



Fall Behavior Recognition Algorithm in Video Surveillance Based on Feature and Deep Learning

Hai-jing Zhou^(✉)

Chongqing Vocational Institute of Tourism, Chongqing 409099, China
zhouhaijing554@yeah.net

Abstract. Aiming at the problem of low recognition rate and accuracy rate of human fall behavior recognition algorithm in current video surveillance, the human behavior feature extraction module is relatively backward. To solve this problem, a fall behavior recognition algorithm based on feature and deep learning is designed. The video image preprocessing is completed by dilation and erosion. The covariance matrix of image features is constructed to extract the features of human fall behavior. The standard image database is constructed, and the deep learning algorithm and neural network are used to complete the human fall behavior recognition in video surveillance. So far, the design of human fall behavior recognition algorithm in video surveillance based on feature and deep learning is completed. Through the example test, the application effect of the feature algorithm is better than that of the traditional algorithm.

Keywords: Deep learning · Feature learning · Behavior recognition · Intelligent video

1 Introduction

With the rapid development of economy at home and abroad, people's life has undergone earth shaking changes, and the process of urbanization has accelerated significantly. However, a series of problems also follow. With the rapid growth of urban population, the number and proportion of left behind elderly and left behind children are increasing, and the health and safety problems of the elderly are becoming more and more prominent. Accidents or deaths occur in the elderly due to falls every year. In addition, with the increase of urban population, the urban environment is becoming more and more complex, and sudden violent incidents such as theft, mob fighting and terrorist attacks often occur, which has brought serious interference to social public order and brought great losses to the people. Governments of all countries pay more attention to the prevention of public security. As an important prevention tool, video surveillance has been widely used in public security, transportation, banking and other important departments [1]. At present, intelligent monitoring is mostly used in the family, through monitoring the activities of children and the elderly to carry out management, to avoid

the safety problems caused by physical discomfort falls. Through literature research, it can be seen that the analysis of abnormal behavior of moving human body is based on intelligent video monitoring, using computer vision technology, taking networking and intelligence as the development direction, to analyze and identify the human target behavior in the video sequence, so as to deal with and alarm the abnormal behavior in time. Greatly improve the ability of video surveillance, and save human and material resources in the monitoring work, at the same time, intelligent analysis can reduce the probability of missing and false positives [2].

At present, there are a lot of problems in human fall behavior recognition in video surveillance, which lead to abnormal recognition results and can not achieve the expected effect. In view of the drawbacks of human fall behavior recognition in traditional video surveillance, it is necessary to realize intelligent video surveillance to meet the needs of society and realize computer simulation of human brain. Realize the automatic analysis and understanding of the video monitoring screen, and timely alarm for the abnormal events, so as to facilitate the staff to deal with them in time, so as to minimize the loss [3]. In recent years, many experts and scholars have done in-depth research on the analysis and understanding of video information, and intelligent video monitoring technology has been further developed. Video understanding is to use computer to process video data and obtain effective information. There are two key elements of people and objects in video, and people as one of the objects of concern makes the behavior recognition of people in video surveillance become the focus of many application fields [4]. Abnormal behavior recognition is one of the specific subclasses, which is directly related to people's interests, so it is of greater significance for the study of human abnormal behavior. Therefore, in this study, feature learning and deep learning are applied to the human fall behavior recognition algorithm in traditional video surveillance, and the human fall behavior recognition algorithm in video surveillance based on feature and deep learning is designed. In order to obtain better recognition results based on the traditional recognition algorithm, and provide technical support for the development of video surveillance.

2 Design of Human Fall Behavior Recognition Algorithm in Video Surveillance

Before the beginning of this study, the problems existing in the use of traditional fall behavior recognition algorithm in video surveillance are analyzed. The deep learning algorithm and feature learning algorithm are applied to this study. The process of human fall behavior recognition in optimized video surveillance is shown in Fig. 1.

According to Fig. 1, the research content is set as follows:

This paper first introduces the research background and significance of this topic, and then introduces the research status in this field at home and abroad. And the existing methods are analyzed, and the research content and organization structure of this paper are introduced.

