CrossMark

Multiple feature set with feature selection for anomaly search in videos using hybrid classification

A. Srinivasan[1] · **V. K. Gnanavel**[1]

Abstract An examination of abnormal activities in video scenes is a very difficult task in computer vision community. An efficient anomaly detection technique to detect anomalies in crowded scenes is presented in this paper. It uses Multiple Feature Set (MFS) to represent a piece of rectangular region of predefined size in a video frame called as patch with Hybrid Classification (HC) using Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) classifiers for anomaly detection. The MFS contains a combination of the following types of features; gray intensity values, gradient edge features and texture energy map. The predominant features are selected from MFS by using a model of t-test feature selection method and are classified by HC model made up of GMM and SVM classifiers. The UCSD video clip database is used for performance analysis of MFS-HC system and compared with other approaches. Results show that MFS-HC provides better results than other approaches.

Keywords Anomaly detection · Feature selection · Multiple features · Hybrid classification · GMM · SVM

1 Introduction

The accurate detection of abnormal events from crowded scenes is required for video surveillance. An anomaly detection system aims in identifying an anomalous event, with a low likelihood occurrence in the surveillance videos. Thida et al. [16] discussed a method of spatio-temporal Laplacian Eigen map model for detection of anomaly activities in videos. The detection is done by monitoring the variations occurred in spatio-temporal local motions. The usual behavior of the crowd is characterized by using the model constructed with the different

✉ A. Srinivasan
    asrini30@gmail.com

1   Department of Computer Science and Engineering, Misrimal Navajee Munoth Jain Engineering
    College, Anna University, Chennai, Tamilnadu, India

activity representatives. The constructed model helps in detecting the anomalies occurred in both local and global contexts of the crowd region.

The global motion information modeling of optical flow for anomaly detection is discussed by Benabbas et al. [2]. It understands the dominant motion magnitude and orientations from the obtained information to detect the foremost motion pattern. Then, a segmentation algorithm is used to make the scenes appears at different patterns with same speed and motion direction. Leyva et al. [8] described wake motion descriptor for video anomaly detection. The motion patterns which never occur previously are used to find the anomalies. Also, the relative change in the size of an object in the scene is compensated by using the perspective grid.

Shreedarshan & Selvi [14] analyzed the crowd behavior for anomaly detection by using the estimated optical flow with adaptive swarm intelligence. At first, the optical flow is generated from the input images which contain information of background, foreground and higher image intensity region. Then the motions are observed by using the streak lines and the optical flows. Then they are analyzed with the help of particle swarm optimization for anomaly detection. Sparsity based classification for anomaly detection is described by Mo et al. [12]. In the sparsity model, the computational cost is reduced by using a low rank structure on the sparse vectors/matrices coefficients.

The anomaly detection approach of Xiao et al. [19] is capable of performing both in multiscale and real-time detection. A local coordinate factorization model is used to know whether the video volumes belong to the spatio-temporal, temporal and spatial anomalies. An approach for learning the video events at each pixel is discussed by Javan Roshtkhari & Levine [5]. Using the dominant behaviors in the videos, a codebook model is constructed for dominant temporal and spatial events independently.

A method of probabilistic framework to detect anomalies in video is discussed by Saligrama & Chen [13] which identify the local spatio temporal anomalies. Based on a set of optimal decision rules, it detects the anomalies even if they occur inside a small region. The optical flow measurements are used for unusual event detection by [1]. At fixed spatial locations, optical flow measurements are extracted to detect anomaly instead of tracking objects in the scene.

A Mixture of Probabilistic Principal Component Analysis (MPPCA) is discussed by Kim & Grauman [6] for anomaly detection. The optical flow patterns are trained by MPPCA and then modeled by space-time Markov random field. The localization and detection of an abnormal event in videos are obtained by a Social Force (SF) model by Mehran et al. [11]. The crowd behaviour is modeled by using the interaction forces of individuals. Then, the frames are classified by a bag of word technique into normal event or anomaly.

The anomalies present in the complex scenes are detected by a Mixture of Dynamic Textures (MDT) by Mahadevan et al. [9]. The joint models of appearance and dynamics are used to represent the crowd patterns in videos. Then, the patterns are learned by expectation and maximization algorithm. Structure Analysis (SA) based anomaly detection is discussed by Yuan et al. [20]. At first, it detects the pedestrian and then structural context descriptor is used to represent the individuals in the frame.

