Optical Flow-Based Weighted Magnitude and Direction Histograms for the Detection of Abnormal Visual Events Using Combined Classifier

Gajendra Singh, National Institute of Technology, Jalandhar, India
Rajiv Kapoor, Delhi Technological University, India
Arun Khosla, Dr. B. R. Ambedkar National Institute of Technology, Jalandhar, India
https://orcid.org/0000-0001-8571-7614

ABSTRACT

Movement information of persons is a very vital feature for abnormality detection in crowded scenes. In this paper, a new method for detection of crowd escape event in video surveillance system is proposed. The proposed method detects abnormalities based on crowd motion pattern, considering both crowd motion magnitude and direction. Motion features are described by weighted-oriented histogram of optical flow magnitude (WOHOFM) and weighted-oriented histogram of optical flow direction (WOHOFD), which describes local motion pattern. The proposed method uses semi-supervised learning approach using combined classifier (KNN and K-Means) framework to detect abnormalities in motion pattern. The authors validate the effectiveness of the proposed approach on publicly available UMN, PETS2009, and Avenue datasets consisting of events like gathering, splitting, and running. The technique reported here has been found to outperform the recent findings reported in the literature.

KEYWORDS
Abnormality Detection, K-Means, KNN, Optical Flow, Semi-Supervised, Surveillance Video, WOHOFM

INTRODUCTION

Security is a major concern for everyone at public places and hence there is an increase in demand of video surveillance systems. These cameras based video surveillance systems generate a huge amount of data but there are limited number of skilled persons to watch and analyze this data. One cannot solely rely upon human observer because a long time may pass before a suspicious event takes place and human attention may not have remained focus on task in such situations which can lead to an event of interest being missed. So to avoid such situations an automated system is needed that can analyze such huge amount of data and trigger alarm in abnormal events.

Abnormal event detection algorithms generally two step process viz. extraction of features and identification of pattern on extracted features. Feature extraction is a process of transforming raw data into different primitives that describes the scene characteristics in more discriminative way. Feature extraction methods are mainly classified into two classes: an object based and pixel based.

DOI: 10.4018/IJCINI.20210701.oa2

This article, published as an Open Access article on April 23rd, 2021 in the gold Open Access journal, the International Journal of Cognitive Informatics and Natural Intelligence (converted to gold Open Access January 1st, 2021), is 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.
Object based methods extract information about arrangement of pixels and trajectory based methods are one of the object based methods (Basharat et al., 2008), where abnormal events are identified by objects trajectories. Trajectories based methods extract trajectories of each object in the scene and make model based on the trajectory statistics. But these type of methods are not able to perform well in crowed scene because there are too many objects in the scene and tracking of each object is difficult due to occlusion.

Pixel based methods describes the scene by low level features. One such method is based on crowd flow which uses information about motion of crowd. In (Wang et al., 2014; Mishra et al., 2020) abnormal event detection is carried out by modelling optical flow, which tracks change in arrangement of pixel and does not require object tracking. Optical flow based algorithms are capable of easily and precisely modeling highly crowded scene. Feature extraction is followed by identification of patterns to differentiate normal and abnormal events. Pattern identification broadly classified into supervised, semi supervised and unsupervised techniques.

In this paper, a new anomaly detection framework is proposed for video surveillance systems using semi- supervised learning process. For feature extraction optical flow is used. Feature vector consists of both magnitude and direction weighted by and energy function, which robustly describe even small change in movement or in direction and the distribution of activity is modelled by combined classifier (KNN and K- Means). Due to robust features and combined classifier our method works well in challenging conditions with respect to other state of the art methods. Rest of the paper is organized as follows, related work provides an overview of related work on abnormality detection in video surveillance systems. Proposed method section gives an overview of proposed method and describe the process of extraction of useful information from the scene by using weighted oriented histogram of optical flow direction (WOHOFD) and weighted oriented histogram of optical flow magnitude (WOHOFM). This section also describes, how to train our model for detection of abnormal events in the scene. The performance results of our proposed method are described in result section. Finally paper is concluded in last conclusion section.

