

CROWD ABNORMAL BEHAVIOUR DETECTION USING DEEP LEARNING

Riddhi Sonkar^{1,3,*}, Sadhana Rathod^{2,**}, and Renuka Jadhav^{3,***} Deepali Patil^{4,****}

¹Department of Information Technology, Ramrao Adik Institute of Technology, Navi Mumbai, India

²Department of Information Technology, Ramrao Adik Institute of Technology, Navi Mumbai, India

³Department of Information Technology, Ramrao Adik Institute of Technology, Navi Mumbai, India

⁴Department of Information Technology, Ramrao Adik Institute of Technology, Navi Mumbai, India

Abstract. Crowd analysis has become an extremely famous research point in the territory of computer vision. Computerized examination of group exercises utilizing reconnaissance recordings is a significant issue for public security since it permits the identification of hazardous groups and where they're going. We all see how many problems are faced because of the crowd. In our country, many terrorists are there. They plant a bomb in a crowded area which causes a lot of injuries. Thieves are mostly found or always leave in crowded areas so they can easily get an advantage of the crowd. In that situation, crowd analysis is very important. This paper presents the design of the deep learning architecture that provides control over the crowd behavior that will help to avoid violence or any other act which occurs because of the crowd which causes harmful effects to the society. So we are proposing a system that detects abnormal behavior of crowds using deep learning techniques.

Keywords— Crowd analysis, pre-processing, object tracking, CCTV, Machine Learning, CNN, event behavior recognition.

1 Introduction

A crowd is countless individuals assembled in a scattered or uncontrollable manner. It is a gathering of individuals assembled in a specific area. Crowds vary in various situations, for example, the crowd in a sanctuary will be unique in relation to the crowd in a shopping zone. The setting where the word 'crowd' is utilized shows the kind of gathering as far as size, duration, composition, inspiration, cohesion, and nearness of people.

Nowadays, so as to ensure social security, video surveillance frameworks whose equipment cost is low, have been broadly applied out in the public places, for example, air terminals, metro stations, traffic intersections, and schools, and so on. In any case, the greater part of the current video surveillance frameworks are simply utilized as a record framework, which can't identify and dissect an unusual occasion naturally, and it is inconceivable for an individual to watch the screen consistently. [1]

Since the observing offices are ordinarily mounted out in the public places, abnormal events discovery in crowd has been basic and has become another region of enthusiasm for the wise checking framework look into network the outcomes might be utilized for public security purposes [2].

Then again, traditional methods for individual behavior analysis can't be suitable for crowded scenarios due to

the occlusion phenomenon. Video surveillance in packed territories is getting progressively huge for public security. Analyzing these crowd behaviors is crucial so any basic circumstance can be controlled before it turns out to be more serious. Videos of crowded scenes present testing issues in computer vision [3]. Video surveillance assumes an essential job in the present society in guaranteeing security in the region where huge quantities of individuals are relied upon to accumulate. An expansion in security threats has utilized surveillance systems in practically all the region. For instance, CCTV cameras are installed in the roads, shopping edifices, temples, stadiums, and so on., to guarantee security. Crowd behavior analysis encourages from numerous points of view to all the more likely comprehend crowd elements and related individuals behavior create reconnaissance or crowd control systems, organize and compose open spaces and improve computer animation models utilized with regards to video games or special effects.

The structure of society has become more complex than it used to be. The people living in society have become intolerant towards sensitive issues persisting in society [4]. This has led to increased instances of public violence. Moreover, the weak security measures have made violence break out in the crowd easier. The present practice is physically checking video takes care of from a few sources. Video analytic permits the the automatic detection of events of interest, however it faces numerous difficulties as a result of non-inflexible crowd movements and impediments. The algorithms developed for rigid ob-

*e-mail: riddhisonkar20@gmail.com

**e-mail: sadhanarathod98@gmail.com

***e-mail: renuka24jadhav@gmail.com

****e-mail: deepali.patil@rait.ac.in

jects are ineffectual in managing crowds. The aim of this work is to develop a system which can classify normal and abnormal behaviors in crowds using a real-time video surveillance system to monitor crowded urban environments using machine algorithm techniques [5]. This system provides rein in crime because the recorded footage is substantial evidence against the criminal and because of crowd analysis will help security agencies to stop the crowd criminal activities like riot etc.