In this paper, an efficient anomaly search in videos is presented based on MFS-HC. The main contribution is the development of MFS-HC system for anomaly detection. A combination of features from raw pixels, gradient map and texture map is used. From the MFS, only selected features are given to HC (SVM and GMM) which classifies the given video frame as normal or anomaly. The rest of this paper is as follows. MFS-HC based system design for anomaly detection is discussed in section 2. In section 3, the results of MFS-HC approach are discussed and in section 4, conclusions based on the obtained results are given.

2 System design

The framework of Video Anomaly Detection (VAD) by MFS with HC scheme is shown in Fig. 1. The main objective of anomaly search in videos is to identify or recognize whether the given video frame contains anomaly or not. To achieve this, a three stage VAD system is designed which consists of preprocessing, feature extraction with feature selection and classification stages. In the first stage, the frames are extracted from the video and with the help of background subtraction the motion vectors are estimated by [15]. The identified motion regions are given as an input to the second stage which extracts the dominant features by the feature extraction algorithms. In this work, multiple features such as raw pixels, gradient map, and texture energy map are extracted and fused to form an initial feature vector. Then, an absolute t-test approach is applied on the initial feature vector in order to select the dominant features. The selected features from the video frames are given to the third stage which classifies the given video frame as normal or anomaly. The classification is performed by means of the HC scheme formed by GMM and SVM classifiers.

2.1 Preprocessing

Preprocessing is the first stage which improves the performance of any classification system. In MFS-HC system, the following preprocessing steps frame separation and motion estimation are performed. From the inputs video clip, video frames are separated and stored as an image. Let us consider, a video clip C consists of N number of video frames F, $C = \{F_1, F_2, F_3 \ldots \ldots$
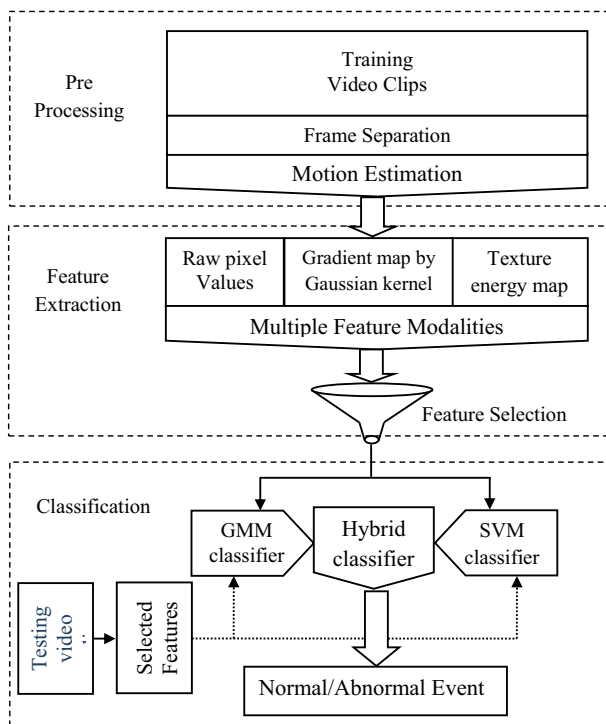

Fig. 1 Framework of VAD by MFS with HC

F_N}.Then the background subtraction is applied to estimate motion in the current frame F_i using the information in the previous frame F_{i-1}. The estimated motion regions are used for the extraction of the MFS features. The pseudo code of preprocessing steps is as follows:

Preprocessing Procedure;

Input:

Video clip C consists of N number of frames; $C = \{F_1, F_2, F_3 \ldots \ldots F_N\}$;

Output:

$Fp = \{F_{m1}, F_{m2}, F_{m3} \ldots \ldots F_{mN}\}$: Motion estimated video frames;

Process:

Loop for each frames $F_i \in FS$

$\qquad F_{i-1}$ is considered as a reference frame;

$\qquad F_i$ is the current frame;

$\qquad F_{mi} = F_i - F_{i-1}$; Motion estimation by background subtraction

$\qquad$ Store the estimated motion into F_p;

End Loop;

2.2 MFS extraction

In any classification or pattern recognition system, it is very important to extract the dominant features from the inputs for training the classifier module. In this work, three different types of features are extracted, and the combined feature space is called as MFS. The extracted features are gray intensity values, gradient edge features, and texture energy map.

2.2.1 Gray intensity values

In any gray scale image, the intensity of pixel values varies from 0 (black) to 255 (white) which is a scalar value. The pixel intensity variation will give some information that might help to detect anomalies in a frame. Thus, the gray intensity value is considered as one of the features for MFS-HC system.

