Video-Based Abnormal Human Behavior Recognition—A Review

Oluwatoyin P. Popoola, *Member, IEEE*, and Kejun Wang

***Abstract*—Modeling human behaviors and activity patterns for recognition or detection of special event has attracted significant research interest in recent years. Diverse methods that are abound for building intelligent vision systems aimed at scene understanding and making correct semantic inference from the observed dynamics of moving targets. Most applications are in surveillance, video content retrieval, and human–computer interfaces. This paper presents not only an update extending previous related surveys, but also a focus on contextual abnormal human behavior detection especially in video surveillance applications. The main purpose of this survey is to extensively identify existing methods and characterize the literature in a manner that brings key challenges to attention.**

***Index Terms*—Anomaly detection, behavior modeling, human action recognition, video surveillance.**

I. Introduction

MACHINE learning and understanding of human actions is a complex, diverse, and challenging area that has received much attention within the past ten years (2001–2011). Human action detection, motion tracking, scene modeling, and behavior understanding (human activity recognition and discovery of activity patterns) have received a lot of attention in the computer-vision and machine-learning communities. Applications have been in—but not limited to—video surveillance, human–computer interfaces, and multimedia semantic annotation and indexing.

Intelligent visual surveillance has got more research attention and funding due to increased global security concerns and an ever increasing need for effective monitoring of public places such as airports, railway stations, shopping malls, crowded sports arenas, military installations, etc., or for use in smart healthcare facilities such as daily activity monitoring and fall detection in old people's homes. Often times, the objective is to detect, recognize, or learn interesting events which contextually may be defined as "suspicious event" [1], "irregular behavior" [2], "uncommon behavior" [3], "unusual activity/event/behavior" [4]–[11], "abnormal behavior" [12]–[22], "anomaly" [23]–[26], etc.

Manuscript received May 9, 2011; revised September 15, 2011; accepted November 22, 2011. Date of publication January 12, 2012; date of current version October 12, 2012. This work was supported in part by the National High Technology Research and Development Program of China under Grant 2008AA01Z148, the Science Fund for Distinguished Young Scholars of Heilongjiang province under Grant JC200703, and the Funds for Science and Technology Development Creative Research of Harbin city under Grant 2007RFXXG009. This paper was recommended by Associate Editor E. Trucco.

The authors are with the Pattern Recognition and Intelligent Systems Laboratory, College of Automation, Harbin Engineering University, Harbin 150001, China (e-mail: toyinpopoola@ieee.org; wangkejun@hrbeu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCC.2011.2178594

In [27], it was suggested that of the three video surveillance research directions namely detection and tracking, human motion analysis, and activity analysis (parsing temporal sequences of object observations to produce high-level descriptions of agent actions and multi agent interactions) "...activity analysis will be the most important area of future research in video surveillance." This projection appears no less true today as research publications in this field over the last decade show. The use of closed-circuit television (CCTV) cameras to capture and monitor scenes by human agents has become ubiquitous. Although video footage capturing devices are more affordable and popular in today's world, available human resources to monitor and analyze the footage are quite limited and sometimes not cheap. In many situations where surveillance cameras are used, it is common to find poor monitoring due to human factors like fatigue. The CCTV operators suffer boredom because in most cases, nothing "strange" or something that catches the attention occurs in the scene.

It is desirable to have systems that perform intelligent realtime detection of "interesting behavior" to the human agent. The challenge, however, is that these events are rare and occur relatively infrequently (sometimes with very undesirable negative consequences). To aid human agents, efforts are being made to design intelligent surveillance systems that are capable of learning what normal behavior is and are able to distinguish between what is normal or abnormal within the context (because a normal behavior in one context may be abnormal in another).

Recently, in [28] and [29] general surveys have been published on the subject of anomaly detection. Our work is related but different in a couple of ways from them. For example, the authors of [28] carried out an impressive broad survey on anomaly detection within diverse research areas and application domains. They, however, reviewed only three papers that are related to video data applications. Much more has been done in this challenging yet exciting area during the same period under review, and this paper provides a much more in-depth and extensive coverage. In the review of [29], the focus was anomaly detection in intrusion detection and prevention systems. The main applications were systems that are deployed to ensure computer/systems network security. Our survey on anomaly detection is not about network intrusion systems but in detecting abnormal behaviors in different contexts of human activity. First, we provide a more focused review on video surveillance applications, which have attracted the most significant research attention in computer-vision applications.

TABLE I
KEY POINTS OF PREVIOUS RELATED SURVEYS

First Author, Year	*Main Focus / Contribution*
[52] Aggarwal, 1994	General non-rigid motion
[53] C. Cedras, 1994	Motion detection using motion led displays
[54] C. Cedras, 1995	Motion and trajectory tracking
[32] Aggarwal, 1997	Human motion analysis: body parts, multiple tracking
[55] Bobick, 1997	Machine perception of motion
[56] Pavlovic, 1997	Hand gestures for human computer interaction
[33] Aggarwal, 1999	Single & multi-camera tracking from image sequence
[34] Gavrila, 1999	2-D & 3-D approaches to human activity recognition
[57] T. B. Moeslund, 2001	Motion capture, tracking, pose estimation & recognition
[58] L. Wang 2002	Human motion detection & activity understanding
[59] Buxton, 2003	Generative models; scene understanding
[60] J. J. Wang, 2003	Tracking body parts to model behavior
[61] Aggarwal, 2004	Recognition
[62] Weiming Hu, 2004	Visual surveillance in dynamic scenes
[63] Valera, 2005	Automated surveillance systems
[64] D. A. Forsyth, 2006	Human motion
[35] Yilmaz, 2006	Tracking
[65] Moeslund, 2006	Initialization, tracking, pose estimation, recognition
[44] Poppe, 2007	Model-based /free approaches for motion
[38] Gandhi, 2007	Behavior modeling of pedestrian protection
[41] Krüger Volker, 2007	Action recognition
[36] Pantic et. al, 2007	Behavior understanding in HCI context
[66] Ko, 2008	Video surveillance, behavior analysis
[67] Kumar, 2008	Surveillance; data fusion in multi modal framework
[68] Turaga, 2008	Human activity detection actions and activities
[37] Zhou, 2008	Motion tracking for rehabilitation
[40] Enzweiler, 2009	Pedestrian detection systems
[42] Wei, 2009	Human motion recognition
[69] G. Lavee, 2009	Video event understanding (abstraction & modeling)
[39] Geronimo, 2010	Pedestrian protection systems
[43] Xiaofei Ji, 2010	View-invariant representation
[31] Candamo, 2010	Human behavior recognition in transit applications
[45] Poppe, 2010	Image representation & action classification methods

We choose the "anomaly" terminology because it registers well with human perception. The aim is to give a panoramic view on the area and highlight key approaches, techniques, and problems that have been addressed during the period under review. Within the scope of human action recognition, our main interest is in discovery of abnormal events and not necessarily learning of different types of activities. The abnormality-detection problem is interesting because it has wide-ranging applications from security monitoring to healthcare for the elderly people [30]. Intuitively speaking, monitoring of human activity has the latent objective to detect and take action when unusual activities take place. The intervention may either be to confront the trespasser (e.g., persons straying into restricted areas or attempting a car theft) or coming to the aid of a patient going astray or has fallen down in an old people's home. The scope of application is indeed broad, hence a broad taxonomy. Thus, the use of the term "behavior" in this paper will generally include what different authors have referred to as "actions," "events," or "activities." We shall use these interchangeably without worrying about inconsistencies of technical definitions.

The organization of this paper is as follows: Section II is a panoramic summary of related work in the general area of human motion analysis to point the reader to the key contributions of previous surveys. In Sections III and IV, we discuss the contextual characteristics of anomalies, scene modeling as well as behavior abstraction and representation. Sections V–VII contain our intuitive characterization of the research to highlight peculiar issues under different contexts, to which the researcher needs to pay attention. Finally, the concluding part—Section VIII—points out important observations and areas that need further research.

II. Related Surveys and Taxonomy

Previous surveys and the number of conferences, papers, and good surveys (see Table I) on human motion analysis and behavior recognition show that it is a highly researched area. Fig. 1 shows the frequency of publications in the area of human abnormal behavior detection published between 2002 and 2011 and reviewed in this survey. In [31], it was noted that behavior recognition publications in the past three years are three times as many as found in all of the related publications before 2005. Video surveillance, which involves acquiring and processing visual data from a scene, to detect target(s) along time and space for purpose of recognizing interesting situations and perhaps generate alarms, has been a particularly hot topic. It typically begins with change detection and motion information capture for moving targets (using tracking or nontracking methods), to enable successive high-level event analysis. Oftentimes, people pose it as a pattern-learning problem that deals with the

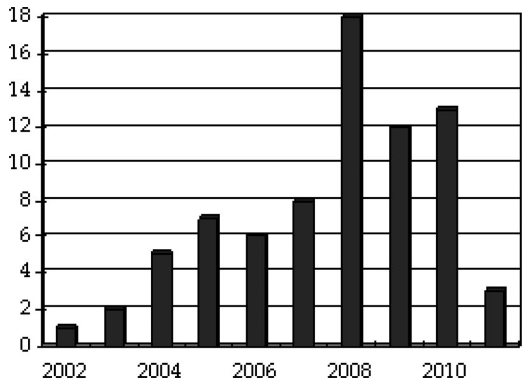

Fig. 1. Frequency of publications in the area of human abnormal behavior detection between 2002 and 2011 as reviewed in this survey. Note the rising trend in research interest in the area particularly spiking to new levels from 2008.

classification of video object behavior by finding good matches either with *a priori* known templates of behavior or learning and forming statistical models of the behavior types from time-varying feature data.