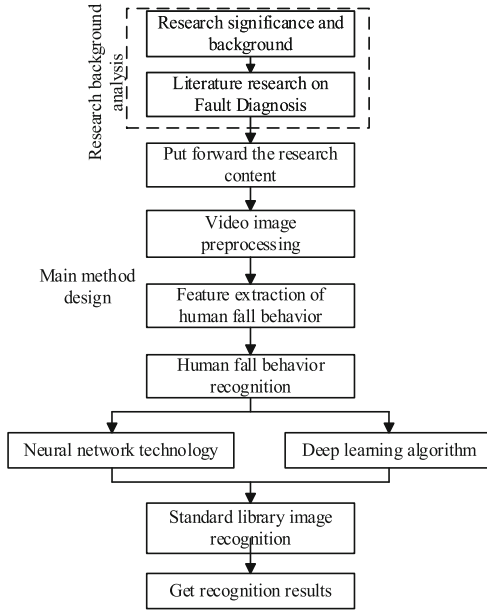


Fig. 1. Construction process of human fall behavior recognition algorithm in video surveillance

Related theory and technology. Firstly, the human behavior recognition algorithm is introduced to lay a theoretical foundation for the following research work. After that, it introduces the theory of deep learning and feature learning in detail, and then introduces the commonly used classification algorithms and deep learning tools.

Human behavior recognition. The calculation of characteristic covariance matrix is studied, and the covariance matrix is transformed into log Euclidean space vector. The bag of words model is improved for human behavior modeling, and the traditional framework is replaced by spectral clustering dictionary learning and local constraint linear coding. The spatial pyramid model is combined to pool features, and nonlinear support vector machine is used to classify human behavior.

Experimental demonstration and analysis. The corresponding example is set to verify the algorithm proposed in this paper, and this method is compared with the traditional method to determine the difference of the use effect of different methods.

Summary and prospect. The paper summarizes the work done in this paper, analyzes the shortcomings of the current algorithm, and puts forward the corresponding improvement direction.

2.1 Video Image Preprocessing

Digital image processing is the use of digital processing technology, image processing methods and technologies, such as noise removal, image enhancement, image restoration, image segmentation and so on [5]. In today’s society, with the rapid development of information technology, the computer processing ability has been greatly improved, and the development of mathematics is also very rapid. The application demand of people in various fields of production and life is also growing. As the most intuitive way

to obtain and express information, image is still the focus of people's research. These factors promote the rapid development of digital image processing technology. In this study, we will mainly use morphology to preprocess the human body image in the video. According to the research goal, we will use expansion operation and corrosion operation to complete the preprocessing process.

Dilation operation is a processing method that combines the background pixels connected with the known target region into the target region by using structural elements, so that the boundary of the target region expands to the outside. The algorithm can fill the holes in the target area and eliminate the noise in the target area. Set the monitoring image set as X and the structural element as Y . mathematical morphology image processing uses Y to process X . If Y is used to dilate X , denoted as $X \oplus Y$ and \oplus as dilation operator, then the formula of dilation operation (1) is as follows:

$$X \oplus Y = \{a|(Y)_{ab} \cap A \neq \emptyset\} \quad (1)$$

The formula (1) shows that the expansion process is to first map the structural element Y itself to the origin, and then translate a . Using Y to expand X is the sum operation of all Y and X after translation a . the operation result is not an empty set, that is, there is at least one non-zero common element. According to this explanation, formula (1) can also be written in the form of formula (2)

$$X \oplus Y = \{|a|(\bar{Y}) \cap (\bar{X}) \neq \emptyset\} \quad (2)$$

Using formula (2), the expansion operation of video image is completed. Then, the image erosion operation is completed. The function of erosion operation is just opposite to that of expansion operation. It is a processing method that uses structural elements to eliminate some boundary points of the target area and make the boundary of the image target area shrink inward. This process can eliminate small and meaningless isolated noise points and segment two small connected objects. If Y is used to corrode X , denoted as $X \cdot$, and \cdot is denoted as corrosion operator, then the formula of corrosion operation is as follows:

$$X \cdot Y = \{a|(Y)_a \subseteq X\} \quad (3)$$

Formula (3) shows that the result of corrosion process can be interpreted as all the points a contained in X after the structural element Y is translated into a . After image preprocessing, the acquired image is enhanced. In this study, the representative smoothing filtering method of spatial domain enhancement method is mainly used [6]. The basic idea of smoothing filtering method is to weaken or eliminate the high-frequency component of the image, enhance the low-frequency component of the image, so as to remove the random noise in the image and smooth the image. According to the research results, the neighborhood average filtering method is selected to complete this part of processing.

