

## Crowd behavior analysis using MoDTA approach

Savitha. C<sup>1</sup>, Dr. Ramesh. D<sup>2</sup>

<sup>1</sup>Department of Electronics and Communication, Sri Siddhartha Institute of Technology, India

<sup>2</sup>Department of Computer Science and Engineering, Sri Siddhartha Institute of Technology, India

---

### Article Info

#### Article history:

Received Sep 20, 2018

Revised Feb 19, 2019

Accepted Mar 3, 2019

---

#### Keywords:

Detection

Modta

Template detector

Tracking

---

### ABSTRACT

In order to analyze the behaviors of human, significant extent of work has been carried out in the video surveillance applications. While considering the crowded scenes, the adopted features are crafted manually which have a great side to detect anomaly. It requires prior information and is hard to extract from complex video scenes and also it involves huge computational costs. In this paper, we are proposing multi-observational detection and tracking approach (MoDTA) that is based on observational filter. The MoDTA initially acquires people location in an image, so that it can detect conviction value at pointed locations which generally increases with respect to people density. In the phase of tracking, MoDTA computes the multiple observed weight values and individual features, also advection particle is used at motion model in order to facilitate the dense scenario tracking. Coefficient of correlation is used as template detector and the function of template detector is to estimate the upcoming object. Our proposed MoDTA is compared with other existing detection and tracking methods in order to evaluate the system performance.

Copyright © 2019 Institute of Advanced Engineering and Science.  
All rights reserved.

---

### Corresponding Author:

Savitha. C,

Department of Electronics and Communication,

Sri Siddhartha Institute of Technology,

Tumakuru, India.

Email: savithagkc@gmail.com

---

## 1. INTRODUCTION

It is identified that the abnormal behavior recognition and group behavior analysis are the major difficulties in the video surveillance system, and to address this problem several researchers have worked on it. Lots of efforts have been put in for anomaly detection [1] and it shows the importance of it. Therefore, subtopic like feature representation is considered to be highly important for detailed description. Feature representation forms an indispensable basis which is extremely correlated with detection approach. Since the analysis of crowd scene requires fundamental models of crowd, initially the crowd knowledge acquisition from crowd dynamics should be summarized before the crowd detection model. Although a variety of representation models and approaches have been proposed, yet there is no any accepted general solution for analysis of crowded scene. Here we consider to detect the anomalous behavior in videos.

The explicit event based approach uses supervised model [2] where the abnormality of behavior can be learnt by using training set. But the problem with this type of approach is that the detection of abnormality generally depends upon training dataset (i.e., previously collected dataset). An approach based on detecting a specific abnormality manually is developed for particular applications such as detecting threats for the 'cargo' video observation system [3]. Unsupervised method (i.e. not dependent on any prior knowledge and training) can identify abnormal behaviors. However this category generally detects only the simple abnormal scenarios such as bicycles and cars among pedestrians. In this approach they generally analyze the acceleration and optical flow that is very different among these objects [4].

A generalized trajectory based approach first sets the scene of interest for crowd and then splits it into different objects. Afterwards the objects are monitored through their behavior in video sequences which forms the extracted feature trajectories [5]. Zone based approach, deep learning based model [6], temporal spatial path search [7], string kernels clustering [8], single class-SVM (Support Vector Machine) [9] are applied to compute anomalies in trajectories. The type of methodology depends on the nature of object and the kind of people who are tracked. The performance of tracking is affected by using rapid motion and low video resolution. The aim of using global pattern based approach is not only to track and detect individually in a scene, but try the methodology to get low/medium features for video scene and to analyze the feature for whole entity [10]. In general, these features can be used for optical flow and spatial temporal gradients, moreover there are some approaches which are quite effective while dealing with group activity such as motion influence map [11], stationary map [12], Gaussian regression [13], global motion map [14], salient motion map [15], energy model [16], PCA-model [17], Gaussian mixture model [18], Social force model [19] etc.

Tracking and detection of crowd abnormality is a very challenging activity due to the randomly changing crowd dynamics. But prediction trackers are not modelled to handle differentiating individual scenarios. In this paper, we are proposing multi-observational detection and tracking approach (MoDTA) that is based on observational filter. The MoDTA initially acquires the people location in an image and detects conviction value at pointed locations, which generally increases with respect to people density. In the phase of tracking, MoDTA computes the multiple observed weight values and individual features are added with scenarios. Also the advection particle is used as motion model in order to facilitate the dense scenario tracking. The function of template detector is to compute the estimation of the upcoming object. Coefficient of correlation is used as a template detector. Here we use UMN dataset for detecting the abnormality in a given scene, moreover our proposed MoDTA is compared with other existing detection and tracking methods.

## 2. LITERATURE SURVEY

In some recent years the auto scene analysis and understanding has attracted lots of research attention in the community of computer vision [20] and the major application can be seen in intelligent surveillance which has replaced the traditional way of video surveillance. Though several approaches have been developed for recognizing, tracking and understanding of various objects behavior in video scene, they are mainly modelled for commonly used scenes at low population density. But whenever crowded scenes are analyzed the difficulty arises due to the large number of objects involvement which not only causes the failure in detection but also in tracking. Thereby increases the computational complexity. The group scenarios [21] are major entities which comes from the crowd. Therefore understanding the properties of group level is very important.

