Mahalanobis Distance for Unusual Event Detection

Md. Haris Uddin Sharif, Sahin Uyaver, Md. Haidar Sharif

***Abstract*—In this paper, we address a simple but effective approach to detect and locate unusual events in videos. It does not depend on individual subject tracking. It is based on the statistical treatments of the spatiotemporal information (STI) of a set of points of interest within a region of interest over time. The achieved STI is clustered and Mahalanobis distances are calculated. Member and non-member groups, are resulted with 3σ rule, help to estimate cluster distance. Each frame corresponds to either usual or unusual frame groups. Cluster distances are summed up to get an effective distance, which is then normalized and fitted with polynomial to know its correspondence. If it corresponds to unusual group, then its region of the highest cluster distance indicates the most visual attended part and so on. Reported experimental results represent its performance.**

***Index Terms*—Unusual event, Cluster distance, Mahalanobis distance, χ^2 distribution, 3σ rule**

I. Introduction

Event detection in surveillance videos is an essential task for both private and public places. As large amount of video surveillance data makes it a backbreaking task for people to keep watching and finding interesting events, an automatic surveillance system is firmly requisite for the security management teams. Video event can be defined to be an observable action or change of state in a video stream that would be important for the security management team. Events would vary greatly in duration and would start from two frames to longer duration events that can exceed the bounds of the excerpt. Some events occur frequently or usually e.g., in the airport people are putting objects, getting objects, meeting and discussing, splitting up, and etc. Conversely, some events go off suddenly or unusually e.g., in the airport a person is running, falling on the escalator, going to the forbidden area, and etc. Both types of event detection in video surveillance is an important task for public security and safety in areas e.g., airports, malls, banks, metros, pedestrian subways, stations, city centers, hospitals, hotels, schools, concerts, parking places, sporting events, political events, and mass meetings. In spite of the concerted effort of computer vision research community, intelligent surveillance systems have not yet attained the desirable level of applicability and robustness. This is mainly due to the algorithmic assumptions as well as the huge amount of video data analysis.

There are some works [6], [26], [17], [11], [27], [28], [3], [2], [10], [4], [1], [9], [18] which detect unusual events or activities mainly in crowd flows. A system for automatically learning motion patterns for anomaly detection and behavior prediction based on a proposed algorithm for robustly tracking multiple objects can be seen in [6]. Authors in [26] detected events which have never occurred or occur so rarely that they are not represented in the clustered activities. The method includes robust tracking, based on probabilistic method for background subtraction. But the robust tracking method is not adapted to crowd scene, in which it is too complex to track objects. A spatial model to represent the routes in an image has been developed in [17]. But the system cannot differentiate between a person walking and a person lingering around, or between a running and a walking person. A method for detecting nonconforming trajectories of objects has been proposed in [11]. Authors in [27], [28] addressed the problem of modeling video behavior captured in surveillance videos for the applications of online normal behavior recognition and anomaly detection without any manual labeling of the training data set. The similarity between behavior patterns are measured based on modeling each pattern using a dynamic Bayesian network. The natural grouping of behavior patterns is discovered through a spectral clustering algorithm. The effectiveness and robustness of the approach has been tested on data sets collected from both indoor and outdoor surveillance scenarios. Authors in [3], [2] combined HMM, spectral clustering and principal component for detecting crowd emergency scenarios. The methods were experimented in simulated data only. Using a supervised SVM method, authors in [10] proposed an approach which makes a step toward generic and automatic detection of unusual events in terms of velocity and acceleration. The problem of detecting irregularities in visual data has been addressed by [4]. Authors in [1] proposed a holistic method, which is suitable to detect flow instabilities from the events, e.g., marathon, religious festival, etc. by identifying changes in the segmentation. An approach to detect abnormal situations in crowded scenes by analyzing the motion aspect without tracking subjects singly can be viewed as [9]. The approach was tested with real videos but it does not localize abnormality. In the framework of [18], an individual is subject to long-ranged forces and the dynamics follow the equation of motion, similar to Newtonian mechanics. A grid of particles is placed over the image and it is advected with the space-time average of optical flow. By treating the moving particles as individuals, their interaction forces are estimated using social force model. Interaction forces localize the abnormalities on the frame. Reporting results are good except some false alarms. Authors in [23] proposed a simple method based on Mahalanobis metric for eccentric event detection. But it does not localizes abnormality on the video frame.

