



A novel real-time fall detection method based on head segmentation and convolutional neural network

Chenguang Yao¹ · Jun Hu¹ · Weidong Min^{1,2} · Zhifeng Deng³ · Song Zou³ · Weiqiong Min⁴

Received: 17 January 2020 / Accepted: 6 May 2020 / Published online: 3 June 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

As the computer vision develops, real-time fall detection based on computer vision has become increasingly popular in recent years. In this paper, a novel real-time indoor fall detection method based on computer vision by using geometric features and convolutional neural network (CNN) is proposed. Gaussian mixture model (GMM) is applied to detect the human target and find out the minimum external elliptical contour. Differently from the traditional fall detection method based on geometric features, we consider the importance of the head in fall detection and propose to use two different ellipses to represent the head and the torso, respectively. Three features including the long and short axis ratio, the orientation angle and the vertical velocity are extracted from the two different ellipses in each frame, respectively, and fused into a motion feature based on time series. In addition, a shallow CNN is applied to find out the correlation between the two elliptic contour features for detecting indoor falls and distinguishing some similar activities. Our novel method can effectively distinguish some similar activities in real time, which cannot be distinguished by some traditional methods based on geometric features, and has a better detection rate.

Keywords Computer vision · Fall detection · Head segmentation · Real-time image processing · Time series motion features

✉ Weidong Min
minweidong@ncu.edu.cn

Chenguang Yao
411727318008@email.ncu.edu.cn

Jun Hu
hujun@ncu.edu.cn

Zhifeng Deng
2690589616@qq.com

Song Zou
383191279@qq.com

Weiqiong Min
1228584642@qq.com

¹ School of Software, Nanchang University, Nanchang 330047, China

² Jiangxi Key Laboratory of Smart City, Nanchang 330047, China

³ School of Information Engineering, Nanchang University, Nanchang 330031, China

⁴ School of Tourism, Jiangxi Science and Technology Normal University, Nanchang 330038, China

1 Introduction

Nowadays, with the aggravation of the aging problem in the world, caring for the health problems of the elderly has become a field of increasing importance, especially the health problems of the elderly living alone. Falling is the primary reason of death among seniors due to injuries, as increasing age, weakened muscle strength and the emergence of chronic diseases all increase the risk of falls [16]. When living alone seniors lose their ability to save themselves because of falls, they may miss the best opportunity for treatment, which can put their lives at risk. Therefore, a real-time intelligent surveillance system is demanded for the fall detection task in order to ensure the health of people, especially the seniors.

These considerations have attracted many researchers to propose the intelligent surveillance system which can automatically detect falls and keep real-time performance [14]. With the arrival of 5G era, Internet of things presents a paradigm growth [9, 18, 19], which makes it possible to reduce the internal cost of installing the intelligent fall detection environments at home. The real-time indoor fall detection

methods based on computer vision only need to install some cameras indoor, which can automatically locate, track the objects and detect falls by analyzing the motion of people. Although wearable sensor-based methods bring high accuracy [15], the elderly often forget or are unwilling to wear the sensors, which makes them difficult to popularize in practice. The biggest advantage of the static vision system is that people do not have to wear any sensor, many fall detection methods are based on the video, and they always achieve good accuracy performance. For example, Xu et al. [34] proposed a fall detection method based on 3D skeleton data obtained from the Microsoft Kinect and adopted long short-term memory (LSTM) networks for fall detection. Yao et al. [38] also applied 3D skeleton data, which was obtained from the Microsoft Kinect, and built a fall detection model named the human torso motion model. Zerrouki et al. [40] detected falling incidents based on the human silhouette shape variation in the video monitoring. Min et al. [13] firstly applied scene analysis in fall detection based on a deep learning method and then detected the spatial relation between the human and the furniture to detect some falls in a special scene.

At present, the fall detection method based on computer vision has a low accuracy in similar activities. In addition, deeper structures lead to high computational complexity and consume much time. Therefore, a novel real-time indoor fall detection method based on computer vision by using geometric features and convolutional neural network (CNN) is proposed. The main contributions of this work are summarized as follows.

1. The method segmenting the head and the torso is proposed to extract the geometric features of them, respectively, which is applied to solve the instability of traditional methods based on geometric features.
2. We propose a shallow CNN, which can learn enough features to achieve a satisfied accuracy. Meanwhile, a shallow structure can converge fast and keep real-time performance with low computational cost.

The rest of this paper is organized as follows. In Sect. 2, we give an overview of the current state-of-the-art fall detection methods related to our method. The head segmentation method and the feature extraction method are shown in Sect. 3. Section 4 describes the fall detection method based on CNN. The experimental results are discussed in Sect. 5. Finally, we provide the conclusion, and the future research directions are discussed in Sect. 6.

2 Related works

In recent years, deep learning [37, 41], cloud computing [20–25], big data [26, 41] and computer vision [10, 29] have been the hot research topics.

Deep learning attracts board attention because it can achieve good performances in different visual tasks [30]. Wang et al. [31] applied an algorithm composed of three networks to improve the segmentation performance from synthetic data to real scenes. Zhou et al. [43] applied a new deep neural network, which improved the performance of semantic segmentation. Real-time performance is an important ability in many deep learning tasks. Zhou et al. [42] applied a lightweight network for the real-time image semantic segmentation. Yang et al. [36] adopted a shallow network to ensure real-time performance and high recognition rate. In the field of fall detection, real-time performance is also very important. Once the fall detection loses real time, the person who falls cannot be found in time, and the fall detection also loses significance. Therefore, increasingly researches focus on real-time fall detection task, and many methods based on computer vision are proposed. However, due to the complexity of visual content and the similar properties of falls with other ordinary human activities, there are still many challenges in fall detection [7]. The main challenges include how to reduce the computational complexity and improve the accuracy. Many algorithms can achieve high accuracy in fall detection, but they need to consume much time [2, 28].

