A Review of Anomaly Detection in Automated Surveillance

Angela A. Sodemann, Matthew P. Ross, and Brett J. Borghetti

***Abstract*—As surveillance becomes ubiquitous, the amount of data to be processed grows along with the demand for manpower to interpret the data. A key goal of surveillance is to detect behaviors that can be considered anomalous. As a result, an extensive body of research in automated surveillance has been developed, often with the goal of automatic detection of anomalies. Research into anomaly detection in automated surveillance covers a wide range of domains, employing a vast array of techniques. This review presents an overview of recent research approaches on the topic of anomaly detection in automated surveillance. The reviewed studies are analyzed across five aspects: surveillance target, anomaly definitions and assumptions, types of sensors used and the feature extraction processes, learning methods, and modeling algorithms.**

***Index Terms*—Abnormal behavior, anomaly detection, automated surveillance, behavior classification, machine learning.**

I. Introduction

IN RECENT years, a wealth of research has been undertaken in the domain of human behavior classification in automated surveillance. Behavior classification involves the categorization or classification of perceived behavioral events by an algorithm. This research effort has been driven by an increased concern for security and safety, coupled with an overabundance of available surveillance data relative to the amount of manpower available to process it.

Anomaly detection in automated surveillance is a subset of behavior classification problems reduced to a two-class or one-class classification problem. In the anomaly detection in automated surveillance process, sensors in an environment collect data representing the behavior of surveillance targets, with some behaviors assumed to be anomalous. The raw sensor data are then subjected to a feature extraction process. The resulting features become the input to a modeling algorithm, in which a learning method is applied to determine the normal or anomalous state of the observed behavior. Fig. 1 illustrates the relationship between the key aspects of the process of anomaly detection in automated surveillance. The process illustrated in Fig. 1 can be implemented either real time or offline, in which case the feature extraction and modeling is applied to recorded sensor data. This review will primarily focus on real-time implementations; see the review by Saykol *et al.* [1] for additional information specific to the offline case.

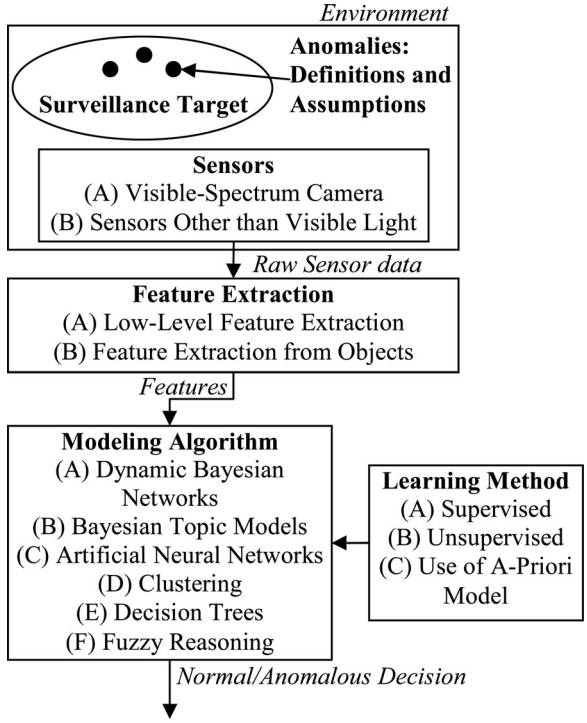


Fig. 1. Diagram of the flow from environment to anomaly detection, illustrating the organization of this review.

Automated anomaly detection is highly useful in reducing the amount of data to be processed manually by directing attention to a specific portion of the data, to the exclusion of the vast amounts of irrelevant data. However, the problem of anomaly detection is greatly open to interpretation, and research efforts are scattered not only in approach, but also in interpretation of the problem, assumptions, and objectives. This review will attempt to bring synergy to these disparate efforts by evaluating the problem formulations and solution methods applied in anomaly detection research as applied to automated surveillance.

Although existing surveys treat topics related to anomaly detection in automated surveillance, none satisfactorily treats the subject itself. The related topic of categorizing the genre of produced video was covered in a review by Brezeale and Cook [2] in 2008, while the more focused topic of understanding specific events in video data was addressed in a review by Lavee *et al.* [3] in 2009. Buxton [4] presented a 2003 survey on understanding

Manuscript received May 24, 2011; revised November 21, 2011, March 14, 2012, and June 22, 2012; accepted August 6, 2012. Date of current version December 17, 2012. This work was supported by the U.S. Department of Defense. This paper was recommended by Associate Editor H. Liu.

A. A. Sodemann is with the College of Technology and Innovation, Arizona State University, Phoenix, AZ 85069 USA (e-mail: angela.sodemann@gmail.com).

M. P. Ross is with the U.S. Air Force Academy, Colorado Springs, CO 80840 USA (e-mail: matthew.ross@us.af.mil).

B. J. Borghetti is with the Air Force Institute of Technology, Dayton, OH 45433-7765 USA (e-mail: brett.borghetti@afit.edu).

Digital Object Identifier 10.1109/TSMCC.2012.2215319

dynamic scene activity. Hu *et al.* [5] produced a 2004 survey of automated visual surveillance which focuses on the two areas of motion detection and object tracking, offering only a short overview of behavior understanding. Ko [6] generated a survey on behavior analysis in video surveillance for homeland security applications. Citing [5], Ko noted that "relatively little research has been reported on reliable classification and understanding of human activities from video image sequences" [6]. The survey by Ko focuses on individual components involved in video processing for automated surveillance: background modeling, object segmentation, object classification, object tracking, behavior analysis, and related supporting tasks. Dee and Hogg, in their 2008 review of real-world surveillance [7], include a section reviewing anomaly detection. However, Dee and Hogg refer their readers to the earlier reviews [4], [5] for further treatment. Haering *et al.* [8] also produced a broad overview of surveillance in 2003, and Raty [9] a survey of surveillance systems in 2010. None of these reviews sufficiently treat the task of anomaly detection.

Anomaly detection, which is also known as outlier or novelty detection, is a widely studied topic that has been applied to many fields including medical diagnosis, marketing, network intrusion, and many applications other than automated surveillance [10]. Many good reviews of anomaly detection can be found in the literature, including the broad overview of the various uses and types of anomaly detection by Chandola *et al.* [11] in 2009.

None of these previous surveys adequately treat the large body of work that addresses the confluence of these two fields: anomaly detection and surveillance, despite the large body of research that has been produced on this specific topic. In this study, we seek to create a more focused review of recent publications on anomaly detection in automated surveillance. Thus, we will be able to address in-depth those particular challenges involved with this topic. This review will focus on the broader problem formulations and assumptions applied in anomaly detection in automated surveillance studies, rather than primarily providing a review of specific pattern-classification methods.

This paper will address the key process elements written in bold-face type in Fig. 1, with subtopics corresponding to the normal-face type. Section II will address the various surveillance targets commonly considered in anomaly detection research. Section III will review definitions and assumptions about the concept of *anomaly* as applied in various studies. In Section IV, the variety of sensors employed will be described, and the feature extraction process will be addressed. Learning methods will be reviewed in Section V, and a summary of machine learning and classification algorithms that have been applied to anomaly detection in automated surveillance will be presented in Section VI. Conclusions and discussions will be given in Section VII, followed by a discussion of areas in need of future work in Section VIII.

II. Surveillance Target

The *target* of surveillance is the entity or entities upon which the surveillance operates. In other words, the target consists of those entities among which the anomaly detection method aims to detect anomalies. Examples of common targets include individuals, crowds, automobile traffic, and inanimate objects. A Venn diagram illustrating the relationship between automated surveillance, anomaly detection, and the four most common surveillance targets is shown in Fig. 2.

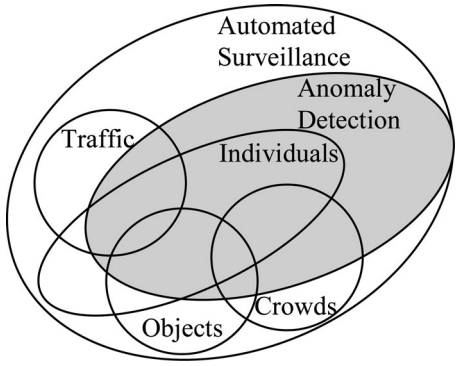


Fig. 2. Venn diagram of anomaly detection methods in automated surveillance classified by their targets, illustrating that some methods do not involve anomaly detection and that some methods apply to more than one type of target.

As illustrated in Fig. 2, there is a large body of research addressing automated surveillance of individuals, traffic, objects, and crowds that do not include the process of anomaly detection. Some of this work is broader, seeking to classify the behaviors or the environment into additional classes other than anomalous or normal. Other studies address specific tasks of raw data processing, such as object identification, motion detection, or background subtraction without attempting to perform any manner of behavior or scene understanding. Such studies will not be included in this review. Rather, this review includes those studies which can be categorized within the shaded region of Fig. 2: studies addressing anomaly detection in automated surveillance with targets including traffic, individuals, crowds, and objects. Note also in Fig. 2 that there is a distinction made between individuals and crowds. Although both of these targets consist of people, the methods used to detect anomalies among individual people are distinct from those used to detect anomalies among crowds of people. Thus, this review will treat crowds and individuals separately. In this review, efforts to detect anomalies are classified by the targets upon which the method was tested by the researchers. If the study claimed that the method is useful for many different targets, but only verified the method on a single type, only the verification target is included here.

A large body of work on anomaly detection for individuals has been applied to ensuring the health and safety of elderly or infirm individuals in a nursing home, hospital, or home setting [12]–[18]. Other studies focus on detecting anomalies in behavior indicating lawbreaking or breach of security [19]–[22] or events indicating a safety problem [23], [24]. A few studies have unspecified applications, and focus on detection of events that are rare, but may not have any specific meaning. For example, a study by Wang *et al.* [25] considers behavior such as an individual rushing into the scene or suddenly bending down while walking to be anomalous.

Studies which target traffic generally intend to detect either traffic law violations or safety issues such as an accident or congestion. An interesting study by Hayashi and Yamada [26] aims to predict the dangerous driving behavior of turning across oncoming traffic.

The third most common target of anomaly detection in automated surveillance is crowds. These studies [27]–[32] typically attempt to detect anomalous crowd motion indicated by a significant change in kinetic energy or pace, although a unique study by Mehran *et al.* [33] detects crowd motion which violates the social force model as anomalous.

