Suspicious Action Detection in Intelligent Surveillance System Using Action Attribute Modelling

Manisha Mudgal*, Deepika Punj and Anuradha Pillai

Department of Computer Engineering, JC BOSE UST YMCA Faridabad, Haryana, India
E-mail: mudgal.05.manisha@gmail.com; deepikapunj@gmail.com; anuangra@yahoo.co.in
*Corresponding Author

Abstract

Research in the field of image processing and computer vision for recognition of suspicious activity is growing actively. Surveillance systems play a key role in monitoring of sensitive places such as airports, railway stations, shopping complexes, roads, parking areas, roads, banks. For a human it is very difficult to monitor surveillance videos continually, therefore a smart and intelligent system is required that can do real time monitoring of all activities and can categories between usual and some abnormal activities. In this paper many different abnormal activities has been discussed. More focuses is given to violence activity like hitting, slapping, punching etc. For this large human action dataset like UCF101, Kaggel is required. This paper proposes a method to model violence actions using Gaussian Mixture Model with Universal Attribute Model. Super action vector is calculated using UMA. To represent every SAV in few significant attributes, factor analysis is performed and result gives a low dimensional relevant action vectors.

1 Introduction

In today's world cameras are popularly chosen for addressing various security concerns in both public and private sites. Surveillance systems can be used to monitor occurrence of activities in these places. Abnormal activities are like hitting, fighting, snatching, punching, fire, attacks etc. Normal activities are like jogging, running, walking, handshaking that can be performed by humans at public places. To monitor these activities video surveillance is increasing day by day. These videos capture all kind of activities of human. In semi-automatic surveillance systems, a human expert is required that can continually monitor and analyse the video. These semi-automatic systems are costly and not very reliable. It is very difficult for a guard to sit and watch videos to prevent occurrence of any abnormal activity. Therefore a smart and intelligent system is required that can not only monitor the activities but can also analyse between normal and abnormal. These smart systems or fully automatic systems can also warn the security agencies if some activities are abnormal.

These intelligent systems can also be used to save public places from explosive attacks done by luggage bags at public places like railway stations. It is very difficult for a guard to monitor crowded places and detect suspicious objects. Therefore a smart and intelligent system is required that can detect un-attempted stationary objects (shown in Figure 1(a)).

These smart and intelligent surveillance systems can also be placed in hospitals to monitor patients. If activities like falling, shouting or any other abnormal behaviour of elderly patient (shown in Figure 1(b)) occurs then these smart systems can visualize and analyse videos and can also alert doctors about the condition of patient.

Intelligent systems can also be used to monitor illegal parking (shown in Figure 1(c)), traffic flow. Illegal parking of cars or other vehicle can cause traffic jams on road. Some wrong activities like taking wrong U turn (shown in Figure 1(d)) or driving on wrong lane can cause accidents. With the help of deep learning, a smart and intelligent system can be developed that can alert traffic police about these.

Fire not only damages the infrastructure but can also take lives of human being. Many fire detecting sensors have been developed. These detectors need to be placed very close to the fire. Therefore such sensors cannot be placed in open spaces like park, roads etc. These sensors also cannot inform about fire (shown in Figure 1(e)) growing rate, location and size. An intelligent video system is a best answer to all these problems and can save

Figure 1   (a) Abandoned object, (b) patient falling out of bed, (c) illegal parking, (d) wrong U turn, (e) fire (f) fighting.

lives of many human beings by alerting the fire extinguishing agencies on time.

An intelligent system is also demanded to monitor violence activities such as hitting, fighting (shown in Figure 1(e)), slapping, punching etc. Earlier on the complaint of victim, police or security agencies use to check the footage. With intelligent video surveillance systems, these security agencies will be alerted at the same time of violence occurrence. This can be very helpful in reducing the crime rate and can warn police at the same moment.

These vision based devices are becoming more and more interesting by providing real time information and saving the world from conduct of abnormal acts.