A classical idea of fall detection method is to analyze the shape variance of people in each frame of the video. Normally, a 2D global geometric shape is used to represent the person's motion and then extract some geometric features to judge the fall. There are two classical geometric shape representations in the previous literature: Minimal External Bounding Box and Approximated Ellipse. 1. Minimal External Bounding Box: Min et al. [12] extracted the geometric features of people by the bounding box such as the ratio of the height to width to detect the falling in different directions. Liu et al. [11] fused three features—the human aspect ratio, the effective area ratio and the center variation rate. This method reduced misjudgment and increased the fall detection rate. Chen et al. [1] detected the height and the aspect ratio of the human body in multiple falls which could reduce the false alarm rate by fusing multi-features. 2. Approximated Ellipse: Rougier et al. [28] applied the approximated ellipse to represent the shape of pedestrian in each frame of a video sequence and detected falls by analyzing the human shape deformation. Yu et al. [39] extracted the human contour from the video and fitted with an ellipse. The information of shape and position was combined to describe the features of shape contour. Then, they used an online one-class support vector machine to distinguish

different activities. Fan et al. [5] located the human body by a minimum area-enclosing ellipse and then developed a normalized directional histogram around the center of the ellipse to represent the human posture by multi-directional statistical analysis. Last, a set of features was extracted to feed into a directed acyclic graph support vector machine to distinguish different human postures. There is also another geometric representation: Chua et al. [3] used three points to present a person and extracted motion features to detect falls. This method achieved high accuracy for human fall detection in real-time indoor video sequence.

The geometric features are useful in fall detection, which provide a lot of information about human posture. Min et al. [12] applied the bounding box to represent the human shape. However, the feature information extracted from the bounding box is not as accurate as the ellipse [27, 32]. Meanwhile, it cannot distinguish some similar activities. Because the bounding box will make a great shape change when pedestrians suddenly stretch out their arms in the course of normal walking, which will lead to misjudgment. Although the ellipse fitting can effectively reduce this problem and remove the slender objects carried by pedestrians, some special activities like sitting down brutally and squatting down brutally can easily be misjudged as falling. Furthermore, the human motion is a highly non-rigid activity, and the rules of movement in different parts of the human body are different. For some highly similar activities, it is inaccurate to use one traditional whole geometric shape to represent the whole shape of a human. To address this problem, a novel real-time fall detection method based on the head segmentation is proposed.

Considering that the amplitude of the head motion is huge during the fall, we extract the head motion as new features. Therefore, the method segmenting the head and the torso is proposed, and the two different ellipses are applied to represent the head and the torso, respectively. Then, three features including the long and short axis ratio, the orientation angle and the vertical velocity are extracted from the two different ellipses, respectively, and fused into a motion feature based on time series. Lastly, a shallow CNN is trained to find out the correlation between the two elliptic contour features to detect falls and distinguish some similar activities.

3 Head segmentation and motion features extraction

3.1 Foreground detection

In the foreground detection part, Gaussian mixture model (GMM) [33] uses multiple Gaussian models to represent the features of each pixel. Each pixel is regarded as a variable. Before the foreground detection, the background is

trained at first, and the GMM is used to simulate the background in each frame. Then, in the test stage, the GMM is updated after a new frame of image is obtained, and each pixel in the current image is used to match the GMM. If the matching is successful, it is considered as the background, otherwise it is the foreground. Then, the foreground extraction, the shadow suppression method is applied to suppress the shadow. After that, there may be some voids and noises in the image, and the dilation and the corrosion operations are used to solve this problem.

3.2 Head segmentation

When one traditional ellipse is used to fit the whole human contour, it cannot effectively reflect the difference between some similar activities. It would increase the false alarm rate and lead to misjudgment. In order to improve the distinguish ability of our method to the similar activities, the importance of the head motion is considered, because the amplitude of the head motion is huge during the fall. Therefore, the head segmentation method is proposed, and the two different ellipses are applied to fit the head and the torso of the human, respectively.

3.2.1 Head pre-location

A head pre-location method is proposed to approximately locate the head position, which is described in Fig. 1.

In Fig. 1, the foreground is obtained from the input frame. Then, the head is segmented by the proportion of the head to height. Lastly, the approximate position of the head can be obtained by the bounding box fitting.

3.2.2 Head tracking

After the head pre-location, the mean shift tracking method [4] is used to track the head. This algorithm has low calculation complexity, and the target can be real-time tracked when the target area is known. Meanwhile, it is also insensitive to the edge occlusion, the target rotation, the deformation and

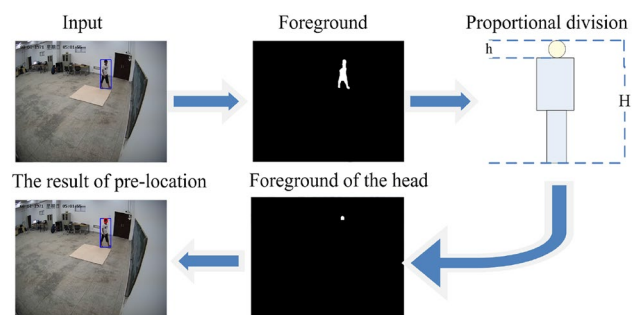


Fig. 1 The pre-location of the head

Fig. 2 The results of tracking head. **a–f** The tracking results under different angles and different people

