

Human Crowd Behavior Analysis Based On Graph Modeling and Matching In Synoptic Video

B.Yogameena^{#1}, K.Sindhu Priya^{*2}^{#1}Department of ECE, Thiagarajar College of engineering, Madurai, Tamilnadu, India.^{*2}Department of ECE, Thiagarajar College of engineering, Madurai, Tamilnadu, India.

ABSTRACT— Huge video dataset captured by using various resources like surveillance cameras, webcams.etc. It is very time consuming to watch whole video manually. In this paper, Video synopsis is used to represent a short video while preserving the essential activities for a long video. In the existing methodology, usually a single moving object is splitted into a few small pieces in a continuous activity. For that Gaussian mixture model is used to detect compact foreground against their shadows. But in high-density crowds background subtraction fails due to occlusion. After detecting the occlusion, background is subtracted. Tracking is done by Centroid of the silhouette. For video synopsis more fluent tubes are generated by concatenation. Then each isolated region is considered as a vertex and a human crowd is thus modeled by a graph. After modeling the graph by using Delaunay triangulation and Convex-Hull, graph matching algorithm is developed to detect the problem of behavior analysis of human crowds. Experimental results obtained by using extensive dataset show that the proposed algorithm is effective in detecting occlusion and anomalous events using video synopsis.

KEYWORDS— Video Synopsis, Fluent Tube, Delaunay Triangulation, Occlusion, Graph Modeling

I.INTRODUCTION

To prevent criminal behaviors, now a day's video surveillance systems have become more and more popular in many public places like airports, train stations, bus stand, etc. People are usually the main objects of interest in surveillance tasks. In public places, people usually move towards some orientation(s) and thus form some crowd(s). A crowd is defined as two or more

people in a small spatial region or interacting with each other. The moving dynamics of a crowd contain rich information about human activities. If the motion trajectories of crowds are out of predictions, it usually means that some abnormal events have occurred. The events induced by the interactions within crowds are also extremely important since the abnormal behaviors of crowds would usually result in some massive damages. In the literature, Andrade *et al.* [1] used optical flows to describe crowd and to extract features from optical flows in an unsupervised manner. Cheriyyadat and Radke [2] present a survey of algorithms of computer vision that deal with crowds of people and review model-based crowd analysis algorithms, in which some type of human model is applied to tracking or segmentation. In crowded scenarios to detect typical motion patterns. Ryan *et al.* [3] proposed a scene independent approach that can count the number of people in crowded scenarios without any training. A "global scaling factor" is used to compensate for camera angle and distance, in each scene based on a reference person. Chan *et al.* [4] developed a privacy-preserving system to estimate the inhomogeneous crowd's attributes. The approach does not depend on feature tracking or object detection. Jiang *et al.* [5] employed the contextual anomaly concept for crowd analysis. An unsupervised approach is used to follow in their system, that automatically identify important contextual information from the crowd video and to detect the blobs corresponding to contextually anomalous behaviors. Mehran *et al.* [6] used social force model

to detect and localize unusual events in a crowd of people. To detect occlusion, Jiang et al. [7] employ integer linear programming, however the of targets number needs to be known. To overcome this limitation, Berclaz et al. [8] introduce sink locations and virtual source to initiate and terminate trajectories. Binary optimization problems are a common trait of these works that are usually solved to global optimality by relaxing them to linear programs (LPs). Most of the cameras work 24 hours a day. It is very time-consuming to watch whole video manually and difficult to find interesting activities in long videos. Video synopsis aims at producing a brief description of the original video while extracting key activities, which makes a full view of the video content quickly [9]. Video synopsis provides effective summary of videos in quicker time. Hence the contribution of the video synopsis is to adapt abnormal behavior analysis.

II.METHODOLOGY

First, the input video is taken and it has been checked for any occlusion is present. Because, in high-density crowds background subtraction fails due to occlusion. After occlusion detection, background is estimated using GMM and then foreground is detected using background subtraction. To reduce the length of the long video, video synopsis is used. Then the graph is to be modeled for the human crowd. For modeling the graph, Delaunay Triangulation is used to connect all the vertex of the connected component. Then the event can be detected by comparing the graph in subsequent frames.