In paper [22], they aim at identifying the modelling problem to differentiate the captured video of the surveillance application into normal and abnormal. Moreover, a new framework is generated in order provide an anomaly detection and automatic behavior profiling, which depends on group clustering. The behavior analysis can be efficiently applied at crowded scenes with random distribution and at crowd density which is potentially necessary in several applications such as abnormal event detection, crowd video classification and crowd dynamic monitoring in the security surveillance. The event based abnormality detection is the procedure to obtain the abnormal scenario which is compared with bulk of usual events and the major challenge present in this is the dynamic variations in scenes and large structured redundancy in videos surveillance. In order to overcome these problems in [23], they have proposed a framework for detection of abnormality and localization in video scenes that depend on the constrained locality-constrained affine subspace coding (LASC) and updating approach. In this study, they have used LASC in order to reconstruct the sample model through its nearest top-k subspaces which is originated through segmenting the usual sample space by clustering approach. A sample of large reconstruction cost is detected by setting up a threshold value. In this paper [24], they have introduced a model and an algorithm with a case study to validate them and to optimize the pattern recognition, event detection and abnormality identification under a real-life video surveillance.

The work includes observing the human nature patterns in an over-all continuously changing nature and adapt with the time, rather being static. In existing works there are some limitations which are identified and accordingly the dynamic clustering algorithm is used in [25] to overcome the drawbacks. Congruently, they have proposed maintaining the concept of two different data-sets in parallel such as abnormal plane and normal plane in order to acquire successfully the task of learning. Moreover, the practicable scenario of the model has been demonstrated by considering real-life cases. Computing the abnormality for crowded scenes is becoming difficult and critical in internet services and cloud environment due to end-to-end user

experience and quality of service. However, the vast changes in behavior of metric streams has increased the challenges, which may diagnose the frameworks based on threshold with normal or stationary assumption. Generally complex models demand extensive offline training. These approaches are inclined to unauthentic false-alarms in the online settings, therefore the metric streams experience quick contextual changes via known reference point. Initially, they predict a property of underlying temporal stream through adaptive learning and afterwards they apply robust statistical control charts in order to identify deviations [26].

To handle the obstruction and complex scenes several approaches uses extracted features from motion cue and low-level appearance. Optical flow and texture flow are used for object tracking. Several frequently used features include histogram oriented gradients, histogram based optical flow and 3D 'spatial-temporal-features'. Spatial-temporal-features are used to provide Markov field model. Normal behavior statistics detects the crowd abnormal behavior. In paper [27] they have employed multi-scale feature of histogram based optical flow that incorporates sparse representation via detection through reconstruction cost.

In paper [28], they have proposed a novel approach that uses Markov random field and local-optical-flow to detect the unusual events. In paper [29] they have developed a methodology to determine the unusual behavior by using statistical aggregates. An enhanced model for determining the unusual behavior through texture, size and foreground is proposed in [30]. To describe the size of spatial-temporal the sparse behavior is represented to identify abnormal activity. A method is developed for unusual behavior detection that is based on Markov random field (MRF) and they have employed optical flow acceleration and histogram optical flow as the behavioral feature [31]. In paper [32] they have used integrated model from motion cues and appearance. In this paper [33] a novel descriptor has been proposed for detection of abnormality through enhancing the acceleration concept by using hybrid optical flow. However, these adopted features are crafted manually that have a great side to detect anomaly. Though it requires prior information, it is hard to extract from complex video scenes and also involves huge computational costs.

### 3. MULTI-OBSERVATIONAL DETECTION AND TRACKING APPROACH (MODTA)

#### 3.1. Density Informed Energy Formulation

Here, we assume the conviction value  $c(b)$  to detect a person  $B$  at location  $B_i$ . In an image, location number is given as  $i = 1 \dots, A$ . In order to compute the person density, initially the people number at per pixel is estimated. The estimated  $E(B_i)$  is in a window having size  $d$  at the location  $B_i$ . The major aim is to get the people location in an image so that it detects conviction value at pointed locations. Conviction value given by  $E$  increases with respect to people density, and also it prevents from appropriate overlapping detections. The detection process is encoded in the image by using a  $A$ -vector which is given as  $x \in \{0,1\}^A$ . When  $x_i$  is considered to be one then  $B_i$  detection is switched on otherwise it is zero and the difficulty in detection can be formulated to minimalize. Decreasing density error ( $F_{ED}$ ) implies minimalizing the differences in obtained density estimations using  $E(b)$  estimator. The minimization of cost function can be written as;

$$\min_{[x \in \{0,1\}^A]} \left( \frac{\alpha \|E - Gx\|_2^2}{F_{ED}} + \frac{x^H K x}{F_P} - \frac{C^H x}{F_S} \right) \quad (1)$$

From (1),  $F_S$  shows the high certitude values at locations to detect people from person detector and is indicated via  $x_i = 1$ . In (1),  $F_P$  shows the valid configuration for selecting the detection of non-overlapping and this can be obtained via setting  $K_{ij} = \infty$ , which happens when the location detected at  $b_i$  and  $b_j$  have optimized area overlap ratio, otherwise it is considered to be zero.

