Research on Video-Based Human Action Behavior Recognition Algorithms

Haifei Si¹,²,a, Xingliu Hu¹,* and Yizhi Wang¹,b

¹College of Intelligent Science and Control Engineering, Jinling Institute of Technology, Nanjing, Jiangsu, 211169, China
²College of Automation, Harbin Engineering University, Harbin, Heilongjiang, 150001, China

*Corresponding author e-mail: jimmy080@jit.edu.cn, a sihaif@jit.edu.cn, b w_yz@jit.edu.cn

Abstract. Human action recognition is an important part of intelligent video processing, which has high research value and broad application prospects. This paper takes the video monitoring of the elderly living alone as an example to recognize human behavior, and chooses the video of indoor background taken by a single camera as the research object. The background subtraction method is used to extract moving objects from video, and the extracted moving objects are pretreated. Then, according to the overall outline of the object or the obvious features of a part of the object, the motion features are obtained to understand and identify the movement of the monitored object, such as walking and falling. Experiments show that the proposed algorithm has better processing effect for actual video.

1. Introduction

In recent years, the trend of population aging in China has been aggravating, and the daily guardianship of the elderly living alone has gradually become a hot social concern. Through intelligent video surveillance of the elderly, family members can find the elderly fall in the best rescue time and take first aid measures, which can reduce the injury of falls to the elderly [1]. There are many ways to detect falls of elderly people. An acceleration sensor is added to the wearing device of the elderly to calculate the acceleration in three directions perpendicular to the space [2, 3]. Compared with the normal threshold set in three directions beforehand, if it exceeds the threshold set, it indicates that there is a fall phenomenon. Using the gyroscope sensor detection angle, judge whether it is more than 70 degrees, if it is larger, there may be fall phenomenon [4]. Mathie and Estudillo-Valderrama use waist accelerator and biomedical sensor to detect human body fall and collect data. When human body has abnormal movement, the movement change will be great, which will make the acceleration of a part of the human body increase [5]. Using the observation platform to monitor the trajectory of head movement, the acceleration data detected are compared with the preliminary threshold, combined with the sound detection sensor detection and the infrared sensor installed on the wall to detect the human body posture and speed, so as to realize the fall detection of the elderly [6, 7]. Hsu et al. used variable triangles to divide human body shape into triangular grids, and proposed two important posture features: skeleton and center. Then, the corresponding matching algorithm was used to recognize
human body posture [8, 9]. In this paper, action recognition is realized at the algorithm level. The program can read the video sequence and quickly detect the moving objects and their motion features in the video.

2. Human Behavior Feature Extraction

How to accurately extract the motion target is the most important problem in human behavior recognition. There are several difficulties in extracting the human motion feature, such as the variety of action categories, the change of environmental background, and the change of time.

2.1. The Minimum External Rectangular Width-Length Ratio

In order to reduce the complexity and computational complexity, and reduce the spatial dimension, we can extract the moving object, determine the human body’s main axis, and then cut out the minimum outer rectangle from the human body’s main axis and the direction perpendicular to the main axis. It is found that the ratio of width to length of the smallest outer rectangle varies with the change of movement, and it has certain regularity. The aspect ratio is the ratio of the width to the length of the outer box of the human body. As shown in Fig. 1.

![Fig. 1 Width-Length ratio of external rectangle when one walks.](image)

By observing Fig. 1, it is found that each organ is limited by muscles and bones during walking, and will not produce a large range, but the aspect ratio shows a certain regularity. This figure shows the aspect ratio. When walking normally, the human body’s width is smaller than the length so that the aspect ratio is less than 1.

2.2. Detection of center of gravity

In the feature extraction method based on the external contour of human body, the center of gravity can represent the posture of human body. When human is standing, the center of gravity is the highest, while when falling, the center of gravity is close to the ground. The calculation of the center of gravity of the human body’s external contour can be done by using the coordinate points of the pixels on the external boundary [10, 11]. The specific steps are as follows:

(1) Find a point on the boundary line of human body. Because the image is grayed, the gray value and density of the pixels in the boundary of human body are the same, and then take two points along the boundary line, $P_0$, $P_1$; the area enclosed by the boundary line is $S_1$, and the line segment is obtained
by calculating \( n \) according to formula (1). \( P_0P_n \) the line \( L_1 \) where the line segment is located is deduced.

(2) Another difference \( X_0, X_n \), according to the method of the first step, the line \( L_2 \) is obtained.

(3) Find the intersection point of two straight lines, namely the center of gravity.

In this paper, the elderly fall recognition algorithm takes the center of gravity of human body in different situations as the detection feature. Through several experiments, it is found that the center of gravity of human body is the highest when walking normally, which belongs to the normal walking situation, while the center of gravity of human body is the lowest when falling, which belongs to the abnormal situation.