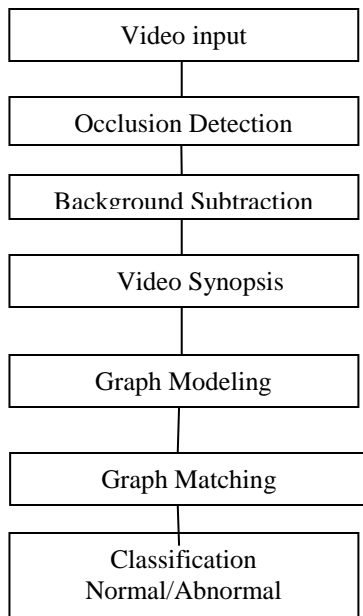


Fig 1 shows Flow diagram for human crowd classification

A. Detection of Occluded People

Tracking multiple objects in crowded scenes necessarily leads to the occlusion problem. Discussing

this problem and comparing the various capabilities, which cause occlusion problem in some frame work existing video tracking systems would be much easier. There are three different types of occlusion like: (1) usually in dense crowded scenes, targets frequently occlude each other causing inter-object occlusion (2) a target may move behind static objects like trees, rock or road signs, that are all examples of common scene occluders; (3) depending on the type of object, extensive articulations, orientation or deformations changes which leads to self occlusion. All three types of occlusion reduce – or completely suppress –evidence of the image for a target’s presence. In this paper, the focus is on the inter-object challenge occlusions. According to the situations where dynamic targets occlude each other, the main task is to overcome the difficulties that arise from the complex dependence between a target’s visibility and of several other targets trajectories that could potentially block the of sight line. Global target visibility is proposed to overcome this problem.

The total area of image is thus given as $Roi(x) dx$. To compute the relative area of target i which is occluded by target j,

$$\frac{1}{\int O_i(x)dx} \int O_i(x)O_j(x)dx \tag{1}$$

we assume here that target j is in target front i, then we will address the general case below. By using the relative area we define the target visibility as given in Eq. (2), then the visibility is not differentiable with respect to the object positions of X_i or X_j , which includes gradient-based optimization methods before processing Here we propose to use a Gaussian “indicator” function $N_i(x) := N(x; c_i; c_i)$ to address the issue. Like before, we compute the overlap area by integrating the two products of the indicator functions, where as Gaussians

$$Z_{ij} = \int N_i(X). N_j(X)dx \tag{3}$$

Besides differentiability, the Gaussians choice allows this integral to be determined in closed form. $Z_{ij} = N(c_i; c_j ;C_{ij})$ with $C_{ij} = c_i + c_j$. we denote it using the normalized Gaussian

$$V_{ij} = exp\left(-\frac{1}{2}[c_i - c_j]^T C_{ij}^{-1}[c_i - c_j]\right) \tag{4}$$

which is differentiable with respect to c_i and c_j and it has the desired property that $V_{ij} = 1$ when $c_i = c_j$. Due to the Gaussians symmetric property, it is considered that $V_{ij} = V_{ji}$.

B. Background Subtraction

To perform foreground extraction, Gaussian mixture models is used to represent the background [10]. The probability that an observed pixel, $x = (x, y)$, will have an intensity value R_t at time t , is estimated by K Gaussian distributions as follows,

$$P(R_t) = \sum_{l=1}^K \frac{\omega_{l,t}}{(2\pi)^{1/2}} e^{-\frac{1}{2}(R_t - \mu_l)^T \Sigma_l^{-1} (R_t - \mu_l)} \tag{5}$$

Where $\omega_{l,t}$ and μ_l are the weight and the mean of the l^{th} distribution, where Σ_l is the Gaussian's covariance matrix that is assumed to be diagonal. To update this model, each pixel value in a new video frame is processed to determine if they matches any of the existing Gaussian distributions at that pixel's location. If a match is confirmed for the l^{th} distribution at time t , then a Boolean variable, $M_{l,t}$, is set. At each time instant, each distribution for the weighting factor is updated according to

$$\omega_{l,t} = (1 - \alpha)\omega_{l,t-1} + \alpha(M_{l,t}) \tag{6}$$

where α is the learning rate that controls the speed of learning. The actual Gaussians at each pixel location are updated as follows

$$\mu_{l,t} = (1 - \beta)\mu_{l,t-1} + \beta R_t \tag{7}$$

In addition to GMM-based background subtraction, to isolate foreground blobs we use three-frame temporal differencing. Specifically, frame differencing is performed by using

