Abnormal Human Behavior Detection in Videos: A Review

Huiyu Mu

College of Information and Electrical Engineering, China Agricultural University, Beijing 10083, P.R. China;
e-mails: B20183080630@cau.edu.cn, yuangang_cau@163.com, B20193080656@cau.edu.cn

Ruizhi Sun

College of Information and Electrical Engineering, China Agricultural University, Beijing 10083, P.R. China;
Scientific Research Base for Integrated Technologies of Precision Agriculture (Animal Husbandry), the Ministry of
Agriculture, Beijing 10083, P.R. China; e-mail: sunruizhi@cau.edu.cn

Gang Yuan, Yun Wang

College of Information and Electrical Engineering, China Agricultural University, Beijing 10083, P.R. China;
e-mails: B20183080630@cau.edu.cn, yuangang_cau@163.com, B20193080656@cau.edu.cnn

Corresponding author: sunruizhi@cau.edu.cn

Modeling human behavior patterns for detecting the abnormal event has become an important domain in recent years. A lot of efforts have been made for building smart video surveillance systems with the purpose of scene analysis and making correct semantic inference from the video moving target. Current approaches have transferred from rule-based to statistical-based methods with the need of efficient recognition of high-level activities. This paper presented not only an update expanding previous related researches, but also a study covered the behavior representation and the event modeling. Especially, we provided a new perspective for event modeling which divided the methods into the following subcategories: modeling normal event, prediction model, query model and deep hybrid model. Finally, we exhibited the available datasets and popular evaluation schemes used for abnormal behavior detection in intelligent video surveillance. More researches will promote the development of abnormal human behavior detection, e.g. deep generative network, weakly-supervised. It is obviously encouraged and dictated by applications of supervising and monitoring in private and public space. The main purpose of this paper is to widely recognize recent available methods and represent the literature in a way of that brings key challenges into notice.

KEYWORDS: abnormal detection, video surveillance, behavior representation, event modeling.

1. Introduction

In modern intelligent video surveillance systems, the researches mainly focus on: abnormal detection, virtual reality and video stitching [8, 100, 102], in which the abnormal detection has an important application scenarios and it has raised much attention within the past twenty years [9, 82]. Abnormal detection in video is vital to ensure security in both internal spaces with the outside (e.g. campus, waiting halls and shopping malls). With the numbers of cameras have been widely installed, the task of supervising multiple monitors by security staff becomes much more difficult as the human decreased concentration and fatigue. In addition, abnormal events are extremely rare and infrequent that makes the supervision task more difficult and challenging.

Anomaly detection in video processes the parse temporal sequences of object observations to generate high-level descriptions of agent actions and multi-agent interactions. Detecting abnormality events require building the complex visual patterns, and some patterns can be learned with the long-term temporal relationship and causal inference [79]. In fact, some previous reviews of intelligent video surveillance systems have been published on the subject of abnormal detection [14, 130]. Our survey is relevant but differences reflect in many ways from them. For example, [130] carried out an impressive broad survey on discussion of the creation of intelligent distributed automated surveillance systems. However, these surveys only focus on one perspective of anomaly detection, which are not detailed or systematic enough. Furthermore, with the development of anomaly detection in video surveillance in recent years, the deep generative models which include the variational auto-encoder (VAE) [69], generative adversarial networks (GANs) [49] and other methods has become an important domain. It is necessary to obtain more detailed analysis. We summarize our main contributions as follows:

1　Most of the existing studies either focus on a particular application domain or specific contexts of human activity [2, 28]. Our work aims to provide a comprehensive outline of the advance researches in abnormal human behavior detection as well as several applications these techniques are used.

2　Recently, a lot of novel methods for detecting abnormal behavior in videos with excellent performance

(e.g. deep learning [60, 136]) have been provided. We survey these researches and classify them into organized framework for better understanding and facilitating the reader to view and retrieve the text.

Our work presents an extensive and structured review of anomaly detection technology in video surveillance. This review is structured in five sections. In Section 2, the definition of anomalies in videos and related surveys are presented. Section 3 and its subsections illustrate the representative approaches. In Section 4, the popular datasets and performance evaluation of previous works are provided. Finally, the conclusions and comments for further research are presented in Section 5.

2. Definition and Related Surveys

2.1. Definition

Anomalies are also known as abnormalities, discordants or outliers in the data mining and statistics subject. With the difference of the nature of input data (e.g. sequential data: video, voice and protein sequences. non-sequential data: images, age and other data), it has been used for diverse set of tasks (e.g. video surveillance, image analysis, healthcare, sensor networks). However, the anomalies in video surveillance are a little different than the way of data mining and statistics research, which the anomalies in video surveillance needs to take into account the surrounding environment and inference about the type of event in a scene. For example, while it is "normal" to people to walk across a pedestrian walkway during the traffic lights is green, such type of motion activity is viewed as "abnormal" when the traffic lights change to red. Essentially, it needs to provide high semantic level information, which involves specific context, scenes and temporal-spatial information [126, 132]. Abnormal behavior detection can be view as a kind of high-level operation of image processing, in which the logical information is extracted from input video data. Fig. 1 shows some different kinds of anomalies in various contexts.

Given this backdrop, the definition of an abnormal event in [147] presented the abnormal as deviating from the normal model. Valantinas et al. [112] defined as events

Figure 1

Examples of anomaly event in various contexts

that were unusual and signify irregular behavior. To date, there is no universally accepted definition for abnormal event detection. Among many studies, abnormal behavior can be classified into the following categories:

1. One or more behaviors that are explicitly specified. Such as designating falls as abnormal behavior [68].

2. Abnormal events that deviate qualitatively from what is considered to be normal. Such as only walking is normal in a scene, the running, falling or loitering is regarded as anomaly [29, 152].

3. The events happen with a low frequency (probability). Namely they are nature rare, unexpected, or out-of-the-ordinary [46, 70].

4. The change of posterior probability is higher. The observer will have a priori prediction of the future development for a new event after the relevant evidence or background is taken into account. The event is regarded as anomaly if there is a bias form the priori prediction [121].

2.2. Related Surveys

Video surveillance, which contains capturing and processing visual data from a scene, to detect objects along time and location for the purpose of cognizing interesting situations, has been attracted more research attention. Several previous surveys and conferences were published (See Table 1). The frequency of publications in the topic of anomaly detection in video (Timespan: 2000.1-2020.9) is shown in Fig. 2. There are lots of studies explain how such technologies can help in social security concerns and monitor of public places. Some surveys have emphasized deep learning based methods [19, 71, 107]. For example, Ben et al.

**Table 1**

Key point of previous related surveys

References	Main focus
[56]	Visual surveillance in dynamic behavior
[73]	Behavior analysis
[75]	Video event understanding
[103]	Action classification
[66]	Human activity recognition
[26]	Deep learning, video feature representation
[5, 47]	Deep learning, human activity recognition
[16]	Fixed and moving cameras
[98, 99]	Human behavior detection

[14] did a comprehensive review of abnormal behavior recognition, similar to ours, which were grouped into the behavior representation and the behavior modeling. The methods in crowd surveillance videos were surveyed in [2]. Some of the reviews represented problem- or application-specific work, e.g., fixed and moving cameras [16], 2D and 3D approaches [30], components of a surveillance system [129], spatio-temporal interest point [80]. We provide a complete overview of state-of-the-art human behavior detection surveys.

Although a lot of works reviewed in abnormal behavior detection for intelligent video surveillance, there were shortages of comprehensive outline of the advance researches. For instance many researches focus on a particular application domain, e.g. dynamic behavior [56], crowded scence [2], deep learning [26] and so on. However, those methods with spatio-temporal inter-

Figure 2
Frequency of publications in the topic of anomaly detection in video (the source is Web of Science, timespan: 2000.1-2020.9.)
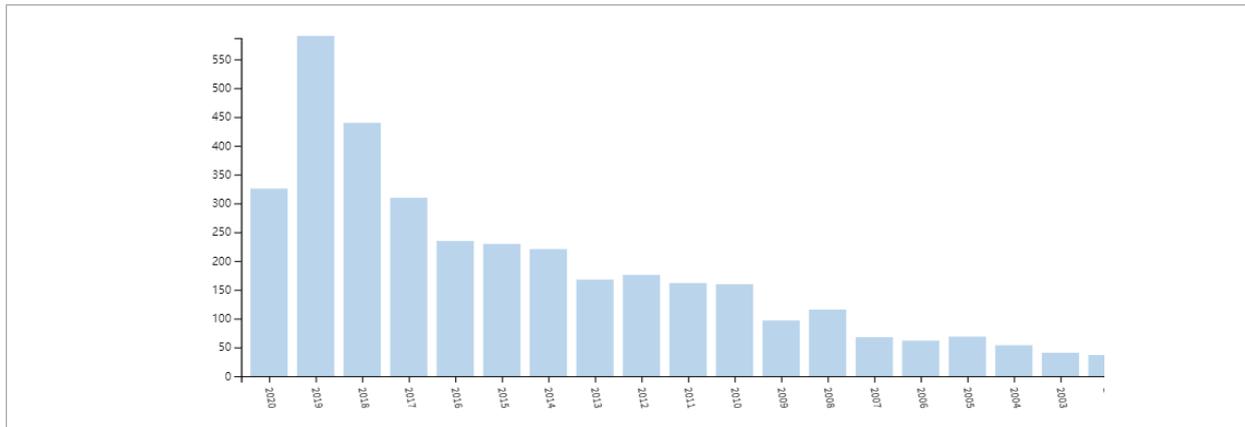
est point or contextual information are insufficient to describe and detection ongoing human activities with complex structures. In this review, we concentrate on high-level activity recognition methodologies designed for the analysis of abnormal human behavior detection and discuss recent research trends in activity recognition. We wish that our survey bridges this gap.