To develop a smart system that can recognize above mentioned abnormal behaviours, most of the researchers have followed these steps:

1. ***Subtraction of background*:** It is very important to foreground the object by detecting the changes in the sequence of frames. For this one of the powerful mechanisms is subtraction of background.
2. ***Detection of Object*:** Object detection is one of the most important task. It can be done through tracking or non-tracking based approaches.
3. ***Extraction of features*:** Feature extraction of objects like shape and motion is done to through different algorithms to identify objects. These feature vectors are then passed to classifiers as input.
4. ***Classification of Object*:** This is done to classify objects of videos. For classification of objects different algorithms can be used like SVM, Face Recognition.
5. ***Analysis of Object*:** After Object recognition analysis of object activity is conducted. These are then compared with different threshold values.

This paper presents us with motivation and different applications of these surveillance systems. Later on issues and challenges faced are discussed. Research work done related to intelligent surveillance system is also discussed. In next sections general steps followed for classification of suspicious activities and different datasets measures are discussed. In late, a proposed framework is discussed for violence action detection.

2 Motivation and Application of this Intelligent Systems

Intelligent Surveillance systems for suspicious activity detection are very important to prevent theft cases, explosive attacks, Fire at sensitive areas, fights, accidents on roads.

This smart and intelligent video surveillance system can protect these sensitive areas from suspicious activities:

1. In College and University Campus: These smart and intelligent systems can be placed in college and university campus to prevent fight and for the safety of assets.
2. Airports: Airports are one of the most sensitive areas of any country. If a real time check through video surveillance is done then safety of passengers, airport will increase.

3. Railway, Bus Stations: Railway and Bus stations are targeted by terrorist of acts. With the help of smart video surveillance system, railway stations, bus stands, parking area can be monitored and activities can be detected at real time.
4. Hospitals: In hospitals doctors can monitor patients remotely with this. In case of elderly patients falling, vomiting, fainting or occurrence of any other abnormal activity can be informed on real time to doctors.
5. Banks: Banking sector need more security as anyone with arms can conduct robbery. If an intelligent surveillance system is installed then police can be informed on time and robbery can be stopped or prevented on time.

3 Issues And Challenges Faced

While developing a good intelligent system for detection of suspicious activities, some issues or challenges can be faced:

Changes Due To Illumination: Nature is quiet unpredictable, sometimes illumination occurs due to change in weather, it can also occur during day-night change. And Illumination in video can be a challenge for video analysis.

Object Shadow: Shadow of an object can create problem while tracking an object and can change the appearance of that object.

Noise in Video: Noise can also act as a problem in video analysis. Noise can be of any kind like rain, dust, waving created by branches of tree.

Huge Crowd: Object detection from a huge crowd is very difficult. In more crowded areas detection of violence, theft, slapping, hitting is very difficult task.

Blurred Objects: It is very difficult to find features of Blurred Objects and its get very difficult to recognise.

Occlusion of Object Partially or Fully: It is very difficult to identify the partially or fully occluded objects.

Poor Resolution: If the resolution is not good then it becomes very difficult to detect foreground objects from videos. It becomes difficult to classify the objects as boundaries of the object are not very clear.

Processing at Real Time: The most challenging task is to develop a real time system. Sometimes video with complex background takes more time in processing and tracking of object may take time.

4 Research Work Done in the Field of Suspicious Activity Recognition Through CCTV Camera Videos

This section covers the work done till date in the field of Suspicious human activity done.

In 2016 **Nam** [5] for detecting abandoned and stolen objects has used spatio-temporal features. For the removal of ghost images and for stable tracking, adaptive background modelling has been used. To detect stolen object or abandoned object, spatio-temporal relationship is determined between the stolen or abandoned object and moving human. Tracking Trajectories has been employed to reduce rising of wrong alarms. Vector matching algorithm has been employed to detect the occluded objects.

In 2015 researcher **Dimitropouloes** [18] proposed a method for real time fire detection to model behaviour of fire by using spatio-temporal features like color probability, spatio temporal energy, flickering.