2.2.2 Gradient edge features

The gradient edge features provide some useful information about the directional changes. It is obtained by the convolution between the image and a Gaussian Kernel. The following equations give the Gaussian kernel and gradient computation.

$$Gradient = I(m, n) \times \left[\left(\frac{\partial K_\sigma}{i} \right)^2 + \left(\frac{\partial K_\sigma}{j} \right)^2 \right] \tag{1}$$

$$K_\sigma(m, n) = \exp\left(-\frac{m^2 + n^2}{2^* \sigma^2} \right) \tag{2}$$

where I is an image and $I(m,n)$ is the intensity value of the image I at location (m,n). The center of gradient changes is (i,j) and σ is the standard deviation. From the gradient changes in Eq. 1, gradient magnitude is computed and used as one of the features for MFS-HC system.

$$Gradient_M = \max\left[\sqrt{Gradient_\sigma(i.j)}\right] \tag{3}$$

2.2.3 Texture energy map

Texture is one of the important features for many computer vision applications. A set of nine 5×5 masks [7] is used to extract texture energies that measure the variations in the fixed size window. These masks are generated from five 1D vectors which are shown in Fig. 2.

The product of a 1D vector and other vectors or itself produces sixteen convolution masks in 2D. These masks are applied to the motion estimated frame to find the texture energy map by using Eq. 4.

$$E_k[r, c] = \sum_{j=c-7}^{c+7} \sum_{i=r-7}^{r+7} |F_k[i,j]| \tag{4}$$

where $E_k[r, c]$ is the row and column of the input images $F_k[i,j]$ is the filtered images with the kth mask at pixel$[i,j]$ and C is the co-efficient. The result of application of Eq. 4 is also a full image corresponds to k^{th} mask. The sixteen energy maps are reduced to only nine energy maps by combing the symmetric pairs such as S5R5/R5S5, E5R5/R5E5, E5S5/S5E5, L5R5/R5L5, L5S5/S5 L5, and L5E5/E5L5 with its average. More information about Laws texture energy map is found by [7]. The pseudo code for the extraction of MFS is as follows:

Extraction of MFS Procedure;

Input:

$Fp = \{F_{m1}, F_{m2}, F_{m3} \ldots\ldots F_{mN}\}$: Motion estimated video frames;

Output:

$MFS = \{F_1, F_2, \ldots\ldots F_N\}$; Frame features

Process:

Loop for each frames $F_{mi} \in Fp$

      $f_1 = $ Gray intensity values of F_{mi}

      $f_2 = $ Gradient edge features of F_{mi} by Eqn. 3

      $f_3 = $ Texture energy map of F_{mi} by Eqn. 4

      $F_i = [f_1 \, f_2 \, f_3]$

      Store the estimated motion into MFS ;

End Loop;

2.3 Feature selection

The extracted features available in MFS will have some redundant features which may affect system accuracy. Thus, a feature selection approach is employed to reduce the
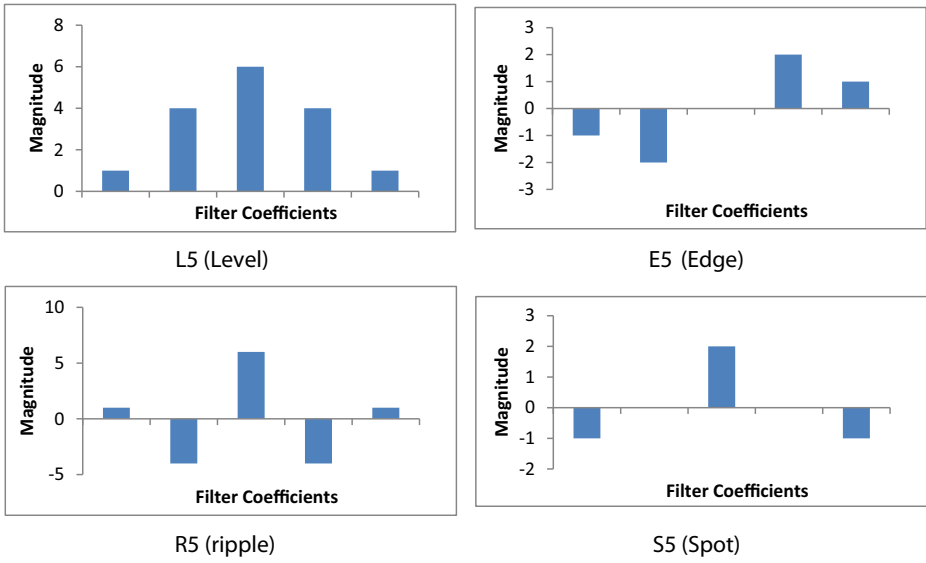
Fig. 2 1D vectors used to make masks in Laws texture