Neighborhood average filtering method is also called mean filtering method, and its implementation process is: first, average a pixel and all pixels in its neighborhood. Then the average value is used to replace the value of the corresponding pixel in the output image to achieve smooth filtering. Set the original image to the 5×5 template type.

Using this model, the calculation process is completed. The specific formula (4) is as follows:

$$s(x, y) = \frac{1}{N} \sum_{k \in D} c(i, j) \tag{4}$$

In formula (4), it is assumed that D is the set of neighborhood pixels with (x, y) as the center, and N is the total number of pixels in the neighborhood. Each pixel is calculated according to formula (4). And the obtained gray value is assigned to the corresponding pixel to enhance the image. The above processing content is integrated to complete the video image preprocessing process.

2.2 Feature Extraction of Human Fall Behavior

In digital image processing, the covariance matrix can represent the image regional features, such as the regional covariance matrix [7]. It can also represent video sequences, such as spatiotemporal 3D feature covariance matrix. This paper mainly studies the application of spatiotemporal 3D feature covariance matrix in human behavior recognition algorithm. Suppose that $A(x, y, t)$ represents a continuous video sequence or a local region in the video sequence, (x, y, t) is the horizontal, vertical and temporal coordinates of the pixel, and $M := \{(x, y, t)^T | x \in [1, O], y \in [1, R], t \in [1, T]\}$ pixel region is defined as. Where O is the region pixel width, R is the region pixel height, T is the region frame length, and $|M|$ is the total number of pixels in region M . The eigenvector of the pixel (x, y, t) is $j(x, y, t)$, let the covariance $W_M : \text{cov}(J)$, where $J := j(M)$, its eigencovariance matrix is:

$$W_M : \text{cov}(J) = \frac{1}{|M| - 1} \sum (j(x, y, t)^T - \alpha_J)(j(x, y, t)^T - \alpha_J)^T \tag{5}$$

In formula (5), $\alpha_J = \sum \frac{1}{|M| - 1} j(x, y, t)$ is the average of eigenvectors. If the $j(x, y, t)$ dimension of the eigenvector is s , then the W_M dimension of the covariance matrix is $s * s$ and is symmetric. The gradient optical flow feature is added to the covariance matrix to improve the accuracy of feature acquisition. Gradient is a kind of image sharpening operator, which is used to enhance the edge of image and highlight the mutation of image gray level. The edge information of moving human body can be enhanced by weakening the area with slow gray change. Gradient is a sharpening filter based on the first order differential of the image to complete the analysis of human behavior. According to this setting, the above setting is optimized, and the image gradient is as follows:

$$\text{grad}(j(x, y, t)) = [J_x(x, y, t), J_y(x, y, t)]^2 \tag{6}$$

In formula (6), $J_x(x, y, t)$ and $J_y(x, y, t)$ are the first-order differential of the image in the x, y direction, and the gray image can be smoothed by using vectors $[-2, 0, +2]$ and $[-2, 0, +2]^T$ as templates respectively. Then the gradient amplitude $E(J(x, y, t))$ and $\beta(E(J(x, y, t)))$ directions of the image are as follows:

$$\begin{cases} E(J(x, y, t)) = \sqrt{J_x^2(x, y, t) + J_y^2(x, y, t)} \\ \beta(E(J(x, y, t))) = \tan^{-1} \frac{J_y(x, y, t)}{J_x(x, y, t)} \end{cases} \tag{7}$$