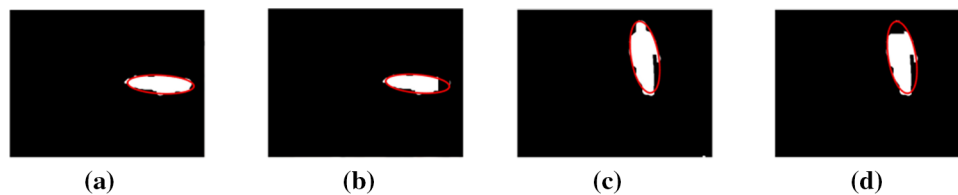
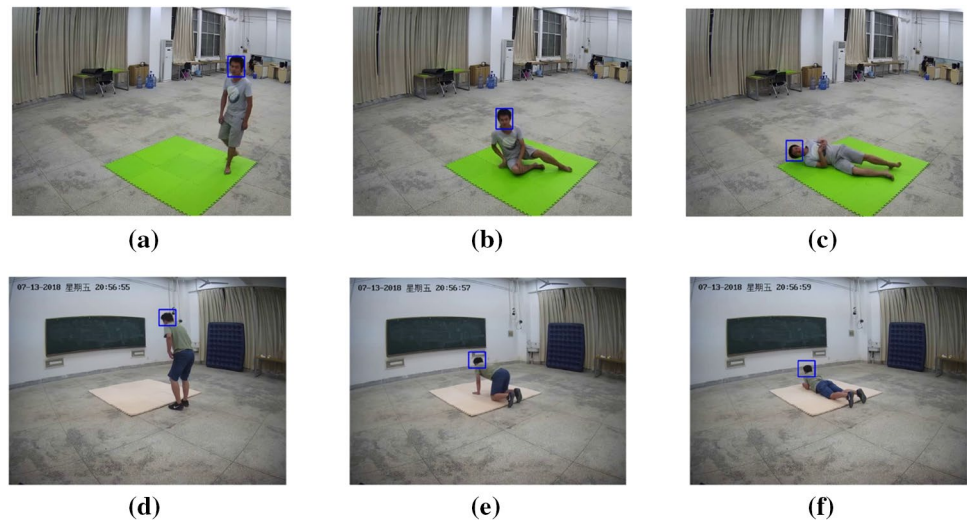


Fig. 3 The whole human body and the torso fitting results based on the traditional method. **a, c** The whole human body ellipse fittings with the traditional method. **b, d** The torso ellipse fittings with the traditional method

the background motion, so the positioning of the target will be more accurate. The target model and the candidate model of this method are calculated based on the distributions of the target region and the candidate region, respectively. Then, the similar function is used to measure the similarity between the initial frame target model and the candidate model of the current frame, and the candidate model which maximizes the similar function is selected. The mean shift vector of the target model is obtained, which is the vector of the target moving from the initial position to the correct position. The mean shift algorithm will converge to the real position of the target and achieve the goal of tracking by iteratively calculating the mean shift vector.

Firstly, the approximate position of the head can be found out by the head pre-location. Then, the head is tracked by the mean shift tracking method. The results of our method are shown in Fig. 2.

3.2.3 Ellipse fitting

After the head tracking stage, the two ellipses are used to fit the head and the torso, respectively. As shown in Fig. 3, the traditional ellipse fitting method [17] is used to fit the torso of the human, but this method cannot effectively reflect the difference between the whole human and the torso of the

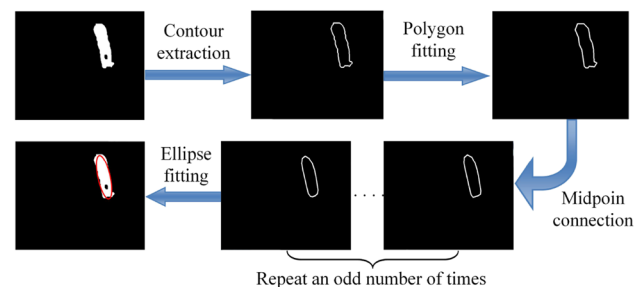


Fig. 4 The torso ellipse extraction diagram

human. Therefore, we modify the torso contour to achieve a compact torso elliptical contour. Firstly, the torso contour is fitted by polygon. Secondly, each side of the polygon is connected to their midpoint and this operation is repeated an odd number of times. The shape of the torso contour will be an ellipse. Lastly, the torso is fitted in this way to obtain a more compact ellipse representation. Figure 4 describes the torso ellipse extraction diagram. Figure 5 shows the results of the torso ellipse fitting. Figure 5a, c shows the results of the ellipse fitting by the traditional method, while Fig. 5b and d shows the results of the torso fitting by our method. Compared with the traditional method, a compact torso ellipse is obtained. This method is also used for the head

Fig. 5 The results of the ellipse fitting based on our method. **a, c** The ellipse fittings based on the traditional method. **b, d** The ellipse fittings based on our method

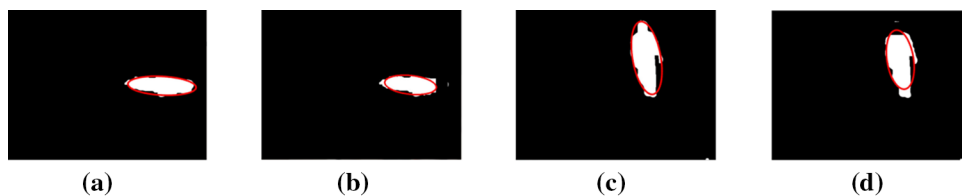


Fig. 6 The fitting results of the different actions. **a–d** The different actions scenarios. The *blue ellipse* represents the head, the *green ellipse* represents the torso, and the *red ellipse* represents the whole body

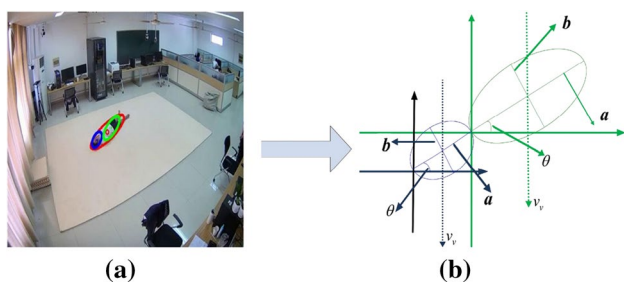
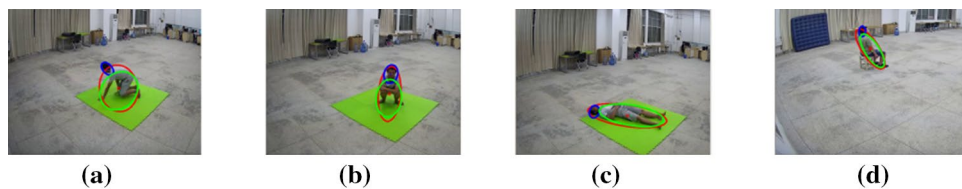