$$\Delta_t = |I_t(x) - I_{t-1}(x)| \oplus |I_t(x) - I_{t+1}(x)| \tag{8}$$

By using the results of Eq. (9) are threshold to remove false positive pixels to obtain accurate foreground regions. Specifically, we use an adaptive threshold function $median(\Delta_t)$ to determine a moving target's pixels $Q_t(x, y)$,

$$Q_t(x, y) = \{\Delta_t(x, y) | I_t(x, y) \geq median(\Delta_t)\} \tag{9}$$

To obtain the final foreground regions, the background subtraction and temporal differencing results are combined by applying a binary "OR" operation to the foreground map derived by background subtraction. This combination ensures that moving foreground objects can be identified by using the result.

C. Tube Generation

By using GMM method, more compact foreground region can be detected to build fluent tubes for video synoptic. If one moving object is segmented into few small pieces, and then tube will be disconnected, it leads to sudden interruption of objects in the video synoptic, therefore, centroid based tracker algorithm [11] is used to concatenate tube fragments and generate fluent tubes. Where x_t and y_t are the hidden state and observation the data at time t and given various observation $y_{0:t} = \{y_0, \dots, y_t\}$ to filtered estimation of x_t is $p(x_t | y_{0:t})$

$$p(x_t | y_{0:t}) \propto p(y_t | x_t) \int p(x_t | x_{t-1}) p(x_{t-1} | y_{0:t-1}) dx_{t-1} \tag{10}$$

Where $p(x_t | x_{t-1})$ is the transition model determining how objects can move between frames and $p(y_t | x_t)$ is the observation model to represent the likelihood of objects. A set of M weighted particles $\{x_t^i\}_{i=1 \dots M}$ is used to determine the distribution of filtering $p(x_t | y_{0:t})$

$$p(x_t | y_{0:t}) \approx \sum_{i=1}^M w_t^i \delta(x_t - x_t^i) \tag{11}$$

The particles $\{x_t^i\}_{i=1 \dots M}$ which are taken from $p(x_t | x_{t-1})$ and the weight w_t^i is taken from the data likelihood $p(y_t | x_t^i)$. The prior distribution for the process of tracking in an image patch is based on the HSV histogram. The observation model is

$$p(y_t | x_t) \propto e^{-\lambda D^2(h_0 h_t(x_t))} \tag{12}$$

The tracker is used to concatenate the disconnected fragments, for a set of initial trajectory fragments.

III. MODELING HUMAN CROWDS

Each isolated region is considered as a vertex and a human crowd is thus modeled by a graph. After modeling the graph, graph matching algorithm is developed [12].

D. Delaunay Triangulation

Each connected component is represented by its centroid and then a set of vertices of a human crowd P . It can be denoted as $P = \{P1, P2, \dots, Pn | Pi = (Xi, Yi)\}$, where (Xi, Yi) is the coordinate of the i^{th} connected component. To construct a graph for a human crowd, two properties of graph theory is used. The first is the

uniqueness of the graph. It means one-to-one mapping that the same set of vertices would correspond to a single graph. According to this property, Delaunay triangulation is used to systematically connect the vertices of the connected component since the union of all simplices in the triangulation is the convex hull of the vertices. The second property is that the constructed graph should be the largest convex hull covering all the vertices. According to this property, convex hull is used to cover

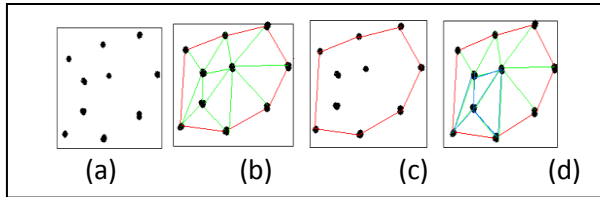


Fig. 2. (a) Black dots denote an individual set, (c) graph is denoted by a Triangles set. (d) Alternations person.(b)Largest convex hull covering the outer vertex after the process of triangulation with an extra vertex added.

the triangulation. With such property, the topology variation of the graph would explicitly show the global changes in the graph. Therefore, a Delaunay triangulation for a set P in the plane is a triangulation, say $DT(P)$, such that no point in P is inside the circum-circle of any triangle in $DT(P)$ [12]. When an additional vertex is seen in the graph, alternation of the triangulation would be limited around the extra vertex in the local area rather than re-triangulation through all vertices. The triangulation process is shown in Fig.2.

