Expert Systems With Applications 91 (2018) 480–491

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Review

Abnormal behavior recognition for intelligent video surveillance systems: A review

Amira Ben Mabrouk*, Ezzeddine Zagrouba

University of Tunis EL Manar, Higher Institute of Computer, Research Team on Intelligent Systems in Imaging and Artificial Vision (SIIVA) – Lab LIMTIC, Aryanah 2036 Tunisia

A R T I C L E I N F O

Article history:
Received 30 April 2017
Revised 10 September 2017
Accepted 11 September 2017
Available online 11 September 2017

Keywords:
Computer vision
Video surveillance system
Behavior representation
Behavior modeling

A B S T R A C T

With the increasing number of surveillance cameras in both indoor and outdoor locations, there is a grown demand for an intelligent system that detects abnormal events. Although human action recognition is a highly reached topic in computer vision, abnormal behavior detection is lately attracting more research attention. Indeed, several systems are proposed in order to ensure human safety. In this paper, we are interested in the study of the two main steps composing a video surveillance system which are the behavior representation and the behavior modeling. Techniques related to feature extraction and description for behavior representation are reviewed. Classification methods and frameworks for behavior modeling are also provided. Moreover, available datasets and metrics for performance evaluation are presented. Finally, examples of existing video surveillance systems used in real world are described.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction and related work

Human action recognition in videos is an active field in computer vision which is attracting more research attention in recent years. This topic becomes very important for many applications such as video surveillance, scene modeling and video content annotation and retrieval. Several previous surveys about human motion detection and analysis (Aggarwal & Cai, 1999, 1997; Ji & Liu, 2010), behavior analysis and understanding (Pantic, Pentland, Nijholt, & Huang, 2007; Teddy, 2008) and activity recognition (Shian-Ru et al., 2013) were published (Table 1). Recently, Dawn, Debapratim, Shaikh, and SoharabHossain (2015) and Hassan et al. (2014) reviewed several methods based on computer vision techniques to recognize simple activities performed by a single person such as running and walking. Bux, Plamen, and Zulfiqar (2017) reviewed techniques relative to the different phases of human activity recognition which are object segmentation, feature extraction and representation and activity classification. Particularly, Sarvesh and Anupam (2013) and Mishra and Bhagat (2015) presented techniques for motion analysis and activity recognition in video surveillance applications. Indeed, the detection of an abnormal behavior in video surveillance is essential to ensure safety in both outdoor and indoor places such as train stations and airports. In fact, abnormal behavior detection is a particular problem of human action recognition. With the increasing number of surveillance cameras, the task of supervising multiple monitors by security agents becomes very difficult due to the human inattention and fatigue. Besides, abnormal events are relatively rare and don't occur frequently. This make the supervision task more complex and challenging. Therefore, there is a growing demand for an intelligent video surveillance system that detects, automatically, an abnormal behavior and rises an alarm.

In fact, previous surveys of video surveillance systems were provided. For example, Valera and Velastin (2005) presented an overview of automated surveillance systems for anomaly detection. Oluwatoyin and Kejun (2012) provided a survey of abnormal human behavior methods in video surveillance applications within different contexts. More recently, Zablocki, K., D., and R. (2014) reviewed different techniques used by intelligent surveillance systems to monitor public spaces. Abnormal event detection methods in crowd surveillance videos were surveyed by de Campos (2014) and Teng et al. (2015). In this paper, we extensively review the existing methods that are used in video surveillance applications and we highlight the current advances in the field of abnormal behavior detection.

The objective of an intelligent video surveillance system is to detect efficiently an interesting event from a large amount of videos in order to prevent dangerous situations. Generally, this task requires two video processing levels as shown in Fig. 1. The first one consists of two steps. First, low level features, aiming to

* Corresponding author.
E-mail addresses: amira_benmabrouk@yahoo.fr (A. Ben Mabrouk), ezzeddine.zagrouba@uvt.tn (E. Zagrouba).

http://dx.doi.org/10.1016/j.eswa.2017.09.029
0957-4174/© 2017 Elsevier Ltd. All rights reserved.

Table 1
Recent related surveys.

Main focus / Topic	*Reference*
Three phases of human activity recognition.	(Bux et al., 2017)
Human action detection and labeling.	(Dhulekar, T., Chitte, & Pardeshi, 2017)
Anomaly detection based on smart home sensors.	(Bakar, Hemant, Hasanm,& Mukhopadhyay, 2016)
Gesture recognition systems.	(Siddharth & Anupam, 2015)
Actions recognition based on spatio-temporal points.	(Dawn et al., 2015)
Human motion detection.	(Mishra & Bhagat, 2015)
Crowded scene analysis.	(Teng et al., 2015)
Actions recognition in crowded surveillance scene.	(de Campos, 2014)
Human action recognition.	(Hassan et al., 2014)
Intelligent video surveillance systems in public spaces.	(Zablocki et al., 2014)
Human activity recognition and understanding.	(Sarvesh & Anupam, 2013)
Human activity recognition.	(Shian-Ru et al., 2013)
Abnormal human behavior recognition.	(Oluwatoyin & Kejun, 2012)
Behavior analysis for homeland security applications.	(Teddy, 2008)
Automated video surveillance systems.	(Valera & Velastin, 2005)
Human motion analysis.	(Aggarwal & Cai, 1999; 1997)

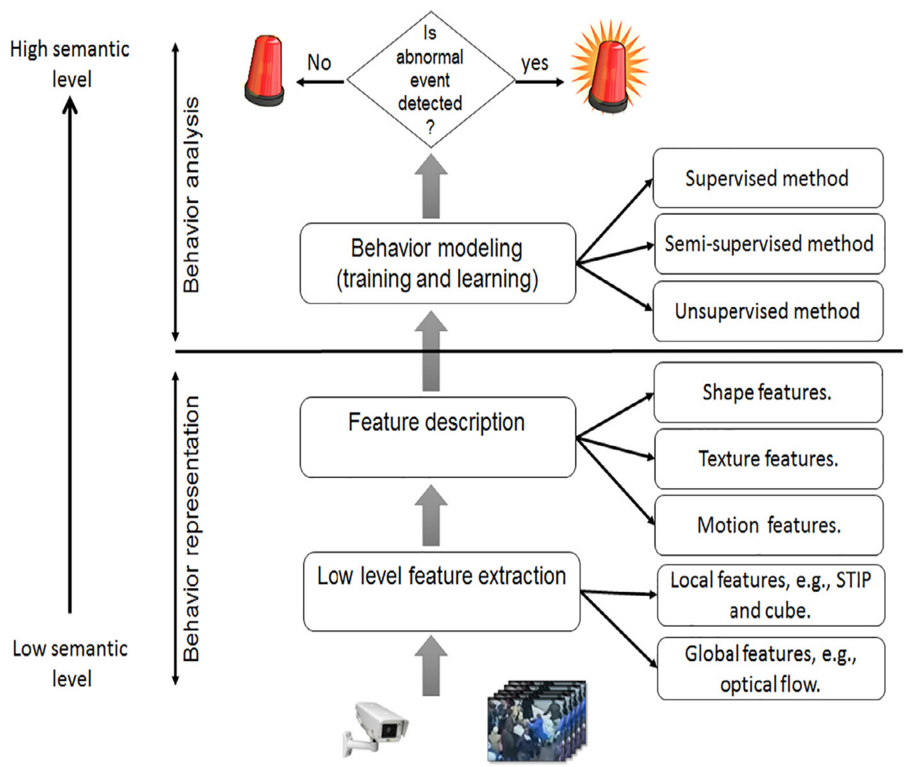

Fig. 1. Intelligent video surveillance system.

detect the interest region in the scene, are extracted. Then, primitives based on low level features are generated to describe the interest region. The second level provides semantic information about the human action and determines whether the behavior is normal or not.

The rest of this paper is organized as follows. In Section 2, the most important techniques for behavior representation including feature extraction and description are presented. Different frameworks and classification methods for behavior modeling both in crowded and uncrowded scenes are reviewed in Section 3. In Section 4, the most popular datasets used to evaluate a video surveillance system are first presented. Then, a performance evaluation of previous works are provided. Examples of existing video surveillance systems in real world are described in Section 5. Finally, a conclusion and a discussion of this review are provided in Section 6.

2. Behavior representation

Behavior representation is the low level processing step of behavior analysis. It aims to capture relevant features describing the target object in the video. It consists of two steps. First, the interest region in the scene is detected based on low level features. Then, a description for this region (target object) is provided. In fact, this level is difficult and challenging because it influences significantly the understanding of the interest object behavior. Indeed, the major challenge in behavior representation is to find suitable features that are robust to many transformations such as changes

Table 2
Popular features for behavior representation.

Feature types	*Description*	*Comments*	*References*
Optical flow based features (histograms, variation, acceleration, etc), motion information.	Extraction of statistical proprieties from optical flow vector for motion characterization.	-Global features. -Computational extensive.	(Cho & Kang, 2014; Chundi et al., 2015; Cong et al., 2013; Gnanavel & Srinivasan, 2015; Gu et al., 2014; Hajananth et al., 2014; Huang & Chen, 2014; Huo, Gao, Yang, & Yin, 2014; Leyva et al., 2014; Nannan et al., 2015; Rasheed et al., 2014; Rezaee et al., 2015; Tao et al., 2017; 2015; Thi-Lan & Thanh-Hai, 2015; Wang & Dong, 2014; Wang & Snoussi, 2014; Zhang, Lu, Zhang, & Ruan, 2016)
Interest points : STIP, CSTIP, MoSIFT, etc.	Salient points detected in both space and time domains. Representation of significant motion variations corresponding to irregular actions.	-Local features. -Low computational time. -Sensitive to noise. -Not adapted to crowded scene.	(Bellamine & Tairi, 2015; 2016; Bermejo et al., 2011; Mabrouk & Zagrouba, 2017; Ming-yu et al., 2010; Xu, Gong, et al., 2014; Zhao et al., 2015)
Spatio-temporal volume, cube, blob, etc.	Obtaining the third (temporal) dimension by gathering consecutive frames.	-Local features. -Memory intensive. -Sensitive to parameters (number of consecutive frames, size of the cube, etc)	(Rao et al., 2014; Li et al., 2015; Songhao et al., 2016; Wang & Xu, 2015; Xu, Song, et al., 2014; Yogameena & Priya, 2015)
Shape (silhouette, HOG, rectangle, bounding box, etc).	Describing the shape of the moving object in sequential frames. Abnormal behavior corresponds to shape change detection.	-Widely used for falls detection (shape change). -Not adapted to crowded scene.	(Aslan et al., 2015; Miao & Song, 2014; Nguyen et al., 2014; Zhao & Li, 2014)
Texture (moments, GLCM, MDT, wavelets, etc).	Extraction of local patterns (texture features) for each moving object included in a bounding box, rectangle, etc.	Adapted for crowd monitoring (detecting changes of patterns).	(Rao et al., 2014; Li et al., 2014; Mahadevan et al., 2010; Miao & Song, 2014; Wang & Xu, 2015; Zhang, Dong, et al., 2014)
Object tracking and trajectory extraction.	Tracking the moving object using trajectory (coordinates in each frame), optimization algorithm, etc.	-Adapted for tracking a single person. -Not adapted for crowded scene. -Background must be static.	(Arroyo et al., 2015; Ce et al., 2013; Conci & Lizzi, 2009; Himanshu et al., 2015; Ko & Yoo, 2013; Lai et al., 2009; Leach, Sparks, et al., 2014b; Ngo, Do, & Nguyen, 2016; Rajkumar et al., 2017; Su et al., 2014; SungChun & Ram, 2014; Zhang et al., 2008; Zhang, Lin, et al., 2014)

in the background and on the object appearance. In Table 2, the most used features for behavior representation are presented.

