Abnormal Behavior Learning Based on Edge Computing toward a Crowd Monitoring System

Yiming Miao, Jun Yang, Bander Alzahrani, Guoguang Lv, Tarik Alafif, Ahmed Barnawi, and Min Chen

Abstract

Abnormal behavior poses a great threat to social security and stability. The resulting violence or crime leads to terrible consequences. How to utilize reasonable means to predict the dangerous intentions of massive crowds and prevent the potential hazard to the public is significant for social security. A crowd monitoring and management system is an effective way to detect abnormal behavior. In this article, we release unmanned aerial vehicles as well as fixed ground devices to achieve multi-level and multi-modal behavioral sensing on a massive crowd, deploy a hybrid model in edge cloud to extract global features from behavioral data of a massive crowd, and then utilize these global features to construct decent classification algorithms for action recognition and behavioral semantic cognition. With the cooperation of behavioral data and cognitive algorithms, we can understand the instantaneous emotions of the crowd. On the basis of behavioral data and the emotional state of the crowd, the correlation between daily behavior and any dangerous intention of a massive crowd has been revealed by utilizing behavioral big data analysis, which is a key foundation for predicting people's dangerous intentions. Finally, we conduct a case study of abnormal behavior detection based on pix2pix and continuous video frames. The experimental results show that the performance of our method is better than other algorithms in both public datasets and the customized Hajj dataset. The proposed novel pattern for the effective learning of a massive crowd is validated to effectively eliminate some of the possible dangers caused by abnormal behavior.

Introduction

Abnormal behavior, as a common social phenomenon in life, is regarded as the external manifestation of physiological, psychological, or mental abnormalities. In a broad sense, behavior that deviates from expectations and norms is called abnormal behavior. There are many causes for abnormal behavior, including cultural differences, relief from distress, lack of thought or feeling, perceiving the world differently, and so on. Psychologists who study abnormal behavior have further classified the criteria of abnormal behavior, including violation of social norms, statistical rarity, personal distress, and maladaptive behavior. If most of the above criteria are met, a behavior can be judged as abnormal. Common abnormal behaviors include general deviant behaviors (e.g., violating social order and moral standards) and serious deviant behaviors (e.g., criminal behavior, suicide, fights, and violent events). General abnormal behavior has the potential to develop into serious abnormal behavior. Therefore, whether it is general or serious, frequent or large-scale abnormal behavior will cause great harm to social security and stability [1].

In view of the harm caused by abnormal behavior, it is important to improve the efficiency of prevention and control on social security, reduce the occurrence of violent injury events [2], study the dangerous intention behind the abnormal behavior, as well as formulate relevant early warning and intervention strategies. In recent years, with the continuous development of the economy, population, and infrastructure construction, it has become an important means to ensure the safety of people's lives and property to deploy crowd monitoring systems indoors and outdoors. One important function of a crowd monitoring system is early warning while detecting abnormal behavior. The traditional crowd monitoring system based on computer vision technology mostly adopts the deployment of closed-circuit television (CCTV) for crowd surveillance. This kind of system has many disadvantages in monitoring mode, blind areas, equipment deployment, image transmission [3], decision analysis [4], and so on.

An unmanned aerial vehicle (UAV) is a kind of aerial monitoring device, which has the advantages of high efficiency, free motion from space and terrain, diverse observation angles, as well as fast interactive response. We should know any dangerous intention of a crowd in advance to prevent the abnormal behavior of the crowd, which can be achieved by establishing a hybrid model with behavioral features extracted from videos collected by UAVs. With the development of the Internet of Things, 5G, big data analysis, and edge computing, scholars all over the world are exploring the utilization of smart surveillance systems

Yiming Miao is with School of Data Science, The Chinese University of Hong Kong, Shenzhen (CUHK-SZ), China, and School of Computer Science and Technology, Huazhong University of Science and Technology, China; Jun Yang is with Wuhan University of Technology, China; Bander Alzahrani and Ahmed Barnawi are with the Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia; Guoguang Lv and Min Chen (corresponding author) are with School of Computer Science and Technology, Huazhong University of Science and Technology, China; Tarik Alafif is with Jamoum University College, Umm Al-Qura University, Makkah, Saudi Arabia.