In [32] and [33], an overview of methods was presented for action recognition, recognition of body parts, and body configuration estimation. In [34], methods were grouped into 2-D approaches with or without explicit shape models and 3-D approaches. Some of the surveys reviewed problem- or application-specific work, e.g., tracking [35], human–computer interactions [36], rehabilitation [37], pedestrian protection systems [38]–[40], transit applications [31], while others are more general. In [41], different levels of complexity were highlighted in the review of action representation, recognition, synthesis, and understanding. Very recent surveys of vision-based human action recognition [31], [42], [43] emphasize 3-D pose from individual image in a sequence, while the authors of [44] and [45] present research that aimed at making the correct inference and understanding of human action patterns. They pointed out that research on semantic description of human behaviors in complex unconstrained scenes remains an open issue, and behavior patterns that are constructed by self-organizing and self-learning for unknown scenes are a desirable future research direction. In [46], a schema of the progression of research from people detection to behavior analysis was given. In another recent survey, the authors of [31] were focused on the surveillance of human activities in the context of transit applications. The recognition methods that are reviewed include single person (e.g., loitering), multiple-person interactions (e.g., fighting and personal attacks), person–vehicle interactions (e.g., vehicle vandalism), and person–vision facility/location interactions (e.g., object left behind and trespassing). In [44], the methods that are used in human action recognition were classified into global and local representation. Some surveys have emphasized outlier detection. For example, [47] and [48] present statistical approaches to novelty detection, while [49] and [50] did a comparative review of various outlier detection techniques. A broad review of anomaly-detection techniques for numeric as well with symbolic data is found in [51]. The concepts upon which the algorithms are based straddle across several disciplines such as data mining, statistics, machine learning, spectral theory, information theory, etc. Some of the questions that motivate research in this area are as follows.

1) The desired level of supervision for the system; i.e., supervised, semisupervised, unsupervised.
2) The types of features that can adequately capture the spatial and temporal signatures of the different behaviors in the scene. The fewer the measures that can unambiguously characterize and represent the original object behavior, the better.
3) How do we deal with noise in the features of interest and still ensure robustness. These features could be local features or global features, pixel based, blob based, frame based, or in groups of frames.
4) How to generate a compact lossless representation of the video clips using these features and deal with the dimensionality issues.
5) What appropriate similarity measures to use to compare different descriptors and classifiers.

This list is by no means exhaustive.

Previous surveys have not effectively reported the broad work done on video-based anomaly detection for human behavior. This is the crux of the matter in this survey.

III. Characteristics of Anomalies

Inference about "what is going on" in a scene, leading to behavior recognition or classification, is motivated by behavior summarization from pose, movement, gesture, etc. We assume that for any specific context, there is a notion of what constitutes normal behavior and, conversely, abnormal behavior. For example, while it is "normal" to people (players) to run across a football pitch during a football match, such type of motion activity is viewed as "abnormal" if it was observed on the same pitch during a marching parade.

In the literature, the term "behavior" is generic and often refers to the observable actions of agents such as persons, or other moving objects in the scene. Interestingly, abnormal or unusual patterns are somehow the "interesting" things that catche the attention of human observers, and often quite easy to identify. Such salient behaviors are so because they are different from the regular patterns in that context. Thus, anomalies are temporal or spatial outliers—events not conforming to learned patterns. They "stand out" as different relative to the context of their surrounding in space and time. Essentially, there is an underlying objective to model both the appearance and dynamics of the normal events in order to detect the presence of, and identify the spatial location of any anomaly present in the scene. Fig. 3 shows a sample of different scenes, where anomalies are detected. Given this backdrop, in [17] a definition of an abnormal event was proposed as "an action done at an unusual location, at an unusual time" or "events that are fundamentally different in appearance or having an unusual order of events." Such unusual activities are by nature rare, unexpected, atypical [5], and out-of-the-ordinary [4], [70], [71]. These characteristics make modeling for the purpose of detection of abnormal

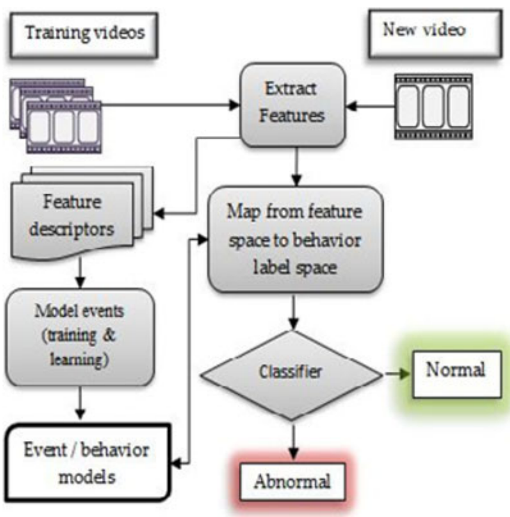

Fig. 2. General process of feature-based modeling and detection of anomalies in video sequences.

behavior, which is a nontrivial challenge. In any given scene, the anomalies can be spatially localized, spatially distributed, and temporally persistent or have some form of structure in seeming unstructured scenarios. What motivates the design of an anomaly-detection algorithm is to be able to spot when and where they occur with as little as possible false alarm.

IV. Modeling Scene Behavior—Paradigm Shift

We can view abnormal behavior detection as a type of high-level operation of image understanding, where logical information is extracted from input image sequences and used to model behavior. Fig. 2 shows a sketch of the general process.

Over the years, there has been a paradigm shift from rule-based to statistical-based methods to achieve a robust framework to conceptualize semantically meaningful scene behaviors. Rule-based systems use predefined rules to define normal or abnormal activities [72]–[74]. They perform well but are limited to detecting only those specific types of predefined anomalies. They are also very limited in terms of robustness and scalability especially for unseen events in the scene. In order to overcome these limitations, researchers utilize probability-based statistical methods to build the activity models in a data-driven fashion. These statistical methods can also be subdivided into those that first learn a model of normal behavior and use that as a basis to detect anomalies and those that automatically learn the normal and abnormal patterns from the statistical properties of the observed data, either batch or online. Any event deemed as being out-of-the-ordinary is an anomaly [75]. One drawback in creating a model of normal activity that is based on some observed behavior profile, however, is that humans can perform the same activity in a variety of ways. There will then be instances when normal behavior will look abnormal, which can trigger a false alarm (or false positive FP) in the monitoring system. Minimizing false-positive rates and false-negative rates, thus, become important challenges in abnormal activity detection.

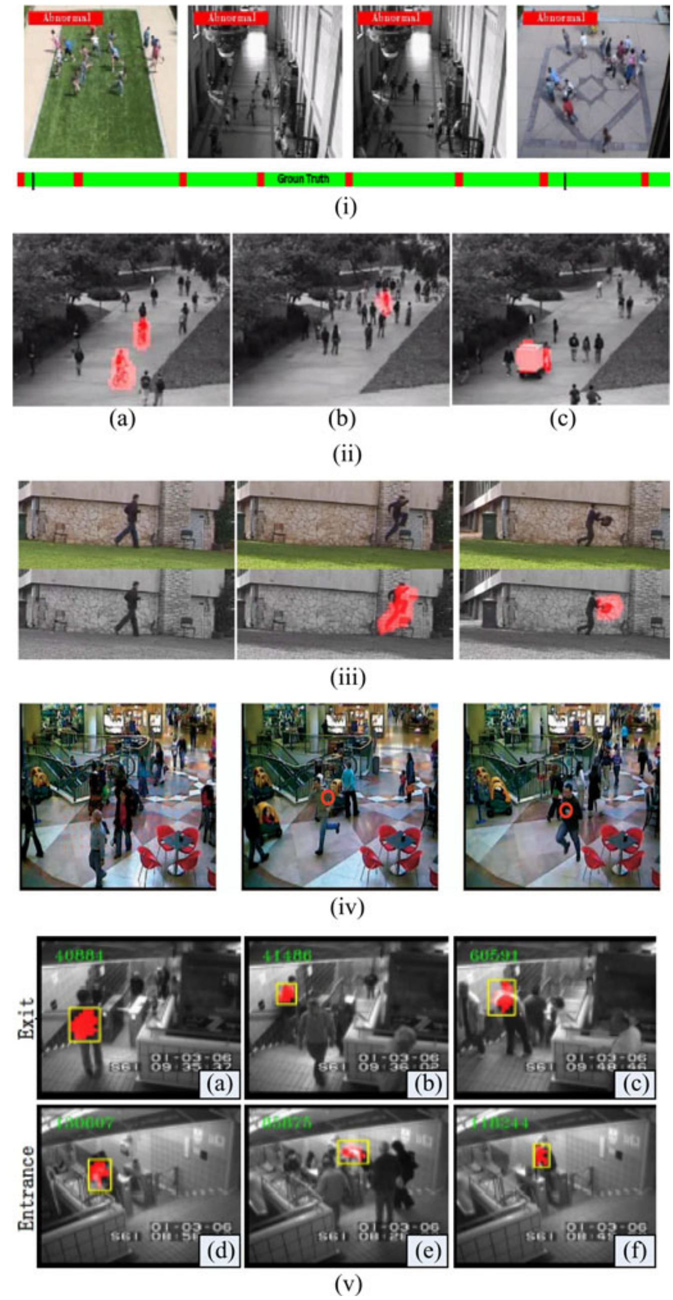

Fig. 3. Different types of anomalies in various contexts. (i) UMN dataset: anomaly in crowd movement caused by panic [16]. (ii) UCSD dataset: anomalies are (a) biker, (b) skater, and (c) vehicle in a pedestrian walkway [138]. (iii) Suspicious behavior: different from walking or jogging [102]. (iv) Unusual speed: running in a shopping mall [109]. (v) Subway: red masks in the yellow rectangle indicate the location of anomaly such as going in wrong directions of exit (top row) and entry (bottom row) areas. E and F show "no-payment" events [118].