In this review, no studies were found which addressed inanimate objects alone as the target, since most anomaly detection efforts are associated with human behavior. Consideration of both individuals and traffic together in the same scene is a common topic addressed by a number of studies that attempt to account for the presence of both vehicles and pedestrians on a road scene. Methods that allow for both individuals and inanimate objects as simultaneous targets is also a widely addressed topic, given the potential benefits of detecting behaviors such as objects changing hands or individuals abandoning baggage. A much more sparsely addressed topic is methods that allow for either individuals or crowds to be in the scene. This is a difficult topic because the methods typically applied to the two are fundamentally different: Individual targets are typically surveilled using object-based feature extraction methods, while crowd targets are addressed with low-level feature extraction methods (this will be addressed further in Section IV). Since crowd targets and individual targets differ only in the number of humans present in the scene, further research into methods that can be applied to both crowd targets and individual targets seems to be a necessary area for further study.

III. Anomaly Definitions and Assumptions

Formulation of an approach to anomaly detection requires the application of definitions and assumptions about anomalous behavior. The definitions and assumptions made vary with the target of the surveillance effort, as well as the overarching goals considered by the researchers. The definitions and assumptions, in turn, affect the methods subsequently applied to perform the anomaly detection, as well as the choice of sensors and feature extraction. The task of defining the concept of *anomaly* can be at once challenging and critical to the success and robustness of the anomaly detector.

There are three common definitions and assumptions of anomalous behavior applied in research:

1) anomalous events occur infrequently in comparison to normal events;
2) anomalous events have significantly different characteristics from normal events;
3) anomalous events are events which have a specific meaning.

Within the first category, there are several studies defining anomalous events as events that are not sufficiently represented within the data available for modeling [23], [34]–[36]. Xiang and Gong [37] specify that anomalies of this type must be distinguished from noise, which can cause a similar effect resulting in a false positive. Li *et al.* [38] specifically address the case that an anomaly is a rare event, although not necessarily significantly distinct from normal events. A study by Li and Han [21], attempting to detect attacks on a network, relies upon the assumption that the number of attackers is much less than the number of honest users. A unique study by Hospedales *et al.* [39] defines anomalous behavior according to *visual saliency* or irregularity from the concept of Bayesian surprise (a significant difference between posterior and prior beliefs of the observer after the event has occurred).

Application of the frequency-based definition of anomalies often results in a high rate of false positives, since many events which would typically be considered normal might nevertheless occur infrequently. Success in anomaly detection with this assumption also requires the availability of many data samples for modeling. Every event which is not to be flagged as anomalous must be well represented in the data used to create the model.

Studies within the second category of assumptions define anomalous events as those which are significantly distinct from the normal events, regardless of their representation in the data available for modeling or general frequency of occurrence [40], [41]. Studies have applied this definition by considering anomalies as having a large distance from data representing normal events [42] within feature space, or having low probability given a learned normal model [43]. Other studies consider as anomalous those events which exhibit a difference in an event duration [16], or position and trajectory [44]. A significant limitation of the second category of assumptions is the inability to detect anomalies which are not significantly distinct from normal events. This is a particular problem when a target is specifically attempting to conceal an anomalous action, such as a behavior associated with a planned crime or terrorist activity.

Methods within the third category of anomaly assumptions are sometimes able to detect anomalies that the second category could miss. The third category defines anomalous events *a priori* as specific meaningful actions or occurrences. This is a specific instance of the more general field of *event classification*, which attempts to ascribe meaning to surveillance video by labeling events. As applied to anomaly detection, event classification aims to identify and label specifically those events which are dangerous, threatening, or illegal. Examples from the literature reviewed here include stealing an item [19], dangerous events such as animals or pedestrians on a highway [45], and unattended electrical devices [13]. *A priori* definitions of anomalous events are also applied in security applications to detect specific security violations such as an intrusion [20], [46] or leaving or picking up an item [47]. Although this assumption succeeds in detecting those anomalies which are difficult to distinguish from normal events, this assumption has the drawback of being very narrow in the range of events which are detectable. This type of anomaly detection is able to detect only the single event it is designed for, and is unable to detect unexpected or different types of anomalies.

A few studies attempt to capitalize on the benefits of each definition without the drawbacks by applying definitions and assumptions which fit within multiple or unique categories.

Xiang and Gong [48] consider anomalous those events which occur infrequently, but note that the definitions of anomalies are highly dependent on context and can change over time. Their model is designed to update over time by augmenting the normal behavior model with previously rare behaviors which increase in frequency of occurrence. This approach alleviates the high false-alarm rate due to changes over time, but does not address the high rate of false alarms due to rare normal events. A unique assumption by Dee and Hogg [7] is that anomalies are those events which cannot be explained by an external model. In the study by Dee and Hogg, the authors apply a psychological model of intention to ignore events which can be explained as an expected action of a human pursuing a goal. The success of this method is dependent upon the accuracy of the external model as well as the applicability of the model to the target environment.

IV. Sensors and Feature Extraction

Sensors are selected in consideration of steps in the anomaly detection process such as feature extraction and modeling method. For example, given the task of detecting fear in human subjects (as in [49]), it may be easier to detect this by audio sensor (recording loud, high-pitched speaking) than by visual sensor (recording facial expression). Indeed, some anomalies are undetectable by some sensors; for example, low-resolution sensors may not detect facial expression at all. Availability, reliability, and cost also impact the choice of sensor. All sensors have inherent advantages and limitations. For example, occlusion (the temporary loss of an entity due to its being blocked from sensor view) is a problem with most visual sensors and rarely a problem with audio sensors.

Raw sensor data typically require preprocessing for feature extraction before these are classified as normal or anomalous. Choice of features to be extracted and methods used in the extraction varies primarily by sensor type and by target. This section discusses the range of sensor types that have been considered for anomaly detection in automated surveillance research, the problems and benefits of each sensor type, and the types of features extracted.

A. Visible-Spectrum Camera

The visible-spectrum camera is the most common sensor applied in automated surveillance anomaly detection, due to its wide availability. However, a good deal of preprocessing of data is often necessary to extract useful information from a visible-spectrum camera, and the applications are limited to detection of anomalous behavior that is visibly distinguishable from normal. Visible light cameras are also sensitive to illumination variations, which can cause false detection of anomalies.

Approaches to anomaly detection with a visible light camera are affected by field of view and resolution of the camera. Field of view ranges from as wide as a square kilometer to as narrow as a single square meter. Fig. 3 shows the range of fields of view typically considered in anomaly detection research, along with characteristic images.

Because field of view, resolution, and surveillance target are interdependent variables, this review defines resolution categories ranging between 1 and 5 by the resolution of possible targets, as indicated in Table I.

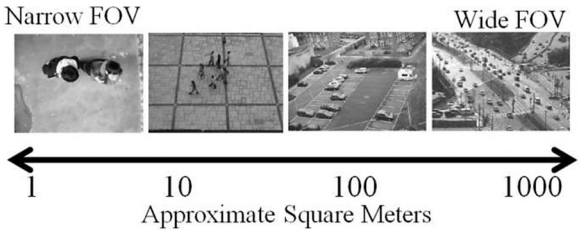


Fig. 3. Illustration of the approximate fields of view in square meters of reviewed studies utilizing a visible light camera as primary sensor. Images (left to right) from [7], [30], [50], [51].

TABLE I
Comparative Resolution Values by Approximate Number of Pixels

Resolution Value	# Pixels in Cars	# Pixels in People	# Pixels in Limbs	# Pixels in Faces
1	10^1	< 1	< 1	< 1
2	10^1-10^2	10^1	< 1	< 1
3	10^2-10^3	10^1-10^2	10^1	< 1
4	$> 10^3$	10^2-10^3	10^1-10^2	10^1
5	$> 10^3$	$> 10^3$	10^2-10^3	10^1-10^2

The values in Fig. 3 and Table I are referenced in Section VII with a list of the reviewed papers which have considered each value. Among the literature, there has been significantly greater interest in fields of view and resolutions in the middle of the scale rather than at the extremes. This is likely due to the much wider availability of data in these ranges. However, certain significant applications in this field, such as wide-area surveillance, are limited to data streams of wide field of view and low resolution. Further research is needed in the study of these extreme cases.

An attempt to account for a wider range of field of view and resolution has been made in the use of multiple cameras. Some multiple camera studies assume that the cameras have overlapping fields of view [36], [52], while others treat the camera setup as independent images [53]. An advantage of the use of multiple overlapping cameras is the possibility of compensating for occlusion through use of multiple views of the same visual area [54]. A primary challenge with the use of multiple cameras is the difficulty of correlating observed events between cameras. Anjum and Cavallaro [52] address this issue by matching extracted trajectories across multiple cameras to determine which trajectories are continuous. Ermis *et al.* [36] use pixel busy-idle periods to match activity between cameras, extending this method to the multicamera case. A study by Wang *et al.* [55] considers nonoverlapping cameras. In this study, it is assumed that there is no correspondence between the cameras. Rather, each individual video feed is treated as an independent source of information, providing unique features.

Feature extraction methods in studies which employ a visible light camera as the primary sensor can be broken down into

1) those methods which first perform some method of target tracking or identification, extracting high-level features from the identified targets, and 2) those that extract low-level features directly from the image at the pixel level. An example of the first method is a system which detects and tracks individual vehicles from full motion video, while an example of the second method is a system that determines the frequency and rate of change for each pixel over subsequent frames to build a map of motion levels in the scene.

Type of feature extraction applied with visible light video is related to the type of target: all of the papers reviewed here which address crowd anomaly detection apply some form of low-level pixel-based feature extraction, while object extraction and tracking is a much more common method to be applied to anomaly detection in individuals. This is due to the difficulty involved in *image segmentation*, the separation of an image into multiple individual objects, when the image consists of multiple overlapping objects. However, this difference in feature extraction method applied to individuals compared with crowds causes difficulty in defining a single anomaly detection method that is equally successful whether a crowd or an individual is in the scene.

1) Low-Level Feature Extraction: Low-level feature extraction is the process of detecting and characterizing motion, color, and other fundamental image properties. The low-level feature extraction approach has the advantages of being robust to image-processing difficulties, such as occlusion, which affect tracking accuracy. Since no objects are extracted from the image, this approach is able to operate successfully even with large numbers of targets in view. However, this approach also results in less specific information about the scene and has only been used to detect anomalies involving many or all of the targets in the scene.

