

Crowd Management and Monitoring using Deep Convolutinal Neural Network

Pratiksha Singh

Department of Computer Science Engineering, Madan Mohan Malviya University of Technology, Gorakhpur, India

A K Daniel

Department of Computer Science Engineering, Madan Mohan Malviya University of Technology, Gorakhpur, India

Corresponding author: Pratiksha Singh, Email: pratikshasingh1212@gmail.com

India is counted as one of the most populated countries in the world. A lot of crime is also increasing due to the increasing population, as criminal activities are more frequent in a crowded place. Being crowded is also facing a lot of diseases. Therefore, crowd management and monitoring are very important therefore viewed from the security, crowd management, and monitoring plays a very important role in identifying group/ individual's behavior in a crowd using video and image sequence for counting the person and detection of such misbehavior elements, This paper proposed a model for crowd management and monitoring person counting as object detection techniques. This paper proposed Deep Convolutional Neural Network, and Support Vector Machine. The data set are taken from Mall, Kumbh Mela, and UCFD. The performance of the model using training and testing of data is improved.

Keywords: Crowd management, Crowd monitoring, Crowd detection, Radio frequency identification, Support vector machine, Deep convolutional neural network.

1 Introduction

India is second in the world in increasing population rate of humans. And the rate of human population has been increasing very rapidly in the last few times and years, that is why crime has also increased due to increasing congestion in the field of development [1], [8]. Basically, the crowd is a group of organized or different people in a group and is inspired by common goals. Several methods of technology are used to count congestion. There are mainly two types of crowd displays, structure crowds and unstructured crowds [7]. In many areas such as markets, towns, college campuses, hospitals, airports, stadiums, shopping malls and cultural and religious places [1], crowds are monitored by video cameras. Computer vision congestion events have been the subject of a core research. This paper decodes both profitability and cutting-edge methods. It is very essential to understand the behavior of the crowd from the research point of view, because based on this we use different methods and technologies. It is only through the behavior of the crowd that activities in the social space are shown such as crime activities, Terrorist attack etc. Several methods of algorithm have been used in computer visions such as crowd monitoring, crowd behavior, detect on the person in the crowd, and find out the crowd [10]. Today many wireless devices and sensors are available such as radio frequency identification is one of them and it has been used for efficient crowd management, Deep Convolutional Neural Network and Support Vector Machine in efficient crowd management and monitoring both.

Monitoring the crowded area and place monitoring the crowd is a task full of challenges and difficulties, explaining a wide variety of activities. Identify the behavior of a crowd becomes unpredictable for some time. The crowd may be involved under the same act or for the same event and due to non-restructuring of this behavior of the crowd, crowd management takes some time apart [17]. Because human behavior is different even if it is related to the same purpose or does not recognize that the main objective is dangerous [9]. Humans and traffic cause jams and risks. Corona, swine flu, bird flu etc. viruses are also likely to spread as these viruses are caused by skin-to-skin contact status [5]. It is very important to understand their behavior even to tell riots and terrorist attacks in a public place due to the mob. Several methods and techniques have been described to secure the individual and the environment and the developed environment. Several computer vision algorithms exist to describe congestion behavior by tracking, but sometimes this algorithm also fails for some reason [18], [19].

2 Crowd Model

The crowd model has been classified into three categories: crowd management, crowd monitoring (behaviour, localisation, and counting), and crowd detection.

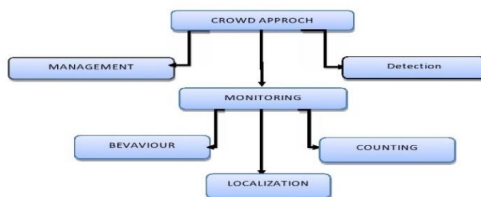


Fig. 1. Defined crowd Models

A) Crowd Management

Crowd management is a collected and proven plan and a direction towards the gradual progress of events where a lot of people gather. Also, the field of crowd management has improved a lot throughout the year. Real-time has been used to manage and monitor [9]. A large amount of people gathers while going to Hajj, and the round robin algorithm has been structured to control the crowds in Hajj. Crowd management is not just for the event but including shopping malls, plazas, airports, etc all locations. Practical groups of people can be used as a part of crowd management

with the aim of making them direct and limited, this is called controlling congestion and is essential for social security.

B) Crowd Monitoring

CCTV cameras are used to monitor congestion, but CCTV cameras are also unable to cover certain areas such as side of the wall. This is the biggest loss of CCTV cameras. Many techniques are used to monitor congestion such as crowd counting, crowd localization, crowd density and crowd behavior etc.

