A Review of Multiple-Person Abnormal Activity Recognition

Yangyue Zhou and Miaolei Deng
College of Information Science and Engineering, Henan University of Technology, Zhengzhou, China
Email: 564185081@qq.com, dengmiaolei@haut.edu.cn

***Abstract*—In recent years, computer vision technology capable of detecting human behavior has attracted more and more attention. Although it has been widely used in many applications, accurate and effective human motion recognition is still a challenge in the field of computer vision. This paper presents a review of the latest research methods for multi-person human anomalous action recognition. The article deeply analyzes the calculation method of activities and briefly describes popular datasets. Discussing the unresolved issues, it provides new ideas for future research.**

***Index Terms*—video surveillance, human action recognition, action detection**

I. Introduction

The field of computer vision and pattern recognition involves target detection, tracking, activity recognition, image fusion and so on, among which human behavior detection is one of the most interesting problems. This is because it is the basic application in many fields, such as intelligent video surveillance and environmental home monitoring [1], [2], human-computer interaction [3], human pose estimation, human tracking, image or video annotation, and identity recognition [4], etc. At first, people's research focused on the analysis of single individuals, but now turned to the analysis of groups [5].

The key to determine a good human action recognition is robust human action modeling and feature representation. Feature representation and selection is a classic problem in computer vision and machine learning [6]. The feature representation of human motion in video is very different from that in image space. It not only describes the appearance of human in image space, but also extracts the change of appearance and posture. The problem of feature representation has been extended from two-dimensional space to three-dimensional space. In recent years, researchers have proposed a variety of motion representation methods, including global and local features based on temporal and spatial changes [7]-[9], trajectory features based on key point tracking [10], [11], motion changes based on depth information [12]-[14], and action features based on human posture changes [15], [16]. As some researchers have successfully applied deep learning to image classification and target detection, many researchers have also applied deep learning to the field of human motion recognition.

Through activity recognition, the operations involved in a specific scene can be identified. The information is obtained by observing the action and environmental conditions of the target human object. Human activities can be divided into two categories: single person behavior recognition and multi-person behavior recognition. Researchers have applied various existing methods to abnormal activity recognition.

In this paper, the research status of multi-person abnormal motion recognition technology is reviewed, including motion detection methods and behavior datasets. The arrangement of the article is as follows. In the second section, we summarize the related work in the field of abnormal human behavior recognition, which will help readers understand the main contributions of previous surveys. In the third section, we summarize the datasets which have been published in recent years. Finally, the fourth section focuses on the special observation and possible research direction of abnormal activity identification, which provides further research for the research in the field of abnormal human activity identification.

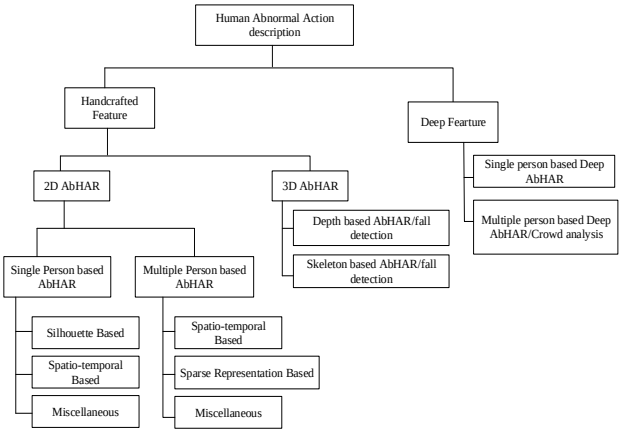

Figure 1. Taxonomy of Abnormal Human Activity Recognition (AbHAR).

II. Multiple Persons Abhar

A. Spatio-Temporal Based Approaches

Singh and Mohan [17] used graph formula and graphics kernel support vector machine of video activity.