2.3. Change rate of human behavior and movement

The change rate of human action is the change rate of the aspect ratio of the rectangle of human boundary detected between adjacent frames. If the aspect ratio of two adjacent frames varies greatly, it shows that the human body produces actions different from the current behavior, at this time, it is necessary to study the changes. In this topic, the study of fall phenomenon can be combined with the change of the center of gravity at this time. If the acceleration of human body posture change is larger, it may be a fall phenomenon. If the change is slow, it may be a bending behavior\[^{12,13}\].

2.4. Skin color region detection

Recognition of human features can also use the inherent skin color aggregation characteristics of the human body, using skin color detection method, so the skin color region detection of image sequence is mainly to detect the exposed parts. Research shows that in the process of video image processing, human skin color has a certain clustering in \( YCrCb \) color space, usually concentrated in the following areas:

\[
\{ P | 133 < Cb(P) < 173,77 < Cr(P) < 117 \} \tag{1}
\]

\( YCrCb \) color space was originally used for transmission of color video signals. Compared with traditional \( RGB \) color space, its main feature is that its color component group is guaranteed by red component \( Cr \) and blue component \( Cb \), and its brightness component \( Y \) is independent. This enables the \( Y \) component of the image to independently represent the gray information of the image, which is equivalent to the traditional and black-and-white television picture quality. On the other hand, if there is no \( Y \) component, the imaging effect of the same kind of target at different brightness can be eliminated. It is for this reason that the skin color which eliminates the brightness information gets a good aggregation in \( YCrCb \) and is often used in skin color detection. On the basis of previous studies, we know that we can use the following formula (2) to transform the traditional \( RGB \) image into the \( YCrCb \) image we need:

\[
\begin{align*}
Y &= 0.299R + 0.587G + 0.114B \\
Cb &= 0.5 - 0.4187G - 0.0813 + 128 \\
Cr &= -0.1687R - 0.3313G + 0.58B + 128 
\end{align*}
\tag{2}
\]

A simple skin color detection method is to filter the image directly by pixels according to the clustering of skin color, leaving the pixels within the skin color range. Then the connected region is obtained from the result of binarization, and the skin color region conforming to the size specification is left. Finally, the skin color region conforming to the target rule is determined according to the other detection rules of the target.

In this paper, human behavior is identified based on human shape features, including the change of width-length ratio of rectangle, the extraction of center of gravity, the change rate of human posture and so on \[^{14,15}\].
3. Experiments and Analysis of Human Behavior Recognition
The experiment of this paper is to detect whether the old people have fallen or not. In this paper, data sets are made in different environmental scenarios to detect whether the program application can correctly detect human behavior in these environmental scenarios. In the experiment, human behavior was captured from the corridor, and these video sets were subdivided into frames. First, fall video was read, then the size of the video was read, and a for loop was made. The video was divided into several image sequences, and each image frame was named in sequence. The flow chart of the program is shown in Fig. 2.

![Program flow chart](image)

**Fig. 2** Program flow chart.

3.1. Detection results of moving targets
Through background subtraction method, the experiment carried out target detection for human walking and falling, as shown in Fig. 3 and Fig. 4.

![Images](image)

**Fig. 3** Human walking chart on corridor.
Fig. 3 and Fig. 4 show that there are differences between the 76th frame image and the background image when a person walks, so background subtraction is used to cut out and save the differences, and the same is true when a person falls down. The pixels without difference are assigned 0, and the pixels with difference are assigned 1, so the white pixel blocks are obtained and the binary image of the moving object is obtained.

Firstly, gray image sequence is processed to simplify calculation, and then the number of frames of moving objects is recycled to find out whether each frame deviates from the pre-saved background. As long as there is deviation and the deviation is greater than the threshold of 25, the subtracted background is saved and displayed in the window.

3.2. Morphological treatment results
From the two pictures above, we can see that although the differential images of human walking and falling are obtained by background subtraction, there are still some individual white pixels around the human body, and there are some filling holes inside the human body. We use the expansion mathematical morphology operation to fill the holes, and use corrosion to eliminate the surrounding pixels. The resulting effects are shown in Fig. 5 and Fig. 6.
As can be seen from Fig. 5, the white pixels between human legs are removed by corrosive algorithm, and the holes in human back are filled by expansion algorithm to get the image of before edges. From this image, the outline of human body can be seen more clearly. Fig. 6 shows that when human falls, the shadow pixels on the ground are human, and most of the shadow pixels are corroded by expansion. Elimination, this operation is conducive to the following edge detection.

3.3. Experimental results of edge detection and external rectangle
We use the edge detection method mentioned above to extract the edge of the binary image of human walking and falling. On the basis of edge extraction, we get the outer rectangle of human body and find the change of aspect ratio. The resulting effects are shown in Fig. 7 and Fig. 8 (a), and in Fig. 8 (b).

