FALL DETECTION BASED ON MOTION HISTORY IMAGE AND HISTOGRAM OF ORIENTED GRADIENT FEATURE

Qi Feng, Chenqiang Gao, Lan Wang, Minwen Zhang, Lian Du, Shiyu Qin

Chongqing University of Posts and Telecommunication, Chongqing, China
Chongqing Key Laboratory of Signal and Information Processing
E-mail: FengQi.FQ@outlook.com

ABSTRACT

In recent years, the aging of population is one of the problems that many countries need to face. Along with the increasing proportion of elderly people living alone, there are more indoor but fatal accidents. Fall is one of these common and dangerous accidents for the elderly. Thus timely rescue after falls becomes particularly important, especially for elderly people who live alone. With the development of computer vision technology and the popularity of home surveillance, the fall detection algorithm based on video analysis provides a good solution to this problem. In this paper, we propose a new fall events detection algorithm. Our algorithm gets sub-motion history image by mapping faster R-CNN detected bounding boxes to motion history image, then extracts histogram of oriented gradient features, and finally uses support vector machine for fall classification. Proved by experiment, Our approach achieves very high recall rates and precision rates in a dataset of realistic image sequences of simulated falls and daily activities.

Index Terms— Fall detection, Faster R-CNN, Motion history image, HOG feature

1. INTRODUCTION

In recent years, with the development of socio-economic, changes of living style, miniaturization of family structure and other factors, sociologists named these families which do not have children at home or whose children are not around the elderly as 'empty nest families'. The number of empty nests is increasing, and the situation makes the rescue for elderly after fall becomes more difficult. The latest World Health Organization reports that about 28% to 35% of older people aged over 65 have fallen each year, compared with 32% to 42% for older people over 70 years of age [1] . The mortality rate caused by falls sharply increased with age, people aged 65 and above, 57% of female and 36% of male deaths caused by falls.

Therefore, timely empty nesters fall event detection has become extremely important, and it is also a difficult problem that the whole world dedicated to solve. With action detection developed tremendously, some effective methods of action detection were proposed [2, 3, 4]. And dataset is an important part of machine learning. Some useful datasets were published, such as unconstrained interaction dataset [3] and in frared action dataset [4]. As a branch of action detection, fall event detection was also developing quickly. A reliable fall detection algorithm can detect the fall event in time, thus providing valuable time for the rescue. Therefore, we propose a fall detection algorithm based on motion history image (MHI) [5] and histogram of oriented gradient (HOG) [6] feature. At first, we use faster R-CNN [7] to detect person in each RGB frame and output the bounding boxes. And then, we track each object and MHI is generated by the same squence of RGB frames. After that, we generate sub-MHI by mapping the bounding boxes, which is the output of faster R-CNN, to the corresponding MHI. We track the ROI in the bounding boxes and extract the features. Last, the feature vector obtained in the previous step is input to the classifier and perform the fall classification.

The rest of this paper is organized as follows: Related work reviews some classic methods in section 2. In section 3, we introduce our proposed method and some technical details. Experiment results are represented in section 4 and we conclude this paper in section 5 finally.

2. RELATED WORK

There are many ways to detect the fall event, and the detection mechanism of the fall event is studied from the different aspects of the human body data sensing mode, the platform and the system network architecture, and some achievements have been made. From the perspective of perception, the current fall detection scheme is divided into the following categories: environment based fall detection, wearable sensor based fall detection and video based fall detection. With the

This work is supported by the National Natural Science Foundation of China (No.61571071), Wenfeng innovationand start-up project of Chongqing University of Posts and Telecommunications (No.WF201404), Undergraduate research training program (No.A2014-39), the Research Innovation Program for Postgraduate of Chongqing (No.CYS17222).

