ATLANTIS PRESS

Advances in Engineering Research (AER), volume 116

International Conference on Communication and Electronic Information Engineering (CEIE 2016)

Falling-action analysis algorithm based on convolutional neural network

Wei Liu[a], Jie Guo[b], Zheng Huang[c] and Dong Qiu[d]

School of Information Security and Engineering, Shanghai Jiao Tong University, Shanghai, China

Email: [a]Vincent_Liu@sjtu.edu.cn, [b]guojie@sjtu.edu.cn, [c]huangzhengsjtu@126.com, [d]qiuwd@sjtu.edu.cn

This paper proposes a deep learning method – convolutional neural network to analyze human falling-action in video surveillance, so that we can recognize the falling-action of human body accurately in the shortest time. Firstly, vibe algorithm is used to extract the foreground and some methods of image preprocessing are employed to optimize the moving target. Then the moving target is fed into the convolutional neural network which extracts the features of various actions (including sitting, crouching, bending, falling) and classifies these actions. It is proved by experiments that our method is accurate and competitive compared with the current method to falling-action recognition.

Keywords: Falling-Action Analysis; Deep Learning; Convolutional Neural Network.

1. Introduction

An undeniable truth is that the aging problem is getting more and more serious in recent years, it bears no delay to solve puzzles like who should take responsibility for the accidental falling of the old or the falling of the elderly living alone at home.

At present, there are three kinds of solutions to the falling action detection. The first one is that we can detect the falling-action by the wearable sensor-based which can extract the physiological signals of the body and get the acceleration of the human body [1]. Then about the second one, some researchers make the judgments by the sound collected by the sensor [2] when the subject in the study falls down. The former is inconvenient and uncomfortable for people to wear the clothes with the sensor and the later can't get a better accuracy with the complexity of audio source and category.

Finally, the falling-action can be detected on the basis of the video surveillance. In order to conquer the drawbacks of the above mentioned systems, camera-based system is used to distinguish the behavior by extracting the

silhouette area features, shape features, posture features and so on [3] –[5]. For example, in [3], Yu et al. proposed a method which detected the falling behavior by the features of silhouette area and used the SVM (support vector machine) [6] to train and classify the action. By these methods, the computation complexity of image is high and the accuracy of classification is still a problem to solve.

This paper mainly consists of two parts. In the first part, the vibe is used to extract the foreground and we can get the ratio of the height and width of the target to comprehend the abnormal behavior which is likely to fall, such as sitting, crouching, bending. In the second part, the convolutional neural network is employed to extract the features of image and classify the action. The reason why we choose the convolutional neural network to analyze the falling-action is that it can get richer and more useful information by extracting the local features and we can get better classification result by training larger image sets.

2. Falling-Action Detection Algorithm

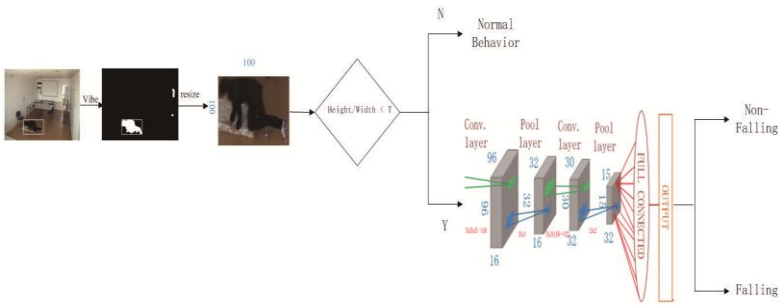

Fig. 1 Framework of falling-action detection algorithm

This detailed process about our algorithm is shown in Figure 1. Our algorithms consist of extracting and optimizing moving target, feature extraction, training and classification. Firstly, we detect the accessed real-time video stream by the vibe [7], extracting the foreground moving target, and then take a series of image processing methods. Secondly, the ratio of the height and width of the minimum circumscribed rectangle of human body is extracted, and we compare it with proper threshold to judge whether someone is likely to fall or non-fall. Lastly, we use convolutional neural network [10] to distinguish the falling and non-falling.

