Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 6, No. 1, 2023

Public Place Crowd Transaction Monitoring System

Zhize Wang

South West Minzu University, Chengdu 610000, China

Abstract: Currently, the phenomenon of abnormal movement in public spaces by groups is becoming increasingly prominent, leading to issues concerning public flow and safety. The escalating problems of high crowd density, the presence of controlled dangerous items, and unexpected group activities highlight the necessity for timely detection in public settings. Timely identification of such scenarios will facilitate prompt responses and assistance from relevant government departments. Exploring how artificial intelligence technology can aid urban management personnel in effectively detecting abnormal group behaviors is crucial. Having the ability to swiftly and efficiently evacuate crowds in emergency situations holds significant practical importance. This paper employs deep learning methodologies to assist urban management personnel in efficiently monitoring crowd density and detecting abnormal behaviors. The aim is to maintain crowd density within reasonable limits and enable rapid and effective crowd evacuation in emergency situations. Detection of abnormal group behaviors typically involves methods based on global features, extracting feature patterns like optical flow from entire video segments and constructing corresponding histograms. Given that automatic classification of crowd patterns involves sudden and abnormal changes, a novel method is proposed to extract motion "textures" from dynamic STV (Space-Time Volume) blocks formed from real-time video streams.

Keywords: Neural Networks; Deep Learning; Anomalous Behavior; Clustering Algorithms.

1. Introduction

In recent years, the issues of abnormal movement among groups in public spaces have become increasingly prominent, intensifying concerns regarding public flow and safety. The timely detection of events involving excessively dense crowds, controlled dangerous items such as metal knives, and unexpected group activities can prompt swift responses and aid from relevant government departments. This, in turn, enhances public safety and reduces harm to individuals and property. The rapid evolution of artificial intelligence (AI) technology in the realm of social sciences and technology has garnered substantial attention and interest across academia and various industries. Notably, with the explosive growth in datasets, data processing capabilities driven by Moore's Law in hardware technology, and exponential computational power, deep learning and machine learning methodologies have played pivotal roles in the field of AI learning.

As China's urbanization progresses and accelerates, the migration of populations towards cities continues to increase. Consequently, the issue of large crowds congregating in public spaces has become more prevalent, especially during weekends, holidays, and peak times when the public tends to gather in areas like tourist spots, subways, and transportation hubs. In today's cities, managing crowd density and preventing anomalous behaviors within these populations are imperative. Incidents of abnormal group behaviors in these areas can lead to severe consequences for both the public and relevant authorities. For instance, the 2014 stampede incident at the Shanghai Bund highlights the criticality of monitoring abnormal crowd behaviors and crowd density in public spaces for everyone's safety.

This paper primarily focuses on studying methods for detecting abnormal group behaviors. It categorizes the current algorithms for abnormal group monitoring [14] into two classes: tracking-based methods and non-tracking methods. Tracking-based approaches detect anomalies by tracking specific targets. This method records certain trajectories and identifies anomalies as deviations from normal trajectories [15]. While suitable for non-congested scenarios, it is inadequate for monitoring abnormal movements in congested public places or crowd density control systems. It focuses solely on spatial errors, failing to detect abnormalities in object appearance or movement along a normal trajectory.

Non-tracking methods typically rely on underlying features such as motion (optical flow or gradients) and texture features to analyze group data. These methods are adaptable to both congested and non-congested groups. For instance, a study [16] presents a research approach based on the particle energy distribution according to optical flow characteristics. It defines groups by constructing a co-occurrence matrix and evaluates group movement using the entropy and contrast of the co-occurrence matrix.

This paper proposes a deep learning-based method for detecting abnormal crowd behaviors using PyTorch. Leveraging the effectiveness of training, neural network optimization, and parameter tuning, it constructs suitable models. Experimental results demonstrate that this method can effectively accomplish the majority of abnormal behavior detection tasks.