To represent the target object, different aspects may be described such as the shape (Aslan et al., 2015; Nguyen et al., 2014; Zhao & Li, 2014) and the texture (Zhang, Dong, Li, & Li, 2014). For example, Li, Mahadevan, and Vasconcelos (2014) and Mahadevan, Li, Bhalodia, and Vasconcelos (2010) presented a method for the detection of a crowd abnormal behavior that is based on mixtures of dynamic textures (MDT) model. Miao and Song (2014) used gray level co-occurrence matrix (GLCM), HU invariant moments and histogram of oriented gradients (HOG) to extract respectively texture, shape and motion features from the video. Rao, Gubbi, Rajasegarar, Marusic, and Palaniswami (2014) used also GLCM based features to detect crowd anomaly. Spatio-temporal texture features are used by Wang and Xu (2015) to detect abnormal behaviors in a crowded scene based on wavelet transformation. However, the most important aspect that is used to represent the interest object in the video is motion information. Multiple features are used to detect and describe an object moving across the time. Those features can be classified into local and global features.

Local features are detected in a predefined region in the frame. This region may be represented by an interest points or a local area. For instance, Bermejo, Deniz, Bueno, and Rahul (2011), Ming-yu, Lily, Padmanabhan, Alexander, and Rahul (2010), and Xu, Gong, Yang, Wu, and Yao (2014) described the local motion of the interest objects using Motion Scale-Invariant Feature Transform (MoSIFT) which is extended from the popular SIFT descriptor (Lowe, 2004). Li, Wu, Xu, Guo, and Feng (2015) detected local anomalies based on spatio temporal video cube analysis. A local abnormal behavior detection method based on spatio-temporal blobs extraction is proposed by Songhao, Juanjuan, and Zhe (2016). Indeed, the local abnormal blob is detected by using a statistical model. Other works focused on detecting Spatio Temporal Interest Points (STIP) to extract local features from the input video. In fact, STIP points which are proposed by Laptev (2005) are detected in both space and time domains. They are salient points representing significant motion variations which correspond to irregular action. For example, Zhao, Yu, Jie, and Nikola (2015) extracted local spatio-temporal descriptor named HNF around STIP points. In fact, HNF is a combinational descriptor vector of HOG and Histogram of Optical Flow (HOF) which are used to describe respectively the appearance and the action. Bellamine and Tairi (2015, 2016) proposed a color version of STIP feature, named Color STIP (CSTIP). It consists in introducing the color information around each STIP point to detect motion in the video.

Global features are used to describe motion in the entire frame. Optical flow features are commonly used to extract global motion information (Huang & Chen, 2014; Rasheed, S.A., & A., 2014). For example, an approach based on moving particles extracted using optical flow is proposed by (Gu, Cui, & Zhu, 2014) to detect abnormal behavior in a crowded scene. Hajananth, Fookes, Denman, and Sridharan (2014) proposed two optical flow based features : optical flow acceleration and histogram of optical flow gradients to represent events. In fact, histograms based on optical flow are usually used for abnormal behavior recognition (Mahadevan et al., 2010). A descriptor based on optical flow named Multiscale Histogram of Optical Flow (MHOF) is proposed by Cong, Yuan, and Liu (2013).

MHOF is combined with Edge Oriented Histogram (EOH) to obtain motion context for anomaly detection in crowded scenes by Gnanavel and Srinivasan (2015). Other works used Histogram of Optical Flow Orientation (HOFO) descriptor to distinguish between normal and abnormal events. For example, Wang and Snoussi (2014) recognized abnormal events in video surveillance using the HOFO feature calculated on both the original image and the foreground obtained by background subtraction process. In the same way, Jaechul and Kristen (2009) used HOFO feature and introduced the magnitude information to detect abnormal events in crowded scene.

We note that optical flow is also widely used for violence detection which is a particular problem of abnormal behavior recognition. For instance, Tao et al. (2015) presented a fast and robust method to detect and locate violence based on Gaussian Model of Optical Flow (GMOF) and OHOF descriptor. Hassner, Itcher, and Kliper-Gross (2012a) proposed a descriptor named Violent Flows (ViF) to detect violence in crowded scene. In fact, this feature is based on optical flow magnitude changes to identify violence in real time. Gao, Liu, Sun, Wang, and Liu (2016) proposed an extension of ViF called oriented ViF (OViF). This descriptor uses both magnitude and orientation information obtained from optical flow in order to characterize, accurately, the action. Tao et al. (2017) proposed a Motion Weber Local Descriptor (MoWLD) for violence detection. In fact, MoWLD is based on both optical flow information and Weber Local Descriptor (WLD) which is successfully used for face recognition (Li, Gong, & Yuan, 2013; Wang, Li, Yang, & Liao, 2011).

Motion information, for object tracking, can be extracted using optimization algorithms, e.g. Ant Colony (Lai, Chang, & Zhong, 2009) and Particle Swarm Optimization (PSO) techniques (Conci & Lizzi, 2009; Zhang, Hu, Maybank, Li, & Zhu, 2008). In fact, Zhang et al. (2008) incorporated the temporal continuity information which is required for the tracking process into the PSO algorithm in order to make it suitable for the tracking of a moving object. Besides, motion can be estimated using several methods such as Kalman filter based tracking method (Rezaee, Haddadnia, & Delbari, 2015), motion vectors obtained from the video (Chundi, Jianbin, Wei, Tong, & Peiqin, 2015), motion matrix using tracking features relative to the object blob (Wang & Dong, 2014), optical flow variation (Cho & Kang, 2014), wake motion descriptors (Leyva, Sanchez, & Chang-Tsun, 2014) and Motion History Image (MHI) method (Thi-Lan & Thanh-Hai, 2015).

Particularly, we note that the trajectory of the moving object is widely used to determine whether the behavior is normal or abnormal. Several previous works analyzed the behavior of the target object based on its trajectory (Arroyo, Yebes, Bergasa, Daza, & Almazan, 2015; Himanshu, Maheshkumar, Neelabh, & Mukherjee, 2015; Su, sheng Yin, Hailong, & Zhiyong, 2014). For example, Ce, Zhenjun, Qixiang, and Jianbin (2013) proposed a method to detect abnormal behavior based on the object trajectory analysis. First, they constructed a dictionary using the trajectories of normal behaviors. Then, they classified each tested trajectory into a normal or an abnormal one. Leach, Sparks, and Robertson (2014b) proposed an anomaly detection method using the pedestrian trajectories. Moreover, Zhang, Lin, et al. (2014) presented an abnormal activity detection method based on blob trajectory optimization process. Using the tracked person trajectory, a system which is structured hierarchically into two levels is proposed by SungChun and Ram (2014). The first level, namely low level process, analyzes the trajectory and generates a real-time alarm when a suspicious event is detected. In the second level (high level process), the proposed system verifies whether the detected event is triggered by humans or not. Rajkumar, Arif, Prosad, and Pratim (2017) are based on object trajectories to extract high level features for surveillance scene segmentation.

Although local features represent accurately the local motion in the video, they may not produce significant information about the action when there is too much motion. On the other hand, global features provide holistic information of the whole scene but they generally give irrelevant information in case of cluttered background and noise. Based on both local and global features, Mabrouk and Zagrouba (2017) proposed a spatio-temporal descriptor, named Distribution of Magnitude and Orientation of Local Interest Frame (DiMOLIF), for violence detection. Indeed, DiMOLIF is based on the bivariate distribution estimation of the optical flow magnitude and orientation around STIP points.

3. Classifying abnormal behavior recognition methods

Abnormal behavior detection in video surveillance is a challenging task in computer vision and has seen lately important advances. The low level processing stages allow detecting and describing the moving object in the scene. However, those steps do not allow to understand the type of the action performed by the moving object or to determine if its behavior is normal or not. Since there are multiple works proposed that are related to abnormal behavior recognition in video surveillance, we aim in this review to group those papers according to :

1. Modeling frameworks and classification methods.
2. Scene density and moving object interaction.

3.1. Modeling frameworks and classification methods

Recognizing an abnormal behavior depends on the proposed framework and the method used to classify behaviors. Given the type of the samples required for the learning process (normal or abnormal), classification methods can be categorized into supervised, semi-supervised and unsupervised methods. In Table 3, we describe and compare the three categories of classification methods.

Supervised methods aim to model normal and abnormal behaviors via labeled data. They are generally designed to detect specific abnormal behaviors predefined in the training phase such as fighting detection (Bermejo et al., 2011; Guang, Fu, Li, & Geng, 2014), loitering detection (Gomez et al., 2015) and falling detection (Stone & Skubic, 2015). Several supervised methods are proposed in the literature aiming to detect an interesting event in a video. One of the most popular is the Bag of Words (BOW) approach (Foggia, Percannella, Saggese, & Vento, 2013). It consists on representing each video (or each frame) using a histogram of words (local image features, trajectory, etc). First, a dictionary of words is constructed. Then, the histogram is computed by counting the frequency of each word within the dictionary in the video. Indeed, the BOW approach is generally used with the support vector machine (SVM) classifier which is an efficient tool for aggressive behavior detection (Bermejo et al., 2011) and crowd anomaly recognition (Cho & Kang, 2014; Ramin, Alexis, & Mubarak, 2009). Kim et al. (2016) proposed an abnormal behavior recognition algorithm based on human body parts estimation using geodesic graph and SVM classifier. However, the performance of the proposed algorithm is highly affected by the detection of the zone representing the human body specially in case of adjoining persons.