Optical flow is the velocity distribution of the image brightness mode, which can be generated by the relative motion between the target and the observer, so it can represent the spatial distribution of the target and the change of the corresponding distribution. In the past few decades, hundreds of scholars have been studying the calculation of optical flow, such as Lucas Kanada sparse optical flow method and Horn-Schunck dense optical flow method [8]. In the research, Horn-Schunck method is used to calculate the optical flow, where $J(x, y, t)$ is the brightness of the pixel at (x, y, t) , and $(g(x, y, t), h(x, y, t))$ is the corresponding optical flow vector. Assuming that the incident light is uniform and the reflected light is smooth, the brightness of a specific pixel is fixed

$$\frac{mJ}{mt} = \frac{\chi J}{\chi x} \frac{mx}{mt} + \frac{\chi J}{\chi y} \frac{my}{mt} + \frac{\chi J}{\chi t} = 0 \quad (8)$$

Let $g = \frac{mx}{mt}$, $h = \frac{my}{mt}$, then formula (8) can be written as $J_x g + J_y h + J_t = 0$, where J_x , J_y and J_t denote partial derivatives in x , y , t directions. Using the above formula, more reliable data of human body characteristics can be obtained. Due to the diversity of human behavior recognition, the background is complex and changeable. In order to ensure good recognition accuracy, behavior features must accurately describe human behavior, try not to be affected by external factors, and maintain strong robustness. Human contour is invariant to raster, color and texture in different scenes. Therefore, the human behavior extracted from the contour must have good robustness to these changes, and can well represent the human behavior.

2.3 Human Fall Behavior Recognition

According to the above design content, in this study, the deep learning algorithm hybrid neural network technology is used to complete the human fall behavior recognition process. Character behavior may occur in different areas of the video frame. The convolution characteristic graph contains the spatial structure of the input. Therefore, this study further proposes a trajectory attention map, which can locate the moving target area in the video on the convolution feature map. Attention map is widely used to locate the target area in video. It has the same length and width as convolution feature map. Convolution feature map and attention map in each frame are corresponding to each other. At present, attention graph is trained through end-to-end network, so it needs a lot of training data. And in a very complex environment is not so accurate [9, 10].

The trajectory attention graph proposed in this study is calculated by dense trajectories and improved dense trajectories. Dense trajectories can describe people's trajectories in complex environment, so these trajectories are always distributed in the moving target area of video. In order to ensure that the recognition algorithm in this study can get high accuracy and reliability analysis results. In this study, the improved dense trajectories will be used to analyze the human motion trajectory.

Improved dense trajectories is an improved version of dense trajectories, which can overcome the influence of camera motion in video. It assumes that the relationship between two adjacent frames can be described by a transformation matrix. In order to calculate the value of the transformation matrix, it first calculates the surf feature and optical flow feature in two adjacent frames. Then the RANSAC (RANDOM SAMPLE

Consensus) algorithm is used to match these features. So the value of transformation matrix can be calculated. The first human motion track from the frame can be expressed as:

$$\{e_l^t, e_l^{t+1}, \dots, e_l^{t+n-1}\} = \{(x^t, y^t)(x^{t+1}, y^{t+1}), \dots, (x^{t+n-1}, y^{t+n-1})\} \quad (9)$$

In Eq. (9), n is the length of the trajectory, and all the improved dense trajectories in frame t can be expressed as. $\{e_1^t, e_2^t, \dots, e_l^t\}$. Based on the dense trajectory points calculated from each frame, the trajectory attention map can be further calculated. The size of the pixel value in the track attention graph is determined by counting the number of track points in the receptive field corresponding to each pixel in the track attention graph. The track attention graph of a video can be expressed as $C \in R^{HWT}$, where H and W are the height and width of the track attention graph, and T is the time length of the video. $b_{i,j}^t$ is the value of position (i, j) in the track attention graph corresponding to the t th frame. The value of $b_{i,j}^t$ is calculated as follows.

$$b_{i,j}^t = N_{e^t \in \partial_{i,j}^t}(e^t) \quad (10)$$

In Eq. (10), $\partial_{i,j}^t$ is the receptive field corresponding to $b_{i,j}^t$ pixel in the trajectory attention map. N is the statistical function, which can count the number of trajectory points in the receptive field. Then the track attention map is normalized by Eq. (11), and the normalized track attention map can be expressed as.