- Crowd counting entails counting a person present in an area and informing him of the number of individuals present in that area using a variety of technological approaches. [1].
- The term "crowd location" refers to where a person exists and how many persons are in that area, And to indicate which person has a specific place.
- The density of congestion means how many people are present in a person and an area and what the distance between them is and how much more or less [8].
- Behavior of crowd tells the activities of the crowd what the reason for the gathering of the crowd is and what the main objective is. For example, if a person goes to the airport, it means that the activity there are a traveling and place, and if a person going to cultural place is mean that cultural events are happening in that place and crowd behavior technologies have been used support vector machine and deep learning techniques [11][12].

C) Crowd Detection

Determines a number by identifying individuals located at a location. The model that detects this function is called the identification model. Some researchers have described detecting crowd attendees by vision-based and wavelet template techniques in which we call HWD (head wavelet detection) [8]. This technology has been used for the main function of HWD technology, to tell the head size a specific description of the attribute, and SVM was used primarily to classify the presence of the head [21]. And some technology and algorithms have also been analyzed that identify the area of the human shoulder and head. The size of a human's shoulder and head area is like an omega (Ω). And to identify humans in the crowd, the ROI is the filter and then the effect is based on different approaches in less congestion and successful and more density congestion [8].

3 Methodology

There are various methodologies by which the object can be identified in a group set and the group using various techniques.

A) Radio Frequency Identification

RFID (Radio Frequency Identification) has been used to identify crowded individual. It has been used in every area for indoor and outdoor. It has been used to count the ruse of real-time. RFID perform the people identify through of radio frequency and radio frequency provide unique tag chip identification card and a card is identified through the wireless sensor and identified to detect via of GPS tracking device [13][18][15]. RFID chip are available in various sizes, sensor network for reading and sensing chip would have some serious economic considerations in case of irregular event like Kumbha [5]. RFID technology has even been successfully utilized for identifying purposes.

B) Support Vector Machine

Support vector machine (SVM) algorithm for popular supervised machine learning algorithm, the crowd density management introduced a support vector machine technique using higher-order singular value decomposition (HOSVD). [16]. it is also classifying the regression problem for crowd control is SVM algorithm the data set display a range for training and testing dataset through supervised learning.

C) Deep Convolutional Neural Network

Many methods of model work inside of deep learning such as Recurrent Neural Network (RNN), Auto encoders (AE) and Artificial Neural Networks (ANN) but Convolutional Neural Networks (CNN) very special model. This model is useful for computer vision and image analysis [8]. CNN is an assortment for Deep neural network and classifies by identifying of special feature of the image. It provides a comprehensive look for visual objective and in this application the validation of image and video such as Computer Vision, Medical Image analysis (MIA), image classification and natural language processing (NLP) are composite. CNN display the mathematical function and his calculate congestion. In a process known as Feature Extraction, a convolution tool isolates and identifies the different characteristics of a picture for analysis.

There are three-layer Pooling layer, Convolutional layer, and Full connected layer [20], and comprised of two additional parameters like dropout and active function and with this help to show its result.

4 Proposed Model

This proposed model consists of data set as input and counting and localization as the techniques used for analyses and given objective.

A) Counting

However, a description of the person present in a video and image and its status cannot be fully communicated. Nevertheless, we have used a variety of applications to crowd have tried to identify size and person. It breaks the technical and image into a small image and calculates the distance and average density between the person present in each image and the detector while many places are such where image calculation is a very challenging task [27]. The area that can be captured cleanly by image and video. The exact average density of the areas of that is possible. Additionally, person counting, and density is also possible through serial images. This work is displayed by map, so and map is used all over the world to calculate ground areas. Calculating the people present in the crowd is also a complex task; MAE and MSE try to make it comfortable.

Average absolute error (MAE) determines the absolute average distance between real and approximate data, although this high prediction does not punish mistakes. The average distance squared between actual data and estimated data is measured by the mean square error (MSE). Major mistakes are highlighted in this section.

$$MAE = \frac{1}{N} \sum_{j=1}^N |V_j - V'_j| \dots\dots (1)$$

$$MSE = \sqrt{\frac{1}{N} \sum_{j=1}^N |V_j - V'_j|^2} \dots\dots (2)$$

The total number of data points N, prediction value from Vj and then and jth truth value from V'j and MAE is mean absolute error and MSE is mean squared error value.

