a$ sub-image block.

Step 5: Finally, combine all the sub-images into a complete image to get the decrypted image.

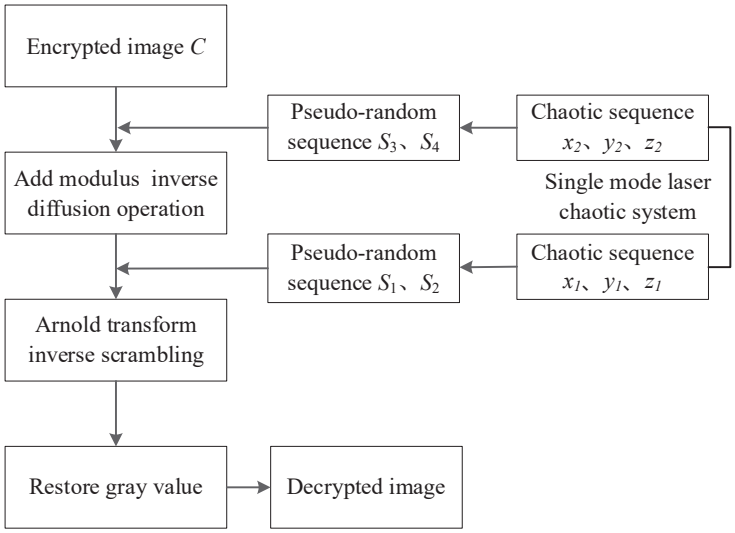
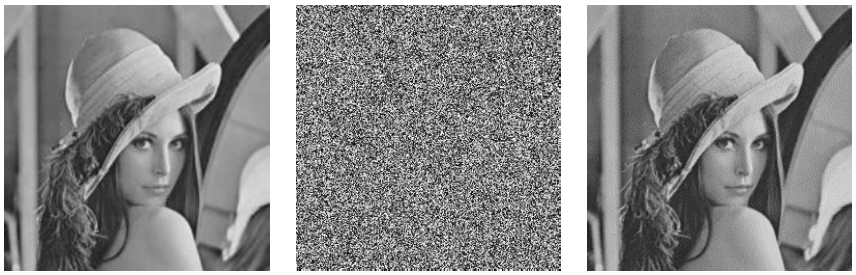


Fig.6 Decryption process

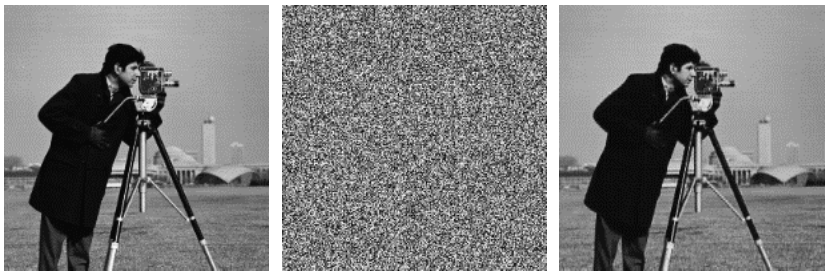
4 Algorithm test results and performance analysis

4.1 Algorithm test results

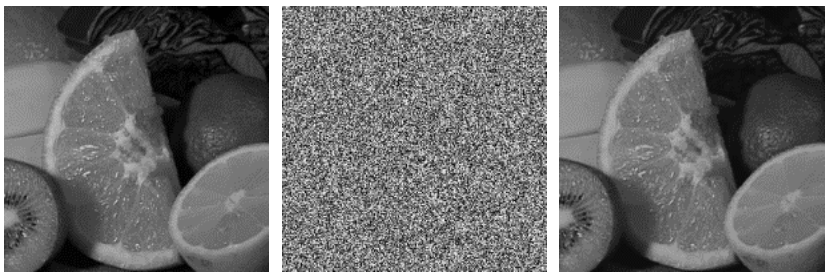
In this paper, 256×256 grayscale images are selected as the object to test the algorithm performance. The encryption and decryption results obtained by this scheme are shown in Fig.7. As can be seen from Figure 7, the encryption is effective and can effectively mask the information in the plaintext image. At the same time, the decryption algorithm can completely decrypt the ciphertext correctly. Since the security performance of image encryption determines whether the algorithm can effectively resist external attacks, it is necessary to analysis the performance of the encryption scheme through methods such as key sensitivity, histogram analysis, correlation analysis, and so on.