In 2015 **Tripathi** [10] developed a framework to detect suspicious activities happing on ATM installations like fight with user, money snatching, attack on user and framework will raise an alarm to warn the authorities. For reducing dimensions PCA was used and for classification SVM was used.

In 2012 **Wiliem** [20] presented an approach for contextual information based automatic suspicious behaviour detector. In this author used three main things: data stream clustering algorithm was used which enables the system to update knowledge continuously according to information coming from video, context space modelling and inference algorithm.

Adam In **2008** [16] presented non tracking based real time framework for suspicious activity detection which was robust and can work in crowded scenes. Instead of object tracking, it monitored the low level measurements. So, this has a limitation of not doing sequential monitoring.

5 A General Framework For Suspicious Activity Recognition

This section presents the general steps that are followed while detecting any kind of suspicious activities. This framework will work for abandoned objects, theft incidents, fire, falling of a human, illegal parking, violence detection. Most of the researchers follow these steps for recognition of suspicious acts along with different algorithms and approaches to improve the performance.

5.1 Foreground Object Detection

Extraction of foreground object is a very important step. It is an initial step for detection of suspicious activities. Subtraction of background is performed to detect the changes in the frames and to do the extraction of foreground object. Moving objects in a video are considered as foreground objects and static objects are considered as video background.

For detection of moving object any of the two methods can be followed –

1. Background modelling
2. Change detection

For moving foreground object detection, researchers have used many different methods to extract activities like robbery, fights, punching, slapping, snatching, and falling. Earlier background modelling method was used using a single Gaussian than multi modal distribution was implemented using mixture of various Gaussians. With time new advancements came and it leads to the development of new approaches. Background subtraction is a very common technique which can detect moving object by differencing between the current frame and background model.

Many researchers have tried to detect stationary foreground objects using different methods. In this dual background modelling techniques can also be applied to detect abandoned objects.

For removing noise, illumination and shadow effect to detect the foreground object is a very difficult task to do. These can create problems in identification of objects that can also lead to false classification. Several researchers have applied different techniques to remove these effects and improve the quality of footage. To reduce these effects color normalization, Phong Shading, Radial Reach Filter, Gaussian smoothing, fuzzy histogram color can be used.

5.2 Object Tracking

Tracking of Object is also a difficult task in computer vision field. For object tracking a trajectory is created over time by tracing the position of object in sequential frames. Object representations that are used for tracking object are object contours, geometric shapes, points, articulates etc. Noise, complex shapes of objects, partial occlusion of objects sometimes create problems in tracking of objects.

5.3 Extraction of Features

It is very important to select appropriate features for automatic detection of suspicious activities from videos.

5.4 Classification of Activities

After extracting moving and static foreground objects, the classification of object is implemented to classify between normal and abnormal behaviour. For classification several researchers have used different methods like Support Vector Mechanism, K-NN, Neural Network, and Multi-SVM.
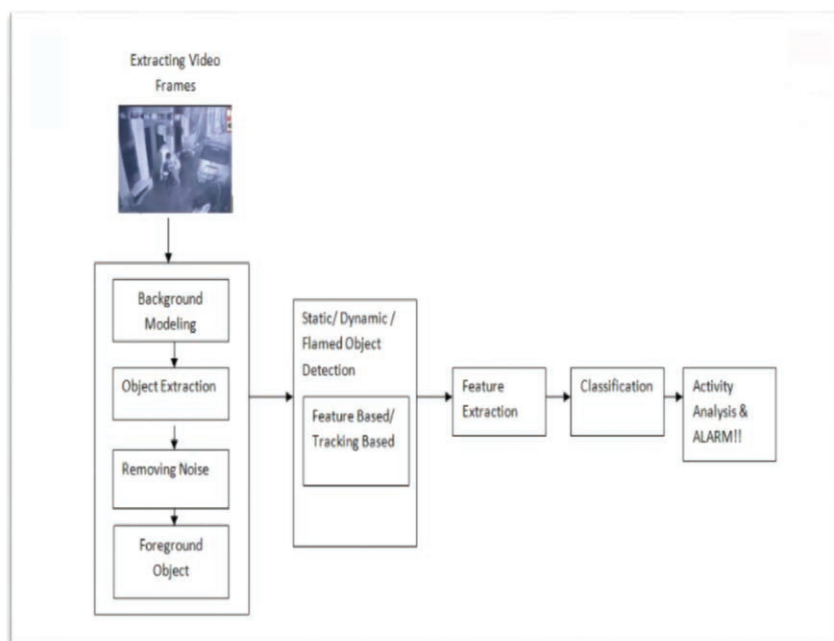
Figure 2   General framework of Intelligent Surveillance system.