Manuscript received November 3, 2020; revised March 5, 2021.

doi: 10.18178/joig.9.2.55-60

In the video, the interaction of entities is represented as the geometric relation graph between spatiotemporal interest points. The vertex of the graph is the spatiotemporal interest point, and the edge is the relationship between the appearance and the dynamic around the interest point. In order to classify activities into normal types or abnormal types, they use binary support vector machine with graph kernel. These graphical kernels provide robustness to slight topological deformation when comparing two graph markets, which may be caused by noise in the data. Kerola *et al*. [18] proposed another graph based method, which regards actions as sequences of graphs. This method is 3D invariant and suitable for single view and multi view human activity recognition. [19] Another method proposed by Chong *et al*. successfully applied spatiotemporal techniques to abnormal activity recognition.

The human body is a 3D space-time surface that can perform certain activities in video. Bakheet and Al Hamadi [20] proposed a variant of Histogram of Oriented Gradients (HOG) feature which is a fuzzy Histogram of Oriented Lines (f-HOL). This new feature has no effect on small geometric transformations and illumination changes. Wang *et al*. [21] decomposed each video into short videos, and each segment was represented by local spatiotemporal statistics using visual word bag. Unlike the generalized time warping, which is equally important for different parts of the sequence of interest, they proposed a temporally-weighted GTW (TGTW) algorithm, which aims at the early part of incomplete active focus video.

Spatiotemporal trajectories are used to retrieve local information. A technique extracts features corresponding to dense trajectories to represent shape, appearance and motion information [22]. It introduces a new feature descriptor to extract moving boundary histogram and other information. In order to deal with the problem of unstructured camera movement, a method based on feature trajectory is proposed [23]. In order to explore the spatiotemporal relationship between trajectories, Zhang *et al*. [24] developed a technique. This descriptor encoding method is based on the relationship between space, time and feature foreground of different trajectories. It can create the distance parameter of perceptible offset to solve the defocusing problem of dense tracks, so as to achieve more accurate trajectory matching.

Singhal *et al*. (2018) [25] proposed an action recognition algorithm based on local binary pattern (LBP). LBP features were extracted by spatiotemporal relationship, and the normalized features were classified by random forest classifier. At the same time, the author also focuses on reducing the descriptor value by standardizing the calculated histogram box. The implementation of this method can be used for action recognition in small places such as ATM room. The results show that the standardized version of LBP exceeds the traditional LBP descriptor, with an average accuracy of 83%. In 2019, Rodrigues *et al*. [26] proposed a multi time scale model to capture the time dynamics under different time scales. In particular, for a given input attitude trajectory, the model predicts the future and the past in different time scales. The model is multi-layered, and the middle layer is responsible for generating predictions corresponding to different time scales, and combining these predictions can detect abnormal activities. Rodrigues also provides an abnormal activity data set for research, which contains 483566 annotated frameworks (https://rodrigues-royston.github.io/Multi-timescale_Trajectory_Prediction/).

B. Sparsse Representation Based Approaches

Sparse model provides descriptor with unique recognition ability, which makes sparse representation become a common method of abnormal behavior detection. In 2016, Li *et al*. [27] defined the Histogram of Maximal optical Flow Projection (HMCFP) descriptor, which was used to optimize the dictionary and calculate the "L1" norm of the Sparse Reconstruction Coefficient (SRC), so as to detect crowd anomalies in the test frame. The combination of spatiotemporal features and Laplacian sparse representation was reported by Zhao *et al*. [28] in 2015 to identify the normalized combination vector HNF (HOG+ HOF).