Semi supervised methods need only normal video data for training and can be divided into rule based and model based approaches. The first category aims to develop a rule using normal patterns. Then, any sample that does not fit this rule is considered as an outlier (anomaly). For example, (Lu, Shi, & Jia, 2013) proposed a rule based method using sparse coding to detect abnormal behaviors. Although, they achieved good result

Table 3
Comparison of classification methods categories.

Method	*Description*	*Advantages*	*Limitations*
Supervised	Building a model of normal and abnormal behaviors via labeled data.	-Good results in detecting known abnormal behaviors. -Easy to run and to understand.	-Designed to detect specific behaviors, e.g. fighting, loitering and falling. -Strongly depends on the training data. -Unknown anomalies cannot be detected. -Learning all abnormal behaviors is not practical in real world.
Semi-supervised : rule based method	-Rule construction using normal patterns. -Any new sample that does not fit the rule is an anomaly.	-Easy to perform and to interpret. -Very expressive. -Close to human reasoning.	High memory and computational complexities.
Semi-supervised : model based method	-Build a model representing the normal behaviors. -Any new sample that does not respect the model is an anomaly.	-Model is easy to generate and to understand. -A new instance is rapidly classified.	-Sensitive to multiple parameters. -Unknown normal data can be identified as abnormal (false alarm).
Unsupervised	Learning using statistical proprieties extracted from unlabeled data.	-Fast and easy to perform. -No prior knowledge required.	-Time consuming for result interpretation. -Based on the assumption that abnormal behaviors are very rare compared to normal ones.

within short execution time (150 frames per second), their result is highly affected by the threshold value. Other works rely on the construction of some rules in order to classify the behavior into normal and abnormal one. A fall detection system based on rules extracted using shape features is proposed by Nguyen et al. (2014). Besides, Tani, Lablack, Ghomari, and Bilasco (2015) used rules that where obtained through an ontology based-approach to detect abnormal events in video surveillance. Castro, Delgado, Medina, and Ruiz-Lozano (2011) combined information from multiple sources (audio, video and sensor) and proposed a rule-based adaptive system using fuzzy logic for intrusion detection. Using fuzzy rules, Albusac, Vallejo, Castro-Schez, Glez-Morcillo, and Jiménez (2014) proposed a method for normality analysis of moving objects in order to detect abnormal situations, e.g. high speed. A framework based on fuzzy clustering method and several auto-encoders is presented by Chen, Tian, Zeng, and Huang (2015). In fact, this framework consists of two phases: a training phase and a testing one. In the training phase, trajectories of moving objects are extracted and grouped using the fuzzy clustering technique. Then, a set of auto encoders, obtained by training each cluster, are used in the testing phase to detect anomaly. Also, Acampora, Foggia, Saggese, and Vento (2015) proposed a human behavior analysis system that is hierarchically structured into multiple layers using a neural network-based fuzzy inference system.

In the model based methods, abnormal patterns correspond to instances that deviate from the model representing the normal behaviors. Markov Random Field (MRF) model, Gaussian Mixture Model (GMM) and Hidden Markov model (HMM) (Ouivirach, Gharti, & Dailey, 2013) are the most used models. For example, Hajananth et al. (2014) detect anomalies using GMM based MRF technique. Other classification methods are based on Gaussian Model (Zhao & Li, 2014). For instance, an approach for abnormality detection using Gaussian Process based model is proposed by Nannan et al. (2015). First, low level features are extracted using HOF to describe the patterns motion. Then, the Gaussian process model is built to produce normal behaviors distribution which will be used later to detect anomaly in videos. In Feng, Yuan, and Lu (2017), a deep GMM is used to learn normal patterns. Authors in Kai-Wen, Yie-Tarng, and Wen-Hsien (2015b) presented a hierarchical abnormal event detection and localization framework using Gaussian Process Regression (GPR) based method and STIPs features.

Unsupervised methods aim to learn normal and abnormal behaviors using statistical proprieties extracted from unlabelled data.

Fig. 2. Abnormal behaviors performed by a single person. (i) falling (ii) person in wrong place (iii) loitering.

For example, Alvar, Torsello, Miralles, and Armingol (2014) proposed an abnormal behavior method using an unsupervised learning framework based on Dominant Set. Weiya, Guohui, Boliang, and Kuihua (2015) presented an unsupervised kernel framework for anomaly detection based on feature space and support vector data description (SVDD) (Tax & Duin, 2004). Table 4 summarizes those methods and frameworks.

3.2. Scene density and moving object interaction

The density of the scene corresponds to the number of persons present within it. The choice of techniques to use in order to characterize the behavior is directly influenced by the scene density. Thus, the moving object in the scene can be a small number of persons (may be a single person) or a group of persons. Therefore, we distinguish two types of scenes. The first type, called uncrowded scene, is characterized by the presence of one or a few number of persons at the same time within the camera field. The second type is called crowded scene since it contains many persons. Table 5 shows a grouping of different abnormal behaviors based on scene density and moving object interaction.

3.2.1. Uncrowded scene

In this type of scene, we are interested on detecting an abnormal behavior performed by one or a few number of persons that are present within the camera field. When there is only one person in the scene, three major abnormal behaviors are generally considered which are falling detection, loitering and being in a wrong place (Fig. 2). The first topic which is human falling detection is an interesting task and several works proposed systems that are used to ensure safety and security especially

Table 4
Frameworks and classification methods for abnormal behavior detection.

Modeling frameworks and classification methods	*References*
Bag of words (BOW) approach	(Bermejo et al., 2011; Cho & Kang, 2014; Li et al., 2015; Ramin et al., 2009)
Mixture of Gaussian model	(Zhao & Li, 2014)
Gaussian process based method	(Nannan et al., 2015)
Gaussian process regression	(Kai-Wen et al., 2015b)
Gaussian Model of Optical Flow	(Tao et al., 2015)
Gaussian Mixture Model (GMM), deep GMM	(Feng et al., 2017; Gu et al., 2014; Rasheed et al., 2014)
GMM based Markov random field technique	(Hajananth et al., 2014)
Dynamic patch grouping (DPG) method	(Yang, Junsong, & Yandong, 2013)
Dominant set clustering method	(Alvar et al., 2014)
Probabilistic Latent Semantic Analysis based model	(Pathak, Sharang, & Mukerjee, 2015)
Sparse reconstruction model	(Ce et al., 2013; Cong et al., 2013; Li et al., 2015)
fuzzy inference system	(Acampora et al., 2015; Wiliem et al., 2012)
fuzzy clustering technique	(Albusac et al., 2014; Castro et al., 2011; Chen et al., 2015)
Neural network	(Acampora et al., 2015; Rasheed et al., 2014)
Convolutional Neural Networks model	(Chunhui et al., 2014; Sabokrou, Fayyaz, Fathy, & Klette, 2016; Shifu et al., 2016)
Unsupervised kernel learning & Hierarchical framework	(Kai-Wen et al., 2015b; Weiya et al., 2015; Xiao, Zhang, & Zha, 2015; Xu, Song, et al., 2014)
Multiple SVM classifier	(Huang & Lee, 2014)
One class SVM	(Aslan et al., 2015; Cho & Kang, 2014; Chundi et al., 2015; Guang et al., 2014; Hassner et al., 2012a; Miao & Song, 2014; Ming-yu et al., 2010; Wang & Snoussi, 2014; Xu, Gong, et al., 2014; Zhang, Lin, et al., 2014)
SVM classifier and Bayesian theory	(Kai-Wen, Yie-Tarng, & Wen-Hsien, 2015a; Kim et al., 2016)
Support Vector Data Description (SVDD) technique	(Weiya et al., 2015)
HOG -SVM	(Thi-Lan & Thanh-Hai, 2015)
Rule based method	(Nguyen et al., 2014; Tani et al., 2015)
Sparse coding technique	(Huo et al., 2014)
Algorithm based on the adjacency matrix and optical flow	(Chen & Huang, 2011)
Bayesian tracking model	(Su et al., 2014)
Hidden Markov Model	(Acharya & Gantayat, 2015; Ouivirach et al., 2013)
Coupled Hidden Markov Model	(Gunduz, Temizel, & Temizel, 2014)
Genetic algorithm	(Miao & Song, 2014)
KNN	(Gnanavel & Srinivasan, 2015)
Hyperspherical Clustering technique	(Rao et al., 2014)

Table 5
Scene density and moving object interaction based grouping.