Secondary processes can also be applied to extract additional features derived from the primary low-level features. A particular benefit of this approach is the potential to account for time dependences. Pruteanu-Malinici and Carin [43], in their study on use of infinite hidden Markov models (iHMMs) in video anomaly detection, review several of these methods, including shift-invariant wavelets, independent component analysis, and invariant-subspace analysis.

Low-level features can be used as direct inputs to anomaly detection algorithms. Several studies [27]–[29], [32], [33], [56] apply optical flow methods for this purpose. Using this technique, anomaly detection algorithms operate on global properties, such as total kinetic energy, of the area in view. Some multicamera studies [36], [53] match low-level features between cameras. Ermis *et al.* [36] match dwell times of foreground pixels (after motion-based background subtraction), and Li *et al.* [53] match the motion in regions from optical flow, without initial background subtraction.

A unique approach by Dong *et al.* [57] combines both low-level feature extraction and feature extraction from objects. In this study, an HSV (hue, saturation, value) image is created in which the image color represents motion data extracted from visible-spectrum video. Objects are then identified within the HSV image. Thus, identified objects do not correspond to physical objects, but to regions with similar motion characteristics. Finally, features such as color, texture, and shape are extracted from the objects and used in the anomaly detection. In a similar approach by Saykol *et al.* [47], pixels in full-motion video are converted to a low-resolution grid which encodes information about the motion of corresponding image regions. The grid regions hold values indicating moving, stopping, merging, or splitting.

Low-level features are often used as a preliminary step to object-based feature extraction. A low-level feature extraction method such as background subtraction is applied to differentiate between foreground objects and the background, where foreground objects are defined as objects that are of interest, and the background is defined as objects that are not of interest. Object-based features are then extracted from the identified foreground objects.

2) Feature Extraction From Objects: To extract more specific information about an individual target than is possible with the low-level feature extraction method, the Feature Extraction from Objects method is more typically employed. The choice of features to be extracted from objects depends upon the target, the type of anomaly expected, and the environment. Field of view and resolution are also key factors in the choice of features; the lower the resolution on the target object, the more limited the potential feature set.

The most common features to be extracted from objects are the position and trajectory of the object's centroid. These two features are sufficient in many studies targeting individuals to determine violation of a restricted area [58]–[60], specific anomalous behaviors such as running or falling [61], or unusual paths indicating loitering or confusion [62]–[64]. Position and trajectory are useful in traffic anomaly detection to detect tailgating [64], or illegal traffic maneuvers [65], [66]. Trajectory and position of objects can also be used to extract related features such as relative distance between objects, acceleration, and motion energy [67]. Adding size of object as a feature allows distinguishing between cars and pedestrians, and is frequently applied to detection of anomalies at intersections and crosswalks [68]–[71].

In some studies with a high resolution targeting individuals, the extracted features are limb angles [72]–[74]. These angles can be used to determine posture or infer a specific action the target is taking. In a study by Lee *et al.* [74], entropy of the limbs is used to determine if the behavior is anomalous. Other studies do not consider limb angles, but extract features about other specific body parts or regions. Zhou and Wu [75] perform a color transformation to YIQ (yellow, in-phase, quadrature) color to identify exposed skin on the hands and then extract the trajectories of the hands to determine if a target is shoplifting. The YIQ color space is the standard in National Television System Committee television transmission and provides higher contrast among human skin tones [75]. Chen *et al.* [76] extract the History Motion Image of regions of the object image corresponding to the arms and legs of the target. Wu *et al.* [77] consider only the torso angle in order to determine if an individual is running or falling down (actions previously defined as anomalous).

Some object-based features do not require tracking. Information about human posture can be gathered by considering features of an extracted object's perimeter. Zhang and Liu [78] "unwrap" the boundary of an identified human silhouette to create a posture signature feature, while Lao *et al.* [79] consider the vertical and horizontal projections of a silhouette. The target's enclosing rectangle is another object-based feature that does not require tracking. Guo and Miao [80] create a system that is able to recognize six postures and four activities from the ratio of black to white pixels in the enclosing rectangle of a human silhouette. Tang *et al.* [50] address the detection of crime in an elevator from the length and width of the enclosing rectangle of the elevator passengers. Xiang and Gong [37] add the filling ratio of foreground pixels to a feature vector that also includes the dimensions, centroid location, and first-order moments of objects to differentiate between typical and atypical behaviors of individuals. In addition to a bounding-box, Chen *et al.* [81] also apply a convex-hull approach to recognize swimming strokes and detect drowning behaviors.

B. Sensors Other Than Visible Light

Some studies apply sensors other than visible light cameras to anomaly detection. A number of these studies make use of the collective information from many simple sensors distributed throughout the environment. One advantage of this approach is the ability to cover a wider area within the environment than is possible by the restricted field of view of a camera. Mishra *et al.* [20] propose the distribution of simple sound and light sensors along a national border to detect intruders. This study also notes that such sensors, due to their simplicity, could easily be powered from the environment such as from solar power or ambient vibrations. Such simple sensors also have the advantage of being less intrusive to targets who may be sensitive to privacy, such as individuals in a home or nursing environment, as is noted by Sawai and Yoshida [82].

Audio signals are particularly useful for determining information about an individual target's emotional state. Clavel *et al.* [49] make use of this advantage to detect the expression of fear in a target as an indicator of anomalous behavior. Wu *et al.* [77] process audio data to detect behavior such as crying or groaning. This study also makes use of audio information to direct the attention of a visible light camera to an area of interest to gather additional information.

Feature extraction from sensors other than visible light cameras presents a wide range of possibilities due to the variety of data types available. Andersson *et al.* [27] extract low-level motion data about individuals in the scene independent of the background by applying optical flow to the image from an infrared camera. The studies by Wu *et al.* [77] and Clavel *et al.* [49] both extract the Mel Frequency Cepstral Coefficients, a type of frequency decomposition particularly suited for audio signals, from audio signals as features. Stoecklin [83] and Adhiya *et al.* [84] extract flow duration, octets per flow, packets per flow, IP addresses, TCP and UDP ports, and other network characteristics as features to detect anomalies in the computer network. Sawai and Yoshida [82] and Park *et al.* [85] extract features such as states and time between state changes from simple motion sensors to detect anomalous behavior of elderly individuals at home, while Chae *et al.* [86] extract similar low-level information, including time and location, from radio-frequency identification (RFID) tags.

V. Learning Method

This section will review the learning methods applied to behavior modeling and anomaly detection. Learning method is closely related to the method for anomaly detection. For example, a learning method that involves supervised learning of normal behavior will employ an anomaly detection method of detecting events which do not fit the learned normal model. Therefore, this section will also review the anomaly detection methods which are associated with learning methods.

Learning methods can be broadly considered to belong to one of three primary groups:

1) supervised learning;
2) unsupervised learning;
3) *a priori* knowledge application.

These groups have recently been subdivided into many new related groups, such as weakly supervised, semisupervised, etc. In this review, we are considering a more broad generalization so that any method that requires any labeling of training data will be considered to be within the supervised learning group. Methods that require training data but do not require labels are considered unsupervised, and those that do not require any training data are considered to involve *a priori* knowledge application.

A. Supervised Learning

Supervised learning refers to a method of modeling in which an algorithm learns a model from a representative set of data, known as *training data*, in which each element of the training data is labeled with its true class membership. During the training phase, the algorithm learns relationships between the data elements and the corresponding class membership. After the training phase, the algorithm can determine the class membership of new data elements not present in the training dataset.

Supervised learning methods for anomaly detection can be subdivided into four training methods:

1) training from data consisting of only normal events;
2) training from data consisting of only anomalous events;
3) training from data consisting of both normal and anomalous events which are labeled;
4) training from data consisting of multiple classes of labeled events.

Methods 1 and 2 are single-class classification formulations, method 3 is a two-class classification problem, and method 4 is multiclass. Note that methods 3 and 4 here have specified that labels are required to be assigned to the training data. This is to distinguish from unsupervised learning methods, which do not require labels to be affixed to the training data. This comment is also made to distinguish from methods 1 and 2, in which all training data are drawn from a single class, and thus, labeling is trivial.

1) Normal Events are Learned: The most common approach to anomaly detection in automated surveillance within supervised learning is to train an algorithm on normal events, and then classify as anomalous all those events which fall outside of the learned class. This approach has the benefit of not requiring any training data from anomalous events, which are often poorly represented within the data available for training. However, this approach may also suffer from a high rate of false positives, since any event not sufficiently represented in the training data will be detected as anomalous.

Some methods of learning from normal training data involve only the learning of a threshold which defines normality. The learning of a threshold may alleviate the high false positive rate by generalizing the model to a single parameter. However, this approach fails to capture the intricacies that may exist in distinguishing between normal and anomalous events. Jodoin *et al.* [51] record the maximum amount of activity occurring in various locations during the normal-only training. Afterward, any event which causes the amount of activity to exceed the threshold limit for that area is detected as anomalous. A similar method by Zweng and Kampel [32] makes use of a *hitmap* which records the pixel locations of activity in the foreground during the training. The hitmap is used to generate an *unexpectedness map* recording pixel locations where motion did not occur during training. The intensity of motion in the unexpectedness map is used to trigger anomaly detection when activity occurs in an unexpected zone. Dong *et al.* [57] also use a supervised learning method to learn a threshold for detection of anomalies. In this study, a single score is generated which takes into account motion speed, orientation, duration, and shape. A normal range for the score is learned from training data, and any future observation violating the normal range is detected as anomalous.

A more complex approach than the learning of a threshold involves the learning of a multidimensional model of normal events within the feature space. Dong *et al.* [65] introduce the *directional motion behavior map* as a descriptor of objects' normal motion properties. Benezeth *et al.* [87] create a cooccurrence matrix indicating when and where pixels have active motion labels. The *normal* cooccurrence matrix is learned by a Markov random field model; then, violations of the model are detected as anomalies. Several studies [41], [44], [88] involve learning the spatiotemporal relationships between objects in normal training samples. A study by Yao *et al.* [88] involves learning normal spatiotemporal relationships between objects in terms of the objects' sources, sinks, and tracks between the sources and sinks. If the probability of an unseen track is low given the learned model, it is recognized as an anomaly. Two studies [23], [89] apply a learned normal model to generate an anomalous model. In the unique study by Yin *et al.* [23], a normal model is learned with a one-class support vector machine (SVM); then, the normal model is used to generate anomalous models with an unsupervised kernel nonlinear regression method.