6 Data Sets and Evaluation Measures

6.1 Dataset

Violence detection dataset mainly consists of four types of video sequences of fight scenes. In some videos people meet, fight and run away. In other videos

two or three people meet, fight and then one fell down and second person runs away. This can also possible that people meet, fight and then chase by running behind each other. For this UCF101 or Kaggel dataset can be used. In these data sets realistic action videos have been collected from online video store spaces like YouTube.

For traffic MIT traffic dataset is available and for parking I-LIDS parking dataset also provides sequence of videos of illegal parking.

Similarly for theft videos Bank Dataset can be used. For fire detection videos are also available on FireSense dataset. In videos of falling there can be sequences of activities like walking, falling little forward or backward, then complete falls, sideway falls, falls due to balance loss.For these videos CAVIAR video dataset can be used.

6.2 Evaluation Measures

Evaluation of Intelligent video surveillance system for checking performance for different activities like theft, violence detection, illegal parking, accident, fall detection, fire detection is one of the most important tasks.

Many quantitative accuracy test measures have been used by researches like:

Recognition Accuracy: For recognition accuracy of different activities measures like:

$$\text{Accuracy}(\%) = \frac{(T^p + T^N)}{(T^p + T^N + F^p + F^N)}$$

Here Tp represents True positive which means suspicious activity detection as suspicious by classifier.

TN represents True negative means non suspicious activity detections as non-suspicious.

Fp represents false positive means classification of non- suspicious as suspicious.

FN represents false negative which means suspicious as non-suspicious.

Precision, Recall are used as experimental evaluators. In which Precision represents True alarms % and Recall represents detected event %.

$$\text{Recall } (\%) = \frac{(T^p)}{(T^p + F^N)}$$

$$\text{Precision } (\%) = \frac{(T^p)}{(T^p + F^p)}$$

7 Proposed Framework For Violence Detection

In the proposed model similarity between violence activities to normal action is exploited to train a large UAM which move around attributes across all actions.

Universal Attribute Model (UAM). UAM is not dependent on labelled violence videos for training.

7.1 Universal Attribute Model (UAM) Construction

For every video has random action process. For sampling function, there is a need of pdf of parameters that describes the action process. For estimating that pdf GMM that is Gaussian Mixture Model can be used. For this the mixtures must be large to properly accommodate the intra action variances occurred in different videos.

This model can be represented as

$$P(x1) = \Sigma_{c=1}^{C} w_c N(X1l\mu_c, \sigma_c).$$

Here mixture of weight w_c follow the constraint

$$\Sigma_{c=1}^{C} w_c = 1$$

μ_c, σ_c represent (means and covariance) for UAM c mixture. X_l is a feature of X video clip which is represented as $x_1, x_2, x_3 \ldots, x_L$. These features can be Motion Boundary Histogram descriptor or Histogram of optical Flow and for each separate UAM is trained during evaluation.