Main topic	*Type of the interaction*			*Type of the scene*		*References*
	No Interaction	One to one Interaction	Group Interaction	Crowded	Uncrowded	
People monitoring / Falling detection	√				√	(Aslan et al., 2015; Bian et al., 2015; Nguyen et al., 2014; Rezaee et al., 2015; Stone & Skubic, 2015; Thi-Lan & Thanh-Hai, 2015; Ye et al., 2014)
Suspicious behavior detection (e.g. chasing, following)	√				√	(Arroyo et al., 2015; Chundi et al., 2015; Huang et al., 2014; Ouivirach et al., 2013; Takai, 2015; Tran et al., 2014; Wiliem et al., 2012)
Loitering	√				√	(Gomez et al., 2015; Ko & Yoo, 2013)
Being in the wrong place	√				√	(Delgado et al., 2014)
Violence in the elevator		√			√	(Guang et al., 2014; Yujie & Zengfu, 2016)
Aggressive behaviors (kicking, punching, fighting, etc)		√		√	√	(Arroyo et al., 2015; Bermejo et al., 2011; Déniz et al., 2014; Gao et al., 2016)
Escape panic / crowd anomaly			√	√		(Cong et al., 2013; Gu et al., 2014; Hassner et al., 2012a; Huang & Chen, 2014; Nannan et al., 2015; Ramin et al., 2009; Santhiya et al., 2014; Wang, Fu, et al., 2016a; Wang & Xu, 2015; Wang et al., 2014; Wang, He, Wu, Xie, & Li, 2016; Yogameena & Priya, 2015)
Special areas monitoring (pedestrians walkway, subway station, roads, etc)			√	√		(Alvar et al., 2014; Rao et al., 2014; Chaker et al., 2017; Chen et al., 2015; Cho & Kang, 2012; 2014; Hajananth et al., 2014; Huo et al., 2014; Kai-Wen et al., 2015a; Leach, Baxter, et al., 2014a; Leach, Sparks, et al., 2014b; Li et al., 2014; Liu, Li, & Ji, 2014; Yang et al., 2013)

for elderly people and for persons living alone. Ye, Li, Zhao, and Liu (2014) proposed a falling detection system for aged people based on a new architecture for wireless sensor networks using 3-axis acceleration sensor. In Bian, Hou, Chau, and N. (2015), a human body parts tracking method is proposed to detect falls of elderly people. They used only one depth camera which makes the approach work even in darkness. Furthermore, Stone and Skubic (2015) proposed a two-stages falling detection system using the Microsoft Kinect (Wenbing, 2016) and decision trees. In Huang, Tian, Wu, and Zhou (2014), an algorithm that is able to recognize abnormal behavior of solitary seniors based on information obtained from an intelligent space namely ISUS (Intelligent Space for Understanding and Service) is proposed. Besides, a system was proposed by Wiliem, Madasu, Boles, and Yarlagadda (2012) that uses contextual information to detect suspicious behavior in video surveillance. This system contains three major components which

Table 6
Available datasets for video surveillance systems evaluation.

Dataset	*Videos*	*Properties*
UCSD (Mahadevan et al., 2010)	-Peds1 : 70 videos (34 for training, 36 for testing), 238 × 158 pixels. -Peds 2 : 30 videos (16 for training, 14 for testing), 360 × 240 pixels.	-Normal events correspond to only pedestrians in the walkways. -Abnormal events belong to no pedestrians in the walkways such as bikers and skaters.
UMN (Ramin et al., 2009)	-11 videos from three different indoor and outdoor scenes. -Resolution = 320 × 240 pixels.	-Normal scenarios contains people that are walking in different directions. -Abnormal scenarios correspond to crowd dispersion.
Violent Flows (Hassner et al., 2012a)	-246 videos (123 violence and 123 no violence). -Resolution = 320 × 240 pixels.	-Collected from Youtube. -Designed for violence detection in crowded places.
Action movies (Bermejo et al., 2011)	200 videos (100 violence and 100 no violence).	-Videos from action movies. -Abnormal behaviors correspond to one to one fight.
Hockey Fight (Bermejo et al., 2011)	-1000 videos (500 violence and 500 no violence). -Resolution = 720 × 576 pixels.	Designed for violence detection in uncrowded scene (ice hockey rink).
Web dataset (Ramin et al., 2009)	-20 videos (8 for training and 12 for testing). -Different resolutions.	-Normal scenarios correspond to normal pedestrians. -Abnormal scenarios are clash, escape panic, fight, etc.

Fig. 3. Samples of violence in uncrowded scene.

Fig. 4. Samples of abnormal behaviors in crowded scene.

are context space model, data stream clustering algorithm and inference algorithm. Tran, Le, and Tran (2014) proposed a new system to monitor lonely people (aged person, patient living alone, etc) by exploiting both image and audio information in video to detect abnormal events.

Another interesting topic in uncrowded scene is loitering. Loitering is defined as the act of being for a long period in a particular public space without any goal such as a person having a bag in an airport and staying for a much time without any purpose. This act is abnormal and several works proposed different techniques to detect the occurrence of this unusual event. For example, Gomez et al. (2015) proposed a two-stages loitering detection system based on sequential micro patterns. Those micro patterns are repeated actions performed by an individual characterizing the loitering behavior and they are obtained with the Generalized Sequential patterns (GSP) algorithm. Also, a method to detect loitering in video surveillance using direction history of the moving object trajectory and Inverse Perspective Mapping (IPM) method is proposed by Ko and Yoo (2013).

Furthermore, many people are dying every day because they were in the wrong place such as a pedestrian in the road or a person crossing the railway of a train. An effort has been made in this topic to prevent such actions and capture the presence of a human being in a wrong place. For instance, Delgado, Tahboub, and Delp (2014) proposed a method to detect, automatically, abnormal events on train platforms such as a person falling to the track bed region or pushed by another one. First, the moving objects are detected and located relatively to the edge of the track bed. Then, the motion information is used to detect the presence of trains.

On the other hand, when the scene contains a few number of persons, it is more interesting to detect violent actions such as fighting, kicking and punching between two persons (Fig. 3).

Several previous works were focused in detecting aggressive behaviors in video surveillance. For example, Déniz, Serrano, Bueno, and Kim (2014) proposed a method for detecting violence in videos (fighting in prisons, psychiatric centers, etc) based on the presence of high acceleration between two consecutive frames. In fact, they noticed that a large acceleration implies the existence of blur in the frame that can be modeled by an ellipse. Consequently, their system aims to detect such ellipse which means that an aggressive behavior has occurred. Moreover, Ouivirach et al. (2013) proposed an approach for suspicious event detection (e.g. fight) based on behavior models which are updated with each new observed sequence using Hidden Markov model (HMM). Contrarily, fights are automatically detected based on Convolutional Neural Networks model by Chunhui, Shouke, Ming, Weiguo, and Baozhi (2014). Guang et al. (2014) proposed a method to detect violence and fighting in the elevator based on motion information and SVM classifier. First, they combined corner kinetic energy based optical flow features and movement characteristics to describe the violent behavior. Then, they used the SVM classifier to detect violence in the elevator. Yujie and Zengfu (2016) proposed a real time violent behaviors, in the elevator, detection methods. First, they extracted and counted the number of humans presented in the elevator. Then, they used the MHI entropy image to detect violence.

3.2.2. Crowded scene

In this type of scene, including a group of persons, it is not possible to track and analyze the behavior of each person individually. This is due to the occlusion and the small number of pixels representing every person in the frame. So, it is better to model the interaction between people in order to detect an abnormal crowd behavior. Several previous works proposed abnormal behavior detection methods in crowded scene based on the interaction between people. Fig. 4 shows some abnormal crowd behaviors.

Table 7
Summary of performance evaluation results; (A) accuracy, (ERR) equal error rate, (AUC) area under ROC curve.

Paper	*Used datasets*				*Rate (%)*			*Comments*
	UMN	*UCSD*	*CV*	*HF*	*A*	*EER*	*AUC*	
(Mabrouk & Zagrouba, 2017)			√		85.83		89.25	Better results than (Gao et al., 2016; Hassner, Itcher, & Kliper-Gross, 2012b)
				√	88.60		93.23	
Alvar et al. (2014)		√				26.00		Peds 1
Yang et al. (2013)		√				23.00/ 24.80	47.10/ 86.80	Peds1/ Peds2
Zhang et al. (2016)	√						99.30/ 96.90/ 98.80	Three scenes. Better results than Ramin et al. (2009).
Xu, Gong, et al. (2014)			√		89.05			-
				√	94.30			
Bermejo et al. (2011)				√	91.70			Three methods are proposed, only the best result is kept.
Li et al. (2015)	√						93.00	-
		√				21.00/ 20.00		Peds 1/ Peds 2
Tao et al. (2015)			√				86.59	–
					82.79			
Nannan et al. (2015)		√					85.70/ 89.20	Peds 1/ Peds 2
						21.00/ 18.00		
	√						90.70	Average over three scenes.
Chen et al. (2015)	√						97.00	-
Xiao et al. (2015)		√				10.00/ 10.00		Peds 1/ Peds 2
Chunhui et al. (2014)				√	91.00			-
Hassner et al. (2012a)			√		81.30			-
							85	
				√	82.90			
Leyva et al. (2014)		√				19.00/ 16.00		Peds 1/ Peds 2
Gu et al. (2014)	√						97.90/ 98.66/ 99.11	Scene 1/ Scene 2/ Scene 3
Cong et al. (2013)	√						99.55/ 97.10/ 97.40	Scene 1/ Scene 2/ Scene 3
		√				20.00		Peds 1
							48.70	
Hajananth et al. (2014)		√				14.90/ 4.89		Peds 1/ Peds 2
							90.80/ 97.90	
Wang and Snoussi (2014)	√						98.45/ 90.37/ 98.15	Scene 1/ Scene 2/ Scene 3
Huang and Chen (2014)			√		86.37			-
				√	86.90			
Huo et al. (2014)	√						99.00/ 98.00/ 99.00	Scene 1/ Scene 2/ Scene 3
		√				27.00/ 8.00		Peds 1/ Peds 2
Weiya et al. (2015)	√						98.00	-
		√				19.00		
							50.30	Peds 1
Kai-Wen et al. (2015b)		√				23.70		Peds 1
							83.80	

Group interaction includes behaviors performed by multiple people such as disorder caused by group panic and violence in football stadium. Many works were focused on detecting unusual events happening in crowded scenes. For example, Ramin et al. (2009) proposed a method to detect crowd abnormal behavior using social force model which estimates the interaction force between individuals. First, they computed the social force model and then they used the BOW approach to classify events as normal and abnormal. Cho and Kang (2012) proposed a method for abnormality detection using multiple social behavior models which are determined based on optical flow and particle advection. Cho and Kang (2014) used static and dynamic agents to characterize the group interaction. Static agents aim to observe the individual behaviors by calculating optical flow variation. Dynamic agents compute group interaction using social force model. Santhiya, Sankaragomathi, Selvarani, and Kumar (2014) are based on motion detection and analysis to describe the anomaly. A social group detection method in crowded scene based on two features which are gaz direction and visual attention is proposed by Leach, Baxter, Robertson, and Sparks (2014a). Those two features are used to specify the intention of the person in the video. Hassner et al. (2012a) detected the crowd violence in real time based on a new Violent Flows descriptor. Using this method and fluid Mechanics, Wang, Gao, He, Wu, and Li (2014) proposed an improved algorithm to detect abnormal behaviors in crowded scene. Chaker, Aghbari, and Junejo (2017) introduced an unsupervised framework based on social network model to capture the crowd interaction and the scene dynamics. The crowd behavior is detected using the adjacent flow position estimation by Wang, Fu, and Liu (2016a).