Digital Object Identifier: 10.1109/MNET.014.2000523

0890-8044/22/$25.00 © 2022 IEEE

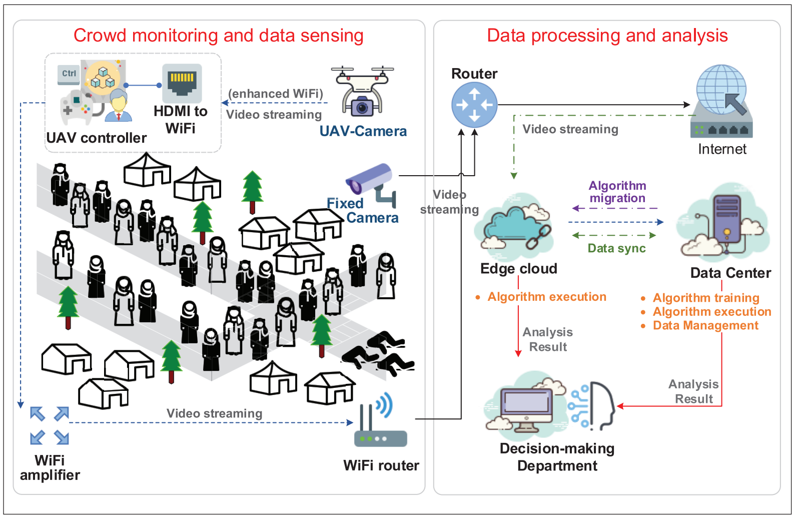

FIGURE 1. The multi-level and multi-sensor architecture of a crowd monitoring system.

to perform object tracking and action recognition, on the basis of which abnormal behavior information of a crowd can be extracted to prevent possible danger. For example, Wu *et al.* utilized motion video to obtain human gait pattern, which is the basis of user identification [5]. Yun *et al.* use audio and video bimodal data to detect the abnormal behavior of students on campus in order to facilitate the real-time intervention of teachers to eliminate potential dangers [6]. By utilizing surveillance UAVs and the ScatterNet-based hybrid deep model, Singh tried to achieve real-time detection of criminal violence in a target area, which should be used for smart surveillance under high-risk scenes such as criminal activity, drug dealing, and border patrolling [7]. Facebook's research team uses RGB images to build a highly dense human pose estimation algorithm, which could achieve superior performance under outdoor scenes and be used for augmented reality [8]. These \ works mainly focus on model construction and performance evaluation on the basis of a standard or self-built dataset; the application and optimization of models in particular domains have not been explored in depth.

Therefore, in the crowd monitoring system with fixed cameras, adding air monitoring equipment, that is, UAVs equipped with ultra-high-definition (UHD) cameras, is a diversified way to expand the monitoring range and video acquisition angle. In addition, due to the uncontrollable weather, terrain, and environmental factors, an infrared camera is also an auxiliary tool to make up for the low-quality video data collected by a UHD camera. Furthermore, by deploying other sensors on the ground, such as ultrasonic sensors and smoke alarms, environmental data can be better collected due to crowd activities. Finally, in order to meet the real-time demand of crowd monitoring, the introduction of edge computing in a multi-sensor system is necessary to reduce transmission delay and enhance local computing ability.

In this article, we develop integrated artificial intelligence (AI) and edge computing based solutions for massive crowd management where UAVs are widely utilized to carry sensing devices such as UHD, and multiple ground sensors such as CCTV and infrared cameras are used to provide diverse data of crowd and environment. Based on the above architecture, we can collect and utilize multi-modal data of human action and facial expression to obtain the emotional state, behavioral state, and behavioral semantic state of a crowd. On this basis, we finally produce a behavioral big data portrait of a crowd and establish the prediction strategy of any dangerous intention. The aim is to arm event operators in massive crowd gatherings such as the multi-million people annual Hajj Pilgrimage to Mecca with powerful intelligent tools toward enhanced decision making in real time.

Even developed multi-level and multi-sensor architecture, viewpoint, background, light, and occlusion may still increase the difficulty of face detection and behavior recognition from surveillance data. Under such circumstances, how to extract effective features from low-quality videos and establish a behavioral recognition model are key issues we seek to solve in this article. Another challenge we pay attention to is how to construct a behavioral big data portrait of a crowd to predict any dangerous intention. Our article investigates corresponding works and achieves three contributions as follows to solve the above three key issues:

- *Senseless all-day monitoring based on multi-level and multi-sensor architecture:* The combination of UAV monitoring, fixed device monitoring, and edge-computing-based data transmission can realize 24/7 surveillance of a target to collect behavior data comprehensively.
- *Cognition of behavioral semantics based on multi-modal data:* We establish a behavioral dataset of a crowd, with which we design a hybrid model to enhance the performance of feature extraction in order to solve the problem of behavior semantic cognition in the mutable environment.
- *Prediction of dangerous intention based on abnormal behavior learning:* By fusing cognitive data over a period of time, we establish the behavioral big data portrait of a crowd, which can correlate behavioral information with possible dangerous intentions to guide decision making.

Our work can promote the application of intelligent surveillance on danger prevention in crowds. The remainder of our article is organized as follows. The next section discusses the design issues of the overall system architecture; following that, we give the construction requirements of a behavioral dataset. We then introduce the modeling means of an abnormal behavior detection model and a hybrid deep model, and the implementation method of a behavioral big data portrait. Next, we show a case study of our method through an abnormal behavior detection experiment. The final section summarizes the whole article.

System Architecture

Figure 1 shows the architecture of the proposed multi-level crowd monitoring system with multiple sensors. The three-layer structure is introduced below:

- *The monitoring network consists of multiple sensors.* Two kinds of fixed cameras, CCTV and infrared, have been utilized to monitor crowds by video data collection. Considering the monitoring blind area, UAVs equipped

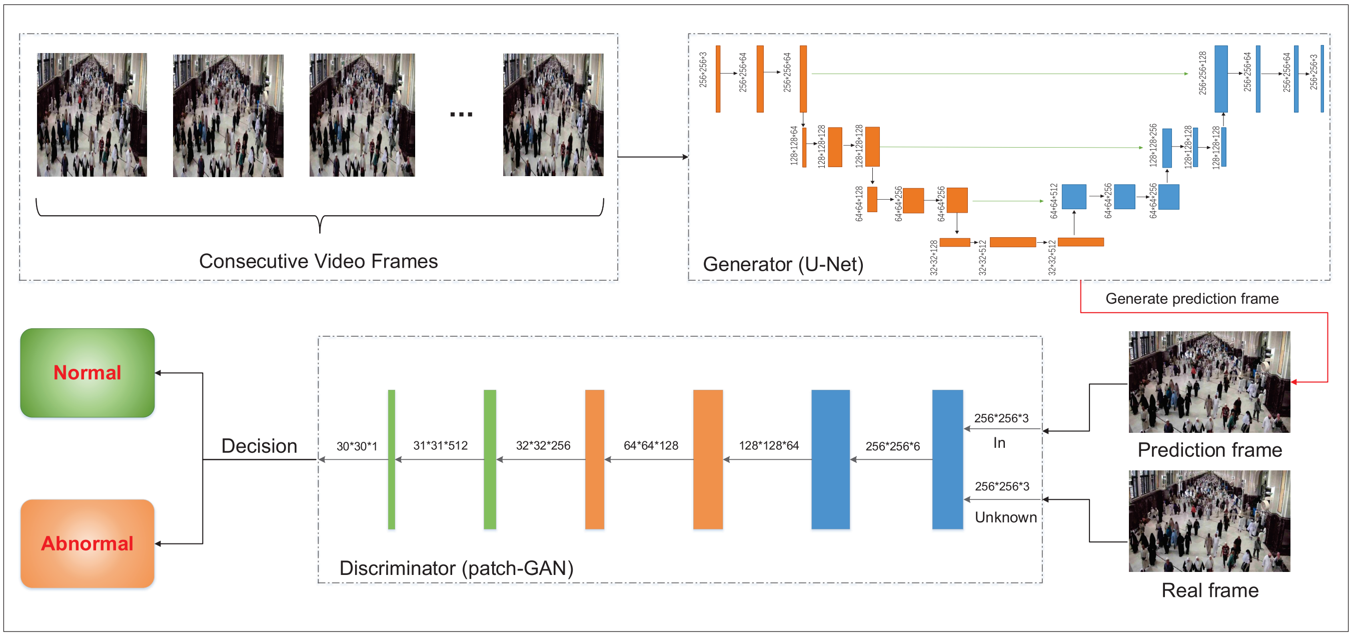

FIGURE 2. Abnormal behavior detection based on pix2pix and continuous video frames.

with UHD cameras are used to track and monitor targets in outdoor environments.

- *The edge cloud provides edge intelligence.* Considering the inherent defects, devices only collect data and transfer data to the edge cloud, which performs data preprocessing, feature extraction, action recognition, and behavior semantic analysis on the basis of a hybrid deep model. The results produced by a model built in the edge cloud will be sent to a decision making department in real time.
- *The cloud data center performs crowd data management and deep algorithm construction.* When providing more powerful services, the storage and computation power of the edge cloud is not enough. Under such circumstances, the edge cloud forwards video data to a cloud data center, which can store and manage the data, optimize the hybrid deep model constantly, and construct a behavioral big data portrait. On this basis, the cloud analyzes the probability of recent dangerous intentions of a crowd, and gives guidance on danger prediction and intervention. The hybrid deep model enhanced by the cloud is synchronized to the edge regularly for upgrading the corresponding model in the edge cloud.

Dataset

With the monitoring video, the hybrid deep model is used to obtain the multi-modal characteristics of crowd behavior data. On this basis, the model for analyzing the behavior pattern and semantics can be constructed. With analysis of this model, we can understand the emotional expression in behavioral activities, which can be used as an important basis for predicting abnormal events. Based on the multi-modal information of the human body and facial expression, we aim to acquire the information of emotional state, behavioral state, and behavioral semantic state, and construct the behavioral semantic recognition model for multi-modal feature extraction.