By Laplacian sparse representation and maximum pooling, the minimum feature error is more descriptive. It is worth noting that the detection accuracy of the proposed method is 93% for both global and local abnormal activities in UMN dataset. For trajectory Sparse Reconstruction Analysis (SRA) (Li *et al*., 2012 [29]), the control point features of cubic B-spline curves are extracted from normal motion trajectory sequences to establish normal dictionary set. In the test phase of abnormal event detection, sparse linear reconstruction coefficients and residuals are helpful to judge normal or abnormal events. However, this method has limitations, one of which is that its performance is highly sensitive to trajectory control point parameters. Liu *et al*. Recently (2017) [30], using dual sparse representation and dynamic dictionary update mechanism to observe crowd movement, which dynamically increases the size of dictionary for a small set of training samples. Dominant Set (DS) has become a powerful technique for detecting any abnormal behavior in an unsupervised framework (Alvar *et al*., 2014 [31]). It provides a locally adaptive boundary to represent the unknown data points sparsely. Experimental results show that compared with other clustering algorithms (KNN, Gaussian mixture, fuzzy k-means), DS produces the smallest overall error rate.

C. Depth Based Approaches

In 2018, Tripathi *et al*. [32] reported the evolution of detection methods for abnormal events in crowded scenes from shallow to deep. The survey emphasized four attributes of the crowd: crowd counting, crowd movement detection, crowd tracking and crowd behavior understanding. At present, group analysis has always been an important topic in the research of public domain. In the face of realistic challenges such as occlusion and messy background in crowded scenes, if the manual method does not provide a solution, it provides a better

solution in the depth model (i.e., CNN, LSTM, automatic encoder, RNN).

In order to minimize the severity of continuing injuries caused by attack related violence, this can be addressed by reducing detection time. In 2017, kaelon *et al*. [33] proposed an automatic abnormal crowd detection method, which can be used to detect violence in public places. Crowding often occurs in public places, which leads to personal behavior being blocked by other people. In order to solve this problem, kaelon also proposes a real-time descriptor, which uses the time summary of Gray Level Co-occurrence Matrix (GLCM) features to simulate group dynamics by encoding the change of group texture. The author also introduces a measure of Inter-Frame Uniformity (IFU) and proves that compared with other types of crowd behavior, the appearance change of violence is not consistent. Experimental results show that the method proposed by kaelon has low computational cost and high receiver performance score (0.9956 on UMN dataset).

In 2019, Lei *et al*. [34] proposed a multi-analysis method for human abnormal behavior in complex scenes. Firstly, by using the similarity measure of social force model, the abnormal behavior is roughly distinguished from the large monitoring area, and then it is analyzed accurately. Based on the multi sum analysis of three-frame difference algorithm, it is used for intrusion detection, left luggage detection and trajectory recognition. This method has some advantages on UMV and CAVIAR datasets. The experimental results show that the method is better than the existing methods.

In 2019, Gao [35] and others proposed a video sequence feature extraction algorithm based on particle filter to give early warning when abnormal events occur. The whole process includes feature sequence generation and particle filter tracking. In order to represent the features of video, an L2 norm extractor based on optical flow is designed. The particle filter then tracks these feature sequences. The occurrence of abnormal events will result in the offset of feature sequence and the tracking error of PF, otherwise, it will allow the computer to understand and define the occurrence of anomaly. The experimental results show that the accuracy of the algorithm in frame level detection reaches 90%.

There are more and more abnormal phenomena in the indoor and outdoor (these abnormal phenomena may be theft, destruction of public property, or even attack the innocent), which requires an accurate and robust action recognition system. Aiming at these phenomena, in 2020 omnia people [36] proposed a new algorithm based on machine learning paradigm to detect human behavior and mark it as normal or abnormal. The algorithm first tests two different human detectors (cascaded target detector and Faster Region Convolutional Neural Network for Human Detection (frcnnhd)). Two detectors were trained using widely available datasets. After that, the detected human body image is extracted to form a video patch representing human motion. In the process of action recognition, the author uses the motion history image to extract the static features of motion, and then uses Support Vector Machine (SVM) to classify the action. Finally, the action with low recognition is marked as "abnormal behavior".