For finding pdf, maximum aposteriori is adopted which enhances the contribution of attributes of that clip. This MAP gives feature vector x_1 and $p(c|x_1)$ is likelihood which draws feature x_1 for c$^{\text{th}}$ mixture.

$$P(c|x_l) = \frac{w_c p(x_l|c)}{\Sigma_{c=1}^{C} w_c \, p(x_l|c)}$$

Here w_c is mixture's prior probability.

$$P(c|x_l) = \frac{w_c p(x_l|c)}{\Sigma_{c=1}^{C} w_c \, p(x_l|c)}$$

This probability $p(c|x_1)$ is used to calculate x clip's Zeroth and Baum-Wech first order statistics.

$$n_c(x) = \Sigma_{l=1}^{L} p(c|x_1),$$
$$F_c(x) = (1 \div (n_c(x)))\Sigma_{l=1}^{L} p(c|x_1)x_1$$

To calculate adapted weights and means for every c mixture of UAM:

$$\widehat{wc} = \alpha n_c(x)/L + (1-\alpha)w_c$$
$$\widehat{\mu}c = \alpha F_c(x)/L + (1-\alpha)\mu_c$$

To compute $(Cd\ x_1)$ dimensional super action vector means are concatenated. And are represented as $s(x) = \widehat{\mu}_1, \widehat{\mu}_2, \widehat{\mu}_3 \ldots, \widehat{\mu}_c]t$ but it may contain attributes that does not contribute to the chip. Therefore to low dimensions, Super action vector is intrinsically low dimensions by decomposition.

To low dimensions the super action vector is decomposed as $s = m + Tw$.

Here m is a super vector, T is total variability $(Cd\ X\ r)$ size matrix. w is r dimensional action vector. The action vector posterior distribution is as

$$P(w|x) \propto P(x|w)N(0,1) \propto exp((-1/2)(w - L(x)^t\ M(x)(w - L(x)))).$$

The $L(x)$ matrix $= M^{-1}(x)T^t\sum^{-1}\tilde{s}(x)$. Here $\tilde{s}(x)$ is centred super vector.

The Baum Wech statistics first order around our Universal Action Accommodation model mean can be calculated using $\widehat{Fc}(x) = \sum_{l=1}^{L} = 1\ p(c|x_l)(x_l - \mu_c)$.

The centred super vector can now be expressed as $\hat{S}(x) = [F \sim 1(x),\ F \sim 2(x) \ldots F \sim C(x)]t$.

Matrix M is represented as $M(x) = \text{Identity} + T^t\sum^{-1}N(x)T$, Identity is the identity matrix and $N(x)$ diagonal matrix.

The covariance matrix and mean matrix of posterior distribution are

$$\text{Covariance}(w(x), w(x)) = M^{-1}(x),$$
$$E[w(x)] = M^{-1}(x)T\ ^t\Sigma^{-1}s \sim (x)$$

Iteratively the posterior covariance and mean are estimated using EM algorithm [21] in step E and in step M same is used to update T and Σ. The total Variability matrix are initialized randomly, T and rank r are chosen. Then mean and covariance is calculated using above equations of both.

After M steps T and Σ matrices are estimation and action vector of the video clip is shown by means of its posterior distribution.

$$W(x) = (I + T^t\Sigma^{-1}N(x)T)^{-1}T^t\Sigma - 1\check{s}(x)$$

This vector is now used to train the classifier to detect violence activity.

8 Experiment and Results

8.1 Violence Dataset Description

The surveillance camera videos are mostly of low resolution. So, it's challenging to activity to detect the person. As violence incidents are frequently seen in campus or canteen area in colleges. Some violence footages are

**Figure 3**   Violence detection from framework.

collected form and clips of violence are also available on UFC101, Kaggle. The clips are divided into 6 second clips.

8.2 Feature Descriptor Effects

HOF and MBH are two action vectors state of art. The performance in classification of these vectors is shown in Table 1 for different UAM mixtures. Different classifiers like KNN, SVM are used for evaluation.

Table 1 Performance of Action vector for Violence Activity using different classifiers

