Research on Coding Method of Microscopic Video Signal Based on Machine Learning

Hai-xiao Gong[1,2](✉) and Jie He[1,2]

[1] School of Data Science and Software Engineering, Wuzhou University, Wuzhou 543002, China

[2] Guangxi Colleges and Universities Key Laboratory of Image Processing and Intelligent Information System, Wuzhou University, Wuzhou 543002, China

Abstract. At present, the commonly used microscopic video signal coding methods have poor processing ability and low coding accuracy. Therefore, this paper proposes a new micro video coding method based on machine learning technology. Firstly, the video coding processing architecture is established by intra prediction, inter prediction, transformation, quantization, entropy coding and loop filtering, and then the coding processing is realized by image segmentation, intra prediction and inter prediction, and the depth decision method is used for depth analysis. The experimental results show that this method can effectively improve the processing ability of microscopic video signal coding, and at the same time enhance the coding accuracy.

Keywords: Machine learning · Microscopic video signal · Video coding processing · Depth decision

1 Introduction

With the continuous updating of mobile terminal technology, smart phones with stronger performance and more functions have become the choice of many consumers. If the two smart phone operating systems represented by Android and IOS make smart terminals more valuable, then the emergence of various types of video application software has substantially changed the way video signals are obtained [1, 2]. Video chat, remote meetings, and various live broadcast software have changed the way people work and live, and real-time video applications have attracted more and more attention. It can be determined that the current low-resolution video will slowly disappear in people's lives. This type of video mainly includes two formats of 360P (480 × 360) and 480P (640 × 480). And 720P (1280 × 720) HD video and 1080P (1920 × 1080) Full HD video will become the mainstream video formats in the future. At the same time, international manufacturers, mainly Samsung, have also introduced displays that support 4K (3840 × 2160). With the continuous deepening of research and development and the reduction of material costs, high-definition displays supporting 2K (2560 × 1440) and 4K (3840 × 2160) will be greatly popularized [3, 4].

S. Liu and X. Ma (Eds.): ADHIP 2021, LNICST 417, pp. 110–123, 2022.
https://doi.org/10.1007/978-3-030-94554-1_10

The improvement of intra coding depth algorithm and inter coding depth algorithm has become an effective means to reduce the complexity of hevc coding signal. At present, one of the main research areas in this field is to simplify the process of data depth coding [5].

Machine learning is closely related to computational statistics, so it is also called statistical learning method and is an important branch in the field of artificial intelligence. Machine learning hopes that through continuous learning and adjustment of computers, it will eventually enable it to realize human learning capabilities. The essence of machine learning is spatial search and the generalization of functions. The main step is to use the features extracted from the data as input, abstract and construct a certain data model, obtain the information in the data, and use the obtained model to solve the unknown data analysis and prediction. According to the predicted results, the appropriate machine learning model is continuously selected until the generalization ability satisfactory to the unknown data can be obtained.

In HEVC, the improvement of the coding depth algorithm is considered to be the key technology of current HEVC to reduce the complexity of the coded signal to achieve practicality. As a recent research hotspot in academia, machine learning has gained a large range with its powerful learning and prediction capabilities Applications. Therefore, in order to improve the processing ability and coding accuracy of the micro-video signal coding processing method, this paper proposes a new micro-video signal coding processing method based on machine learning technology. First by frame prediction, interframe prediction, transform, quantization and entropy coding and processing module based video coding such as loop filtering architecture, then through the image block segmentation, frame prediction, interframe prediction coding for processing, and carries on the deep analysis of reference depth decision method, thus effectively improve the ability of handling the microscopic video signal encoding, and strengthened the coding accuracy.

2 Microscopic Video Signal Encoding Processing Architecture Based on Machine Learning

The hybrid coding framework based on machine learning is adopted. The framework mainly includes intra prediction, inter prediction, transformation, quantization, entropy coding and loop filtering. The detailed coding framework is shown in Fig. 1.

In Fig. 1, the original video sequence on the left is input into the encoder and then divided in the order of image group, frame, slice and coding unit CU. These different levels of division levels can make full use of the information of the coded unit and reduce the coded data. the amount. The encoder will have a set of initial coding parameters before specific coding, which are assigned to GOP level, frame level and coding block unit level [6–8]. However, in the actual encoding process, the encoding may not be performed as expected. When each layer is encoded, the parameters will be recalculated according to the encoding quality of the reference image and the complexity of the current image [9–11].