4. Performance evaluation

The performance evaluation step is essential not only to measure the efficiency of a proposed system, but also to compare it to other ones. There are multiple evaluation projects regarding video surveillance systems. For instance, TRECVID (Awad et al., 2016) is an international campaign in the field of the video retrieval evaluation. It provides the surveillance event detection (SED) task which aims to evaluate the performance of a surveillance system in terms of real time events detection. In fact, the evaluation stage must report how many abnormal behaviors were recognized and

Table 8
Existing video surveillance systems.

Project	*Main goal*	*Environment*	*Architecture*	*Technologies*
CROMATICA	Passengers' surveillance.	Public transport.	Distributed surveillance system.	Video analysis + wireless data transfer.
PRISMATICA	Passengers' surveillance.	Urban public transport.	Multi-sensor distributed system + centralized approach.	Camera network + wireless transmission system + audio surveillance system.
VSAM	Continuous monitoring (24h).	Urban environment.	Distributed networks of video sensors.	Moving Object geo-location + tracking over multiple camera views + gait analysis.
ADVISOR	Human operators assistance (automatic selection, image annotation, etc).	Transport infrastructures (e.g. subways).	Modular and scalable architecture.	Video data recording + storing annotations for interest images.
VIGILANT	Pedestrians monitoring.	Parking lot.	Distributed multi-camera system.	Software agents for people tracking across multi-camera.
AVITRACK	Aircraft activities and movements monitoring.	Aircraft parking zone.	Multi sensors (scene tracking) + data fusion (scene understanding).	Video tracking + 3D representation map + artificial intelligence (scene interpretation).
HESPERIA	Object classification, unusual movement detection.	Public infrastructures (water tanks, telecommunications stations, etc) and public spaces (airport, train stations, etc).	Middleware communications solution interconnecting three parts : video + audio + sensor network.	Crisis management system (CMS) + cognitive video and audio + augmented reality system.

how many false alarms were produced using public video datasets. In the next subsections, we present first the widely used datasets for abnormal behavior recognition. Then, we define the popular metrics used to evaluate the performance of a video surveillance system.

4.1. Datasets

Th high number of proposed methods for abnormal behavior recognition shows that it is widely studied area. There is a grown demand for public datasets to use for video surveillance system evaluation. In fact, those datasets can be categorized into crowded and uncrowded scenes. The first type consists of videos containing often violent actions such as punching and kicking. The second type of datasets contains videos describing the interaction between a group of persons acting abnormally such as panic escape. In Table 6, the popular datasets that are used in the works cited in this paper are listed. For other public datasets, the reader can refer to these websites.[1,2,3,4,5,6,7]

4.2. Evaluation metrics

Several metrics are provided to evaluate a video surveillance system. The two commonly used criteria are the Equal Error Rate (EER) and the Area Under Roc curve (AUC). Those two criteria are derived from the Receiver Operating Characteristic (ROC) Curve which is highly used for performance comparison. EER is the point on the ROC curve where the false positive rate (normal behavior is considered as abnormal) is equal to the false negative rate (abnormal behavior is identified as normal). For a good recognition algorithm, the EER should be as small as possible. Conversely, a system is considered having good performances if the value of the AUC is high. An other used recognition measure is the accuracy (A) which corresponds to the fraction of the correctly classified behaviors. In Table 7, we summarize the performance evaluation of previous interesting papers whose results are available.

1 https://www.kaggle.com/datasets.
2 http://www.svcl.ucsd.edu/projects/anomaly/UCSD_Anomaly_Dataset.tar.gz.
3 http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi.
4 http://www.openu.ac.il/home/hassner/data/violentflows.
5 http://visilab.etsii.uclm.es/personas/oscar/FightDetection/Peliculas.rar.
6 http://visilab.etsii.uclm.es/personas/oscar/FightDetection/HockeyFights.zip.
7 http://crcv.ucf.edu/projects/Abnormal_Crowd/Normal_Abnormal_Crowd.zip.

5. Existing video surveillance systems

To ensure human safety, there is an immediate need for intelligent video surveillance systems that control private and public places and detect dangerous situations. For each system, there is a specific architecture depending on the environment (indoor/ outdoor). For example, Vallejo, Albusac, Castro-Schez, Glez-Morcillo, and Jiménez (2011) proposed a multi-agent architecture in order to deploy a surveillance system for urban traffic monitoring. A Multiscale architecture of a tracking system is presented by Hampapur et al. (2005). The proposed system aims to detect and track moving objects in order to identify their nature (animal, car, people, etc). In fact, depending on the system architecture, several research projects have been developed to improve the surveillance task and to assist human operators. Some of them are funded bythe European Union (EU). For example, CROMATICA (CROwd MAnagement with Telematic Imaging and Communication Assistance) (Tomasi & Kanade, 1991) and PRISMATICA (PRo-active Integrated Systems for security MAnagement by Technological Institutional and Communication Assistance) (Murray & Basu, 1994) are EU funded projects for passengers surveillance in public transport. AVITRACK is designed to monitor aircraft in an airport. The Video Surveillance And Monitoring (VSAM) project (Collins et al., 2000) is funded by the U.S. government for moving object detection, localization and classification. There is other existing surveillance systems used in real world such as ADVISOR (Annotated Digital Video for Surveillance and Optimised Retrieval) (Siebel & Maybank, 2004), VIGILANT and HESPERIA (Garcia et al., 2007). In Table 8, we present and compare the aforementioned systems.

6. Conclusions and discussions

In this review, we studied the different levels of a video surveillance system which are behavior representation and behavior modeling. First, we surveyed the most popular methods used for features extraction and description. Then, we provided a comprehensive overview of different classification methods and frameworks for behavior modeling. Moreover, we presented the most challenging datasets and evaluation metrics used for video surveillance systems evaluation. Finally, we exhibited some existing intelligent video systems in real world.

Despite the significant progress in the field of abnormal behavior recognition, there are some limitations that make it more difficult and challenging. In fact, the choice of features that are used to characterize the moving object is a difficult task because it influences significantly the description and the analysis of the behavior. For example, describing the behavior when the background of the scene changes frequently or when new objects appear suddenly in the scene is a difficult task. Besides, the appearance of the moving target may change due to multiple factors, such as clothing (short dress, coat, boot, sandal, etc) and scene place (outdoor/indoor, lift/stairs, etc). Therefore, choosing features that are robust to scene transformations (rotation, occlusion, cluttered backgrounds, etc) and less sensitive to the changes in the object appearance is essential in order to capture relevant and discriminative information about the moving object behavior. Furthermore, most abnormal behavior recognition algorithms suppose that the moving object is on the front of the camera. However, in reality the viewpoint is arbitrary. To overcome this limitation, there are works that use multiple cameras to capture different views for the moving object and, then, combine them. Despite the fact that those algorithms are effective and give good results, they are very sophisticated, time consuming and not suitable for real time applications. On the other hand, an observed behavior may have several interpretations depending on the context in which it is performed, the time and the place of the action. For example, running is a daily activity most people perform, but when a person is running on the road, it is considered as an abnormal behavior that has to be triggered. To overcome the aforementioned limitations, proposed systems used huge amount of training data including all possible scenarios. To deal with the important size of data, it is become a trend to use cloud computing which allows advanced algorithms, e.g. deep learning to work efficiently on larger datasets. In fact, due to their deep architectures, the use of deep learning algorithms has rapidly grown in order to obtain much larger capacity of learning.

References

Acampora, G., Foggia, P., Saggese, A., & Vento, M. (2015). A hierarchical neuro-fuzzy architecture for human behavior analysis. *Information Sciences, 310*, 130–148. doi:10.1016/j.ins.2015.03.021.

Acharya, B. R., & Gantayat, P. K. (2015). Recognition of human unusual activity in surveillance videos. *International Journal of Research and Scientific Innovation (IJRSI), 2*, 18–23.

Aggarwal, J., & Cai, Q. (1999). Human motion analysis. *Computer Vision and Image Understanding, 73*(3), 428–440. doi:10.1006/cviu.1998.0744.

Aggarwal, J. K., & Cai, Q. (1997). Human motion analysis: a review. In *Proceedings IEEE nonrigid and articulated motion workshop* (pp. 90–102). doi:10.1109/NAMW.1997.609859.

Albusac, J., Vallejo, D., Castro-Schez, J., Glez-Morcillo, C., & Jiménez, L. (2014). Dynamic weighted aggregation for normality analysis in intelligent surveillance systems. *Expert Systems with Applications, 41*(4), 2008–2022. http://dx.doi.org/10.1016/j.eswa.2013.08.097.

Alvar, M., Torsello, A., Miralles, A. S., & Armingol, J. M. (2014). Abnormal behavior detection using dominant sets. *Machine Vision and Applications, 25*(5), 1351–1368. doi:10.1007/s00138-014-0615-4.

Arroyo, R., Yebes, J. J., Bergasa, L. M., Daza, I. G., & Almazan, J. (2015). Expert video–surveillance system for real-time detection of suspicious behaviors in shopping malls. *Expert Systems with Applications, 42*, 7991–8005.

Aslan, M., Sengur, A., Xiao, Y., Wang, H., Ince, M. C., & Ma, X. (2015). Shape feature encoding via fisher vector for efficient fall detection in depth-videos. *Applied Soft Computing*. doi:10.1016/j.asoc.2014.12.035.

Awad, G., Fiscus, J., Michel, M., Joy, D., Kraaij, W., Smeaton, A. F., et al. (2016). Trecvid 2016: Evaluating video search, video event detection, localization, and hyperlinking. In *Proceedings of TRECVID 2016*. NIST, USA.

Bakar, U. A. B. U. A., Hemant, G., Hasanm, S. F., & Mukhopadhyay, S. C. (2016). Activity and anomaly detection in smart home: A survey. In *Next Generation Sensors and Systems* (pp. 191–220). Springer International Publishing. doi:10.1007/978-3-319-21671-3_9.

Bellamine, I., & Tairi, H. (2015). Motion detection using color structure-texture image decomposition. In *Intelligent systems and computer vision (ISCV)* (pp. 1–8). doi:10.1109/ISACV.2015.7105545.

Bellamine, I., & Tairi, H. (2016). Motion detection using color space-time interest points. *Proceedings of the mediterranean conference on information & communication technologies 2015: MedCT 2015 Vol. 1* (pp. 97–105)). Springer International Publishing. doi:10.1007/978-3-319-30301-7_11.

