FALL DETECTION USING CONVOLUTIONAL NEURAL NETWORK WITH MULTI-SENSOR FUSION

Xu Zhou[1], Li-Chang Qian[1], Peng-Jie You[1], Ze-Gang Ding[1], Yu-Qi Han[1]

1, School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
bitzhouxu@foxmail.com, qlc009@sina.com, ypjaus@163.com, z.ding@bit.edu.cn, yatsihan@126.com

ABSTRACT

In this paper, a fall detection method is proposed by employing deep learning and multi-sensors fusion. Continuous wave radar and optical cameras are used simultaneously to capture human action information. Based on the abstraction of both the microwave and optical characteristics of the captured information, multiple convolutional neural network (CNN) is used to realize the information training and fall action recognition. Due to the fusion of multi-sensor information, the overall performance of the fall detection system can be improved remarkably. Detailed experiments are given to validate the proposed method.

Index Terms— human action, fall detection, radar, optical camera, convolutional neural network, fusion

1. INTRODUCTION

Fall and the resulting injuries, are serious public health threatens for the elderly, especially for those who live alone. The major consequences caused by falls include hip fractures, traumatic brain injuries, and upper limb injuries, etc. [1]-[3]. After falling, the elderly living alone cannot stand up without support from a caregiver. They may lose consciousness and lie on the floor for extended long period of time, which can lead to serious complications, including hypothermia, dehydration, and even death [1]-[3]. Therefore, immediate and effective fall detection is urgently necessary.

In the past decades, many researches focused on developing secure and reliable systems to monitor elderly people’s daily activities including fall events. These systems can be roughly classified into wearable devices and non-wearable devices. Wearable devices such as accelerometers[4]-[5], gyroscopes [6], and wearable cameras [7] can monitor the fall action by connecting to the body. Obviously, the drawback of the wearable devices is that the subjects have to wear the sensors all day, which can be very uncomfortable and inconvenient. In addition, the wearable devices often require the batteries to be charged periodically.[1]

In non-wearable devices, single or multiple sensors such as optical cameras [8], microphones [9], or microwave radars [10], are applied in the target environments to detect fall. However, the performance of optical camera monitors are quite sensitive to light intensity, and the microphones require quiet environments to ensure the detection accuracy. In contrast, microwave radar is much more adaptable to bad environment due to not being performance-affected by light and sound noise. However, the microwave characteristics of human action are not robust enough due to the range and Doppler resolution limits of radar system. To improve the fall detection performance and the environmental adaptability, different kinds of sensors can be used simultaneously to obtain and fuse multi-sensor information. However, by now, multi-sensor system for fall detection based on radar and optimal sensor are seldom mentioned in the existing literatures [11].

In this paper, a new fall detection system based on the combination of optical camera and radar is designed. In the proposed multi-sensor detection system, the short time Fourier transform (STFT) is employed to obtain the time-frequency (TF) micro-motion characteristics in radar [12]-[14] and optical camera is used to acquire the picture sequence of human actions. Furthermore, CNN is adopted for the classification and recognition. Specifically, two kinds of CNNs based on Alex-Net [15] and single shot multi-box detector(SSD) Net [16] are employed to classify the TF features, respectively. The Alex-Net based CNN is a relatively small CNN consists of five convolutional layers, and is quite fit for coarse training and recognition of TF outputs with clear background in radar. And the SSD based CNN is a deeper network using bounding box to train and predict. Besides, by further employing SSD to the picture sequence captured by optical camera, the fall detection performance can be improved due to the calculation of the aspect ratio variation of object bounding box. Furthermore, the detection results of the radar and optimal camera are fused

978-1-5386-1737-3/18/$31.00 ©2018 IEEE

to ensure low false alarm, which makes the fall detection system more efficient and robust.

The remainder of this paper is organized as follows: Section II describes the proposed method and the input data. Besides, the steps of fall detection are given. In section III, the deep learning based approach and data fusion strategy for fall detection is discussed. In section IV, the experimental results are given to demonstrate the effectiveness and robustness of the proposed method. In section V, the conclusion is drawn.