IV. GRAPH-BASED EVENT DETECTION

Then abnormal events are detected by the variation of graph topology by using adjacency matrix, sub-graph analysis, description of shape deformation, and global graph matching.

A. Adjacency Matrix of a Graph

If node i and node j are connected, $A_G(i,j)$ or $A_G(j,i)$ is nonzero; otherwise, $A_G(i,j)$ and $A_G(j,i)$ are zero. Adjacency matrix is used to identify the connectivity of the graph. The matrix $A_G(i,j)$ is the adjacency matrix of a graph G and is a symmetric (0,1)-matrix with zeros on the main diagonal. When the graph is connected the adjacency matrix of a graph is indecomposable precisely. Delaunay triangulation is constructed by connecting the graph $G(V, E)$ that none of V is disconnected.

Without generality loss, let G and H are the two constructed graphs with the relations $E(H) \subseteq E(G)$ and $V(H) \subseteq V(G)$. H is then a sub graph of G . If G and H satisfy the conditions $E(H) \subseteq E(G)$ and $V(H) \subseteq V(G)$, H is named a spanning sub graph of G . The sub graph induced by nonempty set S of vertices in G is that sub graph H with vertex-set S whose edge set consist of those edges of G that join two vertices in S ; it is denoted by $G(S)$. A sub graph H of G is induced if $H = \langle V(H) \rangle$.

1) Corollary: If H is an induced sub graph of G , if $\mu_1 \geq \mu_2 \geq \dots \geq \mu_m$ are the eigenvalues of the matrix related to the graph H and if $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ are the eigenvalues of G , then $\lambda_{i+n-m} \leq \mu_i \leq \lambda_i$, for $i=1,2,\dots,m$. Sub graph analysis is important since the graph at time t with slight changes of one or few vertices present additionally or absent would still be a sub graph of graph at time $t - 1$. Based on the theorems mentioned above, the similarity between consecutive graphs can be formulated as the difference between the eigenvalues of adjacency matrix A . If the two set of eigenvalues from consecutive graphs are of the same number and the same values, these two graphs are said to be graph isomorphism. If consecutive graphs satisfy the corollary, these two graphs are regarded as similar but not exactly the same. Since G_t and G_{t-1} do not satisfy the relation $E(G_t) \subseteq E(G_{t-1})$ the corollary cannot be used to determine if they are of the sub graph relationship.

C. Description of Shape Deformation

A shape deformation of graph is employed to describe the graph variation of local changes by using triangle matching method [13].

1) Triangle Matching Method: After processing the Delaunay triangulation, to obtain a triangle set Δ_t of $G_t(V, E)$, where $\Delta_t = \{\delta_{t1}, \delta_{t2}, \dots, \delta_{tk}, \delta_{ti}, (v_{i1}, v_{i2}, v_{i3}), i = 1 \sim k\}$, k is the number of triangles of $G_t(V, E)$ and $v_{ij} \leq V(G_t)$, $j = 1 \sim 3$ are the three vertices of the triangle. Let Δ_{t-1} and Δ_t be the triangle set of graph $G_{t-1}(V, E)$ and $G_t(V, E)$, respectively. The centroid of the triangle Δ_{t-1} and Δ_t is written in the form of

$$M_i^{t-1} = \frac{v_{i1}^{t-1} + v_{i2}^{t-1} + v_{i3}^{t-1}}{3}, i = 1,2,3, \dots, p, \dots, k_{t-1} \quad (13)$$

$$M_j^t = \frac{v_{j1}^t + v_{j2}^t + v_{j3}^t}{3}, j = 1,2,3, \dots, q, \dots, k_t \quad (14)$$

The triangle-pair δ_p^{t-1} and δ_q^t that is of the shortest distance can thus be defined as

$$(\delta_p^{t-1}, \delta_q^t) = \operatorname{argmin} (d(M_i^{t-1}, M_j^t)) \quad (15)$$

Therefore, to find the similarity between δ_p^{t-1} and δ_q^t is

$$S_{p,q}^t = \left(1 - \frac{|e_{p12}^{t-1} - e_{q12}^t|}{\min(e_{p12}^{t-1}, e_{q12}^t)}\right) \times \left(1 - \frac{|e_{p23}^{t-1} - e_{q23}^t|}{\min(e_{p23}^{t-1}, e_{q23}^t)}\right) \times \left(1 - \frac{|e_{p31}^{t-1} - e_{q31}^t|}{\min(e_{p31}^{t-1}, e_{q31}^t)}\right) \quad (16)$$