	UAM Mixture					
	HOF			MBH		
Classifier	256	512	1024	256	512	1024
SVM	99.6	99.6	99.8	99.8	99.5	99.4
ESDA	99.1	99.2	99.2	99.2	99.3	99.3
K-NN	99.4	99.3	99.4	99.6	99.5	99.5
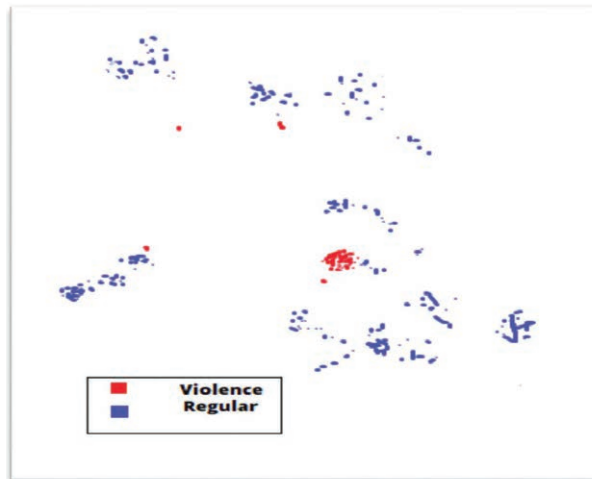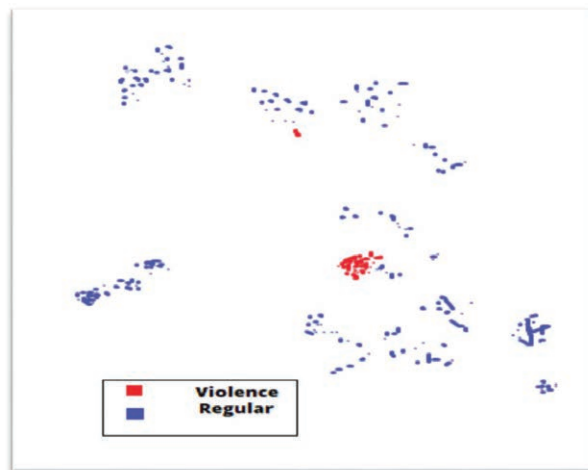
Figure 4 Showing Violence action vector using HOF features.

9 Conclusion

Crime is one of the biggest problems in today's world. To decrease the rate of crime, there is a need for smart and intelligent surveillance system. This paper initially presented us with steps that are generally followed for any suspicious video classification. Then different datasets where explained in detail for abnormal activities. This paper also presents a framework for violence activity detection. For this action vector is used. A large GMM called Universal Attribute Accommodation model is used to learn about all action attributes of violence. Factor analysis is done to remove redundant attributes and violence datasets were used for evaluation. And results shown that action vector works better than state of art feature vectors.IN future this technique can be used with live streaming cameras for real time alerts.

References

[1] Rougier C, Meunier J, St-Arnaud A, Rousseau J (2011). Robust video surveillance for fall detection based human shape deformation. IEEE Trans Circuit Syst Video Technol 21(5):611–622.

[2] Seebamrungsat J, Praising S, Riyamongkol P (2014). Fire detection in the buildings using image processing.Third ICT international student project conference (ICT-ISPC), 2014, IEEE, pp. 95–98.

[3] Z. Wang, Y. Wang, L. Wang, Y. Qiao, Codebook enhancement of vlad representation for visual recognition, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 1258–1262.

[4] Y. Zhang, H. Lu, L. Zhang, X. Ruan, Combining motion and appearance cues for anomaly detection, Pattern Recognition 51 (2016) 443–452.

[5] Nam Y (2016). Real-time abandoned and stolen object detection based on spatio-temporal features in crowded scenes. Multimed Tools Appl 75(12):7003–7028.

[6] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[7] J. F. Henriques, R. Caseiro, P. Martins, J. Batista, High-speed tracking with kernelized correlation filters, IEEE Transactions on Pattern Analysis and Machine Intelligence 37(3) (2015) 583–596.

[8] D. G. Lee, H. I. Suk, S. K. Park, S. W. Lee, Motion influence map for unusual human activity detection and localization in crowded scenes,

IEEE Transactions on Circuits and Systems for Video Technology 25(10) (2015) 1612–1623.

[9] Zin TT, Tin P, Toriu T, Hama H (2012b). A probability-based model for detecting abandoned objects in video surveillance systems. In: Proceedings of the world congress on engineering, vol 2.