978-1-5386-2159-2/17/$31.00 ©2017 IEEE

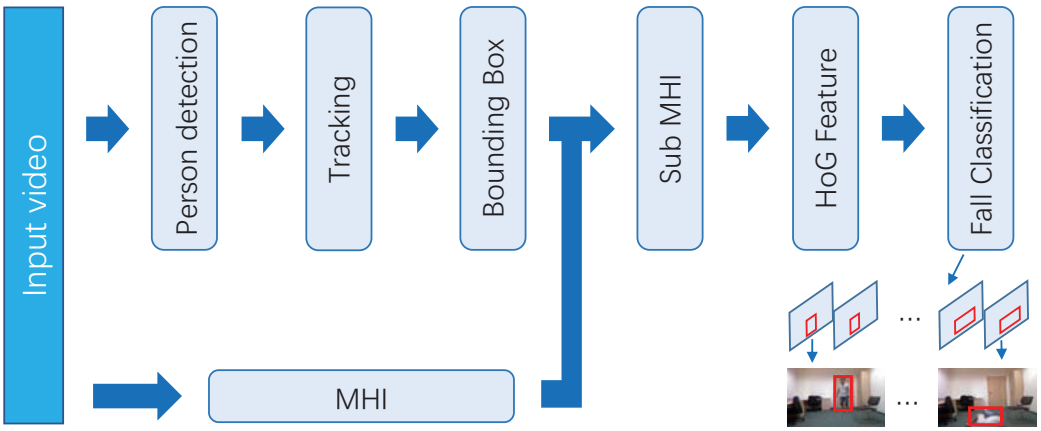

Fig. 1. Pipeline of our approach.

rapid development of video surveillance system and machine learning, video analysis based fall detection shows its unique advantages.

Vinay et al [8] proposed a fall detection algorithm based on finite state machine. The article converted the state machine to achieve fall detection by the changing of aspect ratio, inclination and gradient of the bounding box. A method based on ellipse analysis was introduced by Homa et al in [9]. They extracted features of the ellipses which is a approximation of a human shape. This method achieved a high precision, but recall was not enough for fall detection. Jia-Luen Chua et al [10] proposed a method also based on human shape analysis, but they used a bounding box which is divided into upper portion and lower potion to represent a human shape instead of a ellipse. Their precision was as well as Homa's method, but got a better recall rate. Shengke et al [11] proposed a fall detection method based on principal component analysis (PCA) network , they got the ROIs and the corresponding subimages by background subtracting, then labled every subimages by a trained PCA net. Finally, the labels corresponding to the image squence were treated as a feature vector input to the SVM for fall classification. A fall detection method based on motion history image and human shape features was proposed by Caroline et al [12] . The motion history image used in this method can be a good representation of the intensity of the motion which is effective for fall detection. Eric et al [13] proposed a method based on depth images. They calculated the fall confidence by five features such as speed, acceleration and ground projection changing rate et al. Xin Ma et al. [14] proposed a fall detection method based on the extream learning machine. They also used Kinect as a sensor and got the foreground in each frame by Gaussian mixture modeling. And then extracted the curvature scale spatial feature of foreground objects, and used this feature to establish BoG model for action recognition. Samuele et al. [15] designed a fall detection system that incorporates 3D information and wearable devices information. This method determined the fall by fusing the human skeleton information provided by Kinect and the data obtained by wearable devices on the waist and wrist. This method performanced well, but due to the limit effective range of skeleton information provided by official kinect SDK, it is not widely used in reality.

3. OUR APPROACH

In the proposed method, we apply feature learning methods to detect a fall action. The pipeline of our fall detection framework based on MHI and HOG is shown in Fig.1. We detect person by faster R-CNN firstly, and tracking each object. For each bounding box obtained in previous step, we extend it to the size of latest 10 frames in order to fitting motion history images. And then we map all bounding boxes to the corresponding MHIs which are generated by the same RGB image squences. And we extract HOG feature of each sub-MHI in bounding boxes. At last, using a trained SVM for fall classification.

3.1. Person Detection and Tracking

Person detection is the most basic step in the overall fall detection algorithm, many proposed methods are using background subtraction to get the foreground objects. But there are some problems when using background subtraction, such as ghost area. To avoid these problems, our method uses the faster R-CNN for person detection.

Faster R-CNN is a fast and accurate object detection network proposed by Ren Shaoqin [7]. We obtain a high precision by using a pretrained VGG16 model.

