

# Crowd Abnormality Detection Using Optical Flow and GLCM-Based Texture Features

Ruchika Lalit, Guru Gobind Singh Indraprastha University, India

 <https://orcid.org/0000-0002-9171-5014>

Ravindra Kumar Purwar, Guru Gobind Singh Indraprastha University, India

 <https://orcid.org/0000-0002-2207-2684>

## ABSTRACT

Detection of abnormal crowd behavior is one of the important tasks in real-time video surveillance systems for public safety in public places such as subway, shopping malls, sport complexes, and various other public gatherings. Due to high density crowded scenes, the detection of crowd behavior becomes a tedious task. Hence, crowd behavior analysis becomes a hot topic of research and requires an approach with higher rate of detection. In this work, the focus is on the crowd management and presenting an end-to-end model for crowd behavior analysis. A feature extraction-based model using contrast, entropy, homogeneity, and uniformity features to determine the threshold on normal and abnormal activity has been proposed in this paper. The crowd behavior analysis is measured in terms of receiver operating characteristic curve (ROC) and area under curve (AUC) for UMN dataset for the proposed model and compared with other crowd analysis methods in literature to prove its worthiness. YouTube video sequences are also used for anomaly detection.

## KEYWORDS

Activity Map, Crowd Analysis, Crowd Anomaly Detection, Feature Parameters, GLCM, Optical Flow

## INTRODUCTION

Human stampedes are considered as one of the most feared crowd disasters. These stampedes occur in highly crowded places such as religious pilgrimages, musical events, professional sporting games and many more. Further, the stampedes can occur due to abnormal events such as fire or explosions. Recently, a study (Rodrigues, 2016) reported the mortality in mass gatherings which shows that between 1980 and 2012, around 350 human stampedes have been reported, causing 10,000 death and 22,000 injuries, approximately.

During the last decade, several human stampedes have occurred causing deaths and injuries to agglomerated humans. Recently, automatic crowd analysis has gained attraction from the research community to reduce crowd disasters and mishaps (Kang, Ma & Chan, 2018). Crowd behavior analysis include applications such as detection of escape behavior, panic events in crowded scenes due to natural disaster, chaotic acts, violent activities, traffic management, and surveillance (Kang et al., 2018). In this context of crowd analysis, computer vision-based system (Yogameena & Nagananthini, 2017) has widely been adopted where crowded scene images and videos are analyzed automatically. Recently, pedestrian tracking (Shen, Sui, Pan, & Tao, 2016) and crowd behavior analysis (Marsden, McGuinness, Little, & O'Connor, 2016) have been introduced to improve the surveillance systems.

DOI: 10.4018/JITR.2022010110

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited

In past individual activity analysis has widely been studied and currently, several promising solutions using deep learning concepts are available in literature. Some of these methods for activity recognition are (Wei, Jafari, & Kehtarnavaz, 2019), multi-stream CNN (Tu et al., 2018), deep convolutional networks (Kamel et al., 2018) and many more (Kong & Fu 2018; Qiu, Sun, Guo, Wang, & Zhang, 2019). Despite of significant techniques of individual activity analysis, crowd activity analysis has always been a challenging act, because in crowd scenes, many peoples are located at different positions and move in different directions. Whenever, any abnormal activity triggers, the people move abruptly in all possible directions which creates complexity to analyze the movement. The appropriate movement of crowd will help in detecting the possible anomaly that can alert the security system to take the suitable decision to handle it. Generally, the anomaly detection is categorized as local and global anomaly. According to local anomaly, the behavior of individual differs from the other individual present at the crowded scene. On the other hand, the global anomaly considers the behavior of group in the crowded scene. Figure 1 shows the representation of local and global anomaly detection. In figure 1(a), which represents the local anomaly, movement of one person is studied whereas in figure 1(b) movement of a group of persons is studied for identification of global anomaly.

Figure 1. (a) Local crowd anomaly (b) Global crowd anomaly



Another way of categorizing the crowd behavior is macroscopic and microscopic approach (Zhang, Zhang, Hu, Guo, & Yu, 2018). The macroscopic approaches model the entire crowd as a single entity where each pixel of frame is considered as particles and particle features are extracted and modelled to analyze the crowd behavior such as fluid-dynamics models and network-based models. One such macroscopic approach using swarm particle based optimization has been suggested in (Qasim, & Bhatti, 2019). On the other hand, the microscopic techniques consider crowd as collection of numerous individuals. According to these techniques, each individual is detected and tracked. However, these techniques are suitable to handle the small-scale crowd because in highly dense scenarios, these methods fail to detect and track the individuals due to occlusion and spatio-temporal complexities of video sequences.

The rest of the paper is organized as follows: section II presents the review of existing techniques of crowd behavior analysis, section III presents architecture of proposed model, section IV presents comparative experimental analysis, and finally, section V presents the concluding remarks.

## BACKGROUND

In this section various techniques of crowd abnormal behavior detection are discussed. The global crowd behavior detection techniques are studied here. Various deep learning, CNN based, optical flow based techniques have been used in literature for abnormality detection. Some of the CNN based work are (Ravanbakhsh, Nabi, Mousavi, Sangineto, & Sebe, 2018) developed CNN based plug and play model for abnormal crowd behavior analysis. This model adopts the existing CNN and combined with low-level optical flow. According to this scheme, the input video frames are analyzed to generate temporal CNN pattern and optical flow. The temporal patterns are processed through the binary fully convolutional net. Later, binary and TCP maps are generated. Similarly, the optical flow maps are generated and combined together to detect the anomaly.