In this paper, we have extended the idea of [23], which now detects and localizes unusual activities or events from video frames, based on the statistical treatments of the spatiotem-

poral information (STI) of a set of interest points within a region of interest over time. The obtained STI is clustered and *Mahalanobis distances* are calculated for each cluster. Those distances are then classified into two groups namely *member* and *non-member* based on *three sigma rule*. Distances in *non-member* group of a cluster are collected to represent its distance called *cluster distance* Σ_c. All cluster distances are summed up to get an effective distance for presenting each frame which is then normalized and poly-fitted to obtain a decision whether it falls into usual or unusual frame groups. If a frame belongs to unusual category, then its region which posses the highest cluster distance demonstrates the most visual attended part of the frame and so on. Our main goal is to introduce a holistic method (e.g., [1], [18], etc.) free from individual subject tracking to detect and localize unusual events from video frames. A common aspect of our work and those of [3], [2], [9] is that there is enough perturbation in the optical flow pattern in case of emergency. Our approach would be deemed as a further enhancement of these works with some senses. We profit from a different course by estimating and analyzing of the STI of video frames to detect unusual events directly without tracking subject singly. Thus it is effective for the high density mover scenes as well as low density scenes. It has further advantages: (i) it detects all events in videos where motion variations are important as compared to previous events; (ii) it does not expect efficient learning process and training data but would look for a prior cut-off; (iii) it localizes unusual activities on the video frames in a sorted order based on distance metrics of the clusters; (iv) it does not expect low-level change detection algorithms.

The rest of the paper is organized as follows: Section II illustrates the detailed implementation steps of the proposed approach; Section III demonstrates detection abilities of the proposed detector; and Section IV makes conclusion.

II. Implementation Steps

A. Region of interest (RoI) estimation

Irrespective of indoor and outdoor video surveillance, RoI makes the video processing faster. Based on applications and type of videos, RoI would extend from few parts of a video frame to the whole frame. We can estimate region of interest automatically by building a color histogram which is built from the accumulation of binary blobs of moving subjects, which were extracted following foreground segmentation method [12], [23]. RoI improves the quality of the results and reduces processing time which is an imperative factor for real-time applications.

B. Features estimation

There are a wide variety of approaches to detect corners or points of interest. Normally, a response function is calculated at every pixel in the image and an interest point is a local maxima pixel corresponded to the response function. A common response function is to detect points of interest. In 2D images, this is a spatial points of interest. In video, a spatiotemporal points of interest is a spatial points of interest in video images

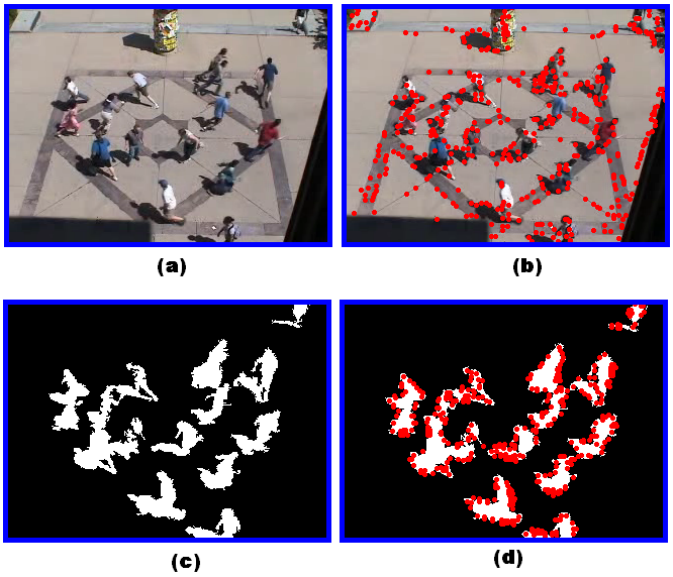

Fig. 1. Images belong to (a) and (c) are the camera view and foreground estimation, whereas (b) and (d) point to ponts of interest obtained by normal and modified Harris points of interest detectors for (a) and (c), respectively.

whose velocity vector is changing. The Harris corner detector [8] is a famous point of interest detector due to its strong invariance to rotation, scale, illumination variation, and image noise [22]. It is based on the local auto-correlation function of a signal, where the local auto-correlation function measures the local changes of the signal with patches shifted by a small amount in different directions. A discrete predecessor of the Harris detector was depicted by Moravec [19], where the discreteness refers to the shifting of the patches. The extended Harris corner detector [13] embodies a very simple and elegant algorithm. Though the extended Harris corner detection is simple and elegant, in practice, spatiotemporal points of interest are quite rare. This has been proved troubling in detection and recognition tasks observed by Lowe [15].

