Support Vector Machine-Based Human Behavior Classification in Crowd through Projection and Star Skeletonization

B. Yogameena, E. Komagal, M. Archana and S. Raju Abhaikumar
Department of Electronics and Communication Engineering,
Thiagarajar College of Engineering, Madurai, Tamil Nadu, India

Abstract: Problem statement: Detection of individual's abnormal human behaviors in the crowd has become a critical problem because in the event of terror strikes. This study presented a real-time video surveillance system which classifies normal and abnormal behaviors in crowds. The aim of this research was to provide a system which can aid in monitoring crowded urban environments. **Approach:** The proposed behaviour classification was through projection which separated individuals and using star skeletonization the features like body posture and the cyclic motion cues were obtained. Using these cues the Support Vector Machine (SVM) classified the normal and abnormal behaviors of human. **Results:** Experimental results demonstrated the method proposed was robust and efficient in the classification of normal and abnormal human behaviors. A comparative study of classification accuracy between principal component analysis and Support Vector Machine (SVM) classification was also presented. **Conclusion:** The proposed method classified the behavior such as running people in a crowded environment, bending down movement while most are walking or standing, a person carrying a long bar and a person waving hand in the crowd is classified.

Key words: Surveillance, human behavior modeling, projection, star skeletonization

INTRODUCTION

Security of citizens in public places such as Hotels, markets, airports and railway stations has increasingly become a crucial problem world widely after September 11, 2001. The fundamental problem in visual surveillance systems is detecting human presence, tracking human motion, analyzing the activity and assess abnormal situations completely automatically. Based on this motivation, crowd modeling technology has been under development to analyze the video input which is constantly crowded with humans, as well as to ready to act against abnormal activities emerge. The aim of the study is to classify human normal and abnormal behaviors in crowds using projection and star skeletonization for a surveillance system which is based on SVM.

Related work: There have been a number of video surveillance systems, for multiple people detection and tracking (Ali *et al.*, 2006; Schiele *et al.*, 2008). Tracking crowded Environment (Zhao and Nevatia, 2004). All these systems can detect and track multiple people. Different methods have been applied to recognize human actions, representing human activities by a set of pose and velocity vectors (Ben-Arie *et al.*, 2002) and using the position and velocity of body parts for learning (Lee and Xu, 2004) and employs Principal Component Analysis for feature selection and Support Vector Machine for classification of human behaviors (Wu *et al.*, 2005). Thus, the majority systems are not designed for classifying normal and abnormal behaviors.

In this study, a novel approach using geometric projections, skeleton features for SVM learning system for indoor outdoor is presented. In this implementation foreground segmentation is accomplished in real-time using the adaptive mixture of Gaussians method proposed by Atev *et al.* (2004). The area that a foreground region occupies on the ground plane in the world is one of the cues used to distinguish between individuals and groups. Additionally, for regions that are labeled as groups, the projected area is used to estimate the number of people in the group and to separate as individuals. The abstract describes the projective geometry used to project a foreground blob into world coordinates and approximate its area (Kilambi *et al.*, 2008).The boundaries are extracted from foreground blobs and star skeletonization is used for feature selection (Fujiyoshi and Lipton, 1998). Using these features SVM classifies the normal and

Corresponding Author: B. Yogameena, Department of Electronics and Communication Engineering,
Thiagarajar College of Engg, Madurai, Tamil Nadu, India

abnormal behaviors, such as running people in a crowded environment, bending down movement while most are walking or standing, a person carrying a long bar and a person running and waving hand in a very crowded environment.

This study varies from SVM used in (Wu *et al.*, 2005) by means of feature selection to the input of Support Vector Machine for classification of human behaviors. In this system, the learning approach based on the border information of the blobs is proposed. Experimental results demonstrate that the approach is robust in classifying human abnormal behaviors.