We use a more detailed taxonomy and compared each approach category. For example, differences between handcrafted features approaches and learned features approaches are discussed in our review. We compare the abilities of event modeling and detection methods within each class as well, pointing out what they are advantages and disadvantages. Furthermore, we discuss the public datasets used by the systems, and compare the different evaluation metrics and performance on the datasets which some previous reviews have not focused on.

3. Representative Approaches

In this section, we provide a comprehensive outline of the representative approaches in abnormal behavior detection techniques. The related reasearches mainly involved two stages: *Feature extraction* (it ranges from handcrafted features and learned features) *and quantization, Event modeling and detection*. The former is used to obtain the information description of target scene. For the latter, different from other surveys, we tackle the existing modeling approaches with four different categories. It is the key point that we will explain here. It is organized by following structure described in Fig. 3.

3.1. Feature Extraction and Quantization

The central issue in feature extraction and quantization level is extracting effective and discriminatory

Figure 3
Taxonomy of existing approaches for abnormal behavior detection
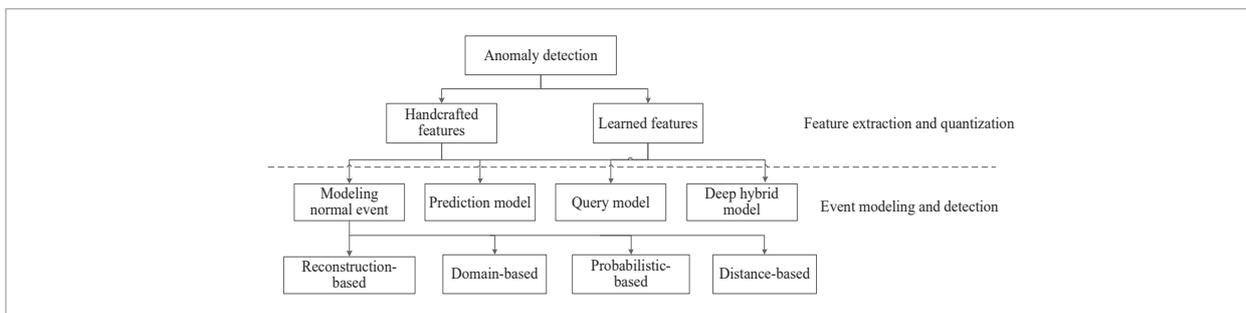
features to match the volumes, trajectories or other forms. Moreover, these features should be invariant to transformation, rotation, illumination, etc. [5]. According to the multifarious methods of feature selection, it ranges from handcrafted features and learned features. Handcrafted features imitate the quality of human vision, which are used to distinguish sensitivity between interested and non-interested areas in vision. In general, handcrafted features contain specific physical meaning [145]. For example, scale-invariant feature transform (SIFT) and histogram of oriented gradient (HOG) could reflect the motion variation and shape information of images respectively [155]. Furthermore, a frame is 2-D data formulated by projecting a 3-D real world scenario, and it covers spatial information (e.g., shapes and locations) of video objects. A video is a sequence of the corresponding 2-D frames placed in a timed sequence. In order to better adapt to the high dimensional characteristics of video data, scholars have extended it to the 3-D scale, i.e. 3-D SIFT and 3-D HOG. Learned features (e.g. deep learning) are a set of techniques that allow the detection systems automatically discover the representations needed for detection or classification task from raw data [157]. Most of them combine multiple features to enhance the effects of the feature extraction. Interested application areas in [8], it integrated the Moravec corner point detection and the scale-invariant feature transform (SIFT) feature extractor. Deep learning and spiking neural networks are excellent methods in machine learning fields and so on. Wu

et al. [145] proposed a fast sparse coding network, and the two-stream neural network was used to extract spatial-temporal fusion features (STFF). Spatial-temporal approaches are those that represent features by analyzing the space-time volumes of video data. The most common strategy is constructing a model 3-D space-time volume representing the information on the videos. The most commonly used features for behavior representation are presented in Table 2.

3.2. Event Modeling and Detection

With the development of abnormal detection research, the methods of event modeling and detection have transferred from rule-based to statistical-based [160]. The rule-based methods are used to assign certain behaviors or building model as abnormal. It can be quite valid in situations where normal behaviours are well-defined and constrained. However, in real world video data, the number of different normal behaviors category can easily surpass what are considered as suspicious [23]. Statistical based methods are sufficient for learning the statistical properties of behavior pattern and they are in favor of describing suspicion events. Hence, the statistical-based methods are preferred in our research. Fast and accurate abnormal event detection is greatly valuable in a large number of scenarios. The core is recognizing the type of behavior performance by the target in video. Accordingly, we devide the event modeling and detection approaches into several categories: modeling normal event, prediction model, query model, and deep hybrid model. The details are described as follows.

1 Modeling Normal Event

In anomaly detection, the idea of event modeling for normal training dataset is a generally adopted device. At this phase, extracting features of training dataset from normal events are used to build a normal event model, which scholars deal with it as a pattern learning problem [117]. Specifically, finding comfortable matches with priori behavior pattern (template), or learning statistical models of the behaviors in video. The sketch of common modeling normal event scheme is shown in Fig. 4. Those methods can be categorized into reconstruction-, domain- probabilistic- and distance-based methods. Some excellent algorithms are presented in Table 3.

Table 2

Popular features for feature quantization

References	Feature types	Comments
[139, 157]	Optical flow	Handcrafted features
[40]	SIFT	
[81]	HOF	
[62]	Social force	
[110]	HOG	
[50, 150]	N-Grams	
[146, 149]	Spatial-temporal	Learned features
[118]	GAN	
[39, 152]	Auto-encoder	

Figure 4

The sketch of common modeling normal event in video sequences
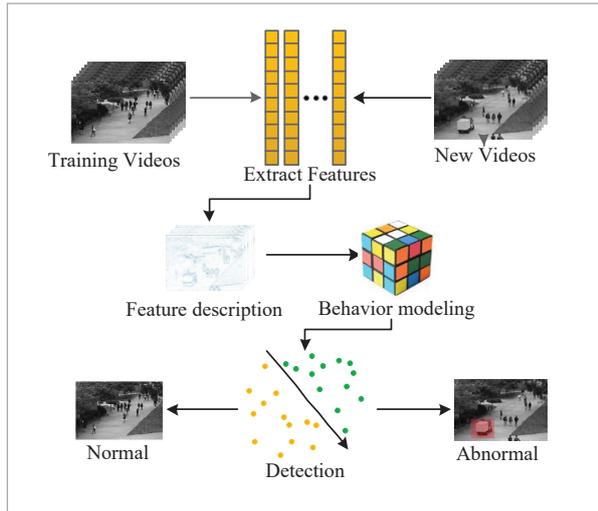
Table 3

The common methods in modeling normal event

Ref.	Description	Comments
[79]	Hierarchy of MDT models	Reconstruction-
[35]	Discrete cosine transform	
[161]	Non-negative matrix factorization	
[72]	A taxonomy of domain anomalies	Domain-
[120]	Graph kernel SVM	
[139]	Nonlinear one-class SVM	
[55]	Bayesian hierarchical method	Probabilistic-
[113]	Entropy approach	
[157]	Motion-field shape description	
[119]	Group motion features	Distance-
[22]	Clustering-driven deep AE	