above mentioned problem occurred in VAD system. A statistical test is used to determine whether the features of normal event and anomalies are significantly different or not. As the MFS-HC system is a two class problem, a simple *t*-test is used by [18] based on the means of features of two groups; normal event and anomalies. It is given by

$$t(x) = \left(\overline{y}_1(x) - \overline{y}_2(x) \right) / \sqrt{\left(s_1^2(x)/n_1 + s_2^2(x)/n_2 \right)} \tag{5}$$

where $\overline{y}_1(x)$, $\overline{y}_2(x)$, $s_1^2(x)$ and $s_2^2(x)$ are the means and standard deviations of the two groups of samples; normal event and anomalies respectively.

Feature Selection Procedure;

Input:

$MFS_1 = \{F_1, F_2, \ldots F_N\}$; **Features of normal events**

$MFS_2 = \{F_1, F_2, \ldots F_N\}$; **Features of anomalies**

Output:

$RF = \{R_1, R_2, \ldots R_N\}$; **Rank features**

Process:

Loop for each features $F_i \in MFS_1 \& MFS_2$

      Find the means of F_i in $MFS_1 \& MFS_2$

      Find the standard deviations of F_i in $MFS_1 \& MFS_2$

      Apply Eqn. 5 to find the t_i value

End Loop;

Sort t_i in ascending order to find the top ranked features

2.4 Hybrid classification

The classification is achieved by using two most popular approaches; SVM and GMM classifier. The former one is a discriminate classifier and the later one is a generative model classifier. To achieve more accuracy and improve the VAD system performance, the result of both classifiers is fused together.

GMM classifier classifies the given event as normal or anomaly by computing the posterior probability using the testing features with training database. In general, an event is described by $G_m = \{\gamma_1, \gamma_2, \gamma_3, \ldots\ldots\gamma_M\}$ with M Gaussian models. Expectation and Maximization (EM) is employed to compute the M Gaussian models and its relative weights by [3]. The conditional probability is given by

$$p(T|\nabla) = \sum_{i=1}^{M} c_i \cdot \gamma_i(T) \tag{6}$$

where $\gamma_i(T)$ and c_i are the N-variate Gaussian function and mixture weights respectively. The best - fit event is computed using Bayes rule and EM algorithm for testing features by finding the posterior probability [3].

SVM classifier classifies the given event by constructing hyper plane which separate the features of normal event and anomalies with maximum margin. Let us consider a testing features t, the decision function O is defined by

$$t \in class\ 1 \text{ when } O(t) \geq +1 \text{ if } c_i = +1 \quad i = 1, 2, 3 \ldots n \tag{7}$$

$$t \in class2 \text{ when } O(t) \leq -1 \text{ if } c_i = -1 \quad i = 1, 2, 3 \ldots n \tag{8}$$

where n is the number of features. The decision function O is $O(x) = w.\ t + b$ where w and b are the weight and bias value respectively. To make the computation on original data, the dot product in the decision function is replaced by a kernel functions. More information about SVM classification can be obtained by (Panu [4]).

As SVM and GMM classifiers have their own advantages and demerits, an effective VAD system is designed by combing these classifiers using a weighted voting method. The weights for both classifiers are obtained by calculating the accuracy of selected training samples randomly.

3 Analysis of MFS-HC system

To evaluate the performance of MFS-HC system, the publically available dataset known as the UCSD [10, 17] database is used. It consists of many video clips of crowded scenes with