(a) Original Lena image (b) Encrypted Lena image (c) Decrypted Lena image



(d) Original Camera image (e) Encrypted Camera image (f) Decrypted Camera image



(g) Original Fruits image (h) Encrypted Fruits image (i) Decrypted Fruits image

Fig.7 Encryption and decryption test of plaintext image

4.2 Key space

The key space of the image cipher system should be large enough so that the encryption system can resist brute force attacks. The password length should be at least 128bit. In this algorithm, the key is about 292bit, so the key space is about 2^{292}. It can be seen from the comparison of Tab.1 with other encryption algorithms that the encryption system in this paper has a large key space, so it can withstand all kinds of brute force attacks.

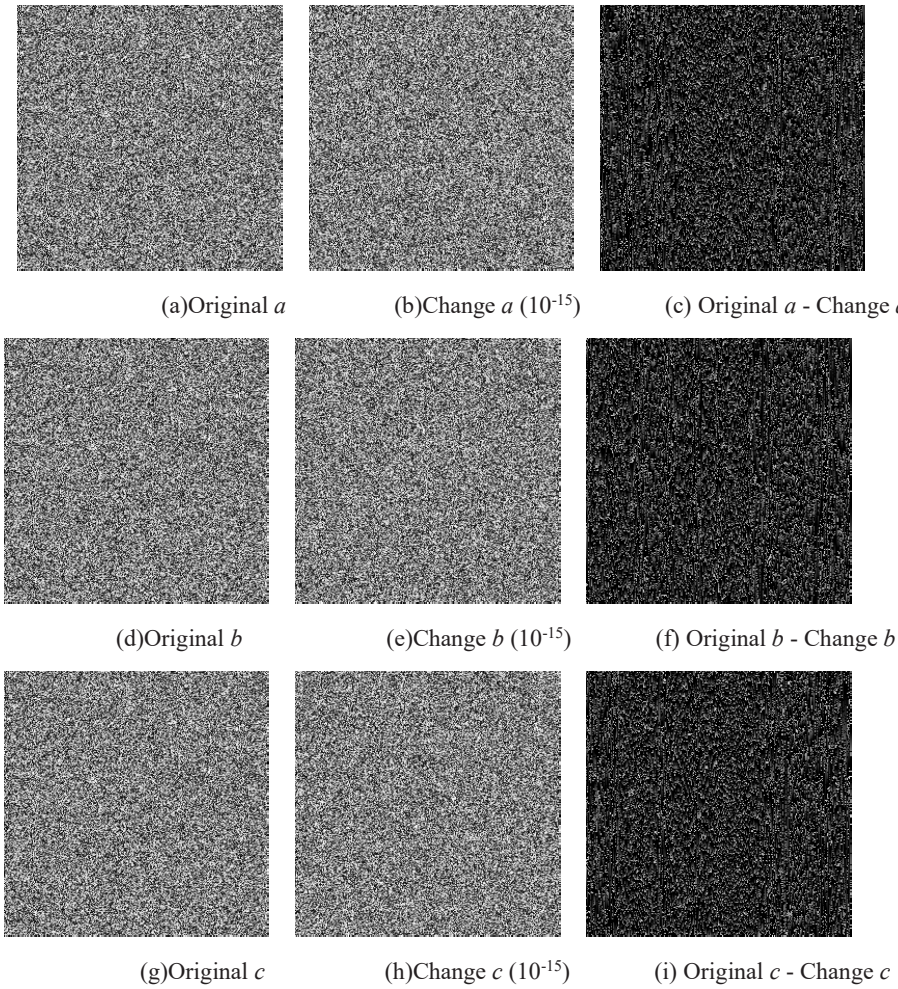
Tab.1 Key space comparison

Our	Ref.[19]	Ref.[21]	Ref.[22]
2^{292}	2^{128}	2^{128}	2^{213}

4.3 Key sensitivity analysis

Key sensitivity is an important characteristic of encryption algorithms. Due to the slight change of the key, the encryption effect will be very different. The degree of this difference can be evaluated by calculating the difference between two different ciphertexts. In this algorithm, sensitivity analysis is performed on keys *a*, *b*, and *c*. When the parameter has a slight change of 10^{-15}, the difference between the new ciphertext image and the original ciphertext image is shown in Fig.8. Fig.8 (c, f, i) shows that when the parameters change slightly, there is a difference between the obtained ciphertext image and the original ciphertext.