– *Reconstruction-based.* Reconstruction-based methods are used to compute the deviation between the test data and the normal patterns which are similar outlier detection problem [18]. Calculating the deviation of a behavior pattern from the other could be done in different ways [146]. For example, Li et al. [79] used a joint feature representation, and a hierarchy of mixture of dynamic textures models. Results show that the method achieves the anticipated goal when compared with the state-

of-the art. For the sudden illumination changes, Cermeno et al. [18] matched the label scenes with previously learned examples. Several previous works are focused on finding a group of basis to represent normal data and recognize data with high reconstruction error, e.g., sparse coding [64, 74], auto-encoder (AE) [63]. Specifically, sparse coding methods, for an input video data $X=\{x_1, x_2, ..., x_n\}$, and the over-completed dictionary $D \in R^{d \times k}$ are constructed using the normal data, where $d \ll k$, k is the basis number of dictionary. The optimization function is

$$\min_{D,A} \| X - DA \|_F^2 + \lambda \| A \|_{M_1}, \tag{1}$$

where $A = \{a_1, a_2, ..., a_m\} \in R^{d \times k}$ is the sparse representation of X. We can use the sparse linear combination of the basis in dictionary D to reconstruction the test data $y \in R^d$

$$a^* = \arg\min \frac{1}{2} \| y - Da \|_2^2 + \lambda \| a \|_1, \tag{2}$$

where $a^* \in R^d$ is the reconstruction weights of data y, and the cost of the reconstruction coefficient y is

$$SRC = \frac{1}{2} \| y - Da^* \|_2^2 + \lambda \| a^* \|_1. \tag{3}$$

Sparse coding are mainly divided into fixed base dictionary (e.g. wavelets, the discrete cosine transform, DCT [35]) and learning dictionary (e.g. generalized principal component analysis (PCA); the method of optimal directions (MOD) [37]; and K-singular value decomposition (K-SVD) [64]. The method with fixed base dictionary cannot be the best match by analyzing the structure features of images, so the noise conditions are difficult to extract features effectively. The latter can generate different dictionaries for different types of signals and it has strong adaptive ability. Li et al. [78] introduced a trajectory sparse reconstruction analysis (SRA). The defect of this method is that the detection performance influenced by the control point parameter. Luo et al. [86] built a temporally coherent sparse coding method. However, the dictionary is trained with only normal events and it is generally over-complete, which cannot ensure the expectation.

Thanks to deep learning methods, recent researches are able to make the best of large-scale datasets and powerful computing resource. Deep auto-encoder (AE) is widely used for training encoding-decoding neural networks by minimizing the reconstruction errors [85, 142]. For an input $x \in R^d$ and the corresponding output y_i, it is usually trained to reconstruct the training model at the output of the network and to minimize the reconstruction error e

$$e(x, y) = \frac{1}{2N} \sum_{i=1}^{N} \| x_i - y_i \|_2^2 . \tag{4}$$

However, it ignored the additional constraints and the mapping is the identity [48, 51]. Some works used the denoising AEs to circumvent such limitation [136, 94]. They also use deep architectures to learn a compressed representation for the training data, by reducing the number of hidden units [60]. Such operation may ignore the 2-D structure in videos and the features result in redundancy in the parameters of the network. To cope with this issue, convolutional AE (CAE) architecture is proposed which the weights are communion among all spatial position in the input [95]. The loss function is shown as:

$$e(x, y, W) = \frac{1}{2N} \sum_{i=1}^{N} \| x_i - y_i \|_2^2 + \lambda \| W \|_2^2, \tag{5}$$

where λ is the parameter for the regularization term $\| W \|_2^2$.

Other relates works are proposed with the 3-D information to analysis both temporal and spatiotemporal irregularities in videos [152]. Wang et al. [142] combined deep AE network with 3-D CNN to model the spatio-temporal information in videos. Most of the AEs based researches was built to train the spatiotemporal representation in videos [114]. However, spatiotemporal irregularities are difficult to analysis in the video frames, as they are commonly not properly defined and do not occur frequently in videos. Khan et al. [68] used an adversarial learning framework which contained the spatio-temporal AE and spatio-temporal convolution network. Yan et al. [153] worked on a two-stream recurrent VAE and each stream could achieve modeling the probabilistic distribution of the normal samples by the recurrent VAE. However, it failed in pixel-level detection scheme.

Deep learning-based methods make a breakthrough in anomaly detection by employing deep features in reconstruction. Song et al. [124], the authors introduced an AE combine with attention model to build normal patterns. For solving the video sequences containing images that were never seen before, Slavic et al. [123] introduced a VAE in videos. However, it is hard to assure the anomaly data with a larger reconstruction error because of the strong extensive ability of neural networks. Suppose the novel behaviors as anomaly is probably one-sided for practical surveillance applications. Such as the campus scene, riding a bike is novel behavior since it has not appeared in normal model. However, it should not be treated as an anomaly. Rather than computing the deviation, Xu et al. [149] proposed an adaptive intra-frame classification network in which the one-class deviation problem was translated into a multi-class classification problem. Yong et al. [154] put forward a deep neural network (DNN) and Zhao et al. [159] proposed a spatio-temporal AE network.

We note that Markov random field (MRF), Gaussian mixture model (GMM) and hidden Markov model (HMM) are also widely used for anomaly detection in videos [61, 96]. Take GMM for example, given the normal data could be linked with at least one Gaussian component of Gaussian mixture model, while the abnormal data could not belong to any Gaussian component. It is usually to train a regression model using the training data. When new video data are mapped to the regression model, the reconstruction error is regarded as the abnormality score [39]. It commonly can be divided into three steps [42]: 1. Select a Gaussian mixture $G \sim$ Category (π); 2. Obtain a latent vector $v \sim N(\mu_G, \sigma_G^2 \mathbf{I})$; 3. Calculate the reconstruction result $\mathbf{x}'$, $p(v, G) = \pi_G N(v|\mu_G, \sigma_G^2)$, $[\mu_x; \log \sigma_x^2] = f(v; \theta)$, $\mathbf{x}' \sim N(\mu_x, \sigma_x^2 \mathbf{I})$, where k is the number of components of the mixture, π is the prior probability, $N(\mu_x, \sigma_x^2 \mathbf{I})$ is Gaussian distribution parameterized by means μ_x and covariance σ_x^2. It relies on observation variables and tries to retrieve those observation variables which are already fixed at beginning. On the other hand, variants of GMM like Dirichlet based mixture GMM models [61] and adaptive GMM [133] do not just depend on observations and pertain to longer interaction between observations. To alleviate the shortcomings of GMM, a deep GMM is used in [42].

– *Domain-based.* Domain-based methods commonly

state as a region of the normal videos, based on the distributed character of video data, to describe the domain of the normal sample [72]. One favorite tool in domain-based method is support vector machines (SVMs). It is a typical algorithm for forming a margin's boundary. The ideal hyperplane is to represent the largest separation (or margin) between different classes [93]. The partitions between classes of normal activities have also been learned using kernel SVM [104, 120]. Most of them are to define a boundary for the normal samples according to the structure of the training data which are used to repress the domain of the normal categorize. Similarity, kernel one-class SVM is an efficient tool for abnormal behavior recognition [137], and it also be extended to nonlinear kernel form [139]. What is remarkable here is the complexity corresponding to the calculation of the kernel functions.

Another way is considering the different depths of field for the same scene. Specifically, one object is closer to the video surveillance, greater movement will be detected, and when it is far from the video surveillance, the detected movement will be smaller. This problem is illustrated in Fig. 5, the pedestrian in the red rectangle is moving at a constant speed. However, the detected speeds in Figures 5(a)-5(b) are different which may lead to different motion patterns of objects. To address this problem, several previous works built block-wise modeling and trained normal event model for each block [88, 141]. For example, Cong et al. [27] proposed a novel feature descriptor named multi-scale histogram of optical flow (HOF) which partitioned the frame into a few basic units. Although the block-wise based methods achieve excellent performance, they may lead to another problem: the resolution of a frame is $M \times N$, and dividing into $n \times n$ blocks with k pixels overlapping, so there will produce $\{(M-k)/(n-k)+1]\} \times \{(N-k)/(n-k)+1]\}$ features. To training large numbers of data will cause expensive time cost and waste storage space. In order to overcome the limitations of the above methods, Shi et al. [117] built a coarse scale and region-wise method, in which the blocks with similar depth were incorporated into one region and shared one normal-events model. However, the performance of those works not satisfactory for real-world surveillance applications due to the limitation of few samples and computation.

Figure 5

The movement inconsistency of the same object in different locations

(a)

(b)
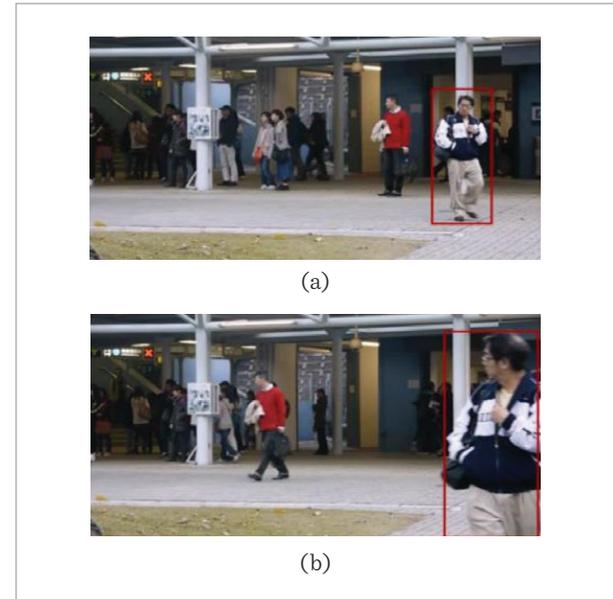
– *Probabilistic-based.* The methods of probabilistic-based are commonly used based on the amount of information that test video should contain, match videos by modeling a probability distribution, and the event not match to the normal pattern is judged to be an abnormal event. For an input $x \in R^d$ and the corresponding $P_x(x)$ is the distribution of normal data X. Hypothesis testing model H_0: x complies with the probability distribution $P_x(x)$; H_1: x complies with the outside probability distribution $P_x(x)$ and when the $P_x(x) < \alpha$, refused H_0, and accept H_1, α is a standardized parameters for unknown distributions.