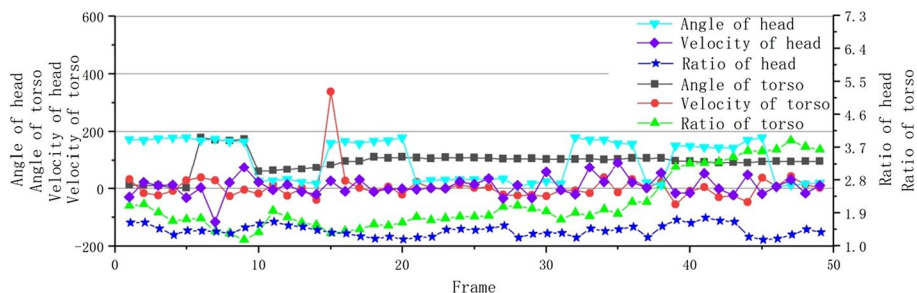
Fig. 7 A feature extraction sample. **a** The *red ellipse* is the whole body ellipse fitting; the *blue ellipse* is the head ellipse fitting; and the *green ellipse* is the torso ellipse fitting. **b** Schematic diagram describing the characteristics of falls

fitting. In Fig. 6, it shows the results of the head and the torso ellipse fitting. These blue, green and red ellipses represent the head, the torso and the whole body, respectively. By observing the effect of different actions on ellipse fitting, it can be concluded that the two ellipses fit human body more accurate than one ellipse.

3.3 Motion features extraction

After the two ellipses fit the head and the torso, respectively, the silhouette features and the velocity feature are extracted

Fig. 8 The motion feature based on time series



from each of them. Therein, the silhouette features are the inclination angle of ellipse Θ and the ratio of the long and short axis of the ellipse $\rho = a/b$. When the people’s action changes, the angle Θ and the ratio ρ both change. Once a fall occurs, the velocity in vertical direction will change rapidly. The velocity in the vertical direction of the ellipse center is extracted as Eq. (1):

$$v_v = \sqrt{(y_n - y_{n-1})^2 + (x_n - x_{n-1})^2} * F * \sin \Theta \tag{1}$$

where v_v represents the velocity in vertical direction; (x_{n-1}, y_{n-1}) is the coordinate center of the $n - 1$ th frame; (x_n, y_n) is the coordinate center of the n th frame; F represents the number of frame per second; and $\sin \Theta$ represents the sine value of the inclination angle of ellipse.

A feature extraction diagram is presented in Fig. 7, where in Fig. 7b, a and b represent the long and the short axis of the ellipse, respectively; Θ represents the inclination angle of ellipse; v_v represents the velocity in the vertical direction of ellipse contours center. These six extracted features are fused into a motion feature based on time series. The motion feature is shown as Fig . 8

4 Real-time CNN-based fall detection

In order to find out the correlation relationship between the head ellipse and the torso ellipse during the fall, deep learning is used to learn the motion features. However, the deeper network structures adopted by predecessors have the problems of the high computational cost and the low convergence [6, 8, 40]. Therefore, we choose a shallow CNN structure to solve these problems. With the above preprocessing methods, the target of the image is segmented from the background for the shallow CNN can learn enough features to achieve a satisfied accuracy. Compared with the deeper structures, the shallow structure can converge fast and keep real-time performance with low computational cost.

4.1 Convolutional neural network

CNN is a deep and feed-forward artificial neural network structure [35], which is widely applied to analyze image. The biggest advantage of the CNN structure is that it can optimize the weight of CNN through a large amount of training dataset without tedious manual operation, so as to achieve accurate classification. The main compositions of CNN are described briefly as below:

- *Convolutional layer* Take the input raw image convoluted with a many trainable filters (or called convolutional kernel) and additive bias vectors to obtain multiple mapping feature maps.
- *Pooling layer* In generally, behind the convolutional layer, it is used for down-sampling to reduce the dimension of the feature. The two most traditional pooling methods are max pooling and mean pooling.

- *Fully connected layer* After the original image is processed by multiple convolutional layers and pooling layers, the output features are compressed into a one-dimensional vector and used for classification. In this layer, other features can be added to this one-dimensional vector.

4.2 CNN-based fall detection

In this paper, the shallow CNN structure shown in Fig. 9 is applied to train and learn the motion features based on time series. There are 74 training videos and 28 test videos, while the learning rate of this structure is set as 0.00001 and the number of epochs is 500. Specifically, firstly, 196 filters of size 1×12 are used in the convolutional layer to learn the three divided feature maps based on time series to obtain a rich feature representation of the data. There is only one layer in the convolutional layer. Then, after ReLU activation function is applied to the 196 feature maps, the max pooling layer of size 1×4 is used to reduce four times for dimensionality. The feature maps output by pooling layer is flattened and stacked together with some statistical features, (e.g., mean value) to obtain 1024 features through the fully connected. Finally, those features of the fully connected layer output are passed to the softmax function, which calculates the last classification. This model is trained to minimize the cross-entropy loss function which is augmented with the l_2 -norm regularization of CNN weights. The back-propagation algorithm is used to calculate the gradient, and the modified method of stochastic gradient descent is used to optimize the network parameters.

Fig. 9 The structure of the proposed shallow CNN

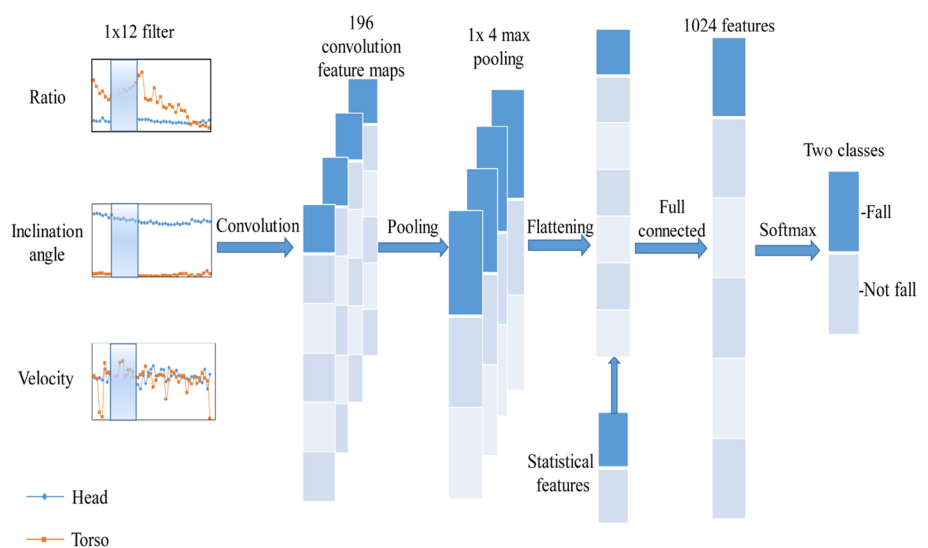


Fig. 10 Different activities on our self-collected dataset. **a** Sitting down. **b, c** Crouching down. **d** Walking. **e** Squatting down. **f–h** Different falling down

