Artif Intell Rev
DOI 10.1007/s10462-017-9545-7

Suspicious human activity recognition: a review

Rajesh Kumar Tripathi[1] · **Anand Singh Jalal**[1] · **Subhash Chand Agrawal**[1]

Abstract Suspicious human activity recognition from surveillance video is an active research area of image processing and computer vision. Through the visual surveillance, human activities can be monitored in sensitive and public areas such as bus stations, railway stations, airports, banks, shopping malls, school and colleges, parking lots, roads, etc. to prevent terrorism, theft, accidents and illegal parking, vandalism, fighting, chain snatching, crime and other suspicious activities. It is very difficult to watch public places continuously, therefore an intelligent video surveillance is required that can monitor the human activities in real-time and categorize them as usual and unusual activities; and can generate an alert. Recent decade witnessed a good number of publications in the field of visual surveillance to recognize the abnormal activities. Furthermore, a few surveys can be seen in the literature for the different abnormal activities recognition; but none of them have addressed different abnormal activities in a review. In this paper, we present the state-of-the-art which demonstrates the overall progress of suspicious activity recognition from the surveillance videos in the last decade. We include a brief introduction of the suspicious human activity recognition with its issues and challenges. This paper consists of six abnormal activities such as abandoned object detection, theft detection, fall detection, accidents and illegal parking detection on road, violence activity detection, and fire detection. In general, we have discussed all the steps those have been followed to recognize the human activity from the surveillance videos in the literature; such as foreground object extraction, object detection based on tracking or non-tracking methods, feature extraction, classification; activity analysis and recognition. The objective of this paper is to provide the literature review of six different suspicious activity recognition systems with its general framework to the researchers of this field.

✉ Anand Singh Jalal
asjalal@gla.ac.in

Rajesh Kumar Tripathi
rajesh.tripathi@gla.ac.in; rajeshkumar.tripathi@gmail.com

Subhash Chand Agrawal
subhash.agrawall@gla.ac.in

[1] Department of CEA, IET, GLA University, Mathura, Uttar Pradesh, India

Published online: 28 February 2017

Keywords Abandoned object · Theft detection · Fall detection · Accidents · Violence · Fire detection

1 Introduction

Suspicious Human Activity Recognition from Video Surveillance is an active research area of image processing and computer vision which involves recognition of human activity and categorizes them into normal and abnormal activities. Abnormal activities are the unusual or suspicious activities rarely performed by the human at public places, such as left luggage for explosive attacks, theft, running crowd, fights and attacks, vandalism and crossing borders. Normal activities are the usual activities performed by the human at public places, such as running, boxing, jogging and walking, hand waving and clapping. Now-a-days, use of video surveillance is increasing day by day to monitor the human activity which prevents the suspicious activities of the human.

An important chore of the video surveillance is to analyze the captured video frames for identifying unusual or suspicious activities in security-sensitive region of any country such as banks, parking lots, department stores, government buildings, prisons, military bases (Gouaillier and Fleurant 2009). Video Surveillance captures images of moving objects in order to watch assault and fraud, comings and goings, prevent theft, as well as manage crowd movements and incidents (Gouaillier and Fleurant 2009). In public places, human performs normal (usual) and abnormal (suspicious or unusual) activities. Normal activities are the usual activities that are not dangerous for the human world but abnormal activities may be dangerous for all over the world. Therefore, an intelligent surveillance system is required that can recognize all the activities and identify the more dangerous and suspicious activities performed by a human being.

There are two types of surveillance system-first is semi-autonomous in which video is recorded and sent for analysis by human expert. Non-intelligent video surveillance requires the continuous monitoring by human, which is very costly, problematic and also very difficult and challenging to watch over all the videos continuously by a guard to prevent the suspicious human activity. Therefore, a second Fully-autonomous surveillance system is required that performs low level tasks-motion detection, tracking, classification and identification of abnormal event.

The goal of the video surveillance is to develop an intelligent video surveillance to replace the traditional passive video surveillance so that abnormal activities performed by human being can be captured; and after analyzing, an alert can be produced through alarms, massages or some other techniques to prevent unusual activities.

There are several abnormal activities such as abandoned object detection, theft detection, health monitoring of patients or elder caring at home (i.e. fall detection), accidents or traffic rule breaking activities such as illegal U-turns, illegal parking and reckless driving detection on road, violence detection such as slapping, punching, hitting, shooting at public places and fire detection requires an intelligent surveillance system that can generate an alarm or alert automatically.

Now-a-days, explosive attacks are more dangerous activity for the public places performed by terrorists. Terrorists target to the more sensitive crowded public areas such as airports, bus stations, railway stations, government buildings and shopping malls. They come to these places and leave their luggage bomb for explosive attacks. It is very difficult for the security guards to watch over the crowded public places and identification of the suspicious objects.

Modern technologies cannot fully prevent such explosive attacks at public places, which are being investigated with cameras. A real-time intelligent video surveillance system can protect to the public places by detecting the left luggage without delay and through raising an alarm to alert the guards to remove that objects. Therefore, a fully automatic effective and efficient intelligent surveillance system is needed to be developed. The Intelligent Surveillance System can detect un-attempted stationary object at public place shown in Fig. 1a.

In recent scenario, snatch theft (shown in Fig. 1b) is a frequent abnormal activities performed by the chain snatchers which is very challenging to detect at public places. Snatch theft abnormal activity attracts the attention of the public, and it needs an urgent reaction to help the victim. To catch the victim, a real-time intelligent video surveillance is required at the public places.

An intelligent video surveillance is also demanded to automatic fall detection of elders at home and patients in hospitals. Mostly, worn-sensor based systems (Willems et al. 2009; Nguyen et al. 2009), are available in the market for the fall detection, these devices are mostly electronic devices that compel to the elder people either to put it into pocket or wear it on the wrist. Normally, these wearable fall detectors have manual help button or accelerometer to detect a fall. However, these wearable fall detectors have a few drawbacks. One of the weaknesses for the fall detectors is that the elderly people can forget to wear devices and help buttons are useless for those people who become unconscious after falling down. The modern advancements in the field of computer vision have brought new solutions to overcome these drawbacks. One of the main advantages of visual-based fall detection is that such system does not require a person to wear anything, and it is less disturbing in comparison to the wearable sensor. Moreover, computer vision system provides more information on the behavior of a person compared with the normal wearable sensors. With this, visual-based home monitoring system is able to provide information on falls and also other activities of daily living behaviors which are useful for health-care monitoring, such as mealtime, and sleep duration. A Human Fall detection image captured by an intelligent visual surveillance system can be seen in Fig. 1c.

An intelligent video surveillance is also demanded to monitor the traffic flow and identification of the behavior of the vehicles. Illegal parking (shown in Fig. 1d–e) causes the jamming of traffic on the road, reckless driving causes accidental injuries or death (shown Fig. 1f–g), traffic rule breaking activities such as illegal U-turns also causes accidents. The real-time recognition of such abnormal behavior on the road can save the life of injured people, it can prevent the illegal parking, as well the accidental injuries or death by providing the medical treatment immediately. Therefore, an intelligent surveillance system can be helpful for the people of the world.

Violence activities such as fighting, slapping (shown in Fig. 1h), vandalism, running people at the public places (shown in Fig. 1i) or schools and colleges are being monitored through the surveillance camera; and after complaint of the victim, captured video is investigated for the crime. However, violence activities cannot be prevented at the same moment. But, an intelligent video surveillance system can recognize such abnormal activity and produce an alarm to alert the Police of that area to stop any violence activity.

In general, the fire disaster (shown in Fig. 1j) frequently causes ecological and economical damage as well as death of many human beings. Therefore, real-time based fire detection and warning is very important. Currently, a lot of sensor-based systems are being used to detect the fire. These sensor detectors must be placed very close to a fire; otherwise fire cannot be detected and it cannot give the information about the fire growing rate, location, size, and so on. Therefore, such fire detectors cannot be successfully applied in open or large spaces. They are not always reliable because energy emission of non-fires or byproducts of combus-

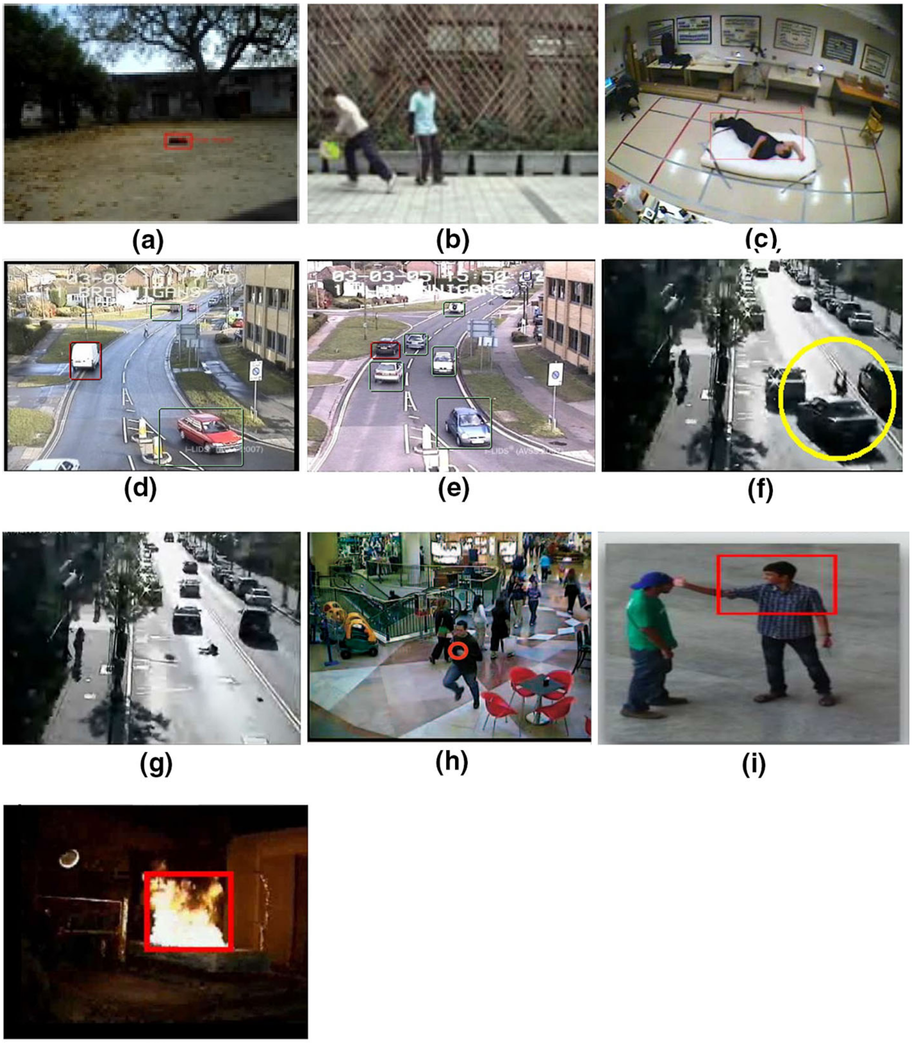

Fig. 1 **a** Abandoned Object (Tripathi and Jalal 2014), **b** Snatch Theft (Chuang et al. 2008), **c** Falling (Yogameena et al. 2012), **d**, **e** Illegal Parking on Road (Guler et al. 2007), **f**, **g** Accident on road (Candamo et al. 2010), **h** Person running in a mall (Adam et al. 2008), **i** Slapping (Penmetsa et al. 2014), **j** Fire Detection (Lai et al. 2012)

tion, which can be yielded in other ways, may be detected by misadventure. This usually causes false alarms. Infrared cameras compared with sensors are used by other fire precaution systems that are relatively reliable but leads to a high cost for surveillance. Therefore, the vision-based approach is becoming more and more interesting to provide more reliable information about fires.

To develop an intelligent surveillance system for recognizing the above mentioned abnormal human activities, many researchers have utilized the following general steps (Candamo et al. 2010; Dick and Brooks 2003):

Table 1 Related literature survey

Author	Paper
Hu et al. (2004)	A survey on visual surveillance of object motion and behaviors
Candamo et al. (2010)	Understanding transit scenes: a survey on human behavior-recognition algorithms
Poppe (2010)	A survey on vision-based human action recognition
Aggarwal and Ryoo (2011)	Human activity analysis: a review
Popoola and Wang (2012)	Video-based abnormal human behavior recognition—a review
Ke et al. (2013)	A review on video-based human activity recognition
Ziaeefard and Bergevin (2015)	Semantic human activity recognition: a literature review

Foreground object detection Background subtraction is a powerful mechanism to detect the change in the sequence of frames and to extract foreground objects (McHugh et al. 2009).

Object detection Object detection in the video frames is done through either the nontracking based approaches or tracking based approaches. Tracking based approach is employed to make the trajectory of an object over time by locating its position in every frame of the video (Yilmaz et al. 2006).

Feature extraction Shape and motion based features of the object are extracted through various algorithms for object identification and sometimes, its feature vector is supplied as input to the classifier.

Object classification Object classification is a mechanism to distinguish the objects available in the video. This process helps to make the distinction between different objects such as human, vehicle etc. There are different techniques to classify the objects such as Support Vector Machine, Haar-classifier, Bayesian, K-Nearest Neighbor, Skin color detection, and Face recognition.

Object analysis After recognizing the objects from the video through classification, activity analysis is performed to compare with the different threshold value to assure abnormal activity.

In the field of human activity recognition, several authors have discussed the progress of literature review. A few papers have mentioned in Table 1 which show the progress in the field of normal and abnormal human activity recognition. But there is very less number of literature reviews have been proposed in the field of suspicious human activity recognition. The contribution of this paper is to present the progress in the field of suspicious human activity recognition such as abandoned object detection, theft case detection, falling detection, accidents and traffic rules breaking detection, violence detection, and fire detection. The progresses in the literature review of the above mentioned suspicious human activities have been discussed with its general frameworks. Researcher of this field can get the more knowledge about the core technologies applied over the different steps to categorize and recognize the human activity.

The rest of the paper is structured as follows: Sect. 2 discusses the motivation and applications of suspicious activity recognition. Section 3 presents the issues and challenges in abandoned object detection, theft detection, falling detection, violence activity detection and fire detection. An overview of the progress in the past decade in the field of abandoned object detection, theft detection, falling detection, violence activity detection and fire detection discussed in Sect. 4. The general framework for suspicious activities detection is discussed in Sect. 5. Section 6 presents the Datasets and Evaluation measures used for abandoned object detection, theft detection, falling detection, accidents and illegal parking detection on road, violence activity detection and fire detection. Finally, the last section presents conclusion and future work.

2 Motivation and applications

Importance of the suspicious human activities recognition from video surveillance is to prevent the theft cases, leaving abandoned objects for the explosive attacks by terrorists, vandalism, fighting and personal attacks and fire in the different highly sensitive areas such as banks, hospitals, malls, parking lots, bus and railway stations, airports, refineries, nuclear power plants, schools, university campuses, borders etc. Intelligent video surveillance protects the following areas from suspicious activities (Yilmaz et al. 2006):

University campus and academic institutions Video surveillance is being used in university campuses and other academic institutions to monitor the activities of students for the safety of assets from theft and vandalism. It also helps to prevent the inappropriate behavior of the students and fighting among the students. It also monitors the perimeter of the university campus, school and academic institutions for the safety of the students and faculties. Video surveillance can be used at the time of examination to monitor the suspicious activity of the students in the examination hall.

Public infrastructure To save population and public infrastructures such as borders, laboratories, prisons, military bases, temples, parking lots; video surveillance is helpful to prevent the theft, vandalism, fighting and personal attacks, increasing crowd, explosive attacks.

Retail trade This is a growing market for the use of video surveillance to detect the suspicious human activity for both the internal such as warehouses, stores and external like parking lots security. Even the small shops are utilizing the cameras to monitor the human activities and to capture the video evidence in case of theft or an incident. In chain stores, much more sophisticated video surveillance systems are set up for centralized monitoring of different locations. Suspicious activity recognition from video surveillance helps to monitor employee fraud and theft, monitor wares and inventory, protecting material goods and infrastructures, protecting staff and clients, monitoring parking lots, vehicles, entries and exits, and emergency situations such as fire.

Airports Airports are high security sensitive areas where the safety of passengers, runway and airplane is the most important in any country. Real-time suspicious human activity recognition system from video surveillance provides high level security to such security sensitive areas.

Railway and bus stations The use of video surveillance at railway and bus stations plays vital role in case of monitoring platforms, routes, parking lots, rails and tunnels. These areas are the prime targets of the terrorists for explosive attacks by leaving a bag containing bomb. Suspicious activity recognition system from video surveillance can recognize the abandoned object and can alarm to remove it from public place for the protection of passengers, personnel and infrastructures.

Banking sector Video surveillance play an important role in banking sectors to provide the security. The presence of cameras prevents to committing the armed robbery and assault. Automated bank machines are prime targets for criminal acts. Surveillance camera helps to detect fraud, for example; the installation of a device to read the magnetic information on bank cards. Intelligent video surveillance can increase monitoring effectiveness in banking sectors. It provides monitoring to all the branches in order to detect suspicious behavior. In ATMs, it also helps to prevent theft cases.

Gaming industries and casinos Suspicious activity recognition from video surveillance can help to detect the cheating, heists, and other crimes. Since monitoring of casino requires watching the activity of human beings in a crowded environment, intelligent video surveillance is an interesting way of helping security personnel.

Hospitals Video surveillance can also be used in hospitals to monitor the patients at home to monitor elder people or children. It can even be found in ambulances to monitor a patient remotely. Video surveillance can monitor the activity of the patients in hospitals and can recognize the suspicious activity such as vomiting, fainting and other unusual activity of the patients.

3 Issues and challenges

To develop an intelligent video surveillance system for the automatic recognition of suspicious human activities; there are various issues and challenges (Yilmaz et al. 2006; Tripathi et al. 2013).

Illumination changes The moving object detection is difficult to process reliably due to dynamic variation in natural scenes such as gradual illumination changes caused by day–night change and sudden illumination variation caused by weather changes. Various illumination effects have been shown in Fig. 2.

Shadow of objects Shadow changes the appearance of an object, which creates problem to track and detect the particular object from the video. Some of the features such as shape, motion, and background are more sensitive for a shadow. Figure 3a shows the shadow with object that will change the shape of the object at the time of tracking.

Noise in the images Sometimes, waving tree branches creates noises that create the problem at the time of recognition of an object from the video.

More crowds To detect the object from more crowded area (shown in Fig. 3b) is very challenging task. In such situation, abandoned object detection, theft detection, violence detection is very difficult.

Partial or full object occlusions In video, sometimes, objects are occluded partially or completely. This creates a problem to identify the object correctly. Partial occluded examples are shown in Fig. 4a–b. In general, there are three types of occlusion which have been shown in Fig. 4c–e.

Blurred objects Segmentation and feature findings of blurred objects are very difficult to identify the particular objects. Figure 3c shows the blurred objects in an image which is very difficult to recognize.