Different probabilities of behaviors cause different attention in many fields of view [151]. To identify the events importance from the point of probabilistic, She et al. [115] proposed a semantics analysis tool of message importance measure (MIM). The MIM is a generalized Shannon information theory.

$$L(\omega, x) = \sum_{x_i} p(x_i)e^{\omega\{1-p(x_i)\}}, \quad (6)$$

where the $p(x_i)$ is a discrete distribution for data x, w is an adjustable parameter. Motivated by the idea of

probabilistic method, Garcia et al. [46] used a GMM to add new behaviors appearing in the environment. It is noted that probabilistic methods approximated the probability density of the normal samples and detected whether a new data comes from the similarly distribution or not. Those methods of judging the potential data density branch out into parametric and no parametric.

The parametric approaches presume the normal samples are created from an underlying parametric distribution, and the concerned parameters of the distribution are measured from the normal data. For example, Yamanaka et al. [151] adopted a binary feature of auto-encoding for detection and it was a low-complexity probabilistic models. Hou et al. [55] introduced the Bayesian hierarchical method to achieve detection. For probabilistic models, anomalous data can be defined as datasets that lie in low density or concentration regions of the domain of an input training distribution, such as probabilistic topic method [70] and hierarchical probabilistic model [4]. However, those methods will cause a larger error when the data do not satisfied the assumed distribution. The no parametric approaches are suitable for above situation with no need for making assumptions, and it has been expand to fit the complexity of video data. This ideology has been used in [156]. Some previous studies analyzed the event in videos based on its trajectory [157, 52], Sadeghi et al. [109] extended it to an on-line based method. Other research utilized the contextual information [110, 144]. Wang et al. [138] proposed a motion information coding algorithm based on image descriptors. It is most likely that developing rules and probability distribution describing behaviors for a complex context would be a difficult task.

– *Distance-based.*

Due to the small types of normal events and similar characteristics, while the types of abnormal events are numerous, some scholars believe that under a certain characteristic space, the distribution of normal events is closely in the feature space and it is distinguishable from abnormal events. The distance-based methods need to satisfy the assumption: the normal samples belong to a number of large and closely cluster while the abnormal samples are the opposite [17, 87]. The detection is performed by matching the input data of them. To better build the data distribution of normal samples, Chang et al. [21] reduced the dis-

tance between video data and the hidden vectors. Ma et al. [87], the authors the authors used a trajectory distance metrics based on recurrent neural network (RNN) to measure similarities and detected anomalies from trajectory data. Actually, most methods approach the abnormal score is the measure of distance of data from center of the sphere, the data points which are far away from center are regarded as anomalous, such as robust invariant distance measures [17] and distance between the cluster centers and a feature vector [11]. The above methods do without a priori knowledge of the data distribution and can work well for the problem of noisy features. These approaches assume the normal data are clustered together, while the abnormal data do the reverse. Based on the motion trajectory of multiple pedestrians, Guo et al. [119] extracted both distance and relative speed between trajectories, and the detection results were based on the spatial relationship. Lin et al. [81] extended it as an online weighted clustering algorithm. Chang et al. [22] proposed a novel clustering-driven deep AE method and both the reconstruction error and the cluster distance were used to evaluate the anomaly. In these methods, only motion or appearance features are used to perform clustering, which make them would not work for complex scenes, such as disorganized motion directions and over-speed objects. In Table 4, we describe and compare the four categories of modeling normal event methods.

2 Prediction model

Prediction methods are used to detect anomaly events by comparing them with their expectation. Learning to future frame prediction in videos includes the building of an inner representation that simulates the image evolution precisely. Thus, in a way, it is content and dynamic. One of the typical representative works was published in [82]. They search the best match for the prediction data and determined how abnormal it is. The main contribution of this paper is combining the appearance constraint and motion constraint with intensity gradient loss and optical flow loss, respectively (see Fig. 6).

Another popular method is long short term memory (LSTM) AE model and it is similar to the future predictor model [128]. The framework of them are show in Fig. 7. The LSTM AE model consists of the encoder LSTM (read the input feature vector) and the decoder LSTM (output the prediction for the feature vector
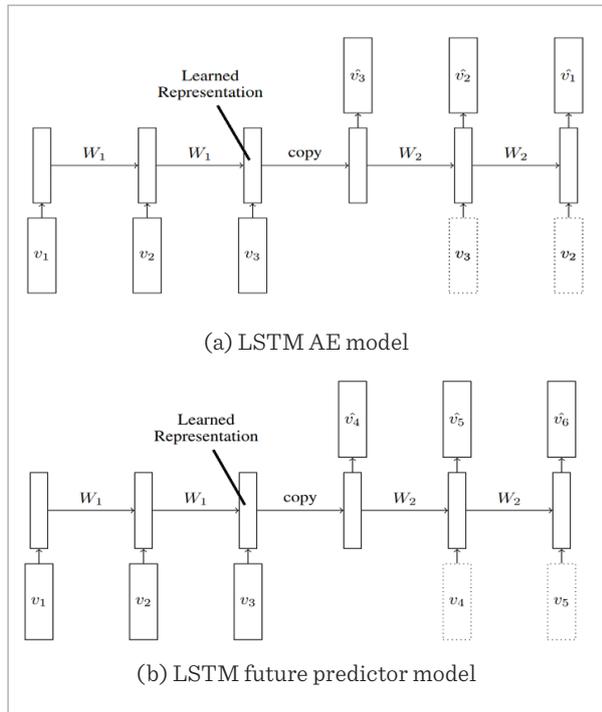
Figure 6

The pipeine of video frame prediction network
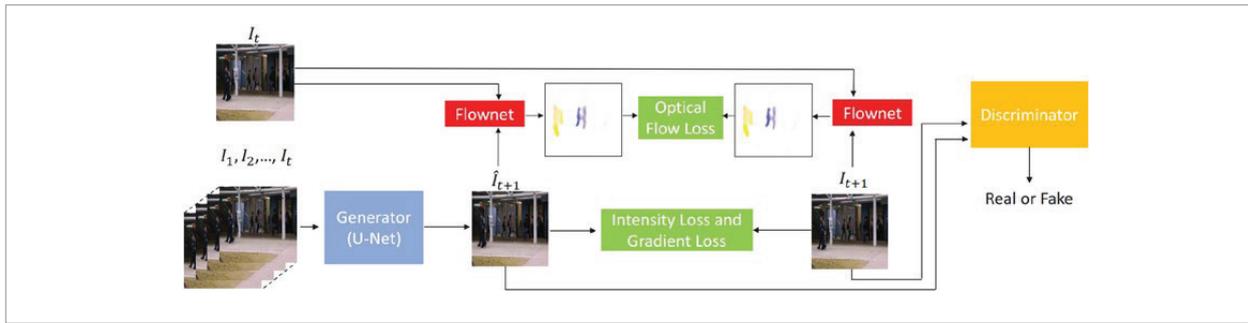
Figure 7

Overview of the LSTM AE model (a) and LSTM future predictor model (b)

(a) LSTM AE model

(b) LSTM future predictor model

create a composite model. The framework of them are show in Fig. 8. However, the precision of object features will decrease with the increase of the training time scale.

Figure 8
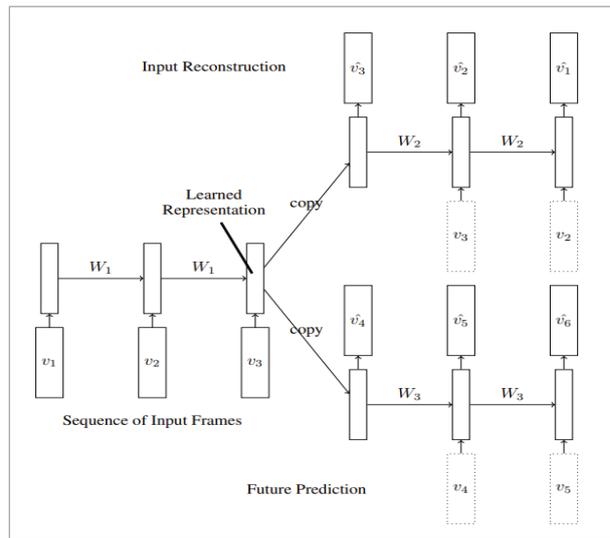
The framework of composite LSTM model [125]

with reverse order). The future predictor model is similar to LSTM AE model, but the decoder LSTM predicts frames that come after the input feature vector. For example, Villegas et al. [134] built a natural video sequence prediction method based on the encoder-decoder CNN and ConvLSTM.