2. Method and Result

2.1. DBSCAN Clustering Algorithm

Detection of abnormal behaviors within crowds hinges on establishing criteria for defining what qualifies as an anomaly. Generally, defining normal versus abnormal behavior often involves a comparative approach. Initially, setting standards for normal behavior allows us to categorize anything contrary to these standards as abnormal. Furthermore, the definition of abnormal behavior naturally fluctuates based on different scenarios. For instance, in some contexts, a sudden acceleration of a crowd might constitute abnormal behavior, while in other situations, an abrupt halt during rapid movement might be considered abnormal. Group counting algorithms provide density maps that encapsulate information about group distribution. Hence, using density maps as a basis for anomaly detection is viable. The DBSCAN algorithm can

identify all dense regions within sample points, significantly reducing computational workload.

In the model design phase, the primary focus is on analyzing density maps using clustering algorithms. The steps involved can be outlined as follows:

1. Convert video into images.
2. Locate and extract coordinates of prominent crowds within the images.
3. Analyze the extracted data using clustering algorithms.
4. Derive data changes over time due to crowd movements.

Through these four steps, the algorithm model outputs the number of clusters and the total number of samples in the largest cluster. Subsequently, the model employs predefined criteria for abnormal behavior—such as sudden rapid movement in a particular direction, sudden dispersal, or sudden aggregation of crowds—to determine instances of abnormal behavior within the crowds.

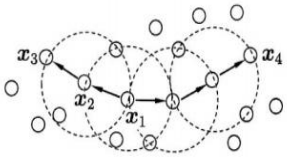

Fig 1. Cluster simulation

Firstly, concerning three different scenarios of normal crowd movement: the phenomenon of crowds walking face-to-face without crossing paths, crowds walking face-to-face with crossing paths, and different crowds walking inward from four directions and crossing paths. The graph below illustrates the trend of the "total samples in the largest cluster" metric.

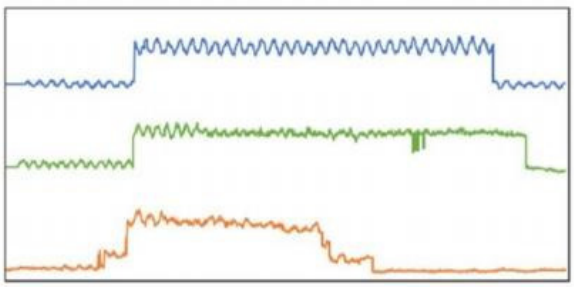

Fig 2. State curve simulation

The upper blue line represents the phenomenon of crowds walking face-to-face without crossing paths, the middle green line represents crowds walking face-to-face with crossing paths, and the lower orange line represents different crowds walking inward from four directions and crossing paths.

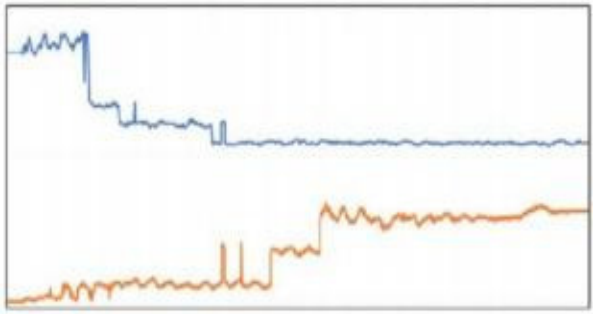

Fig 3. State curve simulation

The graph below illustrates the trend of the "total samples in the largest cluster" metric. The upper blue line represents the phenomenon of crowds dispersing in all directions, showing a continuous decrease in the number of clusters due to the scattering of people in various directions. The lower orange line represents the phenomenon of crowds gathering and moving towards a specific location, indicating a continuous decrease in the "total samples in the largest cluster" due to the dispersion of the crowd away from the gathering point. Simultaneously, due to the congregation of people, the "total samples in the largest cluster" consistently increases.