A differential recurrent convolutional neural network (DRCNN) based deep model is proposed in (Zhuang, Yusufu, Ye, & Hua, 2017). Energy level co-occurrence matrix describes energy level particle distribution to check the abnormality in the video. (Karpathy et al., 2014) presented a CNN based fully automatic group activity recognition model. Group activity recognition is one of the main features of crowd anomaly detection as occlusion major constraint in this analysis. To overcome the speed issue of CNN in patch-based method, (Sabokrou, Fayyaz, Fathy, Moayed, & Klette, 2018) presented a fully convolutional neural network with temporal data. Here, pre-trained supervised network is first converted into unsupervised network so that global anomalies are detected efficiently from the video. (Zhou et al., 2016) captured the features both from spatial and temporal dimensions using spatial-temporal Convolutional Neural Networks (CNN). Both motion and appearance information are extracted for each frame in a continuous fashion, that suppress the local noise and gives better detection accuracy.

Deep learning based techniques are also used in literature. A cubic-patch based technique is proposed by (Sabokrou, Fayyaz, Fathy, & Klette, 2017), composed of a cascade of classifiers, using advanced feature-learning approach. The classifier has two stages with a deep 3D auto-encoder at first stage to identify numerous normal cubic patches. Small cubic patches are operated on, to find the shallow features and interest points. A more complex and deeper 3D convolutional neural network (CNN) is designed in next stage. Deep autoencoder and CNN are divided into multiple sub-stages, to perform as cascaded classifiers. Complex normal patches and background patches are detected at deeper layers. (Smeureanu, Ionescu, Popescu, & Alexe, 2017) proposed a pre-trained CNN fed into a one-class Support Vector Machines (SVM) classifier to extract deep features. The model learns normal activity from training data and apply it to detect abnormal event in videos.

Optical flow based techniques include (Zhang, Zhang, Hu, Guo, & Yu, 2018) focused on energy parameter to analyze the crowd behavior and presented energy-level distribution model for crowd behavior analysis. In this approach, the complete crowd is modeled as particle and optical flow-based method is applied to extract the particle velocities. According to this model flow field texture helps to estimate the motion foreground and linear interpolation is applied to estimate the foreground area and to determine the distance from camera. Finally, consistency, entropy and contrast values are analyzed to detect the behavior. Similarly, Qasim, & Bhatti (2019) introduced swarm optimization-based technique. First of all, for each video frame a 2D variance plane is constructed to represents the optical field flow magnitude of local-spatio temporal neighborhood. Further, 2D variance plane is divided into salient and non-salient clusters using modified ant colony optimization (ACO) algorithm. Video frames with higher variance are represented by salient clusters.

In previous work (Ruchika, & Purwar, 2019) the abnormality detection is performed using Local Binary Pattern and k-means labeling based feed-forward neural network. foreground is identified using background subtraction. LBP features are extracted to find the abnormality in video using k-means clustering based feed-forward neural network. Precision, recall, F1-Score, accuracy are the performance parameters to evaluate the proposed model. Lloyd, Rosin, Marshall, & Moore (2017) presented visual texture analysis-based model to analyze the crowd behavior. This method uses

GLCM features: angular moment, contrast, homogeneity, correlation, and dissimilarity. This study shows that during abnormal behavior, the feature uniformity holds less uniform rate when compared with the normal crowd behavior. Feng, Yuan, and Lu (2017) used PCANet to extract appearance and motion features from 3D gradients. A deep Gaussian mixture model (GMM) is constructed with observed normal events. It is a scalable deep generative model with multiple GMM-layers stacked over each other. This architecture is used to detect the normal and abnormal behavior in the videos.

## PROPOSED MODEL FOR CROWD BEHAVIOR ANALYSIS

In this section, a macroscopic model for crowd behavior analysis is proposed which overcomes the issues of existing techniques. In this model, the optical flow features are extracted from video frames to analyze the crowd motion and presented Gray Level Co-occurrence Matrix (GLCM) features based threshold model to analyze the crowd behavior for static cameras for fixed background. The image pixels are considered as particles and Farneback optical flow (Farneback, 2003) feature is computed to obtain the magnitude of generated flow of particles. The average magnitude is computed for each frame to generate the activity map. Further, this magnitude is used to compute contrast, entropy, homogeneity, and uniformity parameters and for each parameter a threshold is identified to check abnormality. Generally, in normal behavior, these parameters show a minor variation, whereas during abnormal event, these parameters show significant change in their values when compared with their thresholds. Figure 2 presents the flowchart of the proposed model which is summarized in figure 3.

Various steps of the proposed model are explained as follows -

### Video to Frame

The input video is converted into a sequence of frames manually. These frames are then fed to the proposed model for preprocessing.

### Pre-Processing

All the frames are preprocessed by first converting to gray scale images to make computation simpler and Gaussian filter, given in eq. (1) is applied for the smoothening of these images.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

where  $x$  and  $y$  represent coordinates and  $\sigma$  is the standard deviation of the Gaussian distribution.

