Abnormal Attitude Recognition of the Elderly Based on Mobile Devices

Xuri Kou [1, b] and Chengjun Xie [2, a, *]

[1] School of Computer Science and Technolgy, Beihua University, Jilin 132000, Jilin, China

[2] School of Computer Science and Technolgy, Beihua University, Jilin 132000 Jilin, China

[a]Email: abbey1998@sina.com, [b]Email: kouxuri@sina.com

[*]Corresponding author

Keywords: Mobile Devices; The Elderly; Abnormal Attitude; Fusion Feature

Abstract: With the aggravation of aging in China, people pay more and more attention to the safety of the elderly. The combination of the use of mobile portable devices and the study of abnormal attitude identification provides a new idea for people to monitor the life safety of the elderly. This paper designs an abnormal attitude recognition system that uses multi-feature fusion algorithm to extract features based on mobile devices, and compares it with other single-feature recognition algorithms. The advantages and disadvantages of this algorithm are comprehensively evaluated through the comparison of recognition accuracy and recognition efficiency. The result of research shows that the accuracy of the multi-feature fusion was 2.60% higher than the SurfFeatures, which is the best single feature; It is 0.35% higher than Surf&HuFeatures, which is the best dual features. The multi-feature is more accurate in the expression of information, has the highest recognition rate in the experiment, and is more reliable for the protection of the elderly. This paper studies the abnormal attitude recognition of the elderly based on mobile devices, which has important guiding significance for the establishment of human feature extraction and mobile device combination system.

1. Introduction

With the coming of the aging wave, our society will face a considerable burden of providing for the aged. The safety of the elderly is one of the most critical issues. With the popularization of China's family planning policy and the migration of population, the living pattern of the Chinese population is transforming to a small family structure, leading to a large number of older people living alone. Due to the degeneration of the physiological functions of the elderly, there are often some accidents, especially the elderly who live alone. This situation makes the health care for the elderly living alone become a social problem that can not be ignored. In view of this kind of situation, people have done a lot of related researches in recent years, and the most concerned is human behavior recognition based on the video system. The University of Minnesota detected abnormal human behaviors in various scenarios and made early diagnosis of human neuromotor dysfunction[1, 2]. The computer vision research group at the University of Florida in the United States developed the human behavior recognition software Keys, which is based on the feature tree recognition framework and mainly applied to large video databases. At present, the ideal recognition effect has been achieved[3]. The Behave project developed by Edinburgh University in the UK mainly studies how to extract fragments containing abnormal human behaviors from video sequences, which has been applied to the intelligent security monitoring system[4]. Lee et al. used the aspect ratio of the smallest external rectangle of the human body combined with spatial and temporal information, and the continuous multi-frame aspect ratio information is combined to analyze human behavior[5]. Foroughi et al. abstracted each part of the human body into an ellipse and described the movement of the human body through the horizontal and vertical width information of each ellipse. At the same time, they trained the neural network to classify and recognize the extracted feature matrix, to complete the recognition of human behavior[6]. Rougier et al. tracked the 3D position of the human head firstly and recorded the horizontal and vertical

velocity of the human head movement. When the vertical velocity suddenly increased and the horizontal velocity remained unchanged or grew slowly, the human body fell[7-9]. The main difficulty of human behavior recognition based on video system is the extraction of useful features. Finding accurate and efficient feature extraction algorithms is an important research direction.

In this paper, modern portable mobile devices are used as carriers to monitor the abnormal behaviors of the elderly through the multi-feature fusion extraction algorithm, and to alert the guardian in the first time, significantly reducing the risk of death of the elderly. Firstly, this paper introduces the relevant contents of mobile devices and analyzes three common used human body feature extraction algorithms. We propose an extraction algorithm based on multi-feature fusion. The innovative algorithm is comprehensively analyzed through two indicators of recognition accuracy and recognition efficiency. Experimental results show that the algorithm has higher accuracy and efficiency. This paper is a new attempt to extract human features based on mobile devices, which is of considerable significance to the computer visualization industry and the security of the elderly.