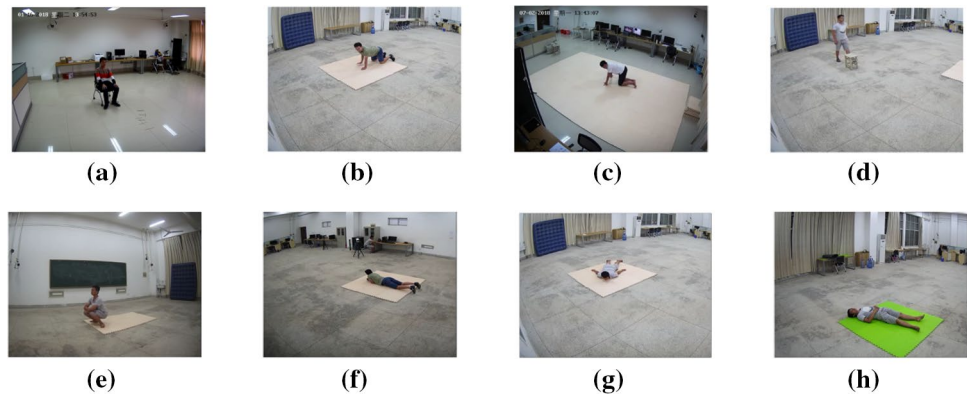


Table 1 The detailed description of experimental data on the self-collected dataset

Experimental data description	Training dataset	Test dataset
Positive sample	14284 (frame)	4247 (frame)
Negative sample	18614 (frame)	5530 (frame)
Total	32898 (frame)	9777 (frame)

Table 2 The test results on self-collected dataset by our method

Fall detection rate (%)	90.5
False alarm rate (%)	10.0

5 Experiment results and analysis

All the experiments are carried out on a laptop PC with Inter(R) Core(TM) i5-4300U CPU @ 1.9GHz and 4GB RAM. In order to test the CNN structure in this paper, we simulate falls and normal daily activities to collect lots of video frame samples. Multiple monocular cameras are used to film 102 short videos from different views and height. These videos include normal activities such as crouching down, walking, squatting down and sitting down, as well as simulating falls in different directions such as backward falls, forward falls and sideways falls. Figure 10 shows different normal activities and simulates falls in various scenes. In the test dataset, there are 30 simulated fall activities and 28 normal activities.

Enough train and test frame samples are collected from the self-collected dataset. The detailed description of experimental data is shown in Table 1, there are 14284 frames positive sample images and 18614 frames negative sample images in training dataset, while the test dataset has 4247 frames positive sample images and 5530 frames negative sample images.

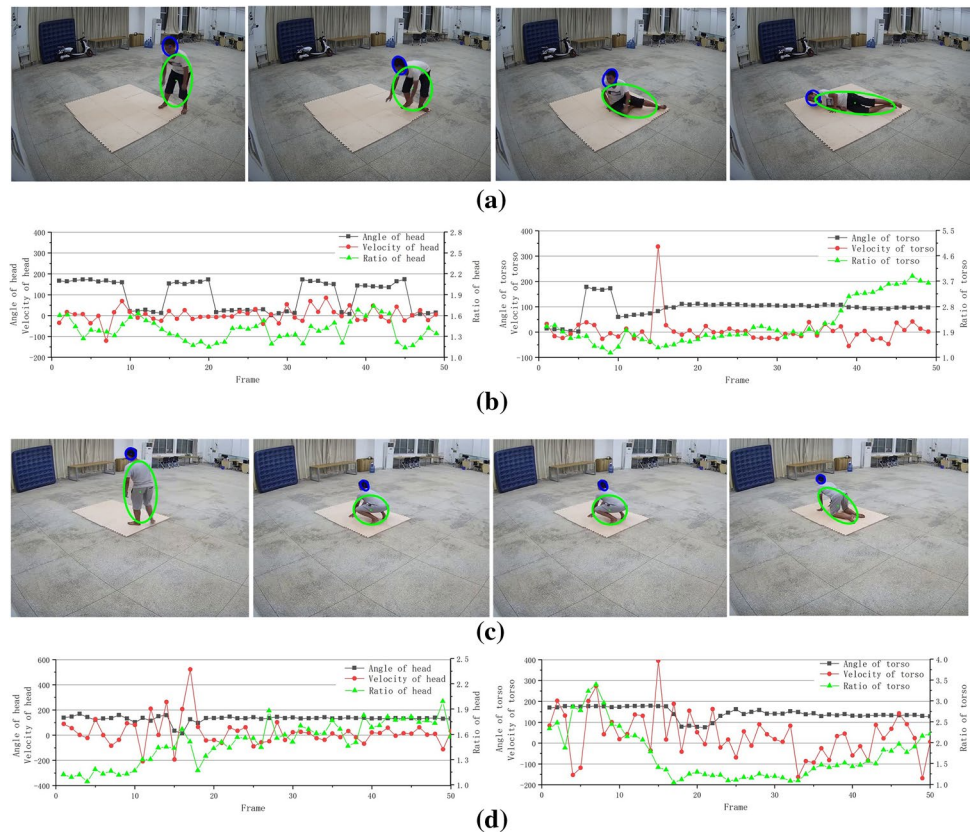
The six feature data are fused into a motion feature based on time series as shown in Fig. 8. Then, they are regarded as the input for training and testing in CNN, and the test results are shown in Table 2. The fall detection rate of this method is as high as 90.5%, and the false alarm rate is as low as 10.0%.