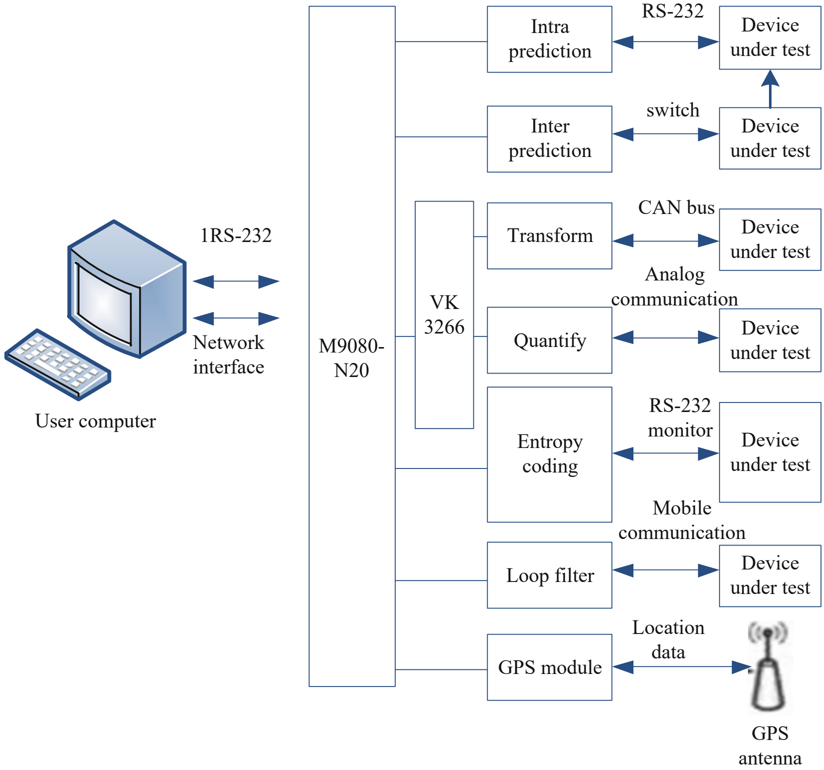

Fig. 1. Architecture of micro video signal coding and processing based on machine learning

It can only be used for reference in the first frame. At the specific coding unit level, each frame is first divided into the largest coding unit with 64 × 64 pixels. Specifically to the coding process of each LCU, the coding unit can continue to be decomposed into prediction unit and transformation unit. The encoder will traverse all the mode selection combinations and decide whether to continue the division by calculating the cost of each combination mode. This process is repeated until all encoding modes are traversed, and the optimal encoding mode is obtained. The encoding mode is stored, and the information compression of the current encoding block is completed through the process of transformation, quantization and direct encoding. After the encoder completes the encoding of the current encoding block or the slice where the encoding block is located, the obtained encoding parameter information and the preset information of the previous encoding block are uniformly packaged, and the necessary network propagation information is added for network transmission. After receiving the corresponding code stream, the receiver carries out the video and decoding process according to the reverse flow of the coding process and the pre-defined syntax rules.

Although Fig. 1 still uses a hybrid coding framework, the main coding modules are the same as those in the previous coding framework. But hybrid coding based on machine learning introduces new coding techniques in almost every module. The purpose of this design is to better handle high-resolution video signals. As a new segmentation mode, the flexible partition structure based on quadtree can represent video content more flexibly and effectively, enabling the encoder to select a more appropriate encoding mode according to the local characteristics of the image. 35 intra-prediction modes including 33-degree prediction, DC prediction mode and Planar prediction mode can better express the local features of high-definition video, match the complex and diverse textures in the video, obtain better prediction results, and achieve effective removal of spatial redundancy purpose. The asymmetric prediction unit division and motion merging technology makes good use of spatial correlation to reduce the motion parameter redundancy between adjacent blocks. The adaptive motion vector prediction technique also uses the correlation between spatial domain and temporal domain to remove the motion parameter redundancy [12, 13]. The sample adaptive compensation technique adds the corresponding offset value to the pixel according to the statistical characteristics of the pixel after the block filtering, which further improves the texture performance characteristics of the image and can improve the subjective quality and compression efficiency of the image.