Before fall classification, tracking is needed. We adopt a tracking method based on the minimum distance matching. First, each track is initialized at the first frame of a video, which is to set the center point of the bounding boxes as the

initial center point. Next, the Euclidean distance between the center points of the bounding boxes in the following frame and the initial central points is calculated. The point with the smallest distance and less than the threshold belongs to the track corresponding to the initial point. If the minimum distance is greater than the threshold, a new target appears. Before calculating the next distance, initial centre points need to update, which is to set centre points in the latest frame as new initial centre points. The tracking method based on minimum distance matching is dependent on the detection results. If an object is not detected in one of the frame sequence, a new but incorrect track will appear in the next frame. In order to avoid this problem, we copy the detection result in last frame to the current frame if there are no consisitent centre points to add to tracks. But when the number of frames without detection results exceeds the threshold, the corresponding object disappears.

3.2. Motion History Image

Each fall is accompanied by a large motion, so we decide to extract the motion information from the image as a basis for the fall classification. Optical flow [16] is widely used in motion detection, but it is not applicable for real-time applications such as fall detection. Thus, we adopt motion history image(MHI) to extract motion information in image sequences. MHI was first introduced by Bobick and Davis[5]. MHI expresses motion information of the object in the form of image brightness by calculating the pixel changes at the same location in the time period. The value of each pixel represents recent motion of the pixel at that position in a video sequences. The closer the time of the last movement is to the current frame, the higher the gray value of the pixel is. Therefore, MHI images can represent the latest actions of an individual during an action, which makes MHI widely used in the field of motion recognition.

Using the frame difference method, the pixel values H (x, y, t) in the motion history image can be calculated by the following update function:

$$H_\tau(\mathbf{t}) = \begin{cases} \tau & \textit{if} \quad D \geq \xi \\ \mathbf{max}(0, H_\tau(t-1) - \delta) & \textit{otherwise} \end{cases} \tag{1}$$

where, H_τ is a function about(x, y, t), which is the location and time of each pixel, τ is the duration, which determines the time range of motion from the number of frames, δ is a decline parameter, ξ is a given difference threshold.

$$D(\mathbf{x}, \mathbf{y}, \mathbf{t}) = |I(x, y, t) - I(x, y, t \pm \Delta)| \tag{2}$$

I(x, y, t) is the value of the pixel on coordinates (x, y) in the *t* th frame of the video sequence, and Delta is interframe distance.

MHI can indicate the speed and direction of motions, and it is very effective for fall detection. Fig.2 shows the contrast of RGB image and MHI image, followed by walking, squatting, stand-falling and sit-falling. By comparison, it can be seen that the difference of MHI between walking, squatting and falling is obvious. The direction of the walk is horizontal in the frames. Although squatting is downward motion, but slower than falling. The moving direction of falling is mainly downward, and it is very fast, which is different from the daily activities.

(a) walking

(a) squatting

(a) stand-falling

(a) sit-falling

Fig. 2. Example of placing a figure with experimental results.

3.3. Fall Classification

Falling is a very common accident, but its speed, direction, and downward characteristics make it distinct from other daily activities. The MHI mentioned in the previous section shows the speed and direction of motion intuitively. In order to give a more accurate feature description of the detected ob-

Table 1. Average precision and recall of different frames.

Frames	MHI-5	MHI-10	MHI-20
Average precision(%)	87.5	96.8	81.2
Average recall(%)	94.3	98.1	91.8

ject, we map the bounding boxes obtained in person detection to MHI, and then extract features from sub-MHI in the boxes instead of extract global features from the whole map. After extracting the features from the sub-MHI, each sub-MHI is classified by SVM. Since a fall is a continuous motion, it does not appear in a single frame or in a small number of frames. We treat the frames as non-falling if the number of continuous frames which are classified as fall by SVM is below the threshold.

4. EXPERIMENTS

4.1. Dataset

Our algorithm is tested on the UR Fall Detection Dataset. This data set is moderately difficult and has a wide range of actions. And Data types are large, including RGB images, depth maps and accelerometer data. We use the part of RGB images only because our algorithm is based on video analysis. This part contains 40 daily activities, such as walking, sitting, squatting, bending and lying down, and also contains 30 simulated falls, including backward falls, forward falls, sideway falls and sit falls. Fig.3 shows examples of some actions in the dataset.