In an attempt to compensate for dependence on context, some studies [73], [78] apply fuzzy logic to learn rules which define the normal class. All events which violate these rules are then detected as anomalous. Stoecklin [83] approaches this challenge in anomaly detection in a computer network by learning both baseline behavior as well as "clouds of natural behavior" which extend the basic knowledge of normality and reduce the incidence of false alarms.

2) Multiple Classes of Events are Learned: In the approach in which multiple classes of events are learned, a behavior classification operation is performed prior to the anomaly detection operation. Anomaly detection is determined using a set of rules about the initial classification. This multiclass approach has the drawback that only the learned events can be recognized reliably; if the set of learned events do not span the domain of all possible scene events, then an unexpected event may be classified incorrectly. Since real scenes rarely consist of only a small set of defined events, this approach may not generalize well outside of a scripted environment.

Chen *et al.* [81] present a study using hidden Markov models (HMMs) to learn multiple behaviors of swimmers: backstroke, butterfly, freestyle, breaststroke, rope grasping, and struggling in the waterway. The last two behaviors are specified as anomalous behaviors to be detected. Similar methods are used by Lao *et al.* [79] to identify robberies and by Foroughi *et al.* [15] to detect falls in the elderly. In Lao's study, several human postures are learned: pointing, squatting, lying down, etc. Specified rules about the temporal relationships between the occurrences of these behaviors define the anomaly.

A unique approach by Andersson *et al.* [27] involves defining seven levels of anomalousness of a crowd scene. Each level is defined by the amount of activity (extracted by optical flow), falling or lying down of individuals, and other indicators of potential problems. Each level is learned and detected independently.

3) Normal and Anomalous Events are Learned: The case where both normal and anomalous events are learned is a two-class approach in which the training data consist of both normal and anomalous labeled examples. This approach tends to work well in the case where anomalous events are both well defined and well represented within the training data. The potential success of this approach is dependent upon previous assumptions about the definition of anomalousness: In the case in which anomalous events are defined as all rare or different events, collecting sufficient training data for anomalous events can be difficult. For this reason, this approach is most commonly applied in the case where anomalous events are defined *a priori* as specific meaningful events.

Albusac *et al.* [69] and Adhiya *et al.* [84] both take the approach of modeling normal events as extensively as possible, and also modeling the most common (or most expected) anomalous events. Any behavior that does not fit within the normal events is labeled as *suspicious*, while any that fits within the anomalous behavior models is detected as anomalous.

Other methods attempt to extensively model both the normal and the anomalous events. Clavel *et al.* [49] and Wu *et al.* [77] both extract features from sound samples labeled as either normal or anomalous. The study by Clavel *et al.* [49] is an extensive study on this topic, considering pitch, intensity, duration, jitter, shimmer, and Mel Frequency Cepstral Coefficients as features to

identify fear characteristics. A Gaussian mixture model (GMM) is employed to model the training data labeled as belonging to either the *fear* class or the *neutral* class.

Other studies make use of this approach in order to detect only specific anomalous behaviors. Zhou and Wu [75] apply a self-organizing map (SOM) to model normal behavior and shoplifting behavior in supermarkets. In this case, the only anomaly to be detected is the put-goods-in-pocket behavior. Mishra *et al.* [20] use a neural network trained with backpropagation to distinguish border intrusion behavior from false positives caused by noise events in the environment. Qing-Wei *et al.* [90] apply this two-class supervised learning approach to traffic accident forecasting. Others [76], [78], [91] identify simple postures and motions (walking, sitting, jogging) as normal, and more complex postures and motions (falling, crouching, jumping) as anomalous. Many instances of both classes are collected as training data, and methods such as HMMs, GMMs, and SVMs are utilized for pattern classification.

In some studies, rules for normal or anomalous classification are learned using fuzzy methods. The method presented by Albusac *et al.* [66] begins by extracting object type, speed, and position from a video sample. These features are then presented to a fuzzy learning algorithm along with normal and anomalous labels to create a set of rules relating type, speed, and position of objects to the normal or anomalous classification. A similar method is used by Safara and Eftekhari-Moghadam [71], using only trajectories of traffic as features.

4) Anomalous Events are Learned: Representing the least common approach, this method is avoided due to the high risk of missed detections for anomalous behaviors which do not fit the pattern learned. However, this approach also is the least likely to produce false positives.

The method of learning only anomalous events is used by Guo and Miao [80] to recognize anomalous behaviors in the elderly at home. In this study, HMMs are used to build models for specific anomalous behaviors, such as leaning or lying down for a longer duration than expected. Wang *et al.* [25] train HMMs to recognize the silhouettes of individuals performing anomalous behavior in an office setting. Zhang *et al.* [67] also use the anomalous event learning approach for a system of retrieving anomalous incidents recorded in surveillance videos. In this study, a user presents an example of an anomalous behavior to be retrieved from records. Features from this example are extracted and used as a single training data point. Possible matches are returned to the user, who labels the correct matches. Features of these matches are then utilized as additional training data to improve the retrieval.

B. Unsupervised Learning

Unsupervised learning methods are those in which both normal and anomalous training data are presented without labels. In this case, the assumption is made that normal events are those which occur frequently and anomalous events those which occur rarely. These methods have the benefit of not requiring labeling of training data, but suffer from the drawbacks associated with the assumption that all rare events are anomalous (see Section III). However, Xiang and Gong [37] note that unsupervised methods may perform better than supervised methods, since human labelers have a tendency to fill in unseen, assumed events and so mislabel training data.

Unsupervised learning methods for anomaly detection typically take a clustering approach, in which unlabeled training data are grouped together by an algorithm. In some studies, events in the training dataset are clustered; then, anomalies are identified by the distance of an unseen data point from the nearest cluster. This method requires the additional assumption that normal and anomalous events are well separated in the feature space, and that the training data consist of mostly normal events. Several studies [28], [29] identify an event as anomalous if the distance between the unseen point and the nearest cluster exceeds a specified threshold. The Battacharyya distance is a common measure for this calculation [29], [62], [70]. The threshold for anomaly detection is sometimes not directly a distance measure. Goshorn *et al.* [92] consider the number of changes required to cause the unseen data point to fit into a trained cluster. Each change is assigned a cost; if the total cost exceeds a specified threshold, then the event is identified as anomalous. Other studies [22], [37], [50], [56], [93], [94] use clustering methods to compute the probability of the unseen event. If the probability is below a threshold, then the event is anomalous. Common methods of computing this probability include the log likelihood [37] and the likelihood ratio test [48].

C. A Priori Modeling

A priori modeling methods do not require training data, either labeled or unlabeled. Instead, these methods apply external knowledge of the problem domain to create models or rules for the normal and anomalous classes. The success of this approach relies heavily upon the accuracy and applicability of the external knowledge to the given target and scene. This approach also has the drawback of being insensitive to any changes in the appearance of anomalies over time; the model is typically assumed to be static. The *a priori* method of anomaly detection can be decomposed into three types of methods:

1) a set of rules pertaining to a feature or feature set is specified;
2) a threshold of a feature is defined;
3) a model constructed using methods unrelated to the surveillance task is applied to define normal behavior.

1) Application of Simple Rules About a Feature: Constructing a set of rules for anomalousness requires the application of specific assumptions about the nature of the anomaly. Hsieh and Hsu [61] consider the trajectories of individual targets in the scene. They apply the assumption that a normal human trajectory is horizontal, and therefore, any trajectory that is vertical is assumed to represent the anomalies of climbing or falling. Vallejo *et al.* [68] make use of a camera at a crosswalk with traffic stoplights. Rules such as *cars must not be moving when the light is red* are applied to detect anomalous events. Others [58]–[60] compare trajectories of people against previously defined restricted areas. If the trajectory crosses the defined area, the behavior is detected as anomalous.

The simple rules approach can also be applied in cases of sensor types other than video or audio. Chae *et al.* [86] make use of an RFID system which reports the tag wearer's identity, location, and a time stamp. Specified rules about certain individuals at specified locations and times are applied to detect anomalies.

2) Thresholding: Some studies apply a threshold defined *a priori* for anomaly detection. Diamantopoulos and Spann [64] identify anomalies in the behavior of cars entering a controlled-access parking lot by thresholding the distance between the cars. If a following car is too close, then it must have violated gate access protocol. The thresholding approach can also be applied to detect temporal anomalies. Two of these studies [82], [85] address anomalousness in the behavior of the elderly at home. If there is a long time lapse between sensor readings, then an anomalous situation is assumed to have occurred. Lee *et al.* [74] apply the assumption that anomalous human behavior exhibits the quality of randomness. They compute the entropy of observed human limbs and centroid. If the entropy exceeds a specified threshold, then the behavior is detected as anomalous. Zhong *et al.* [31] make use of the Markov random field energy for a similar purpose. In this study, a sensor signal is analyzed by wavelet decomposition. If the wavelet decomposition output exceeds a specified threshold, then the behavior is detected as anomalous. A similar approach by Cao *et al.* [30] involves computing the kinetic energy of a crowd. Two types of anomalies are identified with this method: static anomalies, in which the kinetic energy exceeds the specified threshold, and dynamic anomalies, identified when the kinetic energy of the crowd suddenly changes by more than an allowed amount of change.

3) Use of an A Priori Model: Use of previously constructed models has been applied to crowd, traffic, and individual anomaly detection. Mehran *et al.* [33] apply the social force model as a description of normal crowd behavior. If a crowd is observed to behave contrary to this model, the event is detected as anomalous. Sultani and Choi [95] use an intelligent driver model to compare with the current observed behavior of a vehicle. Observations violating the model are detected as anomalous. Oliver *et al.* [96] apply models of interpersonal human interaction to classify new observations of interpersonal interactions.

VI. Modeling and Classification Algorithms for Anomaly Detection

The range of modeling and classification algorithms applied to anomaly detection in automated surveillance is extensive, varying in several fundamental differences of construct. Static versus dynamic, parametric versus nonparametric, and linear versus nonlinear are all valid and significant classification schemes for modeling and classification algorithms. In this section, the modeling and classification algorithms will be broadly grouped together with other methods that are similar in construct. Where applicable, effort is made to also highlight the significant differences between methods along other relevant lines of separation.

There is a wealth of information in the literature focusing on the problem of creating models of observed behavior, without the extension of applying the model to anomaly detection. Research of this nature is not included here. Rather, this section gives an overview of the modeling and classification algorithms specifically applied to anomaly detection in automated surveillance.