2. THE PROPOSED METHOD

In this section we give a brief introduction to the proposed multi-sensor fall detection system and the description of the detection method.

2.1. System description

Fig. 1 shows the designed multi-sensor fall detection system. The microwave radar operates at 24GHz and is attached at a height of 1.5m from the floor. The linear frequency modulation continuous waves (LFMCW) is transmitted and received by the radar. An optical camera (1920x1080 resolution,60 fps) is also used to capture images of human motions frame by frame.

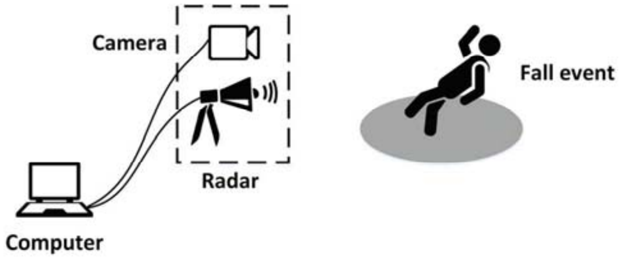


Fig. 1. The proposed fall detection system

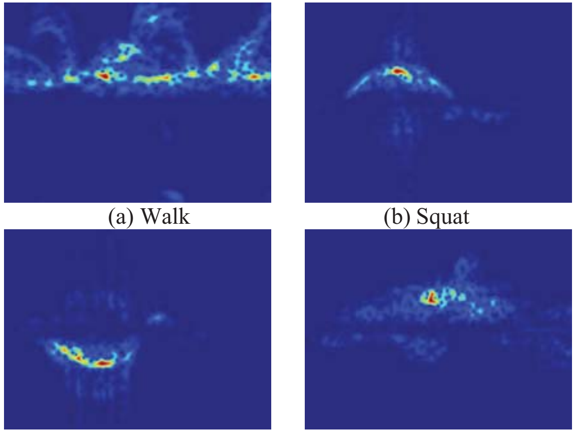
For radar returned signals, the time-frequency (TF) domain is typically used, attributing to its ability to reveal velocities, accelerations, and high-order Doppler information of limbs and various human body parts in motion. Short time Fourier transform (STFT) is traditionally used to analyze the backscattering signals from human subjects [12]-[14], where the TF signals after STFT depicts the distribution of the energy of target echoes varying with time and frequency. The TF signal can be represented as

$$s_{\text{TF}}\left(k,l\right)=\left|\sum_{i=1}^{N_D} s\left(i+\left(k-1\right)\bullet N_D\right)\exp\left(j\frac{2\pi l}{N_D}i\right)\right|, \qquad (1)$$

where $s\left(n\right), n=0,1,2,\cdots,N-1$ is the radar received signal, and N_D is the length of sliding window.

Fig. 2 shows the TF signals of a fall and three common motions: walking, squatting, and standing up. Obviously, the TF signals can be regarded as a kind of images. The other three motions are also chosen because they may be a common cause of fall false alarms. It is obvious that each motion shows different levels of distinction in the TF domains. Walking and standing up are the two most distinguishable motions in the TF domain. Meanwhile, falling and squatting can be confused due to the resemblance of their respective TF features.

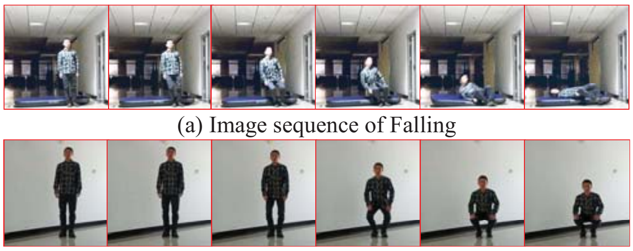
Fig. 3 shows the picture sequence captured by the optical camera. Through CNN processing, the time-varying information can be obtained to help recognize the human motions, which will be discussed in section III.