When one elliptic contour is used to fit the human and detect falls, some special activities like sitting down brutally can easily be misjudged as falling. Facing similar activities like sideward fall and crouching down, this method also has a high false detection rate. Therefore, the discriminating accuracy of these similar activities is tested with our method. Figures 11 and 12 show the two groups of similar activities and their motion features based on time series for proving the feasibility of our method in distinguishing similar activities. Figure 11a, c shows sideward fall and crouching down, while Fig. 12a, c shows backward fall and sitting down. It can be seen that the size and shape of the fitted ellipse are very similar in each group, but Figs. 11b, d and 12b, d show there are obvious differences in the change of the head and the torso features of the two activities. According to our experiments, when the two ellipses are used to represent the head and the torso, respectively, the two similar activities can be effectively distinguished.

To further demonstrate the effectiveness of this method, many extensive experiments are made to compare with some classical methods. The three classical algorithms are achieved in this paper, which are the bounding box ratio analysis approach [33], the ellipse shape analysis approach [27] and Chua's approach [3]. The experimental results of these methods on the self-collected dataset are shown in Table 3.

The ellipse shape analysis approach [27] uses an ellipse to fit the whole person, and ellipse features and motion history images are fused to detect falls. The bounding box ratio analysis approach [33] uses a traditional bounding box to represent a person and detects a fall by analyzing the aspect ratio of the bounding box. Chua's approach [3] uses three points to represent the human body and extracts features

Fig. 11 The results analysis of sideward fall and crouching down activities. **a** Sideward fall. **b** The head and the torso features (sideward fall) based on time series, respectively. **c** Crouching down. **d** The head and the torso features (crouching down) based on time series, respectively



from the three points to detect falls. As shown in Table 3, our method has achieved a detection accuracy of 90.5% and the false alarm rate of 10.0%. The specific fall detection rates are as follows: the bounding box ratio analysis approach (60.0%), the ellipse shape analysis approach (70.0%), Chua’s approach (83.3%) and our method (90.5%). The false alarm rates include our method (10.0%), Chua’s approach (13.8%), the ellipse shape analysis approach (22.2%) and the bounding box ratio analysis approach (25.0%). Compared to other traditional geometric feature methods, the two ellipses we used have achieved a higher fall detection rate and lower false alarm rate. Because the two ellipses fitting the head and the torso, respectively, is closer to the contour of human body than other geometries, which can obtain more accurate motion features, in addition, a shallow CNN is applied to learn the correlation between the two elliptic contour features, which can be accurately distinguished some similar activities.

We have also done enough experiments for real-time test. As shown in Table 4, the used camera can take videos by the rate of 15 fps. Through testing, we have a frame rate of 17.7 fps in the head segmentation section, and a frame rate of 17.0 fps in the motion feature extraction section. In the CNN-based fall detection part, the method we proposed is also excellent enough. We use 9777 frames of the test data, and the test time is 66.3 s, so the proposed method has a good real-time performance.

6 Conclusions

A novel real-time fall detection method is proposed, which is based on the head segmentation and CNN in this paper. This method is different from the traditional single geometric representation approaches. Firstly, the head is segmented from the body, and the two different ellipses are applied to represent the head and the torso,

Fig. 12 The results analysis of backward fall and sitting down activities. **a** Backward fall. **b** The head and the torso features (backward fall) based on time series, respectively. **c** Sitting down. **d** The head and the torso features (sitting down) based on time series, respectively

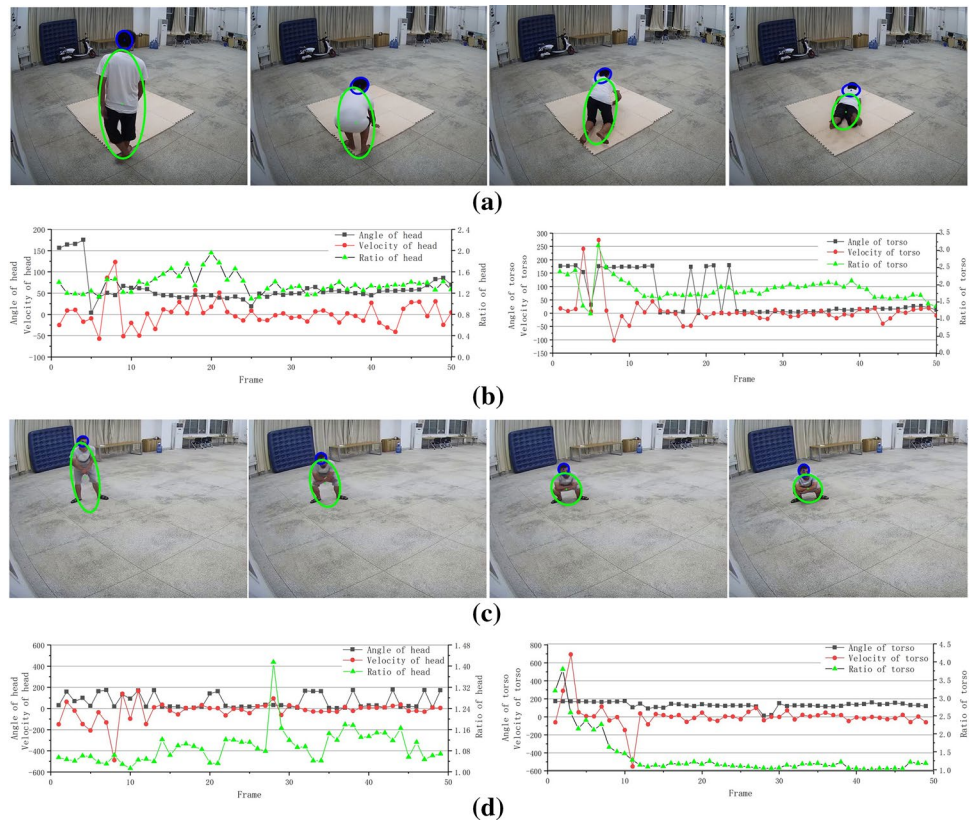


Table 3 The experimental results of these methods on the self-collected dataset

Method	Our method	[27]	[3]	[33]
Fall detection rate (%)	90.5	70.0	83.3	60.0
False alarm rate (%)	10.0	22.2	13.8	25.0

