Crowd Behavior Categorization using Live Stream based on Motion Vector Estimation

Amna Sajid, Sajid Gul Khawaja, Mehak Tofiq
College of Electrical and Mechanical Engineering,
National University of Sciences & Technology
Islamabad, Pakistan
amna971@gmail.com, sajid.gul@ceme.nust.edu.pk, mehaktofiq@yahoo.com

***Abstract*—The detection of anomalies in large crowd is a cognitive task. A proactive approach is required to effectively manage the crowd flow and to accurately detect the erratic behavior of crowd. In this paper, we present an algorithm which observes crowd optical flow in real time and detect any abnormal events in crowds automatically. The system takes the frames at regular intervals through a video camera and processes these frames using image processing techniques. The proposed system further uses certain rules to classify the normal or abnormal activities of crowd. We propose a novel motion vector based technique to detect behavior of the cluster of interest. The features of the motion vectors are analyzed to characterize the crowd behavior. The evaluation of the system is performed using different videos having different crowd behaviors and the results on simulated crowds demonstrate the effectiveness of the proposed system.**

Keywords*—*Motion vector; crowd behavior; real time processing

I. Introduction

With the passage of time, crowd surveillance has become an important issue. Security measures are needed at public places especially at large scale events which require proper crowd flow management. Even for short-range admittance control to a building or location, crowd behavior is tracked to assure the safety and security of individuals. One solution is to install security cameras in many public places which are monitored by security personnel. But being a human, security personnel cannot monitor each and every change in the midst of a large crowd as monitoring through cameras is a cognitive and tiresome task. So, there is a need of automated systems which can keep the complete track of crowd and their activities. In case of any erratic behavior of crowd, system should alarm the security personnel for immediate actions.

For proactive approach to avoid any violence or disruption in crowd, focus should be on the tracking and detecting of movement of crowd. Tracking and detecting the motion of individual is possible but at larger scale the behavior of crowd is of interest rather than individual. A model is required to efficiently process the large data of varying densities and motion. The abnormal behavior of crowd should be detected with less processing time.

For the detection of abnormal behavior of crowd, many systems have been developed on crowd behavior detection in computer vision application. Previously in crowd surveillance domain, work has been done in two categories. First field was the detection of crowd i.e. tracking the flow and density of crowd. Initially, In 1992, Rourke and Bell et al. proposed a new technique for estimating occupancies based on inter-frame image differences showing real-time potential with an implementation on a transputer network in [1]. Then another method was developed to detect density and motion without explicit knowledge of individual shape, size and velocity using edge detection [2]. In 1997, new methods were explored that were based on the relationship between the autonomous virtual humans of a crowd and the emergent behavior originated from it in [3]. In 1998, a motion estimation method and a motionless detection method was developed in [4]. Both methods take into account three difficulties: real time constraint, deformable objects and occultation. Three methods were mainly used to overcome these difficulties which were: Block matching (primitive matching technique), optical flow (differential technique) and Gabor filter (frequential technique). In 2003, a method was proposed that is used for pedestrian-to-pedestrian collision avoidance [5], while assigning goals to pedestrians to make their trajectories smoother and coherent. In [6] Berggren et al. devised a method to detect the crowd in computer games. Another method was developed that did the crowd feature selection and extraction and tracking based on the KLT tracker [7]. The detection of crowd was not enough as these methods does not work well in case of high disruption and occlusion in crowd. Later on, the crowd behaviors were investigated and the anomalies were localized due to individual's abrupt dissipation [8]. Cellular Automata (CA) [9], a new method, based crowd behavior model which mimics movements of humans in an indoor environment. A new method characterized crowd behavior by observing the crowd optical flow and use unsupervised feature extraction to encode normal crowd behavior [10]. The unsupervised feature extraction applied spectral clustering to find the optimal number of models to represent normal motion patterns. The motion models were HMMs to cope with the variable number of motion samples that might be present in each observation window. In this paper, we used a technique to detect the behavior of low density crowd.