RELATED WORK

First thing in suspicious event detection is to extract features which can robustly describe the scene statistics e.g. low level features and high level features. After extracting features from the scene, event modelling or classification of data is done based on extracted features. In event modelling, algorithm learns the behavior or pattern of extracted features and classify whether scene contains an anomalous event or not. Event modeling is generally known as machine learning and can be classified into three major categories: supervised techniques, semi-supervised techniques and unsupervised techniques.

The anomaly detection based on supervised technique requires the labeling of samples for both normal samples and abnormal samples to train the model and give prediction on test samples. These methods are generally train model for specific abnormal state whose features are previously known or set, such as 'U’ turn detection in traffic surveillance scene (Zen & Ricci, 2011; Z et al., 2005).

Semi- supervised techniques train model only for normal samples and these techniques can be further categorized into two sub categories viz. rule based and model based. In rule-based methods, predefined rules are set in during training phase based on the characteristics of scene or extracted features, if a test sample do not follow predefined rules, would be classified/labelled as abnormal/irregular e.g., sparse coding (Cong et al., 2011), online dictionary updating (Zhao et al., 2011) and sparse combination learning (Lu et al., 2013) etc. In Sparse coding method (Cong et al., 2011), a reconstruction cost based method for abnormality detection was proposed, if a sample having larger reconstruction cost will be classified as abnormal sample. In (Boiman & Irani, 2007; Saligrama & Chen, 2012; Hamid et al., 2005; Javan & Levine, 2013) similarity based methods were proposed, where score of a test data is calculated based on how much test sample is similar to the training sample, higher similarity means low abnormality score and vice versa. While on the other hand
model based method attempts to build a model only for normal scene, probability of test samples are calculated with respect to the train model, if probability is low then test sample will be considered as anomalous samples. The commonly used models are Hidden Markov Model (HMM) (Kratz & Nishino, 2009; Zhang et al., 2005; Andrade et al., 2006; Ouivirach et al. 2006) and Markov Random Field (MRF) model (Kim & Grauman, 2009; Benezet et al., 2009), which are used in a wide variety of applications including anomaly detection. Kim et al. (Kim & Grauman, 2009) proposed a Space Time Markov Random Field (STMRF) approach illustrating distribution of regular motion behavior. (Kratz et al., 2009) proposed a distribution based HMM model using the local spatial and temporal motion behavior. Andrade et al. (Andrade et al., 2006) proposed a Multiple Observation Hidden Markov Model (MOHMM), which groups the video into different cluster using spectral clustering and trained the model for each cluster. (Mehran et al., 2009) proposed Social Force Model (SFM) for detection and localization of abnormal behavior of crowd based on interaction forces. (Adam et al., 2008) presented a method for abnormal behavior detection, which observes histogram of movement using optical flow for normal behavior at several fixed-locations, (Wu et al., 2010) modelled chaotic invariants of Lagrangian particle trajectories for normal event to characterize crowded scene. (Cui et al., 2011) proposed an interaction energy potential function which describes the action by spatial arrangement with the surroundings of normal objects changing over time and if a test sample having abrupt fluctuations in function is categories as abnormal. (Sharif et al., 2012) proposed an entropy based method, which calculate entropy of the spatiotemporal data of the interest points to measure randomness in video frame. (Kwon & Lee, 2014) proposed a method for the detection of abnormal events based on predefined energy function whose parameters reflect frequency, causality, and significance of events. (Gu et al., 2014) proposed a method for detection of abnormal events which uses particle entropy to describe the distribution of objects in crowded scenes. (Wang et al., 2019) proposed a semi supervised method based on deep network for the detection of abnormal events. Deep network is used for the extraction of features and SVDD (Support Vector Data Descriptor) is used for classification. The performance of deep network based methods is very good but these methods require high computational power.