Where $(e_{p12}^{t-1}, e_{p23}^{t-1}, e_{p31}^{t-1})$ and $(e_{q12}^t, e_{q23}^t, e_{q31}^t)$ are the edges set for the triangles δ_{i-1p} and δ_{iq} , respectively. By using this equation if we get the result as 1 then we identifies that there is no changes in the triangles. If we get the value less the 1 then there are some changes in the triangles it is represented by green triangles.

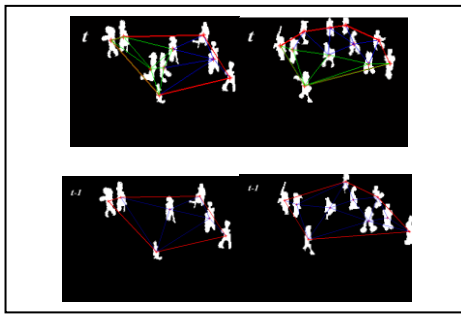


Fig.3 Miss matched triangles in the graph at time t-1 and t is represented by green triangles.

2) *Alignment of Miss-Matched Triangles:* To align the triangles that are not matched between frame pairs, to make proper alignment the cost of the alignment of miss-matched triangles is to be minimized. Let the miss-matched triangles set at time t-1 be Ψ_{t-1} , where $\Psi_{t-1} = \{\phi_1^{t-1}, \phi_2^{t-1}, \dots, \phi_\alpha^{t-1} | \phi_i^{t-1} = (\theta_{i1}, \theta_{i2}, \theta_{i3})\}$, $\Psi_{t-1} \subseteq \Delta_{t-1}$, where α is the number of miss-matched triangles, and $(\theta_{i1}, \theta_{i2}, \theta_{i3})$ is the internal angles set of the triangle ϕ_i^{t-1} . Similarly, the miss-matched triangles set at time t is Ψ_t , where $\Psi_t = \{\phi_1^t, \phi_2^t, \dots, \phi_\beta^t | \phi_i^t = (\theta_{i1}, \theta_{i2}, \theta_{i3})\}$, $\Psi_t \subseteq \Delta_t$ and where β is the number of triangles. The optimal matching for cost C is expressed as the matching between subsets ζ_{t-1} and ζ_t :

$$(\zeta_{t-1}, \zeta_t) = \operatorname{argmin}(C) \quad (17)$$

Therefore, the vertices order can be determined by the x and y coordinates, and thus the respective cost C_X and C_Y can then be obtained by

$$C = 0.5 * C_X + 0.5 * C_Y \quad (18)$$

3) *Global Graph Matching:*

The contour of a human crowd is represented by the convex hull and thus the convex hull variation in shape would reflect the alternations in outer area of the human crowd, i.e., the global changes. Moment invariants have been used to image pattern recognition in a variety of applications due to its invariant features on image like scaling, translation and rotation.

$$m_{p,q} = \sum_x \sum_y I(x,y) x^p y^q \quad (19)$$

Where $I(x, y)$ is pixel intensity in coordinate (x, y) . Where p is the x-order and q is the y-order, where the power to which the corresponding component is taken in the sum as determined as follows,

With Hu's moments, to compare two areas enclosing by convex hulls and to find whether they are similar. To make this effective process, Eq. (22) we have to compute and compare their moments according to a criterion that we provide.

$$M(H_{t-1}, H_t) = \frac{\sum_{i=1}^7 \left| \frac{m_i^{H_{t-1}} - m_i^{H_t}}{m_i^{H_{t-1}}} \right|}{7} \quad (20)$$

$M(H_{t-1}, H_t)$ to compare two areas enclosing by convex hulls.