2.1. *Moving target detection*

The vibe is employed to perform background subtraction frame by frame and the basic idea of vibe [8] is that vibe builds a model for each sample and each model

contains *N* pixels. The schematic diagram of the vibe idea is shown in Figure 2. Taking an example of *p5*, whether the *p5* is foreground or background is up to the similarities between the p5 and the background model. The algorithm is as follows:

(i)Calculate the Euclidean distance *D* between *p5* and each point of each model.

(ii)Compare *D* with the threshold *R*. Record the number *n* of the point if $D < R$.

(iii)Compare *n* with the threshold *N*. If $n > N$, *p5* is the background.

Therefore, what affects the accuracy of building model is the threshold area radius *R* and the sample number *N* within the threshold range. It is proved by experiments that we can achieve good effects when $R = 20$ and $N = 2$.

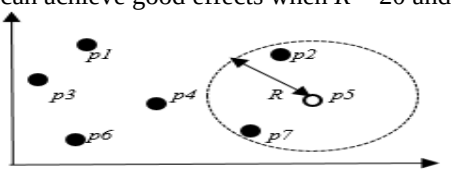

Fig. 2 Vibe model

The realization of the vibe algorithm is mainly divided into two parts: initialization and update [9]. The vibe is initialized by single frame and the update strategy calculates the similarity between each sample and sample in background to judge whether the sample is foreground or background.

2.2. *Classification*

The convolutional neural network [10] is a variant structure of multilayer perception machine, which regards the raw image as the network input. Its kennel idea framework is local receptive field, shared weights and pooling. What the meaning of the local receptive field is that the neurons of the neural network only need to perceive the local image, and we can get the global information by gathering the local information. The shared weights means that if image's characters are no difference, we will share the learning feature of images. And the pooling is the operation which converges and counts the features of different area.

We can see from Figure 1 that our convolutional neural network consists of two convolution layers, two pooling layers and a fully connected layer. In the convolutional layer, back-propagation pass is used to update the weights of the neurons, a learnable convolution kernel execute a convolution to the feature maps of the previous layer, then we can get the output feature maps by a

activation function. Every output maps may be the combinations of a few input maps:

$$x_j^l = f(\sum_{i \epsilon M_j} x_i^{l-1} \cdot k_{ij}^l + b_j^l) \quad (1)$$

M_j is the set of input maps, the extra offset b is given to every output map, k is the convolutional kernel.

In the max-pooling layer, there are N input maps and N output maps, but the output maps get smaller than input maps:

$$x_j^l = f(\beta_j^l down(x_i^{l-1}) + b_j^l) \quad (2)$$

down(.) is a down sampling function which reduces the output image N times in two dimensions. Each output map has its multiplicative bias β and an additive bias b.

3. Experiment Evaluation and Analysis

All training datasets and test datasets are from [11] and UR Fall Detection Dataset [12]. The number of training samples is 602 and 392 test images are absolutely different from the training images. As shown in Figure 1, the input images with the size of 100 * 100 are the input of the first convolution layer, 16 learnable filters of size 5 * 5 * 3 are used to extract feature maps and we can get the resulting maps of size 96 * 96. Then, the feature maps are passed through the max-pooling layer of size 3 * 3 that the height and width are reduced by three 3 times. The theory of the following process is the same as the former.

We analyzed the accuracy of SVM classifier and convolutional neural work classifier. The results are obtained from table 1 that the classification accuracy of the SVM is far worse than the convolutional neural network, the reason why the CNN works better than the SVM is that the CNN extracts local features of the image. The more local features are, the more the image's content and details are.