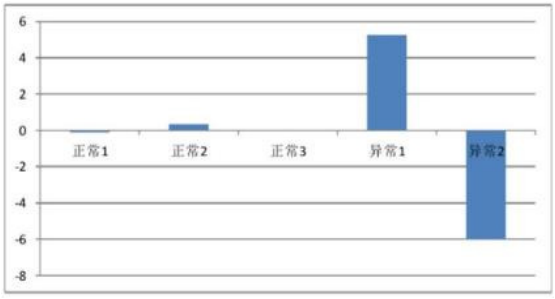

Fig 4. Clustering accuracy rate

The project distinctly differentiates between normal and abnormal situations by defining a "comprehensive chaos value."

2.2. The Identification of Clustering Structures Using OPTICS

Defines ordering points as an algorithm capable of detecting outliers based on the density structure from large-scale datasets. This method is utilized to detect system clusters in spatial data based on density. Its computation not only relates to DBSCAN but also highlights crucial distinctions in identifying various useful system clustering analysis methods based on density.

Core Points: These are points positioned at the center of a cluster and are considered core points if a minimum of MinPts data points exist within their neighborhood.

Core Distance: This represents the minimum radius required to group observed values around core points.

Reachability Distance: This distance metric indicates the shortest length from a data point to a cluster's center point, which cannot be less than the core distance. In the description below, the reachability distance between (c, a) is denoted as x because their distance is less than the unacceptable core distance.

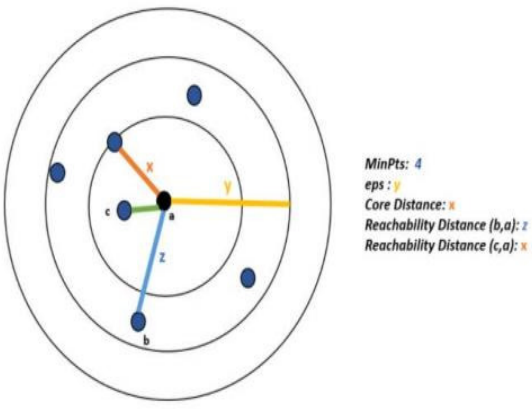

Fig 5. Clustering algorithm

2.3. Crowd Testing

2.3.1. Data Testing

The project conducted training on various datasets, including:

1. UCSD dataset: Consisting of approximately 2000 images with relatively low crowd density.

2. UCF_CC_50S dataset: Containing multiple scenes with varying densities, perspective distortions, and significant variations in the number of individuals across 50 images.

3. WorldExpo 10 dataset: Utilized for cross-scene crowd counting.

4. ShanghaiTech dataset: Employed for large-scale crowd analysis.

For the test set images, coordinates representing human heads were transformed into regions using Gaussian functions. The generated data was saved in CSV format for easy viewing through Excel. Each image corresponds to a dedicated Excel file, comprising a matrix with dimensions of 1024 columns and 768 rows, matching all the original image's pixels. As shown in Figure 5, an area of 15x15 centered around the point (103, 95) contains data, with 225 data points corresponding to the first row of the image. Additionally, the sum of these 225 values closely approaches 1.

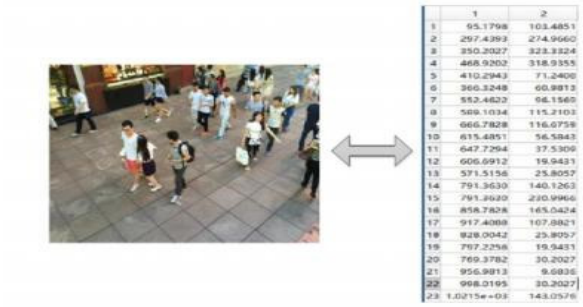

Fig 6. Test set img1 image and corresponding annotation data

This article performed calculations on 300 images from Part B of the ShanghaiTech dataset, which contained a smaller number of individuals. The following graph illustrates the predictive accuracy, with an average accuracy of approximately 0.806. This means that on average, the deviation in detected count, whether overestimation or underestimation, does not exceed 20% of the total count per image.