Table 1 UCSD database description

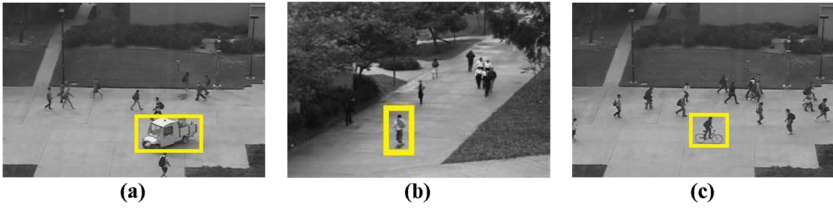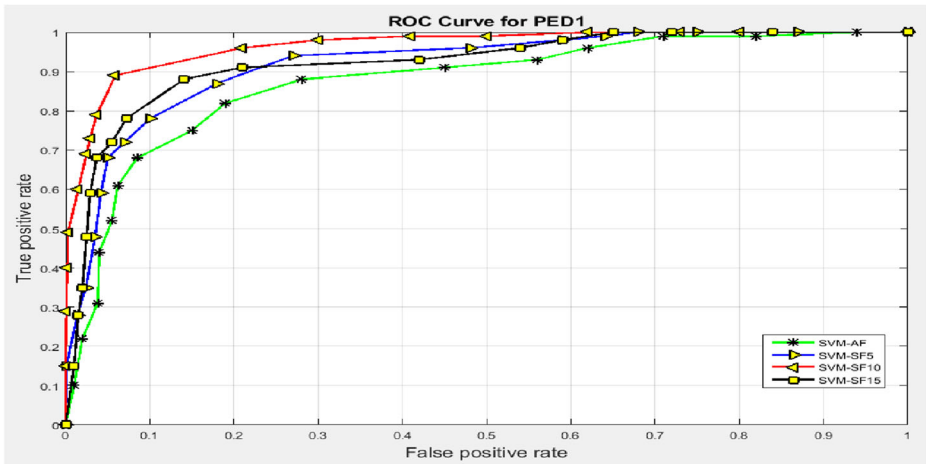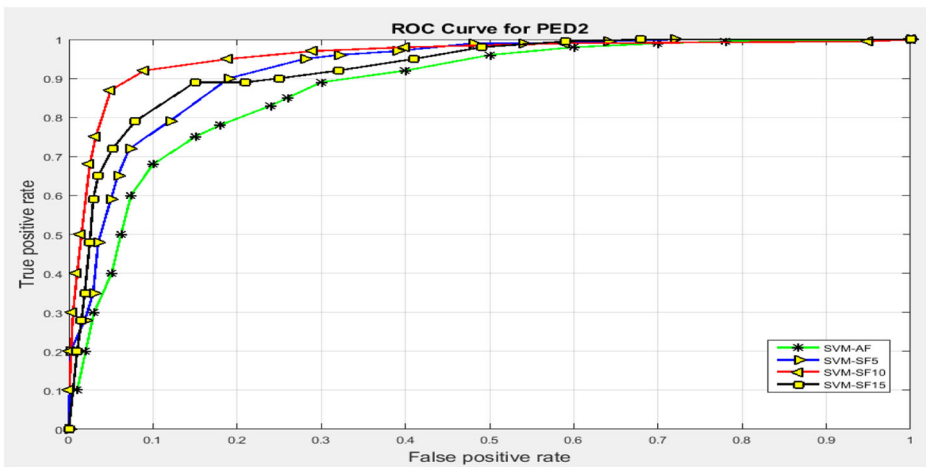| Details	#training videos	#testing videos	#frames per videos	#normal frames	#anomalies frames	Resolution
Ped1	34	36	200	≈5000	≈3400	158 × 238 pixels
Ped2	16	14	200	≈2566	≈2384	320 × 240 pixels

Fig. 3 Anomalies in UCSD database (a) small carts (b) skaters (c) bikers

varying crowd densities. The video clips are used for evaluating the performance of the MFS-HC system. The video footage recorded from every scene was split into large number of clips of around 200 frames. Some important information about the UCSD database is stated in Table 1.


(a)

(b)

Fig. 4 Comparison of ROCs with different SF (a) Ped1 (b) Ped2

All frames in the training video sets of UCSD database are of pedestrians only. Unlike training, the testing videos have abnormal events either in the form of non-pedestrian entries or abnormal movement patterns of pedestrians. The common abnormal events in the testing frames are bikers, skaters, and carts. Ped1 database consists of 34 training and 36 testing video clips whereas Ped2 consists of 16 and 14 video clips respectively. Ped2 database videos have good resolution than Ped1 database videos. The number of frames per second in both databases is 200. The presence of anomalies and its regions are provided in the ground truth information for each clip. Figure 3 shows some anomalies in the testing video clips of UCSD database.