2 Review of Literature

Numerous works have been done related to the crowd behavior analysis using different machine learning techniques. The datasets, algorithms, methods used by the authors and observed results along with the future scope are carried out in finding out efficient methods of crowd abnormal behavior detection with different types of crowded scenes. Here research on crowd behavior analysis has been carried out for several years. This section gives information of the existing system available and gives information of the various techniques used.

Lazaros Lazaridis. [1] proposed new techniques for the detection and characterization of unusual behavior in dense crowd for example the approach used to create the crowd density heat-maps, the extraction of the related optical flow. But when talking about Abnormal behavior detection it should cover the weather condition, lightning condition and camera angle but it is difficult to gather genuine information with abnormal behavior in dense crowds and that's why it will not give the more accurate results.

Vishwa Patel. [2] used object based behavior reorganization method which is transfer on the information on the people that structure the group, which was trying for denser situations. It is just reasonable for low and moderately crowded scenes. So they additionally utilized the second methodology that is all holistic approach regards the crowd as a solitary element, with the goal that the tracking issue is not a challenge. This subsequent methodology is progressively suitable for denser scenes, yet it can't identify abnormal behavior effectively.

Krutika Rohit. [6] presents behavior label distribution method, that can be used for solving problems of mixed behavior. By using behavior label distribution archived public security and safety. But this technique contains so many steps like giving sequence numbers and labels to every behavior and then analyzing the abnormal behavior. This process is time consuming and this system is not real time so that it is not much efficient to handle critical situations.

Swathi H Y. [7] discussed the various techniques utilized for recognizing unusual crowd behavior viz., crowd density estimation, tracking, motion detection and behavior recognition/understanding. They likewise examined the issues in crowd analysis and the answers for these issues. This work is their initial phase in the research of crowd behavior analysis, there are a few issues despite everything being confronted in regards to the efficiency, effectiveness, robustness of the automated surveillance systems for crowd analysis.

Dongping Zhang. [8] proposed another technique for crowd motion behavior detection that is SIFT flow innovation is used for identifying the movement data from the two adjacent frames in the video arrangements, at that point the weighted direction histogram which assumes the job of a factual estimation for the SIFT flow is taken as the contribution for the HMM model.

Marc Van Droogenbroeck. [9] presented a widespread sample based background subtraction algorithm, called ViBe, which consolidates three creative strategies i.e. classification model, background pixel model, neighboring pixel model.

Yong Shean Chong. [10] Used a novel system for representing video information by a lot of basic highlights, which are implied naturally from a long video film through a deep learning approach. In particular, a deep neural network made out of a measure of convolutional autoencoders was utilized to process video frames that caught uncommon structures in information, which gathered, make the video representation. This representation is taken care of into a measure of convolutional brief autoencoders to become familiar with the regular temporal patterns. Right now, neural network in that, all data sources and yields are autonomous from one another.

R. Mehran. [11] proposed another algorithm for discovery and confinement of abnormalities present in crowded scenes recordings by utilizing the SFM association power in blend with the Particle crowd Optimization. This work lies in presenting the enhancement of the social force and performing molecule shift in weather conditions to acquire the improved association power as per the fundamental optical flow field. The entire irregularity detection/localization process is completed with no learning stage. This infers the proposed strategy is very appropriate for genuine situations.

Weiming Hu. [12] introduced a review of ongoing improvements in visual reconnaissance inside a general handling structure for visual observation frameworks. The best in class of existing strategies in each key issue is depicted with the emphasis on the following errands: recognition, following, comprehension and depiction of practices, individual recognizable proof for visual reconnaissance, and intelligent observation utilizing numerous cameras. With respect to the discovery of moving items, it includes ecological demonstrating, movement division and article grouping. Three systems for movement division are tended to: foundation subtraction, fleeting differencing, and optical stream.

But still, there is a need to detect abnormal activity of crowds at an early stage so that it will help us to prevent problems faced because of the crowd. That motivates us to build a system which detects abnormal behavior of crowds.

3 Proposed Methodology

The proposed system consists of planning and executing the system which secured the answers for the existing system required. The task additionally includes an extensive assessment of the system, and is to give proficient se-

curity to Urban by building up an Unsupervised Abnormal Crowd Behavior Detection system. We developed a system which can characterize normal and abnormal behaviors in crowds utilizing a constant video observation system and using CNN algorithm to analyze monitoring crowded urban environments.