Here, the  $F_{ED}$  term used in the (1) is to model the constraints of crowd density by correcting the gap between the values of density, first one computed with density based regression estimator  $E$  and second one computed through detection of switch (i.e., on or off) in  $x$ . In order to evaluate the active detection density in  $x$ , the matrix multiplication is performed, where  $G$  represents the  $A \times A$  matrix with  $G_i$  rows;

$$G_i(l_j) = \frac{1}{2\pi d^2} \exp\left(-\frac{\|b_i - l_j\|^2}{2d^2}\right) \quad (2)$$

Where, the size of Gaussian window ( $d$ ) is centered at  $b_i$  position. Optimizing the cost value  $F_{ED}$  in (1) enables to improve the person detection by correcting certitude value in lower density of image regions.

In a simpler way we can say that  $d$  is same through whole image width, the term  $F_{ED}$  can decrease up-to a multiplicative constant in order to correct gaps between the switches at detection in  $x$ , and the  $D$  density estimator is used to get entire number of people from considered image.

Now we propose a multiscale detector for individual person detection, which includes the score map of dense detection  $c(b)$  which is used in cost equation. The function  $c(b)$  is integrated with the density computation  $E(b)$ . Also it requires geometrics computation for crowd scenes footage that generally involves several moving people on a filmed plane and the plausible positions of peoples.

### 3.2. Detecting and Tracking Approach

In this section, we are providing detection and tracking model, where the multi-observational process is performed and consists of in-scenario ( $z_t^{s-in}$ ), out-scenario ( $z_t^{s-out}$ ) observations with a similarity measurement ( $z_t$ ), which helps to compute the person activity in split and merge case. The major aim is to consider the multi-observational process that can decompose the typical observational model of filter in-to split observation. Therefore the interaction and similarities with individual persons are observed and can be written as;

$$B(z_t|x_t) = b(z_t^i, z_t^p|x_t) \tag{3}$$

$$B(z_t|x_t) = b(z_t^i|z_t^p, x_t)b(z_t^p|x_t) \tag{4}$$

Where, the observation of people interaction with different people is  $z_t^i$  and the individual similarity observation is given by  $z_t^p$ . The independency present at the interaction and similarity, can be written as;

$$B(z_t|x_t) = b(z_t^i|x_t)b(z_t^p|x_t) \tag{5}$$

Therefore, the multi-observations can split the observed interactions into in-scenario ( $z_t^{s-in}$ ) and out-scenario ( $z_t^{s-out}$ ) observations, thus the interaction evaluation is performed as in and out sets of scenarios;

$$b(z_t^i|x_t) = b(z_t^{s-in}, z_t^{s-out}|x_t) \tag{6}$$

$$b(z_t^i|x_t) = b(z_t^{s-in}|z_t^{s-out}, x_t)b(z_t^{s-out}, x_t) \tag{7}$$

$$b(z_t^i|x_t) = b(z_t^{s-in}|x_t)b(z_t^{s-out}, x_t) \tag{8}$$

Hence the observed model can be written as;

$$B(z_t|x_t) = b(z_t^{s-in}|x_t)b(z_t^{s-out}, x_t)b(z_t^p|x_t) \tag{9}$$

In-scenario observation is denoted by  $b(z_t^{s-in}|x_t)$  and it measures the belonging degree associated to that scenario, also the out-scenario observation is denoted by  $b(z_t^{s-out}|x_t)$  and it measures the not belonging degree associated to that scenario. Here,  $m_{x,z}^\theta$  denotes the directional similarity and  $m_{x,z}^{sd}$  denotes normalize spatial distance.

$$m_{x,z}^\theta = (1 + \cos \theta)/2 \tag{10}$$

$$m_{x,z}^{sd} = e^{-\left(n/\min_{q \in p}(q_w, q_h)\right)} \tag{11}$$

Where, ( $sd$ ) denotes the distance value between  $x$  and  $z$ ,  $\theta$  shows the angle difference between motion vectors. The set of boxes of individual detection results in first frame denoted by  $s$  and, the height of bounding box is  $q_h$  and width is  $q_w$  in  $p$  set that is used in normalization purpose. Therefore, the interaction weight can be calculated as;

$$m_{x,z}^i = m_{x,z}^\theta m_{x,z}^{sd} \tag{12}$$

Here,  $B^{s-out}$  and  $B^{s-in}$  describe the number of persons in the out-scenario and in-scenario respectively, out-scenario measures  $m_{x,z}^{s-out}$  and in-scenario computes  $m_{x,z}^{s-in}$  in between the individual value of  $x$  and  $z$ . It can be shown as;

$$m_{x,z}^r = 0, \text{ if } m_{x,z}^\theta > m^\theta \\ \text{'or' } m_{x,z}^{sd} > m^{sd} \quad (13)$$

Otherwise it can be given as;

$$m_{x,z}^r = m_{x,z}^\theta \quad (14)$$

Where,  $m^\theta$  and  $m^{sd}$  shows the threshold standards of dictionary similarities and spatial closeness and hence  $r \in \{s - in, s - out\}$ . The value of  $m_{x,z}^{s-in}$  for an individual can be computed with respect to the persons of  $B^{s-in}$  and  $m_{x,z}^{s-out}$  for an individual can be computed with respect to the persons of  $B^{s-out}$ .