3 Microscopic Video Signal Coding Processing Based on Machine Learning

3.1 Image Segmentation

In this paper, a quadtree based coding partition structure is introduced into the coding standard, which mainly includes coding unit Cu, prediction unit Pu and transformation unit tu. the purpose is to make the encoder better capture the characteristics of high-resolution video signal. In the coding process, the input video sequence is first divided into a series of coding tree units. Although the concept of coding tree is similar to that of macroblock in the previous video coding standard H.264, there is a big difference in the allowed pixel size between them.

The coding unit CU is the most basic unit of coding, and the CTU is the basic unit for dividing the CU. According to the quadtree division strategy and the video content adaptive division mechanism, the CTU can be evenly divided into four identical CUs, and each CU can still be divided in a recursive manner until the most suitable division for the image content is selected structure. CTU is finally divided into multiple CUs, and the size of the CU may be the same as or different from the CTU. The structure of a possible coding unit CU is shown in Fig. 2.

PU is the prediction unit, which is the basic unit to transmit the prediction information such as image index, prediction vector and prediction direction. Its root node is CU. There are two modes of PU division, one is symmetrical and the other is asymmetrical. There are four symmetrical and four asymmetrical modes. The specific segmentation shape of PU is shown in Fig. 3.

The use of an asymmetric rectangular PU partition structure can further match the prediction unit with the object boundary in the image. The PU division is related to the prediction mode. In the intra prediction mode, 2N $\times$ 2N and N $\times$ N PU divisions can be used. At this time, the shape of the PU is all square; in the inter prediction mode, the

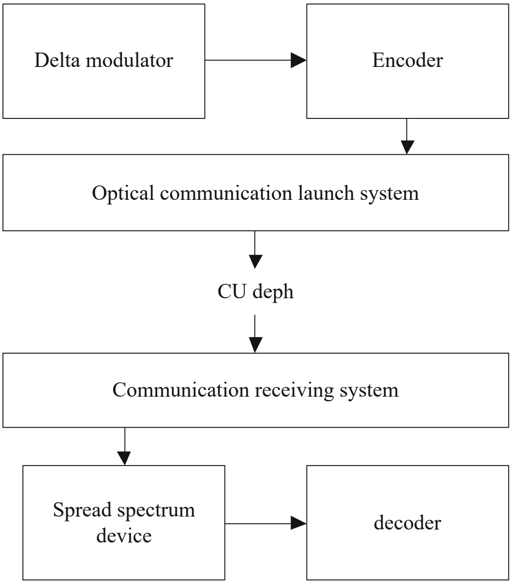

Fig. 2. Structure diagram of coding unit CU

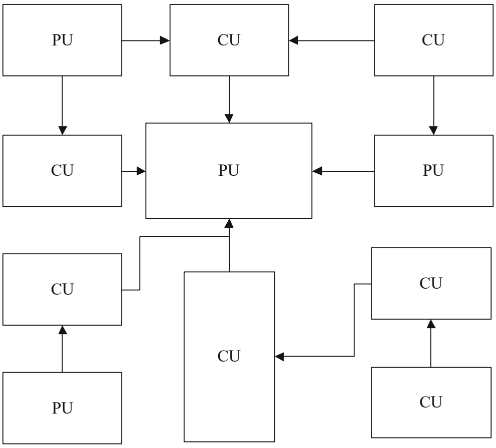

Fig. 3. Pu structure of prediction unit

above 8 types of PU can be used; But when the skip prediction mode is adopted, only the 2N × 2N PU division method can be adopted.

TU is a transformation unit, but its root node is CU. TU is the basic unit used in the process of transformation and quantization, which is used to represent the residual information of the current block after transformation. The shape of TU is square, and the size of TU supported in hevc is 4 × 4–32 × 32, so each CU may contain one or more TU. In the CU block using inter mode coding, TU can exist across the Pu boundary, while in the CU block using intra mode coding, TU cannot exist across the PU boundary, that is to say, TU is limited to a single PU. The former is suitable for all PU sizes, while the latter is only suitable for the case of intra PU size of 4×4.