$$\overline{b_{i,j}^t} = \frac{b_{i,j}^t}{\sum_{i,j} b_{i,j}^t} \quad (11)$$

After using formula (11) to process the human trajectory, the trajectory is used to recognize the human fall behavior. In order to make better use of the information of human falling behavior in video, the result of fusion of two stream model U corresponds to the input of LSTM layer at a time to obtain the information of long-term fighting behavior. LSTM network controls which information is retained and which information is discarded by setting input gate, forgetting gate, output gate and storage unit. At a certain time t , the calculation part is also updated according to the LSTM principle. First, the “forgetting gate” will control whether the last time state f_{t-1} can be transferred to the next layer, and the specific updates are as follows:

$$l_i = \wp(Q_i R_i + P_i h_{n-1} + v_i) \quad (12)$$

The “input gate” can be represented by the following functions:

$$I_i = \wp(Q_j R_i + P_i h_{n-1} + v_j) \quad (13)$$

LSTM uses “forgetting gate” and “input gate” to change the state of storage unit, and uses “output gate” to control the connection between long-term operation information and current motion information, that is, how much storage unit state C is output to the current output value of LSTM. Finally, the final output of LSTM is obtained by using the “output gate” and the state of storage unit.

$$\overline{e_i} = \tan h(Q_{i1} R_i + P_c h_{n-1} + v_c) \quad (14)$$

$$\varepsilon_i = I_i \oplus \varepsilon_{i-1} + I_i \oplus \bar{\varepsilon}_i \tag{15}$$

$$z = \wp(Q_i R_i + P_c h_{n-1} + v_c) \tag{16}$$

$$h_i = z \tanh(\varepsilon_i) \tag{17}$$

where \wp is the activation function of sigmoid. R is the characteristic matrix of the input at time t . Q_i, Q_j, Q_{i1} and Q_{j1} represent the weight matrix from input layer to input gate, forgetting gate, memory cell and output gate respectively. P_i, P_c, P_{c1} and P_{i1} represent the weight matrix from hidden layer to input gate, forgetting gate, storage unit and output gate respectively. v_i, v_c, v_{i1}, v_{c1} respectively represent the offset values of input gate, forgetting gate, storage unit and output gate. According to the above calculation process, the corresponding standard behavior library is constructed, and the secondary behavior library is used to set the human fall behavior recognition process as follows (Fig. 2):

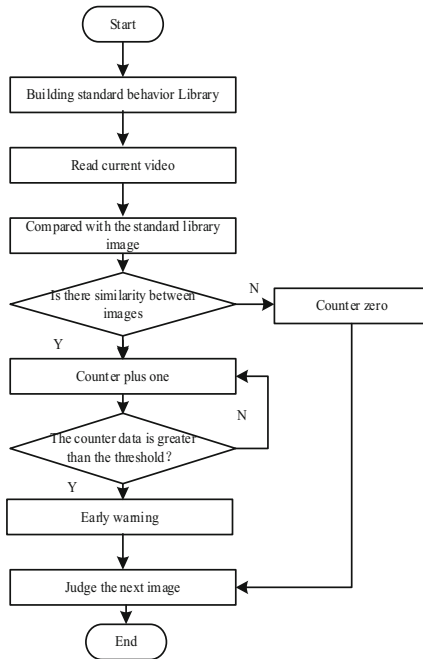


Fig. 2. Human fall behavior recognition process

This paper completes the preprocessing of the video image through the expansion operation and the erosion operation. The image feature covariance matrix was constructed to extract the fall behavior features of the human body, and then a standard image library was constructed on this basis. Use deep learning algorithms and neural networks to monitor human fall behavior. Since the number of images in the standard

library is limited, it is impossible to include all the images of standard behavior. In addition, people's fall behavior is a continuous process, so in this study, when multiple frames appear abnormal continuously. It is considered that the human body has fallen behavior, and then the alarm processing is carried out. By integrating the above settings, the design of human fall behavior recognition algorithm in video surveillance based on feature and deep learning is completed.

3 Example Analysis

3.1 Design of Test Platform

In this study, a human fall behavior recognition algorithm based on feature and deep learning is proposed to verify the effect of this method. In the test process, in order to ensure that the comparison process between this method and the traditional method 1 (reference method [2]) and traditional method 2 (reference method [3]) is not affected by external factors, the test platform is set as shown in the Table 1 shown.

