

An overview of the monitoring of the crowd and abnormal behavior in public places

Xinyu Chen ¹, Hong Zhang ^{2,*}, Tao Wang ¹

¹The College of Nuclear Technology and Automation Engineering, Chengdu University of Technology, Chengdu, Sichuan, China

²The College of Information Science and Technology, Chengdu University of Technology, Chengdu, Sichuan, China

Abstract: In the past, people only rely on the naked eye or sit in front of the monitoring to monitor the crowd and abnormal behavior, which greatly consumes human and material resources. With the increasing prosperity of artificial intelligence, the application of deep learning to solve people's real life problems has become an inevitable trend in the future. The terrorist attack on Kunming railway station in Yunnan in 2014 and the terrorist incident in Hong Kong, China in 2019 all sounded the alarm for public security again. This paper mainly describes the research status of video surveillance system, abnormal face, abnormal behavior, crowd counting direction, and prospects the future development trend.

Keywords: Abnormal behavior, crowd counting, public safety.

1. INTRODUCTION

With the improvement of social material level and the increasing population, the number of daily trips and the number of people gathered in public places show a large-scale growth trend, and public safety has become the focus of social attention. Major public security incidents, such as the terrorist attack at Kunming railway station in Yunnan Province in 2014, the stampede on the Bund in 2014, and the terrorist incident in Hong Kong, China in 2019, have resulted in panic, social unrest and even casualties. People's lives and safety have become the primary focus of attention in public places.

In view of the current situation, although video monitoring is installed in public places, more people watch videos in the monitoring room, and artificial watching can't guarantee to pay attention to emergencies in public places 24 hours a day. China is a country with a large population. Especially during holidays, the tourist attractions, railway stations and downtown areas of cities are full of people. If there is a stampede or even a terrorist attack in such public places, it is difficult for the people who are helpless to deal with it and even lose their lives. If the monitoring system can timely identify crowd crowding or abnormal behavior, it will prompt people to stay away from the place of right and wrong, so as to ensure the safety of more people. With the increase of population, the state's attention

to public security has risen to the point of urgency. Therefore, this article introduces the current research situation in various directions in detail, and looks forward to the future development.

2. RESEARCH STATUS

2.1 Video surveillance system

Video surveillance system has experienced the development process from analog surveillance system, digital surveillance system to intelligent surveillance system. Intelligent video surveillance system belongs to the field of computer vision, which can get rid of the disadvantages of traditional video surveillance systems that require people to keep an eye on the screen at all times, combine video surveillance with deep learning, image processing, software engineering and other fields, realize self-supervised learning, process, track, detect and identify images in videos, and actively respond to abnormal situations in time, thus greatly saving manpower and material resources, saving social resources and improving the accuracy and intelligence of the surveillance system. At present, the research on intelligent monitoring system at home and abroad is mostly used to prevent trampling, crowd counting and other public safety, or to monitor the feedback of vehicles to the traffic control department to help the urban traffic operation.

2.2 Abnormal face

At present, abnormal face recognition technology is usually used in the monitoring of ATM machines and indoor places in banks, which can recognize abnormal people with unclear facial features, such as those who wear masks, sunglasses and sun hats [1] to cover their faces. The recognition methods mainly include comparing the difference between normal face and abnormal face on the basis of head positioning [2]. By locating the eyebrows and centerline of face, the eye region detection method based on face centerline detection or the eye region detection method based on regional gray distribution characteristics is adopted for sunglasses. The mask detection method is based on straight line, and the abnormal face recognition is also carried out by PCA (Principal Components Analysis) algorithm [3]. At present, there are also studies that divide the face region into eye region and mouth region to improve the speed and accuracy of detection.

In the process of eye anomaly detection, a human eye classifier based on Haar feature is adopted, and the eye region can be searched by this classifier. During the abnormal detection of the mouth, the horizontal line detection based on Hough transform combined with classifier is used to search the mouth area [4]. By comparing the difference between the normal face and the detected face, it is judged whether the detected face is abnormal.

2.3 Abnormal behavior

At present, the detection of abnormal human behavior is mainly based on the characteristics of people and their movement points. At present, the main research methods are for the detection of individual abnormal behavior. According to the dense optical flow method and the interpolation method, the moving speed and direction of each moving individual are calculated, and the overall average moving speed and average moving direction in the image are compared with each moving individual. If there is a large difference among individuals, it is determined as abnormal action [5]. For group panic and other group abnormal scene, using the difference between the two consecutive frames of video frame

to find the changed pixel, and carry out the second-order differential of each change pixel to get the corresponding spatial feature point, and calculate the average direction of each spatial feature point field. Combined with the frame difference between two consecutive frames, SVM (support vector) is used Machines) support vector machine does a lot of learning on normal behavior and abnormal behavior samples, and finally realizes the recognition of abnormal behavior.