$$m_i^{H_{t-1}} = \operatorname{sign}(h_i^{H_{t-1}}) \cdot \log|h_i^{H_{t-1}}| \quad (21)$$

$$m_i^{H_t} = \operatorname{sign}(h_i^{H_t}) \cdot \log|h_i^{H_t}|$$

4) *Abnormal Event Detection:* To detect the abnormal event, a multicue measure is used, say P_E , which is the combination of the properties mentioned above is defined as

$$P_E = \max(1 - P_s, P_n) \times [\xi \times P_m + (1 - \xi)P_t] \quad (22)$$

Where ζ is a pre-defined weight for balancing the global moment cost P_m and the local cost of triangle matching is expressed as P_t . Where P_s is the probability of the sub graph relationship between two graphs and is defined as

$$P_s = \frac{R}{\min(n_{t-1}, n_t)} \quad (23)$$

Where R is the number of eigenvalues of $G_t(V, E)$ and $G_{t-1}(V, E)$. P_n is a probability which is used to measure the number nodes that are miss-aligned and is defined by

$$P_n = 1 - \frac{Q}{\max(k_{t-1}, k_t)} \quad (24)$$

where Q represents the number of matched triangles for graphs $G_t(V, E)$ and $G_{t-1}(V, E)$. The graph matching cost by global contour is measured by

$$P_t = \frac{C}{C_{max} - C_{min}} \quad (25)$$

Finally, a temporal sliding window with width Ω is set to 5 empirically. An abnormal event can be detected if the accumulated value P_{E_T} is larger.

$$P_{E_T} = \sum_{t=T-4}^T P_E^t, \quad T \geq 5 \quad (26)$$

If $P_{E_T} \geq 5$: it is abnormal event otherwise normal event

V.RESULTS AND DISCUSSION

The efficiency of the algorithm proposed has been evaluated using Mat lab 10.0. For event detection, the UMN dataset[14] is taken to determine the event detection process. In this paper, the proposed method processes a video of about 25 frames per second for color images at size of 320 x 240. The result of occlusion is shown in figure. The result of occlusion detection is shown in fig.4.

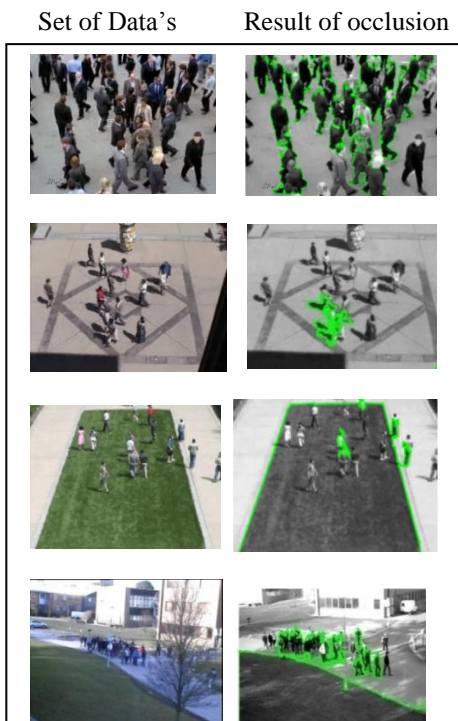


Fig.4, For occlusion detection, we use the some dataset to show the result

1) *Input Frames (UMN Dataset):* The original video is given as input that contains 100 frames and are shown in fig.5.



Fig.5, Input Frames (UMN Dataset)

2) *Background Subtracted Frames:* Background Subtracted frame are shown in Fig.6.



Fig.6, Background Subtracted Frames

3) *Normal activity frames:* After extracting the foreground images, the Graph modeled frames and normal activity is detected using a threshold value of 2.5715 and are shown in Fig.7

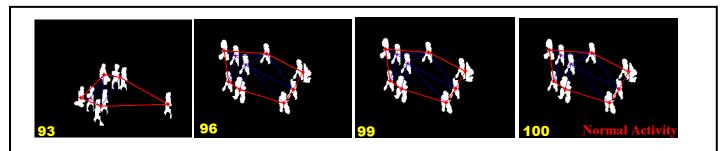


Fig.7, Normal activity frames

4) *Abnormal activity frames:* After extracting the foreground images, the Graph is modeled and abnormal activity is detected using a threshold value of 10.5761 and are shown in Fig.8.

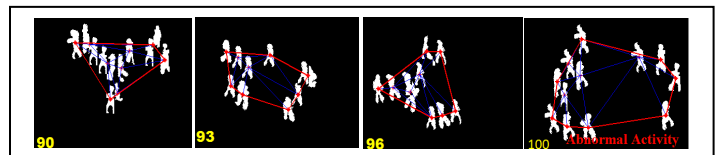


Fig.8, Abnormal activity frames

5) *Performance Analysis:*

Sensitivity Se is calculated using
 $Se = TP/TP+FN$

Accuracy A_c is calculated using

$$A_c = (TP+TN) / (TP+TN+FP+FN)$$

Error Rate E_r is calculated using

$$E_r = (FN+FP) / (TP+TN+FP+FN)$$

Using these equations algorithm efficiency is calculated. Table.1 shows estimated values of the performance analysis parameters. Table.2 shows Summarization of the dataset.