A. Behavior Abstraction and Representation

The abstractions used to represent events and construct behavior model is one actively researched area (see Table III). Because of the huge variety of contexts with peculiar characteristics, there is always a search for more robust and descriptive features, which capture the unique properties of normal behavior. Such features should be invariant to variations in translation, rotation, illumination, etc. The abstractions are either pixel based (pixel-level description with primitives such as gradient, color, texture, motion history image, etc.) or object based (object-level description with primitives such trajectory, size, shape, and speed of object). Other commonly used object-based abstractions are bounding boxes, blobs, and silhouettes. The key to successful modeling, however, lies in the extraction of highly descriptive and discriminative features for a minimal compact representation of scene activity.

In the same vein, features that are used to encode behavior can be "global" or "local," either spatial or temporal or both. Spatial–temporal features have shown particular promise in motion understanding due to its rich descriptive power [76] and, therefore, widely used as a feature descriptor. Object-based abstractions provide direct object-level semantic interpretation of behavior, and most times trajectories from object tracking are used. In [77], the authors observed that motion features provide the strongest clue to detecting abnormal activities. Information such as trajectory, speed, and moving direction are often captured from moving objects. Optical flow is quite commonly used. When motion-information and spatial-location features are used to describe behavior, spatial deviations and unusual speed or inactivity are considered abnormal. Table II shows a list of commonly used features as well as behavior abstraction and representation methods that are found in the literature.

We note particularly the popularity of object tracking and the use of object motion trajectory—a sequence of image coordinates—due to its simplicity [78]. The use of tracking features ranges from single person to multiperson targets. For example, in [2] trajectory of treading track was used to detect abnormal behavior in a supposedly restricted area. The authors of [79] combined trajectory tracking with a model of spatial context to detect unusual activity (a fall). They employed ceiling-mounted, wide-angle cameras with vertically oriented optical axes to capture activity in a supportive home environment. The lack of motion or "inactivity" is a region, where there should be "activity" served as a cue to the occurrence of an anomaly in [80]. Other features that are captured via tracking are body contour [81], centroid track, and figure width of human blob [82]. Some researchers have used motion between consecutive frames, which makes it possible to model correlated behaviors of multiple agents. In [83] and [84], abnormal activity in a group of moving and interacting objects was detected by modeling the changing configuration as a moving and deforming "shape" and continuous hidden Markov models (HMMs) were used to capture the landmark shape dynamics.

Tracking, however, requires effective background subtraction and is limited by factors such as occlusion and shadows, thus restricting its utility to encode complex behavior in the real-world occurrence of anomalies. It is also sensitive to tracking errors even if they occur in only a few frames. This approach also fails when modeling crowded and complicated scenes. Because of the sensitivity of tracking to occlusion and tracking errors, approaches that rely on low-level features and their statistics are used to describe behaviors instead of tracks [17], [25], [71]. Other nontracking approaches extract local motion information at pixel level by using scale-invariant spatiotemporal detectors for feature extraction and description. The authors of [102] detected irregular patterns of behavior in video using an ensemble approach to model video sequences. These nontracking schemes have also been used to model group/crowd behavior (see [16], [23], [71], [109], [110], [116]–[118]).

B. *Classifying Anomaly Detection*

The choice of suitable framework to classify research in video-based human abnormal behavior detection, with such a broad spectrum of application areas, is no mean task. Table IV, for example, is a sample of possible themes along which some of these papers can be grouped. Nevertheless, such categorization may lead to a long ambiguous list since each research paper may have contributions that may be unique to certain aspects but similar to other works when viewed from a different perspective. There is a wide-ranging array of methods proposed to deal with peculiarities of specific applications (and their inherent complexities). Since there are several ways to categorize the research, we attempt in this survey to use a "critical decision" perspective. In other words, the aim is not so much at a definitive partitioning of the literature as to highlight potential issues that help researchers properly situate the problem at hand and decide on what methods to use (or not use) in developing a robust detection system for the particular context. We, thus, proceed with grouping the papers under:

1) training and learning framework;
2) the density of moving targets in the scene (implying different levels of complexity); and
3) contextual types of anomalies and available datasets.

V. Training and Learning Framework

Depending on the amount of prior knowledge and human involvement in the learning process, we may broadly categorize the research in abnormal behavior detection as supervised, unsupervised, and semisupervised.

1) *Supervised:* These methods build models of normal or/and abnormal behavior based on the labeled data [21], [102], [119], [120]. Video segments that do not fit the models are "flagged off" as abnormal. This modeling approach for unusual events' detection is good only if these abnormal events are well defined and there are enough training data. The challenge although in this is how to incorporate long-term scene adaptation. More so, a comprehensive set of all possible scenarios in the real world is impractical, hence, the growing appeal toward dynamic data-driven modeling techniques.
2) *Unsupervised:* These methods utilize the concepts of co-occurrence statistics on extracted features from unlabeled video data. They learn the normal and abnormal patterns from the statistical properties of the observed data. Isolated clusters identified as anomalies. Normal, Poisson, and other distributions are used for statistical modeling for normal patterns [85]. Variants of HMMs and the Bayesian modeling framework have been employed to make

TABLE II
COMMON TOOLS AND PARADIGMS IN THE LITERATURE

Modeling Framework ,Learning Algorithms & Techniques	*References*
Modified Altruistic Vector Quantization algorithm (AVQ)	[86]
Markov Model based	[6], [10], [30], [77], [92]
Explicit State Duration ESD-HMM	[87]
Semi-supervised HMM with Bayesian adaptation	[5]
Gaussian mixture model (+/-) HMM	[88], [128], [21], [24]
Continuous-state HMM	[8], [83]
Coupled HMM	[7], [23]
Infinite Hidden Markov Model (iHMM)	[129]
Multi-observation HMM (MOHMM)	[70]
HDP-HMM	[15], [18]
HMM- SVM	[12]
Markov Random Field (MRF)	[14]
Probabilistic Principal Component Analyzers (MPPCA)	[110]
Bayesian model (and classifier)	[79], [102], [96], [109]
Dynamic Bayesian Network (DBN)	[25]
Clustering (including hierarchical)	[85]
Poisson model	[78]
Graph-cuts	[130]
Normalized cut	[22]
Fuzzy k-means	[108]
k-medoids	[89]
Stochastic	[90]
Fuzzy c-means	[106]
Leader-and-follower	[3]
Co-occurrence statistics & bipartite co-clustering	[4]
Undirected graphical model	[115]
Dynamic Oriented Graph (DOG) & N-ary Trees classifier	[95]
Entropy model	[131]
Support Vector Machines (SVM)	[107]
Multiple Support Vector Machines (SVM)	[81], [99]
One-class SVM,	[15],[30], [94]
Neural network	[98], [113]
Fuzzy	[105], [106]
Rule based	[97]
Distance based function	[1]
Similarity function	[13]
Tread shape analysis	[2]
Probabilistic Latent Semantic Allocation (pLSA)	[9], [17]
Two stage hierarchical pLSA	[103]
Finite State Machines for sequence grammars	[91]
Fuzzy Associative Memory (FAM)	[132]
Principal Components Analysis (PCA)	[114]
Self-Organizing Matrix (SOM)	[93]
Social force model	[16]
Multi-Class Delta Latent Dirichlet Allocation (MC_LDA)	[111]
Markov Clustering Topic Model (MCTM)	[133]
Hierarchical Dirichlet Process (HDP)	[15], [18], [129]
Non-model based	[104]

probabilistic inference on unseen data [9], [15], [17], [18], [25], [70], [71], [80], [95], [121].