2. Advantages of Mobile Devices

With the further integration of IT and communications technologies, the market for smartphones and tablets has boomed[10]. In particular, the functions of smart terminal devices based on iOS, Android and other operating systems are becoming more and more powerful. The mobile Internet industry has developed rapidly and is gradually penetrating into every field of people's life and work. A growing number of technologies have or are about to break free of the constraints of time and space and use smartphones and other mobile devices to perform everyday tasks. Especially with the development of cloud computing, the cloud server is used for auxiliary computing, which dramatically improves the speed of computing and transmission[11]. The mobile device-based processing strategy is shown in figure 1.

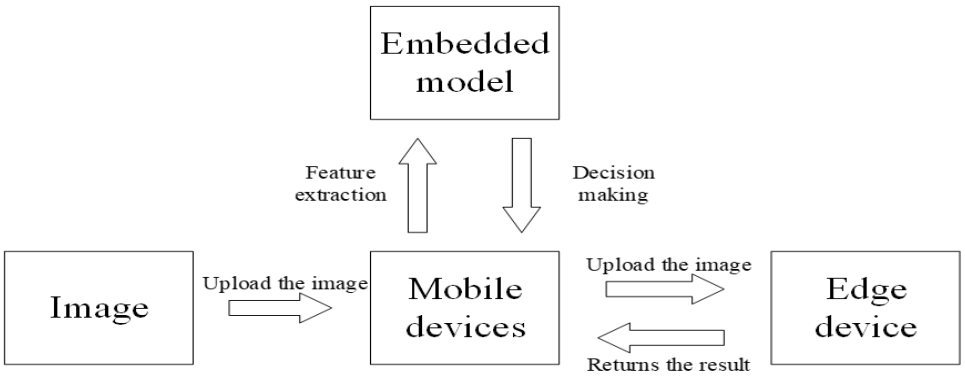

Figure 1. Image processing strategy based on mobile devices

Mobile devices have the following advantages:

(1) Convenience. No matter when and where, the process executor can interact with the system through the mobile terminal, while the traditional system executor is almost confined to the office. Convenience and speed are arguably the most significant changes from the pc-centric past (people had to go to some places where have computers) to the mobile Internet era (the people are the center, everyone has computers).

(2) Flexibility. Internet access is no longer the exclusive right of the connected computer, and people have more choices, wireless Internet access computers, mobile phones, PDAs and so on.

(3) Security. With the continuous development of information processing technology, data security, encryption, identity authentication and other aspects of technology has been quite mature. Some of today's PDAs and smartphones have considerable storage and computing power, and some have the ability to run JAVA programs. Most of the security and encryption technologies that were

implemented on the PC in the past are also available on mobile phones. It is more secure for mobile phone users because the SIM card itself already stores a lot of personal information and is unique.

(4) Interactivity. Mobile devices can not only receive messages but also send them out, enabling point-to-point communication.

(5) Low cost. With the maturity of technology, a variety of mobile devices, mobile Internet is no longer high prices, but gradually toward the civilian, popular. SMS, in particular, is very cheap.

3. Multi-feature Fusion's Extraction of Abnormal Behavior Features

3.1. Silhouette Features of Human Body

3.1.1. Human Height-Width Ratio

According to Vinary's research on human falling, the height-width ratio of the human body can effectively describe the motion characteristics of human falling[12]. In order to simplify the calculation, the ratio of height H and width W of the minimum external rectangle of the silhouette of the human body is defined as the human height-width ratio, which can be expressed by the following formula:

$$ratio = \frac{H}{W} \tag{1}$$

H is the height of the minimum enclosing rectangle, and W is the width of the minimum enclosing rectangle.

When a fall occurs, the height and width of the body change dramatically. However, due to the difference in fatness and thinness of the human body, as well as the distance and angle of the camera equipment from the human body and other factors, we cannot accurately collect the height and width information of the human body. At this point, the ratio of height to width information becomes an excellent descriptive operator, which can well reflect the state of human motion. The specific human height-width ratio is shown in table 1.