The root nodes of PU and TU are both CU, but their partition methods are not limited. This structure greatly increases the flexibility of the encoder. With the concepts of coding tree unit CTU, coding tree block CTB and Quadtree Partition Method, the compression rate of high-resolution video can be further improved.

3.2 Intra Prediction

The principle of intra prediction is to use the correlation in the image space to predict the current pixel value using the reconstructed pixels in the adjacent position, and use the difference between the predicted value and the current pixel value as the basis for subsequent processing to reduce the amount of data transmission. In intra-frame prediction, the design of prediction modes with multiple angles will help remove pixel redundancy more effectively. The intra-frame prediction has been further optimized. It has increased the brightness image intra-frame prediction modes from 9 to 35, including Planar mode prediction, DC prediction, and angle prediction in 33 different directions.

The 35 prediction modes in the microscopic video signal coding process based on machine learning are defined on the basis of the PU, but the specific intra prediction process of the encoder is based on the TU as the basic prediction unit. Microscopic video signal coding based on machine learning makes full use of the directional characteristics of the image in the intra-frame prediction design, and the selection process of its prediction mode is also more complicated. The main process is to first build the MPM list of the current frame; after the list is built, use the Hadamard transform instead of the integer DCT transform to roughly select the 35 intra-frame modes; finally, perform normal coding on the roughly selected 8 modes and the MPM prediction mode. The prediction mode with the least cost is selected as the optimal coding mode, and the current block is compressed and coded.

3.3 Inter Prediction

Video signal sequence has strong temporal and spatial correlation, which is also the theoretical basis of video compression coding. The content similarity of the coding blocks in the same position between the current frame and the previous and subsequent frames is high. Inter frame prediction is an important technology to remove the temporal redundancy of video sequence in the coding standard.

In addition to traditional techniques such as motion estimation and reference frame management, more complex interframe motion vectors and prediction mechanisms, including Merge technology and AMVP technology, are introduced in the process of interframe prediction. The similarity between Merge technology and AMVP technology is that both use spatial correlation and temporal correlation to reduce the redundancy of motion parameters. The whole process mainly consists of two steps:

Step 1: select the motion parameters of the adjacent Pu according to a certain order, and establish the candidate MV list;

Step 2: according to the parameters in the candidate MV list, select the best performance as the motion parameters of the current PU.

Although merge technology and amvp technology are basically the same in the selection of motion parameters. This difference is mainly manifested in the following two aspects:

(a) In merge mode, there is no difference in MV, which is directly predicted by the adjacent PU. In amvp mode, there is a difference value of MV, and the encoder only encodes the difference value;
(b) There are also differences between merge mode and amvp mode in the way of establishing candidate MV list and the length of the list.

4 Coding Depth Decision of Microscopic Video Signal Based on Machine Learning

In order to cope with the high-efficiency coding and decoding of high-definition video and to improve coding efficiency in essence, this article has made a relatively large improvement to the coding technology in H.264/AVC. Including quadtree coding block division, intra and inter coding depth selection algorithms. The main process of this link is to traverse all the programming modes of the existing coding block, and perform the coding and rate-distortion optimization process for each coding mode, and then determine whether to continue the division. The optimal coding mode in the current mode is selected for coding without further division. The coding mode is shown in Fig. 4.

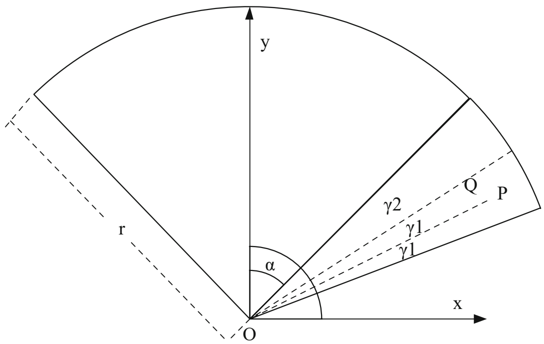

Fig. 4. Video signal coding mode

The essence of video compression is to seek the balance between the video distortion D and the number of bits r of information. The optimal solution is selected by using RDO. This section briefly introduces the video distortion degree and rate distortion optimization technology.