In general, in traditional machine learning, group abnormal behavior detection mainly includes inference based, classification based, clustering and reconstruction methods [6]. The commonly used classification and clustering algorithms are SVM [7-8] support vector machine and neural network. The classical inference algorithms are Markov random field [9] and hidden Markov model [10]. For reconstruction, some studies [11] combine the image features into a feature matrix, use the similarity matrix to reconstruct, and judge whether it is a normal sample or an abnormal sample through the reconstruction error. In addition, some studies [12] introduced population density and population distribution information into the energy model established by anomaly detection. After the vigorous development of deep learning, restricted Boltzmann machine, convolutional neural network, self encoder and full convolutional neural network have become the main research models. According to research [13], 3D convolution neural network model is used to add data information to neural network with spatial dimension and time dimension. Its advantage is that it can not only extract spatial information of data, but also capture motion information between consecutive video frames. Some studies [14] also proposed a spatiotemporal self encoder model combined with neural network to comprehensively judge whether there is abnormal behavior through reconstruction error and set threshold size.

2.4 Crowd counting

At present, there are mainly density estimation methods based on detection, clustering, regional regression and deep learning. The method based on detection is to detect the head, face, pedestrian and so on directly and intuitively to count the number of people [15-16]. The method based on clustering is to cluster the feature points or the tracks of feature points to get the counting results. The commonly used algorithms are iterative optimization and tracking algorithm based on feature invariance and random sampling consistency algorithm [17-18]. The method based on regional regression does not need to obtain the specific location of each individual in the scene, but establishes a regression relationship between the number of feature points and the number of regions to obtain the number of people. The commonly used hybrid dynamic texture model extracts the foreground region, and uses the Gaussian regression model and Bayesian regression model to establish the mapping relationship [19-20].

The existing research based on regional regression mainly uses foreground segmentation, perspective correction, feature extraction and regression counting steps to complete the counting of population [21]. Feature extraction is mainly divided into pixel features, edge features, texture features and feature points [22]. In texture feature extraction, gray level co-occurrence matrix, local binary pattern, directional gradient histogram and other methods are often used. Fast corner and surf feature points are commonly used for feature point extraction, and the weight is weighted by penetration normalization. Regression count is divided into kernel ridge regression, linear regression, Gaussian process regression and random forest regression, and the most commonly used method is Gaussian

process regression.

The density estimation methods based on deep learning can be divided into basic network model, scale adaptive model, context aware model and multi task learning network. At present, research often adopts the idea of extracting image features on each pixel, learning the mapping relationship between pixel features and density distribution map through neural network model training, and finally generating target density map through a large number of training [23]. The advantage of density map is that it not only contains density information, but also can directly get the population number information in any region according to the density. More importantly, the density map contains the distribution information of the target in space, which can provide more information than one-dimensional data. Therefore, in the current research, density estimation based on deep learning is a commonly used population counting method, and based on this method, more in-depth and more accurate research is carried out.

As shown in Table 1, at present, the main technologies of face anomaly include eyebrow and centerline location, eye region detection method based on face centerline detection, eye region detection method based on regional gray distribution features, line based detection method, PCA principal component analysis detection algorithm, eye classifier based on Haar features, and horizontal line detection combined classifier based on Hough transform; At present, the main technologies of abnormal group behavior include optical flow method, inference algorithm based on Markov random field, hidden Markov model, classification clustering algorithm based on SVM support vector machine and neural network, reconstruction error algorithm, restricted Boltzmann machine, convolutional neural network, self coder, full convolution neural network model, 3D convolutional neural network model, and time combined with neural network At present, there are four kinds of density estimation methods based on detection, clustering, regional regression and deep learning. The density estimation methods based on deep learning are the most popular at present, mainly including basic network model, scale adaptive model, context aware model and multi task learning network model.

Table 1 Summary of main research directions and research status

Research direction	Research status
Abnormal face	<ol style="list-style-type: none"> 1. Facial eyebrow and centerline location 2. Eye region detection based on face centerline detection 3. Eye region detection method based on regional gray distribution features 4. Detection method based on line 5. PCA detection algorithm based on principal component analysis 6. Eye classifier based on Haar feature 7. Horizontal line detection combined with classifier based on Hough transform
Abnormal group behavior	<ol style="list-style-type: none"> 1. Dense and sparse optical flow method 2. Inference algorithm based on Markov random field and hidden Markov model 3. Classification clustering algorithm and reconstruction error algorithm based on SVM support vector machine and neural network 4. Constrained Boltzmann machine, convolutional neural network, self encoder, full convolution neural network model 5. 3D convolution neural network model 6. Spatiotemporal self encoder model combined with neural network

Crowd counting	<ol style="list-style-type: none"> 1. There are four kinds of density estimation methods based on detection, clustering, regional regression and deep learning 2. Basic network model, scale adaptive model, context aware model and multi task learning network model
----------------	--

3. SUMMARY

With the continuous improvement of social level, the number of people gathered in public places shows a large-scale growth trend, and the following public security issues such as stampede incidents and terrorist attacks have aroused great social concern. With the rapid development of artificial intelligence, the use of deep learning to solve people's real life problems has become an inevitable trend in the future. This article introduces the current research status of video surveillance system, abnormal face, abnormal behavior and crowd counting. But on the whole, at present, domestic and foreign researches mainly focus on face anomaly, group abnormal behavior and crowd density estimation separately, and have not applied face anomaly, group behavior anomaly and crowd density together in real life scenes.