(a) Walk (b) Squat

(c) Stand up (d) Fall

Fig. 2. TF signals of four human motions:



(a) Image sequence of Falling

(b) Image sequence of Squatting

Fig. 3. Image sequence captured by optical camera

2.2. Steps of the fall detection process

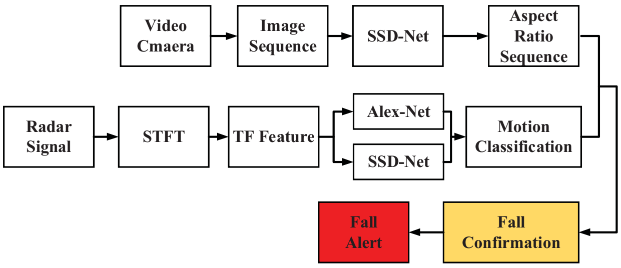


Fig. 4. Fall detection flowchart

The flowchart of the proposed fall detection system is shown in Fig. 4. On the one hand, the radar signals are received and STFT is employed to obtain the TF images of the human motions. Then, the Alex based CNN [15] and SSD based CNN [16] is adopted to classify the TF images respectively, and the results of two CNNs are merged to give the finally detection results. On the other hand, the image sequence captured by optical camera is also processed by another SSD based CNN, and the aspect ratio sequence of the bounding box of the human is provided to help confirm if a fall event is occurring. With the fusion of the radar and optical camera, we can ensure the system detect the fall events more accurately and immediately.

The specific steps of the proposed method for fall detection are described as follow:

Step 1: Receive radar target echoes and image sequence of optical camera.

Step 2: Apply STFT to obtain the TF images.

Step 3: Process the TF images with Alex-Net and SSD-Net, respectively.

Step 4: Merge the results of step 3 and give the final classification results.

Step 5: Process the image sequence with SSD-Net and obtain the aspect ratio sequence of bounding boxes.

Step 6: Fuse the information of step 4 and step 5 to confirm if a fall event is occurring.

3. FALL DETECTION USING DEEP LEARNING

3.1. Training

Three sets of training data are prepared for the training of three CNNs. First, the TF images are warped to 300x300 for the training of Alex-Net. Then, the ground truth bounding boxes are attached on the warped TF images, which is used to train the SSD-Net. In addition, the human motion pictures captured by optical camera is also marked with ground truth bounding boxes to fine tuning the trained SSD-Net downloaded in the GitHub.

3.2. Aspect ratio sequence analysis

We analysis the aspect ratio changing of the bounding boxes frame by frame, and help distinguish the fall and non-fall events. The aspect ratio is defined as

$$Aspect\ Ratio = Height\,/\,Width. \qquad (2)$$

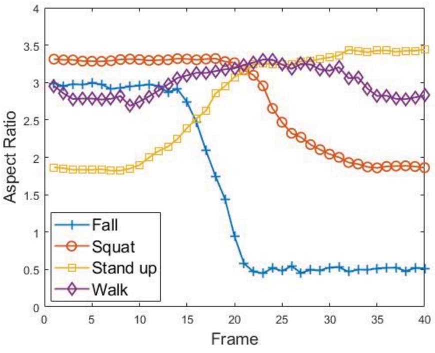
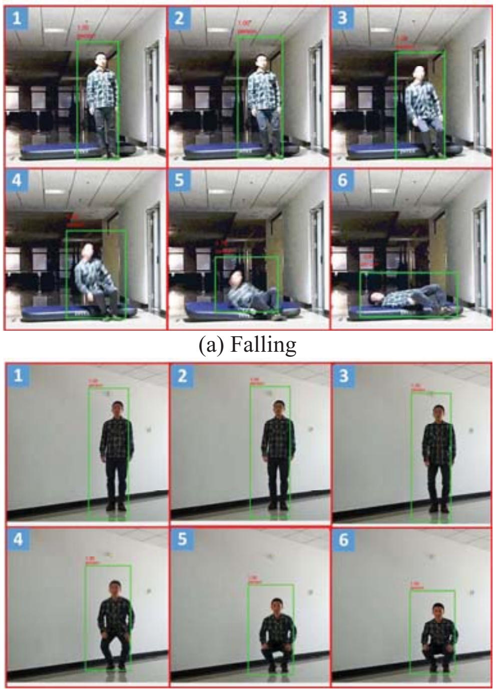