B) Localization

This training does not fully claim the crowd's location will give accurately. Because there is a very challenging task in localizing the crowd. It has been tried as an image and video. If viewed realistically, the same head can match in multiple heads. The person present at one place is checked with every angle to localize the person, and then try to demonstrate a distance estimate. And the effort made has led to confirmation localization [28].

$$F - Measure = \frac{2 \times precision \times Recall}{Precision + Recall} \dots (3)$$

$$Precision = \frac{True\ Positive}{Predicted\ value} \dots\dots (4)$$

$$\text{Predicted} = \text{True Positive} + \text{False Positive} \dots (5)$$

$$\text{Recall} = \frac{\text{True positive}}{\text{Actual Value}} \dots (6)$$

$$\text{Actual value} = \text{True positive} + \text{False Negative} (7)$$

where True Positive, False Positive and False Negative. Normally, box level Precision, Recall, and F-measure are utilized for crowd localization tasks.

C) Dataset

For the result we need some real-world data sets such as crowd videos and crowd pictures. For this we have some data set on the publicly such as the data set of the mall, fair, Kumbh Mela, classrooms, events.

Mall's data are highly used, and they are capture with the help of CCTV camera all around area or have publicly data set This includes 2000 maximum videos and image and the person counted in this 2000 image and video is more the about 62325 more people. But due to the ongoing disease in current time the crowd count is reduced [23].



Fig. 2. Mall dataset-based count person

The UCSD dataset is the first dataset to counting person and this data set is used by installing the camera and the ends of the person in the camera present in it are made by highly resolution video and image in frame or at least 50 frames have been used and it has an overall population of 49885 people. It has mostly been used in place like mall, political areas, and stadium [22].

Table 2. Defined by mean squared error and mean absolute error value UCSD Data set

Method	MAS	MSE
Gaussian process regression	2.23	7.95
Ridge regression	2.24	7.80
Density map + MESA	1.71	---
Count forest	1.61	4.41
Multi-column CNN	1.08	1.34

The Hindu festival Kumbh Mela to be held in India, which is a crowd data set in Allahabad, Nashik, Ujjain, and Haridwar is collected by drones and is a mass pilgrimage held every 12 years and a huge amount of crowd would gather. Whose management and monitoring are very important. According to Allahabad Kumbh mela Data, the mela held in 2013 collects more than 120 million people and then has more than 600 frames in 6 hours and it sets a very large amount of data [25].



Fig. 3. Kumbh Mela based every 6 hours Millen are person count

NWPU data set is a publicly collected data set. It works in deep CNN methods based on small scales. NWPU dataset has a dense congestion limit. NWPU crowd data set has about 2133375 head count based on 5120 images did has gone [26].



Fig. 4. NWPU dataset

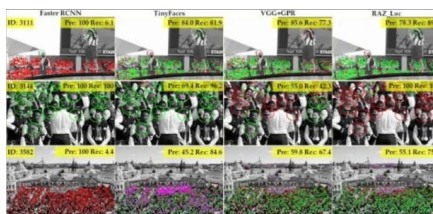


Fig. 5. NWPU person evolution and person count

Shanghai Technology is a very large data set and uses at least 1198 crowd image to count people and it divided into two parts of the data set which is first on his 482 image and the second which is in part there are 716 images, and the first part is divide two-way as testing and training dataset and second images also divide two-way training and testing dataset and it states the best result [24].



Fig. 6. ShanghaiTech Part A, Part B Based person count

The **World Expo'10** data set was first displayed by Sam et al. In this data set, 108 consecutive cameras were installed for surveillance at the World Expo in Shanghai City in 2010. This camera captured 1,132 annotated videos and 3,980 frames. The frame had annotation of the cantor of individuals running continuously on 1, 99, 923 [27][28].

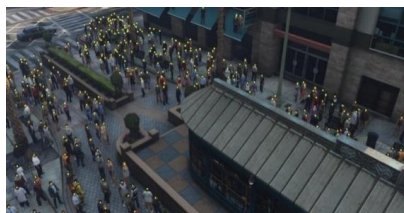


Fig. 7. Would expo 10 dataset to count person with help people head

Table 3. Defined person counting based on MSE and MAS Would Expo'10 data set

Method	MA	MS
	S	E
Gaussian process regression	2.23	7.95
Ridge regression	2.24	7.80
Density map + MESA	1.71	---
Count forest	1.61	4.41
Multi-column CNN	1.08	1.34