Unsupervised techniques do not require labeling of normal or abnormal data samples in advance and abnormal event detection is carried out based on the statics that abnormalities are rare with respect to normal situation. Many of the techniques are based on clustering of data and detect abnormal events based on the calculation of distance between the test data sample and the nearest cluster center, if the distance is more the predefined threshold than test sample is labelled as abnormal.

**PROPOSED METHOD**

This section represents our crowd anomaly detection framework. Our proposed method works in two stages: feature extraction and classification. In this method two main features are extracted: optical flow magnitudes and optical flow directions. Optical flow refers to the visible motion of an object in an image, and the apparent flow of pixels with respect to its neighborhood in an image. It is the result of 3-D motion being projected on a 2-D image plane. In normal crowded scenes, people density is high, the movement of individual is constrained by other people’s movement and hence movement of individuals are generally slow. The speed and direction of individuals do not change so much within a short period. However, in abnormal situations with respect to normal situations, the speed of movement of crowd is very high and the direction also changes very rapidly due to the fear. Hence, both direction and magnitude of optical flow become significant features to describe the crowd movements. People started running or diverge from the places where they were. When we want to consider the crowd area to detect abnormal situation, the view field of cameras are kept quite large.

Flow chart of proposed algorithm based on weighted magnitude and direction histogram with combined classifier for abnormal event detection is shown in Figure 1.
Optical flow is calculated for each frame of video by Lucas-Kanade derivative of Gaussian method proposed by Bruce D. Lucas and Takeo Kanade (Lucas & Kanade, 1981). It assumes that the displacement of patches or objects between consecutive frames is constant or small in a local neighborhood of a point under consideration. Lukas Kanade method divides the image into small sectors and assumes constant velocity in each sector and uses least squares criterion to solve the basic optical flow equations for all the pixels. It computes an estimate of the horizontal and vertical velocity component \([U V]^T\) that minimizes the Eq.(1).

$$\sum_{k \in \Omega} W_k^2 \left[ E_x u + E_y v + E_t \right]^2 \ldots \tag{1}$$

Where \(W\) is a window function that emphasizes the constraints at the center of each section and \(E_x\), \(E_y\) and \(E_t\) are derivatives of image brightness in spatial and temporal dimensions and \(u\) is horizontal optical flow and \(v\) is vertical optical flow. Solution to the minimization problem is given in Eq.(2)
Magnitude and direction of optical flow for each pixel is calculated according to Eq.(3) and Eq.(4) respectively.
Where $M(x,y,t)$, $D(x,y,t)$ are optical flow magnitude and direction value at each spatial location $u_{x,y,t}$ and $v_{x,y,t}$ are horizontal and vertical component of optical flow vector respectively.

Initially, each frame of video is divided into $b$ blocks with 50% block overlapping and further each block is divided into $c$ cells e.g. an image is divided into 4 blocks/frame and each block is further divided into smaller parts or cells of some group of pixels. Then, for each cell, it builds separate magnitude and direction representation. Total feature vector length will be equal to $2^b*b*c*9$ (9 no of bins for each cell).

Figure 2 shows a representation how feature vector is extracted for an image, each small block has one arrow in some direction, arrow represents the optical flow value for particular pixel, arrow length shows the magnitude of optical flow value, more the length more optical flow value (more displacement in arrangement of pixels). Arrow direction shows the direction of displacement of pixel arrangement. WOHOFD and WOHOFM are extracted for each block as described below and histograms of all blocks are concatenated for final feature vector.

Weighted Oriented Histogram of Optical Flow Direction (WOHOFD)

The extraction of WOHOFD provides a histogram $H_{D_{b,t}} = [h_1, h_2, ..., h_c]$ at each time instant $t$, for $b^{th}$ block in the frame(where $c$ is $c^{th}$ cell of $b^{th}$ block), in which each flow vector is binned according to its primary optical flow direction from the horizontal axis and weighted according to its optical flow magnitude. For each cell in frame, Direction angle $D_{c}$ is binned into 9 bins and weighted (Qian et al., 2011; Kaltsa et al., 2015) according to magnitude $M_{Weight}$ given by Eq.(5)

$$M_{Weight_{x,y,c}} = \left( \frac{M_{x,y,c}}{\pi} \right)^2 + \left( \frac{\text{"max}M_{x,y,c}}{\pi} \right)^2$$  \hspace{1cm} (4)$$