### Farneback Optical Flow Estimation

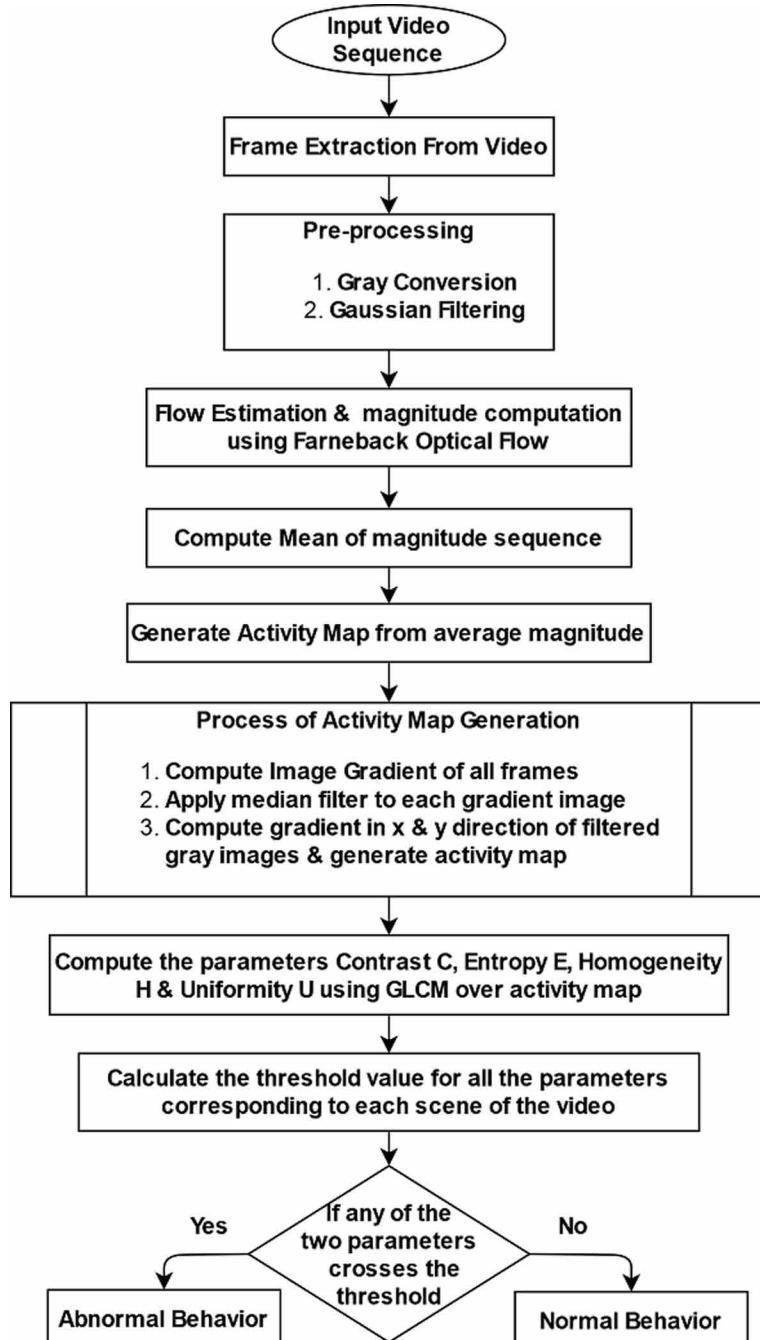
The Farneback Optical Flow (Farneback, 2003), is estimated by representing each frame as quadratic polynomial. The quadratic polynomial can be expressed as:

$$f(x) = xx^T A + xb^T + c \quad (2)$$

Where  $A$  denotes the symmetric matrix,  $x$  denotes the local coordinate,  $b$  denotes the vector and  $c$  denotes the scalar parameter. In a video frame of size  $N \times N$  where the neighborhood of a point  $x = [x, y]^T$  in the frame can be expanded using polynomial basis  $B = \{1, x, y, x^2, y^2, xy\}$  as:

$$E(x, y) = C_1 + C_2x + C_3y + C_4x^2 + C_5y^2 + C_6xy \quad (3)$$

Figure 2. Flow Chart of proposed model



Substituting,  $x = [x, y]^T$  in eq. (2) and comparing it with (3), we get following parameters:

Figure 3. Pseudo Code of the proposed model in MATLAB

<b>Input:</b> Video Sequence
<b>Output:</b> Crowd Behavior Analysis and Abnormal Alert Generation
<b>Step 1 :</b> Read the video data using <i>VideoReader</i>
<b>Step 2 :</b> Obtain the total number of frames <i>F</i>
<b>Step 3 :</b> Initialize the activity detection thresholds for contrast $T_C = 0$ , entropy $T_E = 0$ , and homogeneity $T_H = 0$ uniformity $T_U = 0$ .
<b>Step 4 :</b> Initialize the optical flow using $flow = opticalFlowFarneback$
<b>Step 5 :</b> For $i = 1 : F$ do
<b>Step 6 :</b> Read each frame one-by-one
<b>Step 7 :</b> Apply the following pre-processing on each frame
• Gray conversion
• Gaussian filtering
<b>Step 8 :</b> Estimate the optical flow for the current grayscale frame <i>grayframe</i> , as $Gen_{Flow} = estimateFlow(flow, grayframe)$
<b>Step 9 :</b> Get the magnitude of generated optical flow as $mag = Gen_{Flow}.Magnitude$
<b>Step 10:</b> Compute the mean of obtained magnitude sequence $Magnitude_{mean} = mean(mag)$
<b>Step 11:</b> Generate activity map from the $Magnitude_{mean}$
• Compute image gradient using Robert filter $im_{gradient} = imgradient(Magnitude_{mean}, 'roberts')$
• Apply the median filtering on obtained gradient as $filtered_{region} = medfilt2(im_{gradient})$
• Compute the gradient in x and y direction of median filtered gray image and generate the action map as $Action_{map} = imgradientxy(filtered_{region}, 'centeralldifference')$
<b>Step 12:</b> Consider the $Action_{map}$ to construct the GLCM and compute feature parameters - contrast, entropy, homogeneity, and uniformity.
<b>Step 13:</b> Find the values of threshold $T_C$ , $T_E$ , $T_U$ and $T_H$ for contrast, entropy, uniformity and homogeneity by considering the highest peak point for 1 <sup>st</sup> three parameters while lowest point for homogeneity parameter from graphs shown in figure 4, 5, 6 for all three scenes of the video. These threshold values are chosen from a set of first 400 frames in each video scene.
<b>Step 14:</b> To identify anomaly in a testing frame, its four feature parameters ( $F_C$ , $F_E$ , $F_U$ , $F_H$ ) are computed as discussed above and out of these four values, two values are taken to detect anomaly as follows
<b>Case 1.</b> Homogeneity is taken as one feature (out of two) and if $(F_H < T_H) \&\& (F_C   F_E   F_U > T_C   T_E   T_U)$ i.e. if homogeneity value of testing frame is less than $T_H$ and value of another parameter (it can be Contrast or Entropy or Uniformity) is less than the corresponding threshold value $T_C$ , $T_E$ or $T_U$ , then the testing frame has anomaly.
<b>Case 2.</b> None of the selected feature is homogeneity then their values for testing frame are compared with the threshold values ( $T_C$ , $T_E$ , $T_U$ ) and if both values for testing frame are greater than their corresponding threshold, then testing frame has anomaly
<b>Case 3.</b> If none of the above two cases are satisfies, then testing frame does not have anomaly

$$c = C_1, b = \begin{pmatrix} C_2 \\ C_2 \end{pmatrix} \text{ and } A = \begin{pmatrix} C_4 & C_6 \\ \frac{C_6}{2} & C_5 \end{pmatrix} \quad (4)$$