Fig. 5. Aspect ratio analysis of bounding boxes of four human motions



(a) Falling

(b) Squatting

Fig. 6. Classification results with bounding boxes:

As shown is Fig.5 and Fig.6, the aspect ratio changing of falling is the largest among the four human motion, with the value of aspect ratio varying from more than 1 (3.0) to less than 1 (0.5). On the contrary, the curve of walking is quite smooth with little change. As for the squatting and standing up, the curves show declines and increases respectively, but the aspect ratios are still above 1.0. Therefore, utilizing the difference in the aspect ratio sequence of four human motions, the optical camera can help distinguish the fall and non-fall events.

3.3. Data fusion of multi-sensors

Table 1. Data fusion principles

	TF Images		**Optical Camera Images**	**Detection Results after Fusion**
	Alex-Net	**SSD-Net**	**SSD-Net**	
Single CNN Classification Results	Fall	Fall	Fall	Fall
	Fall	Fall	Non-fall	Fall
	Fall	Non-fall	Fall	Fall
	Fall	Non-fall	Non-fall	Non-fall
	Non-fall	Fall	Fall	Fall
	Non-fall	Fall	Non-fall	Fall
	Non-fall	Non-fall	Fall	Non-fall
	Non-fall	Non-fall	Non-fall	Non-fall

Table 1 shows the fusion principles of the proposed fall detection system. The classification results of the TF images play a dominant role in the fall detection. Furthermore, for the falling results, we believe in the results of SSD-Net more than Alex-net. The main reason is that the SSD-Net can utilize the bounding box to precisely learn the detailed difference of the human behaviors in TF domain. Besides, the optical camera results can help the proposed system further confirm the fall events, which improves the accuracy of fall detection as well as reduces the false alarm rate.

4. EXPERIMENTAL RESULTS

The experiments were performed as shown in Fig. 1. The transmitting frequency of the radar is 24GHz, and the frequency modulation cycle is 1000 Hz. The radar signal and the picture sequence is processed by the computer. Two male subjects and one female subject participated in the experiment. The dataset contained four human motions: falling, walking, squatting and standing up. The training dataset of TF images consist of 300 falls and 600 non-falls (200 walking, 200 squatting, 200 standing up), and the fine-tuning dataset for the SSD-Net in the optical camera contain 400 images. As for the test, 1300 radar signals (time duration 1s-2s) and corresponding image sequences (325 falling, 325 walking, 325 squatting, 325 standing up) are prepared. Both the TF images and camera images have same input size (300x300).

Table 2. Experimental results

Method	mAP
Radar only (Alex-Net only)	97.56%
Rada only (Alex-Net+SSD-Net)	99.63%
Radar + Optical camera (Alex-Net+SSD-Net + Aspect ratio analysis)	99.85%

As seen in Table 2, with the fusion of the radar and optical camera system, the improvement of the accuracy is significant. The experimental results demonstrated the superiority of the deep learning based multi-sensor fusion approach.

5. CONCLUSION

In this paper, we proposed a fall detection system based on deep learning and multi-sensor fusion. A continuous wave radar is adopted to obtain the signals of human motions, and the STFT is used to extract the TF features from the radar signals. Besides, an optical camera is also used to capture the images sequence of the human. Two CNNs is trained to classify the TF images, and one CNN is trained to predict the bounding box variation of the image sequence. Then, the fall detection results are given by jointly decision of the three CNNs results. Experimental results demonstrate the effectiveness and robust of the proposed system.