In the future research, combining these unilateral studies to form a complete crowd anomaly monitoring system, which will be applied to social life, will become the future development trend.

REFERENCES

- [1] Yuan Baohua. Localization and recognition of abnormal faces [D]. Nanjing University of technology, 2005.
- [2] Shao Yeqin. Head localization based on sequence images [D]. Nanjing University of technology, 2004.
- [3] Hong Ziquan, Yang Jingyu. Portrait recognition algorithm based on singular value feature and statistical model [J]. Computer research and development, 1994 (03): 60-65.
- [4] Li Gang, Gao Zheng. Research and development of face detection technology [J]. Computer and modernization, 2003 (04): 7-9 + 12.
- [5] Zhang Chen. Design of crowd abnormal behavior and face alignment algorithm [D]. South China University of technology, 2015.
- [6] Wu Yulin. Research on video based group abnormal behavior detection algorithm [D]. Anhui University, 2019.
- [7] Tung F, Zelek J S, Clausi D A. Goal-based trajectory analysis for unusual behaviour detection in intelligent surveillance [J]. Image and Vision Computing, 2011, 29(4): 230-240.
- [8] Xinyu Wu, Yongsheng Ou, Huihuan Qian, etc. A detection system for human abnormal behavior [C] // Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on .IEEE, 2005: 1204-1208.
- [9] Shunsuke Kamijo, Yasuyuki Matsushita, Katsushi Ikeuchi, etc. Occlusion robust tracking utilizing Spatio-Temporal Markov Random Field model [C] // Pattern Recognition, 2000. Proceedings. 15th International Conference on .IEEE, 2000, 1: 140-144.
- [10] J. Yamato, Jun Ohya, K. Ishii. Recognizing human action in time-sequential images using hidden Markov model [C] // Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92. 1992 IEEE Computer Society Conference on .IEEE, 1992: 379-385.
- [11] Wang, Lijun, Dong, etc. Real-time detection of abnormal crowd behavior using a matrix approximation-based approach. 2012. IEEE.

- [12] G Xiong , X Wu , YL Chen ,etc.Abnormal crowd behavior detection based on the energy model.Proceeding of the IEEE International Conference on Information and Automation Shenzhen,China June 2011.
- [13] Ji, Shuiwang, Xu,etc.3D Convolutional Neural Networks for Human Action Recognition[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,vol.35,no.1,pp.221-231,2013.
- [14] Chong, Yong Shean,Tay,etc.Abnormal Event Detection in Videos using Spatiotemporal Autoencoder[C] /International Symposium on Neural Networks.Springer,Cham,2017:189-196.
- [15] Sheng-Fuu Lin, Jaw-Yeh Chen, Hung-Xin Chao.Estimation of number of people in crowded scenes using perspective transformation[J].IEEE Transactions on Systems Man and Cybernetics - Part A Systems and Humans,2001,31(6):645-654.
- [16] Paul Viola, Michael J. Jones.Robust Real-Time Face Detection[J].International Journal of Computer Vision,2004,57(2):137-154.
- [17] V. Rabaud,S. Belongie.Counting Crowded Moving Objects[J].Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,2006.705-711.
- [18] P Tu , T Sebastian , G Doretto ,etc.Unified Crowd Segmentation.[J].Proceedings of European conference on Computer Vision.Springer,2008.691-704.
- [19] Antoni B. Chan,Zhang-Sheng John Liang,Nuno Vasconcelos.Privacy preserving crowd monitoring: Counting people without people models or tracking[J].Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,2008.1-7.
- [20] Chan, A. B.,Vasconcelos,etc.Counting People With Low-Level Features and Bayesian Regression[J].IEEE Transactions on Image Processing,2012,21(4):2160-2177.
- [21] Liu Xu. Research on target counting method in video surveillance [D]. University of science and technology of China, 2018.
- [22] Ryan, David, Denman,etc.An evaluation of crowd counting methods, features and regression models[J]. Computer Vision and Image Understanding,2015,130:1-17.
- [23] Lempitsky V, Zisserman A..Learning to count objects in images.[J].Proceedings of Advances in Neural Information Processing Systems, volume 1,2010.1324-1332.