Several individuals link with others indirectly for in and out scenarios. For that weights are needed to be filtered in order to provide indirect connection information. That is why we update weights if  $m_{x,z}^r$  becomes zero, but considering individual  $s$  it can be written as;

$$m_{x,s}^r > m^\theta m^{sd} \text{ and } m_{s,y}^r > m^\theta m^{sd} \quad (15)$$

Considering this case, we can set the value as  $m_{x,z}^r = m_{s,z}^r$ . Because  $x$  is linked to  $z$  via  $s$ , after performing refinement operation of individuals-weights the possible event  $u$  is determined by;

$$U(x; z; s) = \begin{cases} P_b, m_{x,z}^{s-in} = 0 \\ V_u, w_{x,z}^{s-out} > 0 \end{cases} \quad (16)$$

If the above condition is not matched then  $U(x; z; s)$  value is none, where  $z \in B^{s-in}$  and  $s$  becomes  $s \in B^{s-out}$ .  $E$  is a function that detects the split activity between  $x$  and  $z$  is  $P_b$  and  $U$  is function that detect the merge activity between  $x$  and  $z$  is denoted by  $V_u$ , moreover the Euclidian distance is used to compute the individual connection  $b(z_t^p | x_t)$ .

Moreover, the motion approach is formed by integrating with state object that is composed of position (i.e.,  $b_x$  and  $b_z$ ) and velocity (such as  $w_x$  and  $w_z$ );

$$x_t = [b_x \ b_z \ w_x \ w_z] \quad (17)$$

$$\pi(X_{t+1}|X_{0:t}) = [1 - \beta] \times \pi_{pA}(X_{t+1}|X_{0:t}, y_{0:t+1}) + \beta \times \pi_{Det}(X_{t+1}|X_t) \quad (18)$$

Where,  $\beta$  determines the relative components weight and it is selected inversely proportional with respect to group density in order to provide higher weighted particle in denser region. Hypothesis part  $\pi_{Det}(X_{t+1}|X_t)$  and  $\pi_{pA}(X_{t+1}|X_t)$  can produce motion vectors by template detector and particle advection, also the higher weight is used at detector template in the sparse regions, so the updated group density is given by;

$$\rho_{sd} = \mu_b / \mu_{op} \quad (19)$$

Where,  $\mu_b$  denotes the number of people present in group scenario,  $\mu_{op}$  shows for outsider person who not intersecting with group scenario and then the  $\beta$  can be written as;

$$\beta = \delta / \rho_{sd} \quad (20)$$

The  $\delta$  denote for the obtained similarity measurement in between the detected template and objects. So the model of linear motion dynamics with the Gaussian noise is given as;

$$x_{t+1}^q = \varphi x_t^q + \mu_r \quad (21)$$

Where,  $\varphi$  denotes the matrix of particle dynamics,  $\mu_r$  is Gaussian noise and  $x_t^q$  is for group state at  $t$  time. Through using velocity vector of every tracked object is engaged for several frames, along with this similarity of existing velocity vector is computed that is based upon the earlier motion of traced object, here we select the object with the vector of consistence velocity, then  $\pi_{PA}(X_{t+1}|X_t)$  is used to compute velocity vector. The function of template detector  $\pi_{Det}(X_{t+1}|X_t)$  is used to compute the  $[w_x w_z]$  estimation for the upcoming object by detection, a coefficient of correlation is used as to be template detector.

**4. RESULT ANALYSIS**

In this section, we are doing result simulation using Matlab-2016b and system configuration Intel i5 Processor, 8GB RAM, 2GB NVIDIA Graphics with windows 10 operating system. Here, we considered very popular data-set for the video surveillance research purpose such as UMN dataset [34], which originally is in avi format. Additionally, it consists of four different data-sets such as courtyard, crowd, corridor and hit-run in the form of video. Each dataset having thirty frames per second. In that we are taking only 2 frames per second, which increases the chances for usability in real-time scenario. In each dataset, there are normal frames and abnormal frames, where normal frames shows the normal activity of crowd and abnormal frames shows the abnormal activity of crowds. To compute the effectiveness of our proposed model here we consider Ground Truth as to be reference.

Figure 1 shows the normal crowd activity at dataset 1 in a courtyard scenario and Figure 2 shows the abnormal crowd activity at dataset 1 in a courtyard scenario. Figure 3 shows the comparison with respect to Ground Truth (GT) at courtyard dataset, where the GT shows the abnormal frame (AF) is started from 37th frame and our MoDTA model detected at 38th frame and so on.



Figure 1. Normal crowd activity at dataset 1 (Courtyard)



Figure 2. Abnormal crowd activity at dataset 1 (Courtyard)

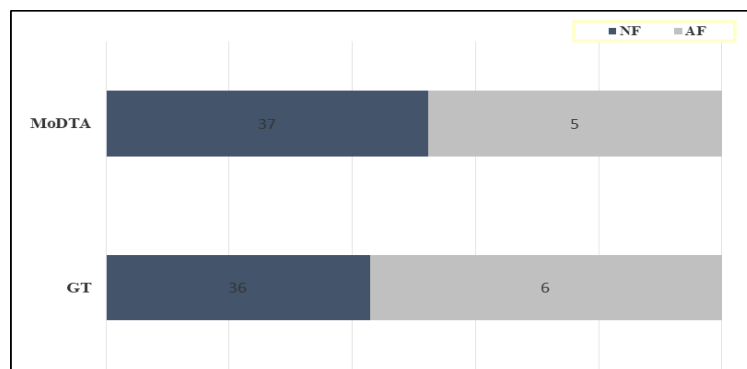


Figure 3. Comparison w.r.t ground truth at courtyard dataset

Figure 4 shows the normal crowd activity at dataset 2 and Figure 5 shows the abnormal crowd activity at dataset 2. Figure 6 shows the comparison with respect to Ground Truth (GT) at crowd dataset,

where the GT shows the abnormal frame (AF) is started from 32th frame and our MoDTA model detects at 33th frame. The normal frame (NF) is detected in GT from 1 to 31 and abnormal frames from 32 to 40. MoDTA detects 1 to 32 as the normal frame and rest of them ie. 33 to 40 as abnormal frame. Figure 7 shows the normal crowd activity at dataset 3 in corridor scenario and Figure 8 shows the abnormal crowd activity at dataset 3 in corridor scenario.