Where

$$\text{"max}M_{x,y,c} = M_{x,y,c} - \max(M_{x,y,c})$$

And

$$M_{x,y,c} = M_{x,y,c} - \overline{M_{x,y,c}}$$

where $\overline{M_{x,y,c}}$ is average value of optical flow magnitude at all pixel positions within a cell. Then for each block, weighted oriented histogram of optical flow direction is obtained by concatenating histogram of cells.

Finally, a feature vector WOHOFD of frame is obtained by concatenating histogram of blocks $H_{D} = [H_{D_1}, H_{D_2}, ..., H_{D_b}]$. Feature vector describes the global movements into successive frames.
Weighted Oriented Histogram of Optical Flow Magnitude (WOHOFM)

The WOHOFM provides a histogram $HM_b(t) = [h_1, h_2, ..., h_c]$ at each time instant $t$, for each block $b$ in the frame, in which each flow vector is binned according to its optical flow magnitude and weighted according to its optical flow direction angle. For each cell $c$ in frame, Magnitude $M_c$ is binned into 9 bins and weighted according to magnitude $D_{Weight}$ given by Eq.(6)

$$D_{Weight_{x,y,c}} = \left( \frac{\max D_{x,y,c}}{\pi} \cdot 10^2 \right) + \left( \frac{\max D_{x,y,c}}{\pi} \cdot 10^2 \right)$$

(6)

Where

$$\max D_{x,y,c} = D_{x,y,c} - \max D_{x,y,c}$$

And

$$D_{x,y,c} = D_{x,y,c} - \overline{D_{x,y,c}}$$

where $\overline{D_{x,y,c}}$ is average value of optical flow direction orientation at all pixel positions within a cell.

Then for each block weighted oriented histogram of optical flow magnitude is obtained by concatenating histogram of cells. Finally, a WOHOFM of frame is obtained by concatenating histogram of blocks $HM = [HM_1, HM_2, ..., HM_b]$. Final feature vector for each frame is obtained by concatenating both histograms $[HD HM]$ that describes the global movements into successive frames.

Classification

After feature extraction, pattern learning is required to analyze the extracted features and find pattern to classify between anomalous event and normal event or expected behavior. Real world crowd behaviors have much more complex distribution, that may not be modeled by single classifier (Gaussian model, SVM (Support vector Machine), and SVDD (Support Vector Data Descriptor) etc.), but combining multiple classifiers can work better for complex scene or data. Because each classifier focus on a selected features or distribution of data and by combining two or more classifiers will combine the strong points of each classifier (Tax & Duin, 2001). Based on this fact, our proposed work combines two one class classifiers: KNN (K-Nearest neighbor) and K-Means. Our training model is trained only on normal instances or normal behavior data. To make the outputs of the different classifiers comparable, confidence or posterior probabilities are estimated. Posterior probabilities ($P_m(t)$) for each test object $t$ for each of the $N$ classes on which the classifiers are trained and limited in the vicinity of 0 and 1. The posterior probabilities are normalized such that:

$$\sum_{m}^{N} P_m(t) = 1$$

(7)

In K-means, data points are described by k clusters and arranged in such way that the mean distance to a cluster is minimized. The objective function can be given as:
\[ O(s) = \min_i (s - c_i)^2 \ldots \] (8)

where \( s \) is data points and \( c_i \) is \( i^{th} \) cluster center. While in KNN test objects are given to the class of the closest mean. Posterior probabilities are calculated by a sigmoid function. This is optimized over the training set using the maximum likelihood rule. KNN is an instance based learning classifier, it is used as base classifier and that performs classification based on the closest data point in feature space.