Bermejo, N. E., Deniz, S. O., Bueno, G. G., & Rahul, S. (2011). Violence detection in video using computer vision techniques. In *Proceedings of the 14th international conference on computer analysis of images and patterns – Volume part ii* (pp. 332–339). Springer-Verlag. doi:10.1007/978-3-642-23678-5_39.

Bian, Z.-P., Hou, J., Chau, L.-P., & N. , M.-T. (2015). Fall detection based on body part tracking using a depth camera. *IEEE Journal of Biomedical and Health Informatics, 19*, 430–439. doi:10.1109/JBHI.2014.2319372.

Bux, A., Plamen, A., & Zulfiqar, H. (2017). Vision based human activity recognition: A review. In P. Angelov, A. Gegov, C. Jayne, & Q. Shen (Eds.), *Advances in computational intelligence systems: contributions presented at the 16th UK workshop on computational intelligence, September 7–9, 2016, Lancaster, UK* (pp. 341–371)). Springer International Publishing. doi:10.1007/978-3-319-46562-3_23.

de Campos, T. (2014). A survey on computer vision tools for action recognition, crowd surveillance and suspect retrieval. In *XXXIV congresso da sociedade brasileira de computacao (CSBC)* (pp. 1123–1132).

Castro, J., Delgado, M., Medina, J., & Ruiz-Lozano, M. (2011). Intelligent surveillance system with integration of heterogeneous information for intrusion detection. *Expert Systems with Applications, 38*(9), 11182–11192. http://dx.doi.org/10.1016/j.eswa.2011.02.165.

Ce, L., Zhenjun, H., Qixiang, Y., & Jianbin, J. (2013). Visual abnormal behavior detection based on trajectory sparse reconstruction analysis. *Neurocomputing, 119*, 94–100. doi:10.1016/j.neucom.2012.03.040.

Chaker, R., Aghbari, Z. A., & Junejo, I. N. (2017). Social network model for crowd anomaly detection and localization. *Pattern Recognition, 61*, 266–281. doi:10.1016/j.patcog.2016.06.016.

Chen, D.-Y., & Huang, P.-C. (2011). Motion-based unusual event detection in human crowds. *Journal of Visual Communication and Image Representation, 22*(2), 178–186. doi:10.1016/j.jvcir.2010.12.004.

Chen, Z., Tian, Y., Zeng, W., & Huang, T. (2015). Detecting abnormal behaviors in surveillance videos based on fuzzy clustering and multiple auto-encoders. In *Multimedia and expo (ICME), 2015 IEEE international conference on* (pp. 1–6). doi:10.1109/ICME.2015.7177459.

Cho, S.-H., & Kang, H.-B. (2012). Integrated multiple behavior models for abnormal crowd behavior detection. In *Image analysis and interpretation (SSIAI), 2012 IEEE southwest symposium on* (pp. 113–116). doi:10.1109/SSIAI.2012.6202466.

Cho, S.-H., & Kang, H.-B. (2014). Abnormal behavior detection using hybrid agents in crowded scenes. *Pattern Recognition Letters, 44*, 64–70. doi:10.1016/j.patrec.2013.11.017.

Chundi, M., Jianbin, X., Wei, Y., Tong, L., & Peiqin, L. (2015). A fast recognition algorithm for suspicious behavior in high definition videos. *Multimedia Systems*, 1–11. doi:10.1007/s00530-015-0456-7.

Chunhui, D., Shouke, F., Ming, Z., Weiguo, F., & Baozhi, J. (2014). Violence detection in video by using 3d convolutional neural networks. In *Advances in visual computing: 8888* (pp. 551–558). Springer International Publishing. doi:10.1007/978-3-319-14364-4_53.

Collins, R., Lipton, A., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., et al. (2000). A system for video surveillance and monitoring. *Technical Report, CMU-RI-TR-00-12, Pittsburgh, PA*.

Conci, N., & Lizzi, L. (2009). Camera placement using particle swarm optimization in visual surveillance applications. In *2009 16th IEEE international conference on image processing (ICIP)* (pp. 3485–3488). doi:10.1109/ICIP.2009.5413833.

Cong, Y., Yuan, J., & Liu, J. (2013). Abnormal event detection in crowded scenes using sparse representation. *Pattern Recognition, 46*(7), 1851–1864. doi:10.1016/j.patcog.2012.11.021.

Dawn, D., Debapratim, Shaikh, & SoharabHossain (2015). A comprehensive survey of human action recognition with spatio-temporal interest point (stip) detector. *The Visual Computer*, 1–18. doi:10.1007/s00371-015-1066-2.

Delgado, B., Tahboub, K., & Delp, E. (2014). Automatic detection of abnormal human events on train platforms. In *Aerospace and electronics conference, NAECON 2014 - IEEE national* (pp. 169–173).

Déniz, O., Serrano, I., Bueno, G., & Kim, T. (2014). Fast violence detection in video. In *VISAPP 2014 - Proceedings of the 9th international conference on computer vision theory and applications, Vol. 2, Lisbon, Portugal, 5–8 January, 2014* (pp. 478–485). doi:10.5220/0004695104780485.

Dhulekar, P., T. , G. S., Chitte, H., & Pardeshi, K. (2017). Human action recognition: An overview. In S. C. Satapathy, V. Bhateja, & A. Joshi (Eds.), *Proceedings of the international conference on data engineering and communication technology: ICDECT 2016, Volume 1* (pp. 481–488)). Springer Singapore. doi:10.1007/978-981-10-1675-2_48.

Feng, Y., Yuan, Y., & Lu, X. (2017). Learning deep event models for crowd anomaly detection. *Neurocomputing, 219*, 548–556. doi:10.1016/j.neucom.2016.09.063.

Foggia, P., Percannella, G., Saggese, A., & Vento, M. (2013). Recognizing human actions by a bag of visual words. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on* (pp. 2910–2915). doi:10.1109/SMC.2013.496.

Gao, Y., Liu, H., Sun, X., Wang, C., & Liu, Y. (2016). Violence detection using oriented violent flows. *Image and Vision Computing, 48 -49*, 37–41.

Garcia, C., Hernandez, Y., Fernandez, D., Ferrer, M. A., Travieso, C. M., Alonso, J. B., et al. (2007). Hesperia: Homeland security technologies for the security in public spaces and infrastructures. In *2007 41st annual IEEE international carnahan conference on security technology* (pp. 221–226). doi:10.1109/CCST.2007.4373493.

Gnanavel, V., & Srinivasan, A. (2015). Abnormal event detection in crowded video scenes. In *Proceedings of the 3rd international conference on frontiers of intelligent computing: Theory and applications (FICTA) 2014: 328* (pp. 441–448). Springer International Publishing. doi:10.1007/978-3-319-12012-6_48.

Gomez, H. F., Martinez, T. R., Arias, T. S., Fernandez, C. A., Sylvie, R., Gonzalez, E. A., et al. (2015). Identification of loitering human behaviour in video surveillance environments. In *Artificial computation in biology and medicine: 9107* (pp. 516–525). Springer International Publishing. doi:10.1007/978-3-319-18914-7_54.

Gu, X., Cui, J., & Zhu, Q. (2014). Abnormal crowd behavior detection by using the particle entropy. *Optik – International Journal for Light and Electron Optics, 125*(14), 3428–3433.

Guang, S., Fu, G., Li, P., & Geng, H. (2014). Violent behavior detection based on svm in the elevator. *International Journal of Security and Its Applications, 8*, 31–40. doi:10.14257/ijsia.2014.8.5.04.

Gunduz, A., Temizel, T., & Temizel, A. (2014). Pedestrian zone anomaly detection by non-parametric temporal modelling. In *Advanced video and signal based surveillance (AVSS), 2014 11th ieee international conference on* (pp. 131–135).

Hajananth, N., Fookes, C., Denman, S., & Sridharan, S. (2014). An mrf based abnormal event detection approach using motion and appearance features. In *Advanced video and signal based surveillance (AVSS), 2014 11th ieee international conference on* (pp. 343–348). doi:10.1109/AVSS.2014.6918692.

Hampapur, A., Brown, L., Connell, J., Ekin, A., Haas, N., Lu, M., et al. (2005). Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking. *IEEE Signal Processing Magazine, 22*(2), 38–51.

Hassan, M., Ahmad, T., Liaqat, N., Farooq, A., Ali, S. A., & Hassan, S. R. (2014). A review on human actions recognition using vision based techniques. *Journal of Image and Graphics, 2*(1), 28–32. doi:10.12720/joig.2.1.28-32.

Hassner, T., Itcher, Y., & Kliper-Gross, O. (2012a). Violent flows: Real-time detection of violent crowd behavior. In *Computer vision and pattern recognition workshops (CVPRW), 2012 ieee computer society conference on* (pp. 1–6). doi:10.1109/CVPRW.2012.6239348.

Hassner, T., Itcher, Y., & Kliper-Gross, O. (2012b). Violent flows: Real-time detection of violent crowd behavior. In *CVPR workshops* (pp. 1–6). IEEE. doi:10.1109/CVPRW.2012.6239348.

Himanshu, R., Maheshkumar, H. K., Neelabh, K., & Mukherjee, J. (2015). Trajectory based unusual human movement identi_cation for video surveillance system. In *Progress in Systems Engineering: 330* (pp. 789–794). Springer International Publishing. doi:10.1007/978-3-319-08422-0_114.

Huang, B., Tian, G., Wu, H., & Zhou, F. (2014). A method of abnormal habits recognition in intelligent space. *Engineering Applications of Artificial Intelligence, 29*, 125–133. doi:10.1016/j.engappai.2013.12.010.

Huang, J.-F., & Chen, S.-L. (2014). Detection of violent crowd behavior based on statistical characteristics of the optical flow. In *Fuzzy systems and knowledge discovery (FSKD), 2014 11th international conference on* (pp. 565–569). doi:10.1109/FSKD.2014.6980896.

Huang, J.-Y., & Lee, W.-P. (2014). A smart camera network with svm classifiers for crowd event recognition. *Lecture Notes in Engineering and Computer Science, 1*, 13–18.

Huo, J., Gao, Y., Yang, W., & Yin, H. (2014). Multi-instance dictionary learning for detecting abnormal events in surveillance videos. *International Journal of Neural Systems, 24*. doi:10.1142/S0129065714300101.