Specific to a single-modal abnormal behavior dataset, the datasets used in this article include the public datasets of UMN and UCSD, and the Hajj dataset collected in Saudi Arabia to monitor whether there are abnormal behaviors affecting the normal walking of people.

UCSD Dataset: A large-scale dataset. It can be used to judge whether there is abnormal behavior by detecting people using various vehicles on the road. It is divided into two scenarios: pedestrian passing and non-pedestrian passing.

UMN Dataset: A small-scale dataset. It is mainly used for determining whether there is abnormal behavior in the picture. For example, people escaping from a scene is judged as abnormal.

Hajj Dataset: A large-scale and relatively dense dataset of people in a channel collected by a UAV. In the Hajj dataset, people who walk in the normal direction are defined as normal behavior, while behaviors that hinder or cause congestion, such as fast running, moving in the opposite direction, and standing, are defined as abnormal.

Hybrid Deep Model and Crowd Big Data Portrait

Abnormal Behavior Detection Based on pix2pix and Continuous Video Frames

The overall model of the single-modal abnormal behavior detection is shown in Fig. 2, which mainly includes video preprocessing, future frame prediction, and confrontation training. In video preprocessing, we divide the original dataset into four frames per second from the continuous video stream. Then we divide the processed data into a training set and a test set. Regarding the training set, in order to learn the characteristics of normal behavior, all frames should be normal behavior. For the sake of making the model more robust, the test set should be composed of a small number of normal behavior frames and a large number of abnormal behavior frames.

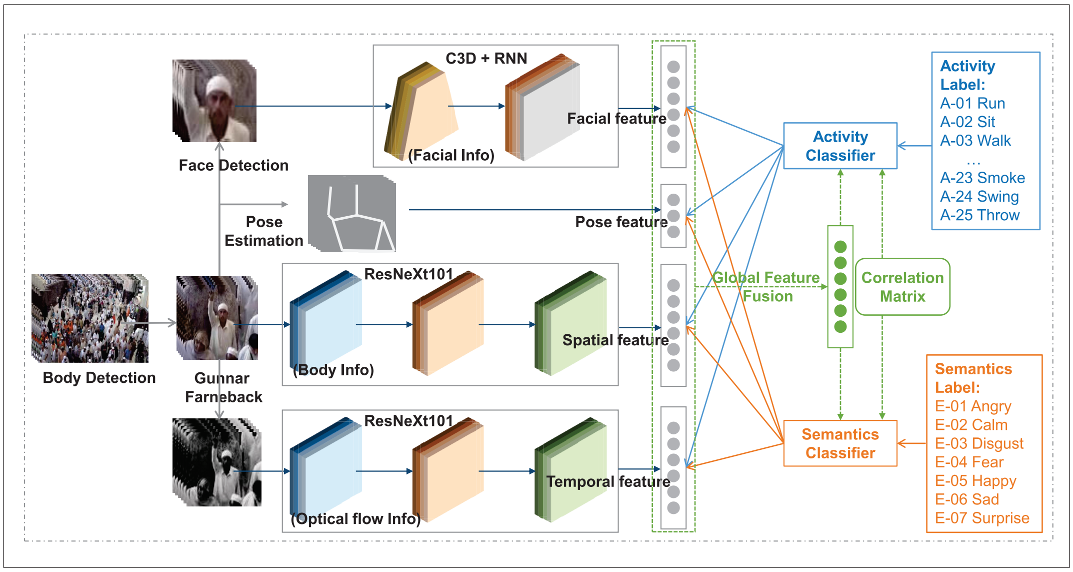

FIGURE 3. The hybrid model for action recognition and behavioral semantic cognition.

In future frame prediction, we input the above processed video frames into the generator structure. In order to comprehensively learn the motion characteristics of normal behavior, the input data should be several consecutive frames. We choose U-Net as the structure of the generator, which can extract the features of the video frame well [1]. U-Net includes down sampling and up sampling. The down sampling is similar to the encoder, which gradually shows the environment information of the video frame by encoding; in contrast, the up sampling is similar to the decoder, which is used to gradually restore the feature information in the video frame.

In confrontation training, for the sake of making the predicted picture more realistic, according to the idea of GAN, we need a discriminator to judge whether the picture generated by the generator is true. The structure of the discriminator is shown in the figure. It can achieve good results by training against the generator. In order to better judge the local details of the image, we use patch-GAN in pix2pix as the discriminator. In the patch-GAN discriminator, the image is first divided into several small patches. Then the authenticity of each small patch is calculated by the discriminator. Finally, the average value of all patches is taken as the result.