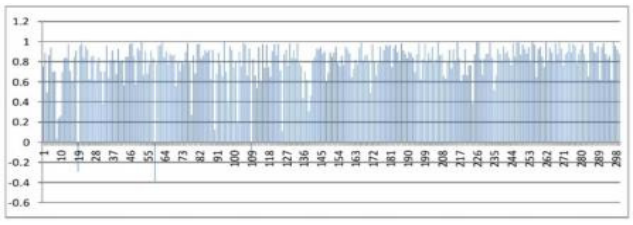

Fig 7. Accuracy diagram of the model

2.3.2. MSCNN Network Construction

Compared to traditional cognitive methods, MSCNN, as a multi-scale convolutional neural network, enables more effective automatic acquisition of information in high-dimensional complex systems through deep learning models. Convolutional neural networks are among the most commonly applied deep learning models, utilizing convolution operations, data pooling, and other techniques to extract features from raw data, and using weighted connections to generate modeling results. The number of convolution operation kernels to some extent affects the efficiency of facial feature extraction. By adopting a multi-scale blob module (similar to the Inception structure), the model can achieve the ability to detect anomalies in public places and monitor crowd movement and density in systems. MSCNN is an improved convolutional neural network that leverages convolution computation kernels of different sizes to extract feature signals from multiple scales, effectively overcoming the adaptive selection issue present in traditional CNN models. The network workflow is as follows.

The foundation of multi-scale information technology is built upon the cornerstone of CNN networks. Its fundamental completion process involves connecting the output characteristics of a central layer or several central layers with the input-output characteristics of the final layer, flattening them to serve as inputs and outputs for the CNN. This, combined with the outputs from higher and lower layers, enables a more comprehensive learning of the entire device degradation process.

By utilizing a multi-scale blob module (similar to the Inception structure), the extraction of relevant features can be achieved. Constructing a Multi-Scale Blob (MSB) structure enhances the diversity of features.

The formation of crowd density maps must undergo Gaussian filtering. A major issue arises when computing in space, as each person represents a pixel, resulting in a very sparse density distribution map. This can lead the model to converge to a completely zero state. However, employing Gaussian data processing helps mitigate the sparse phenomenon to some extent by transforming the density map into a heatmap shape. The calculation of the density map after Gaussian data processing remains constant.

3. Discussion

As urban populations are experiencing exponential growth, the density of crowds in public spaces is continuously increasing. Particularly on weekends and holidays when large numbers of citizens gather in public areas, any sudden abnormal events in these spaces could result in severe consequences. Therefore, the detection of crowd density and monitoring of unusual behaviors in public spaces are crucial for social stability and people's safety.

The continuous development of artificial intelligence technology has become a focal point across various sectors of society. With the explosive growth of data and the constant improvement of hardware technology, deep learning and neural networks have become crucial pillars in many fields.

Exploring how artificial intelligence technology can effectively assist city management personnel in monitoring crowd density, detecting abnormal behaviors, and maintaining crowd density within reasonable limits is of significant practical importance. Ensuring rapid and effective crowd evacuation during emergencies is pivotal. Given the prevalence and high clarity of fixed surveillance cameras, ensuring the accuracy of captured images, this research direction proves feasible.

Regarding the establishment of algorithmic computing platforms, due to the advancement of urban development and the widespread deployment of official surveillance cameras, especially in places with high foot traffic like stations and tourist spots, analyzing the captured footage from stationary camera devices provides clear and highly accurate monitoring. Therefore, analyzing this collected footage becomes a viable direction. For crowd density counting, as abnormal crowd behaviors often occur in an instant, continuous processing and calculation of videos by computing devices are necessary.

Regarding the detection of crowd movement states, a popular method involves direct processing of image data, providing algorithms with good efficacy applicable across

various scenarios, yet this method typically involves substantial computational requirements. Accurately estimating crowd density and obtaining information about crowd distribution from image or video data will play a crucial role in the application of computer vision technology in public safety.