Table 1. Test platform parameters

Use link	Parameter category	Parameter name	Specific model
Video image processing module	Hardware	CPU	Intel core i5 2520M
		CPU frequency	3.6 GHz
		Memory	8 GB
		Hard disk	500 GB*2
	Software	Development language	C language
		Image processing library	opencv
		Programming tools	Anaconda
Video image recognition module	Hardware	CPU	Intel core i5 2520M
		CPU frequency	3.6 GHz
		Memory	8 GB
		Hard disk	500 GB*2
	Software	Computing equipment	Python
		Operating system	Windows10

According to the content in Table 1, complete the construction of the test platform, and introduce the related content of the feature recognition algorithm and the traditional algorithm into the test platform, so as to provide a carrier for the comparative test of the two algorithms.

3.2 Test Data Set Settings

In this test, the traditional dynamic diagram is used as the experimental object. In the original dynamic image, the background area of dynamic image will appear motion blur. At the same time, the background area of the dynamic image contains a large number of vertical black edges. This is because the input video frame contains a large number of horizontal camera jitters, so the dynamic picture will mistakenly regard these background jitters as the motion of the target in the video. Thus, the motion information of the background is encoded into the dynamic image, which greatly reduces the quality of the dynamic image. Therefore, before using this part of the image, the preliminary processing is carried out, and the image types are divided according to the image types, which are used as the source of test data (Table 2).

Table 2. Test image set

Test image set serial number	Number of test images	Contains the number of human fall images
X1	100	25
X2	150	30
X3	200	50
X4	200	120
X5	150	100
X6	100	75
X7	150	26
X8	300	34
X9	250	25
X10	200	30

The above test image data set is used as the test object in this test, and the feature recognition algorithm and traditional algorithm are used to complete the recognition process, and the test results are obtained according to the recognition results.

3.3 Test Plan Setting

In this test, we will complete the application comparison process between the feature recognition algorithm and the traditional algorithm. In order to obtain the test results with high reliability, we set the test index as three parts. They are: recognition rate of human fall behavior, recognition rate of human fall behavior and recognition response time under different clarity. According to the above test indicators, the application effect of the feature recognition algorithm and the traditional algorithm is analyzed to determine the difference between the different recognition algorithms. At the same time, the

design and development direction of the follow-up human body recognition algorithm in surveillance video are controlled according to the test results.

3.4 Analysis of Test Results

Table 3. Recognition rate of human fall behavior(%)

Test image set serial number	Feature recognition algorithm	Traditional method 1	Traditional method 2
X1	90.64	85.08	85.89
X2	90.45	86.05	85.12
X3	90.83	85.38	85.92
X4	91.87	85.89	85.07
X5	90.53	85.91	85.06
X6	90.15	85.84	85.04
X7	91.61	85.23	85.17
X8	90.64	85.12	85.49
X9	90.99	85.53	85.67
X10	91.22	85.32	85.76

The results show that the recognition effect of feature recognition algorithm and traditional algorithm is different for different types of surveillance video images. The feature recognition algorithm can recognize the preset image types effectively and has high recognition accuracy. Traditional algorithms can recognize some image types in the test with high precision, but they can not be effectively recognized for some of the more complex image types. The results of the above test can be determined that the recognition rate of the feature recognition algorithm proposed in this paper is higher and the effect of application is better (Tables 3, 4).

Table 4. Recognition rate of human falling behavior(%)

Test image set serial number	Feature recognition algorithm	Traditional method 1	Traditional method 2
X1	63.62	52.87	50.52
X2	62.51	50.17	53.41
X3	61.46	51.08	54.97
X4	64.47	54.47	51.74

(continued)

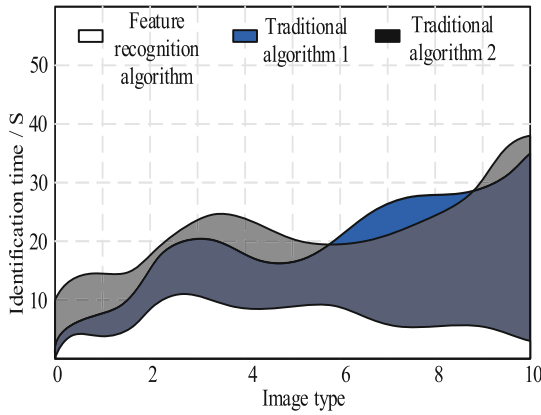
Table 4. (continued)