The MFS-HC is applied on the UCSD database to identify the anomalies present in it. AUC is the performance metric used for the analysis of MFS-HC system with the following definition of False Positive Rate (FPR) and True Positive Rate (TPR). The former one is the percentage of anomalies that are incorrectly classified as a normal event, and the later one is defined as the
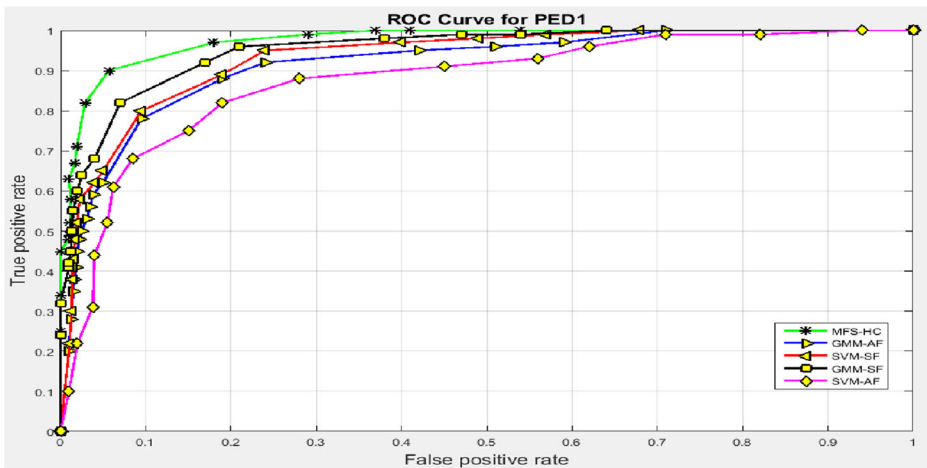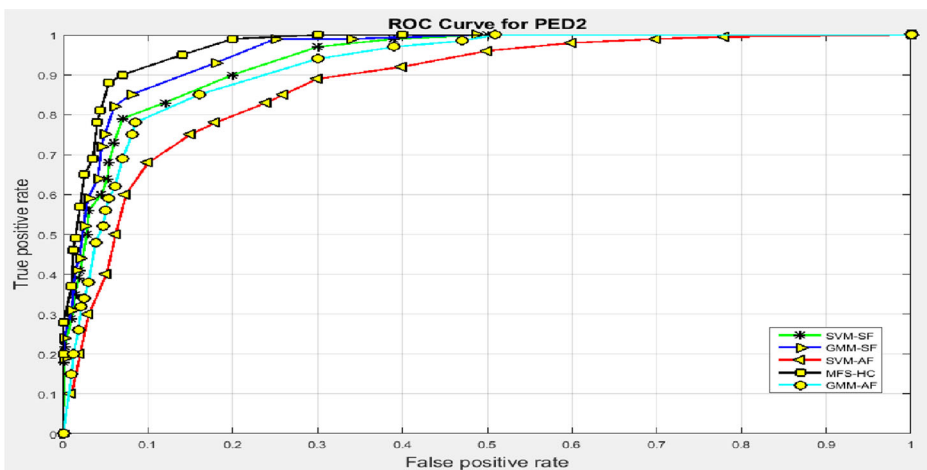

Fig. 5 Comparison of ROCs obtained by SVM, GMM and HC with AF and SF (a) Ped1 (b) Ped2

percentage of anomalies that are correctly classified as anomaly. AUC is measured from the Receiver Operating Characteristics (ROC) curve which is drawn between FPR and TPR.

The performance of MFS features is initially tested with SVM classifier with All Features (AF) and predefined percentage (in multiple of 5) of Selected Features (SF) by using t-test. Figure 4 shows the comparison of ROCs with different SF for Ped1 database.