Subsequently, Srivastava et al. [125] combined the reconstructing the input and predicting the future to

The above-mentioned methods mainly in favor of appearance constraint and predicting future frame directly. Unlike those works, some focuses on predicting transformations required for future frame prediction [24, 131, 135]. To get the better result, Shin et al. [118] proposed a hybrid deep learning model consist of video feature extractor and anomaly detector, while the main limitation was the unseen data. In addition to the frequently used spatial constraints on strength and gradient. Xia et al. [148] developed

a feature prediction framework with a novel temporal attention mechanism. Such spatial and temporal constraints promote the future frame prediction for normal events, and accordingly promote to recognize those abnormal events that do not matching the expectation. Similarity, Chen et al. [25] proposed a framework based on bidirectional prediction, they evaluation the deviation between predictive frame and corresponding ground truth to detect abnormal events. However, the generalization ability of this method need to be improved and the adaptive adjustment strategies for the hyper-parameters require further researches [13, 44]. Recently, a lot of great works are credited with the emergence of generative adversarial network (GAN) [83, 89, 116]. It trained in semi-supervised learning model that have shown excellent promise, even with very few labeled data [116]. The basic idea of GAN is it composed of a generator G (a decoder), and a discriminator D (an binary classifier). D and G are simultaneously optimized by the following two-player minimax game with objective function $V(G, D)$:

$$
\min_{G} \max_{D} V(D, G) =
$$
$$
E_{x \sim P_{data}} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))]. \quad (7)
$$

For the generator G, it is to learn a distribution p over data x with a mapping $G(z)$ of samples z. For the discriminator D, it is a standard CNN that maps a frame to a single scalar value D(·) [31]. A variant of GAN method known as Adversarial AE [49, 106]. It use adversarial training to add a prior knowledge on the latent code learnt with hidden layers of autoencoder that are also bring out to effectively calculate the input distribution. Other works combined attention mechanism which used both "soft attention" and "hard-wired" attention [44], while Batchuluun et al. [13] combined the fuzzy method for behavior recognition. In Table 5, some popular prediction methods are presented.

Table 4

Advantages and disadvantages of modeling normal event methods

Method	Description	Advantages	Limitations
Reconstruction-	Computing the deviation between the test data and the normal patterns.	It could be done in different ways and it is easier to be applied on other scenes.	It is hard to assure the anomaly data with a larger reconstruction error.
Domain-	It commonly state as a region of the normal videos to describe the domain of normal sample.	Eliminating the effect of depth of field on motion amplitude.	Large time consuming and waste storage for modeling lots of normal events.
Probabilistic-	Judging the amount of information that test video should contain.	It does not require clustering or prior assumption in contrast to the existing solutions.	The threshold of anomaly detection was difficult to determine.
Distance-	The normal samples belong to one class while the abnormal samples are the opposite.	It is structurally easy to combine with other features.	It has poor robustness and weak scalability.

Table 5

Popular prediction methods for anomaly detection

| Ref.	Description	Used datasets			FPS	End to end	Platform
		UCSD	Avenue	Other			
[82]	U-Net	√	√	ShanghaiTech	25	No	TensorFlow
[159]	Spatio-temporal AE	√	√	Traffic	-	No	-
[90]	Convolutional LSTM	√	√	Subway Datasets	-	Yes	-
[128]	Two U-Net blocks	√	√	ShanghaiTech	30	Yes	GPU
[148]	LSTM network	√	√	-	-	No	TensorFlow/GPU
[25]	Bidirectional prediction	√	√	-	-	No	GPU
[31]	Dual discriminator GAN	√	√	ShanghaiTech	-	No	-
[106]	Multi-timescale	-	√	ShanghaiTech	25	No	GPU

3 Query model

Abnormal patterns are the "interesting" objects that attract human observers attention to a certain extent, and always easy to recognize. Such salient events are so since they are unlike the regular patterns in that context. The query methods are composing the new video data employing spatiotemporal patches that are extracted from previous data. Thus, the regions in the new data which could be composed from the previous data are considered to be anomalies. Typical algorithm [15] presented a new graph-based Bayesian inference method to detect the patches and a probabilistic graphical model to achieve the inference by composition task. An area in the query frame is considered applicable if it has a large enough contiguous area of support in the video data. New valid frame can be inferred from the database, even though they have never been appeared. The basic concept is shown in Fig. 9. For a query frame (a), we can infer the query from the database (b), the database with the corresponding area of support (c). Finally, we can find an ensembles-of-patch with more flexible and efficient form (d). Related works have been applied in class-based object recognition [41, 43].

Unnatural events are boundless in real world scene, and it is almost unrealistic to gather the total of abnormal events and tackle the problem with a classification method [12]. Some works use the statistical

Figure 9

The basic concept of inference by composition [15]

(a) (b)

(c) (d)
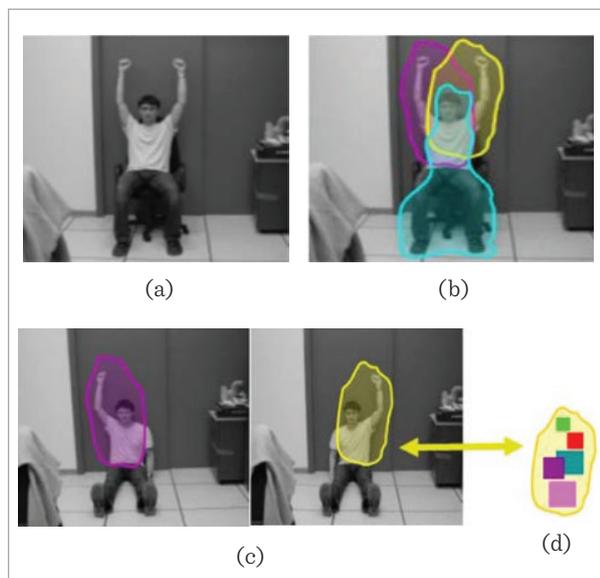
computations [41]. For example, spatial image saliency methods are used by [54], the authors adopted probability estimates and multi-dimensional histogram to find anomalies. The methods of composing new data from previous patches are widely used for a variety of works. Besides, query can be estimated using several methods [34], e.g. Markov network [45] and spatio-temporal patches [143]. For decreasing the computing cost, the dimensionality reduction and nearest neighbor search are applied, which can achieve robust results with small-scale training data. In addition, Leibe et al. [77] built an implicit shape model which combined both identification and segmentation into a probabilistic framework. Sivic et al. [122] added geometric constraints to the non-class-based descriptors. In the review of [158], the authors proposed an improved SVD method to match image pairs. Despite its effectiveness in abnormal detection, methods that fall under this category inherit a problem of difficulty in composing the rules and matches.

4 Deep hybrid model

Deep learning have an advantage of exploring the intricate relationships in multi-dimensional data, and it consistent refreshed the records in many fields, such as computer vision and information retrieval. Meanwhile, machine learning method has been a long period of development, and it explores much knowledge in human behavior detection which the knowledge is universal in the field. At present, the researches begin to focus on the transfer of machine learning methods to deep learning methods, namely deep hybrid models, and it achieves better results,. The most common strategy is to use deep neural network as feature selection, and then the features are input to the classic machine learning algorithms [38]. Specifically, the common used deep neural networks include AE artificial neural networks [7], LSTM neural network [38] and one class neural networks [20]. Given the video data contain a small amount of labeled data with a large amount of unlabeled data, semi supervised methods is applied to the anomaly detection for intelligent video surveillance [57, 121]. For example, Shin et al. [118] proposed a hybrid learning method which the feature extractor trained by GAN and anomaly detection improved by transferring the extractor. In another study, Du et al. [32] proposed a wireless vision sensor network. Hu et al. [58] put forward a spatial-temporal CNN and the deep features were passed in the least squares

Table 6

Popular deep hybrid models for video anomaly detection

| Ref. | Description | Used datasets | | | Platform |
		UCSD	Avenue	Other	
[57]	Faster R-CNN+SVM	√	√	Avenue, Subway	GPU
[121]	CNN+SVM	√	√	-	GPU
[67]	YOLOv3+Flownet2	√	√	ShanghaiTech	-
[65]	DNN+KNN	√		ShanghaiTech	
[108]	FCNs+Gaussian	√	-	Subway	GPU
[118]	3DCNN+LRCN	√	-	Subway	-
[3]	CNN+SVM	√	-	UMN	GPU
[10]	CNN+ADN	√	-	UMN	GPU

SVM to implement classification. Certainly, in many cases, the hybrid learning system has limitations that it is hard to achieve detecting abnormal behaviors by end-to-end learning.