The experimental results indicate that the encryption system in this paper is very sensitive to the keys.



(a)Original a (b)Change a (10^{-15}) (c) Original a - Change a

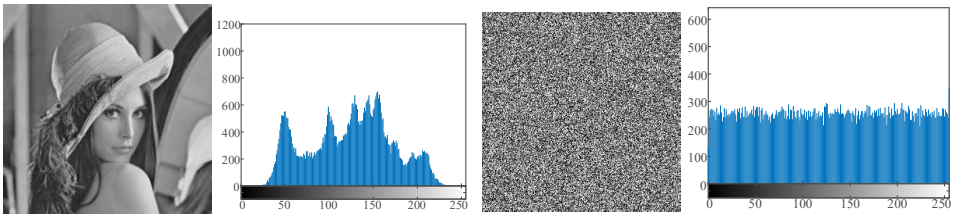
(d)Original b (e)Change b (10^{-15}) (f) Original b - Change b

(g)Original c (h)Change c (10^{-15}) (i) Original c - Change c

Figure.8 Sensitivity of secret keys

4.4 Histogram analysis

The distribution of image pixels can be expressed by a histogram. It can be seen from the simulation results of the “Lena” image in Fig.9 that the histogram of the plaintext image is not flat. In other words, the image contains the key information of the pixel, and the attacker can easily extract the key information of the image. The histogram distribution of the ciphertext image is relatively flat, which better hides the key information of the image. It indicates that the encryption scheme in this work can effectively resist statistical attacks.



(a)Plaintext of Lena (b) Histogram of plaintext (c) Ciphertext of Lena (d) Histogram of ciphertext

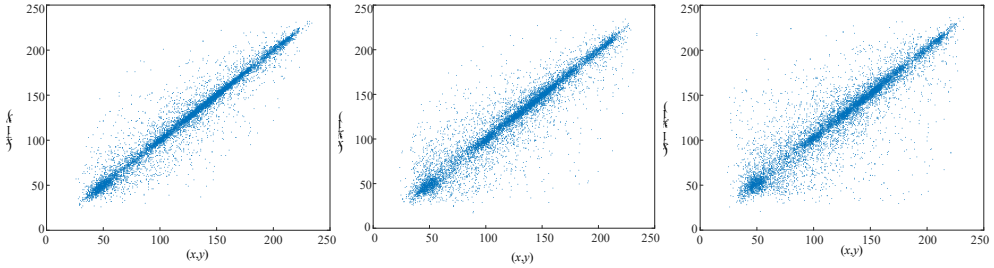
Fig.9 Histogram of original image and encrypted image

4.5 Correlation test

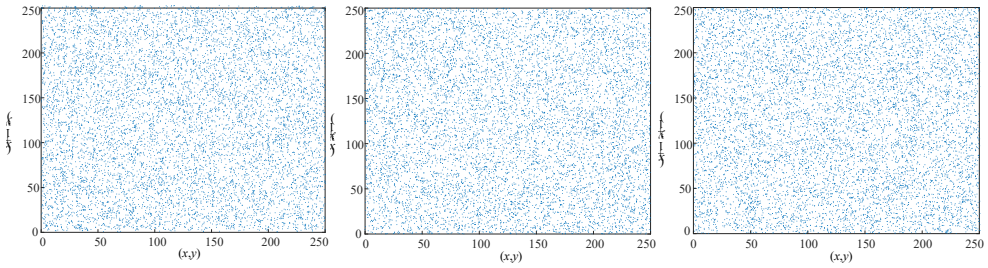
In general, the correlation between adjacent pixels of plaintext images in all directions is very high, so the ability to resist differential attack is weak. Equation (14) can be used to calculate the correlation of adjacent pixels in horizontal, vertical, and diagonal directions.