As the posterior probabilities \( P_{m\left( t \right)}(t) \), \( m = 1: N \), \( n = 1: K \), for \( N \) classes and \( K \) classifiers are calculated, probabilities are combined to a new set of confidence \( Q_m(t) \) for \( m \) class, which are used for final classification. The new probabilities \( Q_m(t) \) for \( m \) class can be given as:

\[ c_m(t) = \text{mean}(P_{m\left( t \right)}(t)) \ldots \] (9)

\[ Q_m(t) = \frac{C_m(t)}{\sum_{m'}Q_{m'}(t)} \ldots \] (10)

Final decision is made by:

\[ D(t) = \arg\max_m Q_m(t) \ldots \] (11)

RESULTS AND DISCUSSIONS

In this section, results of experiments based on the proposed method are presented. To analyze the performance of proposed method UMN dataset, PETS2009 dataset and Avenue Benchmark Dataset (Lu et al., 2013) are used. All simulation is performed in MATLAB and Intel XEON processor with 8GB RAM. The receiver operating characteristic plot (ROC) and Area Under the Curve (AUC) are used as a perform evaluation parameters. ROC curve is plotted between the true positive rate (Sensitivity- the probability of target is 1 same as it’s true value which is also 1) on y axis and the false positive rate (100%-Specificity, the probability of target is 1 however its true value is 0) on x axis, for different cut off values. Each point on the ROC plot shows a sensitivity/specificity pair corresponding to a fixed threshold. A test with ideal segregation (no overlapping between two distributions) has a ROC curve that passes 100% sensitivity and 100 specificity. AUC value generally tells the quality of a classifier. AUC value equal to 0.5 suggest that classifier is randomly predicting the output, while the perfect classifier will have the AUC value 1. For real world application most of the classifier have an AUC value for 0.5 to 1. Our proposed method is also compared with the recent finding reported in the literature (Wang & Snoussi, 2014; Kaltsa et al., 2015; Shi et al., 2010)

UMN Dataset

UMN dataset have three different scenes of crowd escape behavior. This dataset is recorded by still camera, which is mounted at some elevation. One scene is of lawn, another is of indoor, and last one is plaza scene. First scene is picturized outdoor in a lawn; it contains total 1452 frames. Out of total frames, 1156 frames show normal behavior while 296 frames show crowd escape behavior. Indoor
scene contains total 4144 frames, 2986 frames showing normal situation while 1158 frames shows panic situation, Plaza scene contains 1836 normal situation frames, while 306 frames shows abnormal situation. In all three scenes, initially all persons are roaming all over the area at normal speed but after some time they started running due to some abnormal situation. Some frames of UMN dataset are shown in Figure 3, where 1st and 3rd column of figures showing normal activity while 2nd and 4th column of figures showing abnormal activity.

![Figure 3. UMN dataset: Lawn, indoor and plaza scene (first and third column shows normal activity; Second and Forth column Abnormal activity)](image)

Our model is trained only on normal activity frames and tested on rest of the frames. AUC value for all the three scenes of UMN dataset is compared with HOFO (Wang & Snoussi, 2014), HOS (Kaltsa et al., 2015) STCOG (Shi et al., 2010) and given in table 1. ROC curve and higher value of AUC suggest that our proposed method outperforms over other state of the art methods. ROC curves for all three scene are shown in Figures 4, 5, 6.

**PETS2009 Dataset**

The PETS2009 datasets contain three different sequences encompassing crowd situations with increasing complexity of scene. Dataset S1 is mainly for person count and density estimation. Dataset S2 involves people tracking. Dataset S3 addresses crowd flow analysis and abnormal event detection. The original resolution of the PETS2009 dataset is 728 x 576. For obtaining feature each frame is divided into smaller size blocks with 50% overlapping and each block is further divided in smaller size cells. The experiments are performed on view 1 for time sequence 14-17, 14-16, 14-06 and 14-55. Some frames of dataset are shown in Figure 7.