The  $c$  denotes the coefficients represented as  $c = [C_1, C_2, C_3, C_4, C_5, C_6]$  which can be obtained using normalized convolution with polynomial basis  $B$  using equation (4) as:

$$c = (B * W_a W_c B)^{-1} f(x) W_a W_c \quad (5)$$

The image intensity of neighborhood pixels in consecutive frames of input videos can be approximated with the help of coefficients  $A_1, b_1, c_1$  and  $A_2, b_2, c_2$  (which represent A, b, c values of eq.(2) for consecutive frames) which are used to compute pixel displacement d as:

$$d = -\frac{1}{2} A_1^{-1} (b_2 - b_1) \quad (6)$$

The difference between these consecutive frames provides the optical acceleration which is further used for measuring the magnitude of the flow. These magnitude sequences are averaged and used to generate the activity map.

### Activity Map Generation

Process of activity map generation involves the following steps. Firstly, image gradient magnitude RG is computed over the averaged magnitude frames using Robert gradient operator.

$$RG = \sqrt{RG_x^2 + RG_y^2} \quad (7)$$

where,  $RG_x$  and  $RG_y$  are gradients in x and y direction respectively. These gradient images are further smoothened using median filter. Directional gradients are computed using central difference gradient (CD) (Els, Uys, Snyman, & Thoresson 2006) over these filtered image gradients I to evaluate the activity map.

$$CD = \frac{dI}{dx} = \frac{I(x+1) - I(x-1)}{2} \quad (8)$$

Finally, GLCM is computed from this activity map which is used to compute contrast, entropy, homogeneity, and uniformity as

$$Contrast = \sum_i \sum_j (i - j)^2 G_{ij} \quad (9)$$

$$Entropy = \sum_i \sum_j G_{ij} \log G_{ij} \quad (10)$$

$$Homogeneity = \sum_i \sum_j \frac{G_{ij}}{1 + |i - j|} \quad (11)$$

$$Uniformity = \sum_i \sum_j G_{ij}^2 \quad (12)$$

Where,  $G_{ij}$  represents  $(i,j)^{th}$  element of Gray-Level Co-occurrence Matrix.

### Abnormality Detection

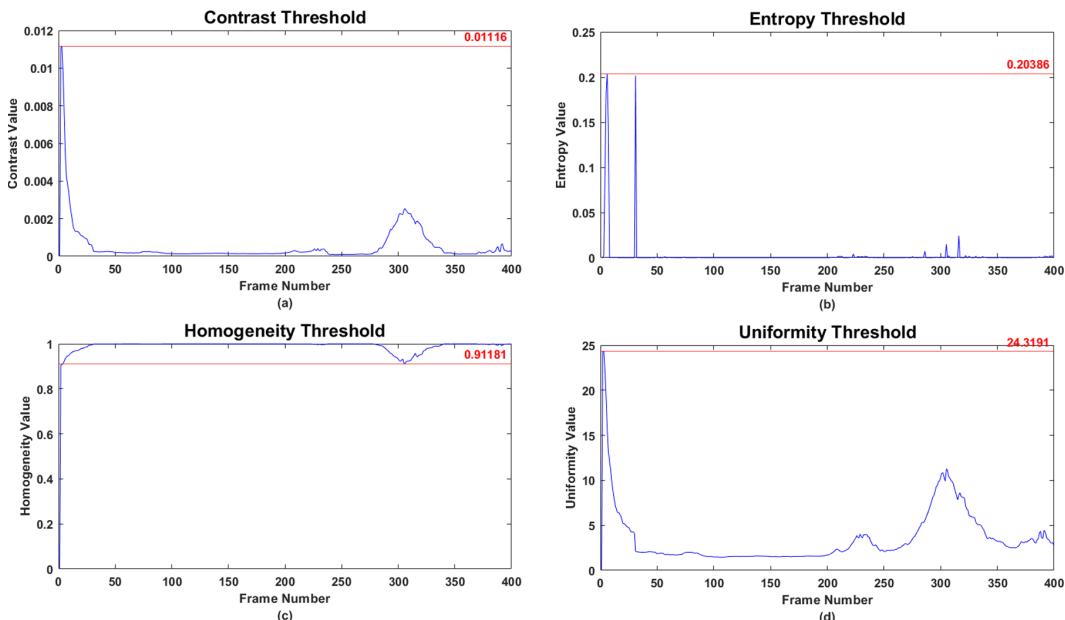
Based on these four parameters, the abnormal behavior is detected. Criterion to determine anomaly is based on the fact that if any of the above two parameters have their values greater than their corresponding thresholds for a particular frame, the frame is identified to have abnormal activity and hence the video. Further, if one or none of the parameter values are more than their corresponding thresholds for all frames of video then video is considered as normal.

### Threshold Computation

Four threshold values are computed, one for each feature parameter in all video scenes separately. For each video scene first 400 frames are taken to compute the threshold. Threshold values are different for each video sequence. The threshold is considered as the first peak of the graph drawn based on each parameter for all the scenes of video. Figure 4-6 shows the threshold values of all the four parameters: contrast, entropy, homogeneity, and uniformity for the three scenes Indoor, Outdoor and Plaza respectively.

Table 1 show the threshold values of four feature parameters of all the scenes of video.