However, our goal is to get high contrast points of interest both in space and time. This identifies the points of interest which are along edges in a video frame and have displacement vector. Explicitly, we wish to get very dense features rather than sparse features resulting from the extended Harris detector. We also expect that the obtained features must contain points of interest with all different kinds of motions. With these aims, instead of points of interest in spatial positions, we extract points along edges with displacement vectors by simply replacing the second moment gradient matrix with gradient magnitudes of x-coordinate, y-coordinate and time-coordinate.

Other way around, there is a potential problem for camera positions and lighting conditions, which allow getting an extremely large number of points of interest that cannot be easily captured and tracked. For example, Fig. 1 (a) is the original video frame with moving subjects, if we apply Harris detector algorithm directly then the output contains lots of unwanted corners as shown in Fig. 1 (b). To avoid

this situation, we prefer to use a background and foreground estimation method before applying modified Harris point of interest detector. An estimated foreground can be derived after background estimation. Ideally, residual pixels obtained on applying background subtraction should represent foreground subjects. Foreground estimation is relatively easy in an indoor environment as the illumination conditions do not change significantly; while in outdoor environment that is much more complicated, as varying weather and sunlight (e.g., shadow of each subject in the Fig. 1 (a)) affect the correct detection of foreground. Some authors have adopted the adaptive Gaussian approach to model the behavior of a pixel [25], [7], [29]. Nevertheless, the background region of a video sequence often contains several moving objects. Consequently, rather than explicitly estimating the values of all pixels as one distribution, we would prefer to estimate the value of a pixel as a mixture of Gaussians [25], [7], [29]. Foreground pixels obtained on applying background subtraction are shown in Fig. 1 (c), in which noise, caused by shadows, which are the result of extremely strong lighting condition.

C. Features tracking

Once we define spatial points of interest, e.g., Fig. 1 (d), we track those points over the next frames using optical flow techniques to get spatiotemporal information. For this, we use the pyramidal implementation of Kanade-Lucas-Tomasi tracker [16], [24], [5]. Upon matching spatial points of interest between frames, we obtain spatiotemporal information, i.e., a $n \times 4$ matrix which is a function of time, explicitly speaking a set of vectors $M(l)(m)$ of n elements which variate in time, where $l = 1, 2, \ldots, 4$; $m = 1, 2, 3, \ldots, n$; i be any feature element in m; $x(1)(i) \mapsto x$ coordinate of a point of interest i; $x(2)(i) \mapsto y$ coordinate of the i; $x(3)(i) \mapsto$ displacement δ_i of the i from one frame to the next; $x(4)(i) \mapsto$ accurate moving direction α_i of the i.

If any feature i in the frame f with its coordinate $P(x_i, y_i)$ and its matched in the frame $f+1$ with coordinate $Q(x_i, y_i)$, it is easy to calculate the change of position (displacement) δ_i of the feature i using Euclidean metric as: $\delta_i = \sqrt{(P_{x_i} - Q_{x_i})^2 + (P_{y_i} - Q_{y_i})^2}$. The accurate moving direction α_i of the feature i can be calculated as: $\alpha_i = atan2(P_{y_i} - Q_{y_i}, P_{x_i} - Q_{x_i})$. Furthermore, we remove noisy features. Noise features are the isolated features which have very big angle and distance difference with their near neighbors due to tracking calculation errors. The resulting features are suitable for the clustering.

D. Features clustering

After suppression of noisy features, we apply *K-means* clustering method on $M(l)(m)$ to accommodate with clusters. Geometric clustering method, K-means, is a simple and fast method for partitioning data points into clusters, based on the work accomplished by Lloyd [14] (so-called Voronoi iteration). It is similar to the expectation-maximization algorithm for mixtures of Gaussians in that they both attempt to find the centers of natural clusters in the data. Image in Fig. 2 presents clustering accomplished by K-means where five clusters namely C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, and C_9 have been enclosed with rectangles. Each cluster contains features of an unknown distribution. We will estimate Mahalanobis distances for each cluster.