A. Dynamic Bayesian Networks

The HMM is the method for behavior modeling and anomaly detection most commonly applied by studies in this review. The popularity of the HMM for behavior modeling is likely due to the nature of temporal dependence inherent in this method. In contrast with many of the other methods applied to anomaly detection, the HMM is able to take into account the inherently dynamic nature of behavior.

The HMM is a structure of nodes connected by transition links representing a time series of states. Each node represents a state which is not directly observable, thus giving the term *hidden* to the method. Rather, at each state, an observation is made. The observation corresponds to a set of probabilities of states. An HMM is defined by matrices encoding the possible states and the probabilities of observations, known as the state transition matrix and the emission matrix, respectively.

The studies in this review which apply HMM modeling methods differ primarily in the states assigned to HMM nodes, the meaning of observations, and the form of the model. Nodes may represent objects' positions [88], velocities, accelerations [26], or postures [79], crowd behavior [27] or local behaviors like standing, leaning, walking, etc. [80]. Observations may be an activity level of a crowd or crowd size [27]. A single model may be a global behavior like sitting down or falling down [80], a general activity class [25] or general normal or abnormal behavior [26], [27].

The state transition matrix and the emission matrix are determined through a training process, commonly accomplished by the Baum–Welch training algorithm [50], [81]. After training of the model, a new set of observations is evaluated, often with the Viterbi algorithm [50], to determine the maximum log-likelihood of a model corresponding to the observation sequence [50], [88]. Although some studies apply a simple first-order left-to-right model [26], [50], many variations have also been applied to anomaly detection in automated surveillance.

Two studies in this review applied the coupled HMM [67], [96]. This model allows for more than one state during a single time interval, enabling more than one HMM to interact. The interaction allows multiple models to influence the behavior of each other. These parallel chains allow for state transitions between the models. For example, a study by Oliver *et al.* [96] uses one HMM for each of two humans in the surveillance video, allowing for the model to account for interpersonal interactions.

The hierarchical context HMM [14] and a cascade of dynamic Bayesian networks [41] are two variations on the HMM in which several models are arranged in a cascading structure, rather than a parallel structure as in the coupled HMMs. The cascading structure has the benefit of allowing for use of

several different models which each has greater sensitivity to a specific type of anomaly. Chung and Liu [14] make use of this method to perform anomaly detection using three reasoning components: spatial context reasoning, behavior reasoning, and temporal context reasoning. Loy *et al.* [41] use a first-order hierarchical HMM to model simple behaviors, and then a multiobservation hidden Markov models (MOHMM) in the second stage to model complex behaviors from the simple ones.

An HMM has a tendency to exhibit high performance on the training data but poorer performance on novel data [12]. This effect is due to the large number of parameters required to define the model encoded within the state transition and emission matrices. The large number of parameters causes the HMM to be vulnerable to overfitting. To compensate for this problem, a number of studies [12], [37], [41], [48], [56] have investigated the use of MOHMM. In the MOHMM, the number of model parameters is reduced by the assumption that each observed feature is independent [12].

A key challenge with the use of HMMs for anomaly detection in automated surveillance is the necessity of determining the appropriate number of hidden states for the model. A key study by Pruteanu-Malinici and Carin [43] addressed this issue by the use of the iHMM. The iHMM is built upon the concept of Dirichlet Processes: A hierarchical Dirichlet process is defined in which each row of the transition and emission matrices of an HMM is modeled as a Dirichlet process [97]. With this method, an infinite number of states are allowed for by use of a small number of *hyperparameters*, each defining many individual parameters collectively. Rather than requiring the model to learn an entire state transition matrix and emission matrix, only the hyperparameters are learned.

In an HMM, there is a nonzero probability that a state will transition back to itself, rather than to a new state. As a result, the state duration follows a geometric distribution if time is discrete, or an exponential distribution if time is considered to be continuous [98]. In order to allow for other distributions of state duration, Duong *et al.* [16] investigate the use of the hidden semi-Markov model. The hidden semi-Markov model allows a variable duration for each state [98]. Duong *et al.* [16] apply a two-layer hidden semi-Markov model in which the top layer is a Markov sequence of switching variables, and the bottom layer is a sequence of concatenated hidden semi-Markov models whose parameters are determined by the switching variable at the top. The bottom layer models low-level activities like spending time at a location or moving between locations, while states in the top layer represent high-level activities like making breakfast or washing dishes.

Luhr *et al.* [17] also apply the hidden semi-Markov model, using the alternate name explicit state duration HMM to detect variations in duration spent at an activity as anomalous. The standard HMM formulation exhibits a property known as *dynamic time warping* which allows different events to be recognized as the same event if they are different only in time scale. Thus, for example, a HMM trained on data of an individual walking slowly will also recognize instances of an individual walking quickly. This integrated accounting for dynamic time warping is generally considered to be an advantage for automated behavior classification. However, Luhr *et al.* [17] consider the dynamic time warping property of HMMs to be a drawback, since significant variations in time spent at an activity should be detected as anomalous. Their use of the explicit state duration HMM successfully compensates for dynamic time warping.

B. Bayesian Topic Models

Latent Dirichlet allocation [99] is a type of Naïve Bayesian modeling method which was originally developed for document modeling and text classification. This method defines several concepts having analogy with the text domain, giving the modeling elements literary terms. A *corpus* is defined as a group of *documents* composed of *words* in which the order of the words is not considered, resulting in a *bag-of-words* approach. Word counts are clustered into topics which are then assigned to documents. Probabilities of topic assignments can then be assigned to similar, unseen documents. These topic models have the advantage of the ability to account for simultaneous interaction of multiple objects or activities [38].

In the study by Sultani and Choi [95], vocabulary is learned by K-means clustering of word counts, while other studies assign a feature to represent words. Varadarajan and Odobez [70] apply probabilistic latent semantic analysis, regarding the location, motion, and size of objects in video as words. Similar to the studies by Li *et al.* [38] and Sultani and Choi [95], video clips are considered as documents, and the entire video is a corpus.

Li *et al.* [38] note that unsupervised topic models are only sensitive to rare behaviors which are visually distinct from the majority of observed behaviors. This study introduces the multiclass delta latent Dirichlet allocation method, which exploits weak supervision to improve the recognition of rare behaviors that are less visually distinguishable from normal behaviors. A video scene is segmented into cells with the average vector of optical flow representing words. This is a multiclass approach: The topics include normal behavior and multiple types of abnormal behavior. In this method, the relationship between learned rare behavior topics and the learned normal behavior topic is utilized to generate new topics representing unseen but plausible rare behaviors.

Relative to the dynamic Bayesian networks, Bayesian topic models have the drawback of ignoring the inherent time dependence of behavior. Hospedales *et al.* [39] address this drawback by introducing the Markov clustering topic model which blends concepts of the dynamic Bayesian networks and the Bayesian topic models. In this study, visual events are considered to be words, simple actions (cooccurring events) are topics, and complex behaviors (cooccurring actions) are document categories, while video clips are documents. The crossover with dynamic Bayesian networks is accomplished by introducing a time-dependent series of behaviors which affects a time series of actions and, in turn, visual words. Temporal modeling is introduced by the Markov components of this approach, enabling the correlation of different behaviors over time, while the Bayesian topic models have the advantage of robustness to noise. Li *et al.* [53] also recognize the importance of

temporal order in their application of latent Dirichlet allocation and introduce an element of temporal order sensitivity to the method.

C. Artificial Neural Networks

The artificial neural network (ANN) is a mathematical formulation for machine learning imitating the biological nervous system. In ANNs, individual nodes known as *neurons* are linked in a network consisting of one or more *layers* of neurons, each fully connected to the neurons before and after it. The first layer receives one or more inputs, and the last layer produces one or more outputs. Layers between the input and output layers are known as *hidden* layers. Each neuron computes a weighted sum of its inputs; its output is a nonlinear function (called an *activation function*) of the weighted sum. The weights of each neuron are learned through a training process, allowing the network to replicate an arbitrary function mapping inputs to outputs. A simple construct of neurons that is able to perform linearly separable classification is known as a *perceptron*.

Lee *et al.* [74] use a multilayer perceptron neural network with a single hidden layer to model human motions. In this study, the inputs to the network are the angles of human limbs and centroid, and the output is a behavior class. A different application of ANNs that also makes use of a single hidden layer is presented in a study by Mishra *et al.* [20]. In this study, an ANN is used to detect border intrusions. Here, the network input is a set of samples from a number of wireless sensors, and the output is a normal/anomalous determination indicating whether an intrusion has occurred.

Other types of ANNs besides perceptrons have also been applied to anomaly detection in automated surveillance. The SOM is a type of ANN that consists of a multidimensional lattice of neurons that utilizes competitive unsupervised learning. The unsupervised learning process causes only a single neuron in the lattice (termed the *winning neuron*) to fire in response to a single input. The result is a mapping of similar input patterns to similar locations of neuron firing, thus reducing the dimensionality between the map input and output.

Feng *et al.* [28] use a SOM consisting of an 8×8 lattice of neurons to identify anomalies in crowd behavior. After training, many samples of crowd video data are presented to the map. For each sample, the distance between the location of the winning neuron and the average location of all winning neurons is calculated. If the distance exceeds the threshold, an anomaly is detected. Zhou and Wu [75] also use a SOM, but with only two output neurons: one for normal behavior, and one for anomalous.

Similar to the SOM concept is the method of altruistic vector quantization (AVQ). Vector quantization is a method of data compression in which input vectors are mapped to a discrete set or *codebook* of representative vectors known as *prototypes* [100]. In the AVQ method, the codebook is learned using a competitive neural network method similar to a SOM. In a study by Mecocci *et al.* [44], a set of normal training vectors representing positions and trajectories of objects in a scene are used to learn a codebook with AVQ. After training, the distance between an unseen vector and the nearest prototype is computed. If the distance exceeds a threshold, the event is classified as anomalous.

Loosely related to the ANN perceptron is the method of SVM. SVMs are a classification method for finding a linear separating hyperplane between two classes of data. Separability between classes is achieved by first transforming the data to a higher dimensional space via use of a kernel. Wu *et al.* [77] and Zhang and Liu [78] use a two-class SVM classifier based on training data from the normal and abnormal classes. Yin *et al.* [23] utilize a one-class SVM classifier to model only the normal data. Here, SVM is used to find a sphere encompassing most normal training data. However, Yin *et al.* note that sparsity of training data cause poor results in the one-class SVM approach.

In an attempt to identify human trips and falls, Foroughi *et al.* [15] adopt a multiclass SVM. They generalize the traditional two-class SVM using a one-against-all method and a one-against-one method. The one-against-all method involves training a set K of one-class SVMs, where $|K|$ is the number of classes to be recognized. In the one-against-one method, each SVM is two-class and thus only two models are trained.

D. Clustering

The GMM is a well-defined method of unsupervised clustering in which each of a known number of clusters is assumed to be drawn from a Gaussian distribution. The mean and standard deviation of each cluster is discovered via an expectation maximization process. The result is a model of the training data consisting of a number of Gaussian distributions each defined by a mean and standard deviation. GMMs are frequently used in image processing applications. However, a few studies reviewed here [34], [45], [49] have also used this method in anomaly detection for automated surveillance.

Tziakos *et al.* [34] and Bouttefroy *et al.* [45] utilize GMMs as a one-class classifier. In this formulation, normal data samples in feature space are used as training data for a GMM. When a new data sample is presented, the probability of the sample occurring according to all learned distributions is calculated. If the probability is below a specified threshold, the event is declared to be anomalous.

Clavel *et al.* [49] use GMMs as a multiclass classifier with both normal and anomalous training data available. Here, a GMM is learned for each of four classes, with two classes representing normal behavior types and two representing anomalous behavior types. An unseen data sample is assigned to the class with the highest probability of generating that sample. A second experiment in the same study introduces additional classes, each represented as an additional GMM, to indicate whether the anomaly has occurred in the past, or is predicted to occur in the future.

E. Decision Trees

A decision tree consists of a structure of cascaded nodes. Each node is connected to a single parent node in the previous layer and a number of children nodes in the succeeding layer. Each node represents a decision, and each connection represents a state and a probability of entering that state.

An N-ary tree classifier is a decision tree in which each node has an equal number of children. Duque *et al.* [60] use an N-ary tree classifier to predict anomalous behaviors. Each layer of the tree represents the state of an object in feature space (position, velocity, object type, and anomaly flag) at an instant of time. The probabilities of the tree connections are learned in a supervised manner from both normal and abnormal training samples. After training, a previously unseen behavior is located on the tree and its probability of entering each connecting state is calculated. If there is a high probability of entering an anomalous state, then the behavior is flagged as anomalous.

The dynamic oriented graph approach to anomaly prediction [58] is similar to the N-ary trees classifier, except that nodes are added to the tree only as needed when a new data sample does not fit any path already present in the model. Thus, this method requires significantly less training time relative to the N-ary trees method.

F. Fuzzy Reasoning

Fuzzy reasoning is a means to account for uncertainty in classification by allowing for a continuous transition between overlapping classes. This classification method provides a more intuitive reasoning result by imitating the ways that human beings reason. In anomaly detection, fuzzy reasoning may be used to manage uncertainty in the anomalous versus normal determination, or in assigning membership to a subclass which determines anomalous behavior.

Albusac *et al.* [66] apply Castro's Algorithm for learning fuzzy rules. In this method, each example of a training set is first converted into a specific classification rule. After all training examples have been converted into specific rules, the rule set is generalized to cover a wide range of possible situations through the method of amplification. This method has a tendency to overgeneralize, or associate rules with situations to which they do not apply. This study introduces a restriction on the amplification process to reduce the tendency to overgeneralize. In another study by Albusac *et al.* [69], fuzzy rules are used to deal with uncertainty in observing an outdoor scene, both in low-level object identification and in classifying a scene as normal or anomalous. Low-level fuzzy sets include temporal relationships such as *before*, *after*, and *during* as well as object type, such as *vehicle* or *pedestrian*. Membership in the low-level fuzzy sets is then used to assign membership to the high-level set, determining if the path of the object is normal or anomalous.

Safara and Eftekhari-Moghadam [71] apply fuzzy rules to convert features such as trajectories and domain information into a human-readable descriptive sentence about the scene. This study also assigns a normal/anomalous decision to a scene based on the scene's fuzzy membership in rules describing violation of traffic laws.

Fuzzy associative memory (FAM) is a type of fuzzy neural network that receives an input and assigns a degree of belonging to a set of rules. Zhicheng and Jun [72] consider angles of human limbs as input to an FAM with three rules defining abnormal movement types. The FAM assigns a degree of membership to each rule and then produces an output determination of anomalous or not based on a specified threshold.

VII. Discussion and Conclusion

We have presented a review examining recent research in anomaly detection in automated surveillance across key aspects of the problem domain, approach, and method. This section will give conclusions from the work and discussion of the range of approaches that have been taken in the literature.

Table II presents a list of the studies in this review according to various aspects of their respective approaches to anomaly detection in automated surveillance.

Within the topic of target, Table II illustrates that the majority of work has been done in the area of anomaly detection in individuals, likely because this target has a wide range of common applications. Table II also reveals the commonality of studies developing single-target methods and the lack of multiple-target studies. This may be due to the difficulty of identifying a single feature extraction method which applies equally well to different target types. However, the lack of methods which can be applied to a wide range of targets may contribute to the lack of widespread use of automated anomaly detection solutions, since each solution has only a specific use and may fail when exposed to the variety of targets and behavior common in realistic scenarios. This lack of cross-target research is most significant in the area of individual/crowd surveillance. This combination of targets is both very common as a realistic scenario and also very difficult due to the difference in feature extraction method applied to each. Further study is needed addressing the applicability of methods to a wider range of surveillance targets in varying environments.

It is evident from Table II that sensor type is dominated by the visible light camera. This popularity is likely due to the ubiquity of surveillance cameras in the environment and the prevalence of such data. However, the relative scarcity of research which considers other sensor types indicates the potential for discovery in this area. Those few studies which have considered other sensor types have shown promising initial results, particularly in situations where visible light cameras would be impractical, such as surveillance of wide areas.

Most of the research in anomaly detection for automated surveillance has considered fields of view which cover the range from approximately the area of a room (10 m) to the area of a medium-sized parking lot (100 m). There is a lack of research considering either the very small or very large fields of view, as might be seen in a satellite image or in a portrait of a human face. A similar trend is seen in the commonality of studies addressing higher resolution images. Although the technology trend is toward wider availability of higher resolution images, data storage and transmission limitations may warrant greater attention given to lower resolution images. Additionally, certain problem domains such as anomaly detection in time-lapse satellite imagery are restricted to low-resolution, wide field-of-view images not currently addressed by the anomaly detection in the automated surveillance literature.

TABLE II
STUDIES REVIEWED ACCORDING TO APPROACHES AND LEARNING METHOD

Process Element	Studies
Target	
Individuals	[7, 12-25, 35-37, 39-41, 43, 45, 46, 48-51, 57-64, 66, 67, 69, 72-82, 85-88, 91-93, 96]
Crowds	[27-34, 46, 56]
Cars	[22, 26, 35, 38-41, 43-45, 51, 58, 60, 64-66, 68, 70, 71, 87, 88, 94-96]
Other/Unspecified	[21, 83, 84]
Sensor	
Visible Light	[7, 12, 14, 15, 17, 19, 22, 25, 26, 28-46, 48, 50, 51, 56-75, 78-81, 87-89, 91-96]
Multiple simple sensors	[20, 23, 82, 85, 86]
Audio	[13, 49, 77]
Infrared	[27]
Feature Extraction	
Low-Level	[19, 27-36, 38, 39, 43, 46, 51, 56, 57, 87, 94, 95]
Object-Based	[7, 12, 14, 15, 17, 22, 25, 26, 37, 40, 41, 44, 50, 57-81, 88, 91-93, 96]
Field of View [approx. m^2]	
1	[15, 50, 91]
10	[12, 14, 17, 19, 25, 29-31, 34, 35, 37, 39, 41, 46, 48, 56, 57, 61, 67, 68, 75, 76, 79, 81]
100	[7, 12, 22, 28, 32, 33, 36, 38, 39, 41, 43, 58, 60, 62, 64-66, 69-71, 87, 88, 94-96]
1000	[41, 51, 58, 68]
Resolution	
1	[40, 51]
2	[7, 22, 36, 38, 60, 64, 65, 71, 88, 94, 96]
3	[14, 28, 31-33, 41, 43, 57, 62, 66, 67, 69, 70, 87, 95]
4	[17, 19, 25, 29, 30, 34, 35, 37, 48, 56, 61, 76, 79, 81, 91]
5	[46, 50, 75]
Anomaly Detection Method	
Supervised	
Normal events learned	[12, 16, 23, 26, 32, 34, 35, 41, 43, 44, 51, 57, 60, 65, 72, 73, 83, 87-89, 96]
Multiple classes of events learned	[15, 17, 27, 57, 63, 66, 71, 79, 81]
Anomalous events learned	[25, 67, 80, 86]
Normal and anomalous events learned	[20, 49, 69, 75-78, 84, 90, 91]
Unsupervised	[12, 14, 18, 21, 22, 25, 28, 29, 36-39, 42, 48, 50, 56, 62, 70, 92-94]
A-Priori	
Simple specified rules about a feature	[46, 58, 60, 61, 64, 68, 82, 85, 86]
Specified feature thresholded	[13, 21, 29-31, 74]
Use of a-priori model	[7, 33, 95, 96]

VIII. FUTURE WORK

The need for future work in this field is dominated by the lack of a sound testing methodology to compare different algorithms and approaches. In order to evaluate the performance of any proposed method for anomaly detection in automated surveillance, a set of test data and a testing methodology must be selected. Each study in this review has applied unique datasets for evaluation of each method, thus precluding the possibility of equally comparing the results of each study. Many of the datasets used are unique to the organization performing the research and are not publicly available. For a list and discussion of a few video datasets that are publicly available, see a previous review on view-invariant human motion analysis by Ji and Liu [101]. There is a need for standardization or common benchmarking approaches against which differing methods can be evaluated. The problem of lack of standardized and meaningful datasets for anomaly detection has been recognized in fields other than automated surveillance, such as in computer network intrusion detection [102]. A common and widely accessible repository of the standardized and meaningful datasets is also needed for sound comparison of different methods. The easy access to standardized datasets provided by a common repository would more easily allow independent evaluation of the various studies.

A significant impediment to achieving standardization in evaluation methodology is the wide range of targets, sensors, fields of view, and resolutions addressed by different studies. This is further complicated by the wide range of problem definitions and target environments considered by each researcher. In order to determine the current state of the field, it is necessary to be able to evaluate a method with respect to its own goals and definition of good performance.