In order to solve the problems of complex background, complex geometric changes and huge amount of data in video human action recognition, in 2019, Dinesh *et al*. [37] proposed an algorithm to identify human behavior in video by using a certain pose. Firstly, a key attitude is extracted by optical flow, and then the feature is extracted by wavelet dual transform (here, Gabor Wwavelet Transform (GWT) and Ridgelet Transform (RT) are used for secondary transform). GWT generates feature vectors by calculating the first-order statistical values of different scales and directions of the input pose, which are robust to translation, scaling and rotation, and uses RT to calculate the direction dependent shape features of human actions. The fusion of these functions provides a reliable and unified algorithm. The accuracy of the algorithm is 96.66%, 96%, 92.75% and 100% respectively on KTH, Weizmann, Ballet-movement and UT-interaction data sets, which shows superior performance compared with other similar latest technologies.

In 2019, Tay *et al*. [38] proposed a CNN based abnormal behavior detection method. This method can automatically learn the most discriminative features related to human behavior from a large number of videos containing normal and abnormal behaviors. His method is an end-to-end solution that can be used to deal with abnormal behaviors under different conditions, including background changes, the number of experimenters (individuals, two people or groups) and a series of different abnormal human activities. In UMI, UTI, HOF, wed and pel, the accuracy rate reaches 100%.

III. DATASETS

With the development of new technologies, the number and content of public datasets for experiments have increased dramatically. Table I gives the details of datasets.

A. UCF-Crime Dataset

In 2018, the UCF-crime dataset constructed by Waqas Sultani *et al*. [39] is a large-scale real-world surveillance video data set. The data set contains 13 kinds of abnormal behaviors that have a significant impact on public safety, including arson, assault, robbery, theft, shooting, explosion, traffic accidents, etc. The dataset contains 1900 long uncut videos, including 1610 training videos and 290 testing videos.

B. ShanghaiTech Dataset

In 2017, W. Luo *et al*. [40] constructed the shanghaitech data set, which is a medium-sized data set, consisting of 437 videos (330 training videos and 107 testing videos), including 130 abnormal events in 13 scenes. The training video and test video cover 13 scenes.

C. UCSD Dataset

In 2010, Mahadevan *et al*. [41] constructed the dataset, which consists of two parts: the UCSD pedestrian 1 (ped1)

dataset and the UCSD pedestrian 2 (ped2) dataset. The UCSD ped1 dataset contains 34 training video clips and 36 testing video clips, as well as 40 unscheduled events, each of which contains 200 frames. All of these anomalies are related to vehicles such as bicycles and cars. The UCSD pedestrian 2 (ped2) dataset contains 16 training videos and 12 test videos with 12 abnormal events. The number of frames in each segment is different. The video consists of pedestrians parallel to the plane of the camera.

D. Avenue Dataset

The avenue dataset was constructed by C. Lu *et al*. [42] in 2013 and contains a total of 37 videos (16 training videos and 21 testing video clips). The video duration of each clip is within one minute and between one and two minutes. Normal scenes include people walking between stairs and subway entrances, while abnormal events are abnormal events such as running, walking in the opposite direction, wandering, etc., and they are staged and captured in one place. In addition, there are few normal patterns in training data.

E. CASIA Dataset

CASIA dataset [43] is a collection of video sequences of human activities captured by different cameras from different perspectives. It is composed of interactive scenes shot by two volunteers. It contains the five most common interactions: 1. Fighting. 2. One person surpasses another. 3. Robbery. 4. One person follows another. 5. Meet and leave. All the videos were shot simultaneously from different perspectives using three static uncalibrated cameras. The views were oblique view, horizontal view and top view. The frame rate of video is 25 frames per second, the frame size is 320 × 240 pixels, and the duration of different activities is 5-30 seconds.