![Figure 7. PETS2009 dataset](image)

The detection results on PETS2009 (time sequence 14-16) are shown in Figure 8. Where persons are running or walking from left to right or right to left direction. A normal situation corresponds to individuals walking at normal speed. While aberrant situation corresponds to the persons started running.
Figure 4. ROC curve for UMN Lawn

Figure 5. ROC curve for UMN indoor scene
Classifier is trained only for normal situation where individuals are walking, training frames are chosen from time sequence 14-16 and 14-06. By our proposed method 94.34% detection accuracy is achieved. ROC curve for detection result on Time 14-16 is shown in Figure 9, AUC value for roc curve of time 14-16 is 0.9788.

The detection result on PETS2009 (time sequence 14-17) are shown in Figure 10, where walking is taken as normal situation and running is taken as abnormal situation.

Detection accuracy for time 14-17 is 91.78% and AUC value is 0.9543. ROC curve is shown in Figure 11.

Detection results for sequence (Time 14-06 and Time 14-55) are shown in Figure 12, where walking on the pedestrian way is taken as normal situation while walking on the grass or other than the pedestrian way is taken as abnormal situation.

Training frames are chosen from sequence (Time 14-06) and tested on sequence (Time 14-55). our method achieves 91.99% accuracy and AUC value is 0.9721, roc curve is shown in Figure 13.
Figure 7. PETS2009 dataset: (left column: normal scene right column: abnormal scene)

Figure 8. Detection results on PETS2009 (Time 14-16)
Figure 9. ROC curve for PETS2009 (Time 14-16)

Figure 10. Detection results on PETS2009 (Time 14-17)
Figure 11. ROC curve for PETS2009 (Time 14-17)

Figure 12. Detection results on PETS2009 (Time 14-55)
Avenue Dataset
The Avenue dataset contains 16 training videos and 21 videos for testing. The videos include a total of 15183 frames for training and 15324 frames for testing. The resolution of video is 640*360. The training videos capture normal situations. Testing videos include both normal and abnormal events. This dataset contains abnormalities like persons running to and fro, persons are going towards such area form wrong direction, where only one way movement is allowed and persons are jumping. Three abnormal detected samples are shown in Figure 14.

This dataset also contains the challenges like minor camera shake (testing video 2, frame 1051 - 1100) presents, a few outliers are included in training data and some normal patterns seldom appear in training data. ROC Curve of detection result on Avenue dataset are shown in Figure 15 and comparison result of our method with (Lu et al., 2013) in table 2.

CONCLUSION
In this paper, a new and effective algorithm for suspicious event detection based on direction and magnitude is proposed. Crowd movement information is represented by movement direction and speed and model is trained on combined classifier (Simple but effective), which utilizes strong points of both classifier and perform very great in anomaly detection. ROC curve and high value of AUC of our model on different datasets show good performance of proposed algorithm over state-of-the-art methods. High detection rate on UMN, PETS2009 datasets and Avenue dataset demonstrate that our proposed algorithm can be used in challenging conditions specially in little bit noisy conditions. It’s effectiveness in challenging crowded scenes, make our method very suitable for a wide variety of video surveillance applications. Our proposed method achieve good accuracy over different crowd datasets, however our algorithm is not suitable for real time application because optical flow calculation of each frame consumes a lot of time of this algorithm. So some different approach to obtain motion information can be used in future to make this algorithm work in real time. So in future, some other addition information can also be utilize e.g. deep learning, shape information etc., to make this algorithm work more efficiently and effectively in more challenging conditions.
Figure 14. Abnormality Detection (Wrong Direction, Running, Jumping) on Avenue Dataset

Figure 15. ROC curve for Avenue Dataset

Table 2. Comparison results of our method with state of art methods (AUC performance of anomaly detection on the Avenue dataset)

| S.No | Method | AUC   |
|------|--------|-------|
| 1    | (Lu et al. 2013) given in (Del Giorno et al. 2016) | 0.809 |
| 2    | Proposed Method | 0.871 |
REFERENCES

Adam, A., Rivlin, E., Shimshoni, I., & Reinitz, D. (2008). Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*(3), 555–560. doi:10.1109/TPAMI.2007.70825 PMID:18195449