Following the achievement of transfer learning to acquire plenty of outstanding features from models pre-trained on large-scale data, deep hybrid models have also adopted these pre-trained methods for feature extracting with desired result [108]. For example, Visual geometry group network 19 (VGGNet-19) were used for transferring learning in [3, 10]. One of the most common approaches among deep learning is the CNN network. This is because CNN are quite effective in processing unstructured raw data [92, 111]. The deep hybrid approaches will keep on expanding the scale of the model to adapt the intricacy of the data. Therefore, it needs lots of data to fit a more efficient model. Some methods of deep hybrid model are presented in Table 6.

4. Performance Evaluation

In the current section, the representative approaches have been reviewed. The performance evaluation is critical to measure the validity of the proposed methods and also to compare it to other methods.

4.1. Datasets

Large numbers of abnormal behavior recognition methods have been proposed which show it is a hot topic, and there is a growing need for the popular datasets to use for intelligent video surveillance system. In this section, we reviewed the common used datasets for abnormal event detection. An overview of all listed datasets is provided in Table 7. We also show the websites of the data source. Some more related new datasets that published in recent years are shown in [6, 33, 36, 105].

4.2. Evaluation Metrics

For better examining the strengths and weaknesses of anomaly event detection methods and the corresponding applicability in different scene/task of interest, it is necessary to evaluate and judge the performance using suitable evaluation metrics. Evaluation of anomaly event detection methods can be divided into different levels: pixel level, frame level, and object level [101]. The basic measures are binary decisions and they are shown in Table 8.

– *Pixel level*. Pixel-level evaluation considers each pixel individually and one detected anomalous frame is anomalous if the percentage at least 40% truly abnormal pixels are detected. The basic measures are TPR, FPR, TNR, and FPR (shown in Table 8).

Frame level. The abnormal detection methods detect a frame as an anomaly if they detect at least one pixel as anomalous. However, they ignore spatial localization of anomalies which may miss the true anomaly and mislabels a normal pixel as an anomaly. In order to avoid the weakness of misses and mislabels, some relate works use the pixel level ground-

Table 7

Some available datasets for video anomaly detection

Dataset	Resolution	Scene	Example anomalies	Dataset link
UCSD Ped1 [108]	238×158	Pedestrian walk path	Bikers, small carts	*http://www.svcl.ucsd.edu/projects/ anomaly/dataset.html*
UCSD Ped2 [108]	360×240	Pedestrian walk path	Bikers, small carts	*http://www.svcl.ucsd.edu/projects/ anomaly/dataset.html*
Subway Entrance [1]	512×384	Subway station	Wrong direction and no pay	*http://vision.eecs.yorku.ca/research/ anomalous-behaviour-data/*
Subway Exit [1]	512×384	Subway station	Wrong direction and no pay	*http://vision.eecs.yorku.ca/research/ anomalous-behaviour-data/*
UMN [91]	320×240	In-out-door scenes	Crowd dispersion	*http://mha.cs.umn.edu/*
Avenue dataset [84]	360×640	CUHK campus	Run, throw	*http://www.cse.cuhk.edu.hk/leojia/ projects/detectabnormal/dataset.html*
ShanghaiTech [82]	856×480	Campus	Bikers and cars	*https://sviplab.github.io/dataset/campus_ dataset.html*
UCF Crime [127]	240×320	Crime	Abuse, arrest	*https://www.crcv.ucf.edu/projects/real-world/*
Violent Flows [53]	320×240	Youtube	Violence	*http://www.openu.ac.il/home/hassner/ data/violentflows*
Action movies [97]	-	Action movies	One to one fight	*http://visilab.etsii.uclm.es/personas/ oscar/FightDetection/Peliculas.rar*
Hockey Fight [97]	720×576	Ice hockey rink	Violence	*http://visilab.etsii.uclm.es/personas/oscar/ FightDetection/HockeyFights.zip*
IITB-Corridor [106]	-	IIT Bombay campus	Protest, unattended	*https://rodriguesroyston.github.io/Multi- timescale_Trajectory_Prediction*
Web dataset [91]	-	Web	Clash, escape panic	*http://crcv.ucf.edu/projects/AbnormalCrowd/ Normal_Abnormal_Crowd.zip*

truth masks valuation scheme. The measures are shown as follows:

Frame is counted as true positive if
$$TP / (TP + FN) \geq \alpha \qquad (8)$$

Frame is counted as false positive if
$$TP / (TP + FN) < \alpha, \qquad (9)$$

where α is set to 40 commonly, which means the scheme requires at least 40% overlap between the prediction and the ground-truth.

Another relates evaluation metric is dual-pixel, which improve the evaluation of frame-level with localization by punishing the number of FP. One frame

Table 8

The basic binary decisions measures (Note: TPR= true positive rate. TNR= true negative rate. FPR= false positive rate. FNR= False negative rate. TP= true positive. TN= true negative. FP= false positive. FN= false negative.)

Criterion	Equation	Note
TPR	TPR= TP / (TP +FN)	Anomalous data detected as anomalous
TNR	TNR= TN / (TN+FP)	Normal data detected as normal
FPR	FPR= FP / (TN+FP)	Normal data detected as anomalous
FNR	FNR= FN/ (TP +FN)	Anomalous data detected as normal

is detected as an anomaly if at least α % of ground-truth anomaly pixels are detected and at least β % of predicted anomaly pixels are detected.

Similarity, intersection-over-union measure is widely used and it is also considered the TP and FP. One frame is detected as an anomaly if the ratio of the number of ground-true anomaly pixels divided by the number of ground-true anomaly pixels is equal or greater than the threshold γ. The measures are shown as follows:

Frame is counted as true positive if
$$TP / (TP + FP + FN) \geq \gamma \tag{10}$$

Frame is counted as true positive if
$$TP / (TP + FP + FN) < \gamma, \tag{11}$$

where γ is set to 50 commonly.

_ *Object level.* Evaluates not only in frame level but also considers spatial location. The correct detection is developed by the detected abnormality area and the true abnormality area, then calculate the Intersection / Union > threshold v.

$$(DAA \cap TAA) / (DAA \cup TAA) \geq v, \tag{12}$$

where v is a given threshold, DAA is detected abnormality area and TAA is true abnormality area.

To describe the operation of the anomaly detection algorithm in various situations, the common operation is calculating the measures in all cases and giving the performance curves with the resulting set, e.g., receiver operating characteristic (ROC) curves and area under the curve (AUC).

ROC is a curve of TPR versus FPR, which indicates the variation trend of the number of correctly detected abnormal change with the number of normal incorrectly detected as abnormal. However, it loses sight of the difference in class sizes which may lead to misjudgment.

AUC is used to interpret and aids to acquire the general trends in ROC or PR curve. Similarity, some other metrics that are used to evaluate the performance, such as the equal error rate (EER): the ratio of false detection when the FPR equals to the miss rate and accuracy (A): A = (TP + TN) / (TP + TN + FP + FN). For an ideal anomaly detection method in video, the EER should be smallest possible. Contrary to AUC, the higher score of AUC, the better performance. Samples of the performance evaluation of previous excellent papers are shown in Table 9.

Table 9

Summary of the performance evaluation results (Note, in table head of "Datasets": 1 is Ped1 data, 2 is Ped2 data, 3 is Subway entry data, 4 is Subway exit data, 5 is UMN data, 6 is Avenue data, 7 is ShanghaiTech data, which are corresponding to the table head of "Comments".)

Ref.	Used datasets					AUC	EER	Comments	Plantform
	1/2	3/4	5	6	7				
[148]	√	-	-	√	-	83.5	13.5	2	TensorFlow / GPU
[57]	√	√	√	√	-	71 / 80 / 88.3 / 98.9 / 84.5	22 / 16 / - / 21 / 5.8	1 / 2 / 4 / 5 / 6	CPU
[153]	√	√	-	√	-	72.7 / 85 / 84.6 / 89.2 / 75	32.4 / 20.4 / 20.6 / 22.1 / 31.4	1 / 2 / 3 / 4 / 6	Titan X / GPU
[140]	√	-	√	-	-	65.2 / 97.0	21 / -	1 / 5	-
[25]	√	-	-	√	-	89 / 96.6 / 87.8	-	1 / 2 / 6	GPU
[76]	√	-	√	√	√	86.7 / 99.6 / 90 / 76.2	-	2 / 5 / 6 / 7	TensorFlow / GPU
[11]	√	-	√	-	-	82.3 / 94.1 / 96.5	21.4 / 13.2 / 14.3	1 / 2 / 5	Matlab / CPU
[121]	√	-	-	√	-	94.6 / 95.9 / 89.3	-	1 / 2 / 6	Caffe / GPU
[145]	√	-	-	√	-	82.4 / 92.8 / 85.5	25.2 / 12.5 / 20.7	1 / 2 / 6	TensorFlow / GPU
[128]	√	-	-	√	√	82.6 / 86.2 / 83.7 / 71.5	- / 10.0 / - / -	1 / 2 / 6 / 7	GPU
[124]	√	√	-	√	√	90.4 / 90.3 / 90.2 / 94.6 / 89.2 / 70.0	15.8 / 15.5 / 22.6 / 9.3 / 17.6 / 36.5	1 / 2 / 3 / 4 / 6 / 7	-
[59]	√	-	√	-	-	96.0 / 96.3 / 99.7	6.2 / 5.5 / 1.65	1 / 2 / 5	-
[149]	√	√	-	√	-	95.1 / 95.3 / 93.4 / 92.2	9.4 / 9.3 / 14.2 / 15.0	1 / 2 / 3 / 4 / 6	GPU

5. Conclusion and Comments for Further Research

Building an intelligent abnormal human behavior detection system in video surveillance is essential to address human fatigue and inattention when monitoring many surveillance scenes over an extended period of time. Indeed, the detection of an abnormal behavior in video surveillance enables the human observers focus on the scenes which are more likely to contain abnormal behaviors. These technologies can support the security agents by monitoring normal behaviors and early detection of abnormal behaviors in large scale scenes.