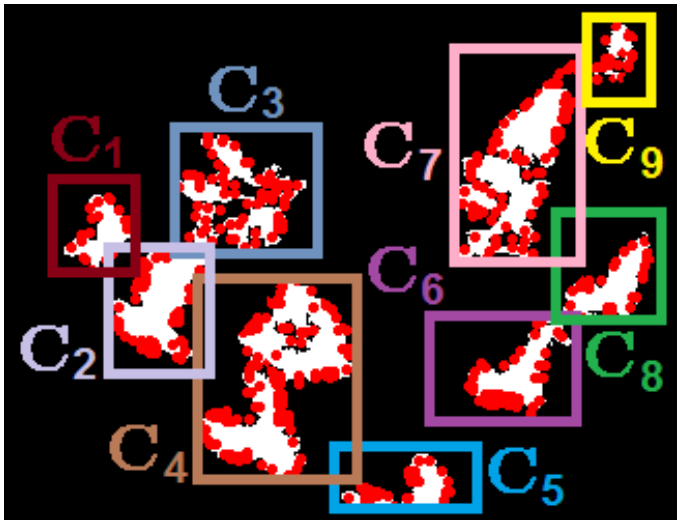

Fig. 2. On clustering features five clusters have been resulted. Each cluster contains features of unknown distribution.

E. Normalization of the data in cluster

A normalized value is a value that has been processed in a way that makes it possible to be efficiently compared against other values. On applying *K-means* clustering method on $M(l)(m)$, we may achieve several clusters. Consider that a cluster c contains spatiotemporal information as a matrix $C_c(i)(k)$ where $c \leq K$, $i \leq l$, and $k \leq m$. For each column of $C_c(j)(k)$, we calculate *mean* $\overline{x_j}$ and *standard deviation* σ_j. Subtracting the average $\overline{x_j}$ from each value in the columns of $x(j)(k)$, and then dividing by the standard deviation σ_j for that column in $x(j)(k)$ generated a new matrix $z_c(j)(k)$ as: $\overline{x_j} = \frac{1}{n}\sum_{k=1}^{n} x(j)(k)$, $\sigma_j = \sqrt{\frac{\sum (x(j)(k) - \overline{x_j})^2}{n-1}}$, $z_c(j)(k) = \frac{x(j)(k) - \overline{x_j}}{\sigma_j}$. All values in $z_c(j)(k)$ are *dimensionless* and *normalized*, hence the new form of $C_c(j)(k)$ yields:

$$Z_c(j)(k) = \begin{bmatrix} z_c(1)(1) & z_c(2)(1) & z_c(3)(1) & z_c(4)(1) \\ \vdots & \vdots & \vdots & \vdots \\ z_c(1)(i) & z_c(2)(i) & z_c(3)(i) & z_c(4)(i) \\ \vdots & \vdots & \vdots & \vdots \\ z_c(1)(n) & z_c(2)(n) & z_c(3)(n) & z_c(4)(n) \end{bmatrix}. \quad (1)$$

F. Calculation of correlation matrix

Correlation is dimensionless while covariation is in units obtained by multiplying the units of each variable. Using $Z_c(j)(k)$, scaling is done by following equations: $c_{pq} = \frac{S_{pq}}{S_p S_q}$, $S_{pq} = \frac{1}{n-1}\sum_{k=1}^{n}\left[z_{c_p}(k) z_{c_q}(k)\right]$, $S_h = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left[z_{c_h}(k)^2\right]}$ where $\{p, q\} \in j$ and $h \in \{p, q\}$. The correlation matrix M_c is formulated as:

$$M_c = \begin{bmatrix} 1 & c_{12} & c_{13} & c_{14} \\ c_{21} & 1 & c_{23} & c_{24} \\ c_{31} & c_{32} & 1 & c_{34} \\ c_{41} & c_{42} & c_{43} & 1 \end{bmatrix}. \quad (2)$$