In Section II, we briefly discuss system overview leading towards the details of the algorithm. In Section III, the experimentation and the results have been discussed. Finally, in Section IV paper is concluded along with directions to the future work.

II. System Overview

The brief description of the algorithm used in this paper is shown in Fig. 1.

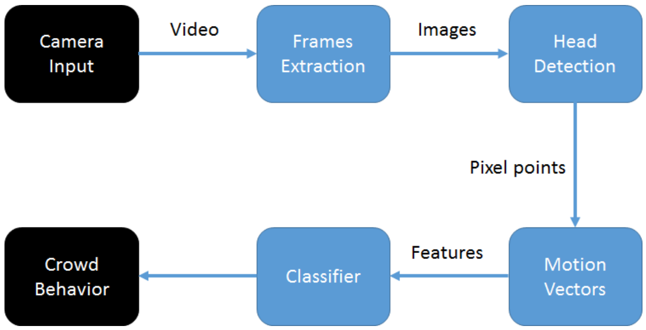

Fig. 1. Block diagram of the proposed model.

Real time crowd behavior detection is very important for security purposes in sensitive areas. We present a system for real time crowd monitoring and detecting their behavior. Firstly, the system takes a video input through a camera by capturing the top view and it extracts the frames from video on regular intervals. The system then applies pre-processing to remove any ambiguities due to varying light or densities. The next module detects the heads of persons present in the video using correlation based method which are further used to form motion vectors. Once the motion vectors are evaluated then they explicitly define the main features of the crowd. Further, on the basis of these features the classifier analyzes the crowd and makes a decision about the behavior.

A. Proposed System

In this section, we present our complete system and explanation of each and every module. The image acquisition module takes a video input from a camera and extracts frames from it. We keep four frames on regular intervals to make a decision for a small duration of time. First step in processing the frames is the detection of foreground pixels which we use to form motion vectors. We apply segmentation to obtain the foreground using a threshold level which is defined as

$$T >= 20 + min(I(x,y)) \quad (1)$$

Where I(x, y) is image and min(I(x,y)) represents the minimum intensity level in the image. Once the foreground is subtracted from background image, the correlation technique is applied to identify the active pixels of heads. A template window of head shape and size is made and that window is correlated with the image to get maxima where the head lies. After that a threshold is applied to get global maxima using following equation:

$$t = \frac{avg(avg(cor(x,y)) + max(cor(x,y))) + max(cor(x,y))}{2} \quad (2)$$

Where cor(x, y) is the correlation matrix and avg and max represent the average and maximum value of correlation. After applying the threshold on the correlation matrix, it is sorted and each value is evaluated as a region by using the concept of Euclidean distance. The first index of each region is selected which is detected as a pixel of head. This index is further used to manipulate the motion vectors.

For the case of dispersed motion there should be a point that is the reference point for an inward or an outward motion. For calculating that reference point, we take four points from two frames and find out the point of intersection by using the line equation given as:

$$y = mx + c \quad (3)$$

Reference point is calculated by the average of all intersection points. To know the direction of crowd, distance from reference point is computed and is compared with the distance calculated in the previous iteration.

After that motion vectors are evaluated. Motion vectors are the result of analysis that tells from where a pixel in the current frame is going to move in the next frame (or from where its coming from the previous frame). The features we selected for computing motion vectors are distance, slope and density of people. Distance is measured by calculating how much the pixel point of each head has displaced in the next frame as compared to the previous frame. Distance is calculated by the Euclidean distance formula given below

$$distance = \sqrt{(x_2 + x_1)^2 + (y_2 - y_1)^2} \quad (4)$$

Where x1 is the value of x-axis pixel point of head in previous frame, y1 is the value of y-axis pixel point of head in previous frame, x2 is the value of x-axis pixel point of head in the present frame and y2 is the value of y-axis pixel point of head in the present frame. Let (x1, y1) be the point of the person in present frame and (x2, y2) be the point of the person in the next frame. Then the slope can be calculated by the following formula:

$$slope = tan^{-1}(\frac{y_2 - y_1}{x_2 - x_1}) \quad (5)$$

The density of people is calculated by dividing the area i.e. 360 into 16 bins where each bin covers 22.5^o. The density of people in each bin is calculated using blob detection after detection of heads in the specific bin. Once all the bins have been evaluated a general sense is generated as normal, less or more dispersion on the bases of previous density and current density in respective bins.