Figure 4. Normal crowd activity at dataset 2 (Crowd)



Figure 5. Abnormal crowd activity at dataset 2 (Crowd)

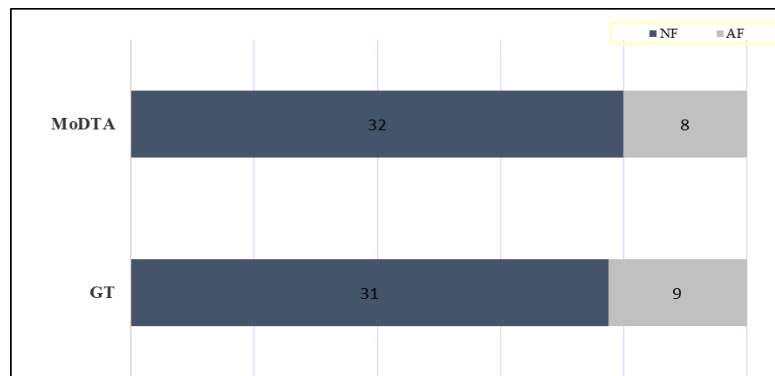


Figure 6. Comparison w.r.t Ground Truth at Crowd Dataset



Figure 7. Normal crowd activity at dataset 3 (Corridor)



Figure 8. Abnormal crowd activity at dataset 3 (Corridor)

Figure 9 shows the comparison with respect to GT at corridor dataset. Here the GT shows the abnormal frame (AF) is started from 21th frame and our MoDTA model detects at 23th frame. The normal frame (NF) is detected in GT from 1 to 20 and abnormal frames from 21 to 34. MoDTA detected 1 to 22 as the normal frame and rest of them 23 to 34 as abnormal frame.

Figure 10 shows the normal crowd activity at dataset 4 in hit-run scenario and Figure 11 shows the abnormal crowd activity at dataset 4 in hit-run scenario. Figure 12 shows the comparison with respect to GT at hit-run dataset, where the GT shows the AF is started from 13th frame and our MoDTA model detects at 15th frame. The normal frame (NF) is detected in GT from 1 to 12 and abnormal frames from 13 to 30. MoDTA detects 1 to 14 as the normal frame and rest of them 15 to 30 as abnormal frame.

ROC is a most common technique to visualize the performance of a classifier. Here, we have used BoF classifier which is a binary classifier because we have only two classes such as normal and abnormal. According to our paper, a crowd scenario can be classified into two categories first one is normal activity and other is abnormal activity. ROC curve is shown in Figures 13, 14, 15 and 16 for different UMN datasets. Figure 17 shows the comparison with respect to different existing techniques for Area Under Curve (AUC %). Our model got 99.62% AUC which compared with existing SFM [35] and we got 4.7% more AUC. Moreover, the proposed model is compared with Chaotic invariants [36], Sparse reconstruction [37], Local statistics [38], MDT [39], and we got 0.22%, 0.02%, 0.12%, and 0.12% more AUC.

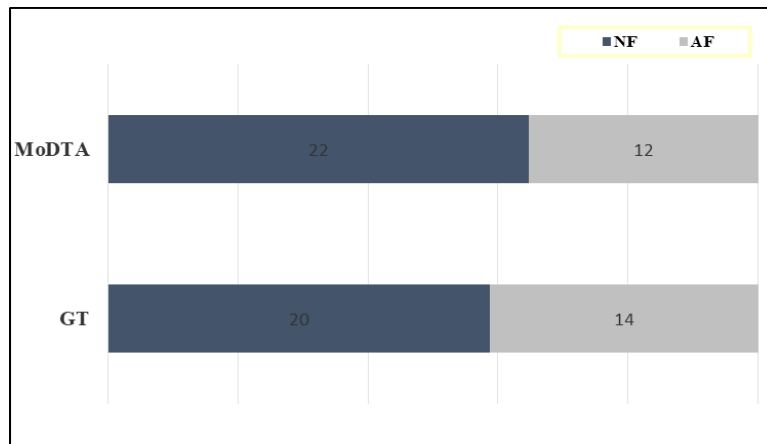


Figure 9. Comparison w.r.t ground truth at corridor dataset



Figure 10. Normal crowd activity at dataset 4 (Hit-run)



Figure 11. Abnormal crowd activity at dataset 4 (Hit-run)



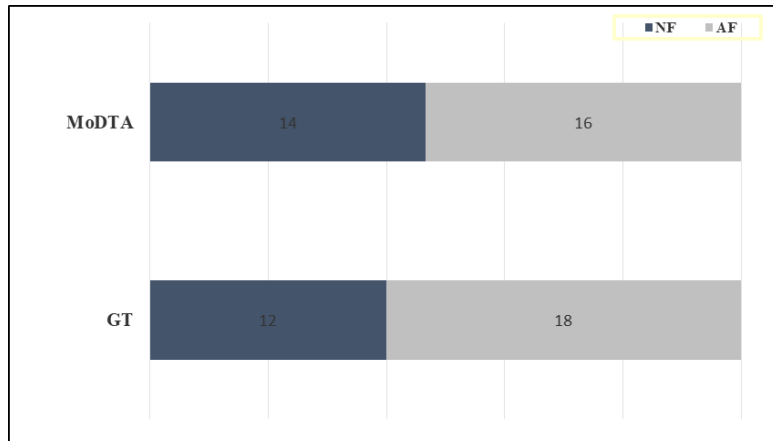


Figure 12. Comparison w.r.t ground truth at hit-run dataset

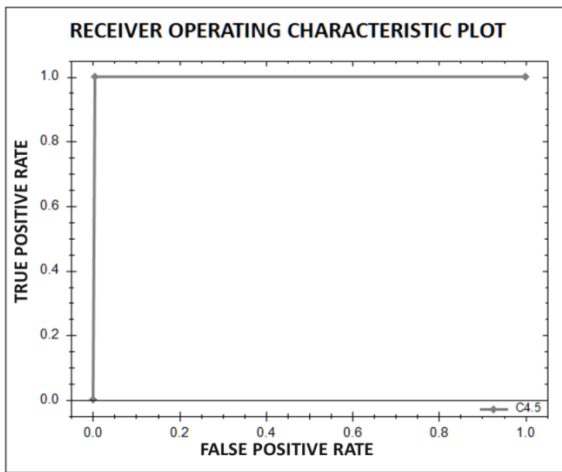


Figure 13. ROC plot for Dataset-1

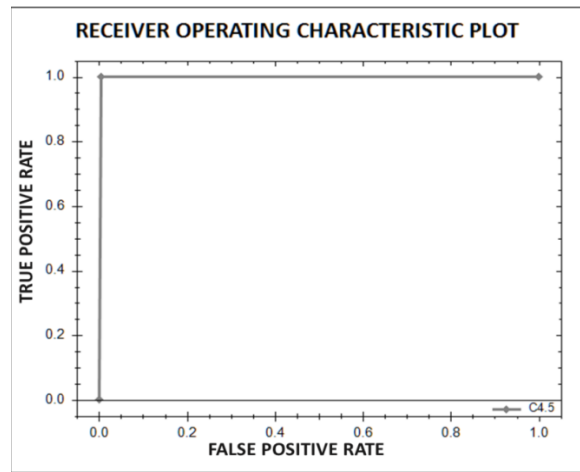


Figure 14. ROC plot for Dataset-2

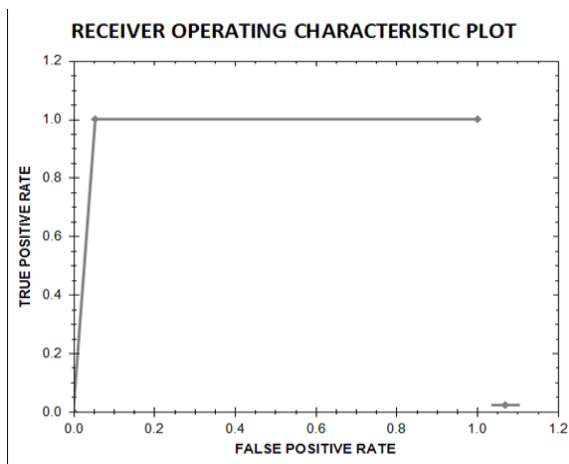


Figure 15. ROC plot for Dataset-3

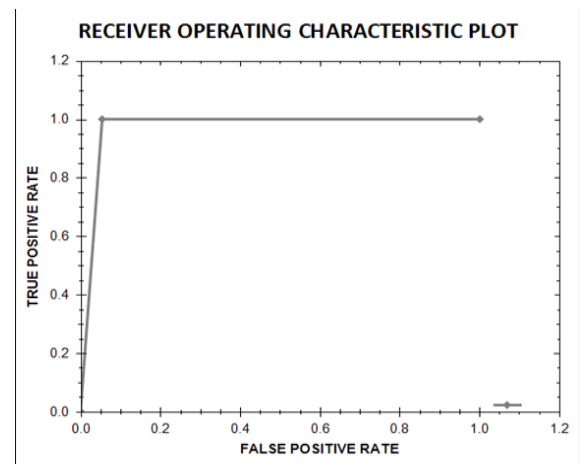


Figure 16. ROC plot for Dataset-4

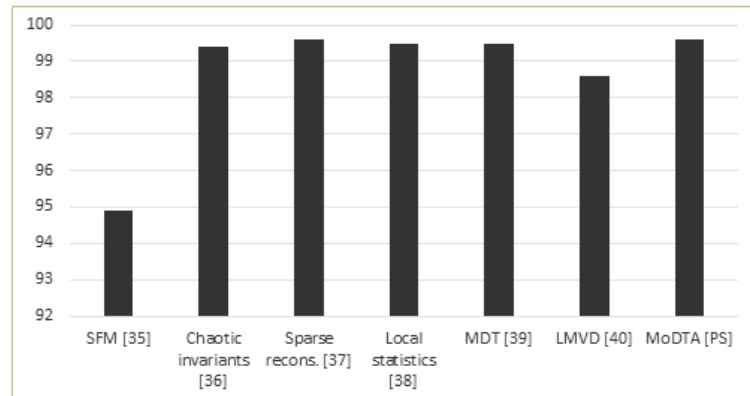


Figure 17. Comparison w.r.t different existing technique

## 5. CONCLUSION

The behavior of crowded scenes is difficult to understand with respect to machine point of view, also the presence of huge diversity and integral complexity in frame makes it very problematic. In order to understand the different behavior of crowd when the crowd-dynamics and crowd-context are changing over a time, it is very much difficult to understand. In this paper, we have proposed the MoDTA that is based on observational filter. The MoDTA initially acquires the people location in an image so that it detects conviction value at pointed locations. MoDTA computes the multiple observed weights values and individual features, so that they can predict the abnormal and normal behavior in a given frame. The proposed approach is applied on the different sets of UMN dataset, afterwards in result analysis we have used BoF classification model to get the AUC with respect to GT, our model got 99.62% AUC which is 1.01% more compared to LMVD [40] that defines the significance of model performance.

## REFERENCES

- [1] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, and S. Yan, "Crowded Scene Analysis: A Survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 367–386, March 2015.
- [2] O. Popoola and K. Wang, "Video-Based Abnormal Human Behavior Recognition -A Review," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 42, no. 6, pp. 865–878, Nov 2012.
- [3] G. J. Burghouts, V. P. Slingerland, H. ten R.J.M, H. den R.J.M, and K. Schutte, "Complex Threat Detection: Learning Vs. Rules, Using A Hierarchy of Features," in *11th IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, 2014, pp. 375–380.
- [4] H. Nallaivarothayan, C. Fookes, S. Denman, and S. Sridharan, "An Mrf Based Abnormal Event Detection Approach Using Motion and Appearance Features," in *11th IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, 2014, pp. 343–348.
- [5] B. T. Morris and M. M. Trivedi, "A Survey of Vision-Based Trajectory Learning and Analysis for Surveillance." *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 18, no. 8, pp. 1114–1127, 2008.
- [6] S. Cosar, G. Donatiello, V. Bogorny, C. Garate, L. O. Alvares, and F. Bremond, "Towards Abnormal Trajectory and Event Detection in Video Surveillance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, pp. 1–1, 2016.
- [7] Poonam G, Shashank B. N, Athri G Rao "Development of Framework for Detecting Smoking Scene in Video Clips", *Indonesian Journal of Electrical Engineering and Computer Science(IJEECS)*, Vol. 13, No. 1, January 2019, pp. 22~26 ISSN: 2502-4752, DOI: 10.11591/ijeeecs.v13.i1.pp22-26.
- [8] D. Tran, J. Yuan, and D. Forsyth, "Video Event Detection: from Subvolume Localization to Spatiotemporal Path Search." *IEEE Transactions on Software Engineering*, vol. 36, no. 2, pp. 404–16, 2013.
- [9] L. Brun, A. Saggese, and M. Vento, "Dynamic Scene Understanding for Behavior Analysis Based on String Kernels," *Circuits & Systems for Video Technology IEEE Transactions on*, vol. 24, no. 10, pp. 1669–1681, 2014.
- [10] Chaveevan Pechsirri, Sumran Phainoun, "Event-Concept Pair Series Extraction to Represent Medical Complications from Texts". *Indonesian Journal of Electrical Engineering and Computer Science(IJEECS)*, Vol. 12, No. 3, December 2018, pp. 1320~1333 ISSN: 2502-4752, DOI: 10.11591/ijeeecs.v12.i3.pp1320-1333.
- [11] O. P. Popoola and K. Wang, "Video-Based Abnormal Human Behavior Recognition: A Review," *IEEE Transactions on Systems Man & Cybernetics Part C*, vol. 42, no. 6, pp. 865–878, 2012.
- [12] D. Lee, H. I. Suk, S. K. Park, and S. W. Lee, "Motion Influence Map for Unusual Human Activity Detection and Localization in Crowded Scenes," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 25, no. 10, pp. 1–1, 2015.