From the comparison of ROCs with different SFs in Fig. 4, it is observed that better performance is achieved by MFS system while using SF values in both Ped1 and Ped2 data's. Also, it is noted that 10% of SF (SF10) provides better result than SF5 and SF15. The performance of MFS system with SVM classifier decreases while increasing the percentage of SF. This is due to that the selected features in SF15 are unable to differentiate the anomalies from normal event. Hence, SF10 is chosen as best percentage of features, and throughout the analysis in this paper, SF10 is used as features for anomaly detection.
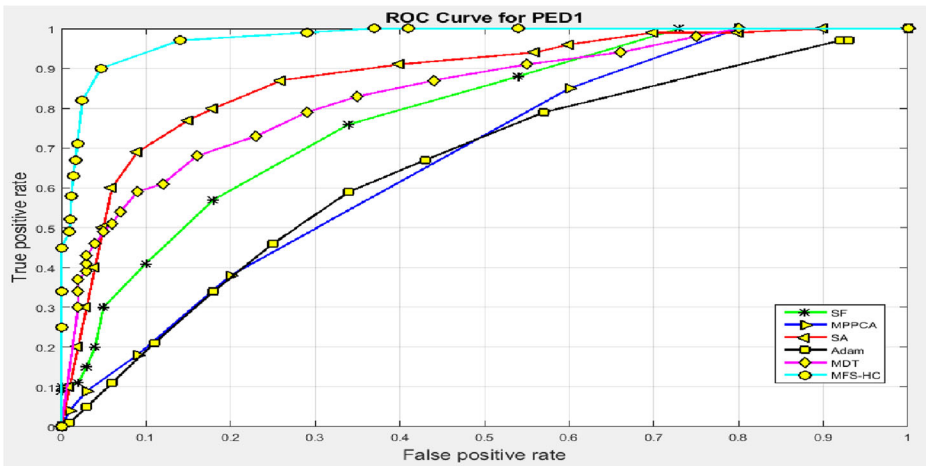
(a)
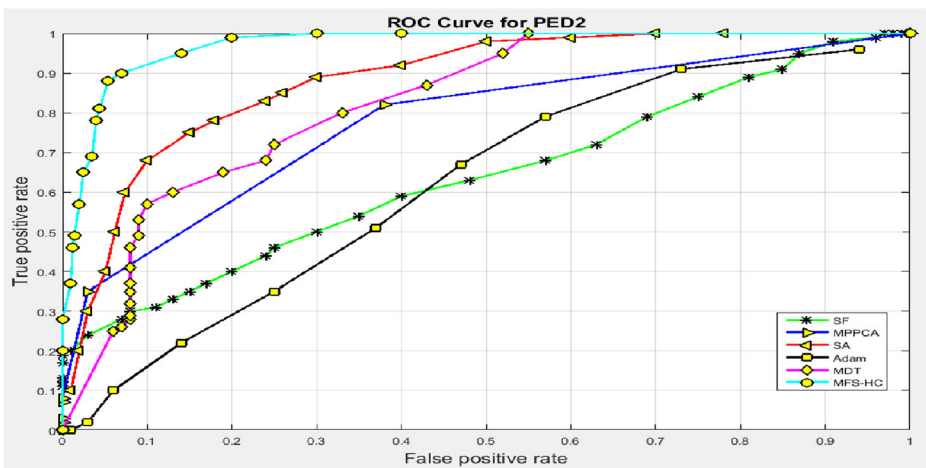
(b)

Fig. 6 Comparison of ROCs of different techniques (a) Ped1 (b) Ped2

In order to further analyze MFC features, a probability model based classifier GMM is combined with SVM classifier. It calculates the posterior probability for the classification of anomaly detection. A mixture of 16 Gaussian models is used for performance evaluation. Figure 5 shows the comparison of ROCs obtained from SVM and GMM with AF and SF. It includes the ROCs of MFS-HC where the decision is made by hybrid the outputs of SVM and GMM classifier.

It is observed from Fig. 5 that the SVM-HC provides better performance than their individual classifier; SVM and GMM performance for both Ped1 and Ped2 databases. Also, the performance of GMM classifier is superior to SVM classifier. The reason for lesser performance of SVM than GMM is due to its weakness on large training datasets. As the SF reduces the training features, the performance of SVM is increased as well in GMM. It is observed from the Fig. 4 that the TPR of MFS-HC for Ped1 and Ped2 database are very high in comparison with GMM and SVM classifier with SF and AF. The TPR of MFS-HC is 0.928 at 0.1 FPR. For the same FPR, the obtained TPR of GMM and SVM with SF are 0.852 and 0.805 respectively. Similarly, the TPR of MFS-HC for Ped2 database is 0.93 at 0.1 FPR which is higher than all other combinations. In order to validate the performance of MFS-HC system, a comparison is made with the following techniques; optical flow measurement (Adam 2008), MPPCA [6], SF [11], MDT [9] and SA [20]. Figure 6 shows the comparison of ROCs of MFS-HC with different techniques in the literature for Ped1 and Ped2 dataset.

It is clearly observed from Fig. 6 that the MFS-HC system outperforms all as the ROCs of MFS-HC system covers more area than other approaches such as optical flow measurement (Adam 2008), MPPCA [6], SF [11], MDT [9] and SA [20]. Among the other approaches, SA provides better result. The TPR of MFS-HC approach is 0.228 is higher than SA for Ped1 database and 0.25 for Ped2 database. The MFS-HC system takes about 8 s to test a frame in UCSD dataset on Windows platform with CPU speed of 3 GHz and RAM size of 2 GB.

4 Conclusion

A video surveillance system for the detection of anomaly events in a crowed video scene is discussed in this paper. It uses patch based extraction of MFS for the motion estimated frame by background subtraction. A feature selection model (t-test) is used to select the dominant features from the MFS. Then, HC module is used for the detection of an anomaly in the given frame. The MFS-HC system is tested by using the UCSD video clips database of the crowded scenes. The TPR of MFS-HC for Ped1 and Ped2 database are 0.928 and 0.93 at 0.1 FPR which outperforms all approaches. In future, real-time monitoring can be achieved through code optimization with graphics processing unit acceleration.