4.2. MHI Frames Setting

As we can see from the introduction of motion history image in the section 3.2, different settings of τ can produce different MHI, which is different frames of action in MHI. As seen in Table 1, the average precision and recall are obtained after 10 cross validation of the different values of the Tao without changing the other conditions, the optimum value is 10. The number of different frames reflect the duration of action. When set the number of frames 5, MHI contains too few motion information that cannot fully express the falling motion. When set it 20, it contain too many frames, in which contain a large part of information other than falling. Thus, 5 and 20 are not ideal settings and 10 is the optimum value.

4.3. Continuous Motion Threshold Setting

After the SVM classifies each target, it is also necessary to filter the consecutive frames. The threshold setting needs to keep recall and precision as high as possible. Due to the particularity of the fall detection, which are the low frequency of occurrence, high real-time requirements and high recall rate

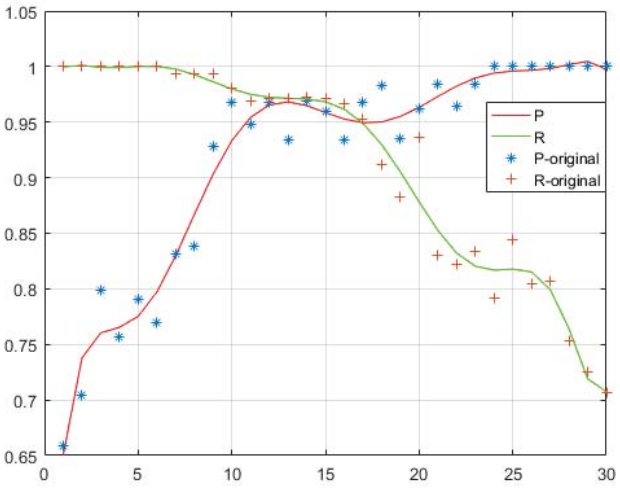

Fig. 4. Continuous Motion Threshold P&R Comparison

Table 2. Definition of TP, FP, FN and TN.

Detection \ Fall	Occurs	Not Occurs
Positive	TP	FP
Negative	FN	TN

requirements, this threshold setting requires considering recall rates preferentially. The recall rates and precision are shown in Fig.4 with the threshold selected from 1 to 30. As we can see from Fig.4, after the threshold 10, the recall rate began to decline rapidly, while the precision rate leveled off. Thus the threshold of continuous motion filter is set to 10.

4.4. Evaluation

To accurately evaluate the method proposed in this paper, we use three criteria which are widely used in the field of event detection and action recognition: Precision, which indicates how many of the positive samples in the forecast is true positive samples, recall, which indicates how many positive samples in all samples are correctly predicted. and F-score, which can balance the effects of precision and recall and comprehensively evaluate a classifier. The definitions of precision, recall and F-score are as follows:

$$P = \frac{TP}{TP + FP} \tag{3}$$

$$R = \frac{TP}{TP + FN} \tag{4}$$

$$F = 2 \times \frac{P \times R}{P + R} \tag{5}$$

where TP, FP, FN and TN are defined in Table 2.

(a) Falls (b) Daily activities

Fig. 3. UR Fall Detection Dataset

Table 3. Average precision, recall and F-score of different methods.

methods	Bounding box ratio analysis	Chua's approach[10]	Human shape and MHI analysis[12]	Ellipse shape analysis [9]	Ours
Average precision(%)	49.9	90.5	88.2	94.3	**96.8**
Average recall(%)	76.9	93.3	83.3	92.8	**98.1**
F-score(%)	60.5	91.9	85.7	93.5	**97.4**
Time per frame(s)	0.12	0.19	-	0.18	0.17

According to the experiments on UR Fall Detection Dataset, the comparison of the experimental results of our method and the results of some other methods is shown in Table 3. Experimental results in Table 3 shows that the proposed method can effectively detect fall events. Our method achieves good results with precision of 96.8% and recall of 98.1%.In time complexity, our method achives 0.17s per frame. As can be seen from the table, although our method is not the fastest one, consider of the highest precision, recall and F-score compared with other methods, our method is an accurate and real-time fall detection algorithm.