$$\begin{cases} E(u)=\dfrac{1}{N}\sum_{i=1}^{N}u_i \\ D(u)=\dfrac{1}{N}\sum_{i=1}^{N}\left(u_i-E(u)\right)^2 \\ Cov(u,v)=\dfrac{1}{N}\sum_{i=1}^{N}\left(u_i-E(u)\right)\left(v_i-E(v)\right) \\ r_{uv}=\dfrac{\left|Cov(u,v)\right|}{\sqrt{D(u)}\sqrt{D(v)}} \end{cases}, \tag{14}$$

where u and v are the gray values of two adjacent pixels, $E(u)$ represents the mean value, $D(u)$ is the variance, $Cov(u,v)$ means the covariance, and r_{uv} is the correlation. It can be seen from the experimental results of "Lena" in Fig.10 that the correlation between adjacent pixels of plaintext images is very high, while that of ciphertext images is very low. To further demonstrate the characteristics of correlation coefficients, the comparison results of correlation coefficients are shown in Tab.2. According to Tab.3, it can be seen that the ciphertext image has low correlation. In other words, the encryption algorithm proposed in this article has better anti-attack performance.



(a) Horizontal of plaintext (b) Vertical of plaintext (c) Diagonal of plaintext



(d) Horizontal of ciphertext (e) Vertical of ciphertext (f) Diagonal of ciphertext

Fig.10 Adjacent pixel distribution of Lena image

Tab.2 Comparison of Lena correlation coefficients

Direction	Plaintext	Ours	Ref.[19]	Ref.[21]	Ref.[22]
Horizontal	0.9698	-0.0005	0.0070	0.0004	-0.0055
Vertical	0.9406	-0.0002	-0.0102	0.0013	0.0075
Diagonal	0.9152	-0.0018	0.0030	-0.0023	0.0187

4.6 Anti-differential attack analysis

Minor changes in image pixels are usually measured by NPCR(rate of change in pixel number) and UACI(uniform mean intensity of change). NPCR and UACI are defined as

$$\begin{cases} NPCR=\dfrac{\sum_{i,j} D(i,j)}{W\times H}\times 100\% \\ UACI=\dfrac{1}{W\times H}\left[\sum_{i,j}\dfrac{\left|c_1(i,j)-c_2(i,j)\right|}{255}\right]\times 100\% \end{cases}, \tag{15}$$

where c_1 and c_2 represent two images of size $W\times H$. When $c_1(i,j)\neq c_2(i,j)$, $D(i,j)=1$ can be obtained. On the contrary, if $c_1(i,j)=c_2(i,j)$, then $D(i,j)=0$. The expectations for NPCR and

UACI were 99.6094% and 33.4635%, respectively. Comparison with different algorithms is shown in Table 3. The comparison results in Table 3 indicate that the proposed encryption scheme has good resistance to differential attacks.

Tab.3. Mean values of UACI and NPCR

Algorithm	Ours	Ref.[19]	Ref.[21]	Ref.[22]
NPCR	99.62%	99.62%	99.59%	99.01%
UACI	33.44%	33.56%	33.50%	33.69%

4.7 Information entropy test

The randomness of image information can be reflected by information entropy. The information entropy can be described as:

$$H(s)=\sum_{i=0}^{2^{n-1}} p(s_i)\log_2\frac{1}{p(s_i)}, \tag{16}$$

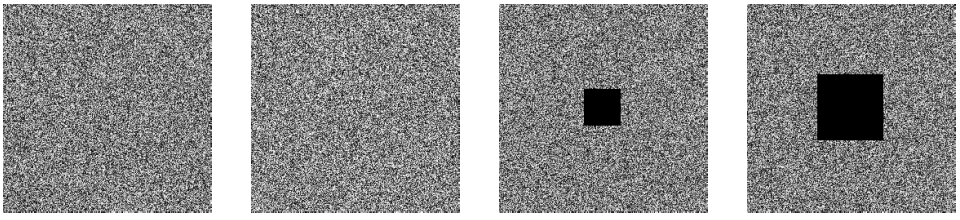
where $p(s_i)$ is the probability of s_i. The expected information entropy of 8-bit grayscale image is 8. According to the comparison of the "Lena" image test results in Tab.4 with the results of other algorithms, the information entropy of the encryption scheme in this work is very close to the ideal value of 8, which has high security.