G. Calculation of Mahalanobis distance $M_d(i)$

Mahalanobis distance is based on correlations between variables by which different patterns can be identified and analyzed. It is a useful way of determining similarity of an unknown sample set to a known one. It varies from Euclidean distance in that it considers the correlations of the data set and is scale-invariant. We calculate the Mahalanobis distance $M_d(i)$ for each row of the normalized matrix $Z_c(j)(k)$ by multiplying the row by the *inverted correlation matrix*, then multiplying the resulting vector by the transpose of the row of the $Z_c(j)(k)$, then dividing the obtained result by the degree of freedom of 4 (number of columns contained in $Z_c(j)(k)$), and finally grasping the square root of the latest result as:

$$M_d(i) = \sqrt{\left[\frac{z(1)(i)\ z(2)(i)\ z(3)(i)\ z(4)(i)}{4}\right][M_c]^{-1}\begin{bmatrix} z(1)(i) \\ z(2)(i) \\ z(3)(i) \\ z(4)(i) \end{bmatrix}}. \tag{3}$$

Samples with an equal $M_d(i)$ lie on an ellipsoid. The $M_d(i)$ is small for samples lying on or close to the principal axis of the ellipsoid. Samples further away from the principal axis have a much higher $M_d(i)$. The larger the $M_d(i)$ for a sample is, the more likely the sample is an outlier.

H. Classification of Mahalanobis distances

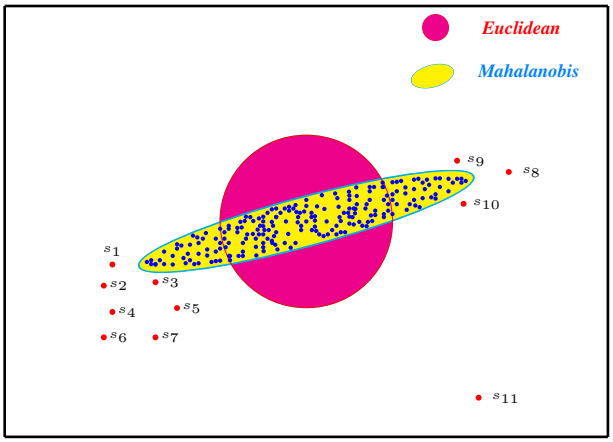

Fig. 3. Mahalanobis metric with respect to Euclidean metric. Red marked samples s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10} and outlier s_{11} belong to *non-member group*. The rest blue marked samples fall into *member group*.

In theory, Mahalanobis squared distance is distributed as a χ^2 statistic with degree of freedom equal to the number of independent variables in the analysis. The χ^2 distribution has only one parameter called the degree of freedom. The shape of a χ^2 distribution curve is skewed for very small degrees of freedom and it changes drastically as the degrees of freedom increase. Eventually, for large degrees of freedom, the χ^2 distribution curve looks like a normal distribution curve. Like all other continuous distribution curves, the total area under a χ^2 distribution curve is 1.0. The *three sigma rule* states that for a normal distribution, about 68%, 95%, 99.7% of the values lie within 1, 2, and 3 standard deviation of the mean, respectively. Clearly, almost all values lie within 3 standard deviations of the mean. Consequently, samples that have a squared Mahalanobis distance larger than 3 have a probability less than 0.01. These samples can be classified as members of *non-member group*. Samples those have squared Mahalanobis distances less than 3 are then classified as members of *member group*. The determination of the threshold depends on the application and the type of samples. In the proposed approach, we settle each $M_d(i)$ goes either *member group* or *non-member group*. Sample with a higher $M_d(i)$ than $\sqrt{3}$ is treated as *non-member group*, otherwise *member group*. *Member group* contains absolutely the samples of a normal event, whereas *non-member group* contains essentially samples of eccentric events (including outliers). Fig. 3 depicts, while Mahalanobis metric produces elliptical cluster where samples are well correlated, Euclidean metric produces circular subsets. The *non-member group* consists of samples s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}, and the outlier s_{11}, while the *member group* groups the rest samples. Presuming in any *non-member group* in a cluster c, which has S samples including outliers where $S \gg outliers$, we sum up their Mahalanobis distances to get a single distance Σ_c for representing the cluster as:

$$\Sigma_c = \sum_{i=1}^{S} M_d(i). \tag{4}$$

I. Estimation of effective distance E_d and normalization

The estimated Σ_c would play a vital role for the decision of abnormality. In the crowd scene in case of abnormal and/or emergencies situations physically there exist sufficient agitation and hence the positions, displacements, and directions of points of interest in the clustering are noticeably different between frames. In such situation, value of Σ_c will be higher as compared to abnormal case. Similarly, as compared to abnormal case, the clustering configurations of normal cases are almost similar between two consecutive frames. Consequently, the value of Σ_c will be small. Theoretically, clustering may be sensitive and a single value of Σ_c from a cluster would not reliable to define abnormality. To minimize this problem, we estimate an effective distance E_d for each frame. To get E_d, we sum up all Σ_c values obtained from all clusters in a video frame as:

$$E_d = \sum_{c=1}^{K} \Sigma_c. \tag{5}$$

This E_d can be used as judgement index for abnormality. For example, if any E_d obtained from a frame exceeds a predefined threshold T_d, then that frame belongs to the frame of abnormal events. But there are two notice problems which prevent it to use directly: (i) E_d is not normalized and (ii) a single E_d of a frame which exceeds T_d is not always a clear evidence of abnormal event frame.

If N_d be the normalized value of E_d, then N_d can be smoothly estimated as:

$$N_d = \frac{1}{2}\left[1 + erf(r)\right],\ erf(r) = \frac{2}{\sqrt{\pi}}\left[r - \frac{r^3}{3} + \frac{r^5}{10} - \frac{r^7}{42} + \dots\right] \tag{6}$$

where erf is an $error\ function$ and $r = \frac{log(E_d)-\mu}{\sigma\sqrt{2}}$. Placing congenial values of μ and σ in Eq. 6, we can explicitly estimate the value of N_d between 0 and 1.

J. Polynomial fitting of N_d values

Any discrete value of N_d which exceeds a predefined threshold T_d is not a clear evidence of unusual event frame, as it may frequently fear that at least one attribute (an outlier) may have been severely corrupted by a mistake or error (e.g., tracking calculation errors, etc.) which would lead an erroneous decision of the usual or unusual event frames. To minimize such problem, a polynomial fitting would be a good solution. Runges phenomenon [21] suggests that lower-order polynomials are normally to be preferred instead of augmenting the degree of the interpolation polynomial, even if some of the badness of this interpolation may be overcome by using Chebyshev polynomials instead of equidistant points. Accordingly, we can apply some lower degree (e.g., 5) of polynomial fitting on the obtained N_d values together with initially considering 5 frames. Consequently, a more reliable, workable, and much less erroneous measures (polynomial fitting distance P_d values) over the originally obtained E_d measures data for a decision of usual and unusual frame can be gained. It is important to define a threshold T_d to differentiate usual and unusual frames. We make a similitude measure between P_d and T_d to reach an explicit conclusion for each frame, i.e., a frame belongs to the frames of *unusual* events *if* $P_d > T_d$, otherwise frequently occurred events. When a frame falls into the frames of unusual events, we sort its cluster Σ_c values to localize unusual activities on the frame according to the sorting values. Normally, the highest Σ_c value posses the most visual attended part on the frame, and so on.

K. T_d Computing

The T_d value can be calculated from large videos that contain exclusively normal motions. The T_d depends on the controlled environment namely the distance of the camera to the scene (also escalator type and position), the orientation of the camera, the type and the position of the camera, lighting system, density of the crowd in working/vacation days/nights/weekends, etc. Deeming these facts, we have minimum one threshold per video stream. If we have N video streams, then we choose at least N thresholds. If the video stream changes, then T_d should be regenerated by means of:

$$T_d = arg \max_{t=1\ldots f} [\ P_d\ ]_t + \frac{1}{\sqrt{\pi}} \sum_{w=0}^{\infty} \left(\frac{P_d}{2w+1} \prod_{v=1}^{w} \frac{-{P_d}^2}{v} \right) \quad (7)$$

where f be the total number of frames. The second part of Eq. 7 is a Gauss error function, which is added for a good estimation of the threshold, is exactly 0.5 at ∞. On affixing the order of the series, Gauss error depends on the P_d. Other way around, for a fixed P_d the value of T_d can be varied by increasing or decreasing the order of the series. Thus, T_d can be affected on both P_d and order of the series.