Tab. 1 Classification accuracy between SVM and CNNs

	SVM	CNNs
Fall	0.77	0.85
Non-fall	0.47	0.96
Accuracy	0.66	0.89

To prove that our method can achieve a better performance, its indicators are shown in table2. There is a brief introduction to the key terms before we learn about the performance of CNNs:

$$\text{precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{recall} = \frac{TP}{TP + FN} \tag{4}$$

TP(true positive) which is predicted as true is true in fact, and the same goes to FP and FN. Precision, recall and f-score are used to measure the performance of our algorithm. The recognition ability of positive sample and distinction capability of the false are represented by recall and precision. Besides, the f-score is the comprehensive description of precision and recall. Obviously, the three values in table 2 are excellent and the algorithm we proposed is suitable to detect falling-action.

Tab. 2 Performance of CNNs

	Precision	Recall	F-score
Fall	0.85	0.98	0.91
Non-fall	0.96	0.68	0.79
Average	0.89	0.88	0.87

4. Summary

In this paper, we propose a method of deep learning – convolutional neural network in the classification of falling-action. The vibe algorithm is used to extract moving target and we can get the optimized foreground image by preprocessing. After these steps the features of the target from the accessed real-time video stream are extracted and classified by the convolutional neural network. Comparing with the support vector machine, we find the convolutional neural network has better performance than the SVM method in the problem of falling-action classification.

Acknowledgments

This research was financially supported by National science and technology support plan (2014BAK06B02).

References

1. Federico Bianchi, Stephen J Redmond, Michael R Narayanan, etc. Barometric pressure and triaxial accelerometry-based falls event detection. Neural Systems and Rehabilitation Engineering, IEEE Transactions, Vol. 18, No. 6, 2010, pp. 619-627.

2. Xiaodan Zhuang, Jing Huang, Gerasimos Potamianos, etc. Acoustic fall detection using gaussian mixture models and gmm supervectors. In Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference (2009), pp. 69-72.
3. Behzad Mirmahboub, Shadrokh Samavi, Nader Karimi, etc. Automatic Monocular System for Human Fall Detection Based on Variations in Silhouette Area. IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, Vol. 60, No. 2, 2013, pp. 427-436.
4. C. Rougier, J. Meunier, A. St-Arnaud, etc. Robust video surveillance for fall detection based on human shape deformation. IEEE Trans. Circuits Syst. Video Technol., Vol. 21, No. 5, 2011, pp. 611–622.
5. Yu, Miao, Rhuma, Adel, Naqvi, Syed Mohsen, etc. A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment. IEEE Transactions on Information Technology in Biomedicine, Vol. 16, No. 6, 2012, pp. 1274-1286,.
6. Yong Hou, A Novel SVM Algorithm and Experiment. International Conference on Computer Science and Electronics Engineering (2012), pp. 31-35.
7. Qin Yinshi, Sun Shuifa, Ma Xianbing, etc. A background extraction and shadow removal algorithm based on clustering for ViBe. Proceedings - International Conference on Machine Learning and Cybernetics, Vol.1 (2015), pp. 52-57.
8. Hernandez, Steven Diaz, De LaHoz, etc. Dynamic background subtraction for fall detection system using a 2D camera. 2014 IEEE Latin-America Conference on Communications (2014), pp. 1-6.
9. Kumar Kshitij, Agarwal Suneeta. A hybrid background subtraction approach for moving object detection. IET Conference Publications (2013), pp. 392-398.
10. Ijjina, Earnest Paul, Mohan, etc. Human action recognition based on recognition of linear patterns in action bank features using convolutional neural networks. Proceedings – 2014 13th International Conference on Machine Learning and Applications (2014), pp. 178-182.
11. Imen Charfi, Johel Miteran, Julien Dubois, etc. Definition and Performance Evaluation of a Robust SVM Based Fall Detection Solution. Signal Image Technology and Internet Based Systems (SITIS), 2012 Eighth International Conference (2012), pp. 218-224.
12. Bogdan Kwolek, Michal Kepski. Human fall detection on embedded platform using depth maps and wireless accelerometer. Computer Methods and Programs in Biomedicine, Vol. 117, No. 3, 2014, pp. 489-501.