Table 1 Human behavior and height-width ratio

Behavior	Walk	Run	Sit	Fall
Ratio	1.8 ~ 3.9	1.9 ~ 3.7	2.3 ~ 2.4	0.33 ~ 0.62

As can be seen from the above table, when the human body is in the state of falling, the ratio of height to width is relatively small compared with other situations, which is easily recognized by the machine. However, a single height-width ratio is not rigorous enough.

3.1.2. Central Change Rate

Although the height-width ratio of the human body has been used as an essential judgment basis, considering that in the video surveillance system, the height and width information is greatly affected by the position and angle of the camera, the concept of the center change rate is introduced[13]. It represents the ratio of the vertical displacement and the horizontal displacement of the human body in the adjacent two frames and can be regarded as the slope of the line corresponding to the position of the center point of the adjacent two frames from a geometric perspective. The specific definition is as follows:

$$K = \frac{y_n - y_{n-1}}{|x_n - x_{n-1}|} \tag{2}$$

(x_n, y_n) and (x_{n-1}, y_{n-1}) are the coordinates of the center point of the current frame and the previous frame respectively, and K represents the slope of the line of the center point. That is the change rate of the human body center.

3.1.3. Effective Area Ratio

Based on the above two bases, the concept of effective area ratio is introduced to reduce the false recognition rate further[14]. This is mainly for some exceptional cases, the human body height-width ratio alone will cause misjudgment. For example, during exercise, due to the extension of the arms or lower limbs, the height-width ratio of the human body at this time will decrease, which is similar to the value of falling, thus causing misjudgment. The effective area ratio E is defined as:

$$E = \frac{S}{S'} \tag{3}$$

E represents the effective area ratio of the human body; S represents the number of pixels that make up the silhouette of the human body, and S' represents the number of pixels that make up the smallest outer rectangle of the human body. It is generally considered that the effective area ratio above 0.3 can be judged as the human body in a non-special state.

The combination of human height-width ratio, center change rate and effective area ratio is helpful to identify abnormal behaviors and reduce false alarm rate correctly.

3.2. Global Features

Hu distance is a global feature descriptor proposed by Hu in 1962[15]. It is widely used in feature recognition and feature matching because of its rotational translation invariance. In this paper, Hu distance feature is extracted for a binary image containing moving human body. Hu distance is composed of 7 description features. It is defined as follows:

$$\varphi_1 = \eta_{20} + \eta_{02} \tag{4}$$

$$\varphi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{5}$$

$$\varphi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{6}$$

$$\varphi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{7}$$

$$\begin{aligned}\varphi_5 = {} & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned} \tag{8}$$

$$\begin{aligned}\varphi_6 = {} & (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ & + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})\end{aligned} \tag{9}$$

$$\begin{aligned}\varphi_7 = {} & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\ & (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned} \tag{10}$$

These seven invariant moment variables constitute a set of 1x7 dimensional feature matrices. Hu features are invariant to rotation, scaling and translation, and concisely summarize image features, which can quickly and accurately describe image contour information. However, for the pictures with rich texture information, the recognition of Hu distance is weak, so for the recognition of more sophisticated image information, in order to achieve the ideal effect, it needs to be combined with the local features of the image to describe.

3.3. Local Features

The global feature and the overall information of the image can be accurately extracted, but the local feature is far from enough. For this reason, many local feature extraction algorithms are proposed and combined with the global feature algorithm to jointly describe the image information, which significantly improves the recognition accuracy. This paper uses SURF algorithm as the description selection of local features. SURF is an improvement of SIFT algorithm[16]. The basic idea of the SIFT algorithm is to find stable feature points in scale space[17]. The specific approach

is first to construct the scale space, then find the local extremum points in the scale space, then determine the main direction of the critical points, and finally generate the descriptors of the critical points and match them.