3) *Semisupervised:* These approaches fall in-between the first two. They learn a model of usual/unusual events using partially labeled data at either the features level or the clips level. Bayesian adaptation techniques can also be used to create models for unusual events in an unsupervised manner [5], [8],] [96], [122]. Table II shows a summary of modeling frameworks, learning algorithms, and techniques that have been used to model normal (and in some cases abnormal) behavior in different contexts. From Table II, it can be seen that Markov-based and clustering-based ap-

TABLE III
SOME METHODS OF FEATURE EXTRACTION AND REPRESENTATION

Feature Types & Representation	*References*
Object or blob trajectory (position, velocity, size, centroid etc.)	[2], [3], [10], [14], [15], [18], [24], [85], [79], [77], [78],[86], [87], [88], [89], [90], [91],[92], [93], [94]
Histograms of (Motion, direction, color, texture, pixel change, n-gram events) , textures of optical flow	[1], [4], [5], [22], [25], [95], [70], [96],[81], [97], [98], [99], [105] [100], [101]
Spatial velocity & curvature	[103]
Spatio-temporal features (including patches, saliency)	[9], [102], [106], [103]
Shape (bounding box, silhouette, image contours)	[81], [104], [7], [105], [106]
Optical flow (including blob boarder, affine from silhouette, pixel level, spatio-temporal volume of force flow),	[107], [6], [108], [109], [110], [16], [111]
Radon High Transform of color-shape information	[112]
R-Transform of silhouette shape	[21]
Pixel behavior profile	[112]
Eigen motion	[113]
Sensor traces	[12], [13], [30]
Invariant Subspace Analysis features	[120]
2D Contour angles	[122]
Inter-frame color distribution	[114]
Visual words (location, motion, size)	[17]
3D Gaussian distribution of spatio-temporal gradients	[23]

TABLE IV
POSSIBLE THEMES FOR GROUPING THE RESEARCH PUBLICATIONS

Research Problems In Focus	*References*
Real-time self-learning in video surveillance	[86]
Stochastic modeling of behavior	[77]
Automatic semantic summarization of human activity	[79]
Automatic segmentation of individual activities Temporal segmentation Semantic scene segmentation	[78] [6] [17]
Unsupervised event modeling in unconstrained environment	[4]
Inference by composition	[102]
Semi-supervised learning	[5], [8]
Unsupervised learning	[22], [95], [88]
Activity or event representation	[1], [21],[115], [114][83]
Similarity Measure	[104]
Multi-camera mining, cluttered environment	[7], [109], [91], [96]
Differentiating ambiguities with accumulated visual evidence	[25]
Incremental and adaptive detection	[103]
Hybrid of generative & discriminative model	[16]

proaches have been widely adapted and extended to solve anomaly-detection problems in video-based applications.

A. Learning From Clusters

When viewed as a clustering problem, the training and learning process for anomaly detection involves grouping similar image/video clip descriptors together and create a finite number of clusters that have unique cluster structures and possible semantic meaning. Local features are extracted and clustered using low-level abstraction to describe activities (e.g., either pixel- or object-level features). A statistical model is built, and a clustering-based outlier detection algorithm is trained using labeled or unlabeled data of either only normal or both normal and abnormal event features. Objects that are not located in the main clusters of a dataset are regarded as outliers, which may represent anomalies. The k-means algorithm is widely used to cluster features. Other improvements overcome the limitations of k-means when applied to behavior clustering such as k-medoids [89], radius-based clustering [123], and ant-based clustering [124]. In [28], it was pointed out that different clustering algorithms can be characterized based on the following key assumptions:

1) normal data instances belong to a cluster in the data, while anomalies either do not belong to any cluster;
2) normal data instances lie close to their closest cluster centroid, while anomalies are far away from their closest cluster centroid;
3) normal data instances belong to large clusters, while abnormal data belong to sparse clusters.

The statistical clustering properties of HMMs and dynamic Bayesian networks (DBNs) are explored in clustering behavior dynamics. These models quantize image, feature into a set of discrete states, and model how states change in time. When used for anomaly detection, there is an assumption that "normal data instances occur in high-probability regions of a stochastic model, while anomalies occur in the low-probability regions of the stochastic model" [28]. After quantization, a classifier is trained on instances of each class of normal activities in a binary classification of normal versus abnormal activities. One favorite tool for such binary classification is support vector machines (SVMs) [15], [30], [125]. The partitions between classes of normal activities have also been learned using multi-SVMs [20], [82]. The authors of [95] suggested the ideals of the optimal behavior classifier as being able to:

1) detect suspicious events with a minimal description of the scene context;
2) perform the detection without the need of a training stage or dataset;
3) be robust to the real-time constraints of the surveillance system;
4) be dynamic (learn and adjusts itself to changes of object behaviors).

B. Dynamic Bayesian Network Models

Bayesian network graphical modeling approaches have gained increasing popularity because they are used to encode probabilistic relationships among variables of interest.

This combined with statistical probability techniques provides a powerful machine-learning framework to effectively encode interdependences between variables, representing causal relationships and combining prior knowledge with data. The complexities depend on the nature of the modeled activities and the number of model parameters that are involved. Learning usually consists of estimating parameters and building a model that aptly represents each of the normal activities. Maximum-likelihood is used in the parameter estimation process but as complexities increase, maximum *a priori* (MAP) estimates and variational methods are employed which incorporate prior knowledge and observed data to estimate the posterior distributions and variables in the models.

Hidden Markov models (HMMs) and its variants are the most popular generative dynamic models used in the literature to model scene activity. This is because HMMs are quite effective in modeling time-varying data and capturing hidden relationship structures between variables in the system. In the Markov model type of modeling, different normal activities are studied at specific intervals, and the probabilities for the sequence of states obtained for these activities. For any previously unseen sequence of states, the larger the probability, the more likely the sequence results from normal activities.

HMMs use hidden states that correspond to different phases in the performance of an action. They model state-transition probabilities and observation probabilities. They can be used to solve three problems such as:

1) evaluate the probability of a sequence of observed events given a specific model;
2) determine the most likely evolution of an abnormal activity (state sequence) that is represented by the HMM;
3) estimate HMM parameters that produce the best representation of the most likely state sequence.

Two assumptions are imposed on HMMs; first, state transitions are conditioned only on the previous state, not on the state history (Markov assumption). Second, observations are conditioned only on the current state; therefore, subsequent observations are considered independent of each other given the current state. The Viterbi algorithm is used to extract the single most likely sequence. When using a single HMM per action, action recognition becomes finding the action HMM that could generate the observed sequence with the highest probability [87]. The authors of [86] observed that to use HMMs, there is need to define a suitable state space superimposed to the scene, thus making it impossible to obtain an adaptive system capable of controlling different scenes (a different state space must be manually defined for each different scenario). However, the authors of [6], [70], and [122] proposed some work around these limitations. Variants of the HMM models have been proposed such as multiobservation HMM [70], continuous hidden Markov models [115], switching hidden semi-Markov model [126], duration HMM in a hierarchical context [127], to address different aspects of the abnormal behavior detection problem. Anomalies have a low probability of support from the Markov chain model of the normal profile. This is a commonly shared perspective to define what constitutes abnormal activities in most of the Bayesian statistical framework.

C. Learning With Generative Topic Models

We specially discuss this approach to abnormal behavior detection because the past couple of years have observed a growing application of latent variable frameworks to model high-dimensional data typical of behavior modeling problems in computer vision. The spike in 2008 of the number of published papers on anomaly detection (see Fig. 1) coincides with the period when majority of topic model-based approaches such as latent Dirichlet allocation (LDA), probabilistic latent semantic analysis (pLSA), hierarchical dirichlet process (HDP), and various extensions appeared in the literature. These generative models have been applied to automatic text analysis in the information retrieval and language-modeling domain but have also been useful for video data. The idea is to learn a generative model of typical behavior using good discriminative features, and then detect and classify abnormal behaviors (outliers) as those that are badly explained by the learned model. In this generative model framework, the problem is set as a supervised classification problem where like in [111], examples of rare behaviors are known and available for training, and explicit models of rare versus typical behaviors are built. Generative graphical model approaches are used to learn and recognize human actions in video, because they can robustly represent sparse spatial–temporal interest points and an unsupervised learning approach [121]. These models share the same fundamental idea that "a document is a mixture of topics" but with different statistical assumptions. In this abstraction, video clips are represented as documents which are random mixtures over latent topics (action category), where each topic is characterized by a distribution over words (video words); see [134] for a fundamental introduction to topic models. Inference (i.e., computing the posterior over the hidden variables: for example, the categories of activities—topics—in the video collection) is done in most cases using Gibbs sampling [a form of Markov chain Monte Carlo (MCMC)] or variational Bayesian inference.

One key property of topic models is the ability to automatically discover meaningful activities from the co-occurrence of visual words. For the problem of anomaly detection under the Bayesian topic models, there is a nice probabilistic explanation in terms of the marginal likelihood of the video clip (motion activity in the clip), and this can be used in scenarios where multiple topics co-occur in a single clip. Unusual events are characterized by low word-topic probabilities or having visual words from existing typical topics, but co-occurring in an unusual and unique combination.

A topic modeling approach to anomaly detection has its roots in the basic LDA model. LDA is a typical standard topic model [134] which has been used to model video clips as being derived from a bag of topics drawn from a fixed (usually uniform) set of proportions. Applications to the problem of anomaly detection are found in [71], [121], and [134]. Other Bayesian modeling approaches that have been explored are pLSA [9], [17], [103], [121], [135] and the use of HDP [15], [18], [71].