Fig. 2 **a** Sudden illumination change by weather changes (pet, 2001). **b** Illumination in night due to the light effects (pet, 2007). **c** Illumination effect in day time (pet, 2007)

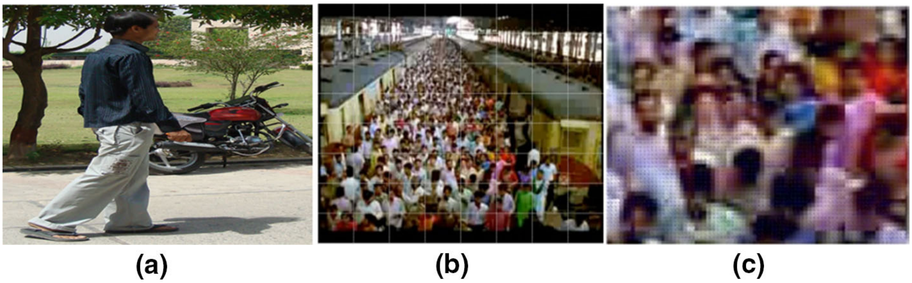

Fig. 3 **a** Shadow effect. **b** More crowd. **c** Blurred image

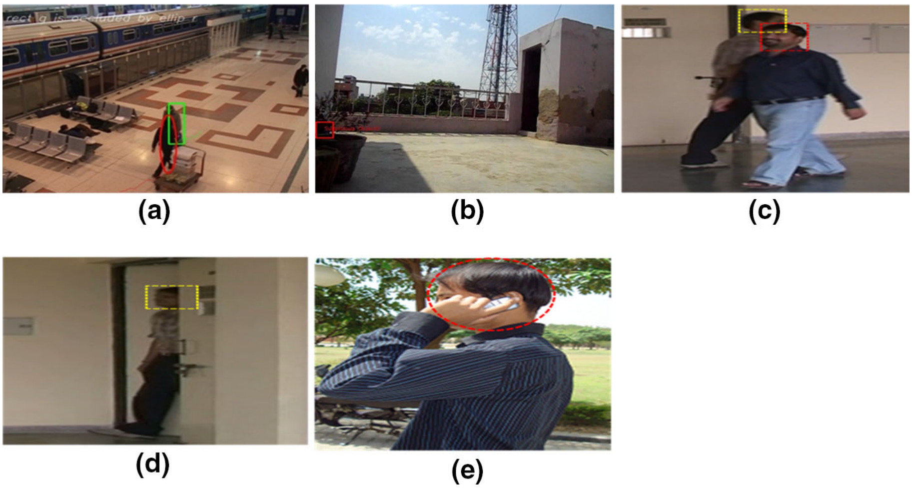

Fig. 4 **a** Partial occluded human beings (Zhou et al. 2010). **b** Partial hidden abandoned object in a flowerpot (Tripathi et al. 2013). **c** Occlusion with other object (Jalal and Singh 2012). **d** Occlusion with background (Jalal and Singh 2012). **e** Self occlusion (Jalal and Singh 2012)

Poor resolution To detect the foreground objects from videos having poor resolution is very challenging task. Object boundaries identification becomes very difficult that causes incorrect object classification.

Real-time processing The more challenging task is to develop a real-time intelligent surveillance system. The videos which have complex background, take more time to process it at the time of foreground object extraction and tracking of the objects.

Static object detection In abandoned object detection, static object detection is a challenging task through the background subtraction because this method detects only the moving objects as a foreground.

4 Researches in suspicious human activity recognition from video surveillance

This section covers the progress in the field of suspicious human activity recognition till date by categorizing them into abandoned object detection from static and moving camera, theft case detection, falling detection for elder caring, accidents and traffic rules breaking detection, violence detection and fire detection.

4.1 Research in abandoned or removed object detection from video surveillance

Abandoned object detection is very difficult in case of highly crowded area, fully occluded objects and sometimes partially occluded objects from single static cameras. Several researchers have worked to detect an abandoned or removed object from the video surveillance to protect the people and public infrastructure from the explosive attacks performed by terrorists. These Abandoned objects may be in any form such as any type of baggage, hidden object behind the wall or other objects, etc. Many works have been done in this field for single static cameras but very a few works have been done for the moving cameras. This section presents the progress in the field of abandoned or removed object detection from static and moving cameras.

Sacchi and Regazzoni (2000) presented a distributed video surveillance system to detect the presence of abandoned objects in unmanned railway stations. In this, an alarm issue is transmitted after recognizing an abandoned object to a remote control center that is located few miles far from the guarded stations. This system employed a direct sequence code-division multiple-access technique to ensure noise-robust and secure wireless transmission links between the remote control center and guarded stations. This system has been developed for monochromatic camera. Performance of this system can be improved in terms of false alarms and misdetections by using colored image sequences but processing time is increased with color image sequences which cause failure of abandoned object detection in real time. Foresti et al. (2002) developed a layered content-based retrieval of video-event shots referring to potentially interesting situations. This approach is able to detect, index and retrieve interesting video-event shots of human activities. Interesting events refer to potentially dangerous situations such as abandoned objects. This system is robust to partial or temporary occlusions between moving objects due to the long memory algorithm that recover object identities after the occlusions. The success rate in video shot detection is 95% for low complexity, 75% for medium complexity and 33% for high complexity and 71% in retrieval to detect abandoned objects. This system is not capable to work well in high complexity videos.

Lavee et al. (2005) developed a framework for analyzing a video for suspicious event detection. In this, low level features are extracted and an event representation for several overlapping subsequences is created for use in event detection. Then, newly created events are compared with a set of predefined events and classified by nearest neighbor algorithm.

Lavee et al. (2007) also developed a framework for detecting the suspicious event from video through the three steps- low-level feature extraction, event classification, and event analysis with one assumption that unlabeled video sequences are known to contain only one event.

Ellingsen (2008) proposed a system to detect dropped objects in which background is modeled through the mean pixel intensity, and standard deviation of each pixel. To detect moving objects in the scene, each frame is subtracted from the mean background image which results a foreground image containing one or more objects. Features such as center of mass, area, minor axis etc. are extracted to measure blobs and these features contain sufficient information to make a description of dropped abandoned object. An automatic classifier of events are required that works on feature vectors and learning mechanisms. Porikli et al. (2008) developed a pixel-wise method that employs dual foregrounds to find abandoned objects, illegally parked vehicles, stopped objects. This method adapts dual background models by using Bayesian update and evidence is obtained from dual foregrounds to achieve temporal consistency. In dual foregrounds, background learning rate can be changed to adjust the absorption of static objects into background. Static objects can be distinguished from the long term background and moving objects by multiple foregrounds of different learning rate. This system has low computational load but it has one shortcoming i.e. it also detects the static person as an abandoned object. Miguel and Martínez (2008) proposed a new method to detect the unattended or stolen objects by fusing three detectors. The detectors are based on the contour, color and shape information for the analysis of the static foreground regions. This method fused low gradient, high gradient, and color histogram features to detect unattended or stolen objects and achieved good precision and recall rate with low, medium and high complex videos in comparison to the single detectors success rate.

Chuang et al. (2009) presented a novel method to detect the abandoned object. To detect suspicious human activity, kernel based object tracking has been used to track the object. Forward-backward ratio histogram and a Finite State Machine have been used to recognize the transferring conditions and provided the 100% accuracy for abandoned object detection. But, Ratio histogram method used 2^{12} color bins to identify the abandoned object which trades off between the efficiency and accuracy. Bhargava et al. (2009) presented a framework to detect the threat that utilizes spatio-temporal and contextual cues to detect baggage abandonment. In this framework, if an un-attempted object is discovered in the video, the system tracks to the previous video frames to recognize owner of that object. The person who carries the object into the scene and sets it down at the location it is found, is considered the owner of the luggage. A background subtraction has been used to detect foreground objects and k-nearest neighbor classifier has been used to classify foreground blobs in the frames as luggage and non luggage class. The k-NN classifier of this system failed to detect the baggage which was very close to a person sitting near it. This positional ambiguity can be removed by adopting fused information of multiple cameras. Li et al. (2009) presented a robust real-time system to detect removed and abandoned objects from video surveillance. This system introduced a pixel wise static object detector and a classifier based on the color richness for detecting removed and abandoned objects. This system has been evaluated on i-LIDS and CAVIAR video datasets. This system is robust to eliminate effect of small repetitive motions like waving trees and can handle occlusions. This system has one shortcoming i.e. it failed in the detection of one abandoned object from one video of CAVIAR in which left bag is occluded by the chair. It also produced false alarms with AB_HARD and PV_HARD video of i-LIDS.

Li et al. (2010) developed a video surveillance system to detect abandoned object that is robust to illumination changes and works well in low quality videos. This system generated two foreground masks with the use of short-term and long-term Gaussian mixture models. Radial Reach Filter (RRF) has been used for refining foreground masks to reduce the influence of illumination changes. Support Vector Machine classifier makes a distinction between left luggage and motionless standing people from static objects. This system failed to detect the object occluded by a person in S4 video sequence of PETS 2006. Evangelio and Sikora (2010) proposed a method to detect the static objects in crowded scenes in which two background models at different learning rates have been used with a finite state machine to recognize the different states. Singh et al. (2010) developed a suspicious activity analysis system through Gaussian mixture model, tracking, and Bayesian inference framework to analyze the events. This system eliminates shadow and noise from the video frames for analysis of objects. The system has been evaluated its performance on its own videos. Detection rate of this approach is 95.78% with 3.74% false acceptance rate. This system detects false positives and false negatives in video sequence due to swinging small objects, bag with little or no protrusion, protruding parts of clothing and camera noise.

Yang and Rothkrantz (2011) developed a method to recognize abandoned object using tracking system. This method used image segmentation, creation of blob, tracking and labeling algorithms to detect abandoned object. This method also used blob splitting and increasing distance between them. If separated blob has no skin color and static, it is considered as an abandoned object. Therefore, the system may fail to detect the skin color abandoned object. Tian et al. (2011) developed a new framework to detect abandoned or removed objects robustly and efficiently by using complement of tracking to reduce the false alarms. This framework employed cascade classifier to detect near-field human whose face is visible, wavelet transform and adaboost learning for shoulder detector of mid-field people whose face is not clearly visible due to poor resolution and pixel height for far-field people. This framework detected a static person as an abandoned object in S3 video sequence of PETS2006 dataset.

SanMiguel et al. (2012) developed a novel approach to discriminate between abandoned and stolen objects based on color contrast along contour of the object at pixel level from video surveillance. Tian et al. (2012) developed a new framework to analyze the foreground. This framework employed phong shading model to handle quick lighting changes, region growing and edge energy methods to classify removed and abandoned objects. This approach utilized a feedback mechanism of interactions between tracking and BGS to handle stopped and slow moving objects to improve the tracking accuracy. The system fails in low contrast situations where the color of the object is very similar to the background, e.g., black bag on a black background. Fan and Pankanti (2012) developed a system to detect abandoned object of large scale to achieve low false positive rates. This method combined structure similarity, region growing, local ternary patterns and phong shading model to extract the features and libSVM to analyze foreground. Abandonment analysis reduces false positives 3% on AB-L2 and 6% on AB-L1dataset. The system generates more false positives on ABL1 dataset. Zin et al. (2012b) proposed a system based on probability to detect abandoned objects in video surveillance. This system employed multiple background models which works better in comparison to the single and dual background model approaches in the detection of abandoned object accurately by handling quick lighting change adaptation and removing shadows. A rule based algorithm classifies abandoned object and static person correctly in crowded environments and also detects very small abandoned objects in low quality videos. Prabhakar and Ramasubramanian (2012) proposed an integrated approach to track an abandoned object and unknown objects. This approach applied frame differencing for background subtraction and morphological filtering for noise removal. Tracking creates the trajectory of each blob and analyzed to decide an abandoned object. Fern'andez-Caballero et al. (2012) presented a work to monitor human activity by local and global finite state machines. Image features based on motion are linked explicitly to a symbolic notion of hierarchical activity through many layers of more abstract activity descriptions. At a low level, atomic actions are detected and fed to hand-crafted grammars for detecting activity patterns of interest.

Chitra et al. (2013) proposed a novel framework to detect occluded object based on blob. Proposed framework detects the occluded objects in video sequences in crowded environment. This framework applied histogram of oriented gradients descriptors (HOG) and support vector machine to detect the pedestrians. This framework provides detection accuracy 80% for high occlusion video sequences. Maddalena and Petrosino (2013) proposed a neural based framework for the static objects and moving object detection. This approach contribution concerns a 3-D neural model for image sequences that automatically adapts the scene changes. This model enables the segmentation of stopped foreground objects against moving foreground objects by handling occlusion. This framework detected all static objects truly excepting AB-hard complex video sequence of i-LIDS. Fan et al. (2013) presented a ranking technique for large scale surveillance to detect the abandoned object with false reduction. In this approach, HL-RANK is a high level attributes ranking technique, worked well but failed to detect one bag of PETS 2006. This system has two false positives due to failures of the tracking approach. Tripathi et al. (2013) applied contour features to detect static objects in the scene and edge based method to detect the partially or fully visible human and non-human object. Non-human stationary object is analyzed for specific time duration and after deciding object as abandoned, an alarm is raised to alert the security.

Pavithradevi and Aruljothi (2014) presented a framework to detect the abandoned object from the video sequence captured from the colored camera. Foreground object extraction has been performed using background subtraction and noises have been removed by using Gaussian filtering with color and gamma correction. Merging and splitting of the objects in crowd are identified with staged matching method. This framework used support vector machine and adjacency matrix based clustering to identify the action of the public in crowd. The features have been extracted by using Gabor algorithm and histogram of gradient. Object direction and inter-object motion features detected the suspicious behavior.

Nam (2016) used spatio-temporal features to detect abandoned and stolen objects in crowded scenes on real-time. Adaptive background modeling has been used for the removal of ghost image and stable tracking. Spatio-temporal relationship is determined between moving human and suspicious object to detect abandoned or stolen object. This method employed a vector matching algorithm to detect partial occluded object and also employed a tracking trajectories to reduce the false alarms. This system can be improved by calculating parameters and threshold automatically using incremental learning rate.

In 2010, one research work has been also proposed for the abandoned object detection from moving camera. Kong et al. (2010) proposed a novel framework for the detection of non-flat abandoned objects from moving cameras with the help of reference video and target video. In this, reference video is recorded from moving camera when there is no suspicious scene in video. The target video is recorded from a camera of the same route. Author has used GPS information to align the two videos to find the corresponding frame pairs. Intersequence alignment highlights all the possible suspicious area by setting a threshold on the normalized cross correlation image of the aligned frame pair of intersequence. Intrasequence geometric alignment and a local appearance comparison between two aligned intrasequence frames remove false alarms in flat areas, remove false alarms caused by high objects, and a temporal filtering step validate the existence of suspicious objects. This system detects 21 suspicious objects out of 23 from 15 video sequences. This system has one weakness that it fails in case of those videos that has almost flat object.

4.2 Research in theft detection from static camera

In the previous section, several researches have been done for the detection and removal of objects. Object removal is also considered as theft case. In this section, we discussed few research work done to detect the chain snatching, robbery in banks, transferring of an abandoned object.

Akdemir et al. (2008) presented a systematic approach to recognize the human activities in banks and airports based on ontology. Author utilized five criteria to design clarity, ontology-coherence, minimal encoding bias, extendibility, and minimal ontological commitment. This system has been evaluated on six bank videos in which four videos consist of bank robbery and two videos consist of normal human activity. Moving objects have been tracked using color based appearance and motion. In this, single-threaded ontology correctly classifies three robbery scenarios but there is one drawback of this system that it fails on one video involving two robbers. Chuang et al. (2008) presented a fuzzy c-means algorithm based on ratio histogram to detect the suspicious activity. The method used GMM to segment the suspicious activity accurately. In this, conventional histogram ratios have been used to detect the object and fuzzy color histogram to deal with color similarity problem. By tracking the transferring conditions, unusual activities are identified.

Chuang et al. (2009) used Forward- backward ratio histogram and a Finite State Machine to recognize the robbery case. The method detected 96% cases of the robbery but forward and backward ratio histogram used 2^{12} color bins to identify the robbery bag completely which trades off between the accuracy and efficiency.

Ibrahim et al. (2010) proposed an approach to compute optical flow of video sequence to detect the snatch theft in pedestrian crowd movement. In this, features have been extracted from the computed optical flow of the video sequence. The event classification is based on the distribution of the optical flow vectors before and after the events using vector matching and SVM classification. The algorithm has a good detection rate of snatch theft events.

Ryoo and Aggarwal (2011) applied a stochastic representation scheme to represent group activities and also developed new hierarchical algorithm for the probabilistic recognition. This algorithm uses probability distribution sampling for detecting a group of thieves stealing an object from another group and a group assaulting a person. This system worked well in case of fight in group–group interaction category but failed in two cases of fight in intra-group interaction. Therefore, an automated learning of group activities can improve the performance of the system.

Ibrahim et al. (2012) presented two-stage decision process by extracting information from optical flow to detect snatch theft abnormal activity from video surveillance automatically. In first step, optical flow result screens the scene for potential criminal activity. After detecting potential crime scene, second stage uses flow pattern statistics to analyze for deciding whether a snatch theft event has occurred or not. Sujith (2014) presented a multiple object detection system and recognition of abnormal behavior to prevent the ATMs crime. To detect human, approach utilized features while classifier should be used for detecting human. In case of partial occlusion, this system may fail to detect the human.

4.3 Research in health monitoring from static camera

This section covers the progress in the field of patient caring in hospitals and caring elder people at homes to assist them independently. Few researchers have worked in this field.

Nasution and Emmanuel (2007) presented an evidence accumulation technique with classifier to detect and record falls as well as other posture based events. The method segments moving objects using adaptive background subtraction approach. In this, adaptive characteristics have been removed to prevent the inclusion of static human as background. Vertical and horizontal histograms of extracted foreground objects and angle between last standing postures with current foreground bounding box have been used as a feature set. Extracted features have been passed to the classifier. Finally, this method uses k-NN classifier and speed of fall to infer the real falling events. Use of k-NN classifier with multiple posture templates has recognition rate of about 90%. In this system, tradeoff is that correct output response will be delayed for an average of 8 frames with evidence accumulation technique. Standing event is wrongly detected as sitting when the segmented silhouette is disturbed by the shadows.

Zhou et al. (2008) developed a framework for automated activity analysis, visualization, and summarization for eldercare video monitoring. At the object level, human detection, silhouette extraction, and tracking algorithm for indoor environments is constructed. At the feature level, an adaptive learning method to estimate the physical location and moving speed of a person from a single camera view without calibration is developed. At the action level, hierarchical decision tree and dimension reduction methods for human action recognition is explored. Thome et al. (2008) proposed a real-time multi-view fall detection system, in which motion is modeled by using layered hidden Markov model because the single view motion analysis is limited by pose classification step that may fail to detect fall direction when people are very close to optical axis. The algorithm detects, tracks, and extracts features independently in each view. The approach performs posture classification using fusion unit. Then, this fusion unit merges the posture analysis to provide a standing or lengthened pose classifier that is efficient in unspecified viewpoints and falling directions. From the pose likelihood estimation, LHMM is used to manage the inference performed by all the cameras jointly. This association deals with sudden changes and is robust to low-level errors. Fall detection rate with single view is 82% while it has been improved with the use of two view system. The robustness of the system can be improved by incorporating good cooperation between views for low level step of algorithm. Foroughi et al. (2008b) developed a novel approach to detect human fall based on human shape variation. In this approach, projection histograms of the segmented silhouette, best-fit approximated ellipse around the human body, and temporal changes of head pose provide a useful cue to detect distinct behaviors. Extracted feature vectors are fed to a multi-class Support Vector Machine for precise classification of motions and determination of a fall event. This approach considers wide range of motions consisting of normal daily life activities, unusual events and also abnormal behaviors. Reliable recognition rate of falling detection is 88.08%. This method cannot detect fall activity in case of multiple elderly people and also cannot handle occlusion.

Chen et al. (2010) proposed an approach which combines posture estimation analysis and motion analysis for the human shape analysis to detect falls. Liu et al. (2010) proposed a falling detection system in which statistical scheme and vertical projection histograms of the silhouette image has been used to reduce the effect of upper limb activities of human body. This approach used the k-NN classification to classify the postures using the difference and height-width ratio of human body silhouettes bounding box. The k-NN classifier and the critical time difference are used to detect fall incident events. This system has fall detection and lying down event detection rate is 84.44%.

Khan and Sohn (2011) developed a system for elderly care monitoring to detect the six abnormal human activities such as chest pain, forward fall and backward fall, faint, vomit and headache from elderly people's daily life. Binary silhouette of the human being is extracted using the probability density function of Gaussian then features of silhouettes are extracted through R-transform. KDA discriminate between the different classes of human activities. HMM is used for training and recognizing the activities with average recognition rate 95.8%.

In binary silhouette, discrimination between the body parts is not observable. The depth silhouette can overcome this limitation of binary silhouette. Rougier et al. (2011) proposed a new method to detect an unusual event human fall through the analysis of human shape deformation in a video sequence. This method used a shape matching technique to track the silhouette of human along the video sequence. Then, the shape deformation is quantified from segmented silhouettes based on shape analysis methods. Finally, human falls are detected using a Gaussian mixture model with shape analysis methods such as procrustes distance and mean matching cost. In some specific conditions, shape analysis methods do not work better.

Yogameena et al. (2012) developed a method to detect the fall by analyzing human shape deformation during a video sequence. This method used Relevance Vector Machine to detect the fall of an individual person based on the results obtained from torso angle through skeletonization. Liu and Zuo (2012) proposed a framework to improve the algorithm to automatic fall detection. The algorithm used three features- effective area ratio, human aspect ratio, and center variation rate to prevent misjudgments. The framework used video of indoor area in which human is far 5–10 m from the camera.

Chua et al. (2013) proposed a visual based fall detection approach with low computational complexity for the analysis of human shape. Median filtering method has been used for the background subtraction. Human body has been detected in three points head, body and legs. The bounding box of the foreground blob is divided into three portions and then centroids are calculated to draw two lines. Each line represents to the distances and orientations of the human body. Ratios of the line distance of two consecutive frames are compared and orientation difference is computed to analyze the body shape of the human. This approach has 6.7% false alarm rate. This system failed in detection of two fall incidents because human body of the person was in a straight line and its ratio distance was computed only 1. Two crouch-down activities were also detected as fall because of the sudden changes in the ratio of distances.

Wang et al. (2016) presented a framework for fall detection system based on automatic feature learning. The training set is formed by using different frames including humans from video sequences of different views. Then, a label of each frame is predicted after training PCANet by all samples. Based on the predicted results of trained PCANet model, an action model is obtained by SVM with the predicted labels of frames in video sequences.

4.4 Research in accidents or illegal parking detection on road from static camera

This section covers the progress in the field of transportation system that automatically monitors the traffic flow and identifies behavior of the vehicles. Few researchers have worked in this field to detect the accidents, traffic rule breaking activities such as illegal U-turns, illegal parking and, reckless driving.

Kamijo et al. (2000) developed an occlusion handling algorithm utilizing spatiotemporal Markov Random Field for traffic images at intersections. The system learns the different event patterns of behavior of each vehicle in the HMM chains and then current event chains are identified using the output of tracking system. This system tracks multiple vehicles at intersections with occlusion and clutter effects at success rate of 93–96%.

Guler et al. (2007) proposed a system to detect stationary foreground objects such as abandoned bag and parked vehicle from video. Author employed a new video tracker i.e. tunnel vision tracker which is also a moving object detector and tracking framework. The main layer of Tunnel vision tracker performs the tracking of moving objects and very fast spatial based detection while the second layer is responsible for the detailed edge, color and region analysis of the objects for higher level tasks; and the purpose of the scene description layer are to produce dynamic background for the scene. The performance has been evaluated on i-LIDS and AVSS-07 datasets for detecting and producing the alarm for parking vehicles in no-parking zones and abandoned object without owner. This system detects vehicles in no parking zone with a small error with ground truth while it has large error in night time due to bright headlights of the vehicles.

Lee et al. (2009) proposed a system for detecting illegally parked vehicle in outdoor environments in real-time. The method employed an image projection technique that reduces the dimension of the data for reducing the computational complexity of segmentation and tracking processes. This system is capable to detect two illegally parked vehicles but failed to detect one illegally parked vehicle due to the arrival of two vehicles in no parking area together and parked also very close to each other.

Jiang et al. (2011) proposed a context aware method to detect anomalies. Three different levels of spatiotemporal contexts are considered through tracking all moving objects in the video. Frequency-based analysis is performed to automatically discover regular rules of normal events. Events deviating from these rules are identified as anomalies. The task of the method is to discover anomalous events from a collection of movement trajectories of vehicles. The system has detection accuracy 92.2% in atomic event, 86.6% in sequential event and 78.5% in co-occurrence event. Foucher et al. (2011) presented a system to detect the three suspicious events at airport that are person running, a person pointing with hand and person leaving an object. To detect the running person, the system adopted a non-parametric approach and accumulates the velocity of tracked object for a long period of time using Gaussian kernel. For the detection of the object put on the floor, long-term and short-term background modeling has been used through Mixture of Gaussians. The approach detected the pointing event based on group of significant spatio-temporal corners in $3 \times 3 \times 3$ cell compound features. The system has been evaluated on a 144 hours video corpus as part of the TRECVID2010 competition. This system generated a large number of false alarms because tracker is noisy and track fragmented blobs. Cui et al. (2011) proposed a method to detect an abnormal event based on local features for traffic surveillance video. Firstly, moving foreground objects are detected and affined with morphological operations. Then each foreground region's area, width-height ratio of outside rectangular, shape factors such as ellipse eccentricity, and pixel moving velocity vector are extracted. Based on these features, the regions are classified into different groups as pedestrian, vehicle or noise region, and their behavior is classified using velocity distribution and trained local features distribution map. Finally, a simple classifier is used to determine states of objects are normal or abnormal. With the rapid development of Intelligent Traffic Surveillance, low complexity and low level abnormality detection method is well fit in early alarm of distributed surveillance system. A new framework based on real-time to detect the traffic accidents using Histogram of Flow Gradient and logistic regression modeling have proposed in Sadeky et al. (2010), Sadek et al. (2010). Benezeth et al. (2011) developed a method for abnormal activity detection using low-level features. In this, illegal U-turns of the vehicle and dropped abandoned baggage have been detected by using co-occurrence matrix and statistical model can be estimated from training video sequence. Markov Random field is the statistical model which is very simple model that accounts for the correlation between time and space pixel activity.

Elhamod and Levine (2013) proposed an automated real-time system to recognize suspicious activities such as fighting, fainting, loitering and abandoned and stolen objects in public transport areas. The system is a complete semantics based behavior recognition approach that depends on object tracking based on color histogram. Codebook background subtraction method has been used to detect the foreground objects. Experiment has been carried out on CAVIAR dataset with the detection precision rate 93% in left bag, 89% in leftbag pickup, 63% in fight one man down video. This system failed to detect abandoned objects in PETS2006S5C3 video sequence because of failure in tracking and classification of object.

4.5 Research in violence activity detection from static camera

This section discusses the work done in the field of Violence activities detection such as vandalism, fighting, slapping, punching, hitting, shooting, peeping etc.

Adam et al. (2008) presented a real-time non-tracking based algorithm for unusual activity (i.e. person running in a mall) detection which is robust and works well in crowded scenes. Algorithm of this system monitors low level measurements in a set of fixed spatial positions instead of tracking to objects. Lack of sequential monitoring is the main limitation of this algorithm.

Wiliem et al. (2012) presented an automatic suspicious behavior detector which utilizes the contextual information. The three main components, a data stream clustering algorithm, a context space model, and an inference algorithm of the system; utilizes contextual information to detect the suspicious behavior. A data stream clustering algorithm enables to the system to update the knowledge continuously from the incoming videos. Inference algorithm combines both the contextual information and system knowledge to infer the decision. The system used two datasets-23 clips of CAVIAR dataset and 2 clips from Z-Block dataset of Queensland University of Technology. This system AUC is 0.778 with 0.144 errors. Ghazal et al. (2012) developed a novel method to detect the vandalism such as theft and graffiti in predefined restricted areas through the videos. The method applied additive Gaussian noise power and background model for the segmentation. A frame differencing is applied in between the current frame and background model. After this process, a low pass filter has been used with adaptive thresholding and morphological edge detection and contour tracing has been utilized to find the color histogram and area key features. Shape and motion features have been used for the tracking. Occlusion and splitting are handled by using set of rules. The approach has frame rate 13 frames per second. Gowsikhaa et al. (2012) presented a real-time method to detect the suspicious activities from surveillance videos in an examination hall. In this, author tried to detect the head position to prohibit copy, entry of any new person in the hall, peeping into another students answer script, passing incriminating materials, and exchanging the seats by students from real-time video. The approach employed adaptive background subtraction with sequential and periodical background modeling to extract the foreground image. This system fails to handle the occlusion situation.

Penmetsa et al. (2014) proposed an autonomous unmanned aerial vehicle visual surveillance system to detect the suspicious human activities such as slapping, punching, hitting, shooting, chain snatching and choking using pose estimation, and appearance of body parts. The system used combination of face detector and upper body detector to improve the efficiency of human detection. Then, a cascade filtering has been used to speed-up the face detection. Hough orientation calculator has been utilized to classify the poses. Orientation features of the human pose is compared with the poses in the suspicious action dataset and it is flagged with the action which matches the best. The system can detect the multiple suspicious activities such as slapping, punching, hitting, shooting, chain snatching and choking with detection accuracy 77.78, 76.67, 79.59, 73.47, 78.26% respectively. This system increases the time complexity and leads to misdetections as the number of people increases in the video frames.

Tripathi et al. (2015) presented a framework to identify the unusual activities happening (money snatching, attack on the customer, fight with customer) at the ATM installations and raise an alarm during any untoward incidence. The method extracted the relevant features from videos by using MHI and Hu moments. In this approach, PCA has been used to reduce the dimension of features and SVM for classification. Analysis has been performed on the basis of MHI window size.

4.6 Research in fire and smoke detection from static camera

In this, we have discussed few research works in the field of fire and smoke detection to prevent the ecological and economical losses.

Chen et al. (2004) proposed a method to raise an alarm after fire detection from video. The method extracted fire pixels and smoke pixels using RGB model based on chromatic and disorder measurement. Decision function of fire pixels is mainly inferred from the saturation of R component and intensity. The realities of extracted fire-pixels are verified by both dynamics of growth and disorder. Based on iterative checking on the growing ratio of flames, a fire-alarm is raised when the alarm condition is met. This approach achieved fully automatic surveillance of fire accident with a lower false alarm rate. A classifier can be applied to improve the reliability of the system by training fire and flame features.

Töreyin and Dedeoglu (2005) developed an algorithm which detects moving pixels and then colored pixels are matched with fire color, if fire color is found then a Hidden Markov Model is applied spatially and temporally to detect whether the fire pixels are flickering or not. Töreyin (2007) employed HMM based flickering model and wavelet based contour modeling approach for fire detection. A weighted majority based online training method has been utilized to adapt the fire detection system to varying conditions in the environment. Celik et al. (2007) developed two models, first for fire detection and second for smoke detection. The first model used the fuzzy logic concepts to replace existing heuristic rules and make the classification more robust in effectively discriminating fire and fire like colored objects. The first model achieves correct fire detection rate up to 99.0% with a 4.5% false alarm rate. Second model used a statistical analysis which is carried out using the idea that the smoke shows grayish color with different illumination.

Gubbi et al. (2009) proposed a new approach to detect the smoke based on wavelets and support vector machine. The method used block based approach in which image is divided into 32×32 blocks. To extract the features, discrete cosine transform and wavelet transform has been used. Then, support vector machine has been used to classify. This approach has an excellent cross validation accuracy of over 90% with specificity and sensitivity of 0.89 and 0.9 respectively is obtained on forest fire videos.

Borges and Izquierdo (2010) proposed a new method for the fire detection that analyzes the frame-to-frame changes of specific low-level features to describe potential fire regions. These features are area, size, color, boundary roughness, surface coarseness, and skewness, within estimated fire regions. Flickering and random characteristics of fire make these features more powerful discriminator. The changes in these features are evaluated, and then results are combined according to the Bayes classifier to decide whether or not fire occurs in that frame. The proposed method has false positive rate 0.68% and a false negative rate 0.028%. Yuan (2010) presented a system for the fire detection and suppression automatically from video surveillance. Fire detection module used the spatiotemporal features such as color and motion in real-time by utilizing the sequential image processing. The fire suppression module consists of control device, mobile device and water gun. On-line experiments performed in a large space hall to show that the integrated system can detect fire in few seconds after the fire was ignited and the fire was suppressed rapidly. But, this system is not capable to detect fire in highly dynamic scenes, outdoor and colorful scenes.

Lai et al. (2012) proposed real-time based flame detection system. Foreground object is detected with the help of YCbCr color clues and motion detection. Background edge model is used to eliminate the noise; to avoid the noise of motion detection in different resolution videos. A fire object is determined by corner flicker rate, compactness, and fire growth rate. The experiment can be performed to any resolution video and complex scene, both outdoors and indoors, such as squares, where people walk around and vehicles pass by. This method can detect the fire accurately and exclude to the undangerous fire. Habiboğlu et al. (2012) proposed a vision based fire detection system that uses color, temporal and spatial information. The system divides the video into spatio-temporal blocks and extracted covariance based features from these blocks to detect fire. The extracted features are trained and tested using Support Vector Machine classifier.

Lei and Liu (2013) designed a structure to detect the fire in coalmines. This structure detects the potential region of fire using frame differencing of video and denoised by median filter. Flame region is extracted by color information. Bayes classifier is employed to recognize fire combined with the dynamic features. This method can greatly improve accuracy of early fire prediction in coalmine. Seebamrungsat et al. (2014) presented a fire detection system based on color conditions and fire growth checking. This system used HSV and YCbCr color models with specified conditions. This system utilized the HSV color model to detect information related to color and brightness and the YCbCr color model to detect information related to brightness because it can distinguish bright images more efficiently than other color models. Fire growth is analyzed and calculated based on frame differencing. The overall accuracy from the experiments is more than 90.0%. Dimitropoulos et al. (2015) proposed an algorithm for fire flame detection in real time which models fire behavior by employing various spatio-temporal features, such as flickering, color probability, spatial, and spatiotemporal energy, while dynamic texture analysis is applied.

5 A general framework for suspicious human activity recognition

In this section, we have presented general framework for abandoned object detection, theft detection, fall detection, accidents and illegal parking detection, violence detection and fire detection (shown in Fig. 5). Suspicious human activity recognition consists of the following important stepladders: Foreground object detection, tracking or non-tracking based object detection, feature extraction, classification; behavior analysis and recognition. Mostly researchers follow up these steps with different algorithms or approaches to improve the recognition accuracy.

5.1 Foreground object detection

Foreground object extraction from the video is an initial and important step of suspicious human activity recognition. Background subtraction is a powerful mechanism to detect the change in the sequence of frames and to extract foreground objects (McHugh et al. 2009). Foreground objects consists of moving objects and newly arrived objects in a video which becomes stationary after some time such as left luggage. But moving objects are considered as the foreground objects while static objects are considered as background of the video in background subtraction techniques. This concept simplifies the moving object detection from a video of static camera but difficult to detect newly arrived stationary objects.

Moving object detection can be performed based on two approaches- background modeling and change detection based approaches (Mukherjee et al. 2014). The change detection

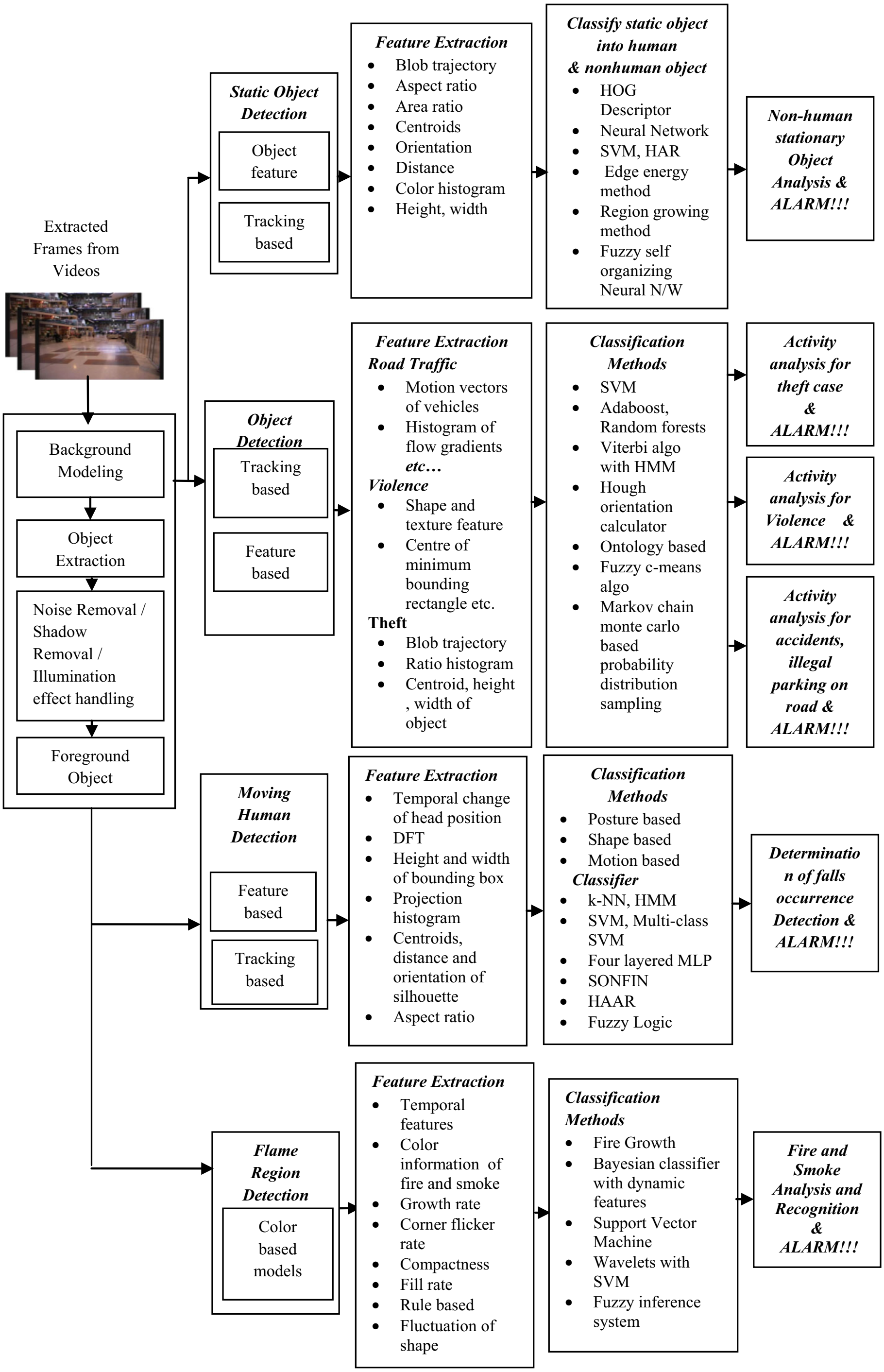

Fig. 5 General framework for abandoned object detection, theft detection, violence detection, accidents and rule breaking vehicles detection on road, falling detection, and fire detection

approaches find the difference between two consecutive frames to recover motion and apply post processing methods to recover the complete object. These methods are faster in respect to execution while lacking in accuracy. Approaches based on Modeling try to generate the background model using some spatial or temporal cues. A reasonably correct background model for the background can help to extract the foreground objects much effectively in comparison to the previous class of methods. These methods can range from very simple to highly complex in implementation and execution.

Newly arrived stationary objects in a video can be dangerous for the human and public place. To extract such stationary foreground objects through the background subtraction is difficult from surveillance video. Researchers applied different methods to extract and identify stationary objects.

5.1.1 Moving foreground object detection

In the last decade, several researchers have worked for the moving foreground object detection from the surveillance video. These methods help in extracting the human activities such as robbery, running crowd, vandalism, fights and attacks, crossing borders, punching, slapping, hitting, chain snatching, falling from the background of the surveillance video. Wren et al. (1997) presented an independently background modeling method at each pixel location using a single Gaussian. It has low memory requirements. Stauffer and Grimson (1999) proposed a most common background model based on Mixture of Gaussians. This method handles multi-modal distributions using a mixture of various Gaussians. Proposed technique cannot model accurately to the background having fast variations with the few (3–5) Gaussians. To solve the previous problem, Elgammal et al. (2000) developed a non-parametric model to model a background which is based on Kemel Density Estimation (KDE) on the buffer of the last n background values. KDE guarantees a smoothed, continuous version of the histogram. Lo and Velastin (2001) proposed temporal median filter background technique. In this technique, author used median value of last n frames as background model. Cucchiara et al. (2003) presented an approach based on mediod filtering, in which mediod of the pixels are computed from the buffer of image frames. Piccardi (2004) presented a review on seven different methods- Temporal median filter, Mixture of Gaussians, Running Gaussian average, Sequential KD approximation, Kernel density estimation (KDE), co-occurrence of image variations, and Eigenbackgrounds based on accuracy, speed, and memory requirements.

Bouwmans (2014) provided a complete survey of traditional and recent background modeling technique to detect the foreground objects from the static cameras video. Background subtraction is a very common technique for the segmentation of foreground objects in video sequences captured by a static camera, which basically detects the moving objects from the difference between the current frame and a background model. In order to accomplish good segmentation results, the background model must be regularly updated so as to adapt to stationary changes in the scene and to the varying lighting conditions. Therefore, background subtraction techniques often do not suffice for the detection of stationary objects and are thus supplemented by an additional approach (Evangelio and Sikora 2010).

5.1.2 Stationary foreground object detection

Suspicious activity recognition includes abandoned object detection to prevent the explosive attacks performed by terrorists. In video surveillance, background techniques consider moving objects as a foreground object and static object as a background. Therefore, when a newely arrived object becomes static then it is absorbed in the background. Several authors used different background subtraction techniques with dual background approach with different learning rate to extract the two foreground objects for detecting the stationary objects of the video.

Porikli et al. (2008) proposed a video surveillance system which uses dual foreground extraction from dual background modeling techniques. Therefore, short-term and long-term background models are created with different learning rates. Through this way, authors were able to control fastest static objects absorption by the background models and detect those groups of pixels which are classified as background by the short-term but not by the long-term background model. A weakness of the proposed system is that the temporarily static objects may also get absorbed by the long-term background model after a given time depending on its learning rate. Dual background modeling technique has been used to detect the abandoned object by several researchers in Porikli et al. (2008), Li et al. (2009, 2010), Evangelio and Sikora (2010), Bangare et al. (2012), Sajith and Nair (2013). Table 2 explores the different background subtraction techniques in the field of abandoned object detection, theft detection, health monitoring, abnormal activities detection on road traffic, violence activity detection and fire detection with the illumination and shadow handling techniques.

5.1.3 Noise removal, shadow removal and illumination handling methods

Detecting the foreground objects without noise, illumination effect, and shadow is a very challenging in area of computer vision. Noise creates problem in the identification of the object, illumination effect causes the false detection, and shadow changes the appearance of the object due to that object tracking becomes very difficult.

Several researchers have utilized different methods to remove the illumination effects, noise, and shadow from the video to minimize the false detections. In the abandoned object detection approaches (Miguel and Martínez 2008; Li et al. 2009; Bhargava et al. 2009; Prabhakar and Ramasubramanian 2012; Singh et al. 2010), researchers used morphological operations to remove the noises from the foreground frames. Tian et al. (2011) used texture information to reduce the false positives, and normalized cross correlation to remove false detection due to shadow. In Li et al. (2010), Radial Reach Filter has been used to reduce the false detected foreground due to the illumination changes and Gaussian smoothing to remove the small holes. In Tian et al. (2012), Fan and Pankanti (2012), Phong Shading Model has been used to handle quick light changes. In Yang and Rothkrantz (2011), Bird et al. (2006), Femi and Thaiyalnayaki (2013), Bangare et al. (2012), color normalization and 2Dconvolution to enhance the image, structure noise reduction algorithm to remove the noise, Mahalanobis distance between the source and background model pixels to handle multimodal backgrounds with moving objects and illumination changes, Gaussian blur to reduce the noise have used respectively. In theft detection method (Chuang et al. 2009), object size has been considered more than 50 pixels to filter out noisy regions. Fuzzy color histogram (FCH) has been used in Chuang et al. (2008) to deal with the color similarity problem.

5.2 Object tracking

Object tracking is an important and challenging chore in the field of computer vision. It helps in generating the trajectory of an object over time with the tracing its position in consecutive frames of surveillance video to analyze the human behavior. Object shape representations employed for tracking are points, object contour, object silhouette, primitive geometric shapes, articulated shapes and skeletal models (Yilmaz et al. 2006). Sometimes,

Table 2 Foreground object detection methods and applied illumination handling and noise removal methods by several researchers

Type of suspicious activity	Reference	Foreground object detection methods	Illumination handling/noise removal method
Abandoned object detection	Bird et al. (2006)	Stauffer and Grimson Mixture of Gaussian	Structural noise reduction algorithm described by Bevilacqua and Bevilacqua (2002) similar to erosion and dilation
	Porikli (2007)	MOG models with Bayesian update	–
	Miguel and Martínez (2008)	Mixture of Average and Running Average Detection Method	Morphological operation-opening
	Porikli et al. (2008)	Pixel wise multivariate Gaussianmodels-long term and short term	–
	Liao et al. (2008)	Six foreground masks are extracted and calculated intersection of these to find abandoned objects	To remove sporadic noisy and irrelevant pixels, filtering operation is carried out on static foreground object mask
	Li et al. (2009)	Two backgrounds foreground mask from Running average method, stationary mask from simple frame difference	Area filtering and morphological operations to filter out noise
	Bhargava et al. (2009)	Background model is constructed by using background initialization algorithm proposed by Chen and Aggarwal (2008)	Morphological operations
	Li et al. (2010)	Two GMMs-long and short term in RGB color space	False detection caused due to illumination changes reduced by Radial Reach Filter method. Gaussian smoothing to remove small holes
	Evangelio and Sikora (2010)	Two mixtures of Gaussians	–
	Singh et al. (2010)	GMM with K-Gaussian distribution	Dilation and erosion

Table 2 continued

Type of suspicious activity	Reference	Foreground object detection methods	Illumination handling/noise removal method
	Tian et al. (2011)	Multi-Gaussian Adaptive Background Models and three Gaussian mixtures	Integrated texture information to remove the false-positive. Normalized cross correlation (NCC) of the intensities at each pixel of the foreground region is calculated between the current frame and the background frame to remove the false foreground caused by shadows
	Yang and Rothkrantz (2011)	Background model from the first 100 frames using codebook method	Color normalization and 2Dconvolution to enhance the image
	Bangare et al. (2012)	Current background and buffered background	Gaussian blur is used to reduce the noise
	Tian et al. (2012)	Stauffer and Grimson-Mixture of Gaussian	Quick lighting changes are handled with Phong shading model
	Fan and Pankanti (2012)	Learning-based approach	Quick lighting changes are handled with Phong shading model
	Zin et al. (2012b)	Probability based background subtraction	Texture and intensity information to handle quick light changes
	Prabhakar and Ramasubramanian (2012)	Frame difference	Erosion and Dilation
	Chitra et al. (2013)	Mixture of Gaussians technique	Basic filtering
	Femi and Thaiyalnayaki (2013)	Improved Multi-Gaussian Adaptive background model	Mahalanobis distance between the source and background model pixels to handle multimodal backgrounds with moving objects and illumination changes
	Sajith and Nair (2013)	Codebook Method-dual background with different frame rate	–

Table 2 continued

Type of suspicious activity	Reference	Foreground object detection methods	Illumination handling/noise removal method
Theft detection	Chuang et al. (2008)	Gaussian mixture model	Color similarity problem is dealt with Fuzzy color histogram (FCH) Han and Ma (2002)
	Chuang et al. (2009)	GMMs	To filter out noisy regions, object size should be more than 50 pixels
	Ibrahim et al. (2010)	Optical flow	–
	Ibrahim et al. (2012)	Optical flow -Horn-Schunck method	–
Health monitoring of Patients or elder caring at home-Falling detection	Thome and Miguet (2006)	Stauffer mixture of Gaussians modeling	Used a color space invariant in luminance to not assign the moving label to shadow pixels
	Juang and Chang (2007)	Background modeling Chien et al. (2002)	Gradient filter is used to eliminate shadow effect. Erosion and dilation operator with 3×3 structuring features are used to remove noise
	Nasution and Emmanuel (2007)	Stauffer and Grimson	–
	Thome et al. (2008)	Stauffer mixture of Gaussians modeling (Stauffer and Grimson 2000)	A color space invariant in luminance handle shadows
	Foroughi et al. (2008b)	Codebook model	–
	Liu et al. (2010)	Frame differencing	Mean filter to smoothen image
	Rougier et al. (2011)	GMM	Used method proposed by Kim et al. (2005) which handles shadows, highlights and high image compression

Table 2 continued

Type of suspicious activity	Reference	Foreground object detection methods	Illumination handling/noise removal method
	Liu and Zuo (2012)	Background subtraction between current frame and background model	Elgammals shadow suppression method Elgammal et al. (2000). To remove noises and holes, corrosion operators and mathematical morphology expansion are used
	Yogameena et al. (2012)	GMM	Gabor filters kernels to find shadow pixels. Morphological dilation followed by erosion removes small holes
	Yu et al. (2012)	Codebook background subtraction	Blobs smaller than 50 pixels are removed as noise
	Brulin et al. (2012)	Single Gaussian distribution	To fill holes and remove isolated pixels, morphological operations were used
	Chua et al. (2013)	Median Filtering Method	–
Abnormal activities on road traffic	Kamijo et al. (2000)	A background frame is modeled by accumulating and averaging an image sequence for a fixed interval of time.	–
	Guler et al. (2007)	Adaptive background model	–
	Sadeky et al. (2010)	Optical flow	–
	Foucher et al. (2011)	Long-term and short-term background modeling was used through Mixture of Gaussians	–
	Elhamod and Levine (2013)	Lab-based codebook background subtraction technique for the segmentation of the blobs of all foreground silhouettes	–

Table 2 continued

Type of suspicious activity	Reference	Foreground object detection methods	Illumination handling/noise removal method
Violence activity detection	Kausalya and Chitrakala (2012)	Partitioning and Normalized Cross Correlation (PNCC) based algorithm	Particle filter eliminate the Isolated points and little blobs. Median and Gaussian filtering remove Random noises
	Gowsikhaa et al. (2012)	Adaptive background subtraction with sequential and periodical adapting modeling.	Gaussian filters to remove noise
	Ghazal et al. (2012)	Combined the segmentation approach in Amer (2005) and the background update in Achkar and Amer (2007)	Significant deviations in object features are used to detect occlusions
	Gracia et al. (2015)	Absolute image difference between consecutive frames	–
Fire detection	Chen et al. (2004)	–	–
	Celik et al. (2007)	–	–
	Yuan (2010)	–	–
	Borges and Izquierdo (2010)	–	–
	Lai et al. (2012)	Three layer background model	Median filter, Gaussian blur and sharpening filter
	Habiboğlu et al. (2012)	–	–
	Lei and Liu (2013)	Frame differencing	Median filtering
	Seebamrungsat et al. (2014)	Frame differencing	A noise reduction algorithm to reduce noise causes false detection

tracking of an object becomes difficult due to noise in the image, partial or full occlusion of objects, complex object shapes, illumination changes, complex object motion, and deformable objects.

According to Yilmaz et al. (2006), there are three tracking categories- kernel tracking, point tracking, and silhouette tracking. Kalman filter (Kalman 1960) is the well known and widely used methods for object tracking with its ease of use and its real-time operation capability. Kalman filter assumes that the tracked object moves based on a linear dynamic system with Gaussian noise. For non-linear systems, methods based on Kalman filter are proposed, such as Extended Kalman Filter, and Unscented Kalman Filter. Kalman filter with a dynamics model of second order derivative has been used in Höferlin et al. (2015). To detect the start and the end of possible snatching events, Kalman filter has been used in Ibrahim et al. (2010). Kalman filter is used when the movement is linear and to overcome this problem particle filter (Kitagawa 1987) focuses on both nonlinear and non Gaussian signals.

Particle filters are an alternative to the Kalman filters due to their excellent performance in very difficult problems including signal processing, communications, navigation, and computer vision. Particle filters recently became popular in computer vision that are especially used for object detection and tracking. In Foucher et al. (2011), particle filtering and blob matching techniques are used to track the objects.

Kernel based tracking has been used in Chuang et al. (2008, 2009), distance, color, and object size has been used for the tracking in Miguel and Martínez (2008), tracking is used to reduce the false alarm rate in Tian et al. (2011), position and region based blob tracking has been applied in Yang and Rothkrantz (2011), tracking is based on the size and location of the blob in Bhargava et al. (2009), centroids, height and width of the object has been used for tracking in Prabhakar and Ramasubramanian (2012).

A shape matching method has been used for tracking in Rougier et al. (2011). A region based tracking is used in Thome and Miguet (2006) and a tracking method based on connected components has been used in Brulin et al. (2012). Liao et al. (2008) used tracking algorithm based on color and human body contour to detect the owner of the abandoned object in the video. Kamijo et al. (2000) developed tracking algorithm using the spatio-temporal Markov random field model. This algorithm models a tracking problem by determining the state of each pixel in an image, and how the states transit along both the xy image axes and the time axes. In Elhamod and Levine (2013), tracking has been done through the blob matching. Partitioning and Normalized Cross Correlation based algorithm is used for tracking in Kausalya and Chitrakala (2012). KLT tracker (Tomasi and Kanade 1991) has been used to track the vehicle in Aköoz and Karsligil (2010).

Most of the proposed algorithms for abnormal activity detection depend on tracking information. These methods do not work in complex environments like scenes involving crowds and large amounts of occlusion. Several researchers have not employed the tracking based abnormal activity detection due to the occlusion, complex object shapes, deformable objects and a fixed camera angle which cause erroneous tracking. Table 3 shows tracking and non-tracking based approaches in different abnormal activity detection.

5.2.1 Feature extraction

Selecting appropriate features plays an important role in an automatic recognition of abnormal activities from video surveillance. The main objective of feature extraction is to find the most promising information in the recorded video.

Table 3 Tracking and non tracking based abnormal activity detection approaches

Works	Category	References
Tracking based approaches	Abandoned object detection	Foresti et al. (2002), Bird et al. (2006), Miguel and Martínez (2008), Ellingsen (2008), Chuang et al. (2009), Singh et al. (2010), Tian et al. (2011), Hsieh et al. (2011), Yang and Rothkrantz (2011), Tian et al. (2012), Prabhakar and Ramasubramanian (2012), Fern'andez-Caballero et al. (2012), Bangare et al. (2012), Zin et al. (2012a), Chitra et al. (2013), Sajith and Nair (2013), Ferryman et al. (2013), Tejas Naren et al. (2014), Pavithradevi and Aruljothi (2014), Höferlin et al. (2015), Nam (2016)
	Theft detection	Akdemir et al. (2008), Chuang et al. (2008), Ryoo and Aggarwal (2011), Ibrahim et al. (2012), Sujith (2014)
	Falling detection	Lin et al. (2005), Anderson et al. (2006), Thome and Miguet (2006), Thome et al. (2008), Foroughi et al. (2008a, b), Rougier et al. (2011), Brulin et al. (2012)
	Abnormal activity in traffic road	Kamijo et al. (2000), Guler et al. (2007), Lee et al. (2009), Jiang et al. (2011), Foucher et al. (2011), Elhamod and Levine (2013)
	Violence detection	Datta et al. (2002), Kausalya and Chitrakala (2012), Ghazal et al. (2012)
	Fire detection	–
Non-tracking based approaches	Abandoned object detection	Sacchi and Regazzoni (2000), Lavee et al. (2005), Lavee et al. (2007), Porikli (2007), Porikli et al. (2008), Li et al. (2009, 2010) Magno et al. (2009), Sternig et al. (2009), Bhargava et al. (2009), Evangelio and Sikora (2010), Fan and Pankanti (2012), Zin et al. (2012b), Maddalena and Petrosino (2013), Femi and Thaiyalnayaki (2013), Beleznai et al. (2013)
	Theft detection	Ibrahim et al. (2010)
	Falling detection	Nasution and Emmanuel (2007), Juang and Chang (2007), Snoek et al. (2009), Liu et al. (2010), Auvinet et al. (2011), Khan and Sohn (2011), Yogameena et al. (2012), Yu et al. (2012), Liu and Zuo (2012), Chua et al. (2013)
	Abnormal activity in traffic road	Sadeky et al. (2010)
	Violence detection	Wiliem et al. (2012), Penmetsa et al. (2014), Gracia et al. (2015)
	Fire detection	Chen et al. (2004), Celik et al. (2007), Yuan (2010), Borges and Izquierdo (2010), Lai et al. (2012), Habiboğlu et al. (2012), Lei and Liu (2013), Seebamrungsat et al. (2014)

5.2.2 Feature extraction for abandoned object detection/theft detection

To detect the static objects in the video is very complex task. Therefore, some features of objects are extracted from video to make distinction between moving and stationary objects.

Blob trajectory: Tracking based approach (Yang and Rothkrantz 2011) generate the blob trajectory after the splitting of the blobs to detect the moving and stationary object in videos.

Dual foreground with different learning rate: In Porikli (2007), Porikli et al. (2008), dual foreground technique has been employed with two different long-term and short-term learning rates. With these two different learning rates, two foreground masks F_L and F_S are created. If $(F_L; F_S) = (1, 0)$, then object is static.

Centroid, height and width of an object Centroid is defined as an average of the pixels in x and y coordinates belonging to the object that can be calculated through the following formula:

$$Cx = \frac{\sum_{i=1}^{n} Xi}{N} \tag{1}$$

$$Cy = \frac{\sum_{i=1}^{n} Yi}{N} \tag{2}$$

Height and width are the Y-axis and X-axis distance. If objects centroid, height and width are same in each frame, then object is found as static. These features are used in Prabhakar and Ramasubramanian (2012).

Gaussian mixtures of background model Tian et al. (2012) used three Gaussian Mixtures of Background Model in which 1st Gaussian distribution models the persistent pixels and represents; to the background pixels, static regions are updated to the 2nd Gaussian distribution and 3rd Gaussian distribution represents to the quick changing pixels.

Ratio histogram Chuang et al. (2008) proposed a ratio histogram method which is based on fuzzy c-means algorithm to find suspicious objects. In Chuang et al. (2009), novel ratio histogram has been used for finding missing colors between two pedestrians if they have interactions. After detecting the missing colors, a color re-projection method finds the location of each carried object easily.

5.2.3 Feature extraction for falling detection

Point features extraction-Centroids, orientation and distance In Chua et al. (2013), three points are drawn on human shape with the help of bounding box around the human. A bounding box is computed around the human, and then bounding box is divided into three portions which represent upper, mid and lower body part. The starting and end point of the bounding box are used to calculate the centroids of the three regions. The coordinates of the centroids are computed by the following formula:

$$C_{Xi} = \frac{1}{N_{Ri}} \sum_{i=1}^{N_{Ri}} X_i \quad i = 1, 2, 3 \tag{3}$$

$$C_{Yi} = \frac{1}{N_{Ri}} \sum_{i=1}^{N_{Ri}} Y_i \quad i = 1, 2, 3 \tag{4}$$

After this, a line is drawn among these three points. Orientation and distances are calculated to analyze the shape.

Silhouette features In Khan and Sohn (2011), features of silhouette are extracted from original videos automatically. R-transform resolves the problem of continuous changing distance of a moving person from two viewpoints. R-transform is used to extract periodic, translation and scale invariant features. The high similarities in postures of different activities are significantly improved by using the kernel discriminant analysis (KDA). KDA is utilized as a non-linear technique to overcome the similarities among different classes of activities.

Human aspect ratio Human aspect ratio (Liu and Zuo 2012) is defined as the ratio of the width of minimum bounded rectangle box to the height of it.

Effective area ratio Effective area ratio (Liu and Zuo 2012) is defined as the ratio of the area of a person in minimum bounded rectangle box to the area of the whole rectangle.

Centre variation rate The distance of two centers of adjacent frames is very big, and the slope will change which is centre variation rate (Liu and Zuo 2012).

Angle between the minimum bounding rectangle length and the vertical direction In Thome et al. (2008), Silhouette of the object is extracted and a minimum bounding rectangle is drawn around the human silhouette. The angle between the vertical direction and MBR length is computed, constituting the input feature for body pose analysis algorithm.

Projection histogram The horizontal histogram is obtained by calculating number of foreground pixels row wise and vertical histogram is obtained by calculating the number of foreground pixels column wise. Angle between the last standing postures with current foreground bounding box is as the feature set for the task. The extracted projection histogram features are used as input for the classifier (Nasution and Emmanuel 2007). This feature has been also used in Juang and Chang (2007), Foroughi et al. (2008a, b).

Approximated ellipse around human body In Foroughi et al. (2008a, b), Yu et al. (2012), projection histogram and approximated ellipse around human body has been used for feature extraction. The approximated ellipse gives information about the orientation and shape of the person in the image. An ellipse is defined by its centre, orientation, major and minor axis length. The center of the ellipse is calculated by computing the coordinates of the center of mass with the first and zero order spatial moments:

$$\bar{x} = \frac{m_{10}}{m_{00}} \; and \; \bar{y} = \frac{m_{01}}{m_{10}} \tag{5}$$

For a continuous image $f(x, y)$, the moments are given by:

$$mpq = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) x^p y^q dxdy \; for \; p, q = 0, 1, 2 \ldots \tag{6}$$

The centroid $(\bar{x}, \bar{y})$ is used to calculate the central moment as follows:

$$\mu pq = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) d(x - \bar{x}) d(y - \bar{y}) \; for \; p, q = 0, 1, \ldots \tag{7}$$

The angle between the horizontal axis and major axis of the person, gives the orientation of the ellipse which can be computed with the central moments of second order:

$$\theta = \frac{1}{2} \arctan \left(\frac{2\mu_{11}}{\mu_{20-}\mu_{02}} \right) \tag{8}$$

Height and width of bounding box around human The ratio of height and width of bounding box and the difference of height and width of bounding box are adopted as the system features in Liu et al. (2010). Length and width ratio is used in Juang and Chang (2007). Width to height ratio feature has been used to detect the fall in Anderson et al. (2006).

Discrete Fourier transform coefficients DFT is performed on a horizontal and vertical projection histogram of segmented people to solve the shifting and scaling problem in Juang and Chang (2007).

Temporal changes of head position In Foroughi et al. (2008a), to localize the head of the human, initially silhouette is enclosed by minimum circumscribed rectangle and topmost detected point of the rectangle is marked in each frame. Then, the absolute difference values in consecutive frames are obtained. The difference in topmost point of the head over successive frame forms this part of feature vector, and apply appropriate threshold on the vertical displacement of topmost point.

5.2.4 Feature extraction for abnormal activity detection on road traffic

Estimate motion vectors of vehicles Once a vehicle region leaves the slit, its shape is updated along the time sequence by algorithm. For this updating, in a vehicle region, motion vectors among blocks are estimated by the algorithm. At each block, a block matching method is employed to estimate its motion vector in Kamijo et al. (2000).

Histogram of flow gradients (HFG) Histogram of Flow Gradients algorithm (Sadeky et al. 2010) is similar to the HOG, but differs in that HFG locally runs on optical flow field in motion scenes. HFG can be implemented computationally faster than that of HOG. The angle and magnitude of the optical flow required to construct HFG are determined by the following formula:

$$\theta = \tan^{-1}\left(\frac{u}{v}\right), \rho = (u^2 + v^2)^{1/2} \tag{9}$$

where θ and ρ are the angle and magnitude of the velocity of flow respectively. An 8-bin histogram of gradient orientations represent to the orientation of flow in the range of $(-\pi, \pi)$.
Color histogram In Elhamod and Levine (2013), the intersection of the color histograms based on the Lab color system, is computed to measure the spectral similarity between a blob and an object. This histogram distance which is the fastest measure to compute, is robust to partial occlusion and has a good discriminative power.

5.2.5 Feature extraction for violence detection

Shape and texture features Shape and texture features are extracted in Kausalya and Chitrakala (2012) for tracking the moving object. Curvelet mainly extracts the features from images and use to compute the similarity values between images so that efficient geometric shape structure-based image retrieval is possible. Edge detection map is used to detect the edge features.
Center of the minimum bounding box (MBB) In Ghazal et al. (2007), the motion of the video object is characterized by the displacement vector of the center of the minimum bounding box of the video object.

5.2.6 Feature extraction for fire and smoke detection

Flame detection through HSV and YCbCr color model In Seebamrungsat et al. (2014), properties of the YCbCr and HSV color models are used to diiferentiate the flame colors from the background. The HSV color model is applied to detect information related to brightness and color. Through the YCbCr color model, information regarding brightness can be extracted due to its more capability to distinguish bright images efficiently than other color models.

Growth rate identification In Seebamrungsat et al. (2014), this method is applied to reduce the false fire alarm due to lighted candles, lighted matches, orange clothes, or other objects with bright orange color in the video sequences. After extracting foreground flame, numbers of white pixels are counted in consecutive frames to know the difference. If the difference is positive then fire grows. Lai et al. (2012) also utilized this feature to recognize fire.

Color histogram The color histogram is used to detect the presence of smoke in videos by several researchers (Cappellini et al. 1989; Aird and Brown 1997; Wieser and Brupbacher 2001; Vicente and Guillemant 2002). Several statistical measures, such as standard deviation and mean are computed to determine the probability of the presence of smoke.

Temporal based technique Time varying features are extracted by using direct differences of the successive frames and wavelet transform of temporal values of pixels. This feature has been used in Cappellini et al. (1989), Aird and Brown (1997), Wieser and Brupbacher (2001), Vicente and Guillemant (2002).

Rule-based techniques (Foo 1996) Knowledge of fire is coded as some rules to infer the presence of smoke.

Dynamic feature- fluctuation of shape (Yuan 2010) To measure fluctuation of shape feature, more attention is paid to shape and area of fire region. Only early warning of fire is significant for fire suppression. When the area of fire region is larger than some threshold, there is possibility of fire events. Fire shape changes frequently over time, while flashlight, shapes of sun, and other artificial light often change slowly.

Color of smoke (Yuan 2010) When the temperature of smoke is low, color of smoke is from white-bluish to white but when the temperature rises just before ignition, color is from black-grayish to black. Color distributions of smoke pixels can be modeled by learning. Color ranges of smoke pixels are specified manually and saturation detection is performed in RGB color space (Yuan 2008).

Corner flicker rate, compactness, fill rate (Lai et al. 2012) It is well known that flame has some corner on its contour in burning stage. Due to the air and wind flow, the corner position will be located in the upper half of the flame region and the position will be constantly flashing. In addition, the space experiments results have done by NASA show that due to the gravity, flame shape is not a circular but it will always have sharp corners. The corners of each object can be acquired by the Harris corner detection algorithm. Each object is compared with the same id object which has been captured in past image. If the corner position of the same id objects in consecutive frames is different then, corner is treated as a dynamic corner otherwise it is a static corner. Flicker rate is defined as the sum of the dynamic corner counts and total corner counts. Compactness is a function that computes the shape feature in geometry. It is used as the relationship between the area and perimeter. In this, the rectangle which encloses the flame region is divided into two smaller equal rectangles. Both smaller rectangles have same perimeter. Then, the flame pixel number is being calculated. After that, compactness of both smaller rectangles is computed. The upper halves of flames area will less than the lower half of flames area because corner must exist in the upper half of flame region. Therefore, the compactness of upper rectangle was bigger than lower rectangle.

$$Compactness = \frac{C^2}{Area - C} \quad where\ C = parameter \tag{10}$$

Fill rate can be defined as per the following formula:

$$FillRate_{id} = \frac{FlameArea_{id}}{Rectangle\ Area_{id}} \tag{11}$$

where Flame Area is the area of flame and Rectangle Area is the rectangle area which encloses the flame region.

5.3 Classification and activity recognition

After finding moving or stationary foreground objects in a frame, the object classification step is applied for the recognition of normal or abnormal behavior. For example, a stationary human and abandoned object at public place will be treated as suspicious objects if there is no knowledge of the object features. Object classification distinguishes to the static human from static abandoned object, fighting from boxing, face from skin color objects, fire from flashlight, sun light, and any artificial light, falling human pose from laying human pose etc. In general, there are three- feature based, motion based and shape-based classification methods. Several researchers have utilized the different features with different classifiers such as SVM, k-Nearest Neighbor, Multi-SVM, Cascade classifier, Neural Network, and HAR to analyze the human behavior and recognition of abnormal activities. Table 4 shows that many researchers have utilized the different classification methods to recognize the abandoned objects and to improve the accuracy by using either tracking or non-tracking based approaches.

Table 5 visualizes the different work done for theft detection with its used datasets, classification methods, and result discussion. In Table 6, research works have been categorized with three different shape based, posture based and motion based classification techniques with result discussion. Table 7 shows accidents, traffic rule breaking detection and violence detection approaches with their classification methods and result discussion. In Table 8, we have discussed the fire and smoke segmentation, detection methods and its result discussion.

6 Data sets and evaluation measures

6.1 Data sets

Data set is one of most important components to evaluate the performance of any system. Evaluating the proposed algorithm against a standard dataset is one of the challenging tasks in video surveillance system. In the recent years, a number of standard datasets are available in different field of abnormal activity recognition.

6.1.1 Abandoned/removed object detection datasets

PETS 2006 (pet, 2006) PETS 2006 dataset designed to evaluate the performance of abandoned object detection algorithms. The ground truth for the testing video sequences of multiple views includes the number of luggage and persons involved in the event, and also spatial

Table 4 Tracking and non-tracking based abandoned object detection approaches with their used datasets, classification methods and detection accuracy, false positive and false negative rate

Category	Works	Datasets	Classification methods	Result discussion
Tracking based approaches	Foresti et al. (2002)	Video sequence recorded in laboratory, Genova—Borzoli Railway station, Genova- Rivarolo Railway station, Italy of 256 × 256 dimension	Multilayer perceptron as neural network	Event detection rate is about 97%. Laboratory false alarm-0.1% and miss-detection-2.0%. Genova—Rivarolo Railway station-false alarm-1.8% and missdetection 3.5%. Genova-Borzoli Railway station false alarm-0.5% and missdetection-2.5%
	Bird et al. (2006)	BSHigh of length- 49 m, 03 s, BSLow of length-48 m, 23 s, MTC8 of length: 1 h, 44 m, 55 s and MTC5 of length-13m, 32s Long-term and short-term logic	Percent event detected-score at 40 s is 40% Percent event detected- score at 80 s is 60%	–
	Miguel and Martínez (2008)	Video Sequences of PETS2006, i-LIDS, Chroma-VSG	Fusion of low gradient, high gradient detector and color histogram detector	Video Complexity Low-99.7%, Medium-93.58% High-76.4% Recall-L-99%, M-76%, H-73% Precision-L-80%, M-60%, H-34%
	Chuang et al. (2009)	466 video of 320 × 240 dimension	Finite State Machines	Detection accuracy is 100.00%. 2^{12}color bins used by trading off between accuracy and efficiency
	Hsieh et al. (2011)	Own dataset	C-means Clustering, Fuzzy Self-Organizing Neural Network	11 out of 12 abnormal activities were correctly detected. False rejection rate is 6%, and false acceptance rate is 8.3%
	Tian et al. (2011)	Video Sequences of PETS2006, i-LIDS, big city onsite test of four views	Cascade classifier trained by 4000 faces and 4000 non-faces. A Haar filter to achieve real-time performance. Adaboost learning and wavelet features for head and shoulders detection	In S3 video sequence of PETS2006, a static person is detected as an abandoned. Some static person is also detected as abandoned in i-LIDS. 11 removed objects are detected out of 12. 87.8% in Big City Onsite Test. False-alarm rate reduced from 44.5 to 20.7% using tracking

Table 4 continued

Category	Works	Datasets	Classification methods	Result discussion
	Tian et al. (2012)	Video Sequences of PETS2006, i-LIDS	Region growingbased method and Edge energy based method	It may fail in low contrast situations where the color of the object is very similar to the background, e.g., black bag on a black background
	Zin et al. (2012a)	PETS 2006 and Own dataset	rule-based classifier for the realtime process	Quantitative analysis is not done. This system can detect the abandoned objects of very small size from videos of low quality
	Sajith and Nair (2013)	PETS 2006 and PETS 2007	HOG Descriptor and Neural Network classifier	In S3 video sequence of PETS2006, a static person is detected as an abandoned object
	Ferryman et al. (2013)	PETS 2006 and SUBITO (Surveillance of Unattended Baggage and the Identification and Tracking of the Owner) dataset	Logic-based inference engine Alarms raised successfully for all tested sequences excepting PETS-S4-3 and PETS-S7-3	–
	Chitra et al. (2013)	Video Sequences of PETS 2006 of 320 × 240 dimension	Support Vector Machine	Video complexity: Low-91.2%, Medium-90.3% High-80%
	Nam (2016)	PETS2006, i-LIDS, PETS2007	Spatio-temporal i.e. Space first detection and time first detection	For i-LIDS-Precision-98.88%, Recall-82.28%, F-measure-82.64%
Non-tracking based approaches	Porikli et al. (2008)	i-LIDS, PETS 2006, and Advanced Technology Center	The evidence statistics used to extract temporarily static region, which may correspond to illegally parked vehicles, abandoned objects, and removed objects from the scene	All Abandoned object, parked vehicle are detected successfully from i-LIDS, PETS2006 and ATC dataset, only 1-1 false alarms are generated in AB MEDIUM, AB HARD video of i-LIDS. One false alarm in ATC-4 and two false alarms in ATC-5 are also generated. A static people for a long time can be detected as an abandoned item.
	Li et al. (2009)	CAVIAR of 384 × 288 and i-LIDS of 720×576 dimension	Color richness in RGB color space, divided into $N_R \times N_G \times N_B$ equal bins. The color richness counts the colors in a region	System missed one abandoned object in video LeftBag_Behind Chair.mpg. System detected all abandoned or removed objects with 2 false alarms in abandoned and 1 false alarm in removed object of AB HARD video. Two false alarms occur in abandoned object detection of PV HARD video

Table 4 continued

Category	Works	Datasets	Classification methods	Result discussion
	Bhargava et al. (2009)	9 sequences of i-LIDS, 6 sequences of PETS2006	k-nearest neighbor uses feature vectors from about 120 negative and 60 positive image samples and with properties eccentricity, size, compactness, orientation	Fails to detect in Test1 i-LIDS and Test5 i-LIDS video sequences
	Li et al. (2010)	PETS 2006 and 2007 of 320×240	HOG feature vectors as a input to the linear SVM and Height-Width Ratio	Detected all the abandoned baggage except dataset S4
	Evangelio and Sikora (2010)	i-LIDS, PETS2006 and CAVIAR	Finite State Machine	Detected all abandoned objects correctly with 5 false detection in AB MEDIUM and 6 false detection in AB HARD. Frame rate (fps) is 15.80 for AB EASY, 13.84 for AB MEDIUM, 13.91 for AB HARD, 15.48 for PETS2006
	Fan and Pankanti (2012)	i-LIDS, AB-L1 and ABL2 of over 120-h video footage having a total of 862 drops. AB-L1 data set was captured in typical urban areas including parks, streets, and indoors as well as subways. ABL2 was captured in 5 train stations, including platforms and indoor scenes	Region Growing and Structure similarity Methods. Binary classifier LibSVM Chang and Lin (2011) trained by 23 extracted features by using LBP (Local Binary Patterns) Tan and Triggs (2007)	Reduced false positives by 6 and 3% on AB-L1 and AB-L2 respectively, with a small loss of accuracy (2%)
	SanMiguel et al. (2012)	ASODds dataset (2011)	Boundary Spatial Color Contrast along the object contour at pixel level discriminate stationary in abandoned orstolen object.	Suitable for real-time video surveillance Category1: Accuracy-96.7% Category2: Accuracy-94.3% Category3: Accuracy-95.1%
	Tripathi et al. (2013)	PETS 2006 and 2007 of 620×480	Edge based object recognition	Detection accuracy is 85.71% for PETS 2006, 100% for PETS 2007 and 94.4% for Own datset
	Maddalena and Petrosino (2013)	i-LIDS dataset, dog sequence is an outdoor sequence of 320 _ 240 dimension (http://www.openvisor.org)	Neural network mapping method	False detections in the AB-hard video sequence due to static people

Table 5 Theft detection approaches with their used datasets, classification methods and detection accuracy, false positive and false negative rate

Works	Purpose	Datasets	Classification methods	Result discussion
Akdemir et al. (2008)	Detection of Bank attacks	Bank dataset Vu et al. (2002)	Single-threaded ontology	At airport, proposed approach detected the 22 passenger embarkation correctly out of 25, 21 passengers disembarkation out of 25, and 2 aircraft arrival out of 2, 1 aircraft departure out of 1 and 4 luggage cart activity out of 5
Chuang et al. (2008)	Robbery and abandoned object detection	Own video sequences	Ratio histogram based on fuzzy c-means algorithm	Quantitative analysis has not discussed
Ibrahim et al. (2010)	Detection of snatch theft	Own snatching and non-snatching video sequences	Distribution of optical flow vectors before and after the events using vector matching and SVM	SVM classify more than 90% of the test data
Ryoo and Aggarwal (2011)	A group assaulting a person and a group of thieves stealing an object from another group	Total of 45 sequences, ten videos for the group assault, and five videos for each of the other group activities	A hierarchical recognition algorithm utilizing probability distribution sampling based on Markov chain Monte Carlo	Stealing accuracy-100%, Group to group fighting accuracy-100%, Intra-group fighting accuracy 60%, Assault accuracy-80%
Ibrahim et al. (2012)	Detection of snatch theft events	Own video sequence for snatch theft and normal activity	SVM with 10 fold where nine from the fold used to train the data and the rest was used to test the system	Average accuracy is 94.56% sensitivity-83.62% and specificity-71.67%

Table 6 Fall detection approaches with their shape based, posture based and motion based classification methods and result discussion

Classification category	Method	Works	Classification methods	Results	Remarks
Human shape analysis	3-D volume of the person	Auvinet et al. (2011)	Vertical volume distribution ratio	Achieved Sensitivity and specificity 99.7% with four or more cameras. Sensitivity decreased down to 80.6% with three cameras	Multi-camera system. A real-time implementation using a GPU reached 10fps with 8 cameras, and 16 fps with 3 cameras
	Bounding Box	Liu and Zuo (2012)	Human aspect ratio, center variation rate, effective area ratio.	Quantitative analysis has not been discussed	Used indoor video sequences
	Line among 3-centroids	Chua et al. (2013)	Distances and orientations of each line are computed for shape analysis	Fall detection rate-90.5%, False alarm rate-6.7%, Execution time per frame (s) 0.19	Used Indoor video(http://foe.mmu.edu.my/digitalhome/FallVideo.zip)
	Ellipse around the human	Foroughi et al. (2008b)	Temporal changes of head position and Multi-class SVM	Forward Fall-90.83% Backward Fall-93.33% Sideway Fall -86.66% Sensitivity-90.27 % Specificity-95.16%	Detected run, walk, limp, stumble, backward, forward, sideway fall, sit down, bend down and Lie down
		Rougier et al. (2011)	Shape deformation features-Full Procrustes distance and the mean matching cost are really discriminant features for classification	For the full Procrustes distance, the best classification error rate is less than 10% for each camera alone, and decreases to 2.7% using a majority vote. For the mean matching cost, majority vote gave 98% accuracy	Uncalibrated multi-camera system using an ensemble classifier to improve. Detected Forward, backward falls, falls when inappropriately sitting down, loss of balance
Posture estimation analysis	Ellipse around Silhouette	Thome et al. (2008)	Silhouette between lengthened and standing postures and Layered HMM	Obtained correct detections rate 82% and false negatives 18%	Real-time, multi-view
		Foroughi et al. (2008a)	Four-layered MLP network with back propagation learning schema	Forward Fall -92.80% Backward Fall-94.40% Sideway Fall -91.20% Sensitivity-92.80 Specificity-97.60	Detected walk, run, stumble, limp, bending, sitting, lying and falling

Table 6 continued

Classification category	Method	Works	Classification methods	Results	Remarks
		Yu et al. (2012)	Multiclass SVM Classification by DAGSVM	97.08% falls can be detected while only 20.8% non falls mistaken as falls	Detected fall, walking around, sitting, bending, and lying on the sofa
	Binary silhouette	Khan and Sohn (2011)	Kernel Discriminant Analysis on R-transform features. k-means clustering algorithm and HMM for activity recognition	Average recognition rate-95.8%	Detected forward fall-92%, backward fall-100%, chest pain-100%,faint-88%,vomit-95%, headache-100%
	Bounding box around	Nasution and Emmanuel (2007)	k-NN and bounding box angle test, temporal information i.e. speeds of fall	Recognition rate-90.0%	Detected the side and fall toward the camera, standing, bending, sitting, lying
		Juang and Chang (2007)	Self-Constructing Neural Fuzzy Inference Network (SONFIN) Classification	Average detection accuracy-97.8%	Detected falling, sitting, lying, bending, and standing
		Liu et al. (2010)	k-NN classify human body postures with k-fold cross-validation	Accuracy rate on fall incident detection is about 82.22%	For k = 3 fold cross validation, average rate is 95.34%
		Brulin et al. (2012)	Viola and Jones' method based on 14 Haar-like filters. boosted classifiers Fuzzy Logic System	Average Accuracy for S4 dataset-72.24% Average Accuracy for S5 dataset-64.93%	Detected falling, sitting, lying, squatting, standing
Motion analysis	posture of moving human	Yogameena et al. (2012)	Relevance Vector Machine is used to detect the fall based on torso angle through skeletonization	Sensitivity: 95.83%, Specificity: 97.5%, Accuracy: 96.67%	–
	Human pose	Thome and Miguet (2006)	Hierarchical Hidden Markov Model	Correct detections 82% and false negatives 18%	–

Table 7 Accidents and traffic rule breaking detection and Violence detection approaches with their classification methods and result discussion

Works	Approach	Classification	Result discussion
Accidents and traffic rule breaking detection approaches			
Kamijo et al. (2000)	Spatio-temporal Markov random field(ST-MRF) model based	HMM	Tracked vehicles rate is 94.6%. Accidents were successfully detected by HMM
Guler et al. (2007)	Tunnel vision tracking based	Set of object features are extracted based on the color edge and shape information of the object for detailed analysis and classification of objects	Parked Vehicle in Easy, Medium, and Night video is detected with a few second differences from the ground truth time. Medium PV sequence effects the end time detection causing a 7 second delay from the ground truth
Lee et al. (2009)	Image projection based and tracking based	Self-Split/Merge event and Merge event	Processing time of each frame is less than 0.2s. This system failed to detect one illegally parked vehicle because two vehicles came to the NP zone together and both parked very close to each other in the NP zone
Sadeky et al. (2010)	Local features of flow gradient orientations and logistic regression modeling based	Pattern classification-Euclidean distance metrics	The recognition rate of the system is 99.6% with 5.2% false alarm rate. Real-time surveillance for accident detection
Aköoz and Karsligil (2010)	Tracking clustering using Continuous HMM with Mixture of Gaussians Rabiner (1989)	Trajectory clustering determines activity patterns. Log-likelihood thresholds segregate normal or abnormal traffic events. Linear multiphase regression applies semantic information to characterize traffic events and collisions	Accuracy for vehicle collision-85%. Accuracy for Nearby passing-89%. Accuracy for Lane deviation-88%. The success rate for the true categorization of the vehicle collisions according to their severity is around 84%

Table 7 continued

Works	Approach	Classification	Result discussion
Violence detection approaches			
Jiang et al. (2011)	Spatial and temporal context based	Viterbi Algorithm with HMM	Detection rate is more than 90% for point anomaly, more than 80% for sequential anomaly, and more than 70% for co-occurrence anomaly
Benezeth et al. (2011)	Low-level feature based	Markov Random Field accounts for direction, speed and size without any intervention	Proposed detected all the abnormal activities with 9.5% false positives
Wiliem et al. (2012)	Contextual information based	Inference algorithm combines contextual information and system knowledge to infer the decision	Proposed system's AUC (Area Under Curve) is 0.778 with 0.144 errors
Ghazal et al. (2012)	Multiple object tracking technique	Detected vandalism using information gathered from object tracks and features	Real-time based system with frame rate of 13 fps. Vandalism detection rate is 96%
Gowsikhaa et al. (2012)	Head motion and contact detection based	Trained Artificial Neural network by Gabor feature extraction to detect face. Skin color for hand detection. Rule based activity classification	Precision and recall for head motion detection is more than 80%. Precision and recall for contact detection is more than 85%
Penmetsa et al. (2014)	Pose estimation based	Hough orientation calculator for pose classification	**Action** / **Precision** / **Recall** / **Accuracy** Slapping 95.65 62.86 77.78 Punching 87.10 58.70 76.67 Shooting 81.82 46.15 79.59 Choking 88.46 63.89 73.47 Snatching 91.97 61.11 78.26
Gracia et al. (2015)	Spatio-temporal features based	Centroids, centroids distances, compactness are used as two variants with three Support Vector Machines (SVM), AdaBoost and Random Forests (RF) classifiers	Processing average time of 0.02391 seconds per frame, it is much faster while still maintaining useful accuracies ranging from 70% to nearly 98% depending on the dataset
Tripathi et al. (2015)	MHI and Hu moments to extract features	PCA reduces the dimensionality of features and SVM carries classification	Average accuracy is 95.73%

Table 8 Fire and smoke detection approaches with their classification methods and result discussion

Works	Fire and smoke segmentation	Fire detection method	Results
Chen et al. (2004)	Segment moving regions by image differencing. RGB chromatic features for fire and smoke detection	Dynamics of growth and disorder	–
Töreyin et al. (2006)	Spatio-temporal wavelet based fire detection	Irregularity of the boundary of the fire-colored region and color variations	Real-time system has average processing time is 16.5 ms/frame. Overall false positive rate is 0.001 and fire detection rate is 1.0
Celik et al. (2007)	YCbCr color, smoke model using RGB	fuzzy inference system	Detection rate-99.00 False alarm rate-4.5%
Gubbi et al. (2009)	Characterize smoke using a block based approach using DCTs and wavelets	Used wavelets along with a non-linear classifier such as support vector machines for smoke detection	Accuracy-88.75% Sensitivity-0.90 Specificity-0.89
Yuan (2010)	Gaussian mixture model and frame difference. spatiotemporal feature-color and motion	Fluctuation of shape Bayesian classifier	Smoke is detected within 7.6s from dry leaves, 3.96s from gasoline without wind and rope with wind
Borges and Izquierdo (2010)	Probability based color analysis	Boundary roughness, area size change, variance, and red channel skewness features are combined according to the Bayes classifier	False-Positive-0.68% False-Negative-0.028%
Lai et al. (2012)	Three inter-frame difference algorithm for foreground objects and YCbCr model for flame	Corner flicker rates, compactness, fill rate and growth rate	Burning foam from scene1 is detected while the flame is not detected in scene2 because flame is mirror flame in scene2
Habiboğlu et al. (2012)	Spatio-temporal covariance descriptor and color model	Support Vector Machine	True detection rates are 96.6, 95.5, and 91.7% for the 34, 55 and 21 parameter respectively. Efficient to process 320 × 240 frames at 20 fps makes it real-time video fire detection system

Table 8 continued

Works	Fire and smoke segmentation	Fire detection method	Results
Lei and Liu (2013)	Potential fire region is detected by using frame differencing of monitor video. Flame region is extracted by color information	Bayes classifier with dynamic features	–
Seebamrungsat et al. (2014)	HSV and YCbCr color models	Fire growth from frame difference	Accuracy was 100% for thirty fire video files
Manjunatha et al. (2015)	Neuro-fuzzy algorithm based	Rule based generic collective model for classification	Accuracy for fire detection and suppression is 99%
Foggia et al. (2015)	Color, shape and flame movements information based	Multi expert system classifier based on a weighted voting rule	Real-time based system Accuracy is 93.55%, false positives is 11.76%

Fig. 6 Four view video sequences of PETS 2006 Dataset

Fig. 7 Four view video sequences of PETS 2007 Dataset

Fig. 8 Three video sequences of i-LIDS 2007 Dataset

relationships between the persons and luggage. The PETS 2006 dataset consists of multi-view video sequences of real scene captured with illumination effect, crowd and luggage left. There are seven different scenarios captured by four cameras from different viewpoints. Figure 6 shows four different views of video sequences of PETS2006.

PETS 2007 (pet, 2007) PETS 2007 dataset designed to test loitering, theft and abandoned object detection. There are 8 video sequences captured by four cameras from different view-points. Two video sequences S7 and S8 are available for abandoned object detection. These video sequences are full of bad illumination and more lighting effects. Figure 7 shows four different views of video sequences of PETS2007.

i-LIDS-abandoned baggage detection (avs, 2007) i-LIDS is a Imagery Library for Intelligent Detection Systems. This dataset consists of unattended bags on the platform of an underground station. There are three videos which have been categorized on the basis of scene complexity (shown in Fig. 8)

Fig. 9 Four video sequences of multiple cameras fall dataset

VISOR (Vezzani and Cucchiara 2010) Video Surveillance Online Repository provides videos for different human actions such as abandoned object, drinking water, jumping, sitting, etc. Nine videos of abandoned object are available out of forty different human action videos.

CVSG (cvs, 2008*)* In this dataset, different sequences have been recorded using chroma based techniques for simple extraction of foreground masks. Then, these masks are composed with different backgrounds. Provided sequences have varying degrees of difficulty in terms of foreground segmentation complexity. Sequences contain examples of abandoned objects and objects removed from the scene.

CAVIAR (cav, 2004*)* CAVIAR dataset consists of a number of video clips which were recorded different activities such as walking people in different lane, leaving bags, fighting, etc.

6.1.2 Theft detection datasets

Bank dataset (Vu et al. 2002) The bank dataset is the collection of six video sequences. Four video sequences out of six video sequences consist of different instances of bank robberies while other two video sequences consist of normal activities in the bank. The bank scenario was captured by a single static camera.

6.1.3 Falling detection datasets

Video sequences (Chua et al. 2013) Chua et al. (2013) acquired video sequences from an un-calibrated IP camera (DlinkDCS-920) through Wi-Fi connection in MJPEG format at a resolution of 320×240. Test video data consist of video sequences of 30 daily normal activities such as walking, crouching down, sitting down, and squatting down, and 21 simulated falls such as forward and backward falls, sideway falls, and falls due to loss of balance (Human fall detection dataset 2014).

CAVIAR video sequences (onf, 2004*)* These datasets contain fall of an individual person from different camera view points. CAVIAR dataset is available at (http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/).

Multiple cameras fall dataset (Auvinet et al. 2010) This dataset consists of walking, standing up, falling, lying on the ground, crounching, moving down, moving up, sitting, lying on sofa, and moving horizontally. There are 24 scenarios of 8 cameras. Figure 9 shows the four video sequences for different type of falls.

6.1.4 Road traffic datasets

TRECVid2010 (tre, 2010) Videos were captured by 5 different indoors cameras at the Gatwick Airport and compressed in MPEG-2 format. The corpus is split between 44 h of test data and 100 h of development data (10 h × 2 h/day × 5 cameras).

i-Lids AVSS-07 parked vehicle detection data set (avs, 2007) I-LIDS provided parked vehicle video sequences (Fig. 10) for illegal parking detection and alarming for the vehicles that stop for 60 s, in the red marked no parking zones.

Traffic videos from the next generation simulation (NGSIM) project (ngs, 2007) This surveillance video monitors a four-way intersection in Los Angeles, California. Each road is a two way road with multiple lanes-some with left turn or right turn lanes. All moving traffic of this area is controlled by traffic lights within the intersection. In this database, detailed trajectory information for all vehicles in this video is available, such as the driving direction, lane information and the velocity of each vehicle at every time of its appearance. This video contains 21,689 frames and 2230 vehicle trajectories are tracked.

MIT traffic data (Wang et al. 2009) MIT traffic data set is for research on activity analysis and crowded scenes. It includes a traffic video sequence which is 90 min long. It was captured by a static camera. The size of the frames are 720 × 480. It is divided into 20 clips. It can be downloaded from (http://www.ee.cuhk.edu.hk/xgwang/MITtraffic.html).

QMUL junction dataset (qmu, 2010) This dataset consists of busy traffic dataset for the behavior understanding and activity analysis. Its length is of 1 h duration with 90,000 frames. It has 360 × 288 dimension with ffdshow mpeg-4 compression codec.

AllGo vision dataset (all, 2015) This dataset consists of illegal parking, speed detection, wrong way vehicle, congestion detection, parking management related video sequences. It also has smoke detection and left baggage detection video sequences. Figure 11 shows the three video sequences of this dataset.

6.1.5 Violence detection datasets

Fight CAVIAR data set (M. 2011) This dataset consists of four video sequences of fighting scenes. In first and second video sequence, two people meet, fight and run away. In third video sequence, two people meet, and then fight, one down and second people runs away. In fourth, two people meet, fight and chase each other.

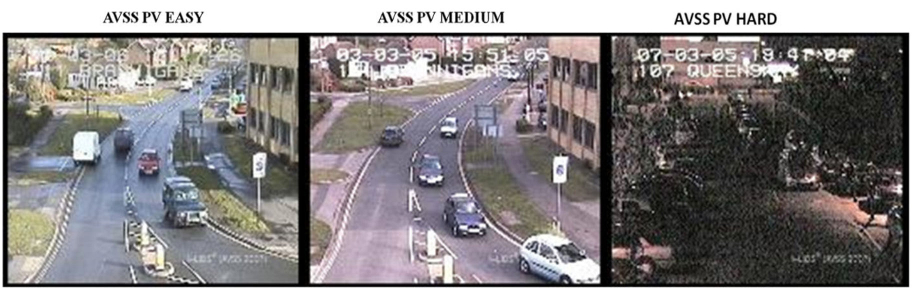

Fig. 10 Four video sequences of i-LIDS parked vehicle dataset

Fig. 11 Three video sequences of AllGo vision dataset

UCF101 (Soomro et al. 2012) The UCF101 is a dataset of realistic action videos collected from YouTube, having 101 categories of actions. UCF101 yields the largest diversity in terms of actions and with the presence of large variations in camera motion, object scale, object pose and appearance, viewpoint, illumination conditions and cluttered background; it is the most challenging dataset to this date. For this case, it is even more challenging since it also contains 50 actions from sports. To our knowledge, this is the most challenging and largest dataset in which a fight detection algorithm has been tested.

6.1.6 Fire detection datasets

Sample fire and smoke video clips (smo, 2009*)* This dataset consists of fire and smoke video clips for evaluating the performance of fire and smoke detection system.

FASTData (fas, 2014*)* This dataset is a collection of resources from the Building and Fire Research Laboratory's Fire Research Division at NIST. These web pages provide links to fire related software, experimental fire data and quick time movies of fire tests that can be downloaded.

DynTex dynamic texture database (Pteri 2012) The DynTex database is a diverse collection of dynamic texture videos of high-quality. Dynamic textures are typically result from processes such as of smoke, waves, a flag blowing in the wind, fire, a moving escalator, or a walking crowd. Many real-world textures occurring in video databases are dynamic and retrieval should be based on both their static and dynamic features.

MESH (mes, 2007*)* This dataset consists of various news videos. In this, fire related videos can be found also for evaluation of fire detection system.

Firesense data set (fir, 2009*)* The Firesense data set contains ten non-fire videos and eleven fire videos.

6.2 Evaluation measures

Evaluation of the performance of an Intelligent Video Surveillance System (IVSS) for abandoned or removed object detection, theft detection, falling detection, abnormal activities on road traffic, violence detection and fire detection is one of the major task to validate the robustness and correctness. The evaluation of different abandoned or removed object detection, theft detection, falling detection, accidents on road, violence detection and fire detection systems can be performed in two ways; firstly quantitatively and qualitatively. Qualitative evaluation approaches are performed on visual interpretation, by looking at processed image yield by the algorithm. It consists of several issues and challenges handling algorithms. Noise removal, illumination handling, shadow removal, partial or full occlusion handling, poor resolution handling etc. improves the qualitative performance of the IVSS. On the other hand, quantitative progress requires a numeric comparison of computed results with ground truth data. Due to the necessity of computing a valid ground truth data, the quantitative evaluation of IVS systems are highly challenging. There are a number of metrics proposed in literature to quantitatively evaluate the performance of an IVS system.

Recognition accuracy Most of the research work in abandoned object detection, theft detection, falling detection, abnormal activity detection on road traffic, violence detection and fire used accuracy for measurement of evaluation. It is defined as follows (Penmetsa et al. 2014):

$$Accuracy(\%) = \left(\frac{T_P + T_N}{T_P + T_N + F_P + F_N}\right) \tag{12}$$

A true positive (T_P) represents suspicious action classified as suspicious by the classifier; a false negative (F_N) represents the classification of suspicious action as non-suspicious; a false positive (F_P) corresponds to the classification of non-suspicious action as suspicious and a true negative (T_N) stands for non-suspicious action classified correctly. In Li et al. (2009), Tian et al. (2011, 2012), Fan and Pankanti (2012), Femi and Thaiyalnayaki (2013), researchers evaluated the performance by True Positive, and False Positive detection. False positives and false negatives have been used to evaluate the performance of fire detection system (Borges and Izquierdo 2010).

Recall, precision, F-measure Penmetsa et al. (2014) used Recall, Precision for their experimental evaluation. In Miguel and Martínez (2008), Fan and Pankanti (2012), Ferryman et al. (2013), researchers have utilized both parameter to evaluate the performance of abandoned object detection systems where the Precision represent the percentage of true alarms and recall represents the percentage of detected events.

$$Recall(\%) = \left(\frac{T_P}{T_P + F_N}\right) \tag{13}$$

$$Precision(\%) = \left(\frac{T_P}{T_P + F_P}\right) \tag{14}$$

$$F\text{-}measure(\%) = \left(\frac{2 \times Precision \times Recall}{Precision + Recall}\right) \times 100 \tag{15}$$

Brulin et al. (2012) computed recall, precision and f-measure to measure the performance of the system.

Sensitivity, specificity Sensitivity and specificity can be defined as follows (Auvinet et al. 2011):

$$Sensitivity(\%) = \left(\frac{T_P}{T_P + F_N}\right) \tag{16}$$

$$Specificity(\%) = \left(\frac{T_N}{T_N + F_P}\right) \tag{17}$$

Yogameena et al. (2012) defined sensitivity and specificity in terms of falling detection system where high sensitivity means majority of falls are detected and a high specificity means that normal activities are not detected as falls. It has been utilized to evaluate the quantitative performance in various research works (Foroughi et al. 2008a, b; Liu et al. 2010; Ibrahim et al. 2012; Yogameena et al. 2012).

Frames per second Real-Time Intelligent Video Surveillance System must have a good execution speed for processing the video frames. Several researchers (Sacchi and Regazzoni 2000; Evangelio and Sikora 2010; Maddalena and Petrosino 2013; Chua et al. 2013) have computed execution speed of their system to decide whether the system would work on real-time or not.

PED and PAT scores (Bird et al. 2006) The PED (Percent Events Detected) score represents the ratio of real alarms in the ground truth that are successfully detected by the module to the total number of alarms in the ground truth. The PAT (Percent Alarms True) score represents the ratio of alarms that correspond to real alarms in the ground truth to the total number of alarms detected by the module. A high PED score indicates that the module detects most objects that should trigger an alarm. A high PAT score indicates that the module rarely triggers false alarms. The formulas of PED and PAT are as follows:

$$PED(\%) = \left(\frac{Number\ of\ real\ alarms\ detected}{Number\ of\ alarms} detected\right) \times 100\% \tag{18}$$

$$PAT(\%) = \left(\frac{Number\ of\ real\ alarms\ detected}{Total Number of alarms} detected\right) \times 100\% \tag{19}$$

Confusion matrix A confusion matrix, also known as an error matrix or a contingency table and is used to envisage performance of a supervised learning algorithm. Each column of the matrix represents the predicted class, while each row represents an actual class. With the help of this matrix, it is easy to see that where system is confusing among classes. It has been utilized in various research works (Nasution and Emmanuel 2007; Foroughi et al. 2008b; Brulin et al. 2012; Habiboğlu et al. 2012). Confusion matrix has been used to compute the overall and average accuracy in Brulin et al. (2012) and to better understand the wrong classification results in Nasution and Emmanuel (2007).

ROC curve In statistics or machine learning, a receiver operating characteristic (ROC) curve is a graphical plot that reveals the performance of a binary classifier. This curve is drawn by plotting the true positive rate against the false positive rate at various threshold settings. Many researchers (Sadeky et al. 2010; Rougier et al. 2011; Wiliem et al. 2012) employed ROC analysis of the performance of the different parameters.

7 Conclusions and future work

In this survey paper, we have discussed the various techniques related to abandoned object detection, theft detection, falling detection, accidents and illegal parking detection, violence detection and fire detection for the foreground object extraction, tracking, feature extraction and classification. In past decades, several researchers proposed novel approaches with noise removal, illumination handling, and occlusion handling methods to reduce the false object detection. Many researchers have also worked for making real-time intelligent surveillance system but processing rate of the video frames is not as good as required and there is no such system that has been developed with 100% detection accuracy and 0% false detection rate for videos having complex background. Much of the attention is required in the following suspicious activities detection:

Abandoned object detection and theft detection Majority of the works have been done for the abandoned object detection from surveillance videos captured by static cameras. Few works detected the static human as an abandoned object. To resolve such problems, human detection method should be very effective and system should check the presence of the owner in the scene, if owner is invisible in the scene for long duration then alarm should be raised. To resolve the problem of theft or object removal, face of the person who is picking up the static object, should match with the owner otherwise an alarm must be raised to alert the security. Future work may also resolve the low contrast situation i.e. similar color problem such as black bag and black background which lead to miss detections. Future improvements may be integration of intensity and depth cues in the form of 3D aggregation of evidence and occlusion analysis in detail. Spatial-temporal features can be extended to 3-dimensional space for the improvement of abandoned object detection method for various complex environments. Thresholding based future works can improve the performance of the surveillance system by using adaptive or hysteresis thresholding approaches. Few works have been also proposed for abandoned object detection from the multiple views captured by multiple cameras. To incorporate these multiple views to infer the information about abandoned object can also be improved. There is a large scope to detect abandoned object from videos captured by moving cameras.

Falling detection Most of the works have been done for fall detection of single person in indoor videos based on human shape analysis, posture estimation analysis and motion based analysis. Future works can include the integration of multiple elderly monitoring which is able to monitor more than one person in the indoor scene. Many elder people go for morning walk everyday in public areas such as parks; to monitor these elder people, a future work can include one or more than one human fall detection from outdoor surveillance videos.

Accidents, illegal parking, and rule breaking traffic detection Several researchers have presented accidents detection, illegal parking detection and illegal U-turn detections from static video surveillance. These systems become incapable to detect these abnormal activities in more crowded traffic on roads. Future works should be based on unsupervised learning of transportation system because of no standard dataset is available for the training.

Violence detection Several research works have been done for the prevention of violence activities such as vandalism, fighting, shooting, punching, and hitting. To detect such violence activities, single view static video camera has been used but sometimes this system fails in occlusion handling. Therefore, a multi-view system has been proposed by few researchers to resolve this problem but it requires important cooperation between all views at the low level steps for abnormal activity detection. Future work may be automatic surveillance system for moving videos. Improvements are required in accuracy, false alarm reduction, and frame rate to develop an intelligent surveillance system for the road traffic monitoring.

Fire detection Future work can include more improvement in accuracy, frame rate, false alarms reduction and also it can be improved to detect far distant small fire covered by dense smoke.

References

Achkar F, Amer A (2007) Hysteresis-based selective gaussian mixture models for real- timebackground maintenance. SPIE Vis Commun Image Process 6508:J1–J11

Adam A, Rivlin E, Shimshoni I, Reinitz D (2008) Robust real-time unusual event detection using multiple fixed-location monitors. IEEE Trans Pattern Anal Mach Intell 30(3):555–560

Aggarwal JK, Ryoo MS (2011) Human activity analysis: a review. ACM Comput Surv (CSUR) 43(3):16

Aird B, Brown A (1997) Detection and alarming of the early appearance of fire using cctv cameras. In: Nuclear engineering internat. fire and safety conference, London, vol 24, p 26

Akdemir U, Turaga P, Chellappa R (2008) An ontology based approach for activity recognition from video. In: Proceedings of the 16th ACM international conference on Multimedia, ACM, pp 709–712

Aköoz Ö, Karsligil M (2010) Severity detection of traffic accidents at intersections based on vehicle motion analysis and multiphase linear regression. In: 13th International IEEE conference on intelligent transportation systems (ITSC), 2010, IEEE, pp 474–479

Allgovision (2015) Advanced video analytics for traffic/parking management. http://www.allgovision.com/traffic-praking.php

Amer A (2005) Voting-based simultaneous tracking of multiple video objects. IEEE Trans Circuit Syst Video Technol 15(11):1448–1462

Anderson D, Keller JM, Skubic M, Chen X, He Z (2006) Recognizing falls from silhouettes. In: 28th annual international conference of the IEEE engineering in medicine and biology society, 2006. EMBS'06. IEEE, pp 6388–6391

Asodds (2011) An abandoned and stolen object discrimination dataset. http://wwwvpu.eps.uam.es/asodds

Auvinet E, Multon F, Saint-Arnaud A, Rousseau J, Meunier J (2011) Fall detection with multiple cameras: an occlusion-resistant method based on 3-d silhouette vertical distribution. IEEE Trans Inf Technol Biomed 15(2):290–300

Auvinet E, Rougier C, Meunier J, St-Anaud A, Rousseau J (2010) Multiple Cameras Fall Data Set. DIRO-Universite de Montrial, Technical Report 1350

AVSS (2007) http://www.eecs.qmul.ac.uk/andrea/avss2007d.html

Bangare PS, Uke NJ, Bangare SL (2012) Implementation of abandoned object detection in real time environment. Int J Comput Appl 57(12):13–16

Beleznai C, Gemeiner P, Zinner C (2013) Reliable left luggage detection using stereo depth and intensity cues. In: IEEE international conference on computer vision workshops (ICCVW), 2013, IEEE, pp 59–66

Benezeth Y, Jodoin PM, Saligrama V (2011) Abnormality detection using low-level co-occurring events. Pattern Recogn Lett 32(3):423–431

Bevilacqua A, Bevilacqua R (2002) Effective object segmentation in a traffic monitoring application. In: ICVGIP 2002 conference proceedings, Ahmedabad, India, Citeseer

Bhargava M, Chen CC, Ryoo MS, Aggarwal JK (2009) Detection of object abandonment using temporal logic. Mach Vis Appl 20(5):271–281

Bird N, Atev S, Caramelli N, Martin R, Masoud O, Papanikolopoulos N (2006) Real time, online detection of abandoned objects in public areas. In: Proceedings 2006 IEEE international conference on robotics and automation, 2006. ICRA 2006. IEEE, pp 3775–3780

Borges PVK, Izquierdo E (2010) A probabilistic approach for vision-based fire detection in videos. IEEE Trans Circuit Syst Video Technol 20(5):721–731

Bouwmans T (2014) Traditional and recent approaches in background modeling for foreground detection: an overview. Comput Sci Rev 11:31–66

Brulin D, Benezeth Y, Courtial E (2012) Posture recognition based on fuzzy logic for home monitoring of the elderly. IEEE Trans Inf Technol Biomed 16(5):974–982

Candamo J, Shreve M, Goldgof DB, Sapper DB, Kasturi R (2010) Understanding transit scenes: a survey on human behavior-recognition algorithms. IEEE Trans Intell Transp Syst 11(1):206–224

Cappellini V, Mattii L, Mecocci A (1989) An intelligent system for automatic fire detection in forests. In: Third international conference on image processing and its applications, 1989, IET, pp 563–570

Caviar fall on floor dataset (2004). http://homepages.inf.ed.ac.uk/rbf/caviardata1/

Caviar Left Bag (2004) http://www.multitel.be/va/cantata/leftobject/

Celik T, Ozkaramanli H, Demirel H (2007) Fire and smoke detection without sensors: image processing based approach. In: 15th European signal processing conference, EUSIPCO, pp 147–158

Chang CC, Lin CJ (2011) Libsvm: a library for support vector machines. ACM Trans Intell Syst Technol (TIST) 2(3):27

Chen CC, Aggarwal J (2008) An adaptive background model initialization algorithm with objects moving at different depths. In: 15th IEEE international conference on image processing, 2008. ICIP 2008, IEEE, pp 2664–2667

Chen TH, Wu PH, Chiou YC (2004) An early fire-detection method based on image processing. In: International conference on image processing, ICIP'04. 2004, IEEE, vol 3, pp 1707–1710

Chen YT, Lin YC, Fang WH (2010) A hybrid human fall detection scheme. In: 17th IEEE international conference on image processing (ICIP), 2010, IEEE, pp 3485–3488

Chien SY, Ma SY, Chen LG (2002) Efficient moving object segmentation algorithm using background registration technique. IEEE Trans Circuit Syst Video Technol 12(7):577–586

Chitra M, Geetha MK, Menaka L, et al (2013) Occlusion and abandoned object detection for surveillance applications. Int J Comput Appl Technol Res 2(6):708–meta

Chua JL, Chang YC, Lim WK (2013) A simple vision-based fall detection technique for indoor video surveillance. SIViP 9(3):623–633

Chuang CH, Hsieh JW, Tsai LW, Ju PS, Fan KC, (2008) Suspicious object detection using fuzzy-color histogram. In: IEEE international symposium on circuits and systems, ISCAS 2008, IEEE, pp 3546–3549

Chuang CH, Hsieh JW, Tsai LW, Chen SY, Fan KC (2009) Carried object detection using ratio histogram and its application to suspicious event analysis. IEEE Trans Circuit Syst Video Technol 19(6):911–916

Cucchiara R, Grana C, Piccardi M, Prati A (2003) Detecting moving objects, ghosts, and shadows in video streams. IEEE Trans Pattern Anal Mach Intel 25(10):1337–1342

Cui L, Li K, Chen J, Li Z (2011) Abnormal event detection in traffic video surveillance based on local features. In: 4th international congress on image and signal processing (CISP), 2011, IEEE, vol 1, pp 362–366

Cvsg (2008) http://www.vpu.eps.uam.es/cvsg/

Datta A, Shah M, Lobo NDV (2002) Person-on-person violence detection in video data. In: Proceedings of the 16th international conference on pattern recognition, 2002, IEEE, vol 1, pp 433–438

Dick AR, Brooks MJ (2003) Issues in automated visual surveillance. In: International conference on digital image computing: techniques and applications

Dimitropoulos K, Barmpoutis P, Grammalidis N (2015) Spatio-temporal flame modeling and dynamic texture analysis for automatic video-based fire detection. IEEE Trans Circuit Syst Video Technol 25(2):339–351

Elgammal A, Harwood D, Davis L (2000) Non-parametric model for background subtraction. In: Computer Vision, ECCV 2000. Springer, Berlin, pp 751–767

Elhamod M, Levine MD (2013) Automated real-time detection of potentially suspicious behavior in public transport areas. IEEE Trans Intel Transp Syst 14(2):688–699

Ellingsen K (2008) Salient event-detection in video surveillance scenarios. In: Proceedings of the 1st ACM workshop on analysis and retrieval of events/actions and workflows in video streams. ACM, pp 57–64

Evangelio RH, Sikora T (2010) Static object detection based on a dual background model and a finite-state machine. EURASIP J Image Video Process 2011(1):858,502

Fan Q, Pankanti S (2012) Robust foreground and abandonment analysis for large-scale abandoned object detection in complex surveillance videos. In: IEEE ninth international conference on advanced video and signal- based surveillance (AVSS), 2012, IEEE, pp 58–63

Fan Q, Gabbur P, Pankanti S (2013) Relative attributes for large-scale abandoned object detection. In: IEEE international conference on computer vision (ICCV), 2013, IEEE, pp 2736–2743

Fastdata (2014) http://fire.nist.gov/fastdata

Femi PS, Thaiyalnayaki K (2013) Detection of abandoned and stolen objects in videos using mixture of gaussians. Int J Comput Appl 70(10):18–21

Fern’andez-Caballero A, Castillo JC, Rodr’ıguez-S’anchez JM (2012) Human activity monitoring by local and global finite state machines. Expert Syst Appl 39(8):6982–6993

Ferryman J, Hogg D, Sochman J, Behera A, Rodriguez-Serrano JA, Worgan S, Li L, Leung V, Evans M, Cornic P et al (2013) Robust abandoned object detection integrating wide area visual surveillance and social context. Pattern Recogn Lett 34(7):789–798

Firesense project protection of cultural heritage (2009). http://www.firesense.eu/

Foggia P, Saggese A, Vento M(2015) Real-time fire detection for video surveillance applications using a combination of experts based on color, shape and motion. IEEE Trans Circuit Syst Video Technol 25(9):1545–1556

Foo SY (1996) A rule-based machine vision system for fire detection in aircraft dry bays and engine compartments. Knowl Based Syst 9(8):531–540

Foresti GL, Marcenaro L, Regazzoni CS (2002) Automatic detection and indexing of videoevent shots for surveillance applications. IEEE Trans Multimed 4(4):459–471

Foroughi H, Aski BS, Pourreza H (2008a) Intelligent video surveillance for monitoring fall detection of elderly in home environments. In: 11th international conference on computer and information technology. ICCIT 2008, IEEE, pp 219–224

Foroughi H, Rezvanian A, Paziraee A (2008b) Robust fall detection using human shape and multi-class support vector machine. In: Sixth Indian conference on computer vision, graphics and image processing, 2008. ICVGIP’08, IEEE, pp 413–420

Foucher S, Lalonde M, Gagnon L (2011) A system for airport surveillance: detection of people running, abandoned objects, and pointing gestures. In: International society for optics and photonics SPIE defense, security, and sensing, p 805610

Ghazal M, Vázquez C, Amer A (2007) Real-time automatic detection of vandalism behavior in video sequences. In: IEEE international conference on systems, man and cybernetics, 2007. ISIC, IEEE, pp 1056–1060

Ghazal M, VáZquez C, Amer A (2012) Real-time vandalism detection by monitoring object activities. Multimed Tools Appl 58(3):585–611

Gouaillier V, Fleurant A (2009) Intelligent video surveillance: promises and challenges. Technological and commercial intelligence report. CRIM Technôpole Def Secur 456:468–561
Gowsikhaa D, Manjunath AS, Abirami S (2012) Suspicious human activity detection from surveillance videos. Int J Internet Distrib Comput Syst 2(2):141–149
Gracia IS, Suarez OD, Garcia GB, Kim TK (2015) Fast fight detection. PLoS ONE 10(4):1–19
Gubbi J, Marusic S, Palaniswami M (2009) Smoke detection in video using wavelets and support vector machines. Fire Saf J 44(8):1110–1115
Guler S, Silverstein J, Pushee IH, et al (2007) Stationary objects in multiple object tracking. In: IEEE conference on advanced video and signal based surveillance. AVSS 2007, IEEE, pp 248–253
Habiboğlu YH, Günay O, Çetin AE (2012) Covariance matrix-based fire and flame detection method in video. Mach Vis Appl 23(6):1103–1113
Han J, Ma KK (2002) Fuzzy color histogram and its use in color image retrieval. IEEE Trans Image Process 11(8):944–952
Höferlin M, Höferlin B, Weiskopf D, Heidemann G (2015) Uncertainty-aware video visual analytics of tracked moving objects. J Spatial Inf Sci 2:87–117
Hsieh CT, Hsu SB, Han CC, Fan KC (2011) Abnormal event detection using trajectory features. J Inf Technol Appl 5(1):22–27
Human fall detection dataset (2014). http://foe.mmu.edu.my/digitalhome/fallvideo.zip
Hu W, Tan T, Wang L, Maybank S (2004) A survey on visual surveillance of object motion and behaviors. IEEE Trans Syst Man Cybern Part C Appl Rev 34(3):334–352
Ibrahim N, Mokri SS, Siong LY, Mustafa MM, Hussain A (2010) Snatch theft detection using low level. In: Proceedings of the world congress on engineering, vol 2
Ibrahim N, Mustafa MM, Mokri SS, Siong LY, Hussain A (2012) Detection of snatch theft based on temporal differences in motion flow field orientation histograms. Int J Adv Comput Technol 4(12):308–317
i-lids dataset for advanced video and signal based (2007) surveillance, AVSS 2007. http://www.eecs.qmul.ac.uk/andrea/avss2007v.html
Jalal AS, Singh V (2012) The state-of-the-art in visual object tracking. Informatica 36(3):227–248
Jiang F, Yuan J, Tsaftaris SA, Katsaggelos AK (2011) Anomalous video event detection using spatiotemporal context. Comput Vis Image Underst 115(3):323–333
Juang CF, Chang CM (2007) Human body posture classification by a neural fuzzy network and home care system application. IEEE Trans Syst ManCybern Part A Syst Humans 37(6):984–994
Kalman RE (1960) A new approach to linear filtering and prediction problems. J Fluids Eng 82(1):35–45
Kamijo S, Matsushita Y, Ikeuchi K, Sakauchi M (2000) Traffic monitoring and accident detection at intersections. IEEE Trans Intell Transp Syst 1(2):108–118
Kausalya K, Chitrakala S (2012) Idle object detection in video for banking ATM applications. Res J Appl Sci Eng Technol 4(4):5350–5356
Ke SR, Thuc HLU, Lee YJ, Hwang JN, Yoo JH, Choi KH (2013) A review on video-based human activity recognition. Computers 2(2):88–131
Khan Z, Sohn W et al (2011) Abnormal human activity recognition system based on r-transform and kernel discriminant technique for elderly home care. IEEE Trans Consumer Electron 57(4):1843–1850
Kim K, Chalidabhongse TH, Harwood D, Davis L (2005) Real-time foreground–background segmentation using codebook model. Real-time Imaging 11(3):172–185
Kitagawa G (1987) Non-gaussian state space modeling of nonstationary time series. J Am Stat Assoc 82(400):1032–1041
Kong H, Audibert JY, Ponce J (2010) Detecting abandoned objects with a moving camera. IEEE Trans Image Process 19(8):2201–2210
Lai TY, Kuo JY, Fanjiang YY, Ma SP, Liao YH (2012) Robust little flame detection on real-time video surveillance system. In: Third international conference on innovations in bio-inspired computing and applications (IBICA), 2012, IEEE, pp 139–143
Lavee G, Khan L, Thuraisingham B (2005) A framework for a video analysis tool for suspicious event detection, pp 79–84
Lavee G, Khan L, Thuraisingham B (2007) A framework for a video analysis tool for suspicious event detection. Multimed Tools Appl 35(1):109–123
Lee JT, Ryoo MS, Riley M, Aggarwal J (2009) Real-time illegal parking detection in outdoor environments using 1-d transformation. IEEE Trans Circuit Syst Video Technol 19(7):1014–1024
Lei W, Liu J (2013) Early fire detection in coalmine based on video processing. Proceedings of the 2012 international conference on communication, electronics and automation engineering. Springer, Berlin, pp 239–245
Li Q, Mao Y, Wang Z, Xiang W (2009) Robust real-time detection of abandoned and removed objects. In: Fifth international conference on image and graphics, 2009. ICIG'09, IEEE, pp 156–161

Li X, Zhang C, Zhang D (2010) Abandoned objects detection using double illumination invariant foreground masks. In: 20th international conference on pattern recognition (ICPR), 2010, IEEE, pp 436–439
Liao HH, Chang JY, Chen LG (2008) A localized approach to abandoned luggage detection with foreground-mask sampling. In: IEEE Fifth international conference on advanced video and signal based surveillance, 2008. AVSS'08., IEEE, pp 132–139
Lin CW, Ling ZH, Chang YC, Kuo CJ, (2005) Compressed-domain fall incident detection for intelligent home surveillance. In: IEEE international symposium on circuits and systems, (2005) ISCAS 2005, IEEE, pp 3781–3784
Liu CL, Lee CH, Lin PM (2010) A fall detection system using k-nearest neighbor classifier. Expert Syst Appl 37(10):7174–7181
Liu H, Zuo C (2012) An improved algorithm of automatic fall detection. AASRI Procedia 1:353–358
Lo B, Velastin S (2001) Automatic congestion detection system for underground platforms. In: Proceedings of 2001 international symposium on intelligent multimedia, video and speech processing, 2001, IEEE, pp 158–161
M E (2011) Caviar dataset 2011: Fight and one man down demo. http://www.cim.mcgill.ca/mndhamod/thesisvideos/caviarfightonemandown.avi
Maddalena L, Petrosino A (2013) Stopped object detection by learning foreground model in videos. IEEE Trans Neural Netw Learn Syst 24(5):723–735
Magno M, Tombari F, Brunelli D, Di Stefano L, Benini L (2009) Multimodal abandoned/ removed object detection for low power video surveillance systems. In: Sixth IEEE international conference on advanced video and signal based surveillance, 2009. AVSS'09, IEEE, pp 188–193
Manjunatha KC, Mohana HS, Vijaya PA (2015) Implementation of computer vision based industrial fire safety automation by using neuro-fuzzy algorithms. Int J Inf Technol Comput Sci 7(4):14–27
McHugh JM, Konrad J, Saligrama V, Jodoin PM (2009) Foreground-adaptive background subtraction. Signal Process Lett IEEE 16(5):390–393
Mesh (2007), multimedia semantic syndication for enhanced news service. In: IST 6th framework programme European Commission Project. http://www.mesh-ip.eu/
Miguel JCS, Mart'ınez JM (2008) Robust unattended and stolen object detection by fusing simple algorithms. In: IEEE fifth international conference on advanced video and signal based surveillance, 2008. AVSS'08, IEEE, pp 18–25
Mukherjee D, Wu Q, Nguyen TM (2014) Gaussian mixture model with advanced distance measure based on support weights and histogram of gradients for background suppression. IEEE Trans Ind Inf 10(2):1086–1096
Nam Y (2016) Real-time abandoned and stolen object detection based on spatio-temporal features in crowded scenes. Multimed Tools Appl 75(12):7003–7028
Nasution AH, Emmanuel S (2007) Intelligent video surveillance for monitoring elderly in home environments. In: IEEE 9th workshop on multimedia signal processing, 2007. MMSP 2007, IEEE, pp 203–206
Nguyen TT, Cho MC, Lee TS (2009) Automatic fall detection using wearable biomedical signal measurement terminal. In: Annual international conference of the IEEE engineering in medicine and biology society, 2009. EMBC 2009, IEEE, pp 5203–5206
Pavithradevi MK, Aruljothi S (2014) Detection of suspicious activities in public areas using staged matching technique. IJAICT 1(1):140–144
Penmetsa S, Minhuj F, Singh A, Omkar SN (2014) Autonomous UAV for suspicious action detection using pictorial human pose estimation and classification ELCVIA Electron Lett Comput Vis Image Anal 13(1):18–32
Pets 2001 benchmark data (2001). http://www.cvg.rdg.ac.uk/pets2001/
Pets 2006 benchmark data (2006). http://www.cvg.rdg.ac.uk/PETS2006/data.html
Pets 2007 benchmark data (2007). http://www.cvg.rdg.ac.uk/pets2006/data.html
Piccardi M (2004) Background subtraction techniques: a review. In: IEEE international conference on systems, man and cybernetics, 2004, IEEE, vol 4, pp 3099–3104
Popoola OP, Wang K (2012) Video-based abnormal human behavior recognitiona review. IEEE Trans Syst Man Cybern Part C Appl Rev 42(6):865–878
Poppe R (2010) A survey on vision-based human action recognition. Image Vision Comput 28(6):976–990
Porikli F (2007) Detection of temporarily static regions by processing video at different frame rates. In: IEEE conference on advanced video and signal based surveillance, 2007. AVSS 2007, IEEE, pp 236–241
Porikli F, Ivanov Y, Haga T (2008) Robust abandoned object detection using dual foregrounds. EURASIP J Adv Signal Process 2008:30
Prabhakar G, Ramasubramanian B (2012) An efficient approach for real time tracking of intruder and abandoned object in video surveillance system. Int J Comput Appl 54(17):22–27

Pteri FSHM R (2012) Dyntex: a comprehensive database of dynamic textures 2012. Pattern Recogn Lett. http://projects.cwi.nl/dyntex/database.html

Qmul junction dataset (2010). http://www.eecs.qmul.ac.uk/ccloy/downloadsqmuljunction.html

Rabiner LR (1989) A tutorial on hidden markov models and selected applications in speech recognition. Proc IEEE 77(2):257–286

Rougier C, Meunier J, St-Arnaud A, Rousseau J (2011) Robust video surveillance for fall detection based on human shape deformation. IEEE Trans Circuit Syst Video Technol 21(5):611–622

Ryoo M, Aggarwal J (2011) Stochastic representation and recognition of high-level group activities. Int J Comput Vision 93(2):183–200

Sacchi C, Regazzoni CS (2000) A distributed surveillance system for detection of abandoned objects in unmanned railway environments. IEEE Trans Veh Technol 49(5):2013–2026

Sadek S, Al-Hamadi A, Michaelis B, Sayed U (2010) A statistical framework for real-time traffic accident recognition. J Signal Inf Process 1(01):77

Sadeky S, Al-Hamadiy A, Michaelisy B, Sayed U (2010) Real-time automatic traffic accident recognition using HFG. In: 20th International conference on pattern recognition (ICPR), 2010, IEEE, pp 3348–3351

Sajith K, Nair KR (2013) Abandoned or removed objects detection from surveillance video using codebook. Int J Eng Res Technol 2:401–406

Sample fire and smoke video clips (2009). http://signal.ee.bilkent.edu.tr/visifire/demo/sampleclips.html

SanMiguel J, Caro L, Martinez J (2012) Pixel-based colour contrast for abandoned and stolen object discrimination in video surveillance. Electron Lett 48(2):86–87

Seebamrungsat J, Praising S, Riyamongkol P (2014) Fire detection in the buildings using image processing. In: Third ICT international student project conference (ICT-ISPC), 2014, IEEE, pp 95–98

Singh R, Vishwakarma S, Agrawal A, Tiwari M (2010) Unusual activity detection for videosurveillance. In: Proceedings of the first international conference on intelligent interactive technologies and multimedia. ACM, pp 297–305

Snoek J, Hoey J, Stewart L, Zemel RS, Mihailidis A (2009) Automated detection of unusual events on stairs. Image Vis Comput 27(1):153–166

Soomro K, Zamir AR, Shah M (2012) Ucf101: a dataset of 101 human actions classes from videos in the wild. arXiv preprint arXiv:1212.0402

Stauffer C, Grimson WEL (1999) Adaptive background mixture models for real-time tracking. In: IEEE computer society conference on computer vision and pattern recognition, 1999, IEEE, vol 2

Stauffer C, Grimson WEL (2000) Learning patterns of activity using real-time tracking. IEEE Trans Pattern Anal Mach Intel 22(8):747–757

Sternig S, Roth PM, Grabner H, Bischof H (2009) Robust adaptive classifier grids for object detection from static cameras. In: Proceedings computer vision winter workshop

Sujith B (2014) Crime detection and avoidance in ATM: a new framework. Int J Comput Sci Inf Technol 5(5):6068–6071

Tan X, Triggs B (2007) Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: Analysis and modeling of faces and gestures. Springer, Berlin, pp 168–182

Tejas Naren TN VKSSLSC Shankar SiddharthKA (2014) Abandoned object detection for automated video surveillance using hadoop. Int J Adv Res Electr Electr Instrum Eng 3:101–107

Thome N, Miguet S (2006) A hhmm-based approach for robust fall detection. In: 9th International conference on control, automation, robotics and vision, 2006. ICARCV'06, IEEE, pp 1–8

Thome N, Miguet S, Ambellouis S (2008) A real-time, multiview fall detection system: alhmm-based approach. IEEE Trans Circuit Syst Video Technol 18(11):1522–1532

Tian Y, Feris RS, Liu H, Hampapur A, Sun MT (2011) Robust detection of abandoned and removed objects in complex surveillance videos. IEEE Trans Syst Man Cybern Part C Appl Rev 41(5):565–576

Tian Y, Senior A, Lu M (2012) Robust and efficient foreground analysis in complex surveillance videos. Mach Vis Appl 23(5):967–983

Tomasi C, Kanade T (1991) Detection and tracking of point features. School of Computer Science, Carnegie Mellon Univ, Pittsburgh

Töreyin BU, Dedeoglu Y et al (2005) Flame detection in video using hidden markov models. In: IEEE international conference on image processing, 2005. ICIP 2005, IEEE, vol 2, pp II–1230

Töreyin BU, Dedeoglu Y, Güdükbay U, Cetin AE (2006) Computer vision based method for real-time fire and flame detection. Pattern Recogn Lett 27(1):49–58

Töreyin BU, et al (2007) Online detection of fire in video. In: IEEE conference on computer vision and pattern recognition, 2007. CVPR'07, IEEE, pp 1–5

Traffic videos from the next generation simulation (2007). http://ngsim.camsys.com/

Tripathi RK, Jalal AS (2014) A framework for suspicious object detection from surveillance video. Int J Mach Intel Sensory Signal Process 1(3):251–266

Tripathi RK, Jalal AS, Bhatnagar C (2013) A framework for abandoned object detection from video surveillance. In: Fourth national conference on computer vision, pattern recognition, image processing and graphics (NCVPRIPG), 2013, IEEE, pp 1–4

Tripathi V, Gangodkar D, Latta V, Mittal A (2015) Robust abnormal event recognition via motion and shape analysis at ATM installations. J Electr Comput Eng 2015. doi:10.1155/2015/502737

Trecvid 2010 evaluation for surveillance detection (2010). http://www.itl.nist.gov/iad/mig/tests/trecvid/2010/

Vezzani R, Cucchiara R (2010) Video surveillance online repository (visor): an integrated framework. Multimed Tools Appl 50(2):359–380

Vicente J, Guillemant P (2002) An image processing technique for automatically detecting forest fire. Int J Therm Sci 41(12):1113–1120

Vu VT, Brémond F, Thonnat M (2002) Temporal constraints for video interpretation. In 15th European conference on artificial intelligence

Wang X, Ma X, Grimson E (2009) Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models. IEEE Trans Pattern Anal Mach Intel 31(2):539–555

Wang S, Chen L, Zhou Z, Sun X, Dong J (2016) Human fall detection in surveillance video based on PCANet. Multimed Tool Appl 75(19):11603–11613

Wieser D, Brupbacher T (2001) Smoke detection in tunnels using video images. NIST SPECIAL PUBLICATION SP, pp 79–90

Wiliem A, Madasu V, Boles W, Yarlagadda P (2012) A suspicious behaviour detection using a context space model for smart surveillance systems. Comput Vis Image Underst 116(2):194–209

Willems J, Debard G, Bonroy B, Vanrumste B, Goedemé T (2009) How to detect human fall in video? In: An overview, positioning and context awareness international conference, POCA

Wren CR, Azarbayejani A, Darrell T, Pentland AP (1997) Pfinder: real-time tracking of the human body. IEEE Trans Pattern Anal Mach Intell 19(7):780–785

Yang Z, Rothkrantz L (2011) Surveillance system using abandoned object detection. In: Proceedings of the 12th international conference on computer systems and technologies. ACM, pp 380–386

Yilmaz A, Javed O, Shah M (2006) Object tracking: a survey. ACM Comput Surv 38(4):13

Yogameena B, Deepila G, Mehjabeen J (2012) RVM based human fall analysis for video surveillance applications? Res J Appl Sci Eng Technol 4(24):5361–5366

Yu M, Rhuma A, Naqvi SM, Wang L, Chambers J (2012) A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment. IEEE Trans Inf Technol Biomed 16(6):1274–1286

Yuan F (2008) A fast accumulative motion orientation model based on integral image for video smoke detection. Pattern Recogn Lett 29(7):925–932

Yuan F (2010) An integrated fire detection and suppression system based on widely available video surveillance. Mach Vis Appl 21(6):941–948

Zhou Y, Benois-Pineau J, Nicolas H (2010) Multi-object particle filter tracking with automatic event analysis. In: Proceedings of the first ACM international workshop on analysis and retrieval of tracked events and motion in imagery streams. ACM, pp 21–26

Zhou Z, Chen X, Chung YC, He Z, Han TX, Keller JM (2008) Activity analysis, summarization, and visualization for indoor human activity monitoring. IEEE Trans Circuit Syst Video Technol 18(11):1489–1498

Ziaeefard M, Bergevin R (2015) Semantic human activity recognition: a literature review. Pattern Recogn 8(48):2329–2345

Zin TT, Tin P, Toriu T, Hama H (2012a) A novel probabilistic video analysis for stationary object detection in video surveillance systems. IAENG Int J Comput Sci 39(3):295–306

Zin TT, Tin P, Toriu T, Hama H (2012b) A probability-based model for detecting abandoned objects in video surveillance systems. In: Proceedings of the world congress on engineering, vol 2