Accurately estimating crowd size, obtaining crowd distribution information from images or videos, and simultaneously monitoring abnormal crowd behaviors are increasingly vital applications of computer vision technology in crowd control and public safety. This research has practically designed a relatively simple yet effective algorithm, indicating significant practical implications at present.

References

[1] Yang Zhe. Crowd Abnormal Behavior detection based on Video surveillance [D]. Yanshan University,2019.

[2] Mo X, Monga V, Bala R, et al. Adaptive sparse representations for video anomaly detection[J]. IEEE Transactions on Circuits and Systems forVideo Technology, 2013, 24(4): 631-645.

[3] Zhang X, Zhang Q, Hu S, et al. Energy level-based abnormalcrowd behavior detection[J]. Sensors, 2018, 18(2): 423.

[4] Zhou Mengxing. Research on Abnormal Behavior detection Algorithm of crowd [D]. Dalian University,2020.

[5] Okusa K, Kamakura T. Human gait modeling and statistical registration for the frontal view gait data with application to the normal/abnormal gait analysis[M]//IAENG Transactions on EngineeringTechnologies. Springer, Dordrecht, 2014: 525-539.

[6] Supreeth H S G, Patil C M. Efficient multiple moving object detection and tracking using combined background subtraction and clustering[J].Signal, Image and Video Processing, 2018, 12(6): 1097- 1105.

[7] Lucas B D, Kanade T. An iterative image registration technique with an applicationto stereo vision[C]. 1981.

[8] G. Csurka, C. R. Dance, L. X. Fan, J. Willamowski, C.Bray. Visual categorization with bags of keypoints. In Pro-ceedings of ECCV InternationalWorkshop on StatisticalLearning in Computer Vision, Grenoble, France,pp.145- 146, 2004.

[9] B. Solmaz, B. E. Moore, M. Shah. Identifying behaviorsin crowd scenes using stability analysis for dynamical sys-tems. IEEE Transactions on Pattern Analysis and Ma-chine Intelligence, vol. 34, no. 10, pp. 2064-2070,2012.DOI:10. 1109/TPAMI.2012.123.

[10] B. Krausz, C. Bauckhage. Loveparade 2010: Automaticvideo analysis of a crowd disaster. Computer Vision andImage Understanding, vol. 116, no.3, pp. 307-319,2012.DOI: 10. 1016/ j.cviu.2011.08.006.

[11] D. Helbing, P. Molnar. Social force model for pedestrian dynamics. Physical Review E, vol. 51, no. 5, pp. 4282-4294, 1995. DOI:10. 1103/PhysRevE.51.4282.

[12] W. Yan, Z. Zou, J. B. Xie, T. Liu, P. Q. Li. The detecting of abnormal crowd activities based on motion vector. Op-tik, vol. 166, pp. 248-256, 2018.DOI:10. 1016/j.ijleo.2017.11.187.

[13] Y. Hao, Y. Liu, J. L. Fan. A crowd behavior feature descriptor based on optical flow field. Journal of Xi′an Uni-versity of Posts and Telecommunications, vol. 21, no. 6,pp. 55-59, 2016. DOI: 10. 13682/j.issn.2095-6533.2016.06.011. (In Chinese).

[14] S. A. Niyogi, E. H. Adelson. Analyzing and recognizing walking figures in XYT. In Proceedings of IEEE Confer-ence on Computer Vision and Pattern Recognition, Seattle, USA, pp. 469-474, 1994. DOI:10. 1109/CVPR.1994.323868.

[15] M. Kass, A. Witkin, D. Terzopoulon. Snakes: Active con-tour models. International Journal of Computer Vision, vol. 1, no. 4, pp. 321-331, 1988.DOI:10. 1007/BF00133570.