A simple quantitative decision tree has been used as a classifier for our proposed system. Over all the decision tree is designed to work on *speed*, *dispersion* and direction of *movement* as the features, whereas the binary categorization of the crowd's behavior has been considered as the output of our decision tree. The behavior of crowd can either be *normal* or *abnormal* depending upon the dispersion of people and their relative speed. E.g. less dispersion at normal or low speed generally points towards normal behavior whereas more dispersion at relative high speed points towards abnormality in the behavior of the crowd.

Finally, any disperse movement of crowd due to riot or chaotic situation can be characterized using the proposed model. The divergence or convergence of crowd at a reference point can be explained with the help of inward or outward direction. In case of severe occlusion, the system displays an immediate alert to indicate the abnormal behavior which can help the video surveillance operator to take necessary steps against disrupted crowd on time.

III. Results and Discussion

The evaluation of proposed system is performed using 80 videos, synthetically generated, capturing different crowd behaviors. The results are determined through real time videos which captured the normal and abnormal behavior of crowd. Frames are taken at regular intervals to classify the trend of crowd. The behavior of crowd is characterized on the basis of speed, motion and direction. In Fig. 2, 3 and 4, different processing steps on frames are shown. The first image is the original image; the second image is obtained after adaptive thresholding. Third image shows the removal of background from the second image. Fourth image is the result of correlation. Fifth image shows the head detected and in the last image, the motion vectors are shown for each individual.

In Fig. 2, the motion of the crowd is straight representing the normal behavior. For abnormal events in crowd, the two cases are categorized on the basis of direction. A reference point is calculated in each frame to determine the inward or outward direction of dispersed crowd. In case of accident or fighting situation, many people approach the scene from different angles which show an abnormal activity. As we can see in Fig. 3 people are approaching a reference point, indicating inward motion of the abnormal crowd. In case of fire emergency, people will scatter in different directions to evacuate the place. As we can see in Fig. 4 crowd is dispersing in outward direction from a reference point.

Out of 80 sample videos, there are 52 videos that are of normal case and 28 videos that are of abnormal case. Out of 52 normal cases 40 are detected as normal and 12 are detected as abnormal while in abnormal case we have 28 videos out of which 20 are detected as normal and 8 are detected as normal. The results obtained after experiments are shown in Table I as confusion matrix.

TABLE I. Confusion Matrix Showing the Results of Proposed System on the 80 Videos

	Normal	Abnormal
Normal	40	12
Abnormal	8	20

IV. Conclusion and Future Work

In this paper, we propose a simplified algorithm for the analysis of crowd behavior in real time environment. The aim is to detect any abnormal event in crowd and to help the video operator to take any precautionary measures on time or before the occurrence of abnormal activity. The heads of crowd are detected as a cluster of interest and this information is used to form the motion vector. The proposed algorithm is tested on the real time video feed. The effectiveness of algorithm is tested on many different scenarios synthetically generated. The results obtained are efficiently detecting the behavior of crowd using features extracted from the motion vectors.

Future work can be done on head detection technique which can be improved using other algorithms. The processing time of results can be reduced further. More features can be added to analyze the speed or motion variations. Acceleration or energy can be estimated to give more promising results.

References

[1] A. Rourke and M. G. Bell, "Video image-processing techniques and their application to pedestrian data-collection," *TORG RESEARCH REPORT*, no. 83, 1992.

[2] S. Velastin, J. Yin, M. Vicencio-Silva, A. Davies, R. Allsop, and A. Penn, "Image processing for on-line analysis of crowds in public areas," *IFAC Transportation systems, Tianjin, Proceedings*, 1994.