VI. Density of Moving Targets

The density of moving targets in a scene influences the choice of techniques to adopt for behavior pattern characterization in video-based abnormal human behavior detection. Thus, the scene can either be viewed as either uncrowded or crowded.

Uncrowded scene: Uncrowded scenes have a small probability of occlusion (single or small number of moving objects).

In such contexts, tracking-based methods are described in [2], [3], [10], [14], [85], [91]–[94].

Crowded scene: For the crowded scenes, anomaly detection can be either pixel-level (localized) or frame-level (global) detection. Tracking fails in such scenes, and so other statistical methods that are based on low-level features are employed. Motion and appearance features are extracted from local 2-D patches or local 3-D spatiotemporal volumes to detect localized anomalies. The state-of-the-art methods include the approaches that are proposed in [16], [23], [71], [109], [110], [116], and [117]. The authors of [138] model normal crowd behavior by modeling the collection of videos as samples from a set of dynamic texture based on a linear dynamical system. In [16], abnormal crowd behavior was modeled using the social force model. LDA was used to threshold the likelihood of a clip, and distinguish abnormal from normal frames. In [23], abnormal events were detected by fitting Gaussian models from spatiotemporal gradient, and the HMM was used to detect abnormal events. In [109], the use of histograms was employed to measure the probability of optical flow for local patches. The authors of [110] model local optical flow using a mixture of probabilistic principal component analyzers and enforce consistency by the Markov random field. The authors of [136] used Lagrangian particle trajectories that are based on optical flow to model crowd scenes and computed a set of chaotic invariant features (maximal Lyapunov exponent and correlation dimension) to detect anomalies. The authors of [137] used a graph-based nonlinear dimensionality reduction method for abnormality detection.

VII. Contextual Types of Anomalies and Available Datasets

There is a mutual correlation between the density of moving targets in a scene, and the definition of abnormal behavior. Types of abnormal behavior vary depending on the characteristics of the scene. For instance, a public place/utility (usually outdoors) or a private (indoor) environment has varying definitions of what is considered abnormal (see Fig. 3). Table V shows different contexts for abnormal behavior detection. Deciding on the most suitable methods to employ depends so much on the types of abnormal behavior(s) of interest. In outdoor unconstrained environments, there is a broad range of possibilities. However, we can group existing research according to the type of anomaly detected. We group these behaviors according to single person or group behavior.

For single-person anomaly (usually based on tracking features), common anomalies are fall including motionlessness [12], [13], [99], unusual speed, [109], wrong direction, etc.

When the scene involves group behavior, it is desirable to have a probabilistic estimation of the abnormal events such as evacuation (rapid dispersion resulting from panic), crowd formation, herding in one direction, local dispersion, and splitting at different time instances as in [139]. Being able to determine the start and end of the events and the transitions between them is very important. The models that are often used are field based, i.e., they utilize features from densities and flux from motions in the scene.

VIII. Conclusion and Comments for Further Research

This review highlights recent trends in the research on video-based human abnormal behavior detection. The following can be inferred about this interesting field of research.

1) The definition anomaly can have some degree of ambiguity within a domain of application.

TABLE V
SCENE DENSITY BASED GROUPING AND SOME AVAILABLE DATASETS

Type of Scene / Datasets	*References*
Single person	
Pedestrian motion monitoring	[10], [14], [16], [86], [88], [132], [133]
Supportive home nursing or elderly people (fall detection)	[4], [12], [13], [79], [98], [113], [99]
Domestic activity monitoring (kitchen)	[87]
Game cheating detection	[4]
Audio-visual activities	[5]
Specialized activity scene (e.g. package delivery)	[115]
Indoor corridor monitoring	[9], [25], [85], [70], [104], [131], [89] , [128], [114]
Predefined daily activities	[1], [30], [105], [97], [106]
Staircase monitoring	[6], [18]
Office behavior	[21]
Specialized behavior monitoring (e.g. drivers	[90]
Elevator cage	[92]
Sparse	
Outdoor interaction	[102], [85], [78], [81], [91]
Street and pedestrian path surveillance	[2], [4], [130]
Special areas (Lobby, convenience store, airport terminal, supermarket	[22], [24], [25], [83], [109], [93]
Entry-Exit zones monitoring	[70], [96], [110]
Relatively Dense	
Crowded outdoor scenes	[23], [110], [107], [103]
Traffic scene (road, train station, subway monitoring)	[109], [131], [108], [112] [94], [77]
Complex human and vehicle interaction	[14], [17], [129], [133]
DATASET Types	
MIT PLIA 1 dataset (domestic chores)	[15]
UMN Dataset, Web dataset	[16] [140]
i-LIDS dataset	[133]
MIT dataset and QMUL dataset	[111]
Terrascope Dataset. (indoor scenes)	[7]
UCSD Anomaly Detection Dataset (pedestrian walk path)	[138]
CAVIAR Database (shopping mall)	[3]
PETS Dataset	[141]

2) Visual behaviors are complex and have much variety in an unconstrained environment.
3) The influence of noisy data, the choice and representation of low-level features, significantly influences the discriminative power of the classifier.
4) Video quality, shadows, occlusion, illumination, moving camera, and complex backgrounds are challenges especially with a single-camera view.
5) "Abnormal" activity can be maliciously adapted to appear as "normal" by the human agents. Thus, the spatiotemporal variations for the same activity can be very high even when performed by the same individual.

These and many more make anomaly detection a difficult task and therefore have no "all-purpose" algorithm that works well for different contexts. Thus, the approach has been to build computational action models into machines (via training/learning) to automatically determine whether a newly observed behavior is normal or not.

Different categorizations are possible for this research area, with each highlighting important characteristics in the problems and proposed solutions.

Much work has been done on effective feature extraction and behavior representation, more efficient modeling algorithms (emphasizing speed, less complexity, and computational requirement), and a general move from detecting abnormal behavior in a controlled single-agent indoor environment to detecting abnormal behavior in outdoors complex and crowded environments. The methods that are adopted for feature representation, learning, and definition of anomaly vary according to the peculiarities of the context. In practice, it may be easy to provide sufficiently many samples of normal activities, whereas it is difficult, if not impossible, to provide all possible examples and types of unusual activity that can happen in the scene. Because abnormal events are rare, many attempt modeling normal behavior (which have abundant data) than using abnormal events as baseline. Another added complexity is that unusual behavior may not necessarily be abnormal, but simply a new instance of a normal behavior (depending on the context).

To highlight some important properties of the contexts in which abnormal behavior detection algorithms are being developed, and the motivations behind the different proposed techniques in the literature, we have discussed the research in terms of the following:

1) the learning framework, namely supervised, unsupervised, and semisupervised approaches;
2) the density of moving targets, emphasizing the properties of crowded and uncrowded scenes;
3) type and definitions of anomalies and available datasets, reporting various contexts and definition of anomalies.

We note that while significant success has been achieved in this domain of research, some more work needs to be done as indicated next.

A. Benchmarks for Evaluating Anomaly-Detection Algorithms

The fewness of datasets that are suitable as benchmarks to train and test contextual anomaly detection poses a challenge to the community. This understandably is because of the rarity and almost infinite variety of abnormal behaviors in real life. However, harnessing the huge amount of footage that is captured by the ubiquitous CCTV cameras in many public places around the world can provide a source for standard datasets, which is useable in different contexts of interests. Such datasets should enable researchers assess how well an anomaly-detection algorithm performs in two important tasks, namely anomaly detection, i.e., "does this frame contain an anomaly or not?" Ground truths for each frame need to be provided in the dataset to evaluate performance on this task. The other is anomaly localization, i.e., "where is the anomaly taking place?" It is important in any anomaly-detection system to not only do well in detecting the presence of anomaly in the scene but identify where it is taking place.

B. Reduced Computational Requirement and Running Time

Although the processing rates of many of the competing techniques are often not stated (which limits comparative analysis), we observed in this survey that most are yet to be deployable in real time despite good performance.

Many of the algorithms employ the popular expectation maximization algorithm in the learning process. This algorithm, however, is known to have some common problems such as slow convergence, ambiguity in the optimal number of components to use, and strong reliance on the initial clustering parameters. New methods are needed to overcome these limitations as well as the video data preprocessing and feature extraction so that these systems can be useful in real-time scenarios. There is still a need for systems that can detect suspicious events with a minimal description of the scene context, and use minimal training dataset, real-time detection with constant learning and updating of the encoded information to changes of object behaviors in the scene.

C. Challenge of Complex Scene

Many anomaly-detection algorithms deal with highly structured scenes. There is need to test performance of these approaches in unstructured situations. More research attention need to be devoted to the development of frameworks that will also effectively deal with the question of scalability of video analysis especially in the real world of cluttered environments with so many moving objects and activities. This remains an open challenge.

Acknowledgment

O. P. Popoola would like to thank Prof. M. MacAlpine (Department of Electrical Engineering, Tsinghua University) and O. A. Ani for their very helpful comments and proofreading of this article.