The remainder of this study is organized as follows: The methodology, Pre-processing, projection and star skeletonization and SVM learning system for classification are explained. Then the experimental results are demonstrated. Finally, the conclusion is presented.

MATERIALS AND METHODS

The overview of the system is shown in Fig. 1. In the foreground detection stage, the blob detection subsystem detects the foreground pixels using a statistical background model. Then, the foreground pixels will be grouped into blobs.

In the segmentation stage, the blobs that contain multiple people are divided into individuals when occlusion happens. Each foreground region is labeled as an individual or a group. These classifications are made based on the projected sizes and velocities of the regions. A tracker is then automatically initialized for each foreground blob that is identified as either an individual or a group.

To detect abnormal behaviors, the contour is extracted for each blob. Then the distances from the centroid to each border point are calculated. The local minimum and maximum points for each blob are selected as features and they formed star skeleton of human body. Then SVM is used to classify the behaviors into two classes: Normal behavior and abnormal behavior.

Behavior analysis using SVM learning:

Pre-processing: It is assumed that people move on the ground plane in the real world. After foreground detection, the Blob is projected to head and ground plane from the camera view point leading to intersected area in world coordinates. Such projection is shown in Fig 2. This eliminates the variation of area with the distance from the camera and shorter objects which will identify only humans. The objective here is to have the head plane low enough so that all the individuals in the

scene are detected by Eq. 1-3. This helps to formulate the method robust to false detections from other objects and shadows (Kilambi *et al.*, 2008).

Individual and group classification: All blobs whose projected area exceeds a threshold are classified either as a group as Shown in Fig. 3 or a large object (bus or car). This threshold is selected to be just under the area corresponding to two individuals in the real world. For all blobs whose area is less than the area threshold, then the region is classified as an individual and correspondingly. If it stays above a velocity threshold, they are classified as vehicles and ignored. The remaining large blobs are classified as groups.
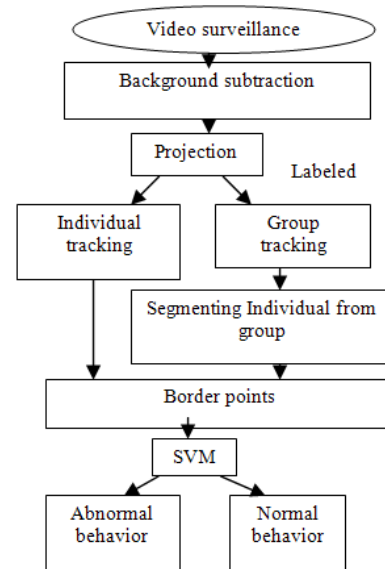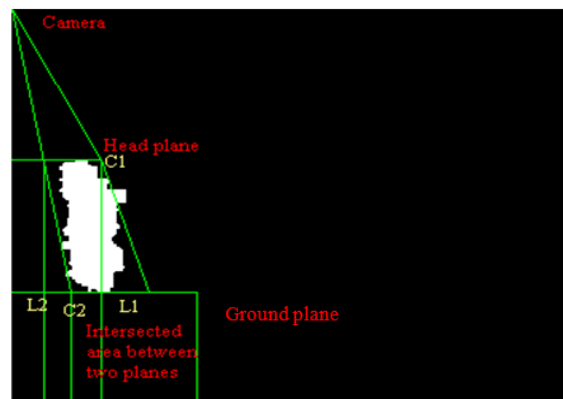