[3] S. R. Musse and D. Thalmann, "A model of human crowd behavior: Group inter-relationship and collision detection analysis," in *Computer Animation and Simulation97*. Springer, 1997, pp. 39–51.

[4] S. Bouchafa, D. Aubert, and S. Bouzar, "Crowd motion estimation and motionless detection in subway corridors by image processing," in *Intelligent Transportation System, 1997. ITSC'97., IEEE Conference on*. IEEE, 1997, pp. 332–337.

[5] C. Loscos, D. Marchal, and A. Meyer, "Intuitive crowd behavior in dense urban environments using local laws," in *Theory and Practice of Computer Graphics, 2003. Proceedings*. IEEE, 2003, pp. 122–129.

[6] R. Berggren, "Simulating crowd behaviour in computer games," Ph.D. dissertation, BSc dissertation. Luleå University of Technology, 2005.

[7] S. Saxena, F. Brémond, M. Thonnat, and R. Ma, "Crowd behavior recognition for video surveillance," in *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 2008, pp. 970–981.

[8] A. Zaharescu and R. Wildes, "Anomalous behaviour detection using spatiotemporal oriented energies, subset inclusion histogram comparison and event-driven processing," in *European Conference on Computer Vision*. Springer, 2010, pp. 563–576.

[9] D. Wang, N. M. Kwok, X. Jia, and F. Li, "A cellular automata based crowd behavior model," in *International Conference on Artificial Intelligence and Computational Intelligence*. Springer, 2010, pp. 218–225.

[10] E. L. Andrade, R. Fisher, and S. Blunsden, "Detection of emergency events in crowded scenes," in *Crime and Security, 2006. The Institution of Engineering and Technology Conference on*. IET, 2006, pp. 528–533.

Fig. 2. (a) Original Frame showing normal behavior (b) Heads Identification (c) Binarization using Adaptive thresholding (d) Correlation with previous frame (e) Head Detection (f) General direction on the bases of Motion Vectors.

Fig. 3. (a) Original Frame showing anormal behavior, people approaching a reference point (b) Heads Identification (c) Binarization using Adaptive thresholding (d) Correlation with previous frame (e) Head Detection (f) Inward direction on the bases of Motion Vectors.

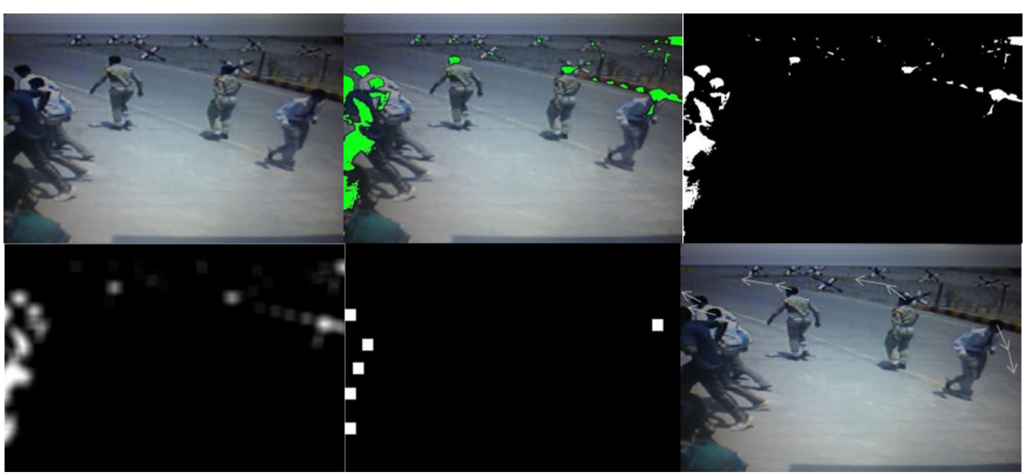

Fig. 4. (a) Original Frame showing abnormal behavior as people are moving away from a reference point (b) Heads Identification (c) Binarization using Adaptive thresholding (d) Correlation with previous frame (e) Head Detection (f) Outwards direction on the bases of Motion Vectors.