Table 4 The experimental results of real-time frame rate

Our method	Head segmentation	Motion feature extraction	CNN-based fall detection
Frame per second (fps)	17.7	17.0	147.5

respectively. Then, three features including the long and the short axis ratio, the orientation angle and the vertical velocity are extracted from the two ellipses in each frame, respectively, and fused into a motion feature based on time series. Finally, a shallow CNN is used to find out the relations between the two ellipses to distinguish some similar activities. Compared with other state-of-the-art methods, our method can effectively distinguish some similar activities while others cannot. Therefore, the detection rate is increased. The experiments also show

that the proposed method has a good real-time performance. In future research, we will look for ways to cope with occlusion in a more realistic indoor environment and explore the possibility of applying this method outdoors.

Acknowledgements This work is supported by the National Natural Science Foundation of China under Grant No. 61762061, the Natural Science Foundation of Jiangxi Province, China, under Grant No. 20161ACB20004 and Jiangxi Key Laboratory of Smart City under Grant No. 20192BCD40002. The authors Chenguang Yao and Jun Hu contributed equally to this paper and shall be considered as co-first authors.

Compliance with ethics requirements

Conflict of interest The authors declare that they have no conflict of interest.

Informed consent All participants were informed and agreed that the study was in compliance with relevant ethical requirements.

References

1. Chen, M.C.: A video surveillance system designed to detect multiple falls. *Adv. Mech. Eng.* **8**(4), 1–11 (2016)
2. Chen, Y.T., Lin, Y.C., Fang, W.H.: A hybrid human fall detection scheme. In: 2010 IEEE International Conference on Image Processing (ICIP), pp. 3485–3488. IEEE (2010)