3.4. Detection of center of gravity
The change of human’s center of gravity when walking can be seen from the figure that the position of human’s center of gravity is basically unchanged. We use the spatial coordinate axis to represent the position of the center of gravity pixels, which are time axis, horizontal axis and vertical axis respectively. By detecting the largest white pixel blocks, we can get the center of gravity, and mark the pixels of each frame. We can see that the center of gravity pixels are in a certain range when walking. There is no big deviation in internal fluctuation. Graph of the center of gravity for walking and falling is displayed, as shown in Fig. 9 and Fig. 10.
In order to detect the center of gravity when falling more clearly, we use two-dimensional coordinates to express it. As shown in Fig. 10, we can see that there is downward fluctuation between 55 and 65 frames, indicating that the center of gravity is low and there is a fall phenomenon. From the video, we can see that there is a fall action between 50 and 70 frames, which verifies the fall detection effect.

4. Conclusion
The purpose of detecting the falls of the elderly is to recognize the human action in the video. At present, whether it is through sensors or using video means to detect the falls of the elderly, each has its advantages and disadvantages. Observing the daily life of the elderly through video may also involve personal privacy issues. Some people are unwilling to adopt video observation method, while the method studied in this paper only extracts the falls. The outline information of the human body, which does not involve its specific privacy information, will be a bright spot for future promotion and application. This paper establishes the algorithm model, optimizes the model further according to the actual work, and implants the robust program into the embedded hardware platform, using the
corresponding hardware advantages, to achieve real-time response and processing, and improve the health and safety of the elderly.

Acknowledgments
This work was financially supported by the Jiangsu Province’s Natural Science Foundation under Grant BK20171114, Youth Academic Leader in Qinglan Project of Jiangsu University in 2017, A joint prospective industry-University-Research Collaboration project of Jiangsu Province under Grant BY2016012-02, National Science Foundation of the Higher Education Institutions of Jiangsu Province under Grant 15KJB520010, Ministry of Education’s Cooperative Education Project under Grant 201701056001.20180112066, Jinling Institute of Technology’s Education Reform Project under Grant 2017JYJG03, Jinling Institute of Technology Talent Introduction Project under Grant Jit-rcyj-201604.

References
[1] Zhou, M. Research on human fall detection algorithm based on computer vision. Shandong University, Jinan, 2013.
[2] Cai, Y., Shang, Y., Tan, Y., Tang, Z., & Zhao, B. Human Action Recognition Based on Deep Learning. In International Conference on Applications and Techniques in Cyber Security and Intelligence, Springer, Cham. 2019, pp. 1595-1600.
[3] Kulbacki, M., Segen, J., Wojciechowski, S., Wereszczynski, K., Nowacki, J. P., Drabik, A., & Wojciechowski, K. Intelligent video monitoring system with the functionality of online recognition of people’s behavior and interactions between people. In Asian Conference on Intelligent Information and Database Systems, Springer, Cham. 2018, pp. 492-501.
[4] Shi, X., Xiong, Q., Lei, L. Research on Fall Detection System Based on Pressure Sensor. Journal of Instruments and Instruments, 2010, 31 (2): 66-70.
[5] Dai, C., Liu, X., Lai, J., Li, P., & Chao, H. C. Human Behavior Deep Recognition Architecture for Smart City Applications in the 5G Environment. IEEE Network. 2019.
[6] Li, D., Liang, S. Design of an elderly fall detection device based on acceleration sensor. Sensors and Microsystems, 2008, 27 (9): 85-88.
[7] B.U. Toreyin, E.B. Soyer, I. Onaran, A.E. Cetin. Falling person detection using multi-sensor signal processing. Signal Processing and Communications Applications, SIU 2007. IEEE 15th. 2007, pp. 1-4.
[8] Mabrouk, A. B., & Zagrouba, E. Abnormal behavior recognition for intelligent video surveillance systems: A review. Expert Systems with Applications, 2018, 91, 480-491.
[9] Zhou, Q. Design of Human Motion Detection System and Research on Fall Prediction Method. East China Normal University. 2014.
[10] Zhao, Y., Tian, G., Yin, J., et al. Human trajectory analysis method based on HMM in family intelligence space. Journal of Pattern Recognition and Artificial Intelligence, 2015, 28 (6): 542-549.
[11] Xu, H., Li, L., Fang, M., & Zhang, F. Movement Human Actions Recognition Based on Machine Learning. International Journal of Online Engineering, 2018, 14(4):193-210.
[12] Yang, S., Yu, Z., Su, B., et al. The method of elderly behavior recognition based on trajectory segmentation. Journal of Qingdao University: Natural Science Edition, 2017, 30 (1): 103-107.
[13] Kong, L., Li, H., et al. Detection of Abnormal Behavior of Elderly People Based on Location and Motion Trajectory. Journal of Computer Engineering and Design, 2012, 33 (2): 735-739.
[14] Tian, G., Yin, J., Yan, Y., et al. Trajectory analysis behavior recognition method based on Mixture Gauss model and principal component analysis. Journal of Electronics, 2016, 44 (1): 143-149.
[15] Ma, Y. Design and implementation of office health analysis system based on Kinect. Harbin, Harbin University of Technology, 2015.