**Figure 4. Indoor Scene (a) Contrast Threshold (b) Entropy Threshold (c) Homogeneity Threshold (d) Uniformity Threshold**

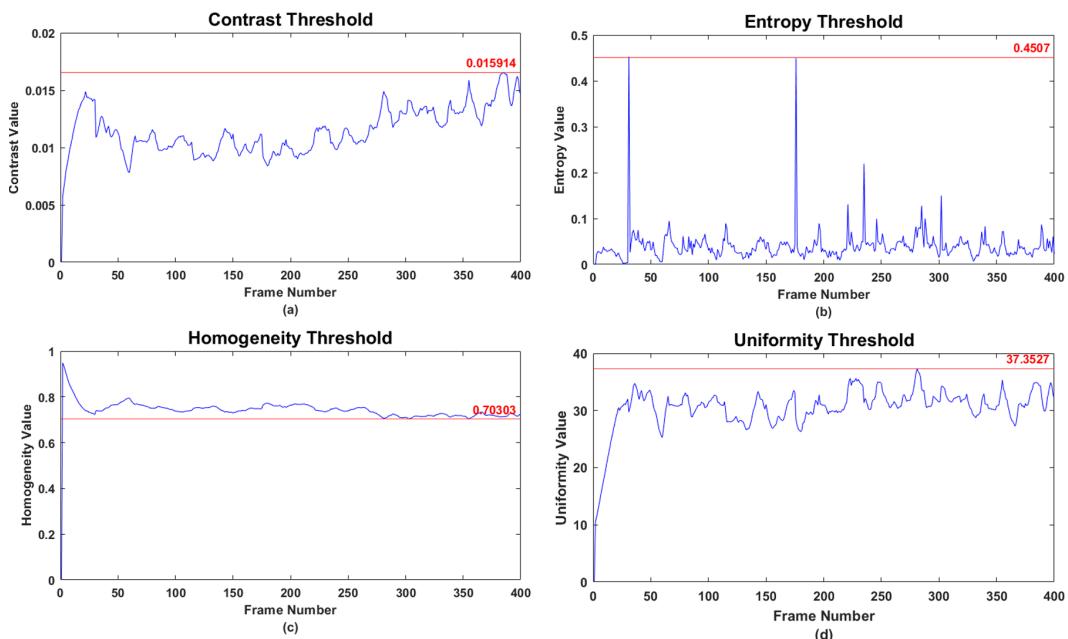


## RESULTS AND DISCUSSION

In this section, experimental results and analysis is done for the proposed model. Simulation has been carried out using MATLAB 19A on an i7 processing unit with 8GB RAM, on window 10 operating system. A universal dataset known as UMN dataset, has been used for it and the performance of the proposed model is compared with six other parallel methods of anomaly detection in video in terms of ROC and AUC and it can be seen that the proposed model performs better than others.

**Table 1.** Threshold values of 4 parameters for all video scenes

	<b>Contrast</b>	<b>Entropy</b>	<b>Homogeneity</b>	<b>Uniformity</b>
<i>Lawn Scene</i>	0.015914	0.4507	0.70303	37.3527
<i>Indoor Scene</i>	0.011116	0.20386	0.91181	24.3191
<i>Plaza Scene</i>	0.01657	1.28051	0.61535	40.9742

**Figure 5.** Lawn Scene (a) Contrast Threshold (b) Entropy Threshold (c) Homogeneity Threshold (d) Uniformity Threshold

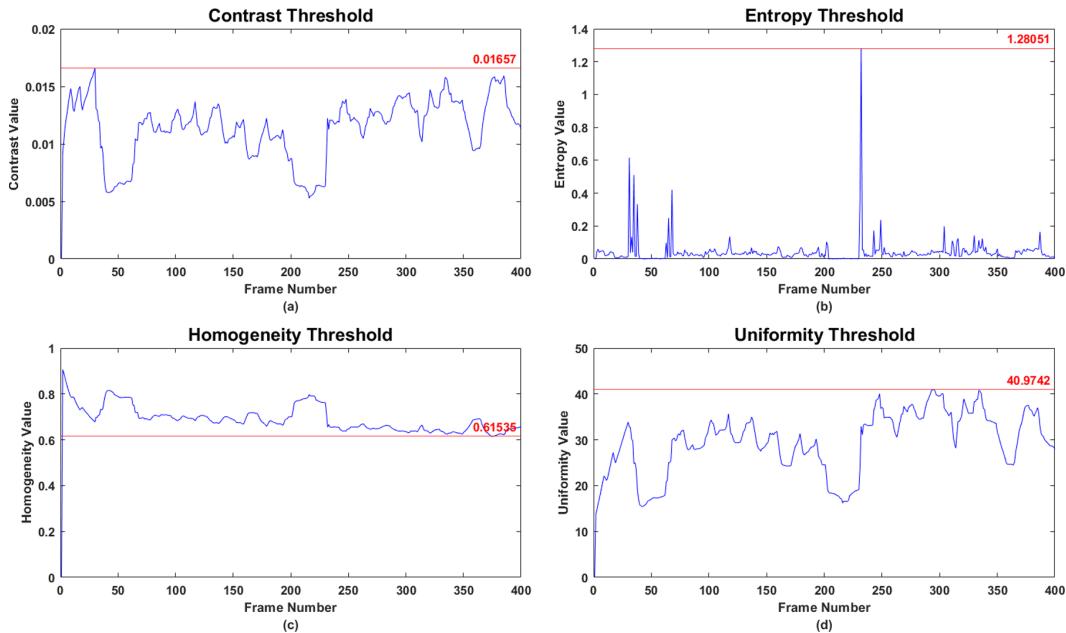
## DATASET USED

*UMN dataset:* It is a publicly available University of Minnesota dataset (“Unusual Crowd Activity Dataset of University of Minnesota”) used by research community. This dataset contains total 11 videos of three different scenes with escape events in indoor and outdoor scenes. Figure 7 shows the sample frames of these sequences of outdoor and indoor scenes.

## Performance Measurement Details

Three different video scenes have been used for experimental results. Four feature parameters are extracted to compute the threshold for each scene to detect the anomaly. Figure 8, shows the anomaly detection for all three scenes. In figure 8(a), value of ground truth for lawn video scene is 524 frame number whereas as per proposed model, anomaly is identified in frame no 521, a difference of 3 frames only. Similarly, this difference between ground truth and predicted frame number for other two scenes is not more than 4. It shows the closeness of anomaly detection for the proposed model. Further, two YouTube video sequences have also been explored to identify possible anomaly as shown in figure 9 and the same has been predicted within a frame difference of 3 in both scenes.