Hybrid Model Based on Multi-Modal Visual Features

Considering the complexity of monitoring the environment and the variability of human action, this section focuses on the model design of action recognition and behavioral semantic cognition. Figure 3 displays the framework of a hybrid model for behavioral semantic cognition, which extracts four different kinds of visual features from surveillance videos. The method for local feature extraction uses the OpenPose algorithm [9] to extract feature data of human pose estimation, facial expression, and body information. It utilizes the Gunnar Farneback algorithm (a kind of high density optical flow algorithm) to extract temporal feature data from videos. Then, in order to achieve the global feature extraction of human facial expression, we construct a recurrent neural network (RNN) + C3D hybrid model, which utilizes the EGC dataset for model pre-training, that is, finishing the initialization of all parameters in the hybrid model. Furthermore, we use a customized dataset to train the RNN + C3D hybrid model based on the action recognition labels and behavioral semantic labels, respectively. As for global feature extraction of human body information and optical flow data, we both use the ResNeXt101 model [10], which completes parameter initialization by performing model pre-training with the ImageNet dataset. Then some mainstream datasets of action recognition are utilized for further model tuning. At last, two single ResNeXt101 models for global feature extraction are generated based on our customized dataset. Based on the above four kinds of global features, two global classifiers for action recognition and behavioral semantic cognition are established independently.

$$p(u) = (1 - u)d_{j1} + ud_{j2} \quad (1)$$

OpenPose Algorithm: It is used to extract local features, that is, key points of human pose estimation, facial expression, and human body. The OpenPose algorithm is a bottom-up human pose estimation algorithm using part affinity fields (PAFs), with which the key point position is obtained at first, and then the global skeleton features are obtained. The algorithm uses pretrain VGG as the infrastructure. Each key point has a channel, and the method of taking the best among multiple Gaussian distributions is used to retain the optimal response of each key point when generating ground truth. Whether a point falls on the limb is determined according to the threshold range. After knowing the location of PAFs and key points, the similarity between two joint points is approximately calculated by using uniform sampling with Eq. 1. Then the OpenPose algorithm uses the Hungarian algorithm to perform optimal matching on adjacent nodes, and finally obtains the skeleton for pose estimation.

Gunnar Farneback Algorithm: When a human observes a moving object, the scene of the object forms a series of continuously changing images on the retina of the eyes. Such information constantly "flows" through the retina (image plane), like a kind

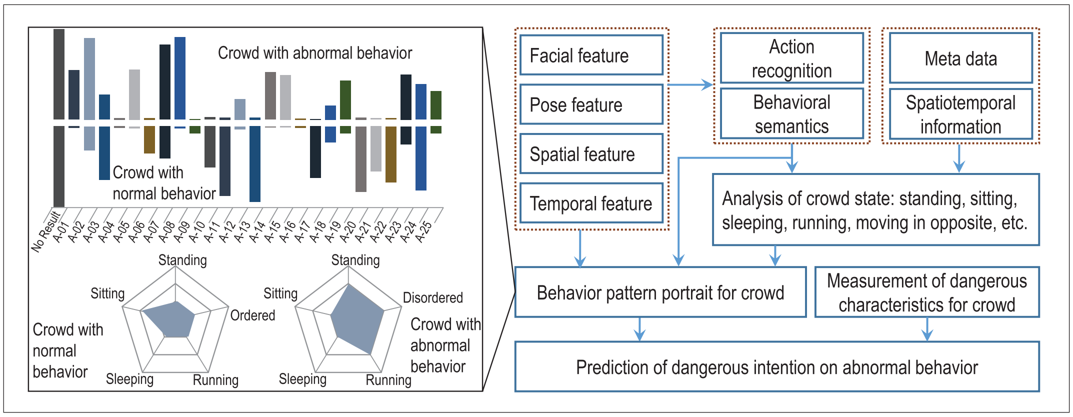
Crowd with abnormal behavior
Crowd with normal behavior
No Result
A-01
A-02
A-03
A-04
A-05
A-06
A-07
A-08
A-09
A-10
A-11
A-12
A-13
A-14
A-15
A-16
A-17
A-18
A-19
A-20
A-21
A-22
A-23
A-24
A-25
Standing
Sitting
Ordered
Sleeping
Running
Crowd with normal behavior
Standing
Sitting
Disordered
Sleeping
Running
Crowd with abnormal behavior
Facial feature
Pose feature
Spatial feature
Temporal feature
Action recognition
Behavioral semantics
Meta data
Spatiotemporal information
Analysis of crowd state: standing, sitting, sleeping, running, moving in opposite, etc.
Behavior pattern portrait for crowd
Measurement of dangerous characteristics for crowd
Prediction of dangerous intention on abnormal behavior

FIGURE 4. The procedure of constructing a big data portrait.