Table 4. Comparison of varies nine real-world public datasets based of frames

Data Areas	Purpose	No of Image	Person head counting	From Use	Device
Mecca madina	Counting and behavior	---	----	Images, videos	CCTV Surveillance
Mall	Counting and detection	2000	62325	Images, videos	CCTV Surveillance
Shanghai tech part A, B	Counting	1198	482,716	Images, videos	CCTV Surveillance
CUHK	Counting	1535	----	Images, videos	CCTV and internet
UCSD	Counting	2000	49885	Images, videos	CCTV Surveillance
NWPU	Counting	5120	21,33,345	Image	Camera
Kumbh mela	Counting and localization	6144	120 million	Images, videos	CCTV Surveillance and Drone
UCF_CC_50	Crowd management	50	1279	Images, videos	CCTV
Would expo'10	Person counting	1132	1199,9233	Image videos	CCTV

5 Performance Analysis of Model and Result

A) Counting

The ground truth image is called a sample image. It states the exact position of the person. Support vector machines have been used to make this image accurate. In this study we have used the data set of the mall. From the point of training, we have selected mall data sets and video sequential images, and the conclusion is displayed in the picture.

**Fig. 8.** Target Image (Mall Data set)**Fig. 9.** Support vector machine base crowd counting

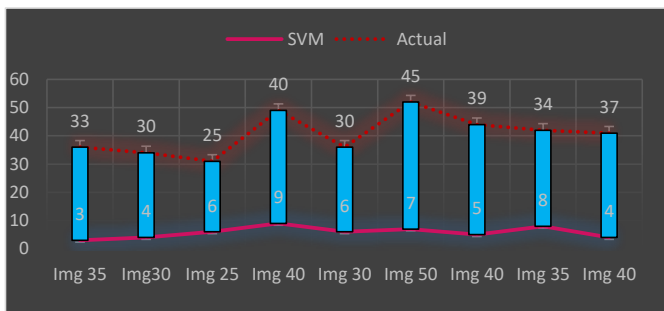


Fig. 10. Actual value and SVM method Comparison based person counting

Table 5. Define same image different-different area count person SVM techniques with image range

Range (Image)	SVM	Actual value
Img1 (35)	3	33
Img2 (30)	4	30
Img3 (25)	6	25
Img4 (40)	9	40
Img5 (30)	6	30
Img6 (50)	7	45
Img7 (40)	5	39
Img8 (35)	8	34
Img9 (40)	4	37

B) Localization

In many applications, as such the technology of high-density crowd tracking has been used. While ground veracity has been used 1–1 to indicate the error of permanent taxation. F-Measure in collaboration with a Greedy Association, have tried local calculations from Recall and Precision. The area of the recall, precision curve is also known as L-AUC. It has been used to display overall performance [27].

Table 6. Localization Accuracy on using different date set

Method	Recall	Precision	F-measure
M-CNN	64%	60%	62%
LSC-CNN	73.55%	75%	74%
CL-CNN	60%	76%	67%
D-CNN	79%	82%	81%

6 Conclusion

Crowd management, monitoring and counting analysis has gained popularity in recent years for identifying individual's behavior and misbehavior in video and image sequence. The analysis of multiple crowd management and monitoring approach had been discussed. Crowd management and monitoring are complicated attributed to reason such as lighting variations in each picture scene. The paper including the finding using the Deep convolution neural network and SVM based crowd management and monitoring technique had been applied for more detailed information. The model is implemented for evaluation of Mall, Kumbh Mela, using would expo'10, Shanghai Tech, NWPU and UCFD dataset.