References

[1] G. Lavee, L. Khan, and B. Thuraisingham, "A framework for a video analysis tool for suspicious event detection," *Multimedia Tools Appl.*, vol. 35, pp. 109–123, 2007.
[2] Y. Zhang and Z. Liu, "Irregular behavior recognition based on treading track," in *Proc. Int. Conf. Wavelet Anal. Pattern Recog.*, 2007, pp. 1322–1326.
[3] A. Wiliem, V. Madasu, W. Boles, and P. Yarlagadda, "Detecting uncommon trajectories," in *Proc. Digital Image Computing: Tech. and Applications*, 2008, pp. 398–404.
[4] H. Zhong, J. Shi, and M. Visontai, "Detecting unusual activity in video," in *Proc. 2004 IEEE Comput. Vis. Pattern Recog.*, 2004, vol. 2, pp. II-819–II-826.
[5] D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan, "Semi-supervised adapted HMMs for unusual event detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2005, vol. 1, pp. 611–618.
[6] J. Snoek, J. Hoey, L. Stewart, and R. S. Zemel, "Automated detection of unusual events on stairs," in *Proc. the 3rd Canadian Conf. Comput. Robot Vision*, 2006, p. 5.
[7] H. Zhou and D. Kimber, "Unusual event detection via multi-camera video mining," in *Proc. 18th Int. Conf. Pattern Recog.*, 2006, pp. 1161–1166.
[8] M. Jager, C. Knoll, and F. A. Hamprecht, "Weakly supervised learning of a classifier for unusual event detection," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1700–1708, Sep. 2008.
[9] A. Choudhary, M. Pal, S. Banerjee, and S. Chaudhury, "Unusual activity analysis using video epitomes and pLSA," in *Proc. 6th Indian Conf. Comput. Vision, Graphics Image Process.*, 2008, pp. 390–397.
[10] F. Jiang, Y. Wu, and A. K. Katsaggelos, "A dynamic hierarchical clustering method for trajectory-based unusual video event detection," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 907–913, Apr. 2009.
[11] K. Hara, T. Omori, and R. Ueno, "Detection of unusual human behavior in intelligent house," in *Proc. 2002 12th IEEE Workshop Neural Netw. Signal Process.*, 2002, pp. 697–706.
[12] Y. Hung, C. Chiang, S. J. Hsu, and C. Chan, "Abnormality detection for improving elder's daily life independent," in *Proc. 8th Int. Conf. Smart Homes Health Telematics*, Jun.22–24, 2010, pp. 186–194.
[13] K. Park, Y. Lin, V. Metsis, Z. Le, and F. Makedon, "Abnormal human behavioral pattern detection in assisted living environments," in *Proc. 3rd Int. Conf. Pervasive Technol. Related to Assistive Environments*. Samos, Greece, Jun. 23–Jun. 25, 2010.
[14] Y. Benezeth, P. Jodoin, V. Saligrama, and C. Rosenberger, "Abnormal events detection based on spatio-temporal co-occurrences," in *Proc. IEEE Conf. Comput. Vision Pattern Recog. Workshops*, Jun.20–25, 2009, pp. 2458–2465.
[15] D. H. Hu, X. Zhang, J. Yin, V. W. Zheng, Q. Yang, "Abnormal activity recognition based on HDP-HMM models," in *Proc. IJCAI*, 2009, pp. 1715–1720.
[16] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Proc. IEEE Conf. Computer Vision Pattern Recog.*, 2009, pp. 935–942.

[17] J. Varadarajan and J. Odobez, "Topic models for scene analysis and abnormality detection," in *Proc. IEEE 12th Int. Conf. Comput. Vision Workshops*, Sep. 27–Oct. 4, 2009, pp. 1338–1345.

[18] X. Zhang, H. Liu, Y. Gao, and D. H. Hu, "Detecting abnormal events via hierarchical Dirichlet processes," in *Proc. 13th Pacific-Asia Conf. Knowledge Discovery Data Mining*, Apr.27–30, 2009, pp. 278–289.

[19] A. Chianese, V. Moscato, and A. Picariello, "Detecting abnormal activities in video sequences," in *Proc. Ambi-Sys workshop on Ambient media delivery and interactive television*, pp. 1–8, 2008.

[20] Y. Chen, G. Liang, K. L. Ka, and Y. Xu, "Abnormal behavior detection by multi-SVM-based Bayesian network," in *Proc. Int. Conf. Inf. Acquisition*, Jul. 9–11, 2007, pp. 298–303.

[21] Y. Wang, K. Huang, and T. Tan, "Abnormal activity recognition in office based on R transform," in *Proc. IEEE Int. Conf. Image Process.*, 2007, pp. I-341–I-344.

[22] C.-K. Lee, M.-F. Ho, W.-S. Wen, and C.-L. Huang, "Abnormal event detection in video using N-cut clustering," in *Proc. Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, 2006, pp. 407–410.

[23] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2009, pp. 1446–1453.

[24] C. Brax, L. Niklasson, and R. Laxhammar, "An ensemble approach for increased anomaly detection performance in video surveillance data," in *Proc. 12th Int. Conf. Inf. Fusion*, 2009, pp. 694–701.

[25] T. Xiang and S. Gong, "Video behavior profiling for anomaly detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 5, pp. 893–908, May 2008.

[26] P. Feng and W. Weinong, "Anomaly detection based-on the regularity of normal behaviors," in *Proc. 1st Int. Symp. Syst. Control Aerosp. Astronautics*, Jan.19–21, 2006, pp. 1041–1046.

[27] R. T. Collins, A. J. Lipton, and T. Kanade, "Introduction to the special section on video surveillance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 745–746, Aug. 2000.

[28] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surveys*, vol. 41, pp. 1–58, Jul. 2009.

[29] A. Patcha and J. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Comput. Netw.*, vol. 51, pp. 3448–3470, 2007.

[30] J. Yin, Q. Yang, and J. J. Pan, "Sensor-based abnormal human-activity detection," *IEEE Trans. Knowledge Data Eng.*, vol. 20, no. 8, pp. 1082–1090, Aug. 2008.

[31] J. Candamo, M. Shreve, D. B. Goldgof, D. B. Sapper, and R. Kasturi, "Understanding transit scenes: A survey on human behavior-recognition algorithms," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 206–224, Mar. 2010.

[32] J. K. Aggarwal and Q. Cai, "Human motion analysis: A review," in *Proc. IEEE Nonrigid Articulated Motion Workshop*, 1997, pp. 90–102.

[33] J. K. Aggarwal and Q. Cai, "Human motion analysis: A review," *Comput. Vis. Image Understanding*, vol. 73, pp. 428–440, 1999.

[34] D. M. Gavrila, "Visual analysis of human movement: A survey," *Comput. Vis. Image Understanding*, vol. 73, pp. 82–98, 1999.

[35] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surveys*, vol. 38, 2006.

[36] M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang, "Human computing and machine understanding of human behavior: a survey,“ in *Proc. ICMI 2006 and IJCAI 2007 Int. Conf. Artif. Intell. Human Computing (ICMI'06/IJCAI'07)*, T. S. Huang, A. Nijholt, M. Pantic, and A. Pentland, Eds. Berlin, Heidelberg: Springer-Verlag, pp. 47–71.

[37] H. Zhou and H. Hu, "Human motion tracking for rehabilitation—A survey," *Biomed. Signal Process. Control*, vol. 3, pp. 1–18, 2008.

[38] T. Gandhi and M. M. Trivedi, "Pedestrian protection systems: Issues, survey, and challenges," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 413–430, Sep. 2007.

[39] A. Geronimo, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1239–1258, Jul. 2010.

[40] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: Survey and experiments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2179–2195, Dec. 2009.

[41] K. Volker, "The meaning of action: A review on action recognition and mapping," *Adv. Rob.*, vol. 21, pp. 1473–1501, 2007.

[42] W. Wei and A. Yunxiao, "Vision-based human motion recognition: A survey," in *Proc. 2nd Int. Conf. Intell. Netw. Intell. Syst.*, Nov. 1–3, 2009, pp. 386–389.

[43] X. Ji and H. Liu, "Advances in view-invariant human motion analysis: A review," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 1, pp. 13–24, Jan. 2010.

[44] R. Poppe, "Vision-based human motion analysis: An over-view," *Comput. Vis. Image Understanding*, vol. 108, pp. 4–18, Oct. 2007.

[45] R. Poppe, "A survey on vision-based human action recognition," *Image Vision Comput.*, vol. 28, pp. 976–990, 2010.

[46] R. Cucchiara, "When multimedia meets surveillance and forensics in people security," in *Proc.2nd ACM Workshop on Multimedia in forensics, security and intelligence*, 2010, pp. 49–50.

[47] M. Markou and S. Singh, "Novelty detection: A review: Part 1. Statistical approaches," *Signal Process*, vol. 83, pp. 2481–2497, 2003.

[48] M. Markou and S. Singh, "Novelty detection: A review: Part 2. Neural network based approaches," *Signal Process*, vol. 83, pp. 2499–2521, 2003.

[49] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artif. Intell. Rev.*, vol. 22, pp. 85–126, 2004.

[50] Z. A. Bakar, R. Mohemad, A. Ahmad, and M. M. Deris, "A comparative study for outlier detection techniques in data mining," in *Proc. IEEE Conf. Cybern. Intell. Syst.*, 2006, pp. 1–6.