Until such methodology is agreed upon, the state of the field and the identification of strategic trajectories for future work cannot be based upon the relative success of competing methods. However, development of a single dataset or repository to cover the entire range of the field as summarized in Table II might prove to be infeasible. In this case, a consistent ontology or taxonomy for subdivision of the problem domain could provide the specificity and structure necessary to define testing datasets. Further research is needed in developing such ontology.

Without a sound testing methodology, future work in the area of learning and modeling algorithms for automated surveillance cannot be determined, because the effectiveness of existing methods cannot be reliably evaluated. If a consistent testing methodology, ontology, and evaluation datasets are defined, additional areas of work in this field would open. Some additional areas for work in this field involve study of less common domains which have not been sufficiently covered in the literature, such as the use of a variety of sensors, fields of view and resolutions at the outside ranges of the scale, and methods that are applicable to a wide range of different targets.

Following successful experimental validation of different methods, work is needed to address issues arising in the deployment of surveillance systems. Since the studies in the literature typically evaluate methods in only a single environment or in controlled conditions, issues such as the ability to generalize

in new environments, scaling to wide-area distribution, and robustness to unexpected events will need to be addressed before the methods can be applied in real-world scenarios.

ACKNOWLEDGMENT

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

REFERENCES

[1] E. Saykol, U. Gudukbay, and O. Ulusoy, "Scenario-based query processing for video-surveillance archives," *Eng. Appl. Artif. Intell.*, vol. 23, no. 3, pp. 331–345, 2010.

[2] D. Brezeale and D. J. Cook, "Automatic video classification: A survey of the literature," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 38, no. 3, pp. 416–430, May 2008.

[3] 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, Appl. Rev.*, vol. 39, no. 5, pp. 489–504, Sep. 2009.

[4] H. Buxton, "Learning and understanding dynamic scene activity: A review," *Image Vision Comput.*, vol. 21, no. 1, pp. 125–136, 2003.

[5] 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.

[6] T. Ko, "A survey on behavior analysis in video surveillance for homeland security applications," in *Proc. Appl. Imagery Pattern Recognit. Workshop*, Oct. 15–17, 2008, pp. 1–8.

[7] H. M. Dee and D. C. Hogg, "On the feasibility of using a cognitive model to filter surveillance data," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveillance*, Sep. 15–16, 2005, pp. 34–39.

[8] N. Haering, P. L. Venetianer, and A. Lipton, "The evolution of video surveillance: An overview," *Mach. Vision Appl.*, vol. 19, no. 5-6, pp. 279–290, 2008.

[9] T. D. Raty, "Survey on contemporary remote surveillance systems for public safety," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 5, pp. 493–515, Sep. 2010.

[10] F. Angiulli, S. Basta, and C. Pizzuti, "Distance-based detection and prediction of outliers," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 2, pp. 145–160, Feb. 2006.

[11] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, 2009.

[12] 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.

[13] S. Moncrieff, S. Venkatesh, G. West, and S. Greenhill, "Incorporating contextual audio for an actively anxious smart home," in *Proc. Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process. Conf.*, Dec. 5–8, 2005, pp. 373–378.

[14] P. Chung and C. Liu, "A daily behavior enabled hidden Markov model for human behavior understanding," *Pattern Recognit.*, vol. 41, no. 5, pp. 1589–1597, 2008.

[15] 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, Graph. Image Process.*, Dec. 16–19, 2008, pp. 413–420.

[16] 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 Recognit.*, Jun. 20–25, 2005, vol. 1, pp. 838–845.

[17] 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 Pac. Rim Int. Conf. Artif. Intell.*, Aug. 9–13, 2004, pp. 983–984.

[18] S. W. Lee, Y. S. Kim, and Z. Bien, "A nonsupervised learning framework of human behavior patterns based on sequential actions," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 4, pp. 479–492, Apr. 2010.

[19] S. Gong and T. Xiang, "Scene event recognition without tracking," *Zidonghua Xuebao/Acta Automatica Sinica*, vol. 29, no. 3, pp. 321–331, 2003.

[20] A. Mishra, K. Sudan, and H. Soliman, "Detecting border intrusion using wireless sensor network and artificial neural network," in *Proc. 6th IEEE Int. Conf. Distrib. Comput. Sensor Syst. Workshops*, Jun. 2010, pp. 1–6.

[21] H. Li and Z. Han, "Catching attacker(s) for collaborative spectrum sensing in cognitive radio systems: An abnormality detection approach," in *Proc. IEEE Symp. New Frontiers Dyn. Spectr.*, Apr. 6–9, 2010, pp. 1–12.

[22] R. J. Morris and D. C. Hogg, "Statistical models of object interaction," *Int. J. Comput. Vision*, vol. 37, no. 2, pp. 209–215, 2000.

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

[24] F. Ziliani, S. Velastin, F. Porikli, L. Marcenaro, T. Kelliher, A. Cavallaro, and P. Bruneaut, "Performance evaluation of event detection solutions: The CREDS experience," in *Proc. IEEE Conf. Adv. Video Signal Based Surveillance*, Sep. 15–16, 2005, pp. 201–206.

[25] Y. Wang, K. Huang, and T. Tan, "Abnormal activity recognition in office based on R transform," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 16–Oct. 19, 2007, vol. 1, pp. 341–344.

[26] T. Hayashi and K. Yamada, "Predicting unusual right-turn driving behavior at intersection," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 3–5, 2009, pp. 869–874.

[27] M. Andersson, J. Rydell, and J. Ahlberg, "Estimation of crowd behavior using sensor networks and sensor fusion," in *Proc. 12th Int. Conf. Inf. Fusion*, Jul. 6–9, 2009, pp. 396–403.

[28] J. Feng, C. Zhang, and P. Hao, "Online learning with self-organizing maps for anomaly detection in crowd scenes," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 23–26, 2010, pp. 3599–3602.

[29] M. H. Sharif, S. Uyaver, and C. Djeraba, "Crowd behavior surveillance using bhattacharyya distance metric," in *Proc. CompIMAGE*, May 5–7, 2010, pp. 311–323.

[30] T. Cao, X. Wu, J. Guo, S. Yu, and Y. Xu, "Abnormal crowd motion analysis," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, Dec. 19–23, 2009, pp. 1709–1714.

[31] Z. Zhong, N. Ding, X. Wu, and Y. Xu, "Crowd surveillance using Markov random fields," in *Proc. IEEE Int. Conf. Autom. Logist.*, Sep. 1–3, 2008, pp. 1822–1828.

[32] A. Zweng and M. Kampel, "Unexpected human behavior recognition in image sequences using multiple features," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 23–26, 2010, pp. 368–371.

[33] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 20–25, 2009, pp. 935–942.

[34] I. Tziakos, A. Cavallaro, and L. Xu, "Local abnormality detection in video using subspace learning," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surveillance*, Aug. 29–Sep. 1, 2010, pp. 519–525.

[35] I. Tziakos, A. Cavallaro, and L. Xu, "Event monitoring via local motion abnormality detection in non-linear subspace," *Neurocomputing*, vol. 73, no. 10–12, pp. 1881–1891, 2010.

[36] E. B. Ermis, V. Saligrama, P. Jodoin, and J. Konrad, "Abnormal behavior detection and behavior matching for networked cameras," in *Proc. 2nd ACM/IEEE Int. Conf. Distrib. Smart Cameras*, Sep. 7–11, 2008, pp. 1–10.

[37] T. Xiang and S. Gong, "Video behaviour profiling and abnormality detection without manual labelling," in *Proc. 10th IEEE Int. Conf. Comput. Vision*, Oct. 17–20, 2005, vol. 2, pp. 1238–1245.

[38] J. Li, T. Hospedales, S. Gong, and T. Xiang, "Learning rare behaviours," *Lecture Notes Comput. Sci.*, vol. 6493, pp. 293–307, 2010.

[39] 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*, Sep. 29–Oct. 2, 2009, pp. 1165–1172.

[40] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000.

[41] C. C. Loy, T. Xiang, and S. Gong, "Detecting and discriminating behavioural anomalies," *Pattern Recognit.*, vol. 44, no. 1, pp. 117–132, 2011.

[42] B. B. Orten, A. A. Alatan, and T. Ciloglu, "Event detection in automated surveillance systems," in *Proc. IEEE 14th Signal Process. Commun. Appl.*, Apr. 17–19, 2006, pp. 1–4.

[43] 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.

[44] 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.*, Jul. 29–31, 2003, pp. 187–192.

[45] P. L. M. Bouttefroy, A. Bouzerdoum, S. L. Phung, and A. Beghdadi, "Local estimation of displacement density for abnormal behavior detection," in *Proc. IEEE Workshop Mach. Learn. Signal Process.*, Oct. 16–19, 2008, pp. 386–391.

[46] S. A. Velastin, B. A. Boghossian, B. P. L. Lo, J. Sun, and M. A. Vicencio-Silva, "PRISMATICA: Toward ambient intelligence in public transport environments," *IEEE Trans. Syst., Man Cybern. A, Syst. Humans*, vol. 35, no. 1, pp. 164–182, Jan. 2005.

[47] E. Saykol, M. Bastan, U. Gudukbay, and O. Ulusoy, "Keyframe labeling technique for surveillance event classification," *Opt. Eng.*, vol. 49, pp. 1–12, 2010.

[48] T. Xiang and S. Gong, "Incremental and adaptive abnormal behaviour detection," *Comput. Vision Image Understanding*, vol. 111, no. 1, pp. 59–73, 2008.

[49] C. Clavel, L. Devillers, G. Richard, I. Vasilescu, and T. Ehrette, "Detection and analysis of abnormal situations through fear-type acoustic manifestations," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 15–20, 2007, vol. 4, pp. 21–24.

[50] 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.

[51] P. Jodoin, J. Konrad, and V. Saligrama, "Modeling background activity for behavior subtraction," in *Proc. 2nd ACM/IEEE Int. Conf. Distrib. Smart Cameras*, Sep. 7–11, 2008, pp. 1–10.

[52] N. Anjum and A. Cavallaro, "Trajectory association and fusion across partially overlapping cameras," in *Proc. 6th IEEE Int. Conf. Adv. Video Signal Based Surveillance*, Sep. 2–4, 2009, pp. 201–206.

[53] J. Li, S. Gong, and T. Xiang, "On-the-fly global activity prediction and anomaly detection," in *Proc. IEEE 12th Int. Conf. Comput Vision Workshops*, Sep. 27–Oct. 4, 2009, pp. 1330–1337.

[54] J. Black, T. Ellis, and P. Rosin, "Multi view image surveillance and tracking," in *Proc. Workshop Motion Video Comput.*, Dec. 5–6, 2002, pp. 169–174.