4.1 Video Signal Distortion

Subjective quality assessment and objective quality assessment are two common mechanisms for measuring video signal distortion. Because the evaluation standard of subjective quality assessment is not easy to be quantified, it is difficult to carry out and its accuracy cannot be guaranteed. Therefore, scholars mostly use objective quality assessment methods. There are many ways to evaluate the objective quality of the video signal, the purpose of which is to evaluate the distortion of the current video signal from multiple objective angles. The main measurement methods are square error, mean square error, absolute error sum and peak signal-to-noise ratio, the specific formula is as follows:

$$SSD = \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} |\hat{f}(x, y) - f(x, y)|^2 \tag{1}$$

$$MSE = \frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} |\hat{f}(x, y) - f(x, y)|^2 \tag{2}$$

$$SAD = \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} |\hat{f}(x, y) - f(x, y)| \tag{3}$$

$$PSNR = 10 \log_{10} \frac{(255)^2 WH}{\sum_{x=0}^{W-1} \sum_{y=0}^{H-1} |\hat{f}(x, y) - f(x, y)|^2} \tag{4}$$

In the above formula, $f(x, y)$ and $\hat{f}(x, y)$ respectively represent the original value and the reconstructed value at (x, y), w is the width of the image in the current video, and H is the height of the image.

4.2 Rate-Distortion Curve

The relationship between bit rate and distortion is a convex function, which also determines the relationship between them, as shown in Fig. 5. For a given rate distortion curve, when the distortion D is given, the required minimum rate R can be obtained, which is only a theoretical value. In practical application scenarios, due to various resource constraints, the optimal theoretical value can not be achieved.

In the actual encoding process, a certain encoding mode will be selected, and the distortion D and the rate R in the corresponding mode will be obtained respectively, which are shown in Fig. 6 as a series of operating points. The outer envelope curve is obtained by connecting the operating points, which is the solid line in Fig. 6. The curve also points out the optimal performance that the encoder can achieve in the actual encoding process. In the actual encoding process, given the code rate R, find the operating point closest to the rate-distortion curve by traversing all the encoding mode costs. This is also the main process of RDO.

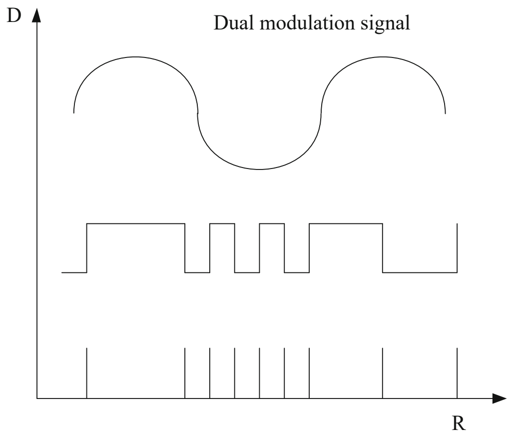

Fig. 5. R-D rate-distortion function curve diagram (R-D curve)

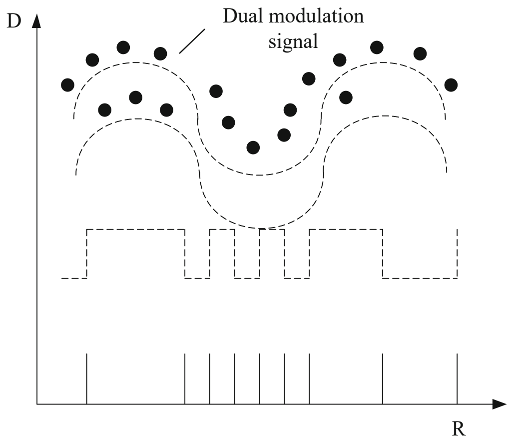

Fig. 6. Schematic diagram of video signal encoding rate-distortion curve

K-means algorithm is used for coding. As the simplest clustering algorithm in machine learning, K-means clustering algorithm is famous for its efficient classification. The main idea of the algorithm is to cluster n point $x_1, x_2, x_3, \ldots, x_n$ into $k(k < n)$ cluster $S = \{\mathrm{S}_1, \mathrm{S}_2, \ldots \mathrm{S}_k\}$, so that each point belongs to the nearest cluster set. The minimum clustering principle is shown in Eq. (5):

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{x_j \in S_i} \left| x_j - \mu_i \right|^2 \tag{5}$$

Among them, μ_i is the center point of each cluster, and x_j represents any point belonging to the cluster set S_i.