[51] M. Agyemang, K. Barker, and R. Alhajj, "A comprehensive survey of numeric and symbolic outlier mining techniques," *Intell. Data Anal.*, vol. 10, pp. 521–538, 2006.

[52] J. K. Aggarwal, Q. Cai, W. Liao, and B. Sabata, "Articulated and elastic non-rigid motion: A review," in *Proc. IEEE Workshop on Motion Non-Rigid Articulated Objects*, 1994, pp. 2–14.

[53] C. Cedras and M. Shah, "A survey of motion analysis from moving light displays," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 1994, pp. 214–221.

[54] C. Cédras and M. Shah, "Motion-based recognition: A survey," *Image Vision Comput.*, vol. 13, pp. 129–155, 1995.

[55] A. F. Bobick, "Movement, activity and action: The role of knowledge in the perception of motion," *Philos. Trans. Roy. Soc. B, Biol. Sci.*, vol. 352, pp. 1257–1265, 1997.

[56] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human–computer interaction: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 677–695, Jul. 1997.

[57] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Comput. Vis. Image Understanding*, vol. 81, pp. 231–268, 2001.

[58] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recognit.*, vol. 36, pp. 585–601, 2002.

[59] H. Buxton, "Learning and understanding dynamic scene activity: A review," *Image Vision Comput.*, vol. 21, pp. 125–136, 2003.

[60] J. J. Wang and S. Singh, "Video analysis of human dynamics—A survey," *Real Time Imag.*, vol. 9, pp. 321–346, 2003.

[61] J. K. Aggarwal and S. Park, "Human motion: Modeling and recognition of actions and interactions," in *Proc. 2nd Int. Symp. 3D Data Process., Visualization Transmission*, 2004, pp. 640–647.

[62] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 3, pp. 334–352, Aug. 2004.

[63] M. Valera and S. A. Velastin, "Intelligent distributed surveillance systems: A review," *IEE Proc.—Vision, Image Signal Process.*, vol. 152, pp. 192–204, 2005.

[64] D. A. Forsyth, O. Arikan, L. Ikemoto, J. O'Brien, and D. Ramanan, "Computational studies of human motion: Part 1. Tracking and motion synthesis," *Foundations Trends Comp. Graphics Vision*, vol. 1, pp. 77–254, 2006.

[65] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Comput. Vis. Image Understanding*, vol. 104, pp. 90–126, 2006.

[66] T. Ko, "A survey on behavior analysis in video surveillance for homeland security applications," in *Proc. 37th IEEE Appl. Imagery Pattern Recog. Workshop*, 2008, pp. 1–8.

[67] P. Kumar, A. Mittal, and P. Kumar, "Study of robust and intelligent surveillance in visible and multi-modal framework," *Informatica*, vol. 32, pp. 63–77, 2008.

[68] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1473–1488, Nov. 2008.

[69] G. Lavee, E. Rivlin, and M. Rudzsky, "Understanding video events: A survey of methods for automatic interpretation of semantic occurrences in video," *IEEE Trans. Syst., Man, Cybern. C*, vol. 39, no. 5, pp. 489–504, Sep. 2009.

[70] T. Xiang and S. Gong, "Incremental and adaptive abnormal behaviour detection," *Comput. Vis. Image Understanding*, vol. 111, pp. 59–73, 2008.

[71] X. Wang, X. Ma, and W. E. L. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian

models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 539–555, Mar. 2009.

[72] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz, and M. Aud, "Linguistic summarization of video for fall detection using voxel person and fuzzy logic," *Comput. Vis. Image Understanding*, vol. 113, pp. 80–89, 2009.

[73] A. H. Nasution and S. Emmanuel, "Intelligent video surveillance for monitoring elderly in home environments," in *Proc. IEEE 9th Workshop Multimedia Signal Process.*, 2007, pp. 203–206.

[74] Y. A. Ivanov and A. F. Bobick, "Recognition of multi-agent interaction in video surveillance," in *Proc. 17th IEEE Int. Conf. Computer Vision*, 1999, vol. 1, pp. 169–176.

[75] N. R. Prasad, J. C. King, and T. Lu, "Machine intelligence-based decision-making (MIND) for automatic anomaly detection," *Proc. SPIE*, vol. 6574, pp. 65740F-1–65740F-9, 2007.

[76] J. C. Niebles and F. Li, "A hierarchical model of shape and appearance for human action classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun.17–22, 2007, pp. 1–8.

[77] S. Rao and P. S. Sastry, "Abnormal activity detection in video sequences using learnt probability densities," in *Proc. Conf. Convergent Technologies Asia-Pacific Region*, 2003, vol. 1, pp. 369–372.

[78] A. Panangadan, M. Mataric, and G. Sukhatme, "Detecting anomalous human interactions using laser range-finders," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2004, vol. 3, pp. 2136–2141.

[79] H. Nait-Charif and S. J. McKenna, "Activity summarisation and fall detection in a supportive home environment," in *Proc. 17th Int. Conf. Pattern Recog.*, 2004, vol. 4, pp. 323–326.

[80] F. Nater, H. Grabner, and L. Van Gool, "Exploiting simple hierarchies for unsupervised human behavior analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 13–18, 2010, pp. 2014–2021.

[81] Y. Chen, G. Liang, K. K. Lee, and Y. Xu, "Abnormal behavior detection by multi-SVM-based Bayesian network," in *Proc. Int. Conf. Inf. Acquisition*, 2007, pp. 298–303.

[82] H.-M. Qian, Y.-B. Mao, and Z.-Q. Wang, "SVM-based abnormal activity detection for home care," in *Proc. 7th World Congr. Intell. Control Autom.*, 2008, pp. 3766–3771.

[83] N. Vaswani, A. K. Roy-Chowdhury, and R. Chellappa, ""Shape Activity": A continuous-state HMM for moving/deforming shapes with application to abnormal activity detection," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1603–1616, Oct. 2005.

[84] N. Vaswani, A. Roy Chowdhury, and R. Chellappa, "Statistical shape theory for activity modeling," in *Proc. 2003 IEEE Int. Conf. Accoustics, Speech, Signal Process.*, Apr. 6–10, 2003, pp. 493–496.

[85] Y. Zhou, S. Yan, and T. S. Huang, "Detecting anomaly in videos from trajectory similarity analysis," in *Proc. IEEE Int. Conf. Multimedia Expo.*, 2007, pp. 1087–1090.

[86] A. Mecocci, M. Pannozzo, and A. Fumarola, "Automatic detection of anomalous behavioural events for advanced real-time video surveillance," in *Proc. IEEE Int. Symp. Comput. Intell. Meas. Syst. Appl.*, 2003, pp. 187–192.

[87] S. Luhr, S. Venkatesh, G. West, and H. H. Bui, "Explicit state duration HMM for abnormality detection in sequences of human activity," in *Proc. 8th Pacific Rim Int. Conf. Artif. Intell.*, Aug. 9–13, 2004, pp. 983–984.

[88] D. Makris and T. Ellis, "Learning semantic scene models from observing activity in visual surveillance," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 35, no. 3, pp. 397–408, Jun. 2005.

[89] S. Calderara, R. Cucchiara, and A. Prati, "Detection of abnormal behaviors using a mixture of von mises distributions," in *Proc. IEEE Conf. Adv. Video Signal Based Surveillance*, Sep.5–7, 2007, pp. 141–146.

[90] P. L. M. Bouttefroy, A. Bouzerdoum, S. L. Phung, and A. Beghdadi, "Abnormal behavior detection using a multi-modal stochastic learning approach," in *Proc. Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process.*, Dec. 15–8, 2008, pp. 121–126.

[91] R. Goshorn, D. Goshorn, J. Goshorn, and L. Goshorn, "Abnormal behavior-detection using sequential syntactical classification in a network of clustered cameras," in *Proc. 2nd ACM/IEEE Int. Conf. Distributed Smart Cameras*, 2008, pp. 1–10.

[92] Y. Tang, X. Wang, and H. Lu, "Intelligent video analysis technology for elevator cage abnormality detection in computer vision," in *Proc. 4th Int. Conf. Comput. Sci. Convergence Inf. Technol.*, Nov.24–26, 2009, pp. 1252–1258.

[93] G. Zhou and Y. Wu, "Anomalous event detection based on self-organizing map for supermarket monitoring," in *Proc. Int. Conf. Inf. Eng. Comput. Sci.*, 2009, pp. 1–4.

[94] C. Piciarelli and G. Foresti, "Surveillance-oriented event detection in video streams," *IEEE Intell. Syst.*, vol. 26, no. 3, pp. 32–41, May/Jun. 2010.

[95] D. Duque, H. Santos, and P. Cortez, "Prediction of abnormal behaviors for intelligent video surveillance systems," in *Proc. 1st IEEE Symp. Comput. Intell. Data Mining*, Apr. 1–5, 2007, pp. 362–367.

[96] X. Zou and B. Bhanu, "Anomalous activity classification in the distributed camera network," in *Proc. 15th IEEE Int. Conf. Image Process.*, 2008, pp. 781–784.

[97] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Fall detection from human shape and motion history using video surveillance," in *Proc. 21st Int. Conf. Adv. Inf. Netw. Appl. Workshops*, 2007, pp. 875–880.