of "flow" of light. Therefore, we also call this phenomenon optical flow. Optical flow expresses the change of image. Due to the information of target motion contained, it can be used by the observer to determine the motion state. The purpose of studying the optical flow field is to approximate the motion field, which cannot be obtained directly from the image sequence. The motion field is actually the movement of objects in the three-dimensional real world; in contrast, the optical flow field is the projection of the motion field on the two-dimensional image plane (human eyes or camera). The optical flow algorithm is a way to find the corresponding relationship between the previous frame and the current frame by using the changes of pixels in the time domain and the correlation between adjacent frames in the image sequence, and to calculate the motion information of objects between adjacent frames. Assuming that the change of optical flow (vector field) is almost smooth, the core problem is the location of the same pixel in the next frame. When performing analysis, the optical flow at the corner point can be determined easily, which is also the most reliable. Second, the optical flow at the boundary can be gotten easily. The optical flow method has a variety of branches. We use a classical dense optical flow algorithm, the Gunnar Farneback algorithm, which is widely used.

ResNeXt101 Algorithm: ResNeXt adopts the idea of VGGNet stacking and the idea of split-transform-merge of Inception at the same time, which has better scalability than VGGNet. The complexity of ResNeXt barely changes when the model is reinforced to increase its accuracy. The novelty of the model is the aggregated transformations, which replace the original ResNet's three-layer convolution block with parallel stacking blocks, and improves the accuracy of the model without significantly increasing the magnitude of the parameters. The performance of ResNeXt is better than ResNet with the same number of parameters. A 101-layer ResNeXt network has the same accuracy as a 200-layer ResNet network, while the computation volume is only half that of the latter. Therefore, we use ResNeXt101 to detect abnormal behavior from videos.

The models discussed above are utilized to extract human features and optical flow features from surveillance videos. Based on the fusion of all kinds of feature data, two classification models for behavior recognition and behavior semantic cognition can be established.

Dangerous Intention Prediction Based on Big Data Portrait

Human action and its behavioral semantics can reflect the instantaneous emotion of people who behave abnormally, but cannot directly reflect their intentions. However, by fusing such instantaneous information over a period of time, to some extent, we can recognize the dangerous intention of a crowd. As shown in Fig. 4, we establish behavioral big data portraits for massive crowds to perform early warning against their possible dangerous actions. The crowd action state includes standing, sitting, sleeping, running, moving in the opposite direction, and so on. Thus, the people under monitoring can be divided into two categories based on the above state level: crowd with normal behavior and crowd with abnormal behavior. The normal crowd gathering scene should be orderly; otherwise, it is abnormal. The big data portrait of a large-scale crowd is based on multi-modal feature data, that is, human action, behavioral semantics, temporal data, as well as location. The value among features is also extracted to construct a high-dimensional vector representation for each person, which maps people into high-dimensional space. On this basis, we build the final model to predict the dangerous intention of a massive crowd.

Case Study

In this section, we test and analyze the abnormal behavior detection model designed earlier as a case study. We compare our model in public datasets and get the best results when the number of consecutive input frames is 4. We also compare the experimental data with some existing methods. The results show that our model has better performance on abnormal behavior detection. Finally, we verify our model in a customized Hajj dataset in the aspect of certain adaptability for relatively large and complex datasets.

Experimental Evaluation Index

In this section, we introduce some of the metrics used in our experiments. We use the area under curve (AUC) to measure the abnormal behavior detection performance of the model. The accuracy performance of AUC is gratifying because it calculates the generalization ability of the model under different thresholds by dynamically adjusting the boundary between the true label and the false label.

AUC is directly proportional to the performance of the model, that is, the closer the AUC is to the maximum, the better the model performance is.

In the data anomaly detection, in order to measure whether the image is abnormal, we use the peak signal-to-noise ratio (PSNR) to compare the similarity between the predicted frame and the real frame. The larger the PSNR is, the more similar the predicted frame is to the real frame. Since the probability of an abnormal image predicted by the model trained with normal data is extremely low, the smaller the difference, the more likely the real frame is to be a normal frame. On the contrary, the real frame is more likely to be an abnormal frame. We use PSNR to convert the image into the corresponding score, adjust different thresholds to determine whether there is an abnormal image, and then compare it with the real label to calculate the AUC of the model.

Comparative Experiments

In this section, we use the public UMN and UCSD datasets to conduct experiments on the model. Considering that the number of consecutive video frames directly affects the ability of the model to learn normal behavior features, we conduct comparative experiments with 2, 4, 6, and 8 as input continuous video frames in UMN and UCSD, respectively. The experimental results are shown in Table 1. Through the comparison, we can see that in different scenarios of UMN and UCSD exposing datasets, when the number of input video frames is 4, the abnormal behavior detection ability of the model is significantly better than that of other input video frames.

Experimental Evaluation

In this section, regarding public datasets of UMN and UCSD, we choose some advanced deep learning methods such as stacked RNN and some traditional methods such as optical flow to compare with our best experimental results. The experimental results are shown in Table 2.