The k-means clustering algorithm can be regarded as a process of continuous iteration, adjustment and optimization. The main steps of the algorithm are as follows:

Step 1: Select k points from all the input n sample points as the initial cluster center points;

Step 2: According to the selected k cluster centers, traverse all points except the k cluster center points, find the cluster center closest to each point, and assign it to the corresponding cluster;

Step 3: When a cluster is completed, that is, all points in the sample have corresponding cluster centers, recalculate the mean of each cluster to adjust the cluster center point, and traverse all the points again to find the nearest cluster Center point, add this point to the corresponding cluster;

Step 4: Repeat Step 3 until the cluster centers of the previous two iterations no longer change.

According to the definition and steps of clustering algorithm, assuming that the final clustering process is completed t times, the time complexity of the whole process is $O(n \times k \times t)$, which requires a lot of calculation. If the online training method is used for k-means clustering, it will cost a lot of time, so this paper chooses the off-line training method for k-means clustering, according to the training center point to judge which cluster the input point belongs to in the actual coding process.

In this paper, the texture complexity of the image is represented by the difference between the pixels in the coding region.

The definition of complexity is shown in Eq. (6):

$$C_{3\times 3} = \frac{1}{3\times 3}\sum_{t=1,j=1}^{i=3,j=3}\left(P_{i,j} - \frac{1}{3\times 3}\sum_{i=1,j=1}^{i=3,j=3} P_{i,j}\right)^2 \tag{6}$$

Among them, $P_{i,j}$ represents the specific pixel value whose current position is (i, j), and $C_{3\times 3}$ represents the texture complexity of the entire block.

Due to the richness and variety of image content, it is impossible to completely determine whether the current CU continues to be divided based on the complexity of the current pixel block. At this time, the complexity of the four sub-blocks needs to be further calculated. For example, for a block with a size of 64 × 64, 64 $C_{8\times 8}$, 16 $C_{16\times 16}$, and 4 $C_{32\times 32}$ need to be calculated in sequence. In the whole process, only 64 8 × 8 area averages need to be calculated, and the combined calculation can be obtained. The average value of other pixels is calculated, and the amount of calculation in this part can be ignored.

If the complexity of the current coding block is small, the complexity of the corresponding four sub blocks is also small, which means that the pixel distribution of the coding block is balanced, and the video content is similar, so it is not necessary to continue to divide. On the contrary, if the complexity difference between the current block and the sub block is large, especially when the complexity difference between the four sub blocks is large, the current coding block should continue to be divided.

5 Experimental Analysis

Podk technology and symmetric encryption algorithm are commonly used in traditional video signal encoding and processing methods. In the process of encrypting the video information, PODK technology mainly uses the accurate data monitoring device to encode the information in real time, but this method has weak data processing capacity; while the symmetric encryption algorithm uses the secret key to establish the communication protocol for the security data of the online communication network, so as to ensure the security of the online confidential information, so as to improve the security Realize coding. However, this algorithm is weak in the perception of offline network information, which can not guarantee the overall perception of network secrets. In order to verify the effectiveness of the method designed in this paper, experiments were designed to compare the application performance of the three methods, so as to verify the feasibility of the method proposed in this paper.

In the process of encoding the confidential information in the video, it is necessary to set the security sensing module in every link of the whole communication system. Security module also contains information isolation, authentication and encryption functions. The way of video information transmission in the network is selective. Therefore, in this experiment, a variety of sensing methods are used to perceive the transmitted data in different channels, so as to judge its security.

The data perception layer in the experiment includes confidential information source, confidential information collection system and confidential security guarantee system. Figure 7 shows the topology of the operating environment of this experiment.