Further, the performance of the proposed model is evaluated in terms of receiver operating characteristic curve (ROC) and compared with three other existing techniques mentioned in (Wang,

**Figure 6. Plaza Scene (a) Contrast Threshold (b) Entropy Threshold (c) Homogeneity Threshold (d) Uniformity Threshold**

& Xu, 2016) as shown in figure 10. It can be seen from the figure that the true positive rate for the proposed model is better than other techniques as rate of false positive increases.

A comparative analysis for UMN dataset in terms of Area under curve (AUC) is presented in table 2. Performance of the proposed model has been compared with other six methods given in (Lloyd et al., 2017).

## CONCLUSION

In this work, a model for crowd anomaly detection for video surveillance is proposed. It performs the abnormal behavior analysis using four thresholds each one for contrast, entropy, homogeneity, and uniformity features which are used to finally detect anomaly in the video scene. Performance is measured in terms of ROC & AUC. The comparative study reported a significant improvement in terms of these parameters than the existing techniques.

## ACKNOWLEDGMENT

Visvesvaraya PhD Scheme of Ministry of Electronics and Information Technology, Government of India, being implemented by Digital India Corporation has supported this research work.

Figure 7. Sample frames of UMN dataset of abnormal crowd behavior

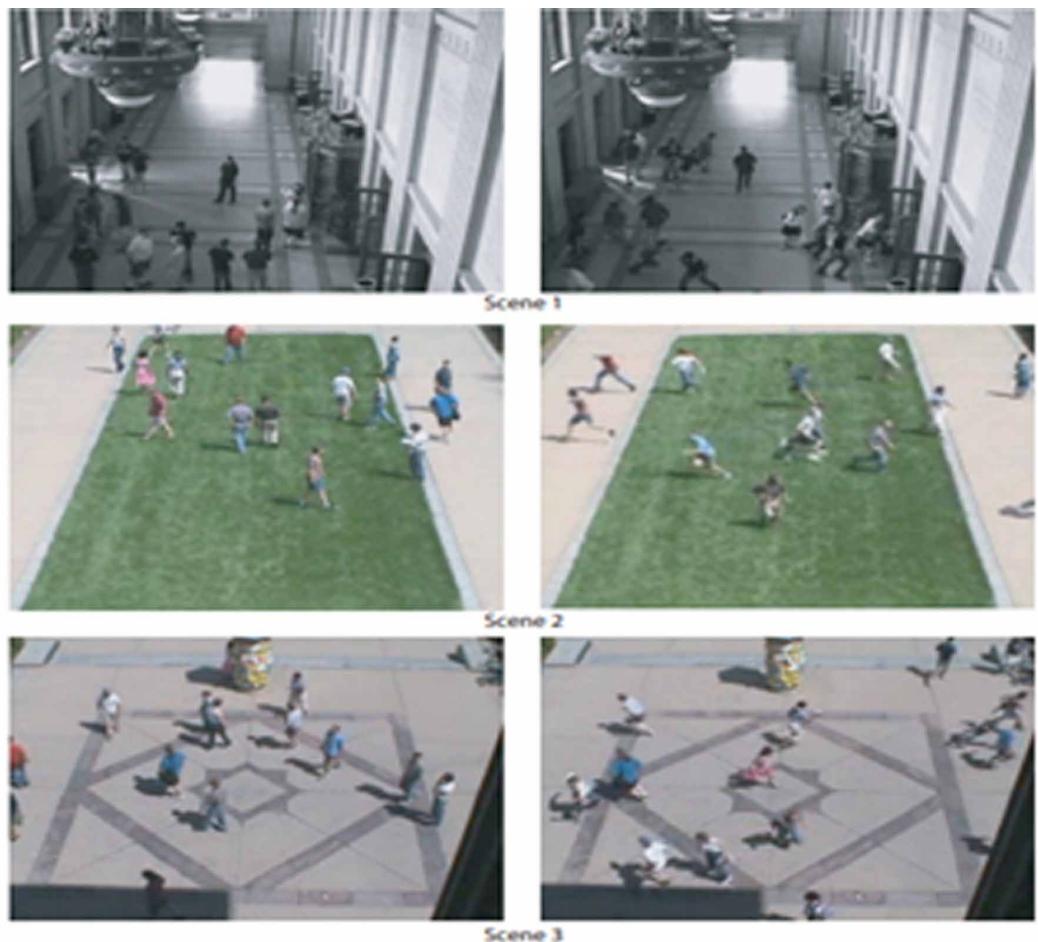


Figure 8. Comparison with ground truth of UMN dataset (a) Lawn Scene (b) Indoor Scene (c) Plaza Scene

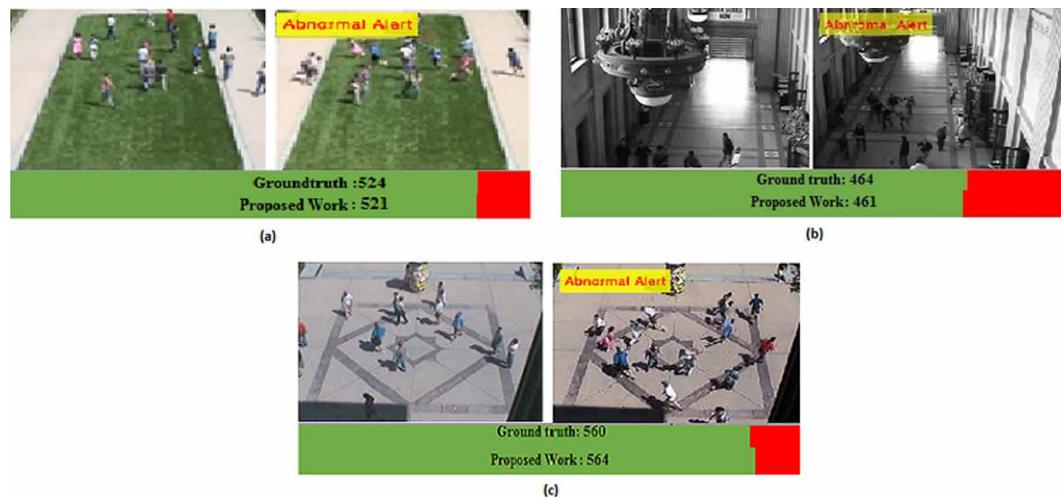


Figure 9. Comparison with ground truth of YouTube sequence (a) Sequence 1 (b) Sequence 2