We can see that the performance of our model is improved compared to the previous methods in terms of AUC in UMN and UCSD datasets, which proves that our model is feasible for abnormal behavior detection. Moreover, the AUC of our model is 73.1 percent in the more complex Hajj population dataset, where the population density is higher and abnormal behavior identification is more difficult. Although the experimental results do not achieve the high accuracy of public datasets, it also has a certain ability to detect abnormal behavior compared to such large-scale and difficult-to-identify datasets. On the whole, our model can reflect the ability of abnormal behavior detection for small and medium-sized datasets, and it also has certain adaptability and robustness for a large, complex, and high population density dataset.

Conclusion

This article explores a novel pattern for crowd management to attain the possible relationship between long-term behaviors and dangerous intentions. UAV surveillance and computer vision technologies are used to continuously analyze crowd behaviors and construct big data portraits. Early warning of dangerous intention is realized through a hybrid model of abnormal behavior detection, action recognition, and behavioral semantic cognition. The results show that our method can significantly improve the detection effect of abnormal behavior, help to distinguish diverse behaviors and tendencies, and effectively manage crowds with different scales.

Num	Ped1	Ped2	Umn1	Umn2	Umn3
2	81.13%	93.95%	97.18%	96.41%	96.26%
4	83.89%	94.13%	97.91%	97.25%	97.63%
6	81.89%	93.86%	97.48%	97.06%	96.81%
8	83.45%	94.56%	95.59%	96.49%	96.36%

TABLE 1. The AUC results of comparative experiments.

Method	UCSD Ped1	UCSD Ped2	UMN	HAJJ
MPPCA [11]	59.0%	69.3%	N/A	N/A
optical-flow [12]	N/A	N/A	84.0%	N/A
Social force(SF) [12]	67.5%	55.6%	N/A	N/A
SF+MPPCA [13]	68.8%	61.3%	N/A	N/A
SFM [12]	N/A	N/A	96.0%	N/A
MDT [13]	81.8%	82.9%	N/A	N/A
Conv-AE [14]	75.0%	85.0%	N/A	N/A
Sparse reconstruction [2]	N/A	N/A	97.0%	N/A
Stacked RNN [13]	N/A	92.2%	N/A	N/A
Unmasking [15]	68.4%	82.2%	N/A	N/A
Our proposed method	83.9%	94.6%	97.9%	73.1%

TABLE 2. The AUC results of the UCSD, UMN, and Hajj datasets.

Acknowledgments

The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number (227). Y. Miao's work was partially supported by the Guangdong-Shenzhen Young Scientists Program of China under Grant 2021A1515110353, and the Shenzhen Institute of Artificial Intelligence and Robotics for Society (AIRS), Shenzhen, China.

References

[1] T. Alafif *et al.*, "Generative Adversarial Network Based Abnormal Behavior Detection in Massive Crowd Videos: A Hajj Case Study," *J. Ambient Intelligence and Humanized Computing*, 2021, pp. 1-12. DOI: 10.1007/s12652-021-03323-5

[2] Y. Cong, J. Yuan, and J. Liu, "Sparse Reconstruction Cost for Abnormal Event Detection," *Proc. IEEE CVPR 2011* 2011, pp. 3449–56.

[3] X. Ding *et al.*, "Crowd Density Estimation Using Fusion of Multi-Layer Features," *IEEE Trans. Intelligent Transportation Systems*, vol. 22, no. 8, 2021, pp. 4776–87.

[4] B. Sirmacek and P. Reinartz, "Automatic Crowd Density and Motion Analysis in Airborne Image Sequences Based on a Probabilistic Framework," *Proc. IEEE Int'l. Conf. Computer Vision Wksps.*, 2011.

[5] Z. Wu *et al.*, "A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 39, no. 2, 2017, pp. 209–26.

[6] S. Yun, Q. Nguyen, and J. Choi. "Recognition of Emergency Situations Using Audio–Visual Perception Sensor Network for Ambient Assistive Living," *J. Ambient Intelligence and Humanized Computing*, vol. 10, no. 1, 2019, pp. 41–55.

[7] S. Amarjot, D. Patil, and S. N. Omkar, "Eye in the Sky: Real-Time Drone Surveillance System (DSS) for Violent Individ-

uals Identification Using ScatterNet Hybrid Deep Learning Network," *Proc. IEEE Conf. Computer Vision and Pattern Recognition Wksps.*, Salt Lake City, UT, 2018.

[8] A. G. Riza, N. Neverova, and I. Kokkinos, "Densepose: Dense Human Pose Estimation in the Wild," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018.

[9] Z. Cao *et al.*, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," arXiv preprint, sarXiv:1812.08008v2, 2019.