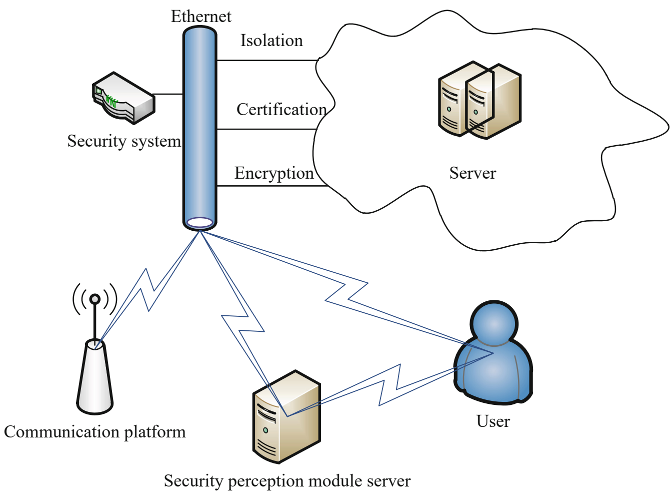

Fig. 7. Topology diagram of experimental operating environment

In the process of the experiment, the video signal coding first needs to transmit the video information to the space network in the system. The transmission process of the video signal in the network is different from that in the processing space. The transmission of the video signal in the network space depends on the security protocol, while the data transmission in the processing space is a secure form of information security isolation Lose. The video signal first enters the acquisition system of related equipment from the perception module of the system, and finally transmits to the data terminal, and arrives at the perception system of video signal through the security processing and security transmission protocol of the data terminal. When the data passes through the isolation module, it encodes the video signal, transmits the encoded information to the intelligent terminal, and then uploads it to the cyberspace. The special server and data processor in the cyberspace encode the information security. After the terminal interface in cyberspace works stably, the staff can observe the working state of the perceptual information data in the system in real time, and the terminal data in different states changes dynamically.

In the experiment, after the environment of the perception system is stable, the video signal can be received and the recognition side coding can be performed. This article first compares the video signal processing capabilities of three different systems, and the results are shown in Fig. 8.

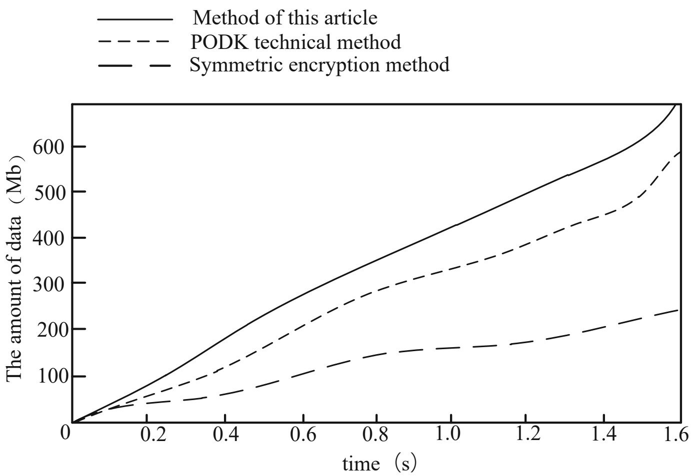

Fig. 8. Comparison of video signal processing capabilities of three different methods

According to the comparison results in Fig. 8, the method in this paper has the strongest processing ability for video signal. This is because this system uses rs245 standard serial port to receive more video signals, and can carry on the preliminary processing to the video information, slow down the later data processing work, and uses machine learning to perceive and process the data to be processed to ensure the security and efficiency of data processing in the form of encryption. However, the two traditional methods mainly use the common function operation mode, which has certain limitations and can not effectively process the real-time video signal.

On this basis, this paper compares the video signal coding accuracy of the three methods, and the results are shown in Fig. 9.