Figure 10. ROC curve analysis for three different scenes of UMN dataset a) Lawn scene (b) Indoor scene (c) Plaza scene

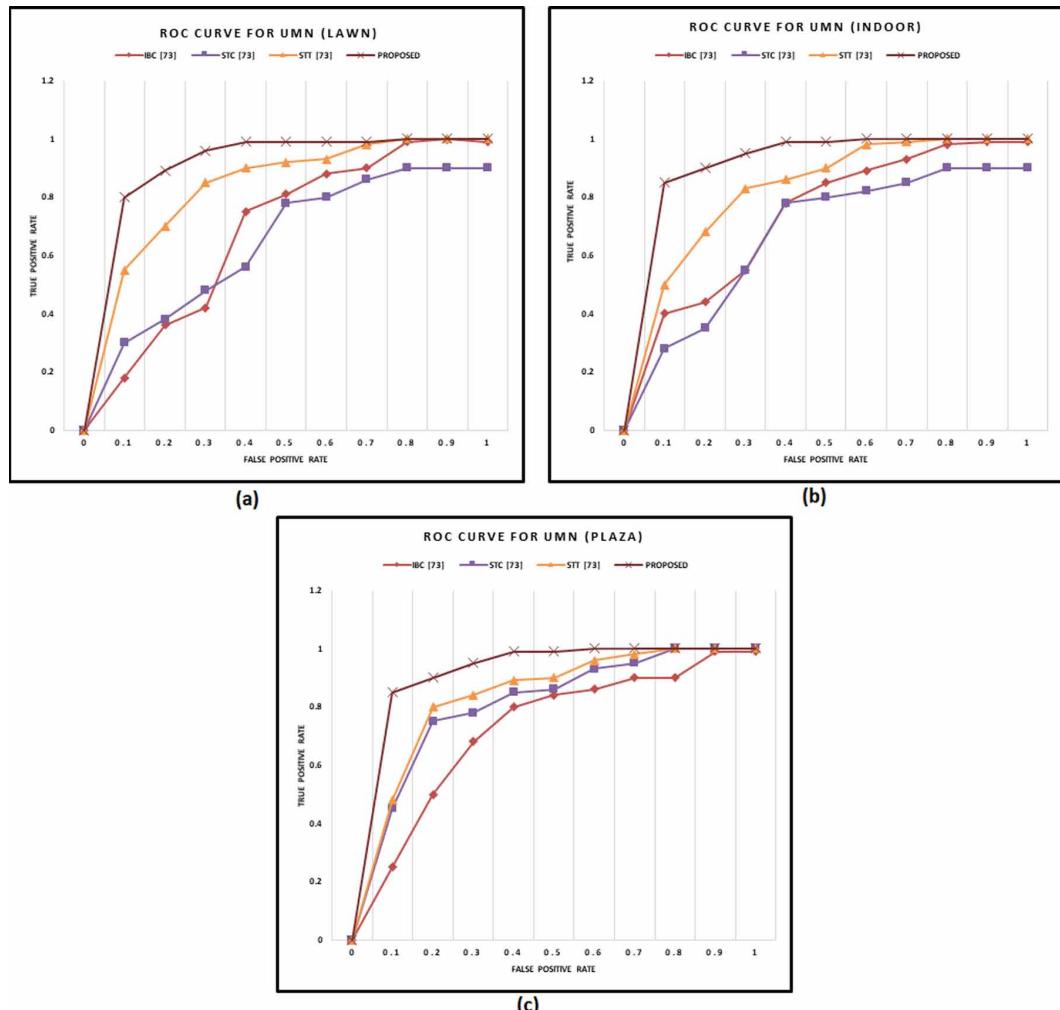


Table 2. Performance comparison in terms of AUC

Technique	Obtained AUC
Optical flow (as cited in Lloyd et al., 2017)	0.84
SF (as cited in Lloyd et al., 2017)	0.96
MDT (as cited in Lloyd et al., 2017)	0.9965
Chaotic invariants (as cited in Lloyd et al., 2017)	0.99
Biswas (as cited in Lloyd et al., 2017)	0.9838
Lloyd (Lloyd et al., 2017)	0.9956
Proposed Model	0.9971