References

- [1] S. Jeevitha and R. Rajeswari, "A Review of crowd counting techniques", *Int. J. Res. Analytical Reviews*, vol. 5, no. 3, pp. 1343-1348, 2018.
- [2] K. Khan, W. Albattah, R. U. Khan, A. M. Qamar and D. Nayab, "Advances and trends in real time visual crowd analysis", *Sensors*, vol. 20, no. 18, pp. 1-12, 2020.
- [3] A. Rani, "The soft crowd management-Special reference to Kumbh-Haridwar", *IOSR J. Computer Eng.*, vol.19, no. 1, pp 99-102, 2017.
- [4] C. Martella, J. Li, C. Conrado and A. Vermeeren, "On current crowd management practices and the need for increased situation awareness, predication, and intervention", *Safety Sci.*, vol. 91, pp. 381-393, 2017.
- [5] M. Yamin and Y. Ades, "Crowd management with RFID and Wireless Technologies" in *First Int. Conf. on Networks & Communications*, pp. 439-442, 2009.
- [6] Wafaa M. Shalash, Aliaa Al. Hazimi, Basme Al Zahrani, "A mobile based crowd management system", *Int. J. Adv. Res. Comput. Commu. Eng.*, vol. 6, pp. 205-215, 2017.
- [7] A. Khan, J. A. Shah, K. Kadir, W. Albattah and F. Khan, "Crowd monitoring and Localization using Deep Convolutional neural network: the review", *Appl. Sci.*, vol. 10, no. 14, pp. 1-17, 2020.
- [8] M. D. Chaudhari and A. S. Ghotkar, "A study on crowd detection and density analysis for safety control", *Int. J. Comput. Sci. Eng.*, vol. 6, pp. 424-428, 2018.
- [9] A. Mohammed, A. Shaery and M. O. Khozium, "Crowd management challenges: Tackling Approach for real time crowd monitoring", *Int. J. Sci. Eng. Res.*, vol. 7, no. 1, pp. 84-88, 2019.
- [10] S. Lamba and N. Nain, "Crowd monitoring and classification: A survey", in *Advances in Computer and Computational Sciences*, S. K. Bhatia et al. Eds. Springer Nature: Singapore, 2017, pp 21-31.
- [11] D. Yimin, C. Fudong, L. Jinping and C. Wei, "Abnormal Behavior Detection Based on Optical Flow Trajectory of Human Joint Points", in *Chinese Control and Decision Conference*, 2019, pp. 653-658.
- [12] T. Wang, M. Qiao, A. Zhu, G. Shan and H. Snoussi, "Abnormal event detection via the analysis of multi-frame optical flow Information", *Front. Comput. Sci.*, vol. 14, pp. 304-313, 2020.
- [13] M. Addelee, R. Curwen, S. Hodges, J. Newman, P. Steggle, A. Ward and A. Hopper, "Implementing a sentient computing system", in *Computer*, vol. 34, no. 8, pp. 50-56, 2001.
- [14] I. Bokun and K. Zielinski, "Active Badges—the Next Generation Industrial Enterprises", 2009.
- [15] S. Granneman, "RFID Chips Are Here", Available http://www.theregister.co.uk/2003/06/27/rfid_chips_are_here Want, R., Hopper, A., Falco, V. and Gibbons, J.: 1992, The active badge location system, *ACM Transactions on Information Systems* 10(1).
- [16] B. Zhou, F. Zhang and L. Peng, "Higher-order SVD analysis for crowd density estimation", *Comput. Vision and Image Understanding*, vol. 116, no. 9, pp. 1014-1021, 2012.
- [17] V. A. Sindagi and V. M. Patel, "A Survey of Recent Advances in CNN-based Single Image Crowd Counting and Density Estimation", *Pattern Recognition Letters*, 2017.
- [18] H. Weiming, et al. "A survey on visual surveillance of object motion and behaviors", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 3, pp. 334-352, 2004.
- [19] B. Tapas, et al. "An Adaptive Codebook Model for Change Detection with Dynamic Background", in *11th Int. Conf. on Signal-Image Technology & Internet-Based Systems*, 2015, pp. 110-116.
- [20] C. Wang, H. Zhang, L. Yang, S. Liu and X. Cao, "Deep People Counting in Extremely Dense Crowds", in *23rd ACM Int. Conf. on Multimedia*, 2015, pp. 1299-1302.
- [21] S. F. Lin, J. Y. Chen and H. X. Chao, "Estimation of number of people in crowded scenes using perspective transformation", *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 31, no. 6, pp. 645-654, 2001.
- [22] A. B. Chan, Z. S. J. Liang, N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking", in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2008, pp. 1-7.
- [23] K. Chen, S. Gong, T. Xiang and C. C. Loy, "Cumulative attribute space for age and crowd density estimation", in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2013, pp. 2467-2474.
- [24] Y. Zhang, D. Zhou, S. Chen, S. Gao and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network", in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2016, pp. 589-597.
- [25] A. Pandey, M. Pandey, N. Singh and A. Trivedi, "KUMBH MELA: A case study for dense crowd counting and modeling", *Multimed. Tools Appl.*, vol. 79, pp. 17837-17858, 2020.
- [26] Q. Wang, J. Gao, W. Lin and X. Li, "NWPU-crowd: A large-scale benchmark for crowd counting", *IEEE Transac. Pattern Analy. Machine Intell.*, 2020.
- [27] C. Zhang, K. Kang, H. Li, X. Wang, R. Xie and X. Yang, "Data-driven crowd understanding: A baseline for a largescale crowd dataset", *IEEE Transac. Multimed.*, vol. 8, no. 6, pp. 1048-1061, 2016.
- [28] C. Zhang, H. Li, X. Wang and X. Yang, "Cross-scene crowd counting via deep convolutional neural networks", in *IEEE Conf. Comput. Vision and Pattern Recog.*, 2015, pp. 833-84.