Table.1 Accuracy Analysis

Dataset	Sensitivity	Accuracy	Error Rate
DS1/pets2009	89.47%	89.40%	5.7%
DS2/UMN/1	86.45%	89.51%	4.6%
DS3/UMN/2	88.90%	90.45%	5.7%
DS5/ UCSD/Peds2	89.56%	90.58%	5.9%

Table.2 Summarization of the Dataset

Dataset	Sequence Length	Frame Size	Event detected
DS1/pets2009	91	768*576 pixels bmp image sequences	Abnormal Event
DS2/UMN/1	1450	228*158 pixels bmp image sequences	Normal Event
DS3/UMN/2	2145	320*240 pixels bmp image sequences	Abnormal Event
DS5/ UCSD/Peds2	200(taken a short clip)	360*240 pixels bmp image sequences	Normal Event

VI.CONCLUSION

In this paper, video synopsis is used to represent a short video while preserving the essential activities for a long video. In the existing methodology, usually a single moving object is splitted into a few small pieces in a continuous activity. For that Gaussian Mixture Model is used to detect compact foreground against their shadows. But in high-density crowds background subtraction fails due to occlusion. After detecting the occlusion, background is subtracted. A centroid based tracker is used to concatenate two trajectories if they belong to the same foreground activity. For video synopsis more fluent tubes are generated. Then each isolated region is considered as a vertex and a human crowd is thus modeled by a graph. After modeling the graph by using

Delaunay triangulation and Convex-Hull, graph matching algorithm is developed to detect the problem of behavior analysis of human crowds. Experimental results obtained by using UMN datasets show that the proposed algorithm is effective in detecting occlusion and anomalous events in quicker time for uncontrolled environment of surveillance videos. The Future work is to avoid collision due to video synopsis among the moving object. Because, this collision amongst foreground blobs merges to each other and it is difficult to analyze the behavior.

REFERENCES

[1] E. Andrade and R. Fisher, "Hidden markov models for optical flow analysis in crowds," in *Proc. 18th Int. Conf. Pattern Recognit.*, vol. 1. Sep. 2006, pp. 460–463Y.

[2] N. Ihaddadene and C. Djeraba, "Real-time crowd motion analysis," in *Proc. Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[3] D. Ryan, S. Denman, C. Fookes, and S. Sridharan, "Scene invariant crowd counting for real-time surveillance," in *Proc. IEEE 2nd Int. Conf. Signal Process. Commun. Syst.*, Dec. 2008, pp. 1–7.

[4] A. B. Chan, Z. Sheng, J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in *Proc. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 17.

[5] F. Jiang, Y. Wu, and A. K. Katsaggelos, "Detecting contextual anomalies of crowd motion in surveillance video," in *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 1117–1120

[6] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 935–942.

[7] H. Jiang, S. Fels, and J. J. Little. A linear programming approach for multiple object tracking. In CVPR, 2007.

[8] J. Berclaz, F. Fleuret, and P. Fua. Multiple object tracking using flow linear programming. In Winter-PETS, 2009.

[9] Pritch, A. Rav-Acha, and S. Peleg, "Nonchronological video synopsis and indexing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no.11, pp. 1971–1984, 2008.

[10] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, vol. 19. Jun. 1999, pp. 23–25.

[11] P. P'erez, C. Hue, J. Vermaak, and M. Gangnet, "Color based probabilistic tracking," *Computer Vision–ECCV 2002*, pp. 661–675, 2002. .

[12] A. Okabe, B. Boots, and K. Sugihara. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. New York, USA: Wiley, 2000

[13] Duan-Yu Chen and Po-Chung Huang, "Visual-Based Human Crowds Behavior Analysis Based on Graph Modeling and Matching," *In IEEE Sensors Journal*, Vol. 13, No. 6, June 2013

[14] University of Minnesota–Crowd Activity Dataset, (2010)[Online]. Available: <http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi>