Anomalous Event Detection in Videos Using Supervised Classifier

K. Seemanthini and S. S. Manjunath(✉)

Dayananda Sagar Academy of Technology and Management, Bangalore, India
be.outstanding@gmail.com, mnj_ss2002@yahoo.co.in

Abstract. Observing and modeling human behavior and activity patterns for detecting anomalous events has gained more attention in recent years, especially in the video surveillance system. An anomalous event is an event that differs from the normal or usual, but not necessarily in an undesirable manner. The major challenge in detecting such events is the difficulty in creating models due to their unpredictability. Most digital video surveillance systems rely on human observation, which are naturally error prone. Hence, this work validates the rising demand of analysis of video surveillance system. The system being proposed here is of minimum requirements with a competitive computational power when compared to the existing ones.

The main objective of this research work is to build up a framework that recognizes small group of human and to detect the event in the video. A combination of feature extraction using Histogram of Oriented Gradient (HOG) and feature reduction with Principle Component Analysis (PCA) is proposed in this work. The knowledge base and video feed for test cases are classified using the Support Vector Machine (SVM) to categorize the event as either anomalous or not based on various parameters.

The experimental result demonstrates that this approach is able to detect anomalous events with a competitive success rate. The framework can be used to identify various events such as anomalous detection of events, counting people, fall detection, person identification, gender classification, human gait characterization etc.

Keywords: Foreground extraction · Human group extraction (HOE)
Visual saliency · Event detection · Classification

1 Introduction

1.1 Overview

Detecting human beings and identifying actions in videos of a surveillance system is gaining more importance due to its wide variety of applications in detecting anomalous events, counting the number of people in a dense crowd, identification of human, traffic safety and surveillance, sports analysis, gender classification, human gait characterization, fall detection for elderly people, etc.

Latest surveillance cameras are installed all around the world daily, as webcams, for surveillance and other purposes. Most of the digital video surveillance systems depend on human observers for detecting and identifying specific activities in the video scenes.

But, there are several limitations in the human capability to monitor simultaneous events in a video surveillance system. Hence, automated human event analysis in video surveillance system has become one of the most effective and attractive research topics in the area of computer vision and pattern recognition.

However, most existing multi-person tracking methods are still limited to special application scenarios. They require multi-camera input, scene specific knowledge, a static background, or depth information, or are not suitable for online processing.

Moreover, there may be both individual and gathering activity in the same scene, it is much harder to speak to and receive such situations. Punctuation models have been generally utilized as a part of the complex visual occasion acknowledgment lately. To apply sentence structure in models or occasion acknowledgment, normally low-level components are firstly removed from features and after that characterized to an arrangement of terminal images, i.e., visual occasion primitives. Hence the proposed work detects and identifies both individual and group events, hence it overcomes the drawback of existing system, and it is also based on cognitive linguistic method which uses unsupervised learning method.

Detecting human object is a difficult task from a machine vision perspective as it is motivated by a wide variety of possible appearance due to changing articulated pose, lighting, clothing and background, but prior knowledge on these limitations can improve the performance of detection. The proposed system detects and captures motion information of moving targets for accurate object classification. Unsupervised classifiers are used for learning method and labels are known, hence the instance such as kick, hug, punch, or any such features are extracted and events are detected. The classified object is being used for high level analysis.

1.2   Objectives

The main objective of this research work is to build up a framework that recognizes the small human group and to detect the event in the video. This framework is utilized for robotized little human gathering occasion discovery inside of social or open spot environment furthermore serves to recognize a fording wrongdoing, in places like Railway station, traffic, collages, office, etc.

1.3   Problem Statement

Detecting human beings and identifying actions in videos of a surveillance system is gaining more importance due to its wide variety of applications in detecting anomalous events, counting the number of people in a dense crowd, identification of human, traffic safety and surveillance, sports analysis, gender classification, human gait characterization, fall detection etc.

The proposed work is used to represent both individual and multiple individuals in an event, hence it overcomes the drawback of existing system, and it is based on cognitive linguistic method which uses unsupervised learning method.

The proposed system is able to identify the events automatically in the video surveillance system. Thus, it reduces the human interaction with the video surveillance system and reports the alerts as the events detected.

1.4   Proposed Methodology

The image is given as an input to the training database. The obtained RGB images are further preprocessed using mathematical morphological method to reduce noise and later converted to grayscale. Features are extracted using HOG descriptor and reduced using PCA. The resultant is stored in a file which is trained using SVM classification. On the other hand the testing dataset is converted to frames, preprocessed, and their absolute difference is evaluated to distinguish background from foreground. Further morphological operation takes place to reduce noise, followed by Feature extraction using HOG and PCA and classified using the SVM Classifier. The key techniques used are

1. Preprocessing using morphological operations.
2. Feature Extraction and Reduction using HOG and PCA.
3. Classification using SVM.

1.5   Applications

Automated anomaly detection has a wide variety of applications. It has huge potential in the field of video surveillance system. Even though video surveillance cameras are installed everywhere, the availability of human resources to monitor the footage is poor. Hence, an automated system will aid in overcoming such human errors. Events such as trespassing can be alerted immediately when an automated system is placed.

Detection of non-human objects in unexpected places aids in betterment of security measures. It helps in person counting in densely crowded places such as those shown in Fig. 1. An automated anomaly detection system may aid in fall detection in the homes

(a) Stadium

(b) Shopping Mall

(c) Airport

(d) University Pass

Fig. 1. Analysis of different scenario through anomalous event detection.

of the elderly. Traffic safety is the major applications of anomaly detection. Detection of speeding vehicles or reporting drivers breaking the law immediately can be achieved using an anomaly detection system. Another growing field is in sports analysis where an automated system might alert the referee or judge in case of actions which may otherwise be overlooked.

2   Literature Survey

Zhaozhuo Xu [1] has introduces a Human-Object Interaction model, and are able to establish methods and systems to recognize events that are dangerous. In this approach, the process of event understanding is based on identifying dangerous objects in possible areas predicted by human body parts. The accuracy of dangerous human events understanding is improved when human body parts estimation is combined with objects detection.

Dongping Zhang [2] presents an approach to identify group level crowds and detect any abnormal activities in them. It incorporates particle motion information calculated using a set of sample images with long trajectories and other properties, into identifying small human crowds in foreground images while in motion. Science of Human behaviour is studied and employed to detect normal and abnormal activity. Attributes such as orientation, velocity and crowd size are used to distinguish between normal and abnormal behaviour.

MyoThida [3] has presented a review of crowd video analysis in this paper. Automation of surveillance has become in crowded places such as shopping malls, railway stations and airports. Providing intelligent solutions to these places is of high priority to computer researchers. The paper provides a thorough review of the existing automation techniques for analyzing complex and crowded scenes.

The merits and demerits of the various modern methods are discussed in detail. Tracking individuals in a crowd is a major topic. It is a highly complex task due to interactions with various other objects present in the crowd.

M. Sivarathinabala [4] proposes an intelligent video surveillance system, which can be remotely monitored and alerts the user in a situation that the system may interpret as an anomaly. The main focus is on monitoring a single person in situations such as a burglary.

A live video is captured and reduced to images. The images undergo preprocessing. Human behaviour analysis plays an important role to detect any anomalous human activity. This is done by comparing existing sample templates with the processed image. If found, the image is stored in the system and an alert is sent as specified by the user either to MMS, SMS or email. The live video is then compressed and a key frame is specified to directly retrieve the required part of the video.

This paper concludes by providing an automated method for surveillance that not only identifies an anomaly but also triggers an alert to the user. It helps in retrieval of

the suspected video by holding key frame values and help in extracting of images of individuals before and after the incident.

Manoranjan Paul [5] throws light on the need for accurately detecting anomalies in videos and its applications in surveillance technologies. Detecting human beings and their actions accurately in a video has various applications such as person identification, fall detection for elderly people, event classification and gender classification.

The authors use the benchmarks set by few existing datasets for comparison and providing their assessment. An intelligent system can capture and detect moving objects in a video. In this study the authors focus on detecting only human beings in general. This in itself is a complex task due to the number of various attributes each person may have such as, clothing, pose, lighting and background.

Detecting objects in a surveillance video is a challenging task due to the low resolution of the video. This paper discusses different methods of object identification and object classification. The various benchmarks are discussed and the applications of human detection in surveillance videos are reviewed.

C. Stauffer [6] proposed a computer vision algorithm for detecting or analyzing the motion of people in crowds. Computer vision algorithm divides background in regions and track the crowds and analyses every movement of people.

D. Ryan [7] develops a scene independent approach that can count the no of people in the crowd. A scene independent counting system can easily be deployed at different place. The counting is been done using a global scaling factor to relate crowd size from one scene to another.

Condition of providing the right heuristic ranking to the individuals, to avoid confusing them with one another. Hence, achieve robustness by finding optimal trajectories over many frames while avoiding the combinatorial explosion that would result from simultaneously dealing with all individuals.

3 System Design

3.1 Overview

The following image represents the system architecture which is further broken down into a clear flowchart in the later segments. The architecture as described in the image consists of testing phase and training phase. In the training phase, the training video is added into the knowledge base after preprocessing, feature extraction and reduction. This dataset is classified into either normal or abnormal event using SVM Classifier. The testing set of the video, which goes through the same operations are classified as abnormal or normal by comparing it to the sample frames of the training videos that are classified already. The output classifies each frame to be either "Normal" or "Abnormal" (Fig. 2).
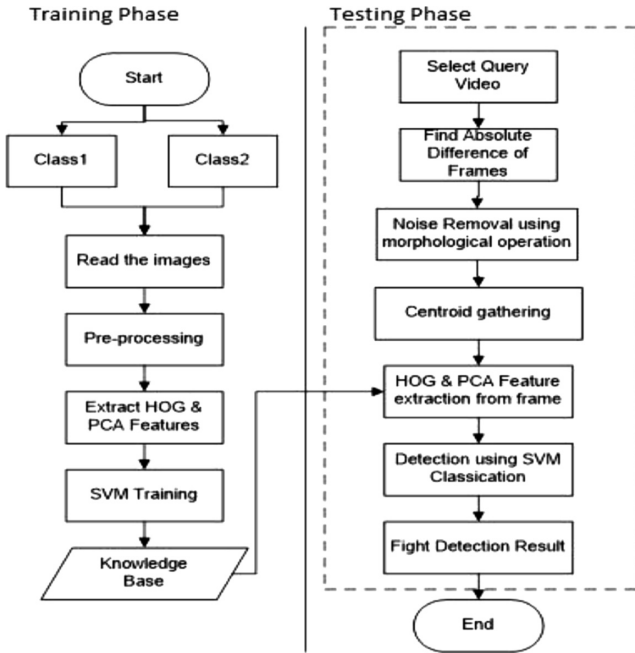
Fig. 2. Overall flow chart of the proposed system

3.2   Preprocessing

Initially the given video is converted into frames. The converted frames are used for further processing. In pre-processing unnecessary noise in the frames are eliminated using morphological operations. In order to get a more accurate difference between the background and the foreground, the image needs to have lesser noise [7].

3.3   Feature Extraction

The main purpose of feature extraction is to extract the image component and to separate the foreground from the background through HOG feature extraction and further reduce the obtained attributes by the method of PCA. This helps in providing faster time for analysis due to a better predictive model, with many similar attributes reduced to a single attribute (Figs. 3, 4 and 5).

3.3.1   Histogram of Oriented Gradients
Histogram of oriented gradients is a feature descriptor used in image processing for detecting objects. This technique counts occurrences of gradient orientation in localized portions of an image. The HOG descriptor is most popularly used for detecting humans in images. The flow diagram of the HOG descriptor is given below (Fig. 6).
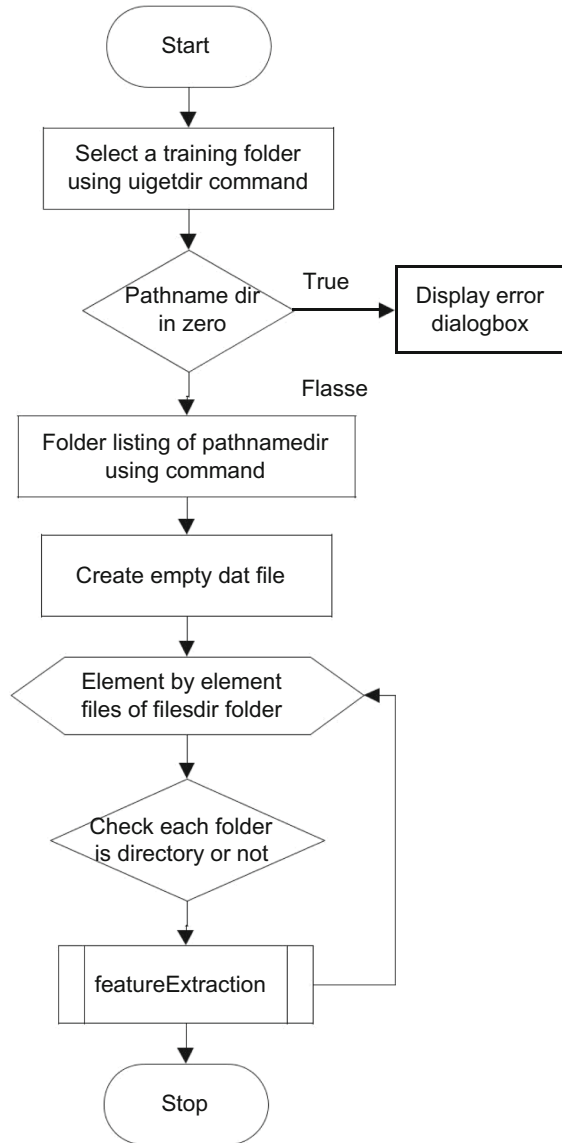
Fig. 3. Feature collection

3.3.2 Principal Component Analysis

PCA is mathematically defined [8] as an orthogonal linear transform that transforms the data to a new coordinated system such that the greatest variance by some projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.
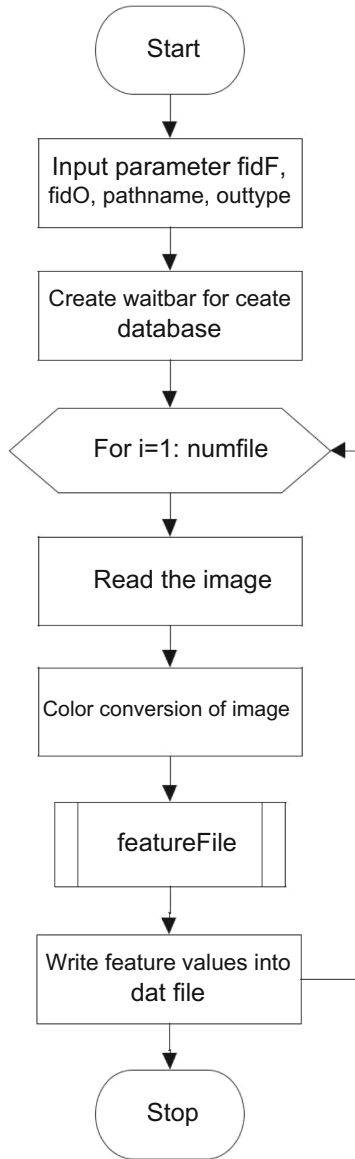
Fig. 4. Feature extraction

PCA applied to image processing in this particular project to reduce the orientation values of HOG that are stored. This helps in improved performance and faster detection and analysis of objects in the frames.
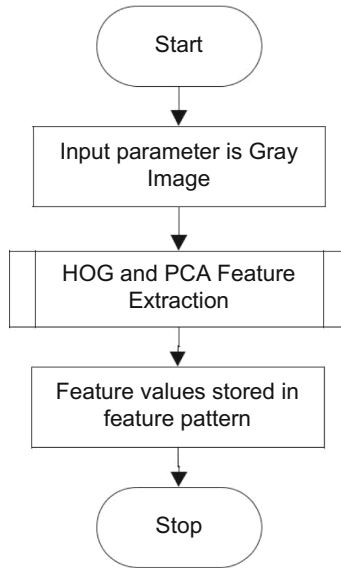
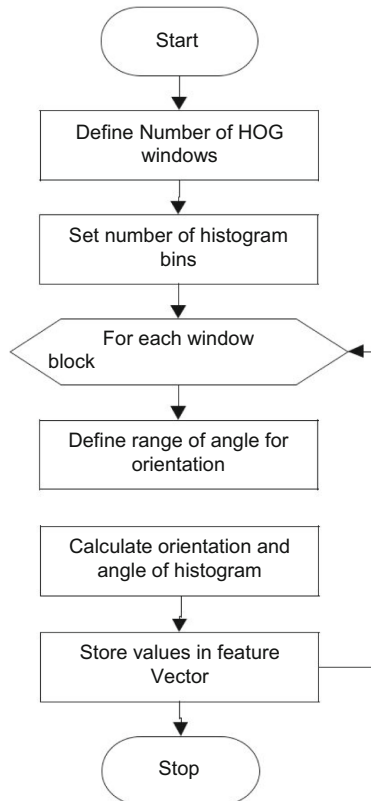**Fig. 5.** FeatureFile subroutine that performs HOG and PCA feature extraction/reduction

Fig. 6. HOG feature extraction

3.4   Support Vector Machine Classifier

Support Vector Machines (SVM) is the most popularly used classification method [11]. It has a wide variety of applications such as classification of text [8], recognizing the facial expression [10], in the analysis of genes [7] and many others. SVM is one of the technique for constructing a linear classifier, which produces a classifiers based on the theoretical foundation [12].

3.4.1   SVM Classification Using "KERNAL TRICKS"

Kernel techniques is used in SVM to solve linearly inseparable problems which transforms data to a high dimensional space. But training and testing large data sets consumes more time. Hence, we can train and test the large data sets using linear SVM without kernels.

The Experimental results proves that the proposed method is beneficial for large-scale data sets. Hence the proposed method can be successfully applied to natural language processing (NLP) applications [6].

3.4.2   Radial Basis Kernel Function for SVM Classification of Images

Radial basis kernel function (RBF) is the most commonly used kernel function in machine learning. RBF kernel is used for various kernelized learning algorithms. In particular, it is most commonly used SVM classification [7].

Consider two samples $\mathbf{x}$ and $\mathbf{x}'$, which represents a feature vectors in some *input space*, The RBF kernel is defined as

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)$$

$\|\mathbf{x} - \mathbf{x}'\|^2$ indicates the squared Euclidean distance between the two feature vectors.
σ is a free parameter. An equivalent, definition involves a parameter $\gamma = \frac{1}{2\sigma^2}$:

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma\|\mathbf{x} - \mathbf{x}'\|^2)$$

As shown in Fig. 7, we use SVM training loading the extracted features which are stored in features.dat. The output frames loaded in output.dat is compared to the trained features that are stored in the knowledge base.

3.5   Data Flow

The following image represents the data flow diagram of the proposed work. The architecture as described in the image consists of a testing phase and a training phase.

In the training phase, the training video is preprocessed, features are extracted and added into the knowledge base. This dataset is classified into either normal or abnormal
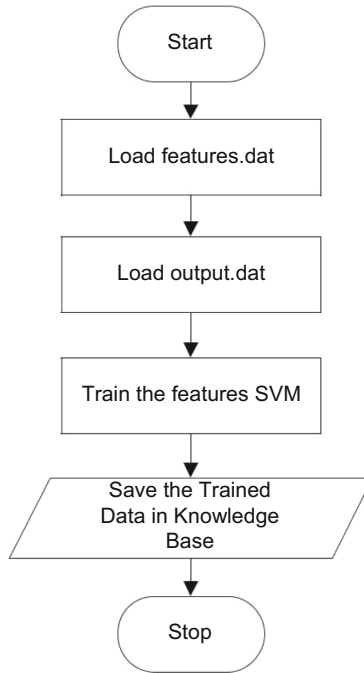
Fig. 7. SVM training

event using SVM Classifier. Even in the testing phase, the video is preprossed and the features are extracted. Later, The SVM classifier, classifies each frame to be either "Normal" or "Abnormal".

The Fig. 8 shows the two phase of data flow diagram. Initially, the image is read in RGB Frame format. During preprocessing, the frames are converted to gray scale format. This resultant format is used for extracting features using HOG and PCA methods. The extracted features are trained using SVM and the trained data is stored in the knowledge base and used for classifying the data under the SVM.

In the testing phase, the input video is broken down into frames and preprossed using morphological operations. Then the foreground objects are extracted by performing the background substraction. This is done by finding the Absolute difference of the frames. Later Feature extraction takes place on the noise free images. The feature extraction procedures include HOG & PCA methods. These features fall under a classification trained under the SVM, as explained earlier.

In the current project, the SVM classifies data as either Normal or Anomalous (Abnormal). The detected region and recognition result is displayed along with the frame.

Fig. 8. Data flow diagram

4   Result Analysis

4.1   Discussion

The dataset used in this project was shot in a Canon D750 camera at 55 mm focal length. The video resolution is adjusted to 380 × 240 pixels at a frame rate of 15. Three different scenarios are evaluated in the datasets. The first two datasets were used to depict anomalous event such as punching and kicking. This also depicts normal scenarios of handshake and hug. The third dataset is of a typical fall detection even that may occur at any old age homes.

4.2   Performance of Our System

The system, when tested with the datasets mentioned above, an error rate of **0.27** was obtained. The error rate was calculated using the formula:

Error Rate = No of False Negatives/(No of False Negatives + No of True Positives)

Here False Negatives refer to the anomalous events that were not identified. True Positives are the anomalous events that were identified correctly.

Table 1 provides a detailed overview of the performance. The SVM classifier was successfully able to classify most of the event as normal or abnormal. The dataset was provided as frames to the classifier with specification for classification process (Fig. 9).

Table 1. Comparison of error rate for the three datasets

	Dataset 1	Dataset 2	Dataset 3	Total
True positives	33	16	20	69
False positives	1	0	7	8
False negatives	8	5	12	25
Err. rate	0.20	0.24	0.38	**0.27**
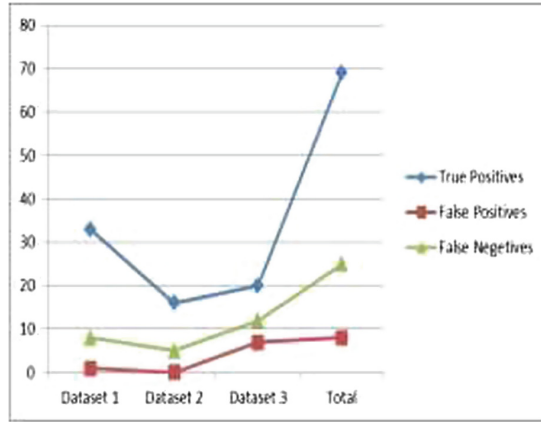
Fig. 9. Performance comparisons for three datasets

4.3   Discussion of Result

The dataset 3 produced the highest number of false negatives because the person of interest's movement and placement towards the camera made it difficult for the classifier to identify the event. The feature extraction was also not optimal due the constant changes the objects of the environment.

It is observed that when this dataset was removed from evaluation the error rate was reduced to only **0.21**. This shows the importance of a static background for our system, which is also a drawback and suggestion for future enhancement.

Table depicts the performance when dataset 3 is removed (Table 2).

Table 2. Comparison of error rate without the challenging video

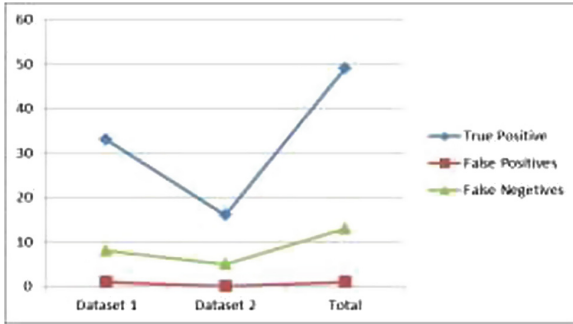|	Dataset 1	Dataset 2	Total
True positive	33	16	49
False positives	1	0	1
False negatives	8	5	13
Err. rate	0.20	0.24	**0.21**

Fig. 10. Performance comparisons without the challenging video.

Fig. 11. Dataset 1: shows sequence of normal and anomalous events

The performance of the system is calculated, with and without the challenging video in Figs. 11 and 12. Figure 13 provides a detailed description of the three datasets (Figs. 10, 14, 15, 16 and 17).
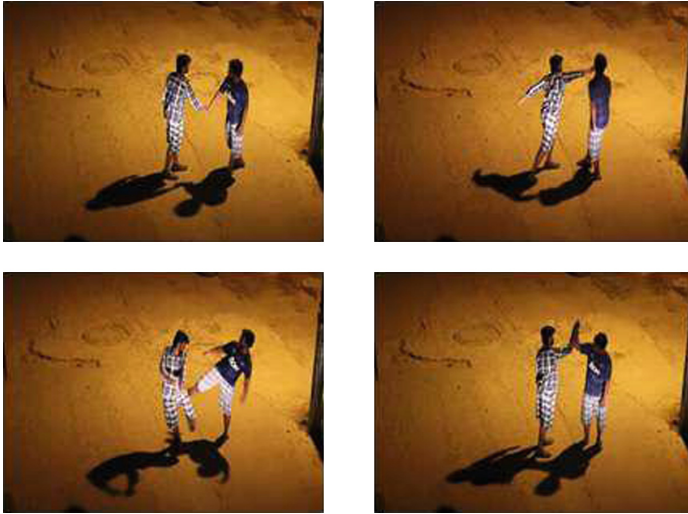
Fig. 12. Dataset 2: handshake, punch, kick, high five events

Fig. 13. Dataset 3: false negative event followed by fall detection event
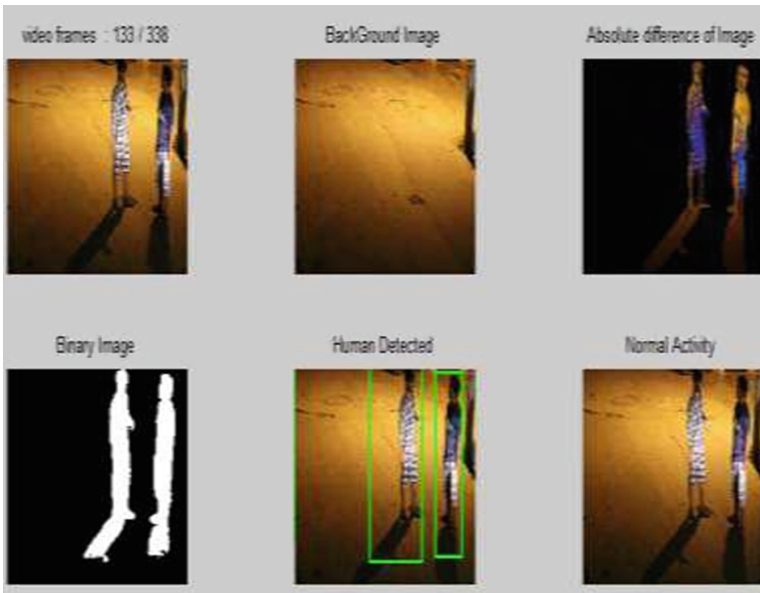
4.4   Output

Fig. 14. Normal event recognition as seen in dataset 1

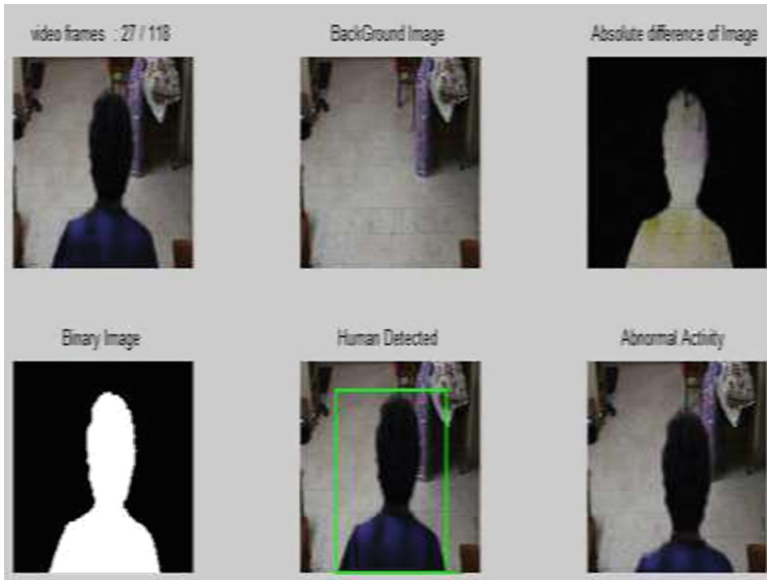**Fig. 15.** Anomalous (or abnormal) event recognition as seen in dataset 1

Fig. 16. False negative events in dataset 3

Fig. 17. Fall detection in dataset 3

5   Conclusion

The automated human event analysis in video surveillance system has become one of the most effective and attractive research topics in the area of computer vision and pattern recognition. The increasing computational power, provides a great environment for improving the existing systems.

The proposed work has provided the satisfactory results as expected. The dataset used in this implementation was taken and designed for a static camera. An error rate of 0.27 was achieved when tested with the given datasets. A better error rate of 0.21 was achieved when a challenging video was removed from the dataset. This method can be further improved and implemented for real time surveillance systems. Hence the study for anomalous event detection can grow further.

References

1. Andriluka, M., Roth, S., Schiele, B.: People tracking by detection and people detection by tracking. In: IEEE Computer Vision and Pattern Recognition (2008)
2. Leibe, B., Schindler, K., Cornelis, N., Gool, L.V.: Coupled object detection and tracking from the static cameras and moving vehicles. IEEE Trans. Pattern Anal. Mach. Intell. **30**(10), 1683–1698 (2008)
3. Li, Y., Huang, C., Nevatia, R.: Learning to associate: HybridBoosted multi-target tracker for crowded scene. In: IEEE Computer Vision and Pattern Recognition (2009)
4. Fradkin, D., Muchnik, I.: Support vector machines for classification. DIMACS Series in Discrete Mathematics and Theoretical Computer Science
5. Michel, P., Kaliouby, R.E.: Real time facial expression recognition in video using support vector machines. In: Proceedings of ICMI 2003, pp. 258–264 (2003)
6. https://www.quora.com/How-does-one-decide-on-which-kernel-to-choose-for-an-SVM-RBF-vs-linear-vs-poly-kernel answer by Charles H Martin, followed up on https://charlesmartin14.wordpress.com/2012/02/06/kernels_part_1/
7. Vapnik, V.: The Nature of Statistical Learning Theory, 2nd edn. Springer, Heidelberg (2013). https://doi.org/10.1007/978-1-4757-3264-1
8. Suriani, N.S., Hussain, A., Zulkifley, M.A.: Sudden event recognition: a survey. Sensors **13**(3), 9966–9998 (2008)
9. Lin, W., Sun, M.-T., Poovendran, R., Zang, Z.: Group event detection for video surveillance. In: IEEE, pp. 2830–2833 (2009)
10. Candamo, J., Shreve, M., Goldgof, D.B.: Understanding transit scenes: a survey of human behavior-recognition algorithms. IEEE Trans. Intell. Transp. Syst. **11**, 206–224 (2010)
11. Zaidenberg, S., Boulay, B., Garate, C., Chau, D.P.: Group interaction and group tracking for video-surveillance in underground railway stations. In: International Workshop on Behavior Analysis and Video Understanding (2011)
12. Zhang, Y., Ge, W., Chang, M.-C., Liu, X.: Group context learning for event recognition. In: Applications of Computer Vision. IEEE, pp. 249–255 (2012)
13. Zhang, D., Chen, F., Tong, C.: Particle motion based abnormal event detection in group-level crowd. J. Converg. Inf. Technol. **7**(14) (2012)
14. Arbat, S., Sinha, S.K., Shikha, B.K.: Event detection in broadcast soccer video by detecting replays. Int. J. Sci. Technol. Res. **3**(5), 282–285 (2014)

15. Tran, D., Yuan, J., Forsyth, D.: Video event detection: from subvolume localization to spatiotemporal path search. IEEE Trans. Pattern Anal. Mach. Intell. **36**(2), 404–416 (2014)
16. Kumar, A.N., Suresh Kumar, C.: Abnormal crowd detection and tracking in surveillance video sequence. Int. J. Adv. Res. Comput. Commun. Eng. **3**(9) (2014)
17. Pooka, N.S.: Suspicious group event detection for outdoor environment. Int. J. Mod. Trends Eng. Res. **0X**(0Y) (2015)
18. Berclaz, J., Fleuret, F., Fua, P.: Robust people tracking with global trajectory optimization. In: IEEE Computer Vision and Pattern Recognition (2006)
19. Chang, Y.-W., Hsieh, C.-J., Chang, K.-W., Ringgaard, M., Lin, C.-J.: Training and testing low-degree polynomial data mappings via linear SVM. J. Mach. Learn. Res. **11**, 1471–1490 (2010)
20. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Mach. Learn. **46**(1–3), 389–422 (2002)
21. Jolliffe, I.T.: Principal Component Analysis. Springer Series in Statistics, 2nd edn. Springer, New York (2002). https://doi.org/10.1007/b98835. XXIX, 487 p. 28 illus. ISBN 978-0-387-95442-4
22. Joachims, T.: Text categorization with support vector machines: learning with many relevant features. In: Nédellec, C., Rouveirol, C. (eds.) ECML 1998. LNCS, vol. 1398, pp. 137–142. Springer, Heidelberg (1998). https://doi.org/10.1007/BFb0026683
23. Paul, M., Haque, S.M.E., Chakraborty, S.: Human detection in surveillance videos and its applications - a review. J. Adv. Sig. Process. **2013**, 176 (2013)
24. Sivarathinabala, M., Abirami, S.: An intelligent video surveillance framework for remote montioring. Int. J. Eng. Sci. Innov. Technol. IJESIT **2**(2), 297–301 (2013)
25. Thida, M., Yong, Y.L., Climent-Pérez, P., Eng, H.-l., Remagnino, P.: A literature review on video analytics of crowded scenes. In: Atrey, P.K., Kankanhalli, M.S., Cavallaro, A. (eds.) Intelligent Multimedia Surveillance, pp. 17–36. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41512-8_2