- [13] S. Yi, X. Wang, C. Lu, and J. Jia, "L0 Regularized Stationary Time Estimation for Crowd Group Analysis," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2219–2226.
- [14] K. W. Cheng, Y. T. Chen, and W. H. Fang, "Video Anomaly Detection and Localization Using Hierarchical Feature Representation and Gaussian Process Regression," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [15] B. Krausz and C. Bauckhage, "Loveparade 2010: Automatic Video Analysis of A Crowd Disaster," *Computer Vision & Image Understanding*, vol. 116, no. 3, pp. 307–319, 2012.
- [16] C. L. Chen, X. Tao, and S. Gong, "Salient Motion Detection in Crowded Scenes," in *Communications Control and Signal Processing*, International Symposium on, 2012, pp. 1–4.
- [17] G. Xiong, J. Cheng, X. Wu, Y. L. Chen, Y. Ou, and Y. Xu, "An Energy Model Approach to People Counting for Abnormal Crowd Behavior Detection," *Neurocomputing*, vol. 83, no. 7, pp. 121–135, 2012.
- [18] Y. J. Lee, Y. R. Yeh, and Y. C. F. Wang, "Anomaly Detection Via Online Oversampling Principal Component Analysis," *IEEE Transactions on Knowledge & Data Engineering*, vol. 25, no. 7, pp. 1460–1470, 2013.
- [19] Y. Yuan, Y. Feng, and X. Lu, "Statistical Hypothesis Detector for Abnormal Event Detection in Crowded Scenes." *IEEE Transactions on Cybernetics*, pp. 1–12, 2016.
- [20] Y. Zhang, L. Qin, R. Ji, H. Yao, and Q. Huang, "Social Attribute-Aware Force Model: Exploiting Richness of Interaction for Abnormal Crowd Detection," *Circuits & Systems for Video Technology IEEE Transactions on*, vol. 25, no. 7, pp. 1–1, 2014.
- [21] B. Zhou, X. Wang, and X. Tang, "Understanding Collective Crowd Behaviors: Learning A Mixture Model of Dynamic Pedestrian-Agents," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2871–2878.
- [22] B. Solmaz, B. E. Moore, and M. Shah, "Identifying Behaviors in Crowd Scenes Using Stability Analysis for Dynamical Systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2064–2070, Oct. 2012.
- [23] G. Palanisamy and T. T. Manikandan, "Group Behaviour Profiling for Detection of Anomaly in Crowd," 2017 *International Conference on Technical Advancements in Computers and Communications (ICTACC)*, Melmaurvathur, 2017, pp. 11-15.
- [24] Fan, Yaxiang, Gongjian, Wen, "Detecting Anomalies in Crowded Scenes Via Locality-Constrained Affine Subspace Coding", *Journal of Electronic Imaging* 2017.
- [25] R. A. A. Rupasinghe, D. A. Padmasiri, S. G. M. P. Senanayake, G. M. R. I. Godaliyadda, M. P. B. Ekanayake and J. V. Wijayakulasooriya, "Dynamic Clustering for Event Detection and Anomaly Identification in Video Surveillance," 2017 *IEEE International Conference on Industrial and Information Systems (ICIIS)*, Peradeniya, 2017.
- [26] O. Ibdunmoye, A. R. Rezaie and E. Elmroth, "Adaptive Anomaly Detection in Performance Metric Streams," in *IEEE Transactions on Network and Service Management*, vol. 15, no. 1, pp. 217-231, March 2018.
- [27] Cong, Y., Yuan, J., Liu, J.: "Sparse reconstruction cost for abnormal event detection". In: *IEEE Computer Vision and Pattern Recognition*, pp. 3449–3456. IEEE, Colorado Springs (2011).
- [28] Kim, J., Grauman, K. "Observe Locally, Infer Globally: A Spacetime MRF for Detecting Abnormal Activities with Incremental Updates". In *IEEE Computer Vision and Pattern Recognition*, pp. 2913–2920. IEEE, Miami Beach (2009).
- [29] Saligrama, V., Chen, Z. "Video Anomaly Detection Based on Local Statistical Aggregates". In *IEEE Computer Vision and Pattern Recognition*, pp. 2112–2119. IEEE, Providence (2012).
- [30] Reddy, V., Sanderson, C., Lovell, B.C. "Improved Anomaly Detection in Crowded Scenes Via Cell-Based Analysis of Foreground Speed, Size and Texture". In *IEEE Computer Vision and Pattern Recognition*, pp. 55–61. IEEE, Colorado Springs (2011).
- [31] Nallaivarothayan, H., Fookes, C., Denman, S. "An MRF Based Abnormal Event Detection Approach Using Motion and Appearance Features". In *IEEE International Conference on Advanced Video and Signal based Surveillance*, pp. 343–348. IEEE, Seoul (2014).
- [32] Wang, Q., Ma, Q., Luo, C.H., et al. "Hybrid Histogram of Oriented Optical Flow for Abnormal Behavior Detection in Crowd Scenes". *Int. J. Pattern Recognit Artif Intell.* 30(2), 1655007 (2016).
- [33] Zhang, Y., Lu, H.C., Zhang, L.H., et al. "Combining Motion and Appearance Cues for Anomaly Detection". *Pattern Recognit.* 51, 443–452 (2016).
- [34] Unusual crowd activity dataset of university ofminnesota available at <http://mha.cs.umn.edu/movies/crowd-activity-all.avi>.
- [35] R. Mehran, A. Oyama, and M. Shah, "Abnormal Crowd Behavior Detection Using Social Force Model," in *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, 2009, pp. 935–942.
- [36] S. Wu, B. E. Moore, and M. Shah, "Chaotic Invariants of Lagrangian Particle Trajectories for Anomaly Detection in Crowded Scenes," in *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, 2010, pp. 2054–2060.
- [37] Y. Cong, J. Yuan, and J. Liu, "Sparse Reconstruction Cost for Abnormal Event Detection." in *Computer Vision and Pattern Recognition*, CVPR, 2011, pp. 3449–3456.
- [38] V. Saligrama and Z. Chen, "Video Anomaly Detection Based on Local Statistical Aggregates," in *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, 2012, pp. 2112–2119.
- [39] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly Detection and Localization in Crowded Scenes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 18–32, 2014.
- [40] H. Fradi; B. Luvison; Q. C. Pham, "Crowd Behavior Analysis Using Local Mid-Level Visual Descriptors," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol.PP, no.99, 2017.