[55] X. Wang, K. Tieu, and E. L. Grimson, "Correspondence-free activity analysis and scene modeling in multiple camera views," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 56–71, Jan. 2010.

[56] E. L. Andrade, S. Blunsden, and R. B. Fisher, "Modelling crowd scenes for event detection," in *Proc. 18th Int. Conf. Pattern Recognit.*, Aug. 20–24, 2006, vol. 1, pp. 175–178.

[57] Q. Dong, Y. Wu, and Z. Hu, "Pointwise motion image (PMI): A novel motion representation and its applications to abnormality detection and behavior recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 3, pp. 407–416, Mar. 2009.

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

[59] M. Shah, O. Javed, and K. Shafique, "Automated visual surveillance in realistic scenarios," *IEEE Multimedia*, vol. 14, no. 1, pp. 30–39, Jan.–Mar. 2007.

[60] D. Duque, H. Santos, and P. Cortez, "The OBSERVER: An intelligent and automated video surveillance system," in *Proc. 3rd Int. Conf. Image Anal. Recognition*, Sep. 18–20, 2006, pp. 898–909.

[61] C. Hsieh and S. Hsu, "A simple and fast surveillance system for human tracking and behavior analysis," in *Proc. 3rd IEEE Int. Conf. Signal-Image Technol. Internet-Based Syst.*, Dec. 16–18, 2007, pp. 812–818.

[62] 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.

[63] T. Jan, M. Piccardi, and H. Gunes, "Suspicious behavior assessment for visual surveillance using neural network classifiers," in *Proc. Int. Conf. Imaging Sci., Syst. Technol.*, Jun. 23–26, 2003, vol. 2, pp. 65–71.

[64] G. Diamantopoulos and M. Spann, "Event detection for intelligent car park video surveillance," *Real Time Imaging*, vol. 11, no. 3, pp. 233–243, 2005.

[65] N. Dong, Z. Jia, J. Shao, Z. Xiong, Z. Li, F. Liu, J. Zhao, and P. Peng, "Traffic abnormality detection through directional motion behavior map," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surveillance*, Aug. 29–Sep. 1, 2010, pp. 80–84.

[66] J. Albusac, J. Castro-Schez, L. Lopez-Lopez, D. Vallejo, and L. Jimenez-Linares, "A supervised learning approach to automate the acquisition of knowledge in surveillance systems," *Signal Process.*, vol. 89, no. 12, pp. 2400–2414, 2009.

[67] C. Zhang, W. Chen, X. Chen, L. Yang, and J. Johnstone, "A multiple instance learning and relevance feedback framework for retrieving abnormal incidents in surveillance videos," *J. Multimedia*, vol. 5, no. 4, pp. 310–321, 2010.

[68] D. Vallejo, J. Albusac, L. Jimenez, C. Gonzalez, and J. Moreno, "A cognitive surveillance system for detecting incorrect traffic behaviors," *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10503–10511, 2009.

[69] J. Albusac, D. Vallejo, L. Jimenez-Linares, J. Castro-Schez, and L. Rodriguez-Benitez, "Intelligent surveillance based on normality analysis to detect abnormal behaviours," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 23, no. 7, pp. 1223–1244, 2009.

[70] 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.

[71] F. Safara and A. Eftekhari-Moghadam, "Knowledge discovery of traffic/people behaviors based on image mining approach," in *Proc. Geom. Model. Imaging-New Trends*, Jul. 5–7, 2006, pp. 255–258.

[72] W. Zhicheng and Z. Jun, "Detecting pedestrian abnormal behavior based on fuzzy associative memory," in *Proc. 4th Int. Conf. Nat. Comput.*, Oct. 18–20, 2008, pp. 143–147.

[73] J. Zhang and Z. Liu, "Abnormal behavior of pedestrian detection based on fuzzy theory," *Moshi Shibie Yu Rengong Zhineng/Pattern Recognit. Artif. Intell.*, vol. 23, no. 3, pp. 421–427, 2010.

[74] C. Lee, K. Lim, and W. Woon, "Statistical and entropy based multi purpose human motion analysis," in *Proc. 2nd Int. Conf. Signal Process. Syst.*, Jul. 5–7, 2010, pp. 734–738.

[75] G. Zhou and Y. Wu, "Anomalous event detection based on self-organizing map for supermarket monitoring," in *Proc. Int. Conf. Inf. Eng. Comput. Sci.*, Dec. 19–20, 2009, pp. 1–4.

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

[77] X. Wu, H. Gong, P. Chen, Z. Zhong, and Y. Xu, "Surveillance robot utilizing video and audio information," *J. Intell. Robot. Syst.: Theory Appl.*, vol. 55, no. 4–5, pp. 403–421, 2009.

[78] J. Zhang and Z. Liu, "Detecting abnormal motion of pedestrian in video," in *Proc. Int. Conf. Inf. Autom.*, Jun. 20–23, 2008, pp. 81–85.

[79] W. Lao, J. Han, and P. H. N. de With, "Automatic video-based human motion analyzer for consumer surveillance system," *IEEE Trans. Consum. Electron.*, vol. 55, no. 2, pp. 591–598, May 2009.

[80] P. Guo and Z. Miao, "A home environment posture and behavior recognition system," in *Proc. Int. Conf. Convergence Inf. Technol.*, Nov. 21–23, 2007, pp. 175–180.

[81] H. Chen, M. J. Tsai, and C. C. Chan, "A hidden Markov model-based approach for recognizing swimmer's behaviors in swimming pool," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Jul. 11–14, 2010, vol. 5, pp. 2459–2465.

[82] K. Sawai and M. Yoshida, "Algorithm to detect abnormal states of elderly persons for home monitoring," *Syst. Comput. Japan*, vol. 38, no. 6, pp. 34–42, 2007.

[83] M. Stoecklin, "Anomaly detection by finding feature distribution outliers," in *Proc. 2nd Conf. Future Netw. Technol.*, Dec. 4–7, 2006, pp. 1–2.

[84] K. P. Adhiya, S. R. Kolhe, and S. S. Patil, "Tracking and identification of suspicious and abnormal behaviors using supervised machine learning technique," in *Proc. Int. Conf. Adv. Comput., Commun. Control*, Jan. 23–24, 2009, pp. 96–99.

[85] 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. Relat. Assist. Environ.*, Jun. 23–25, 2010, pp. 1–8.

[86] H. Chae, T. Lee, and H. P. In, "Situation aware RFID system: Evaluating abnormal behavior detecting approach," in *Proc. 4th IEEE Workshop Software Technol. Future Embedded Ubiquitous Syst.*, Apr. 27–28, 2006, pp. 27–28.

[87] Y. Benezeth, P. Jodoin, V. Saligrama, and C. Rosenberger, "Abnormal events detection based on spatio-temporal co-occurences," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 20–25, 2009, pp. 2458–2465.

[88] B. Yao, L. Wang, and S. Zhu, "Learning a scene contextual model for tracking and abnormality detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit. Workshops*, Jun. 23–28, 2008, pp. 1–8.

[89] 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. Softw.*, vol. 18, no. 3, pp. 527–537, 2007.

[90] Z. Qing-wei, F. Ai-Ying, and X. Zhi-Hai, "Application of support vector regression and particle swarm optimization in traffic accident

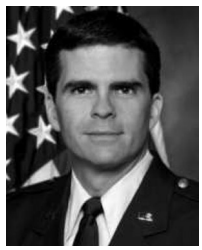
forecasting," in *Proc. Int. Conf. Inf. Manage., Innovation Manage. Ind. Eng.*, Dec. 26–27, 2009, vol. 4, pp. 188–191.
[91] J. Yin and Y. Meng, "Abnormal behavior recognition using self-adaptive hidden Markov models," in *Proc. 6th Int. Conf. Image Anal. Recognit.*, Jul. 6–8, 2009, pp. 337–346.
[92] 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. Distrib. Smart Cameras*, Sep. 7–11, 2008, pp. 1–10.
[93] 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. 15–18, 2008, pp. 121–126.
[94] E. B. Ermis, V. Saligrama, P. Jodoin, and J. Konrad, "Motion segmentation and abnormal behavior detection via behavior clustering," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 12–15, 2008, pp. 769–772.
[95] W. Sultani and J. Y. Choi, "Abnormal traffic detection using intelligent driver model," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 23–26, 2010, pp. 324–327.
[96] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 831–843, Aug. 2000.
[97] M. J. Beal, Z. Ghahramani, and C. E. Rasmussen, "The infinite hidden Markov model," *Adv. Neural Inf. Process. Syst.*, vol. 1, pp. 577–584, 2002.
[98] S. Yu, "Hidden semi-Markov models," *Artif. Intell.*, vol. 174, pp. 215–243, 2010.
[99] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
[100] K. Sayood, *Introduction to Data Compression*. San Francisco, CA: Morgan Kaufmann, 2006.
[101] 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.
[102] M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomaly-based intrusion-detection methods," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 5, pp. 516–524, Sep. 2010.

Angela A. Sodemann received the M.S. degree from the University of Wisconsin-Milwaukee, Milwaukee, in 2006, and the Ph.D. degree from Georgia Institute of Technology, Atlanta, in 2009, both in mechanical engineering.

She is currently an Assistant Professor with the Department of Engineering, College of Technology and Innovation, Arizona State University. Her current research interests include applications of artificial intelligence and machine learning in mechatronics, robotics, and manufacturing.

Matthew P. Ross received the B.S. degree from the U.S. Air Force Academy, Colorado Springs, CO, and the M.S. degree from the Air Force Institute of Technology, Dayton, OH, in 2012, both in computer science.

He is currently an Academic Instructor with the U.S. Air Force Academy. His research interests include artificial intelligence with focus on statistical machine learning, probabilistic modeling, and machine learning.



Brett J. Borghetti received the B.S. degree in electrical engineering from Worcester Polytechnic Institute, Worcester, MA, in 1992, the M.S. degree in computer systems from the Air Force Institute of Technology (AFIT), Dayton, OH, in 1996, and the Ph.D. degree in computer science from the University of Minnesota, Twin Cities, in 2008.

He is currently an Assistant Professor of Computer Science with AFIT. His research interests include artificial intelligence, semiautonomous multiagent systems, and improving human performance in complex data-intensive tasks. He has research experience with statistical machine learning, genetic algorithms, self-organizing systems, neural networks, game theory, information theory, and cognitive science.