[98] H. Foroughi, B. S. Aski, and H. Pourreza, "Intelligent video surveillance for monitoring fall detection of elderly in home environments," in *Proc. 11th Int. Conf. Comput. Inf. Technol.*, 2008, pp. 219–224.

[99] H. Foroughi, A. Rezvanian, and A. Paziraee, "Robust fall detection using human shape and multi-class support vector machine," in *Proc. 6th Indian Conf. Comput. Vision, Graphics Image Process.*, Dec.16–19, 2008, pp. 413–420.

[100] V. Reddy, C. Sanderson, and B. C. Lovell, "Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog. Workshops*, 2011, pp. 55–61.

[101] D. Ryan, S. Denman, C. Fookes, and S. Sridharan, "Textures of optical flow for real-time anomaly detection in crowds," in *Proc. 8th IEEE Int. Conf. Adv. Video Signal Based Surveillance*, 2011, pp. 230–235.

[102] O. Boiman and M. Irani, "Detecting irregularities in images and in video," in *Proc. 10th IEEE Int. Conf. Comput. Vision*, 2005, vol. 1, pp. 462–469.

[103] J. Li, S. Gong, and T. Xiang, "Global behaviour inference using probabilistic latent semantic analysis," presented at the Proc. British Machine Vision Conf., Leeds, U.K, 2008.

[104] C. E. Au, S. Skaff, and J. J. Clark, "Anomaly detection for video surveillance applications," in *Proc. 18th Int. Conf. Pattern Recog.*, 2006, pp. 888–891.

[105] Z. Jun, L. Yushu, and L. Xuhong, "Anomalous detection based on Adaboost-HMM," in *Proc. 6th World Congr. Intell. Control Autom.*, Jun. 21–23, 2006, pp. 4360–4363.

[106] Z. Jun and L. Zhijing, "Detecting irregularities by image contour based on fuzzy neural network," in *Proc.3rd Int. Conf. Innovative Comput. Inf. Control*, Jun. 18–20, 2008, p. 401.

[107] X. Wu, Y. Ou, H. Qian, and Y. Xu, "A detection system for human abnormal behavior," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2005, pp. 1204–1208.

[108] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, "A system for learning statistical motion patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1450–1464, Sep. 2006.

[109] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 555–560, Mar. 2008.

[110] J. Kim and K. Grauman, "Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2009, pp. 2921–2928.

[111] J. Li, T. M. Hospedales, S. Gong, and T. Xiang, "Learning rare behaviours," in *Proc. 10th Asian Conf. Computer Vision*, 2010, pp. 293–307.

[112] E. B. Ermis, V. Saligrama, P. Jodoin, and J. Konrad, "Motion segmentation and abnormal behavior detection via behavior clustering," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 12–15, 2008, pp. 769–772.

[113] H. Foroughi, A. Naseri, A. Saberi, and H. S. Yazdi, "An eigenspace-based approach for human fall detection using integrated time motion image and neural network," in *Proc. 9th Int. Conf. Signal Process.*, 2008, pp. 1499–1503.

[114] W. Wang, P. Zhang, and R. Wang, "Abnormal video sections detection based on inter-frames information," in *Proc. 3rd Int. Conf. Multimedia Ubiquitous Eng.*, Jun. 4–6, 2009, pp. 529–533.

[115] R. Hamid, A. Johnson, S. Batta, A. Bobick, C. Isbell, and G. Coleman, "Detection and explanation of anomalous activities: Representing activities as bags of event n-grams," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog.*, vol. 1, pp. 1031–1038.

[116] Z. Zhong, N. Ding, X. Wu, and Y. Xu, "Crowd surveillance using Markov random fields," in *Proc. IEEE Int. Conf. Autom. Logistics*, 2008, pp. 1822–1828.

[117] M. Husni and N. Suryana, "Crowd event detection in computer vision," in *Proc. 2nd Int. Conf. Signal Processing Syst.*, Jul. 5–7, 2010, pp. V1444–V1447.

[118] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2011, pp. 3449–3456.

[119] F. Nater, H. Grabner, T. Jaeggli, and L. van Gool, "Tracker trees for unusual event detection," in *Proc. IEEE 12th Int. Conf. Comput. Vision Workshops*, 2009, pp. 1113–1120.

[120] Fengjun. Lv and R. Nevatia, "Single view human action recognition using key pose matching and viterbi path searching," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2007, pp. 1–8.

[121] J. C. Niebles, H. Wang, and L. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," *Int. J. Comput. Vis.*, vol. 79, pp. 299–318, 2008.

[122] H. Li, Z. Hu, Y. Wu, and F. Wu, "Behavior modeling and abnormality detection based on semi-supervised learning method," *Ruan Jian Xue Bao/J. Software*, vol. 18, pp. 527–537, 2007.

[123] L. Ballan, M. Bertini, A. Del Bimbo, L. Seidenari, and G. Serra, "Effective codebooks for human action categorization," in *Proc. IEEE 12th Int. Conf. Comput. Vision Workshops*, Sep. 27–Oct. 4, 2009, pp. 506–513.

[124] K. Wang and O. P. Popoola, "Ant-based clustering of visual-words for unsupervised human action recognition," in *Proc. 2nd World Congr. Nature Biologically Inspired Comput.*, 2010, pp. 654–659.

[125] C. Piciarelli and G. L. Foresti, "Anomalous trajectory detection using support vector machines," in *Proc. IEEE Conf. Adv. Video Signal Based Surveillance*, 2007, pp. 153–158.

[126] T. V. Duong, H. H. Bui, D. Q. Phung, and S. Venkatesh, "Activity recognition and abnormality detection with the switching hidden semi-Markov model," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog.*, 2005, vol. 1, pp. 838–845.

[127] P. Chung and C. Liu, "A daily behavior enabled hidden Markov model for human behavior understanding," *Pattern Recognit.*, vol. 41, pp. 1589–1597, 2008.

[128] W. Wang, R. Wang, and Y. Chen, "The abnormal behavior analysis of single person on the road based on region and behavior features," *Proc. SPIE*, vol. 6786, p. 67862N, 2007.

[129] I. Pruteanu-Malinici and L. Carin, "Infinite hidden Markov models for unusual-event detection in video," *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 811–822, May 2008.

[130] I. N. Junejo, O. Javed, and M. Shah, "Multi feature path modeling for video surveillance," in *Proc. 17th Int. Conf. Pattern Recog.*, 2004, vol. 2, pp. 716–719.

[131] H. Hung and S. Gong, "Detecting and quantifying unusual interactions by correlating salient motion," in *Proc. IEEE Conf. Adv. Video Signal Based Surveillance*, 2005, pp. 46–51.

[132] Z. Wang and J. Zhang, "Detecting pedestrian ab-normal behavior based on fuzzy associative memory," in *Proc. 4th Int. Conf. Natural Comput.*, 2008, pp. 143–147.

[133] T. Hospedales, S. Gong, and T. Xiang, "A Markov clustering topic model for mining behaviour in video," in *Proc. IEEE 12th Int. Conf. Comput. Vision*, 2009, pp. 1165–1172.

[134] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[135] J. Zhang and S. Gong, "Action categorization by structural probabilistic latent semantic analysis," *Comput. Vis. Image Understanding*, vol. 114, pp. 857–864, 2010.

[136] S. Wu, B. E. Moore, and M. Shah, "Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2010, pp. 2054–2060.

[137] I. Tziakos, A. Cavallaro, and L.-Q. Xu, "Local abnormality detection in video using subspace learning," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surveillance*, 2010, pp. 519–525.

[138] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2010, pp. 1975–1981.

[139] Y. Benabbas, N. Ihaddadene, and C. Djeraba, "Motion pattern extraction and event detection for automatic visual surveillance," *J. Image Video Process.*, vol. 2011, pp. 7:1–7:15, Jan. 2011.

[140] R. Mehran, A. Oyama, and M. Shah. (2011). UMN dataset. [Online]. Available: http:www.cs.ucf.edu/∼ramin/?page_id = 24#1. UMN_Dataset

[141] University of Reading. (2010). PETS 2010 benchmark data. [Online]. Available: http:www.cvg.rdg.ac.uk/PETS2010/a.html

Oluwatoyin P. Popoola (M'09) received the B.Sc. and M.Sc. degrees in industrial and production engineering from the University of Ibadan, Ibadan, Nigeria, in 1999 and 2003, respectively. He is currently working toward the Ph.D. degree with the Pattern Recognition and Intelligent Systems Laboratory, College of Automation, Harbin Engineering University, Harbin, China.

He joined the faculty at Systems Engineering Department, University of Lagos, Nigeria, in 2004. His research interests include pattern recognition, machine learning, biometric recognition, and industrial engineering.

Kejun Wang received the B.S. degree from Northeast Heavy Machinery Institute, Hebei, China, in 1984, the M.S. degrees from Harbin Shipbuilding Engineering Institute, Harbin, China, in 1987, and the Ph.D. degree from Harbin Engineering University, Harbin, China, in 1995.

He is a Professor with Harbin Engineering University. His fields of research interest include biometric recognition, image processing, neural networks, machine learning, and machine perception.