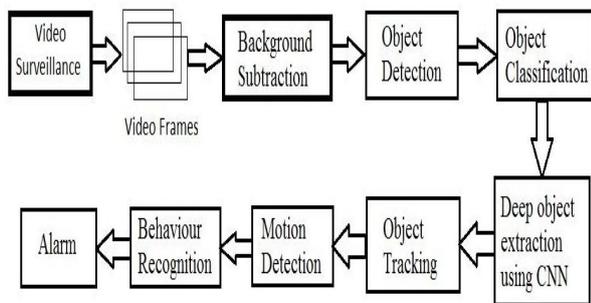


Figure 1. Proposed System

Initially, we did an analysis of existing systems that are available, but after analysis we came to know that there is a need for some real-time system that may overcome the drawbacks of an existing system. So we started research and we found that only CNN is not comfortable for real time behaviour detection so that we use both CNN and KNN in proposed system.

When we give video input i.e. video surveillance to the system, it is converted into video frames which are in “.jpg” format. Background subtraction is a technique for isolating out closer view components from the background and is finished by creating a frontal area veil. This technique is utilized for recognizing progressively moving objects from static cameras. Background subtraction procedure is significant for object tracking. There are a few methods for Background subtraction, we are using ViBe algorithm for that. Then it will detect and classify and track the object. CNN algorithm is used to deep object extraction. Deep object extraction is use for identify human begins in frame. We use KNN for calculate position difference between 2 sequence frames of object. On the bases of euclidean distance we track the object. Then it will analyze the motion of objects on the bases of 3 parameters i.e. speed, angle, direction. Using these parameters if the motion of object is beyond the threshold then it is recognize as abnormal behavior. If the behavior is abnormal then it delivers an alarm. The threshold is changes according to application.

4 Algorithms

Algorithms used to build the proposed system are described as follows:

4.1 ViBe

ViBe is a universal sample-based background subtraction algorithm. Background subtraction is the way toward

isolating out frontal area objects from the Background during a sequence of video outlines. sequence subtraction is a generally utilized methodology for discovery of moving objects from static cameras[9]. This methodology varies from those dependent on the traditional conviction that the most seasoned qualities ought to be supplanted first. At long last, when the pixel is seen as a major aspect of the foundation, its worth is proliferated away from plain sight model of a neighboring pixel. We depict our strategy in full subtleties and contrast it with other foundation subtraction strategies [13]. Productivity figures show that our strategy outflanks later and demonstrated cutting edge techniques regarding both calculation speed and recognition rate. We likewise examine the presentation of a down-scaled variant of our calculation to irrefutably the base of one examination and one byte of memory for each pixel. It gives the idea that even such an improved adaptation of our calculation performs superior to standard procedures.

4.2 Convolutional Neural Network (CNN)

Convolutional neural networks are most useful with very large data sets, large numbers of features and complex classification tasks. A Convolutional Neural Network (CNN) is a type of neural network algorithm that is also used in deep learning and it comprised one or more layers. The CNN algorithm is designed to take images as an input to train the system to give the predictions. It builds the features from the various groups of pixels. CNN’s are simpler to prepare and have numerous less parameters than completely associated with the systems with a similar number of hidden units. Convolutional neural systems comprise of an input layer, an output layer, and hidden layer operations.

A neural system comprises of a few distinct layers, for example, the information layer, in any event one concealed layer, and a yield layer. They are best utilized in object discovery for perceiving examples, for example, edges, shapes, hues, and surfaces. The shrouded layers are convolutional layers right now neural system which acts like a channel that initially gets input, changes it utilizing a particular example/include, and sends it to the following layer. With more convolutional layers, each time another information is sent to the following convolutional layer, they are changed in various ways. For instance, in the first convolutional layer, the channel may distinguish shape/shading in a locale, and the following one might have the option to close the item it truly is , and the last convolutional layer may characterize the article as a pooch. Essentially, as an ever increasing number of layers the info experiences, the more refined examples the future ones can identify.