Tab.4 Information entropy of different algorithms

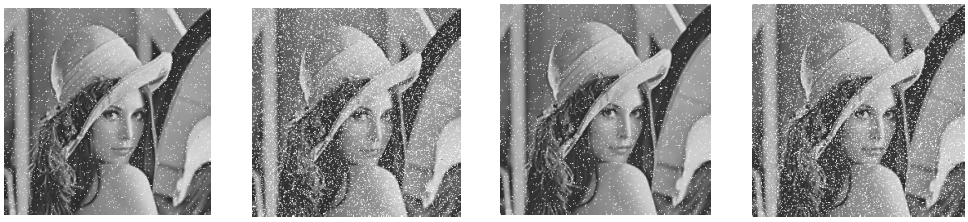
Image	Our algorithm	Ref.[19]	Ref.[21]	Ref.[22]
Lena	7.9960	7.9975	7.9978	7.9564

4.8 Robustness analysis

In the transmission and storage of digital images, different types of noise and data loss are prone to occur. Clipping attacks and noise attacks tests can be used as the evaluation methods of robustness analysis. Figures 11(a)-(d) show the salt and pepper noise attack and occlusion attack of the Lena encrypted image. Figure 11(e)-(h) show the decrypted images of the salt and pepper noise attack and occlusion attack. When the encrypted image loses some data or is blurred by noise, the decryption process can still restore the original image and has a high visual effect. The experimental results indicate that the encryption scheme proposed in this work has good anti-noise and anti-occlusion attack capabilities.



(a) S&P attack wih 1% (b) S&P attack wih 4% (c) data lost with 3% (d) data lost with 10%



(e) Decrypted image of (a) (f) Decrypted image of (b) (g) Decrypted image of (c) (h) Decrypted image of (d)

Fig.11 Robustness analysis

5 Conclusion

An image encryption algorithm based on a laser chaotic system and neural network is proposed in this work. The sensitivity of the laser chaotic system to initial conditions is used to enhance the secret key space and its sensitivity. At the same time, this work combines BP neural network and Arnold transform to effectively reduce the correlation between adjacent pixels of the image and improve the performance of resisting statistics and sensitive attacks. By comparing and analyzing the security with other encryption algorithms, it can be seen that the algorithm proposed in this work has high security. Therefore, this algorithm is suitable for secure communications to protect the secure transmission of digital image information on the Internet and has good practical significance and application prospects.

References

[1] Xu J, Sun B, Yang F, et al. A trademark graphic encryption algorithm based on discrete chaotic system [J]. Journal of Dalian Polytechnic University, 2019, (3).

[2] Mou J, Yang F, Chu R, et al. Image Compression and Encryption Algorithm Based on Hyper-chaotic Map[J]. Mobile Networks and Applications, 2019, (3): 1-13.

[3] Alawida M, Samsudin A, Sen Teh J, et al. A new hybrid digital chaotic system with applications in image encryption[J]. Signal processing, 2019, 160(JUL.): 45-58.

[4] Al. C P E. Understanding Cryptography[J]. Springer-Verlag Berlin Heidelberg 2010.

[5] Hua Z, Xu B, Jin F, et al. Image Encryption Using Josephus Problem and Filtering Diffusion[J]. IEEE Access, 2019, 7: 8660-8674.

[6] Zhu C, Sun K. Cryptanalyzing and Improving a Novel Color Image Encryption Algorithm Using RT-Enhanced Chaotic Tent Maps[J]. IEEE Access, 2018, 6.

[7] Yang F, Luo C, Mou J, et al. Analysis of memory chaotic synchronization based on driver-response synchronization algorithm and coupling synchronization algorithm [J]. Journal of Dalian Polytechnic University, 2019, (3): 229-234.

[8] Yang F, Mou J, Ma C, et al. Dynamic analysis of an improper fractional-order laser chaotic system and its image encryption application[J]. Optics and Lasers in Engineering, 2020, 129.

[9] Wang X, Wang Q, Zhang Y. A fast image algorithm based on rows and columns switch[J]. Nonlinear Dynamics, 2015, 79(2): 1141-1149.

[10] Haibo L, Bin G. Image encryption based on Henon chaotic system with nonlinear term[J]. Multimedia Tools and Applications, 2019, Vol.78 (24): 34323-34352.