3. Chua, J.L., Chang, Y., Lim, W.K.: A simple vision-based fall detection technique for indoor video surveillance. *Signal Image Video Process.* **9**(3), 623–633 (2015)
4. Comaniciu, D., Ramesh, V., Meer, P.: Real-time tracking of non-rigid objects using mean shift. In: 2000 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2142. IEEE (2002)
5. Fan, K., Wang, P., Hu, Y., Dou, B.: Fall detection via human posture representation and support vector machine. *Int. J. Distrib. Sens. Netw.* **13**(5), 1–21 (2017)
6. Fan, Y., Levine, M.D., Wen, G., Qiu, S.: A deep neural network for real-time detection of falling humans in naturally occurring scenes. *Neurocomputing* **260**(18), 43–58 (2017)
7. Feng, W., Liu, R., Zhu, M.: Fall detection for elderly person care in a vision-based home surveillance environment using a monocular camera. *Signal Image Video Process.* **8**(6), 1129–1138 (2014)
8. Fischer, T., Krauss, C.: Deep learning with long short-term memory networks for financial market predictions. *Eur. J. Oper. Res.* **270**(2), 654–669 (2017)
9. Gong, W., Qi, L., Xu, Y.: Privacy-aware multidimensional mobile service quality prediction and recommendation in distributed fog environment. *Wirel. Commun. Mob. Comput.* **2018**, 1–8 (2018)
10. Kendall, A., Gal, Y.: What uncertainties do we need in Bayesian deep learning for computer vision? In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 5574–5584 (2017)
11. Liu, H., Zuo, C.: An improved algorithm of automatic fall detection. *AASRI Procedia* **1**, 353–358 (2012)
12. Min, W., Wei, L., Han, Q., Ke, Y.: Human fall detection based on motion tracking and shape aspect ratio. *Int. J. Multimed. Ubiquitous Eng.* **10**(11), 1–14 (2016)
13. Min, W., Cui, H., Rao, H., Li, Z., Yao, L.: Detection of human falls on furniture using scene analysis based on deep learning and activity characteristics. *IEEE Access* **6**, 9324–9335 (2018)
14. Nez-Marcos, A., Azkune, G., Arganda-Carreras, I.: Vision-based fall detection with convolutional neural networks. *Wirel. Commun. Mob. Comput.* **2017**, 1–16 (2018)
15. Nguyen, T.T., Cho, M.C., Lee, T.S.: Automatic fall detection using wearable biomedical signal measurement terminal. In: 2009 IEEE International Conference of Engineering in Medicine and Biology Society, pp. 5203–5206. IEEE (2009)
16. Ozcan, A., Donat, H., Gelecek, N., Ozdirenc, M., Karadibak, D.: The relationship between risk factors for falling and the quality of life in older adults. *BMC Public Health* **5**, 1–6 (2005)
17. Pratt Jr., W.K., Adams, J.E.: Digital image processing. *J. Electron. Imaging* **16**(2), 633–640 (2007)
18. Qi, L., Chen, Y., Yuan, Y., Fu, S., Zhang, X., Xu, X.: A qos-aware virtual machine scheduling method for energy conservation in cloud-based cyber-physical systems. *World Wide Web* **23**, 1275–1297 (2020)
19. Qi, L., Dai, P., Yu, J., Zhou, Z., Xu, Y.: Time–location–frequency-aware Internet of things service selection based on historical records. *Int. J. Distrib. Sens. Netw.* **13**(1), 1550147716688696 (2017)
20. Qi, L., Dou, W., Chen, J.: Weighted principal component analysis-based service selection method for multimedia services in cloud. *Computing* **98**(1–2), 195–214 (2016)
21. Qi, L., Wang, R., Hu, C., Li, S., He, Q., Xu, X.: Time-aware distributed service recommendation with privacy-preservation. *Inf. Sci.* **480**, 354–364 (2019)
22. Qi, L., Xu, X., Dou, W., Yu, J., Zhou, Z.: Time-aware IoE service recommendation on sparse data. *Mob. Inf. Syst.* **2016**, 1–12 (2016)
23. Qi, L., Yu, J., Zhou, Z.: An invocation cost optimization method for web services in cloud environment. *Sci. Program.* **2017**, 1–9 (2017)
24. Qi, L., Zhang, X., Dou, W., Hu, C., Yang, C., Chen, J.: A two-stage locality-sensitive hashing based approach for privacy-preserving mobile service recommendation in cross-platform edge environment. *Futur. Gener. Comp. Syst.* **88**, 636–643 (2018)
25. Qi, L., Zhang, X., Dou, W., Ni, Q.: A distributed locality-sensitive hashing-based approach for cloud service recommendation from multi-source data. *IEEE J. Sel. Areas Commun.* **35**(11), 2616–2624 (2017)
26. Qi, L., Zhou, Z., Yu, J., Liu, Q.: Data-sparsity tolerant web service recommendation approach based on improved collaborative filtering. *IEICE Trans. Inf. Syst.* **100**(9), 2092–2099 (2017)
27. Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Fall detection from human shape and motion history using video surveillance. In: 2007 IEEE International Conference on Advanced Information NETWORKING and Applications Work-shops, vol 2, pp. 875–880. IEEE (2007)
28. Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Robust video surveillance for fall detection based on human shape deformation. *IEEE Trans. Circuits Syst. Video Technol.* **21**(5), 611–622 (2011)
29. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J.: Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2016, pp. 2818–2826. IEEE (2016)
30. Wang, Q., Chen, M., Nie, F., Li, X.: Detecting coherent groups in crowd scenes by multiview clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(1), 46–58 (2020)
31. Wang, Q., Gao, J., Li, X.: Weakly supervised adversarial domain adaptation for semantic segmentation in urban scenes. *IEEE Trans. Image Process.* **28**(9), 4376–4386 (2019)
32. Williams, A., Ganesan, D., Hanson, A.: Aging in place: fall detection and localization in a distributed smart camera network. In: International Conference on Multimedia, pp. 892–901. ACM (2007)
33. Xiao, H., Wang, X., Li, Q., Wang, Z.: Gaussian mixture model for background based automatic fall detection. In: International Conference on CyberSpace Technology, pp. 234–237. IET (2013)
34. Xu, T., Zhou, Y.: Elders' fall detection based on biomechanical features using depth camera. *Int. J. Wavelets Multiresolut. Inf. Process.* **16**(2), 1840005 (2018)
35. Yang, Z., Leng, L., Kim, B.-G.: StoolNet for color classification of stool medical images. *Electronics* **8**(12), 1464 (2019)
36. Yang, Z., Li, J., Min, W., Wang, Q.: Real-time pre-identification and cascaded detection for tiny faces. *Appl. Sci.* **9**(20), 4344 (2019)
37. Yann, L.C., Yoshua, B., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2016)
38. Yao, L., Min, W., Lu, K.: A new approach to fall detection based on the human torso motion model. *Appl. Sci.* **7**(10), 993 (2017)
39. Yu, M., Yu, Y., Rhuma, A., Naqvi, S.M.R., Wang, L., Chambers, J.A.: An online one class support vector machine-based person-specific fall detection system for monitoring an elderly individual in a room environment. *IEEE J. Biomed. Health Inform.* **17**(6), 1002–1014 (2013)
40. Zerrouki, N., Houacine, A.: Combined curvelets and hidden Markov models for human fall detection. *Multimed. Tools Appl.* **77**(5), 6405–6424 (2018)
41. Zheng, Y., Xu, X., Qi, L.: Deep CNN-Assisted personalized recommendation over big data for mobile wireless networks. *Wirel. Commun. Mob. Comput.* **2019**, 1–6 (2019)
42. Zhou, Q., Wang, Y., Liu, J., Jin, X., Latecki, L.J.: An open-source project for real-time image semantic segmentation. *Sci. China Inf. Sci.* **62**(2), 227101–227102 (2019)
43. Zhou, Q., Yang, W., Gao, G., Ou, W., Lu, H., Chen, J., Latecki, L.J.: Multi-scale deep context convolutional neural networks for semantic segmentation. *World Wide Web* **22**(4), 1–16 (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Chenguang Yao received the B.E. degree in software engineering from Qingdao University in China in 2018. He is a postgraduate at Nanchang University in China now. His research subject is abnormal behavior detection in video surveillance.



Zhifeng Deng received the B.E. degree in software engineering from Chongqing University of Posts and Telecommunications, China, in 2016. He obtained his M.E. degree from Nanchang University in China in 2019. His research subject is abnormal behavior detection in video surveillance.



Jun Hu received the B.E. degree in department of computer science from Jiangxi University in China in 1992. He received the M.E. degree in Applied mathematics from Nanchang University in China in 1997. He received the Ph.D. degree in computer application from Beijing Institute of Technology in China in 2003. He visited the School of Computer Science, Nanyang Technological University, Singapore, from 2012.10 to 2012.12. He visited the Clemson University School of Computer

Science for a year from 2017 to 2018. He is mainly engaged in the research work of artificial intelligence, computational intelligence, trust computing, deep learning and other related fields.



Song Zou received the B.E. degree in material forming and control engineering from Nanchang University in China in 2008. He obtained his M.E. degree in computer software and theory from Zhejiang University in China in 2011. He is a Ph.D. student in the School of Information Engineering in Nanchang University. His research interests include image and video processing, vision-based human action understanding and abnormal behavior detection.



Weidong Min received the B.E., M.E. and Ph.D. degrees in computer application from Tsinghua University, China, in 1989, 1991 and 1995, respectively. He is currently a Professor and the Dean, School of Software, Nanchang University, China. He is an Executive Director of China Society of Image and Graphics. His current research interests include image and video processing, artificial intelligence, big data, distributed system and smart city information technology.



Weiqiong Min received the B.E. degree in international culture and M.E. degree in international trade from Kyushu University, Japan, in 2002 and 2006, respectively. She is now an Assistant Professor, School of Tourism, Jiangxi Science and Technology Normal University, China. Her research interests include image and video processing, hi-tech applications in tourism.