F. IXMAS Dataset

INRIA Xmas Motion Acquisition Sequences (ixmas) dataset [44] is a human activity recognition dataset with constant perspective. In this dataset, 13 activities of daily living were performed by 11 actors three times each. These activities include: nothing, check watch, cross arms, scratch head, sit down, get up, turn around, walk, wave, punch, kick, point, pick up, throw (over head), and throw (from bottom up). The frame rate is 23 frames per second and the frame size is 390×291.

TABLE I. COMPARISON OF VARIOUS DATASETS

Dataset	Resolution	Frame rate	videos	Duration (s)
UCF-Crime	240×320	30	1900	-
UCSD Dataset	238×158	10	138	-
CASIA Dataset	320×240	25	1446	5-30
IXMAS Dataset	390×291	23	1800	1-5

IV. CONCLUSION

In this paper, the research on multi-person abnormal behavior recognition is comprehensively reviewed, and the research methods of action recognition in recent years are summarized, including the methods based on space-time, sparse representation and deep learning. Although the research on human behavior detection has been developed, there are still some problems: 1. The shape and size of objects vary with different frames. 2. Occlusion. 3. Noise and Blur. 4. Brightness and intensity changes. 5. Object's abrupt motion. 6. Projection of 3D world into 2D space. 7. Real time scenario analysis requirements.

In the current research, accurate depth information and bone data can effectively study the human motion features. However, in most real scenes, the data collection platform can only provide RGB data. In the monitoring scene, the depth sensor is not suitable for accuracy and cost. Therefore, on the basis of RGB data, depth data and skeleton data, the integration of multimodal data is a key issue in the research of behavior recognition.

In the aspect of interaction recognition, the interaction between people and objects has a high degree of semantic information (such as carrying dangerous goods, leftover goods and waving hand-held weapons, etc.). Based on multimodal data, modeling the interaction between people and objects, and quickly analyzing the interaction information, has not reached the appropriate accuracy. This is an important direction of human behavior recognition research in the future.

2020 is a special year, with the epidemic spreading all over the world, causing panic and riots among people in various regions (a large number of goods were snapped up in supermarkets; fires were set in front of the government, guns were held, vehicles smashed, demonstrations were held, etc.). It is more difficult to detect all the special items from head to foot. The tracking, detection and identification of wearing protective equipment in government and sensitive areas is also an important research direction in the future, and the establishment of special data sets will become the primary task. At the same time, fast and accurate motion detection is also the key to the success of human motion recognition.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Yangyue Zhou conducted the research, analyzed the data and wrote the paper; Miaolei Deng guided the writing ideas and revision of the paper; all authors had approved the final version.

ACKNOWLEDGMENT

Grateful acknowledgement is made to all the researchers mentioned in the references. And we would like to thank the anonymous reviewers and editors for the constructive and insightful comments for improving the quality of this paper.

REFERENCES

[1] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," *ACM Computing Surveys*, vol. 43, p. 16, 2011.