6. REFERENCES

[1] P. Kannus, M. Palvanen, S. Niemi, and J. Parkkari, “Alarming rise in the number and incidence of fall-induced cervical spine injuries among older adults,” *J. Gerontol. A, Med. Sci.*, vol. 62, no. 2, pp. 180–183, Feb. 2007。

[2] M. E. Tinetti, W.-L. Liu, and E. B. Claus, “Predictors and prognosis of inability to get up after falls among elderly persons,” *J. Amer. Med. Assoc.*, vol. 269, no. 1, pp. 65–70, Jan. 1993.

[3] L. Z. Rubenstein, “Falls in older people: Epidemiology, risk factors and strategies for prevention,” Age Ageing, vol. 35, no. 2, pp. ii37–ii41, Sep. 2006.

[4] C.-F. Lai, S.-Y. Chang, H.-C. Chao, and Y.-M. Huang, “Detection of cognitive injured body region using multiple triaxial accelerometers for elderly falling,” *IEEE Sensors J.*, vol. 11, no. 3, pp. 763–770, Mar. 2011.

[5] Y. Liu, S. J. Redmond, N. Wang, F. Blumenkron, M. R. Narayanan, and N. H. Lovell, “Spectral analysis of accelerometry signals from a directed-routine for falls-risk estimation,” *IEEE Trans. Biomed. Eng.*, vol. 58, no. 8, pp. 2308–2315, Aug. 2011.

[6] Q. Li, J. A. Stankovic, M. A. Hanson, A. T. Barth, J. Lach, and G. Zhou, “Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information,” in *Workshop 6th Int. Conf. Wearable & Implantable Body Sensor Networks, IEEE*, June 2009, pp. 138–143.

[7] K. Ozcan, A. K. Mahabalagiri, M. Casares, and S. Velipasalar, “Automatic fall detection and activity classification by a wearable embedded smart camera,” *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 3, no. 2, pp. 125–136, Jun. 2013.

[8] Z.-P. Bian, J. Hou, L.-P. Chau, and N. Magnenat-Thalmann, “Fall detection based on body part tracking using a depth camera,” *IEEEJ. Biomed. Health Inform.*, vol. 19, no. 2, pp. 430–439, Mar. 2015.

[9] Y. Li, K. C. Ho, and M. Popescu, “A microphone array system for automatic fall detection,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 2, pp. 1291–1301, May 2012,

[10] L. Liu, M. Popescu, M. Skubic, M. Rantz, T. Yardibi, and P. Cuddihy, “Automatic fall detection based on Doppler radar motion signature,” in *Proc. 5th Int. Conf. Pervasive Comput.* Technol. Healthcare, Dublin, Ireland, 2011, pp. 222–225.

[11] E. Cippitelli, F, Fioranelli, E. Gambi, and S. Spinsante. “Radar and RGB-depth sensors for fall detection: a review,” *IEEE Sensors Journal*, vol.17, no.12, pp. 3585-3604, Jun. 2017.

[12] X. Shi, F. Zhou, L. Liu, B. Zhao, and Z. Zhang, “Textural feature extraction based on time-frequency spectrograms of humans and vehicles,” *IET Radar, Sonar Navigat.*, vol. 9, no. 9, pp. 1251–1259, 2015.

[13] R. Ricci and A. Balleri, “Recognition of humans based on radar micro Doppler shape spectrum features,” *IET Radar, Sonar Navigat.,* vol. 9, no. 9, pp. 1216–1223, 2015.

[14] L. R. Rivera, E. Ulmer, Y. D. Zhang, W. Tao, and M. G. Amin, “Radar based fall detection exploiting time-frequency features,” in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process. (ChinaSIP)*, Jul. 2014, pp. 713–717.

[15] A. Krizhevsky, I. Sutskever, and G. Hinton. “ImageNet classification with deep convolutional neural networks,” in *NIPS*, 2012, pp. 1097-1105.

[16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed,C. Fu, and A. C. Berg. “SSD: single shot multibox detector,” In *ECCV*, 2016, pp. 21–37.