In this review, we discussed the various abnormal human behavior detection methods. For each category of anomaly detection techniques, we described the assumption regarding the notion of normal and abnormal data along with its advantages and disadvantages. First, we discussed the definition and related surveys for abnormal event detection in videos. Then we provided a comprehensive overview of representative approaches that covers the feature extraction and event modeling and detection. Finally, we exhibited the most popular datasets and evaluation metrics used for abnormal human behavior detection in intelligent video surveillance. We note that while valid success has been achieved in this interrelated research field, some more work needs to be done as indicated next.

1. To date, there is no universally accepted definition for abnormal event detection.
2. The behaviors are complex and changes frequently in an unconstrained environment.
3. Noise in the raw acquired data, the choice and representation of low-level features, significantly influences the analysis of the behavior.
4. Video shadows and occlusion, scaling of the moving target, light intensity changes, etc. are great challenges with a single-camera.

All these make abnormal behavior detection a difficult task. The relate approaches attempt to build computational action models to automatically identify whether a behavior is normal or not. Specifically, for the suspicious behaviors may have several interpretations, they lie with the context, the time and the place of the event. One possible resolution is using large amount of training data including as much scenarios as possible. For the scene change, the effective solution is choosing features that are robust to scene transformations and less sensitive to the appearance of object. For the visual restrictions of single-camera, some works that use multiple cameras to acquire different views. To handle large amounts of data, it is become a trend to use effective feature selection strategy and deep learning to work efficiently. Furthermore, thanks to the strong learning ability of deep learning method, it achieves the optimal detection result. It is the future direction of development, has broad prospects. Traditional machine learning methods have explored a lot of knowledge related to anomaly detection in video surveillance through a long time of development. These methods are universal in the field, and need to be transferred to the deep learning method to achieve better results. In a word, the anomaly detection in video is still a hot area of research, a possible future survey would be to extend and improve with more mature techniques are put forward.

Acknowledgements

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Adam, A., Rivlin, E., Shimshoni, I., Reinitz, D. Robust Real-Time Unusual Event Detection Using Multiple Fixed-Location Monitors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(3), 555-560. https://doi.org/10.1109/TPAMI.2007.70825

2. Afiq, A. A., Zakariya, M. A., Saad, M. N. M., Nurfarzana, A. A., Khir, M. H. M., Fadzil, A. F., Jale, A., Gunawan, W., Izuddin, Z. A. A., Faizari, M. A Review on Classifying Abnormal Behavior in Crowd Scene. Journal of Visual Communication and Image Representation, 2019, 58, 285-303. https://doi.org/10.1016/j.jvcir.2018.11.035

3. Al-Dhamari, A., Sudirman, R., Mahmood, N. H. Transfer Deep Learning Along with Binary Support Vector Machine for Abnormal Behavior Detection. IEEE Access, 2020, 8(12), 85-95. https://doi.org/10.1109/ACCESS.2020.2982906

4. Alam, R., Petruska, Z. Smart-Energy Group Anomaly Based Behavioral Abnormality Detection. 2016 IEEE Wireless Health (WH), 2016, 1-8. https://doi.org/10.1109/WH.2016.7764554

5. Ali, H. H., Moftah, H. M., Youssif, A. A. A. Depth-Based Human Activity Recognition: A Comparative Perspective Study on Feature Extraction. Future Computing and Informatics Journal, 2017, 3(1), 51-67. https://doi.org/10.1016/j.fcij.2017.11.002

6. Almazroey, A. A., Jarraya, S. K. Abnormal Events and Behavior Detection in Crowd Scenes Based on Deep Learning and Neighborhood Component Analysis Feature Selection. In: Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020), 2020, 258-267. https://doi.org/10.1007/978-3-030-44289-7_25

7. Andrews, J. T., Morton, E. J., Griffin, L. D. Detecting Anomalous Data Using Auto-encoders. International Journal of Machine Learning and Computing, 2016, 6(1), 21-26. https://doi.org/10.18178/ijmlc.2016.6.1.565

8. Bai, Z., Li Y., Chen, X., et al. Real-Time Video Stitching for Mine Surveillance Using a Hybrid Image Registration Method. Electronics, 2020, 9(9), 1-18. https://doi.org/10.3390/electronics9091336

9. Polap, D. Human Machine Interaction in Intelligent Technologies Using the Augmented Reality. Information Technology and Control, 2018, 47(4), 4-37. https://doi.org/10.5755/j01.itc.47.4.21602

10. Bansod, S. D., Nandedkar, A.V. Transfer Learning for Video Anomaly Detection. Journal of Intelligent and Fuzzy Systems, 2019, 36(3), 1967-1975. https://doi.org/10.3233/JIFS-169908

11. Bansod, S. D., Nandedkar, A. V. Crowd Anomaly Detection and Localization Using Histogram of Magnitude and Momentum. The Visual Computer, 2020, 36(3), 609-620. https://doi.org/10.1007/s00371-019-01647-0

12. Bart, E., Ullman, S. Class-Based Matching of Object Parts. In: Conference on Computer Vision and Pattern Recognition Workshop, 2004. https://doi.org/10.1109/CVPR.2004.38

13. Batchuluun, G., Kim, J. H., Hong, H. G., Kang, J. K., Park, K. R. Fuzzy System Based Human Behavior Recognition by Combining Behavior Prediction and Recognition. Expert Systems with Applications, 2017, 81, 108-133. https://doi.org/10.1016/j.eswa.2017.03.052

14. Ben, M., A., Zagrouba, E. Abnormal Behavior Recognition for Intelligent Video Surveillance Systems: A Review. Expert Systems with Applications, 2018, 91(1), 480-491. https://doi.org/10.1016/j.eswa.2017.09.029

15. Boiman, O., Irani, M. Detecting Irregularities in Images and in Video. International Journal of Computer Vision, 2007, 74(1), 17-31. https://doi.org/10.1109/ICCV.2005.70

16. Bour, P., Cribelier, E., Argyriou, V. Crowd Behavior Analysis from Fixed and Moving Cameras. Multimodal Behavior Analysis in the Wild, 2019, 289-322. https://doi.org/10.1016/B978-0-12-814601-9.00023-7

17. Calderara, S., Heinemann, U., Prati, A., Cucchiara, R., Tishby, N. Detecting Anomalies in People's Trajectories Using Spectral Graph Analysis. Computer Vision and Image Understanding, 2011, 115(8), 1099-1111. https://doi.org/10.1016/j.cviu.2011.03.003

18. Cermeno, E., Perez, A., Siguenza, J. A. Intelligent Video Surveillance Beyond Robust Background Modeling. Expert Systems with Applications, 2018, 91, 138-149. https://doi.org/10.1016/j.eswa.2017.08.052

19. Chalapathy, R., Chawla, S. Deep Learning for Anomaly Detection: A Survey. arXiv: Learning, 2019. https://doi.org/10.1145/3394486.3406704

20. Chalapathy, R., Menon, A.K., Chawla, S. Anomaly Detection Using One-Class Neural Networks. arXiv: Learning, 2018.

21. Chang, H., Wang, T., Li, A., Fang, H. Local Hyperspectral Anomaly Detection Method Based on Low-Rank and Sparse Matrix Decomposition. Journal of Applied Remote Sensing, 2019, 13(02), 1-8. https://doi.org/10.1117/1.JRS.13.026513

22. Chang, Y., Tu, Z., Xie, W., Yuan, J. Clustering Driven Deep Autoencoder for Video Anomaly Detection. arXiv: European Conference on Computer Vision, 2020, 1-17. https://cse.buffalo.edu/~jsyuan/papers/2020/ECCV2020-2341-CameraReady.pdf https://doi.org/10.1007/978-3-030-58555-6_20

23. Chaudhary, S., Khan, M. A., Bhatnagar, C. Multiple Anomalous Activity Detection in Videos. Procedia Computer Science, 2018, 125, 336-345. https://doi.org/10.1016/j.procs.2017.12.045

24. Chen, B., Wang, W., Wang, J., Chen, X. Video Imagination from a Single Image with Transformation Generation. arXiv: Computer Vision and Pattern Recognition, 2017. https://doi.org/10.1145/3126686.3126737

25. Chen, D., Wang, P., Yue, L., Zhang, Y., Jia, T. Anomaly Detection in Surveillance Video Based on Bidirectional Prediction. Image and Vision Computing, 2020, 98, 915-924. https://doi.org/10.1016/j.imavis.2020.103915

26. Chong, Y. S., Tay, Y. H. Modeling Representation of Videos for Anomaly Detection Using Deep Learning: A Review. arXiv: Computer Vision and Pattern Recognition, 2015.