Fig. 1: The block diagram of the system

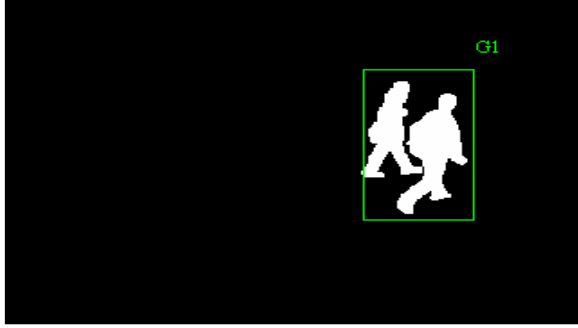Fig. 2: Single human intersected area between the ground plane and the head plane projection
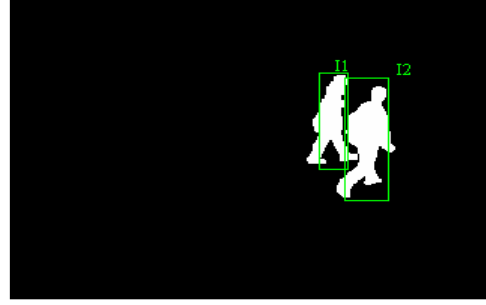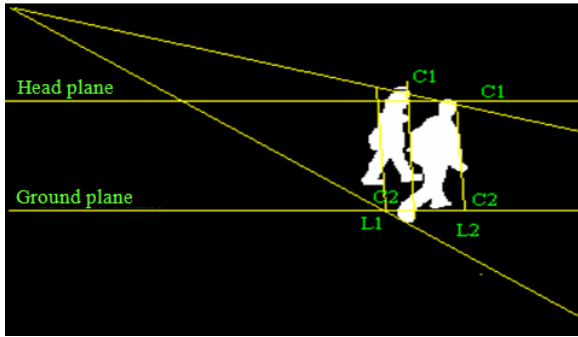
Fig. 3: Group Frame 70

Fig. 4: Group human intersected area between the ground plane and the head plane projection for frame 80

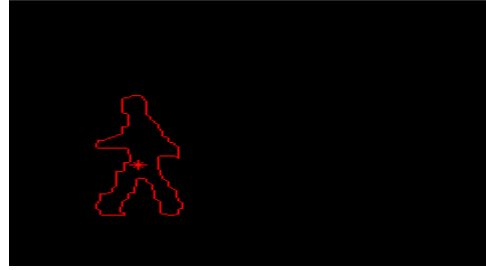The individuals are recognized and labeled as follows:

$$\mathrm{Int\,er\,sec\,ted\,plane} = C1 - C2 \tag{1}$$

$$\mathrm{Head\,plane} = C1 - L2 \tag{2}$$

$$\mathrm{Grand\,plane} = L1 - C2 \tag{3}$$

Given the intersected area of the group being tracked, estimate the count for this current frame using Eq. 4 and is shown in Fig. 4:

$$\mathrm{Count} = \frac{\mathrm{Area}}{\mathrm{K}} \tag{4}$$

K is the individual's intersected area.

According to the estimated intersected area for individual, the group is separated as individuals and is shown in Fig. 5.

Fig. 5: Segmented Individual frame 80

Fig. 6: Border points and centroid

Star skeletonization: After segmentation and tracking, the system collects consecutive blobs for a single person. Then, the contour for a human blob is extracted (Fujiyoshi and Lipton, 1998). As seen in Fig. 6, the centroid (x_c, y_c) of the human blob is determined by using the following Eq. 5 and 6 shown in Fig. 5:

$$X_c = \frac{1}{N} \sum_{j=1}^{N} X_i \tag{5}$$

$$Y_c = \frac{1}{N} \sum_{j=1}^{N} Y_i \tag{6}$$

where, (x_c, y_c) represent the average contour pixel position, (x_i, y_i) represent the points on the human blob contour and there are a total of N points on the contour (Fujiyoshi and Lipton, 1998). The distance d_i from the centroid to contour points is given by Eq. 7:

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \tag{7}$$

From the d_i plot the local maximum and minimum points are collected and star skeletonization is formed as shown in Fig. 7 and their corresponding Plot is shown in Fig. 8. These points are selected as features and fed into SVM learning system to classify normal and abnormal behavior of human.
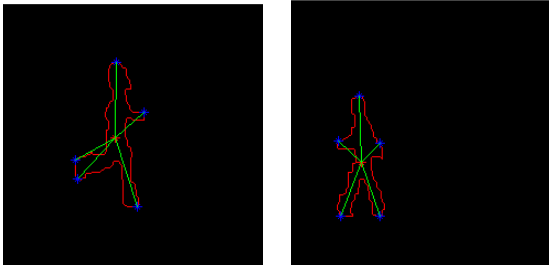
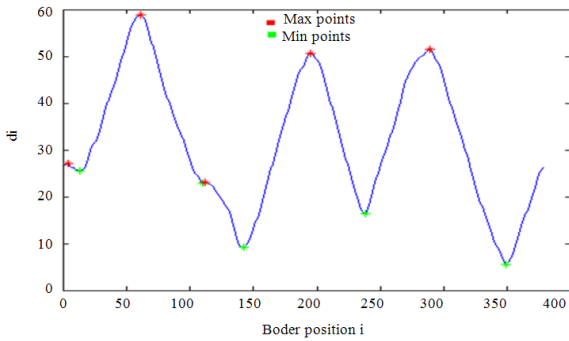Fig. 7: Skeleton points for running and walking

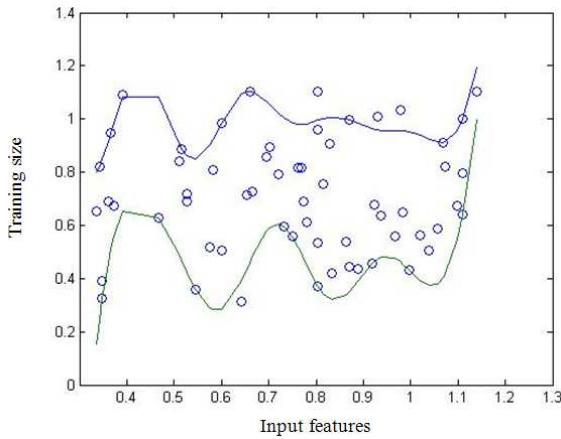Fig. 8: Plot of skeleton peak points

Fig. 9: SVM learning iteration 3 for running and walking

SVM classifiers: The features obtained from star skeletonization are fed into SVM learning system. The support vector machine is a powerful tool for binary classification, capable of generating very fast classifier functions following a training period (Lee and Xu, 2004). Having so many advantages, SVM has achieved great success in pattern recognitionand it applies in many domains, such as hand-writing recognition, face recognition, voice recognition, text classification and image processing (Awad and Motai, 2008; Bauckhage *et al*., 2009).

SVM can be considered as a linear method in a high-dimensional feature space nonlinearly related to the input space. By the use of kernels, all input data are mapped nonlinearly into a high-dimensional features space. Ou and Xu (2003) Using appropriate kernel functions, it is possible to compute the separating hyper planes without explicitly carrying out the map into the feature space. Hence SVM is basically a binary classifier. Several different schemes can be applied to the basic SVM algorithm to handle the K-class pattern classification problem (Crammer and Singer, 2001).

Every sample is partitioned by a series of optimal hyper planes. The optimal hyper plane means training data is maximally distant from the hyper plane itself and the lowest classification error rate will be achieved when using this hyper plane to classify current training set. These hyper planes can be modeled as Eq. 8:

$$\omega_{st} X_i^T + b_{st} = 0 \tag{8}$$

Where:
X_i = Denotes the i th row of matrix X
s and t = Mean two partitions which are separated by an optimal hyper plane and what these partitions mean lies on the construction of multiclass classification algorithms

And the classification functions are defined as Eq. 9:

$$f_{st}(X_i^T) = \omega_{st} X_i^T + b_{st} = 0 \tag{9}$$

Where:
S = t mean two arbitrary classes in K classes
ω_{st} = An n-dimensional weight vector
b_{st} A bias term

Experimental Results an In K-class SVM, K decision functions are constructed for each class. But each of these decision functions is a linear combination of all the training data points belong to that particular class. Figure 9 shows the SVM learning during the 3rd iteration.

RESULTS AND DISCUSSION

The proposed method can detect many abnormal behaviors by using one framework. In the training process, for example, we put the blobs of normal behaviors like standing, walking into one class and put the blobs of abnormal behavior like running into the other class Multi class SVM is applied to classify these behaviors into two classes: Abnormal behavior and normal behavior.

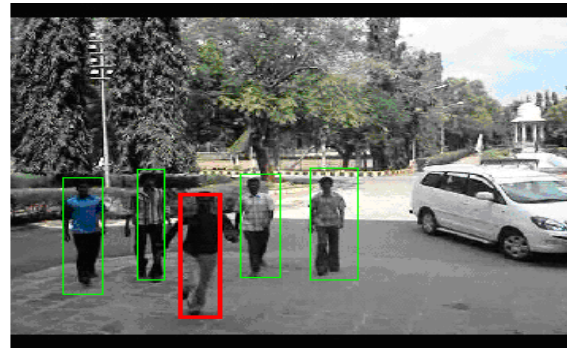Fig. 10: A bending down person normal behavior abnormal behavior
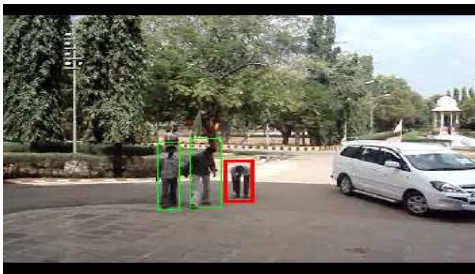

Fig. 11: One person bending down

Fig. 12: One person waving hand

Fig. 13: One person running in the crowd

Fig. 14: Results for benchmark datasets

Table 1: The results of SVM classification

	Success rate (%) (2200 samples)		
Behaviors	SVM	Multi class SVM with PCA features	Multi class SVM with skeleton features
Normal	93.2	94.6	97.2
Abnormal running	82.5	86.2	96.7
Carrying bar	84.3	87.3	93.5
Bending	78.3	83.6	93.4
Waving	86.5	92.4	95.7

After SVM training, the algorithm is tested by using a sequence of benchmark videos as well as videos taken in our college campus which is illustrated in Fig. 10. It shows that a person is bending down while others are standing and they are considered as background in the bench mark video and they are detected as foreground in the campus video as shown in Fig. 11. A person waving his hand and a person running in the crowd are shown in Fig. 12 and 13. As illustrated in the Fig. 14, the individuals are clearly projected and then, the star skeleton is used to obtain the motion cues. The skeleton features are obtained clearly for individual blob like walking, running and jumping. Even in the CMU Pulling dataset the individuals are clearly skeletonized, the abnormal behavior is classified as normal behavior.
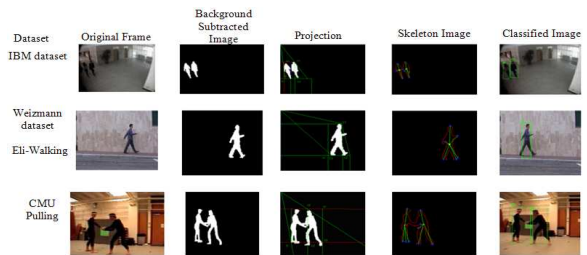
The algorithm has been tested using 2200 samples. The success rate of the behavior classification is shown in Table 1. Occasionally the classification algorithm considers normal behavior as abnormal behavior leading to false alarm. To reduce the false alarm rate, the skeletonization, without using low pass filter to smooth the position Vs distance curve, selecting all local minimum and maximum points as features, the system will reduce the false alarms.