Test image set serial number	Feature recognition algorithm	Traditional method 1	Traditional method 2
X5	62.48	50.31	50.76
X6	63.11	54.15	50.96
X7	63.69	53.75	54.53
X8	62.77	53.58	53.56
X9	62.59	50.25	51.41
X10	64.12	52.31	50.01

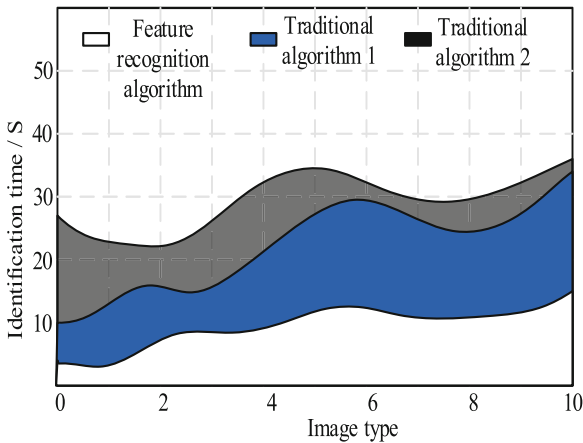
The results of recognition rate obtained by different algorithms are analyzed and the corresponding results are obtained. According to the results, we can see that the recognition rate of different recognition algorithms is different for different types of video images. The feature recognition algorithm uses deep learning algorithm in the design process, and improves the accuracy of behavior recognition to a certain extent, and obtains the recognition accuracy that traditional algorithm can not achieve. In the follow-up research, this algorithm can be used to promote the intelligent development of video monitoring. At the same time, according to the design idea of feature recognition algorithm, the content of traditional algorithm is optimized to improve the application effect of the algorithm.

After analyzing Fig. 3, it can be found that the response time of the feature recognition algorithm is more stable under different image separation rates, and there is no problem of large fluctuation. Under different image resolutions, the recognition time of traditional algorithms is different, and the fluctuation is increased. According to the analysis results, it can be determined that the recognition effect and use experience of the feature recognition algorithm are better.

Through the analysis of all the test results obtained in this test, it can be found that the application effect of the feature recognition algorithm is better than the traditional algorithm, and this algorithm can be used in the future design to obtain more effective recognition results of human falling behavior.



(a) High resolution image



(b) Low resolution image

Fig. 3. Recognition response time under different definition

4 Discuss

With the development of science and technology, people pay more and more attention to public security. As an important preventive tool, video surveillance has been gradually popularized in people's daily life. At the same time, with the application of surveillance video in social life more and more popular, the intelligent video surveillance technology is getting more and more attention. People hope to realize the intelligent video surveillance, and hope that the computer can automatically identify abnormal behaviors such as falling, fighting and so on. With the development of computer vision and deep learning technology, this idea is gradually realized through the method of deep learning. This research mainly designs a human fall behavior recognition algorithm based on feature and deep learning, and makes the following work and contributions.

- (1) This paper summarizes the main algorithms of human fall behavior recognition based on feature and deep learning. Firstly, according to the different types of input data, this paper classifies the current methods of human behavior recognition reasonably, reviews the development history and research status of human behavior recognition based on color video, discusses the shortcomings and defects of current human behavior recognition algorithms, and puts forward many corresponding improvement measures.
- (2) A hybrid model based on feature and deep learning algorithm is proposed to recognize the fall behavior. After simple processing of the video data set, the hybrid model uses the sliding window and weight sharing of convolutional neural network to obtain the local features of the video sequence, which is used as the input of the next layer, and uses the timing to obtain the temporal features of the video data. They make full use of their respective advantages. In addition, the deep learning can automatically extract behavior features, which avoids the process of manual feature extraction. The paper analyzes the fall data of this paper by using traditional pattern recognition method, convolutional neural network, long-term memory network and this model respectively, and verifies that the method is feasible and effective in this paper based on the data set provided in this paper.