[11] Jiang D, Chen Y, Gu X, et al. Efficient and universal quantum key distribution based on chaos and middleware[J]. International Journal of Modern Physics B, 2017, 31(02): 1650264.

[12] Zhou N, Hu Y, Gong L, et al. Quantum image encryption scheme with iterative generalized Arnold transforms and quantum image cycle shift operations[J]. Quantum Information Processing, 2017, 16(6): 164.

[13] Shen Q, Liu W, Lin Y, et al. Designing an Image Encryption Scheme Based on Compressive Sensing and Non-Uniform Quantization for Wireless Visual Sensor Networks[J]. Sensors (Basel), 2019, 19(14).

[14] Yang X, Wu H, Yin Y, et al. Multiple-image encryption base on compressed coded aperture imaging[J]. Optics and Lasers in Engineering, 2020, 127.

[15] Ye G, Pan C, Dong Y, et al. Image encryption and hiding algorithm based on compressive sensing and random numbers insertion[J]. Signal Processing, 2020.

[16] Zhang X, Han F, Niu Y. Chaotic Image Encryption Algorithm Based on Bit Permutation and Dynamic DNA Encoding[J]. Comput Intell Neurosci, 2017, 2017: 6919675.

[17] Wang X, Wang Y, Zhu X, et al. A novel chaotic algorithm for image encryption utilizing one-time pad based on pixel level and DNA level[J]. Optics and Lasers in Engineering, 2020.

[18] Chen X, Hu C J. Adaptive medical image encryption algorithm based on multiple chaotic mapping[J]. Saudi J Biol Sci, 2017, 24(8): 1821-1827.

[19] Li B, Liao X, Jiang Y. A novel image encryption scheme based on logistic map and dynatomic modular curve[J]. Multimedia Tools and Applications, 2017, 77(7): 8911-8938.

[20] Chai X, Zhang J, Gan Z, et al. Medical image encryption algorithm based on Latin square and memristive chaotic system[J]. Multimedia Tools & Applications, 2019, 21.

[21] Wang S, Wang C, Xu C. An image encryption algorithm based on a hidden attractor chaos system and the Knuth–Durstenfeld algorithm[J]. Optics and Lasers in Engineering, 2020.

[22] Abbas A E, Abdulbaqi M A, H. L S. Block image encryption based on modified playfair and chaotic system[J]. Journal of Information Security and Applications, 2020.

[23] Niu Y, Sun X, Zhang C, et al. Anticontrol of a Fractional-Order Chaotic System and Its Application in Color Image Encryption[J]. Mathematical Problems in Engineering, 2020, 2020: 1-12.

[24] Song W, Zheng Y, Fu C, et al. A Novel Batch Image Encryption Algorithm Using Parallel Computing[J]. Information Sciences, 2020: 211-224.

[25] Chen L, Yin H, Huang T, et al. Chaos in fractional-order discrete neural networks with application to image encryption[J]. Neural Netw, 2020, 125: 174-184.

[26] Zhang Y, Chen A, Tang Y, et al. Plaintext-related image encryption algorithm based on perceptron-like network[J]. Information Sciences, 2020, 526: 180-202.

[27] Liu Y, Jiang Z, Xu X, et al. Optical image encryption algorithm based on hyper-chaos and public-key cryptography[J]. Optics and Laser Technology, 2020.

[28] Zhou N, Hu Y, Gong L, et al. Quantum image encryption scheme with iterative generalized Arnold transforms and quantum image cycle shift operations[J]. Quantum Information Processing, 2017, 16(6):164.

Proceedings of the 8th EAI International Conference on Green Energy and Networking

June 6-7, 2021, Dalian, People's Republic of China

GreeNets 2021

www.eai.eu

https://greenets.eai-conferences.org/2021/

ISBN: 978-1-63190-313-7

European Alliance for Innovation

EAI is a non-profit organization with free membership and the largest open professional society for advancing research careers through community collaboration and fair recognition. Members benefit from finding feedback and mentorship for their work and they are guaranteed to be evaluated fairly, transparently, and objectively through community.

ISBN: 978-1-63190-313-7

ISSN: 2593-7642

http://eudl.eu/series/CCER | www.eai.eu