[2] M. Ziaeefard and R. Bergevin, "Semantic human activity recognition: A literature review," *Pattern Recognit.*, vol. 48, pp. 2329-2345, 2015.
[3] L. L. Presti and M. L. Cascia, "3D Skeleton-based human action classification: A survey," *Pattern Recognit.*, vol. 53, pp. 130-147, 2016.
[4] S. N. Paul and Y. J. Singh, "Survey on video analysis of human walking motion," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 7, pp. 99-122, 2014.
[5] M. Shah and R. Jain, *Motion-Based Recognition*, Berlin: Springer, 2013, vol. 9.
[6] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, pp. 1798-1828, 2013.
[7] D. D. Dawn and S. H. Shaikh, "A comprehensive survey of human action recognition with Spatio-Temporal Interest Point (STIP) detector," *The Visual Computer*, vol. 32, pp. 289-306, 2015.
[8] T. V. Nguyen, Z. Song, and S. C. Yan, "STAP: Spatial-Temporal attention-aware pooling for action recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, pp. 77-86, 2015.
[9] L. Shao, X. T. Zhen, D. C. Tao, and X. L. Li, "Spatio-Temporal Laplacian pyramid coding for action recognition," *IEEE Transactions on Cybernetics*, vol. 44, pp. 817-827, 2013.
[10] G. J. Burghouts, *et al.*, "Instantaneous threat detection based on a semantic representation of activities, zones and trajectories," *Signal Image and Video Process*, vol. 8, pp. 191-200, 2014.
[11] H. Wang and C. Schmid, "Action recognition with improved trajectories," in *Proc. the IEEE International Conference on Computer Vision*, Sydney, NSW, Australia, December 2013, pp. 3551-3558.
[12] X. Yang and Y. L. Tian, "Super normal vector for activity recognition using depth sequences," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, June 2014, pp. 804-811.
[13] M. Ye, Q. Zhang, L. Wang, J. Zhu, R. Yang, and J. Gall, "A survey on human motion analysis from depth data," in *Time-of-Flight Depth Imaging. Sensors, Algorithms, and Applications*, Springer Berlin Heidelberg, 2013, pp. 149-187.
[14] O. Oreifej and Z. Liu, "HON4D: Histogram of oriented 4D normals for activity recognition from depth sequences," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, USA, June 2013, pp. 716-723.
[15] M. Li, H. Leung, and H. P. H. Shum, "Human action recognition via skeletal and depth based feature fusion," in *Proc. the 9th International Conference on Motion in Games*, Burlingame, CA, USA, October 2016, pp. 123-132.
[16] X. Yang and Y. L. Tian, "Effective 3D action recognition using eigenjoints," *Journal of Visual Communication & Image Representation*, vol. 25, pp. 2-11, 2014.
[17] D. Singh and C. K. Mohan, "Graph formulation of video activities for abnormal activity recognition," *Pattern Recognition: The Journal of the Pattern Recognition Society*, vol. 65, pp. 265-272, 2017.
[18] T. Kerola, N. Inoue, and K. Shinoda, "Cross-View human action recognition from depth maps using spectral graph sequences," *Computer Vision and Image Understanding*, vol. 154, pp. 108-126, 2017.
[19] Y. S. Chong and Y. H. Tay, "Abnormal event detection in videos using spatiotemporal autoencoder," in *Proc. International Symposium on Neural Networks*, Springer, Cham, 2017, pp. 189-196.
[20] S. Bakheet and A. Al-Hamadi, "A discriminative framework for action recognition using f-HOL Features," *Information*, vol. 7, p. 68, 2016.
[21] H. Wang, W. Yang, C. Yuan, H. Ling, and W. Hu, "Human activity prediction using temporally-weighted generalized time warping," *Neurocomputing*, vol. 225, pp. 139-147, 2016.
[22] H. Wang, A. Klaser, C. Schmid, and C. Liu, "Dense trajectories and motion boundary descriptors for action recognition," *International Journal of Computer Vision*, vol. 103, pp. 60-79, 2013.
[23] S. Singh, C. Arora, and C. V. Jawahar, "Trajectory aligned features for first person action recognition," *Pattern Recognition*, vol. 62, pp. 45-55, 2017.
[24] L. Zhang, Z. Wang, T. Yao, T. Mei, and D. D. Feng, "Exploiting spatial-temporal context for trajectory based action video retrieval," *Multimed. Tools Appl.*, vol. 77, no. 2, 2018.
[25] S. Singhal and V. Tripathi, "Action recognition framework based on normalized local binary pattern," in *Progress in Advanced Computing and Intelligent Engineering*, 2019, vol. 1, pp. 247-255.