CONCLUSION

In this study, a real-time visual surveillance system for classifying human normal and abnormal behaviors is described. To detect abnormal behaviors, skeleton features are extracted and fed into the SVM learning system. Then the processed data are learnt through a

support vector machine to classify the behaviors into two classes. The proposed system has successfully detected the following abnormal behaviors: Person running in the crowd, person bending down and person waving hand while others walk. Experimental results are compared with PCA Method.

REFERENCES

Ali, M.A., S. Indupalli and B. Boufama, 2006. Tracking multiple people for video surveillance. Computer-Vision. http://www.computer-vision.org/4security/pdf/windsor.pdf

Atev, S., O. Masoud and N. Papanikolopoulos, 2004. Practical mixtures of Gaussians with brightness monitoring. Proceeding of the 7th IEEE International Conference on Intelligent Transportation Systems, Oct. 3-6, IEEE Xplore Press, USA, pp: 423-428. DOI: 10.1109/ITSC.2004.1398937

Awad, M. and Y. Motai, 2008. Dynamic classification for video stream using support vector machine. Applied Soft Comput., 8: 1314-1325. DOI: 10.1016/j.asoc.200711.008

Bauckhage, C., J.K. Tsotsos and F.E. Bunn, 2009. Automatic detection of abnormal gait. Image Vis. Comput., 27: 108-115. DOI: 10.1016/j.imavis.2006.10.004

Ben-Arie, J., Z. Wang, P. Pandit and S. Rajaram, 2002. Human activity recognition using multidimensional indexing, IEEE Trans. Patt. Anal. Mach. Intel., 24: 1091-1104. DOI: 10.1109/TPAMI.2002./023805

Crammer, K. and Y. Singer, 2001. On the algorithmic implementation of multi-class kernel-based vector machines. J. Mach. Learn. Res., 2: 265-292. http://citeseerx.ist.psu.edu/viewdoc/download?doi= 10.1.1.69.8716&rep=rep1&type=pdf

Fujiyoshi, H. and A.J. Lipton, 1998. Real-time human motion analysis by image skeletonization. Proceeding of the 4th IEEE Workshop on Applications of Computer Vision, Oct. 19-21, Prinston, IEEE Computer Society, Washington DC, USA, pp: 1-15. DOI: 10.1109/ACV.1998.732852

Kilambi, P., E. Ribnick, A.J. Joshi, O. Masoud and N. Papanikolopoulos, 2008. Estimating pedestrian counts in groups. Comput. Vis. Image Understand., 110: 43-59. DOI: 10.1016/j.CVIU.2007.02.003

Lee, K.K. and Y. Xu, 2004. Modeling human actions from learning. Proceedings of the IEEU/RSJ International Conference on Intelligent Robots and Systems, Sept. 28-Oct. 2, IEEE Xplore Press, USA., pp: 2787-2792. DOI: 10.1109/TROS.2004.1389831

Ou, Y. and Y. Xu, 2003. Learning human control strategy for dynamically stable robots: Support vector machine approach. Proceeding of the IEEE International Conference on Robotics and Automation, Sept. 14-19, IEEE Xplore Press, USA., pp: 3455-4729. DOI: 10.1109/ROBOT.2003.1242124

Schiele, M., S. Roth and B. Schiele, 2008. People-tracking-by-detection and people-detection-by-tracking. Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 24-26, ACM Press, Anchorage, Alaska, USA., pp: 1-8. DOI: 10.1109/CVPR.2008.4587583

Wu, X., Y. Ou, H. Qian and Y. Xu, 2005. Detection system for human abnormal behavior. Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Aug. 2-6, IEEE Xplore Press, USA., pp: 1204-1208. DOI: 10.1109/IROS.2005.1545205

Zhao, T. and R. Nevatia, 2004. Tracking multiple humans in crowded environment. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 27-July 2, IEEE Xplore Press, USA., pp: 406-413. DOI: 10.1109/CVPR.2004.1315192