Jaechul, K., & Kristen, G. (2009). Observe locally, infer globally: A space-time mrf for detecting abnormal activities with incremental updates.. In *Cvpr* (pp. 2921–2928). IEEE Computer Society. doi:10.1109/CVPRW.2009.5206569.

Ji, X., & Liu, H. (2010). Advances in view-invariant human motion analysis: A review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40*(1), 13–24. doi:10.1109/TSMCC.2009.2027608.

Kai-Wen, C., Yie-Tarng, C., & Wen-Hsien, F. (2015a). An efficient subsequence search for video anomaly detection and localization. *Multimedia Tools and Applications*, 1–22. doi:10.1007/s11042-015-2453-4.

Kai-Wen, C., Yie-Tarng, C., & Wen-Hsien, F. (2015b). Video anomaly detection and localization using hierarchical feature representation and gaussian process regression. In *Ieee conference on computer vision and pattern recognition (CVPR)* (pp. 2909–2917).

Kim, H., Lee, S., Kim, Y., Lee, S., Lee, D., Ju, J., et al. (2016). Weighted joint-based human behavior recognition algorithm using only depth information for low-cost intelligent video-surveillance system. *Expert Systems with Applications, 45*, 131–141. http://dx.doi.org/10.1016/j.eswa.2015.09.035.

Ko, J.-G., & Yoo, J.-H. (2013). Rectified trajectory analysis based abnormal loitering detection for video surveillance. In *Artificial intelligence, modelling and simulation (AIMS), 2013 1st international conference on* (pp. 289–293). doi:10.1109/AIMS.2013.53.

Lai, D.-X., Chang, Y.-H., & Zhong, Z.-H. (2009). Active contour tracking of moving objects using edge flows and ant colony optimization in video sequences. In T. Wada, F. Huang, & S. Lin (Eds.), *Advances in image and video technology: Third pacific rim symposium, PSIVT 2009, Tokyo, Japan, January 13–16, 2009. Proceedings* (pp. 1104–1116)). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-92957-4_96.

Laptev, I. (2005). On space-time interest points. *International Journal of Computer Vision, 64*, 107–123. doi:10.1007/s11263-005-1838-7.

Leach, M., Baxter, R., Robertson, N., & Sparks, E. (2014a). Detecting social groups in crowded surveillance videos using visual attention. In *Computer vision and pattern recognition workshops (CVPRW), 2014 IEEE conference on* (pp. 467–473). doi:10.1109/CVPRW.2014.75.

Leach, M. J., Sparks, E., & Robertson, N. M. (2014b). Contextual anomaly detection in crowded surveillance scenes. *Pattern Recognition Letters, 44*, 71–79. doi:10.1016/j.patrec.2013.11.018.

Leyva, R., Sanchez, V., & Chang-Tsun, L. (2014). Video anomaly detection based on wake motion descriptors and perspective grids. In *Information forensics and security (WIFS), 2014 IEEE international workshop on* (pp. 209–214). doi:10.1109/WIFS.2014.7084329.

Li, N., Wu, X., Xu, D., Guo, H., & Feng, W. (2015). Spatio-temporal context analysis within video volumes for anomalous-event detection and localization. *Neurocomputing, 155*, 309–319. doi:10.1016/j.neucom.2014.12.064.

Li, S., Gong, D., & Yuan, Y. (2013). Face recognition using weber local descriptors. *Neurocomputing, 122*, 272–283. doi:10.1016/j.neucom.2013.05.038.

Li, W., Mahadevan, V., & Vasconcelos, N. (2014). Anomaly detection and localization in crowded scenes. In *IEEE transactions on pattern analysis and machine intelligence: 36* (pp. 18–32). doi:10.1109/TPAMI.2013.111.

Liu, Y., Li, Y., & Ji, X. (2014). Abnormal event detection in nature settings. *International Journal of Signal Processing, Image Processing and Pattern Recognition, 7*(4), 115–126. ijsip.2014.7.4.11.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91–110. doi:10.1023/B:VISI.0000029664.99615.94.

Lu, C., Shi, J., & Jia, J. (2013). Abnormal event detection at 150 fps in matlab. In *2013 IEEE international conference on computer vision* (pp. 2720–2727).

Mabrouk, A. B., & Zagrouba, E. (2017). Spatio-temporal feature using optical flow based distribution for violence detection. *Pattern Recognition Letters, 92*, 62–67. doi:10.1016/j.patrec.2017.04.015.

Mahadevan, V., Li, W., Bhalodia, V., & Vasconcelos, N. (2010). Anomaly detection in crowded scenes. In *Proceedings of IEEE conference on computer vision and pattern recognition* (pp. 1975–1981). doi:10.1109/CVPR.2010.5539872.

Miao, Y., & Song, J. (2014). Abnormal event detection based on svm in video surveillance. In *Advanced research and technology in industry applications (wartia), 2014 IEEE workshop on* (pp. 1379–1383). doi:10.1109/WARTIA.2014.6976540.

Ming-yu, C., Lily, M., Padmanabhan, P., Alexander, H., & Rahul, S. (2010). Exploiting multi-level parallelism for low-latency activity recognition in streaming video. In *Proceedings of the first annual acm sigmm conference on multimedia systems* (pp. 1–12). ACM. doi:10.1145/1730836.1730838.

Mishra, S. K., & Bhagat, K. S. (2015). A survey on human motion detection and surveillance. *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), 4*, 1044–1048.

Murray, D., & Basu, A. (1994). Motion tracking with an active camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 16*(5), 449–459. doi:10.1109/34.291452.

Nannan, L., Xinyu, W., Huiwen, G., Dan, X., Yongsheng, O., & Yen-Lun, C. (2015). Anomaly detection in video surveillance via gaussian process. *International Journal of Pattern Recognition and Artificial Intelligence, 29*(06), 1555011. doi:10.1142/S0218001415550113.

Ngo, D. V., Do, N. T., & Nguyen, L. A. T. (2016). Anomaly detection in video surveillance: A novel approach based on sub-trajectory. In *2016 international conference on electronics, information, and communications (ICEIC)* (pp. 1–4). doi:10.1109/ELINFOCOM.2016.7562953.

Nguyen, V. D., Le, M. T., Do, A. D., Duong, H. H., Thai, T. D., & Tran, D. H. (2014). An efficient camera-based surveillance for fall detection of elderly people. In *Industrial electronics and applications (ICIEA), 2014 IEEE 9th conference on* (pp. 994–997). doi:10.1109/ICIEA.2014.6931308.

Oluwatoyin, P. P., & Kejun, W. (2012). Video-based abnormal human behavior recognition - a review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C, 42*(6), 865–878. doi:10.1109/TSMCC.2011.2178594.

Ouivirach, K., Gharti, S., & Dailey, M. N. (2013). Incremental behavior modeling and suspicious activity detection. *Pattern Recognition, 46*(3), 671–680. doi:10.1016/j.patcog.2012.10.008.

Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: A survey. In *Proceedings of the ICMI 2006 and IJCAI 2007 international conference on artifical intelligence for human computing*. In *ICMI'06/IJCAI'07* (pp. 47–71). Berlin, Heidelberg: Springer-Verlag.

Pathak, D., Sharang, A., & Mukerjee, A. (2015). Anomaly localization in topic-based analysis of surveillance videos. In *2015 IEEE Winter Conference on Applications of Computer Vision, WACV 2014, Waikoloa, HI, USA, January 5–9, 2015* (pp. 389–395). doi:10.1109/WACV.2015.58.

Rajkumar, S., Arif, A., Prosad, D. D., & Pratim, R. P. (2017). Surveillance scene segmentation based on trajectory classification using supervised learning. *Proceedings of International Conference on Computer Vision and Image Processing: CVIP 2016, Volume 1* (pp. 261–271)). Springer Singapore. doi:10.1007/978-981-10-2104-6_24.

Ramin, M., Alexis, O., & Mubarak, S. (2009). Abnormal crowd behavior detection using social force model. In *CVPR* (pp. 935–942). IEEE Computer Society. doi:10.1109/CVPRW.2009.5206641.

Rao, A. S., Gubbi, J., Rajasegarar, S., Marusic, S., & Palaniswami, M. (2014). Detection of anomalous crowd behaviour using hyperspherical clustering. In Digital Image Computing: Techniques and Applications (DICTA), 2014 International Conference on (pp. 1–8). doi:10.1109/DICTA.2014.7008100.

Rasheed, S. A. , K., & A. , K. (2014). Tracking and abnormal behavior detection in video surveillance using optical flow and neural networks. In *Advanced information networking and applications workshops (WAINA), 2014 28th international conference on* (pp. 61–66).

Rezaee, K., Haddadnia, J., & Delbari, A. (2015). Modeling abnormal walking of the elderly to predict risk of the falls using kalman filter and motion estimation approach. *Computers & Electrical Engineering*. doi:10.1016/j.compeleceng.2015.03.005.

Sabokrou, M., Fayyaz, M., Fathy, M., & Klette, R. (2016). Fully convolutional neural network for fast anomaly detection in crowded scenes. *CoRR, abs/1609.00866*.

Santhiya, G., Sankaragomathi, K., Selvarani, S., & Kumar, A. (2014). Abnormal crowd tracking and motion analysis. In *Advanced communication control and computing technologies (ICACCCT), 2014 international conference on* (pp. 1300–1304).

Sarvesh, V., & Anupam, A. (2013). A survey on activity recognition and behavior understanding in video surveillance. *The Visual Computer, 29*(10), 983–1009. doi:10.1007/s00371-012-0752-6.

Shian-Ru, K., Uyen, T. H. L., Yong-Jin, L., Jenq-Neng, H., Jang-Hee, Y., & Kyoung-Ho, C. (2013). A review on video-based human activity recognition. *Computers, 2*(2), 88–131. doi:10.3390/computers2020088.

Shifu, Z., Wei, S., Dan, Z., Mei, F., Yuanwang, W., & Zhijiang, Z. (2016). Spatial-temporal convolutional neural networks for anomaly detection and localization in crowded scenes. *Image Communication, 47*, 358–368. doi:10.1016/j.image.2016.06.007.

Siddharth, R., & Anupam, A. (2015). Vision based hand gesture recognition for human computer interaction: A survey. *Artificial Intelligence Review, 43*(1), 1–54. doi:10.1007/s10462-012-9356-9.