[26] R. Rodrigues, N. Bhargava, R. Velmurugan, and S. Chaudhuri, "Multi-timescale trajectory prediction for abnormal human activity detection," in *Proc. IEEE Winter Conference on Applications of Computer Vision*, 2020.
[27] A. Li, Z. Miao, Y. Cen, and Q. Liang, "Abnormal event detection based on sparse reconstruction in crowded scenes," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Shanghai, 2016.
[28] Y. Zhao, Y. Qiao, J. Yang, and N. Kasabov, "Abnormal activity detection using spatio-temporal feature and Laplacian sparse representation," in *Proc. International Conference on Neural Information Processing*, 2015.
[29] C. Li, Z. Han, Q. Ye, and J. Jiao, "Visual abnormal behavior detection based on trajectory sparse reconstruction analysis," *Neurocomputing*, vol. 119, pp. 94-100, 2012.
[30] P. Liu, Y. Tao, W. Zhao, and X. Tang, "Surveillance scene segmentation based on trajectory classification using supervised learning," *Neurocomputing*, vol. 269, pp. 3-12, 2017.
[31] M. Alvar, A. Torsello, A. S. Miralles and J. M. Armingol, "Abnormal behavior detection using dominant sets," *Machine Vision Applications*, vol. 25, pp. 1351-1368, 2014.
[32] G. Tripathi, K. Singh, and D. K. Vishwakarma, "Convolutional neural networks for crowd behaviour analysis: A survey," *The Visual Computer*, pp. 1-24, 2018.
[33] K. Lloyd, P. L. Rosin, D. Marshall, and S. C Moore, "Detecting violent and abnormal crowd activity using temporal analysis of Grey Level Co-occurrence Matrix (GLCM) based texture measures," *Machine Vision Applications*, vol. 28, pp. 361-371, 2017.
[34] L. Cai, P. Luo, and G. Zhou, "Multistage analysis of abnormal human behavior in complex scenes," *Journal of Sensors*, pp. 1-10, 2019.
[35] X. Gao, G. Xu, S. Li, Y. Wu, E. Dancigs, and J. Du, "Particle filter-based prediction for anomaly detection in automatic surveillance," *IEEE Access*, p. 1, 2019.
[36] O. A. Elsayed, N. A. M. Marzouk, E. Atef, and M. A. M. Salem, "Abnormal action detection in video surveillance," in Proc. *International Conference on Intelligent Computing*, 2020.
[37] D. K. Vishwakarma, "A two-fold transformation model for human action recognition using decisive pose," *Cognitive Systems Research*, vol. 61, pp.1-13, 2020.
[38] N. C. Tay, T. Connie, T. S. Ong, K. O. M. Goh, and P. S. Teh, "A robust abnormal behavior detection method using convolutional neural network," in *Computational Science and Technology*, Springer, 2019.
[39] W. Sultani, C. Chen, and M. Shah, "Real-World anomaly detection in surveillance videos," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
[40] W. Luo, W. Liu, and S. Gao, "A revisit of sparse coding based anomaly detection in stacked RNN framework," in *Proc. IEEE International Conference on Computer Vision*, 2017.
[41] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Proc. IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, 2010.
[42] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB," in Proc. *IEEE International Conference on Computer Vision*, 2013.
[43] Y. Wang, K. Huang, and T. Tan, "Human activity recognition based on R transform," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition IEEE*, 2007, pp. 1-8.
[44] D. Weinland, R. Ronfard, and E. Boyer, "Free viewpoint action recognition using motion history volumes," *Computer Vision & Image Understanding*, vol. 104, pp. 249-257, 2006.

Yangyue Zhou was born in Henan, P.R. China, in 1995. She received the B.S.E degree from the College of Science, Henan Agricultural University, in 2018. She is currently pursuing the master's degree in computer technology with the College of Information Science and Engineering, Henan University of Technology, China. Her research interests include artificial intelligence information processing, human action recognition, and deep learning.

Miaolei Deng was born in Henan, P.R. China, in 1977. He received his PhD degree from Xidian University, Xian, P.R. China in 2010. Now, he is an associate professor of Henan University of Technology, Zhengzhou, P.R. China. His research interests currently include information security and cryptographic protocol design.