For a frame t, the output highlight of 5th convolution layer is a 6h6h256 vector. In a standard CNN, the high-dimensional convolution highlights are normally taken care of to fully connected layers for dimensionality decrease, and then input to a classifier. Be that as it may, in FCN, there are no completely associated layers, and we are not preparing a classification task. So we dispose of heavy classifier, for example SVM, and utilize a novel strategy

for dimensionality decrease as well as abnormal recognition.

We encode the high-dimensional convolution includes as a set of binary codes by iterative quantization (ITQ) [13]. ITQ is a proficient unsupervised method for learning binary codes in huge scope picture assortments. ITQ reduces the dimensionality of information by PCA first, at that point finds a rotation of zero-centered data to limit the quantization blunder of mapping this information to binary codes. Preparing the ITQ costs not as much as preparing a classifier, and we can extend the identification task simpler dependent on the binary codes.

4.3 K-Means

K-Means clustering algorithm is an unaided algorithm and it is utilized to section the intrigue zone from the foundation. It groups, or segments the given information into K-clusterd or parts dependent on the K-centroids. The algorithm is utilized when you have unlabeled data(i.e. information without characterized classes or gatherings). The objective is to discover certain gatherings dependent on a closeness in the information with the quantity of gatherings group to by K.

A K-Means clustering algorithm is used for motion detection of the objects which are detected by CNN. We created the cluster of objects pixels. Then we calculate the difference between the position of that object and calculated Euclidean distance on the bases of that we set the threshold value for detecting abnormal behavior of the crowd.

Knn is used to calculate the points distance between two image frame in sequence to track the motion of moving objects.

5 Results

Success of intensive learning network training is subject to the presence of enormous datasets. Large buildings, commented on with the mark of ground truth, the dataset is over the top expensive because of the need of human time and exertion. A decent Dataset needed to detect unusual phenomena in crowded scenes be different, occupying every single imaginable part of the issue; And fine annotate. It should cover different seasons Conditions, power conditions, different camera angles and Crowd density. Furthermore, due to legal and confidentiality rules it is incredibly hard to gather genuine information with unusual behavior in dense crowds. After analyzing some of the most well known datasets related to abnormal crowd behavior, we conclude the UMN dataset [15]: this dataset includes panic congestion behavior in a non-reasonable way. Novel violent [16]: By comparing this dataset with UMN, which is progressively reasonable however it needs crowd density.

The proposed network gives better results than those reported in the literature. Table-I [1] is a confusion matrix of existing system which uses Density Heat maps and Optical Flow for Abnormal behavior detection in crowd

scenes. On the basis of confusion matrix the accuracy of existing system is 77.08% Whereas Table-II is confusion of proposed system. On the basis of confusion matrix the accuracy of proposed system which is based on convolutional neural network and K-Means algorithm is 80.79%.

Table 1. CONFUSION MATRIX OF THE EXISTING METHOD [1]