[16] J. Wang, Z. J. Xu. STV-based video feature processing for action recognition. Signal Processing, vol. 93, no. 8, pp. 2151-2168, 2012.DOI:10. 1016/j.sigpro.2012.06.009.

[17] C. Van Gemeren, R. Poppe, R. C. Veltkamp. Hands-on: deformable pose and motion models for spatiotemporal localization of fine-grained dyadic interactions. EURASIP Journal on Image and Video Processing, vol. 2018, Articlenumber 16, 2018. DOI: 10. 1186/s13640-018-0255-0.

[18] X. F. Ji, Q. Q. Wu, Z. J. Ju, Y. Y. Wang. Study of human action recognition based on improved spatio-temporal fea-tures. International Journal of Automation and Comput-ing, vol. 11, no. 5, pp. 500-509, 2014. DOI: 10. 1007/s11633-014-0831-4.

[19] E. H. Aldelson, J. R. Bergen. Spatiotemporal energy mod-els for the perception of motion. Journal of the Optical So-ciety of America A, vol. 2, no. 2, pp. 284-299, 1985. DOI:10. 1364/ JOSAA. 2.000284.

[20] Y. Iwashita, M. Petrou. Person identification from spatio-temporal volumes. In Proceedings of the 23rd Internation-al Conference Image and Vision Computing, IEEE, Christ-church, New Zealand, 2008. DOI: 10. 1109/IVCNZ.2008.4762086.

[21] S. A. Niyogi, E. H. Adelson. Analyzing and recognizing walking figures in XYT. In Proceedings of IEEE Confer-ence on Computer Vision and Pattern Recognition,Seattle, USA, pp. 469-474, 1994. DOI:10. 1109/CVPR.1994.323868

[22] R. C. Bolles, H. H. Baker, D. H. Marimont. Epipolar-plane image analysis: an approach to determining structure from motion. International Journal of ComputerVision, vol. 1, no. 1, pp. 7-55, 1987. DOI: 10. 1007/BF00128525.

[23] H. H. Baker, R. C. Bolles. Generalizing epipolar-plane im-age analysis on the spatiotemporal surface. In Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, Ann Arbor, USA, pp. 33-49, 1988. DOI: 10. 1109/CVPR.1988.196209.

[24] G. Kuhne G, S. Richter, M. Beier. Motion-based segment-ation and contour-based classification of video objects. In Proceedings of the 9th ACM international conferenceon Multimedia, Ottawa, Canada, pp. 41-50, 2001.DOI:10. 1145/500141.500150.

[25] G. Kuhne G, S. Richter, M. Beier. Motion-based segment-ation and contour-based classification of video objects. In Proceedings of the 9th ACM international conferenceon Multimedia, Ottawa, Canada, pp. 41-50, 2001.DOI:10. 1145/ 500141. 500150.

[26] C. E. Shannon. Amathematical theory of communication. Bell System Technical Journal, vol. 27, no. 3, pp. 379-423, 1948. DOI: 10. 1002/j.1538-7305. 1948.tb01338.x.

[27] K. He, S. X. Wang. Study on denoising of fractal signal based onShannon entropy. In Proceedings of Internation-al Conference on Neural Networks and Signal Processing, IEEE, Nanjing, China, pp. 751-755, 2003.DOI:10. 1109/ICNNSP. 2003. 1279384.

[28] S. Liang, Y. Ma, Y. Y. Huang, J. Guo, C. F. Jia. The scheme of detecting encoded malicious web pages based on information entropy. In Proceedings of the 10th Interna-tional Conference on Innovative Mobile and InternetSer-vices in Ubiquitous Computing, IEEE, Fukuoka, Japan,pp. 310-312, 2016. DOI: 10. 1109/IMIS.2016.82.

[29] Z. J. Zhang, X. N. Wang, L. Sun. Mobile payment anom-aly detection mechanism based on information entropy.IET Networks, vol. 5, no. 1, pp.1-7, 2014. DOI:10. 1049/iet-net. 2014. 0101.