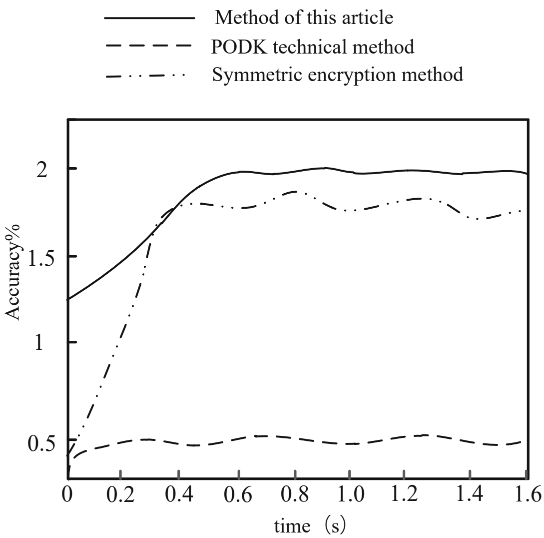

Fig. 9. Comparison of encoding information accuracy under three methods

It can be seen from Fig. 9 that the coding information perception ability of the method in this paper is the strongest, and can perceive coding data with an accuracy of 0.5% in a short time, while the perception ability of the traditional method is about 2% on average. This is because the method in this article uses a self-designed SD flash memory card, which not only has the function of data caching, but also accurately extracts and recognizes data, which increases the work intensity of the coding information recognition module. The method in this article also has an independent coding module that can Independently complete the perceptual operation of encoded information, which greatly improves the perceptual ability of encoded information. Traditional methods mainly perceive the online coding information, but ignore the offline confidential information perception, which leads to a decrease in the average coding ability of the system.

6 Conclusion and Prospect

In this paper, a new micro video signal coding method is designed by using machine learning method. The key technologies and algorithms are designed, and the basic control work is completed by anonymous data controller. The experimental results show that the method designed in this paper has stronger processing ability and higher coding accuracy.

However, due to the limitation of research time and other environmental conditions, the method in this paper still has a series of problems to be improved. In the following research, the method in this paper will be further improved from the perspective of improving the conversion speed between network protocols.

Fund Projects. National Natural Science Foundation of China: The Key Technologies about Fast Coding and Quality Controlling of Fractal Image Compression (61961036).

Basic Ability Improvement Project for Young and Middle-aged Teachers in Guangxi: Research on 3D Terrain Rendering for Large Scene Oblique Photography (2020KY17019).

Natural Science Foundation of Guangxi:Research on the Key Technologies about Decoder for Reliable Transmission of HEVC for Microscopic Video (2020JJA170007).

References

1. Wang, J., Di, Y., Rui, X.: Research and application of machine learning method based on swarm intelligence optimization. J. Comput. Meth. Sci. Eng. **19**(2), 1–9 (2019)
2. Antolinez, F.V., Rabouw, F.T., Rossinelli, A.A., et al.: Observation of electron shakeup in CdSe/CdS Core/Shell nanoplatelets. Nano Lett. **19**(12), 8495–8502 (2019)
3. Zhang, C., Zhou, Y., Guo, J., Wang, G., Wang, X.: Research on classification method of high-dimensional class-imbalanced datasets based on SVM. Int. J. Mach. Learn. Cybern. **10**(7), 1765–1778 (2018). https://doi.org/10.1007/s13042-018-0853-2
4. Zeng, W., Xu, H., Li, H., et al.: Research on methodology of correlation analysis of sci-tech literature based on deep learning technology in the big data. J. Database Manag. **29**(3), 67–88 (2018)
5. Jinnouchi, R., Lahnsteiner, J., Karsai, F., et al.: Phase transitions of hybrid perovskites simulated by machine-learning force fields trained on-the-fly with Bayesian inference. Phys. Rev. Lett. **122**(22), 225701 (2019)
6. Chen, Y., Tao, J., Wang, J., et al.: The novel sensor network structure for classification processing based on the machine learning method of the ACGAN. Sensors **19**(14), 3145 (2019)
7. Qin, F., Xu, D., Zhang, D., et al.: Robotic skill learning for precision assembly with microscopic vision and force feedback. IEEE/ASME Trans. Mechatron. **24**(99), 1117–1128 (2019)
8. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. Complex Intell. Syst. **7**(4), 1895–1917 (2020). https://doi.org/10.1007/s40747-020-00161-4
9. Liu, S., Lu, M., Li, H., et al.: Prediction of gene expression patterns with generalized linear regression model. Front. Genet. **10**, 120 (2019)
10. Liu, S., Bai, W., Zeng, N., et al.: A fast fractal based compression for MRI images. IEEE Access **7**, 62412–62420 (2019)
11. Wu, Q., Zhang, C., Zhang, M., et al.: A modified comprehensive learning particle swarm optimizer and its application in cylindricity error evaluation problem. Int. J. Perform. Eng. **15**(3), 2553 (2019)
12. Zhou, B., Li, H.: Research on video compression coding based on embedded system. Wirel. Internet Technol. **16**(18), 39–41
13. Xue, R.B., Ma, M.Y., Jin, S., et al.: Surveillance video coding algorithm based on object. Mod. Comput. **27**(3), 41–45 (2019)