Siebel, N. T., & Maybank, S. J. (2004). The advisor visual surveillance system. In *in ECCV 2004 workshop applications of computer vision (ACV*.

Songhao, Z., Juanjuan, H., & Zhe, S. (2016). Local abnormal behavior detection based on optical flow and spatio-temporal gradient. *Multimedia Tools and Applications, 75*(15), 9445–9459. doi:10.1007/s11042-015-3122-3.

Stone, E. E., & Skubic, M. (2015). Fall detection in homes of older adults using the microsoft kinect. *IEEE Journal of Biomedical and Health Informatics, 19*, 290–301. doi:10.1109/JBHI.2014.2312180.

Su, J., sheng Yin, G., Hailong, C., & Zhiyong, L. (2014). Multi-target tracking and behavior analysis method for video surveillance applications. *International Journal of Multimedia and Ubiquitous Engineering, 9*, 83–92. doi:10.14257/ijmue.2014.9.8.08.

SungChun, L., & Ram, N. (2014). Hierarchical abnormal event detection by real time and semi-real time multi-tasking video surveillance system. *Machine Vision and Applications, 25*, 133–143. doi:10.1007/s00138-013-0516-y.

Takai, M. (2015). Measurement of complex quantity of monitoring area and detection of high active part of invading object in complex background for surveillance camera system. In *Mechatronics (ICM), 2015 IEEE international conference on* (pp. 522–528).

Tani, M. K., Lablack, A., Ghomari, A., & Bilasco, I. M. (2015). Events detection using a video-surveillance ontology and a rule-based approach. In *Computer vision - ECCV 2014 workshops: 8926* (pp. 299–308). Springer International Publishing. doi:10.1007/978-3-319-16181-5_21.

Tao, Z., Wenjing, J., Baoqing, Y., Jie, Y., Xiangjian, H., & Zhonglong, Z. (2017). Mowld: a robust motion image descriptor for violence detection. *Multimedia Tools and Applications, 76*(1), 1419–1438. doi:10.1007/s11042-015-3133-0.

Tao, Z., Zhijie, Y., Wenjing, J., Baoqing, Y., Jie, Y., & Xiangjian, H. (2015). A new method for violence detection in surveillance scenes. *Multimedia Tools and Applications*, 1–23. doi:10.1007/s11042-015-2648-8.

Tax, D. M., & Duin, R. P. (2004). Support vector data description. *Machine Learning, 54*(1), 45–66. doi:10.1023/B:MACH.0000008084.60811.49.

Teddy, K. (2008). A survey on behavior analysis in video surveillance for homeland security applications. In *Aipr* (pp. 1–8). doi:10.1109/AIPR.2008.4906450.

Teng, L., Chang, H., Wang, M., Ni, B., Hong, R., & Yan, S. (2015). Crowded scene analysis: A survey. *circuits and systems for video technology, IEEE transactions on, 25*, 367–386. doi:10.1109/TCSVT.2014.2358029.

Thi-Lan, L., & Thanh-Hai, T. (2015). Real-time abnormal events detection combining motion templates and object localization. In *Some current advanced researches on information and computer science in Vietnam: 341* (pp. 17–30). Springer International Publishing. doi:10.1007/978-3-319-14633-1_2.

Tomasi, C., & Kanade, T. (1991). Detection and tracking of point features. *Technical Report*. International Journal of Computer Vision.

Tran, D.-D., Le, T.-L., & Tran, T.-T.-H. (2014). Abnormal event detection using multimedia information for monitoring system. In *Communications and electronics (ICCE), 2014 IEEE fifth international conference on* (pp. 490–495). IEEE. doi:10.1109/CCE.2014.6916753.

Valera, M., & Velastin, S. A. (2005). Intelligent distributed surveillance systems: A review. *IEE Proceedings - Vision, Image and Signal Processing, 152*(2), 192–204. doi:10.1049/ip-vis:20041147.

Vallejo, D., Albusac, J., Castro-Schez, J., Glez-Morcillo, C., & Jiménez, L. (2011). A multi-agent architecture for supporting distributed normality-based intelligent surveillance. *Engineering Applications of Artificial Intelligence, 24*(2), 325–340. http://dx.doi.org/10.1016/j.engappai.2010.11.005.

Wang, B., Li, W., Yang, W., & Liao, Q. (2011). Illumination normalization based on weber's law with application to face recognition. *IEEE Signal Processing Letters, 18*(8), 462–465.

Wang, G., Fu, H., & Liu, Y. (2016a). Real time abnormal crowd behavior detection based on adjacent flow location estimation. In *2016 4th international conference on cloud computing and intelligence systems (CCIS)* (pp. 476–479). doi:10.1109/CCIS.2016.7790305.

Wang, J., & Xu, Z. (2015). Crowd anomaly detection for automated video surveillance. In *The 6th international conference on imaging for crime detection and prevention ICDP-15*. doi:10.1049/ic.2015.0102.

Wang, L., & Dong, M. (2014). Detection of abnormal human behavior using a matrix approximation-based approach. In *Machine learning and applications (ICMLA), 2014 13th international conference on* (pp. 324–329). doi:10.1109/ICMLA.2014.58.

Wang, T., & Snoussi, H. (2014). Detection of abnormal visual events via global optical flow orientation histogram. *Information Forensics and Security, IEEE Transactions on, 9*, 988–998. doi:10.1109/TIFS.2014.2315971.

Wang, X., Gao, M., He, X., Wu, X., & Li, Y. (2014). An abnormal crowd behavior detection algorithm based on fluid mechanics. *Journal of Computers, 9*(5). doi:10.4304/jcp.9.5.1144-1149.

Wang, X., He, X., Wu, X., Xie, C., & Li, Y. (2016). A classification method based on streak flow for abnormal crowd behaviors. *Optik - International Journal for Light and Electron Optics, 127*(4), 2386–2392. doi:10.1016/j.ijleo.2015.08.081.

Weiya, R., Guohui, L., Boliang, S., & Kuihua, H. (2015). Unsupervised kernel learning for abnormal events detection. *The Visual Computer, 31*, 245–255. doi:10.1007/s00371-013-0915-0.

Wenbing, Z. (2016). A concise tutorial on human motion tracking and recognition with microsoft kinect. *Science China Information Sciences, 59*(9), 93101. doi:10.1007/s11432-016-5604-y.

Wiliem, A., Madasu, V., Boles, W., & Yarlagadda, P. (2012). A suspicious behaviour detection using a context space model for smart surveillance systems. *Computer Vision and Image Understanding, 116*(2), 194–209. doi:10.1016/j.cviu.2011.10.001.

Xiao, T., Zhang, C., & Zha, H. (2015). Learning to detect anomalies in surveillance video. *Signal Processing Letters, IEEE, 22*, 1477–1481. doi:10.1109/LSP.2015.2410031.

Xu, D., Song, R., Wu, X., Li, N., Feng, W., & Qian, H. (2014). Video anomaly detection based on a hierarchical activity discovery within spatio-temporal contexts. *Neurocomputing, 143*, 144–152. doi:10.1016/j.neucom.2014.06.011.

Xu, L., Gong, C., Yang, J., Wu, Q., & Yao, L. (2014). Violent video detection based on mosift feature and sparse coding. In *IEEE international conference on acoustics, speech and signal processing, ICASSP 2014,Florence, Italy, May 4–9, 2014* (pp. 3538–3542). doi:10.1109/ICASSP.2014.6854259.

Yang, C., Junsong, Y., & Yandong, T. (2013). Video anomaly search in crowded scenes via spatio-temporal motion context. *IEEE Transactions on Information Forensics and Security, 8*(10), 1590–1599. doi:10.1109/TIFS.2013.2272243.

Ye, Z., Li, Y., Zhao, Q., & Liu, X. (2014). A falling detection system with wireless sensor for the elderly people based on ergnomics. *International Journal of Smart Home, 8*, 187–196. doi:10.14257/ijsh.2014.8.1.20.

Yogameena, B., & Priya, K. (2015). Synoptic video based human crowd behavior analysis for forensic video surveillance. In *Advances in pattern recognition (ICAPR), 2015 eighth international conference on* (pp. 1–6).

Yujie, Z., & Zengfu, W. (2016). Real-time abnormal behavior detection in elevator. *Intelligent Visual Surveillance: 4th Chinese Conference, IVS 2016, Beijing, China, October 19, 2016, Proceedings* (pp. 154–161)). doi:10.1007/978-981-10-3476-3_19.

Zablocki, M., K. , G., D. , F., & R. , H. (2014). Intelligent video surveillance systems for public spaces - a survey. *Journal of Theoretical and Applied Computer Science, 8*, 13–27.

Zhang, X., Hu, W., Maybank, S., Li, X., & Zhu, M. (2008). Sequential particle swarm optimization for visual tracking. In *2008 IEEE conference on computer vision and pattern recognition* (pp. 1–8). doi:10.1109/CVPR.2008.4587512.

Zhang, Y., Dong, L., Li, S., & Li, J. (2014). Abnormal crowd behavior detection using interest points. In *Broadband multimedia systems and broadcasting (BMSB), 2014 IEEE international symposium on* (pp. 1–4). doi:10.1109/BMSB.2014.6873527.

Zhang, Y., Lin, W., Zhang, G., Luo, C., Jiang, D., & Yao, C. (2014). A new approach for extracting and summarizing abnormal activities in surveillance videos. In *Multimedia and expo workshops (ICMEW), 2014 ieee international conference on* (pp. 1–5).

Zhang, Y., Lu, H., Zhang, L., & Ruan, X. (2016). Combining motion and appearance cues for anomaly detection. *Pattern Recognition, 51*, 443–452. doi:10.1016/j.patcog.2015.09.005.

Zhao, F., & Li, J. (2014). Pedestrian motion tracking and crowd abnormal behavior detection based on intelligent video surveillance. *JNW, 9*(10), 2598–2605. doi:10.4304/jnw.9.10.2598-2605.

Zhao, Y., Yu, Q., Jie, Y., & Nikola, K. (2015). Abnormal activity detection using spatio-temporal feature and laplacian sparse representation. *Neural Information Processing: 22nd International Conference, ICONIP 2015, November 9–12, 2015, Proceedings, Part IV* (pp. 410–418)). Cham: Springer International Publishing. doi:10.1007/978-3-319-26561-2_49.