Andrade, E. L., Blunsden, S., & Fisher, R. B. (2006, August). Modelling crowd scenes for event detection. In *18th international conference on pattern recognition (ICPR’06)* (Vol. 1, pp. 175-178). IEEE. doi:10.1109/ICPR.2006.806

Basharat, A., Gritai, A., & Shah, M. (2008, June). Learning object motion patterns for anomaly detection and improved object detection. In *2008 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-8). IEEE. doi:10.1109/CVPR.2008.4587510

Benezeth, Y., Jodoin, P. M., Saligrama, V., & Rosenberger, C. (2009, June). Abnormal events detection based on spatio-temporal co-occurences. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2458-2465). IEEE. doi:10.1109/CVPR.2009.5206686

Boiman, O., & Irani, M. (2007). Detecting irregularities in images and in video. *International Journal of Computer Vision, 74*(1), 17–31. doi:10.1007/s11263-006-0009-9

Cong, Y., Yuan, J., & Liu, J. (2011, June). Sparse reconstruction cost for abnormal event detection. In *CVPR 2011* (pp. 3449–3456). IEEE. doi:10.1109/CVPR.2011.5995434

Cui, X., Liu, Q., Gao, M., & Metaxas, D. N. (2011). Abnormal detection using interaction energy potentials. *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 3161–3167.

Del Giorno, A., Bagnell, J. A., & Hebert, M. (2016, October). A discriminative framework for anomaly detection in large videos. In *European Conference on Computer Vision* (pp. 334-349). Springer. doi:10.1007/978-3-319-46454-1_21

Fu, Z., Hu, W., & Tan, T. (2005, September). Similarity based vehicle trajectory clustering and anomaly detection. In *IEEE International Conference on Image Processing 2005* (Vol. 2, pp. II-602). IEEE.

Gu, X., Cui, J., & Zhu, Q. (2014). Abnormal crowd behavior detection by using the particle entropy. *Optik (Stuttgart)*, 125(14), 3428–3433. doi:10.1016/j.ijleo.2014.01.041

Hamid, R., Johnson, A., Batta, S., Bobick, A., Isbell, C., & Coleman, G. (2005, June). Detection and explanation of anomalous activities: Representing activities as bags of event n-grams. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)* (Vol. 1, pp. 1031-1038). IEEE. doi:10.1109/CVPR.2005.127

Javan Roshtkhari, M., & Levine, M. D. (2013). Online dominant and anomalous behavior detection in videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2611–2618). doi:10.1109/CVPR.2013.337

Kaltsa, V., Briassouli, A., Kompatsiaris, I., Hadjileontiadis, L. J., & Strintzis, M. G. (2015). Swarm intelligence for detecting interesting events in crowded environments. *IEEE Transactions on Image Processing, 24*(7), 2153–2166. doi:10.1109/TIP.2015.2409559 PMID:25769154

Kim, J., & Grauman, K. (2009, June). Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2921-2928). IEEE. doi:10.1109/CVPR.2009.5206569

Kratz, L., & Nishino, K. (2009, June). Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1446-1453). IEEE. doi:10.1109/CVPR.2009.5206771

Kwon, J., & Lee, K. M. (2014). A unified framework for event summarization and rare event detection from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(9), 1737–1750. doi:10.1109/TPAMI.2014.2385695 PMID:26353123

Lu, C., Shi, J., & Jia, J. (2013). Abnormal event detection at 150 fps in matlab. *Proceedings - IEEE International Conference on Computer Vision, 2720–2727*. doi:10.1109/ICCV.2013.338
Lu, C., Shi, J., & Jia, J. (2013). Abnormal event detection at 150 fps in matlab. In *Proceedings of the IEEE international conference on computer vision* (pp. 2720-2727). doi:10.1109/ICCV.2013.338

Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *IJCAI’81 Proceedings of the 7th international joint conference on Artificial intelligence*, 674–679.