There are still some deficiencies in this study, which need to be improved.

- (1) For the study of fall behavior, the data set used in the experiment is fixed and single person behavior. In addition, for the research of fall behavior, we use the single person behavior, without considering the group behavior, which has deviation from the actual situation. In the future, we should conduct in-depth research on fall behavior which is closer to the actual scene.
- (2) In the study of behavior in video surveillance, when choosing neural network model, we can consider using deep network model such as Google net, VGg to study.

5 Conclusion

Now, with the development of information technology, video surveillance is more and more common, intelligent video surveillance has been more and more widely concerned, and the analysis of abnormal human behavior based on intelligent video surveillance has achieved certain results. But the moving human body is non rigid, has a high degree of freedom, and the scene of video surveillance is complex and changeable, so the research on human abnormal behavior is still in the analysis of some simple human behavior. Although the human fall behavior recognition method designed in this paper meets the needs of customers, its function is relatively single, and the reliability, scalability and security of the algorithm have not yet reached the standard, and the efficiency in the process of operation is not particularly high. Therefore, in the follow-up work, we hope to develop a more robust platform to make the proposed algorithm more competitive in the current intelligent monitoring platform.

Fund Projects. 1. The 2019 Municipal Education Commission Science and Technology Research Program Project, Project Name: Research on Attitude Adaptive Static Image Human Behavior Recognition (Project Number: KJQN201904601)

2. School level project of Chongqing Vocational Institute of Tourism in 2020: Project Name: Research On Human Action Recognition Algorithm Based On Deep Learning (Project Number: XJKT-2020-12)

References

1. She, X., Su, X.: Research on human fall behavior using CNN and LSTM-based hybrid model. *Appl. Res. Comput.* **36**(12):3857–3859+3868 (2019)
2. Ding, Y., Pu, H., Zhang, J., et al.: Implementation of behavior recognition based on machine vision. In: 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS). IEEE, pp. 1–5 (2019)
3. Chen, Y., Hu, R., Li, S., et al.: Recognition algorithm for human behavior in video based on combined features and SVM. *J. Shenyang Univ. Technol.* **42**(06), 665–669 (2020)
4. Kun, H., Zefan, H.: Falling behavior recognition method based on dynamic characteristics of human body posture. *J. Hunan Univ. (Nat. Sci.)* **47**(12), 69–76 (2020)
5. Huang, Y., Wan, C., Feng, H.: Multi-feature fusion human behavior recognition algorithm based on convolutional neural network and long short term memory neural network. *Laser Optoelectron. Progress* **56**(07), 243–249 (2019)
6. Liu, Y., Zhang, Y., Wang, C., et al.: Behavior recognition algorithm based on RGB-D and deep learning. *Comput. Eng. Design* **40**(06), 1747–1750+1762 (2019)
7. Ma, L., Pei, W., Zhu, Y., et al.: Fall action recognition based on deep learning. *Comput. Sci.* **46**(09), 106–112 (2019)
8. Fang, C., Jia, K., Liu, P.: Identification of taxi violation behavior based on surveillance video. *Comput. Simulat.* **37**(05), 326–331 (2020)
9. Zhou, Q., Xing, J., Yang, Q., et al.: Sequential image deep learning-based Wi-Fi human activity recognition method. *J. Communicat.* **41**(08), 43–54 (2020)
10. Nie, H., Xiong, X., Guo, Y., et al.: Video abnormal behavior identifying algorithm based on deep learning. *Modern Electron. Techn.* **43**(24), 110–112+116 (2020)
11. Liu, S., Liu, X., Wang, S., et al.: Fuzzy-aided solution for out-of-view challenge in visual tracking under IoT-assisted complex environment. *Neural Comput. Appl.* **33**(4), 1055–1065 (2021)
12. Liu, S., Liu, D., Srivastava, G., et al.: Overview and methods of correlation filter algorithms in object tracking. *Complex Intell. Syst.* 1–23 (2020)
13. Liu, S., Liu, D., Muhammad, K., et al.: Effective template update mechanism in visual tracking with background clutter. *Neurocomputing* (2020). <https://doi.org/10.1016/j.neucom.2019.12.143>