5. CONCLUSION

In this work, we propose a fall event detection algorithm with high recall and accuracy. In our proposed method, faster R-CNN can accurately give the target position, and MHI can provide the target speed and direction information. By combining the two advantages, our method performs well in the fall event detection algorithm, which has high recall requirements. In the process of moving target filtering, the threshold is selected by several experiments to obtain the optimum value. The final experimental results are also the average values obtained by cross validation. Experimental results show that, our algorithm is robust, and has high recall and precision in realistic image sequences of simulated falls and daily activities.

In future work, our proposed algorithm can be improved to detect and distinguish different falls by adding action recognition process. Because different fall may have different causes, differentiation and identification of falls can provide more information for the rescue of falls, and thus increase the success rate. In addition, if the fall detection algorithm is used for the elderly who live alone, abnormal event detection can also be derived, such as residential invasion, thereby enhancing the safety of elderly people.

6. REFERENCES

[1] World Health Organization. Ageing and Life Course Unit, *WHO global report on falls prevention in older age*, World Health Organization, 2008.

[2] Chenqiang Gao, Luyu Yang, Yinhe Du, Zeming Feng, and Jiang Liu, "From constrained to unconstrained datasets: an evaluation of local action descriptors and fusion strategies for interaction recognition," *World Wide Web*, vol. 19, no. 2, pp. 265–276, 2016.

[3] Chenqiang Gao, Deyu Meng, Wei Tong, Yi Yang, Yang Cai, Haoquan Shen, Gaowen Liu, Shicheng Xu, and Alexander G Hauptmann, "Interactive surveillance event detection through mid-level discriminative representation," in *Proceedings of International Conference on Multimedia Retrieval*. ACM, 2014, p. 305.

[4] Chenqiang Gao, Yinhe Du, Jiang Liu, Jing Lv, Luyu Yang, Deyu Meng, and Alexander G Hauptmann, "Infar dataset: Infrared action recognition at different times," *Neurocomputing*, vol. 212, pp. 36–47, 2016.

[5] Aaron F. Bobick and James W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 3, pp. 257–267, 2001.

[6] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 886–893.

[7] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[8] Vinay Vishwakarma, Chittaranjan Mandal, and Shamik Sural, "Automatic detection of human fall in video," *Pattern Recognition and Machine Intelligence*, pp. 616–623, 2007.

[9] Homa Foroughi, Baharak Shakeri Aski, and Hamidreza Pourreza, "Intelligent video surveillance for monitoring fall detection of elderly in home environments," in *Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on*. IEEE, 2008, pp. 219–224.

[10] Jia-Luen Chua, Yoong Choon Chang, and Wee Keong Lim, "A simple vision-based fall detection technique for indoor video surveillance," *Signal, Image and Video Processing*, vol. 9, no. 3, pp. 623–633, 2015.

[11] Shengke Wang, Long Chen, Zixi Zhou, Xin Sun, and Junyu Dong, "Human fall detection in surveillance video based on pcanet," *Multimedia tools and applications*, vol. 75, no. 19, pp. 11603–11613, 2016.

[12] Caroline Rougier, Jean Meunier, Alain St-Arnaud, and Jacqueline Rousseau, "Fall detection from human shape and motion history using video surveillance," in *Advanced Information Networking and Applications Workshops, 2007, AINAW'07. 21st International Conference on*. IEEE, 2007, vol. 2, pp. 875–880.

[13] Erik E Stone and Marjorie Skubic, "Fall detection in homes of older adults using the microsoft kinect," *IEEE journal of biomedical and health informatics*, vol. 19, no. 1, pp. 290–301, 2015.

[14] Xin Ma, Haibo Wang, Bingxia Xue, Mingang Zhou, Bing Ji, and Yibin Li, "Depth-based human fall detection via shape features and improved extreme learning machine," *IEEE journal of biomedical and health informatics*, vol. 18, no. 6, pp. 1915–1922, 2014.

[15] Samuele Gasparrini, Enea Cippitelli, Ennio Gambi, Susanna Spinsante, Jonas Wåhslén, Ibrahim Orhan, and Thomas Lindh, "Proposal and experimental evaluation of fall detection solution based on wearable and depth data fusion," in *ICT innovations 2015*, pp. 99–108. Springer, 2016.

[16] Berthold KP Horn and Brian G Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.