## REFERENCES

- Els, P. S., Uys, P. E., Snyman, J. A., & Thoresson, M. J. (2006). Gradient-based approximation methods applied to the optimal design of vehicle suspension systems using computational models with severe inherent noise. *Mathematical and Computer Modelling*, 43(7-8), 787–801. doi:10.1016/j.mcm.2005.08.018
- Farnebäck, G. (2003, June). Two-frame motion estimation based on polynomial expansion. In *Scandinavian conference on Image analysis* (pp. 363–370). Springer. doi:10.1007/3-540-45103-X\_50
- Feng, Y., Yuan, Y., & Lu, X. (2017). Learning deep event models for crowd anomaly detection. *Neurocomputing*, 219, 548–556. doi:10.1016/j.neucom.2016.09.063
- Kang, D., Ma, Z., & Chan, A. B. (2018). Beyond Counting: Comparisons of Density Maps for Crowd Analysis Tasks—Counting, Detection, and Tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(5), 1408–1422. doi:10.1109/TCSVT.2018.2837153
- Kamel, A., Sheng, B., Yang, P., Li, P., Shen, R., & Feng, D. D. (2018). Deep convolutional neural networks for human action recognition using depth maps and postures. *IEEE Transactions on Systems, Man, and Cybernetics. Systems*, 49(9), 1806–1819. doi:10.1109/TSMC.2018.2850149
- Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 1725–1732). doi:10.1109/CVPR.2014.223
- Kong, Y., & Fu, Y. (2018). *Human action recognition and prediction: A survey*. arXiv preprint arXiv:1806.11230.
- Lloyd, K., Rosin, P. L., Marshall, D., & Moore, S. C. (2017). Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (GLCM)-based texture measures. *Machine Vision and Applications*, 28(3-4), 361–371. doi:10.1007/s00138-017-0830-x
- Marsden, M., McGuinness, K., Little, S., & O'Connor, N. E. (2016, September). Holistic features for real-time crowd behaviour anomaly detection. In *2016 IEEE International Conference on Image Processing (ICIP)* (pp. 918–922). IEEE. doi:10.1109/ICIP.2016.7532491
- Qasim, T., & Bhatti, N. (2019). A hybrid swarm intelligence based approach for abnormal event detection in crowded environments. *Pattern Recognition Letters*, 128, 220–225. doi:10.1016/j.patrec.2019.09.003
- Qiu, Z., Sun, J., Guo, M., Wang, M., & Zhang, D. (2019, September). Survey on Deep Learning for Human Action Recognition. In *International Conference of Pioneering Computer Scientists, Engineers and Educators* (pp. 3–21). Springer. doi:10.1007/978-981-15-0121-0\_1
- Ravanbakhsh, M., Nabi, M., Mousavi, H., Sanginetto, E., & Sebe, N. (2018, March). Plug-and-play cnn for crowd motion analysis: An application in abnormal event detection. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1689–1698). IEEE. doi:10.1109/WACV.2018.00188
- Rodrigues Leal Moitinho de Almeida, M. (2016). *Human stampedes: A scoping review*. Academic Press.
- Ruchika, P. (2019). Abnormality Detection Using LBP Features and K-Means Labelling based Feed-Forward Neural Network in Video Sequence. *International Journal of Innovative Technology and Exploring Engineering*, 8(9S), 629–633. doi:10.35940/ijitee.I1100.0789S19
- Sabokrou, M., Fayyaz, M., Fathy, M., & Klette, R. (2017). Deep-cascade: Cascading 3d deep neural networks for fast anomaly detection and localization in crowded scenes. *IEEE Transactions on Image Processing*, 26(4), 1992–2004. doi:10.1109/TIP.2017.2670780 PMID:28221995
- Sabokrou, M., Fayyaz, M., Fathy, M., Moayed, Z., & Klette, R. (2018). Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. *Computer Vision and Image Understanding*, 172, 88–97. doi:10.1016/j.cviu.2018.02.006
- Shen, X., Sui, X., Pan, K., & Tao, Y. (2016). Adaptive pedestrian tracking via patch-based features and spatial-temporal similarity measurement. *Pattern Recognition*, 53, 163–173. doi:10.1016/j.patcog.2015.11.017

- Smeureanu, S., Ionescu, R. T., Popescu, M., & Alexe, B. (2017, September). Deep appearance features for abnormal behavior detection in video. In *International Conference on Image Analysis and Processing* (pp. 779–789). Springer. doi:10.1007/978-3-319-68548-9\_70
- Tu, Z., Xie, W., Qin, Q., Poppe, R., Veltkamp, R. C., Li, B., & Yuan, J. (2018). Multi-stream CNN: Learning representations based on human-related regions for action recognition. *Pattern Recognition*, 79, 32–43. doi:10.1016/j.patcog.2018.01.020
- Unusual Crowd Activity Dataset of University of Minnesota. (2020). Retrieved from <http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi>
- Wang, J., & Xu, Z. (2016). Spatio-temporal texture modelling for real-time crowd anomaly detection. *Computer Vision and Image Understanding*, 144, 177–187. doi:10.1016/j.cviu.2015.08.010
- Wei, H., Jafari, R., & Kehtarnavaz, N. (2019). Fusion of Video and Inertial Sensing for Deep Learning-Based Human Action Recognition. *Sensors (Basel)*, 19(17), 3680. doi:10.3390/s19173680 PMID:31450609
- Yogameena, B., & Nagananthini, C. (2017). Computer vision based crowd disaster avoidance system: A survey. *International Journal of Disaster Risk Reduction*, 22, 95–129. doi:10.1016/j.ijdrr.2017.02.021
- Zhang, X., Zhang, Q., Hu, S., Guo, C., & Yu, H. (2018). Energy level-based abnormal crowd behavior detection. *Sensors (Basel)*, 18(2), 423. doi:10.3390/s18020423 PMID:29389863
- Zhou, S., Shen, W., Zeng, D., Fang, M., Wei, Y., & Zhang, Z. (2016). Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes. *Signal Processing Image Communication*, 47, 358–368. doi:10.1016/j.image.2016.06.007
- Zhuang, N., Yusufu, T., Ye, J., & Hua, K. A. (2017, May). Group activity recognition with differential recurrent convolutional neural networks. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 526–531). IEEE. doi:10.1109/FG.2017.70

*Ruchika Lalit has obtained her M.Tech. (Information Technology) degree from University School of Information Technology, Guru Gobind Singh Indraprastha University (GGSIPU), Delhi. She is pursuing her Ph.D. from University School of Information, Communication and Technology, GGSIP University, Delhi. Her research areas are Image/Video processing, computer vision.*

*Ravindra Kumar Purwar has obtained his M.E. (Computer Science & Engineering) degree from MNREC Allahabad (currently known as MNNIT Allahabad). He has pursued his doctorate from university school of information technology, GGSIP University, Delhi. He is life member of Computer Society of India (CSI) and Indian Society of Technical Education (ISTE). He has various publications in peer reviewed quality international journals and conferences. Image/Video processing, pattern recognition, video security and database management are his current area of research.*