References

1. 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
2. Benabbas Y, Ihaddadene N, Djeraba C (2011) Motion pattern extraction and event detection for automatic visual surveillance. EURASIP Journal on Image and Video Processing 1:1–15
3. Bishop CM (2006) Pattern recognition and machine learning, Springer, chapter 9, vol. 1, pp. 435

4.   Erasto P (2001) Support Vector Machines - Backgrounds and Practice, Academic Dissertation for the Degree of Licentiate of Philosophy, Rolf Nevanlinna Institute
5.   Javan Roshtkhari M, Levine MD (2013) Online dominant and anomalous behavior detection in videos, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2611–2618
6.   Kim J, Grauman K (2008) Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates, IEEE Conference on Computer Vision and Pattern Recognition, pp. 2921–2928
7.   Laws KI (1980) Rapid texture identification. International Society for Optics and Photonics In Image processing for missile guidance 238:376–382
8.   Leyva R, Sanchez V, Li CT (2014) Video anomaly detection based on wake motion descriptors and perspective grids, IEEE International Workshop on Information Forensics and Security, pp. 209–214
9.   Mahadevan V, Li W, Bhalodia V, Vasconcelos N (2009) Anomaly detection in crowded scenes, IEEE Conference on Computer Vision and Pattern Recognition, pp. 1975–1981
10. Mahadevan V, Li W, Bhalodia V, Vasconcelos N (2010) Anomaly detection in crowded scenes, IEEE Conference on Computer Vision and Pattern Recognition, pp. 1975–1981
11. Mehran R, Oyama A, Shah M (2009) Abnormal crowd behavior detection using social force model, IEEE Conference on Computer Vision and Pattern Recognition, pp. 935–942
12. Mo X, Monga V, Baia R, Fan Z, Burry A (2014) Low rank sparsity prior for robust video anomaly detection, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1285–1289
13. Saligrama V, Chen Z (2012) Video anomaly detection based on local statistical aggregates, IEEE Conference on Computer Vision and Pattern Recognition, pp. 2112–2119
14. Shreedarshan K, Selvi SS (2016) An Adaptive swarm optimization technique for anomaly detection in crowded scene, IEEE International Conference on Circuits, Controls, Communications and Computing, pp. 4–6
15. Stauffer C, Grimson W (2000) Learning patterns of activity using real-time tracking. IEEE Trans Pattern Anal Mach Intell 22(8):747–758
16. Thida M, Eng HL, Remagnino P (2013) Laplacian eigenmap with temporal constraints for local abnormality detection in crowded scenes. IEEE Transactions on Cybernetics 43(6):2147–2156
17. UCSD database: http://www.svcl.ucsd.edu/projects/anomaly/
18. Wei Z, Xuena W, Yeming M, Manlong R, James G, John SK (2003) Detection of cancer-specific markers amid massive mass spectral data. Proc Natl Acad Sci 100(25):4666–14671
19. Xiao T, Zhang C, Zha H, Wei F. (2014) Anomaly detection via local coordinate factorization and spatio-temporal pyramid, In Asian Conference on Computer Vision, pp. 66–82, Springer, Cham
20. Yuan Y, Fang J, Wang Q (2015) Online anomaly detection in crowd scenes via structure analysis. IEEE transactions on cybernetics 45(3):548–561

Srinivasan Arulanandam earned his M.E. and Ph.D.degrees in Computer Science and Engineering from Madras Institute of Technology, Anna University, Chennai, before proceeding to complete his post-doctoral studies at NTU, Singapore. With 24 years of teaching and research experience in the fields and one year of industrial experience. He has published and presented over 56 research papers, One Book work in Pearson Education in Cloud Computing and one Indian Patent. He is currently working as a Professor and Head in Department of Information Technology, Misrimal Navajee Munoth Jain Engineering College, Anna University, Chennai. His fields of interests are Digital Image Processing, Distributed Computing. He is a member of the IEEE and ACM and a life member of CSI and ISTE.

Gnanavel Krishnan is currently pursuing the Ph.D. degree in the Faculty of Information and Communication Engineering. He received the B.E degree in Electrical and Electronics Engineering from University of Madras, Chennai and M.Tech degree in Computer Science and Engineering from Dr. M.G.R Deemed University, Chennai. He has 20 years of experience in teaching and research. He has published 6 research articles in National and International journal and conferences. His current research interests include computational methods, algorithms and applications in the field of image processing.