Mehran, R., Oyama, A., & Shah, M. (2009, June). Abnormal crowd behavior detection using social force model. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 935-942). IEEE. doi:10.1109/CVPR.2009.5206641

Mishra, S. R., Mishra, T. K., Sarkar, A., & Sanyal, G. (2020). Detection of Anomalies in Human Action Using Optical Flow and Gradient Tensor. In *Smart Intelligent Computing and Applications* (pp. 561–570). Springer. doi:10.1007/978-981-3-9282-5_53

Ouivirach, K., Gharti, S., & Dailey, M. N. (2013). Incremental behavior modeling and suspicious activity detection. *Pattern Recognition*, 46(3), 671–680. doi:10.1016/j.patcog.2012.10.008

PETS. (2009). *Benchmark dataset*. http://cs.binghamton.edu/~mrldata/pets2009

Qian, H., Wu, X., & Xu, Y. (2011). Dynamic analysis of crowd behavior. In *Intelligent Surveillance Systems* (pp. 119–154). Springer. doi:10.1007/978-94-007-1137-2_8

Saligrama, V., & Chen, Z. (2012, June). Video anomaly detection based on local statistical aggregates. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2112-2119). IEEE. doi:10.1109/CVPR.2012.6247917

Sharif, M. H., & Djeraba, C. (2012). An entropy approach for abnormal activities detection in video streams. *Pattern Recognition*, 45(7), 2543–2561. doi:10.1016/j.patcog.2011.11.023

Shi, Y., Gao, Y., & Wang, R. (2010, August). Real-time abnormal event detection in complicated scenes. In *2010 20th International Conference on Pattern Recognition* (pp. 3653-3656). IEEE. doi:10.1109/ICPR.2010.891

Tax, D. M., & Duin, R. P. (2001, July). Combining one-class classifiers. In *International Workshop on Multiple Classifier Systems* (pp. 299-308). Springer. doi:10.1007/3-540-48219-9_30

UMN. (n.d.). *Unusual event datasets of University of Minnesota*. http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi

Wang, T., Miao, Z., Chen, Y., Zhou, Y., Shan, G., & Snoussi, H. (2019). AED-Net: An Abnormal Event Detection Network. *Engineering.*, 5(5), 930–939. doi:10.1016/j.eng.2019.02.008

Wang, T., & Snoussi, H. (2014). Detection of abnormal visual events via global optical flow orientation histogram. *IEEE Transactions on Information Forensics and Security*, 9(6), 988–998. doi:10.1109/TIFS.2014.2315971

Wu, S., Moore, B. E., & Shah, M. (2010, June). Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 2054-2060). IEEE. doi:10.1109/CVPR.2010.5539882

Zen, G., & Ricci, E. (2011, June). Earth mover’s prototypes: A convex learning approach for discovering activity patterns in dynamic scenes. In *CVPR 2011* (pp. 3225–3232). IEEE. doi:10.1109/CVPR.2011.5995578

Zhang, D., Gatica-Perez, D., Bengio, S., & McCowan, I. (2005, June). Semi-supervised adapted hmms for unusual event detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)* (Vol. 1, pp. 611-618). IEEE. doi:10.1109/CVPR.2005.316

Zhao, B., Fei-Fei, L., & Xing, E. P. (2011, June). Online detection of unusual events in videos via dynamic sparse coding. In *CVPR 2011* (pp. 3313–3320). IEEE. doi:10.1109/CVPR.2011.5995524
Gajendra Singh received the Bachelor degree in Electronics and communication engineering from GLAITM, Mathura and his post-graduation degree in Electronics and communication engineering from NIT, Jalandhar. His research interests are Image Processing, Deep Learning and Machine Learning.

Rajiv Kapoor is professor at Delhi Technological University, Delhi. He was also Principle of AIACT&R, Delhi. He has published hundreds of research papers on Image and signal analysis in journal and conferences. His research interest includes Vision/Speech based Tracking, Activity Recognition Vision/Speech based, Signal Processing, Pattern Recognition, Cognitive Radio.

Arun Khosla is Professor at NIT Jalandhar. His areas of interest are Image Processing, Assisted Technologies and Games for Health. He is Fellow of IETE and Life Member of ISTE and Computer Society of India.