The process of extracting SIFT feature vectors is mainly divided into the following three steps:

(1) Create scale space

Due to the different distance from the camera and other factors, the scale of real objects will change, which often causes significant difficulties in feature recognition of image processing. Therefore, the researchers proposed the concept of scale space to simulate the characteristics of images at different scales. It simulates the imaging principle of human eyes and establishes the image scale space pyramid, that is, with the distance of objects getting further and further away, the image in the image pyramid will become more and more blurred. The multi-scale space helps people to explore the essential information in the image more efficiently and extract the stable key points that can best express the information in the image. Gaussian convolution kernel is usually used to construct scale space and find stable SIFT feature points. Lindeberg et al. proved the uniqueness of the gaussian convolution kernel in scale space transformation field through experiments. The principle of scale transformation using gaussian convolution kernel is as follows: Firstly, a fuzzy template is calculated according to the gaussian distribution, and the template is convolved with the original image to achieve scale transformation. Images are mapped to different scale Spaces according to different selection scales. The distribution equation of the gaussian function is as follows:

$$G(r) = \frac{1}{\sqrt{2\pi\sigma^{2N}}} e^{-r^2/(2\sigma^2)} \tag{11}$$

σ is the standard deviation, which reflects the scale of the image. The larger the value of σ, the smoother the image. R is the fuzzy radius, representing the distance between the point on the template and the center of the template. By changing the size of σ, images of different scales can be formed to complete the construction of multi-scale space. The value of σ is proportional to the blurriness of the scale image, that is, the small scale image has higher definition and mainly reflects the high-resolution detail features in the image. The resolution of large scale image is low, which mainly reflects the general features of low resolution image.

(2) Spatial key point detection

After the scale space is built, the key points need to be located. In this process, we need to detect each point, so that it is compared with the eight points adjacent to the scale and the 18 points in the upper and lower layers, which is a total of 26 points. A point is considered to be a feature point of the image in the DOG scale if it is the maximum or minimum value among the 26 points in this layer of the DOG scale space and in the upper and lower layers.

(3) Key point descriptor generation

A. Rotate the coordinate axis. Rotate the coordinate axis to the main direction of the key point so that the rotation invariance of the feature point is maintained. The corresponding relation of coordinates of feature points before and after rotation is as follows:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \times \begin{pmatrix} x \\ y \end{pmatrix} \tag{12}$$

(x,y) and (x',y') are the coordinates of the feature points before and after rotation respectively, indicating the main direction of the key points.

B. Generate descriptors. The gradient information of all points in the 8x8 neighborhood of the key point was counted. Before the statistics of gradient information, the circular gaussian weighting function should be used to weight the neighborhood space. The closer the key points are, the larger the pixel weight will be, and the higher influence will be on the histogram of the gradient. Next, we take each 4x4 as a seed, then count the histogram of gradient of each seed along with eight different

directions, and generate a 4x8 vector as a descriptor. The process is shown in figure 2.