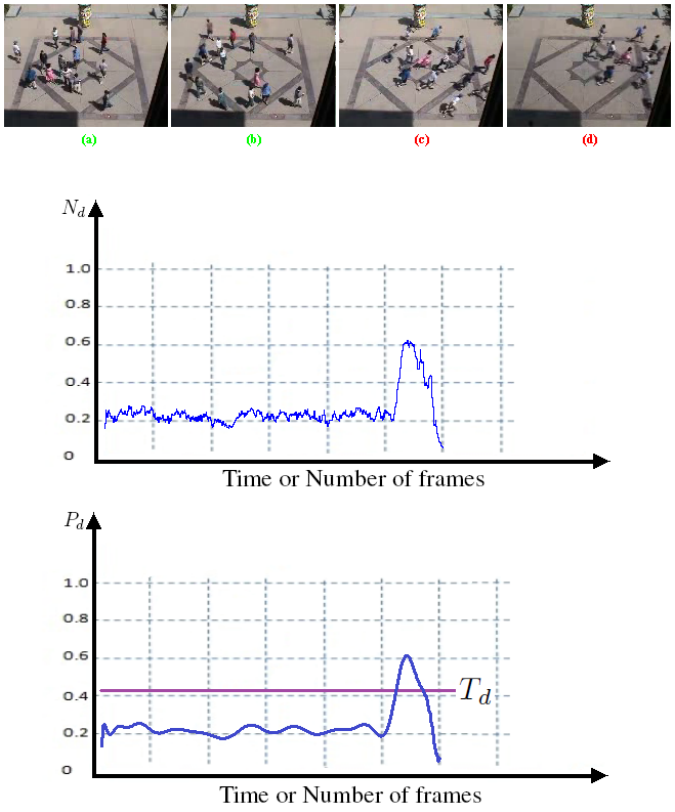

Fig. 4. Detection of unusual activities of crowd from a sample video of UMN dataset. Sample images (a), (b) and (c), (d) indicate the frames of before and during an escape panics circumstances, respectively. First and second graphs represent plotting of N_d and P_d values, consecutively.

III. Experimental Results and Discussion

A. Data set and parameters

To conduct experiments, we have mainly used *UMN data set* [20]. This publicly available data set comprises the videos of 11 different scenarios of an escape event in 3 different indoor and outdoor scenes. Number of frames is 7724 and frame size 320×240 pixels. Each video starts with normal behavior and ends with abnormal behavior. Parameters of μ and σ have been chosen as 0 and 5, respectively. Number of feature points $n = 2000$ and K-means $K = 10$.

B. Performance evaluation

Graphs in Fig. 4 are the output of one of the sample videos from UMN data set detected by the algorithm. Unusual event includes a sudden situation when the group of people start running. Gaussian like curves in time series demonstrate the abnormal events when those groups of people are trying to leave their places with atypical motions (e.g., a sudden run).

We have mainly tested our approach with *UMN data set* [20] and the performance comparison with Sharif et al. [23] has been listed on the Table I. Actually, our approach is an extended work of Sharif et al. [23], the extension includes good feature selection, clustering of the obtained spatiotemporal information, and polynomial fitting. The work of Sharif et

TABLE I
Performance evaluation and comparison

Approaches	Effectiveness	
	Recall Rate	*Precision rate*
Sharif et al. [23]	**82%**	**77%**
Proposed	**89%**	**86%**

al. [23] does not localize the unusual activities on the frame, whereas our approach both detects and localizes unusual activities on the frames. For the same video, the recall and precision rates are better for our approach than those of Sharif et al. [23]. However, both approaches are based on optical flow and do not concern occlusion, which has limited the performance of the algorithms. One possible solution to minimize the detection problem of unusual events which occur with occlusion is to use multiple cameras. Multiple cameras can provide different viewpoints of a region of interest. The engagement of multiple cameras will help to analyze many regions of interests which would be occluded by a single camera. All of our experiments have been conducted on videos obtained from single fixed camera, it would be interesting to test the approach with moving single camera data sets and future work would take into account this aspect as well.

In addition, our proposed approach has few important differences from few others recent and closely related body of works, e.g., Andrade et al. [2], [3]. Like our approach, in the work of Andrade et al. [2], [3] crowd behavior has been characterized at a global level by using the optical flow of the video sequences. Unlike our approach, during the learning stage, a reduced order representation of the optical flow was generated by performing PCA on the flow vectors. Afterwards, top few eigenvectors were used as the representative features and spectral clustering was performed to identify the number of distinct motion patterns present in the video. The features in the clustered motion segments were used to train different HMMs which were then used for event detection in crowds. However, the method does not localize the unusual activities on the video frames. Furthermore, the method was only tested by data obtained from simulation. Model building for simulation is often costly and time-consuming.