Actual/Predicted	Actual	Predicted
Actual	83.3	5.8
Predicted	37.7	61.2

Table 2. CONFUSION MATRIX OF THE PROPOSED METHOD

Actual/Predicted	Actual	Predicted
Actual	51	3.34
Predicted	12	12



Figure 2. Some Abnormal activity is detected

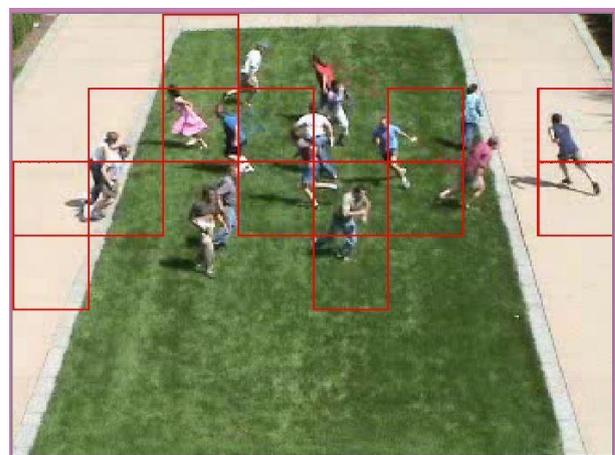


Figure 3. Abnormal activity is detected

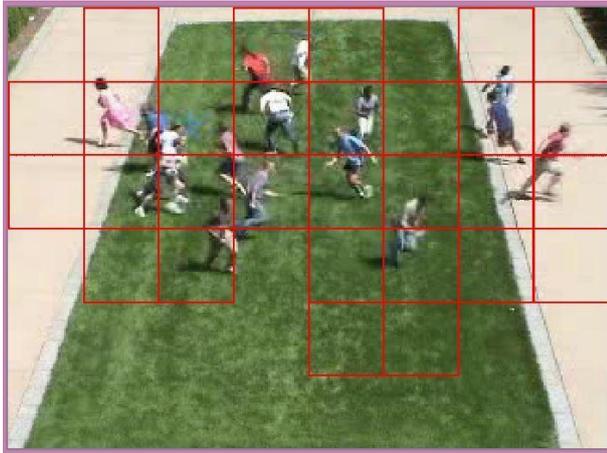


Figure 4. Abnormal Behavior is detected

References

- [1] LazarosLazaridis, AnastasiosDimou, PetrosDaras, "Abnormal Behavior Detection in Crowded Scenes Using Density Heatmaps and Optical Flow" 26th European Signal Processing Conference (EUSIPCO), 2018.
- [2] Vishwa Patel, Narendra Patel, MahaswetaJoshi, "A Review: Abnormal Crowd Behavior Detection Techniques" ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887 Volume 6 Issue II., IEEE, 2018.
- [3] G.Santhiya, K.Sankaragomathi, S.Selvarani, A.Niranjan Kumar, "Abnormal Crowd Tracking and Motion Analysis" IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), 2014.
- [4] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in Computer Vision (ICCV), 2015 IEEE International Conference on. IEEE, 2015, pp. 4489–4497.
- [5] J. R. Medel and A. Savakis, "Anomaly detection in video using predictive convolutional long short-term memory networks," arXiv preprint arXiv: 1612.00390, 2016.
- [6] Krutika Rohit, Ms. Kinjal Mistry and Mr. JenishLavji, "A Review on Abnormal Crowd Behavior Detection" International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017.
- [7] Swathi H Y, Dr. H S Mohana, Dr. G Shivakumar, "Crowd Behavior Analysis: A Survey" International Conference on Recent Advances in Electronics and Communication Technology, 2017.
- [8] Dongping Zhang, KaihangXu, Huailiang Feng, Ye Shen, "Abnormal Crowd Motion Behaviour Detection based on SIFT Flow" International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.9, 2016.
- [9] Barnich, Olivier Droogenbroeck, Marc. "ViBe: A Universal Background Subtraction Algorithm for Video Sequences. Image Processing", IEEE Transactions on. 20. 1709 - 1724. 10.1109/TIP.2010.2101613, 2011.
- [10] Y. S. Chong and Y. H. Tay, "Abnormal event detection in videos using spatiotemporal auto encoder," in International Symposium on Neural Networks. Springer, 2017, pp. 189–196.
- [11] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in Computer Vision and Pattern Recognition, CVPR 2009. IEEE Conference on. IEEE, 2009.
- [12] Weiming, H., et al., "A survey on visual surveillance of object motion and behaviors." Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions, Vol. 34, No. 3, (2004), pp. 334-352.
- [13] Barnich, Olivier; Van Droogenbroeck, Marc (2009). "ViBe: A powerful random technique to estimate the background in video sequences", IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 945–948. doi:10.1109/ICASSP.2009.4959741. ISBN 978-1-4244-2353-8, 2009.
- [14] Hongtao Wei1, Yao Xiao, Ruifang Li, Xinhua Liu , "Crowd abnormal detection using two-stream Fully Convolutional Neural Networks", 10th International Conference on Measuring Technology and Mechatronics Automation , 2018.
- [15] Ucf web videos dataset. http://www.vision.eecs.ucf.edu/projects/rmehran/cvpr2009/Abnormal_Crowd.htmlVI. Downloads. 4, 7
- [16] Detection of unusual crowd activity dataset. http://mha.cs.umn.edu/proj_events.shtmlcrowd. 4, 6
- [17] L. Zeng, X. Xu, B. Cai, S. Qiu, and T. Zhang, "Multi-scale convolutional neural networks for crowd counting," arXiv preprint arXiv: 1702.02359, 2017.
- [18] T. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent flows: Real-time detection of violent crowd behavior," in Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE Computer Society Conference on. IEEE, 2012, pp. 1–6, 2012.