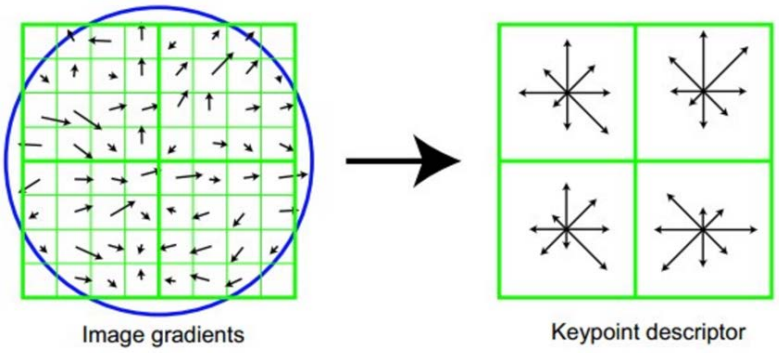

Figure 2. 32 dimensional vector feature descriptor

SURF feature point algorithm is a robust local feature point detection and description algorithm. The algorithm is an improvement of the SIFT algorithm and improves the efficiency of the algorithm. In this algorithm, the Hessian matrix is firstly constructed to generate stable edge points (mutation points) of the image and to lay a foundation for feature extraction. The process of constructing the Hessian matrix corresponds to the gaussian convolution process in the SIFT algorithm. When the discriminant of Hessian matrix obtains the maximum local value, it determines that the current point is a brighter or darker point than other points in the surrounding neighborhood, to locate the location of the key point. In the scale space construction, like SIFT, the scale space of SURF is also composed of O group L layer. The difference is that the size of the images in SIFT is half of that in the previous group, and the size of the images in the same group is the same, but the gaussian blur coefficient used gradually increases. In SURF, the image sizes of different groups are the same, except that the template size of the box filter used between different groups gradually increases. Filters of the same size are used between different layers of the same set, but the blur coefficient of the filters increases gradually.

3.4. Multi-Feature Fusion

In this paper, a total of three features are extracted from the moving human body, which is ProfileFeatures=[Ratio, K, E]. Hu invariant feature HuFeatures= $[\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7]$ represents the global feature of the image, and SURF feature point descriptor SurfFeatures represents the local feature of the image. Because in different images, even if it is the same target, after rotation and other transformations, the number of feature points detected is not the same. In this paper, about 20 feature points are extracted each time to detect the feature points of a single frame image containing a moving human body. According to the SURF feature description method, each feature point is described by a 64-dimensional description vector. Therefore, if N SURF feature points are included in a detection image, an Nx64 dimensional SURF feature descriptor will be generated. This will generate a massive amount of computation in the later behavior classification process, which is not conducive to the application in the real-time monitoring system. Moreover, due to the different number of feature points detected in each frame, it will also cause significant difficulties in the later feature classification process. Therefore, in order to ensure the computing efficiency of the system, the 64-dimensional description vector of SURF feature points detected in a single frame of the image was summed up in bits, and the statistical feature description vector of SURF feature points in a single frame of the image was obtained, Surf Features= $[s_1, s_2, \cdots, s_{64}]$. Finally, the three feature vectors are fused into one feature vector, which can be expressed as MixFeatures=[ProfileFeatures, HuFeatures, SurfFeatures].

4. Experimental Analysis

In order to accurately measure the difference between our proposed multi-feature fusion extraction algorithm and the traditional algorithm, we will use the Weizman human behavior

database for comparative testing. Among them, Weizmann's human behavior database includes 90 videos. In the scene with the camera still, nine people performed ten different actions, including bend, jack, jump, run, side, skip, walk, wave1, wave2, etc.

In this paper, three different types of Features of ProfileFeatures, global Features HuFeatures and local Features SurfFeatures were tested respectively, and then tested in pairs. Finally, they were compared with the fusion operator MixFeatures . The specific detection accuracy is shown in table 2.

Table 2 Comparison of accuracy of each group

A. Comparison of single-feature accuracy

Features	Accuracy（%）
Profile Features	80.77
Hu Features	91.49
Surf Features	95.86

B. Comparison of double-feature accuracy

Features	ProfileFeatures	HuFeatures	SurfFeatures
ProfileFeatures	-	95.23	97.48
HuFeatures	95.23	-	98.11
SurfFeatures	97.48	98.11	-

C. Comparison of multi-feature accuracy

Features	Accuracy（%）
MixFeatures	98.46

As can be seen from the table above, SURF has the highest accuracy in single-feature detection, with a value of 95.86%. Through the use of double feature detection, accuracy all have been significantly improved. It can be said that the recognition rates of all three are satisfactory, with the highest being the Hu&Surf feature, which has a value of 98.11%. The accuracy of multi-feature fusion proposed by us is the highest among all extraction algorithms, reaching 98.46%. Experiments show that the algorithm which we proposed is effective.