[10] Tripathi V, Gangodkar D, Latta V, Mittal A (2015). Robust abnormal event recognition via motion and shape analysis at ATM installations. J Electr Comput Eng 2015.

[11] Mohammad Nakib, Rozin Tanvir Khan, Md. Sakibul Hasan, Jia Uddin, "Crime Scene Prediction by Detecting Threatening Objects Using Convolutional Neural Network', International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), IEEE 2018.

[12] Alkesh Bharati, Dr Sarvanaguru RA, "Crime Prediction and Analysis Using Machine Learning", International Research Journal of Engineering and Technology IRJET), (2018).

[13] Babakura A, Sulaiman MN, Yusuf MA. Improved method of classification algorithms for crime prediction. InBiometrics and Security Technologies (ISBAST), 2014 Aug 26 (pp. 250–255). IEEE.

[14] K. Soomro, A. R. Zamir, M. Shah, UCF101: A dataset of 101 human actions classes from videos in the wild, CoRR abs/1212.0402. URL http://arxiv.org/abs/1212.0402

[15] Mohammad Nakib, Rozin Tanvir Khan, Md. Sakibul Hasan, Jia Uddin, "Crime Scene Prediction by Detecting Threatening Objects Using Convolutional Neural Network", International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), IEEE 2018.

[16] Adam A, Rivlin E, Shimshoni I, Reinitz D (2008) Robust real-time unusual event detection using multiple fixed-location monitors. IEEE Trans Pattern Anal Mach Intell 30(3):555–560.

[17] D. Xu, Y. Yan, E. Ricci, N. Sebe, Detecting anomalous events in videos by learning deep representations of appearance and motion, Computer Vision and Image Understanding 156 (Supplement C) (2017) 117–127, image and Video Understanding in Big Data.

[18] Dimitropoulos K, Barmpoutis P, Grammalidis N (2015). Spatio-temporal flame modelling and dynamic texture analysis for automatic video-based fire detection. IEEE Trans Circuit Syst Video Technol 25(2):339–351.

[19] Ghazal M, Vázquez C, Amer A (2012). Real-time vandalism detection by monitoring object activities. Multimed Tools Appl 58(3):585–611.

[20] Wiliem A, Madasu V, Boles W, Yarlagadda P (2012). A suspicious behaviour detection using a context space model for smart surveillance systems. Comput Vis Image Underst 116(2):194–209.

[21] P. Kenny, G. Boulianne, P. Dumouchel, Eigenvoice modeling with sparse training data, IEEE Transactions on Speech and Audio Processing 13(3) (2005) 345–354.

[22] Debaditya Roy, K. Sri Rama Murty, and C. Krishna Mohan (2018). Unsupervised Universal Attribute Modeling for Action Recognition, IEEE Transactions on Multimedia.

Biographies

Manisha Mudgal is a PHD scholar in Department of Computer Engineering at JC BOSE University of Science and Technology YMCA, Faridabad, India. She has done her M. Tech from M D University Haryana, India. She has successfully published 5 papers in Reputed National and International Journals. Her subjects of interest include Data Mining, Information Retrieval, and Machine Learning.
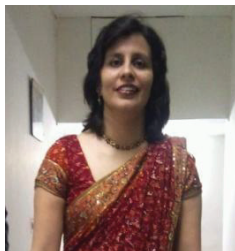
Deepika Punj is working as Assistant Professor in Department of Computer Engineering at JC BOSE University of Science and Technology YMCA, Faridabad, India. She has done Ph.D in Computer Engineering. She is having 14 years of experience in teaching. She has published more than 25 papers in Reputed National and International Journals. Her research interests include Data Mining, Deep Learning, Machine Learning and Internet Technologies.

Anuradha Pillai is an Associate Professor in the Department of Computer Engineering, JC Bose University of Science and Technology, YMCA, Faridabad, Haryana, India. She received Ph.D. in Computer Engineering from Maharishi Dayanand University, Rohtak. She published more than 60 papers in reputed international journals and successfully guided 4 PhD students. Her subjects of interest include Data Mining, Information Retrieval, Hidden web, Web Mining and Social Networks.