IV. Conclusion

We proposed a simple but effective approach to detect and localize unusual activities in videos without the motive to track subjects individually. The approach is based on the statistical treatments of the spatiotemporal information of a set of interest points within a region of interest over time. It does not need low-level change detection algorithms. It uses region of interest to improve the quality of results and to reduce processing time. The reported results show its effectiveness.

References

[1] S. Ali and M. Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. pages 1–6, 2007. 1, 2

[2] E. L. Andrade, S. Blunsden, and R. B. Fisher. Hidden markov models for optical flow analysis in crowds. In *International Conference on Pattern Recognition (ICPR)*, pages 460–463, 2006. 1, 2, 6

[3] E. L. Andrade, S. Blunsden, and R. B. Fisher. Modelling crowd scenes for event detection. In *International Conference on Pattern Recognition (ICPR)*, pages 175–178, 2006. 1, 2, 6

[4] O. Boiman and M. Irani. Detecting irregularities in images and in video. *International Journal of Computer Vision*, 74(1):17–31, 2007. 1

[5] J. Y. Bouguet. Pyramidal implementation of the lucas kanade feature traker. In *A part of OpenCV Documentation, Intel Corporation, Microprocessor Research Labs*, 2000. 3

[6] W. H. et al. A system for learning statistical motion patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1450 –1464, 2006. 1

[7] R. Hammond and R. Mohr. Mixtire densities for video objects recognition. In *International Conference on Pattern Recognition (ICPR)*, pages 71–75, 2000. 3

[8] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–152, 1988. 2

[9] N. Ihaddadene and C. Djeraba. Real-time crowd motion analysis. In *International Conference on Pattern Recognition (ICPR)*, pages 1–4, 2008. 1, 2

[10] I. Ivanov, F. Dufaux, T. M. Ha, and T. Ebrahimi. Towards generic detection of unusual events in video surveillance. In *International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 61–66, 2009. 1

[11] I. N. Junejo, O. Javed, and M. Shah. Multi feature path modeling for video surveillance. In *International Conference on Pattern Recognition (ICPR)*, volume 2, pages 716–719, 2004. 1

[12] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis. Real-time foreground-background segmentation using codebook model. *Real-Time Imaging*, 11(3):172–185, 2005. 2

[13] I. Laptev. On space-time interest points. *International Journal of Computer Vision*, 64(3):107123, 2005. 2

[14] S. P. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–136, 1982. 3

[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 2

[16] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 674–679, 1981. 3

[17] D. Makris and T. J. Ellis. Path detection in video surveillance. *Image and Vision Computing Journal*, 20:895–903, 2002. 1

[18] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In *Computer Vision and Pattern Recognition (CVPR)*, pages 935–942, 2009. 1, 2

[19] H. P. Moravec. Obstacle avoidance and navigation in the real world by a seeing robot rover. In *Technical Report CMU-RI-TR-80-03, Robotics Institute, Carnegie Mellon University & doct. diss., Stanf. University*, 1980. 2

[20] U. of Minnesota. *Usual and Unusual crowd activity dataset of the University of Minnesota, Minnesota, USA*. Available from the URL of: http://mha.cs.umn.edu/movies/crowdactivity-all.avi, 2009. 5

[21] C. D. T. Runge. Ueber empirische funktionen und die interpolation zwischen äqui-distanten ordinaten. *Zeitschrift für Mathematik und Physik*, 46:224–243, 1901. 5

[22] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172, 2000. 2

[23] M. H. Sharif and C. Djeraba. A simple method for eccentric event espial using Mahalanobis metric. In *Proceedings of the Iberoamerican Conference on Pattern Recognition (CIARP)*, volume 5856 of *Lecture Notes in Computer Science*, pages 417–424. Springer, 2009. 1, 2, 5, 6

[24] J. Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition (CVPR)*, pages 593–600, 1994. 3

[25] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition (CVPR)*, pages 23–25, 1999. 3

[26] C. Stauffer and W. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, 2000. 1

[27] T. Xiang and S. Gong. Video behavior profiling and abnormality detection without manual labeling. In *International Conference on Computer Vision (ICCV)*, pages 1238–1245, 2005. 1

[28] T. Xiang and S. Gong. Video behavior profiling for anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):893–908, 2008. 1

[29] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *International Conference on Pattern Recognition (ICPR)*, pages 28–31, 2004. 3