5. Conclusion

As an essential means to protect the safety of the elderly, abnormal behavior detection deserves people's extensive attention. This paper makes a new exploration on the abnormal attitude recognition of the elderly based on modern mobile devices. The introduction of mobile devices is an inevitable trend in the era of the Internet, which expands infinite possibilities for the use of software both in space and time. In this paper, an innovation is made in human feature extraction, and a multi-feature fusion algorithm is proposed to replace the traditional single feature algorithm. The results show that the multi-feature fusion algorithm has higher accuracy and is suitable for the increasingly complex daily environment. The research in this paper provides a new idea for feature extraction of abnormal behaviors, which is of considerable significance to the development of computer vision.

Acknowledgements

This work was supported by Science and Technology Development Program of Jilin City(Grant No. 201851901) and Key Program of Graduate Innovation of Beihua University（Grant No.2018009）

References

[1] Sivalingam, R., Cherian, A., Fasching, J., Walczak, N., Bird, N., Morellas, V., Murphy, B., Cullen, K., Lim, K. & Sapiro, G. 2012 A multi-sensor visual tracking system for behavior monitoring of at-risk children. In *2012 IEEE International Conference on Robotics and Automation* (pp. 1345-1350, IEEE.

[2] Bird, N., Atev, S., Caramelli, N., Martin, R., Masoud, O. & Papanikolopoulos, N. 2006 Real time, online detection of abandoned objects in public areas. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.* (pp. 3775-3780, IEEE.

[3] Shah, M. & Jain, R. 2013 *Motion-based recognition*, Springer Science & Business Media.

[4] Blunsden, S. & Fisher, R. 2010 The BEHAVE video dataset: ground truthed video for multi-person behavior classification. *Annals of the BMVA* **4**, 4.

[5] Lee, Y.-S. & Lee, H. 2009 Multiple object tracking for fall detection in real-time surveillance system. In *2009 11th International Conference on Advanced Communication Technology* (pp. 2308-2312, IEEE.

[6] Foroughi, H., Aski, B. S. & Pourreza, H. 2008 Intelligent video surveillance for monitoring fall detection of elderly in home environments. In *2008 11th international conference on computer and information technology* (pp. 219-224, IEEE.

[7] Rougier, C., Meunier, J., St-Arnaud, A. & Rousseau, J. 2007 Fall detection from human shape and motion history using video surveillance. In *21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07)* (pp. 875-880, IEEE.

[8] Rougier, C., Meunier, J., St-Arnaud, A. & Rousseau, J. 2011 Robust video surveillance for fall detection based on human shape deformation. *IEEE Transactions on circuits and systems for video Technology* **21**, 611-622.

[9] Rougier, C. & Meunier, J. 2006 Fall detection using 3D head trajectory extracted from a single camera video sequence. In *First International Workshop on Video Processing for Security (VP4S-06), June* (pp. 7-9.

[10] Przybylski, A. K. & Weinstein, N. 2013 Can you connect with me now? How the presence of mobile communication technology influences face-to-face conversation quality. *Journal of Social and Personal Relationships* **30**, 237-246.

[11] Wu, G., Talwar, S., Johnsson, K., Himayat, N. & Johnson, K. D. 2011 M2M: From mobile to embedded internet. *IEEE Communications Magazine* **49**, 36-43.

[12] Vishwakarma, V., Mandal, C. & Sural, S. 2007 Automatic detection of human fall in video. In *International conference on pattern recognition and machine intelligence* (pp. 616-623, Springer.

[13] Changlin, Z. 2012 Study and realization of automatic fall detection based on video, Anhui university.

[14] Xiaohui, L. 2015 Study on detection of human falls under complex background, Beijing jiaotong university.

[15] Hu, M.-K. 1962 Visual pattern recognition by moment invariants. *IRE transactions on information theory* **8**, 179-187.

[16] Bay, H., Tuytelaars, T. & Van Gool, L. 2006 Surf: Speeded up robust features. In *European conference on computer vision* (pp. 404-417, Springer.

[17] Lowe, D. G. 2004 Distinctive image features from scale-invariant keypoints. *International journal of computer vision* **60**, 91-110.