[10] S. Xie, *et al.*, "Aggregated Residual Transformations for Deep Neural Networks," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Honolulu, HI, 2017.

[11] J. Kim *et al.*, "Observe Locally, Infer Globally: A Space-Time MRF for Detecting Abnormal Activities with Incremental Updates," *Proc. 2009 IEEE Conf. Computer Vision and Pattern Recognition*, Miami, FL, 2009, pp. 2921–28. DOI: 10.1109/CVPR.2009.5206569.

[12] R. Mehran, A. Oyama, and M. Shah, "Abnormal Crowd Behavior Detection Using Social Force Model," *Proc. 2009 IEEE Conf. Computer Vision and Pattern Recognition*, Miami, FL, 2009, pp. 935–42. DOI: 10.1109/CVPR.2009.5206641.

[13] V. Mahadevan *et al.*, "Anomaly Detection in Crowded Scenes," *Proc. 2010 IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 2010, pp. 1975–81.

[14] M. Hasan *et al.*, "Learning Temporal Regularity in Video Sequences," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 733–42.

[15] R. Tudor Ionescu *et al.*, "Unmasking the Abnormal Events in Video," *Proc. IEEE Int'l. Conf. Computer Vision*, 2017, pp. 2895–2903.

Biographies

Yiming Miao is currently a research assistant professor in the School of Data Science at the Chinese University of Hong Kong, Shenzhen (CUHK-SZ), China. She received her Ph.D, degree from the School of Computer Science and Technology at Huazhong University of Science and Technology (HUST), Wuhan, China, in 2021. She also received her B.Sc. degree from the College of Computer Science and Technology from Qinghai University, Xining, China, in 2016. Her research interests include edge computing, mobile communication systems, the Internet of Things, unmanned aerial vehicles, blockchain, and wireless sensor network, among others.

Jun Yang received his Ph.D. degree from the School of Computer Science and Technology, HUST, in 2018. Then he worked as a postdoctoral fellow in the Embedded and Pervasive Computing (EPIC) Lab at HUST between July 2018 and February 2020. Now he is an associate professor in the School of Information Engineering at WHUT. His research interests include human action recognition, software intelligence, smart healthcare, cloud computing and big data analytics, and others. He has published 30+ papers in *IEEE Communications Magazine*, *IEEE Network*, *Future Generation Computing System*, *Multimedia Tools and Applications*, *IEEE Multimedia*, the *IEEE Sensors Journal*, *Mobile Networks & Applications*, *Computer Communications*, and more.

Bander Alzahrani received his M.Sc. in computer security and Ph.D. in computer science from the University of Essex, United Kingdom, in 2010 and 2015, respectively. He is currently an associate professor in the Faculty of Computing and Information Technology, King Abdulaziz University, Saudi Arabia. He has led more than 10 national research projects and co-authored more than 70 research articles in peer reviewed journals and conferences. His current research interests include WSN, ICN, and secure content routing.

Guoguang Lv is currently a Master's student with the Embedded and Pervasive Computing (EPIC) Laboratory in the School of Computer Science and Technology, HUST. He graduated from Jilin University, China, in 2020. His research interests focus on abnormal crowd behavior, computer vision, and more.

Tarik Alafif received his Ph.D. degree from the Computer Science Department at Wayne State University, Detroit, Michigan, in 2017 and his Master's from Gannon University, Erie, Pennsylvania, in 2011. He is currently an assistant professor in the Computer Science Department at Jamoum University College of Umm Al-Qura University. His main research interests include computer vision, machine learning, deep learning, and large-scale data. He has published several refereed journal and conference papers in these areas.

Ahmed Barnawi is currently a professor in the Faculty of Computing and IT at King Abdulaziz University. He is the managing director of the KAU Cloud Computing and Big Data Research group. He acquired his Ph.D. from the University of Bradford, United Kingdom, in 2005, and his M.Sc. from UMIST, United Kingdom, in 2001. His research interests include big data, cloud computing, and advanced mobile robotic applications. He has published more than 100 papers in peer reviewed journals.

Min Chen [F] has been a full professor in the School of Computer Science and Technology at HUST since February 2012. He is the director of the Embedded and Pervasive Computing Lab, and the director of the Data Engineering Institute at HUST. He is the founding Chair of IEEE Computer Society Special Technical Communities on Big Data. He was an assistant professor in the School of Computer Science and Engineering at Seoul National University before he joined HUST. He is the Chair of the IEEE GLOBECOM 2022 eHealth Symposium. His Google Scholar Citations reached 35,150+ with an h-index of 90. His top paper was cited 3850+ times. He was selected as a Highly Cited Researcher from 2018 to 2021. He received the IEEE Communications Society Fred W. Ellersick Prize in 2017 and the IEEE Jack Neubauer Memorial Award in 2019.