27. Cong, Y., Yuan, J., Liu, J. Abnormal Event Detection in Crowded Scenes Using Sparse Representation. Pattern Recognition, 2013, 46(7), 1851-1864. https://doi.org/10.1016/j.patcog.2012.11.021

28. Decampos, T. A Survey on Computer Vision Tools for Action Recognition, Crowd Surveillance and Suspect Retrieval. In XXXIV Congresso da Sociedade Brasileira de Computacao (CSBC), 2014, 1123-1132. https://doi.org/10.1142/S0218213013500309

29. Deepak, K., Chandrakala, S., Mohan, C. K. Residual Spatiotemporal Autoencoder for Unsupervised Video Anomaly Detection. Signal Image and Video Processing, 2020, 11(4), 1-8. https://doi.org/10.1007/s11760-020-01740-1

30. Dhiman, C., Vishwakarma, D. K. A Review of State-of-the-Art Techniques for Abnormal Human Activity Recognition. Engineering Applications of Artificial Intelligence, 2019, 77, 21-45. https://doi.org/10.1016/j.engappai.2018.08.014

31. Dong, F., Zhang, Y., Nie, X. Dual Discriminator Generative Adversarial Network for Video Anomaly Detection. IEEE Access, 2020, 8(12), 170-176. https://doi.org/10.1109/ACCESS.2020.2993373

32. Du, Y. An Anomaly Detection Method Using Deep Convolution Neural Network for Vision Image of Robot. Multimedia Tools and Applications, 2020, 1-14. https://doi.org/10.1007/s11042-020-08684-1

33. Dwivedi, N., Singh, D. K., Kushwaha, D. S. Orientation Invariant Skeleton Feature (oisf): A New Feature for Human Activity Recognition. Multimedia Tools and Applications, 2020, 79, 1-36. https://doi.org/10.1007/s11042-020-08902-w

34. Efros, A. A., Leung, T. K. Texture Synthesis by Non-Parametric Sampling. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, 2002, 2, 1-6. https://doi.org/10.5555/850924.851569

35. Elad, M., Aharon, M. Image Denoising via Sparse and redundant Representations Over Learned Dictionaries. IEEE Transactions on Image Processing, 2006, 15(12), 3736-3745. https://doi.org/10.1109/TIP.2006.881969

36. Elbayoudi, A., Lotfi, A., Langensiepen, C. The Human Behavior Indicator: A Measure of Behavioural Evolution. Expert Systems WITH Applications, 2019, 118, 493-505. https://doi.org/10.1016/j.eswa.2018.10.022

37. Engan, K., Aase, S. O., Husoy, J. H. Multi-Frame Compression: Theory and Design. Signal Processing, 2000, 80(10), 2121-2140. https://doi.org/10.1016/S0165-1684(00)00072-4

38. Ergen, T., Kozat, S. S. Unsupervised Anomaly Detection with LSTM Neural Networks. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(8), 3127-3141. https://doi.org/10.1109/TNNLS.2019.2935975

39. Fan, Y., Wen, G., Li, D., Qiu, S., Levine, M. D. Video Anomaly Detection and Localization via Gaussian Mixture Fully Convolutional Variational Autoencoder. arXiv: Computer Vision and Pattern Recognition, 2020, 195, 1-12. https://doi.org/10.1016/j.cviu.2020.102920

40. Febin, I. P., Jayasree, K., Joy, P. T. Violence Detection in Videos for an Intelligent Surveillance System Using Mobsift and Movement Filtering Algorithm. Pattern Analysis and Applications, 2020, 23, 611-623. https://doi.org/10.1007/s10044-019-00821-3

41. Felzenszwalb, P. F., Huttenlocher, D. P. Pictorial Structures for Object Recognition. International Journal of Computer Vision, 2005, 61(1), 55-79. https://doi.org/10.1023/B:VISI.0000042934.15159.49

42. Feng, Y., Yuan, Y., Lu, X. Learning Deep Event Models for Crowd Anomaly Detection. Neurocomputing, 2017, 219(219), 548-556. https://doi.org/10.1016/j.neucom.2016.09.063

43. Fergus, R. Object Class Recognition by Unsupervised Scaleinvariant Learning. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, 2, 1-8. https://doi.org/10.1109/CVPR.2003.1211479

44. Fernando, T., Denman, S., Sridharan, S., Fookes, C. Soft + Hardwired Attention: An LSTM Framework for Human Trajectory Prediction and Abnormal Event De-

tection. Neural Networks, 2018, 108, 466-478. https://doi.org/10.1016/j.neunet.2018.09.002

45. Freeman, W. T., Pasztor, E. C., Carmichael, O. Learning Lowlevel Vision. International Journal of Computer Vision, 2000, 40(1), 25-47. https://doi.org/10.1023/A:1026501619075

46. Garcia, J., Varona, L., Leskovsky, P., Nieto, M. On-line Normality Modelling and Anomaly Event Detection Using Spatio-Temporal Motion Patterns. In: 7th International Conference on Imaging for Crime Detection and Prevention, 2016. https://doi.org/10.1049/ic.2016.0070

47. Liao, H. C., Pan, M. H., Hwang, H. W., et al. An Automatic Calibration Method Based on Feature Point Matching for the Cooperation of Wide-Angle and Pan-Tilt-Zoom Cameras. Information Technology and Control, 2011, 40(1), 41-47. https://doi.org/10.5755/j01.itc.40.1.191

48. Gong, D., Liu, L., Le, V., Saha, B., Mansour, M. R., Venkatesh, S., Den Hengel, A. V. Memorizing Normality to Detect Anomaly: Memory-Augmented Deep Autoencoder for Unsupervised Anomaly Detection. arXiv: Computer Vision and Pattern Recognition, 2019, 4, 1705-1714. https://doi.org/10.1109/ICCV.2019.00179

49. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Bing, X., Bengio, Y. Generative Adversarial Nets. In Proceedings of the Advances in Neural Information Processing Systems, 2014, 2, 2672-2680. https://doi.org/10.5555/2969033.2969125

50. Hamid, R., Johnson, A., Batta, S., Bobick, A., Isbell, C., Coleman, G. Detection and Explanation of Anomalous Activities: Representing Activities as Bags of Event N-grams. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, 1031-1038. https://doi.org/10.1109/CVPR.2005.127

51. Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A. K., Davis, L. S. Learning Temporal Regularity in Video Sequences. In: Computer Vision and Pattern Recognition, 2016, 733-742. https://doi.org/10.1109/CVPR.2016.86

52. Hassanein, A. S., Hussein, M., Gomaa, W., Makihara, Y., Yagi, Y. Identifying Motion Pathways in Highly Crowded Scenes: A Non-parametric Tracklet Clustering Approach. Computer Vision and Image Understanding, 2020, 191, 710-722. https://doi.org/10.1016/j.cviu.2018.08.004

53. Hassner, T., Itcher, Y., Kliper-Gross, O. Violent Flows: Real-Time Detection of Violent Crowd Behavior. In: Computer Vision and Pattern Recognition Workshops, 2012, 1-6. https://doi.org/10.1109/CVPRW.2012.6239348

54. Honda, T., Nayar, S. K. Finding "Anomalies" in an Arbitrary Image. In: IEEE International Conference on Computer Vision, 2001, 2, 1-8. https://doi.org/10.1109/ICCV.2001.937669

55. Hou, E., Yilmaz, Y., Hero, A. O. Anomaly Detection in Partially Observed Traffic Networks. IEEE Transactions on Signal Processing, 2019, 67(6), 1461-1476. https://doi.org/10.1109/TSP.2019.2892026

56. Hu, W., Tan, T., Wang, L., Maybank, S. A Survey on Visual Surveillance of Object Motion and Behaviors. IEEE Trans Systems Man and Cybernetics, 1986, 34(3), 334-352. https://doi.org/10.1109/TSMCC.2004.829274

57. Hu, X., Dai, J., Huang, Y., Yang, H., Zhang, L., Chen, W., Yang, G., Zhang, D. A Weakly Supervised Framework for Abnormal Behavior Detection and Localization in Crowded Scenes. Neurocomputing, 2020, 383, 270-281. https://doi.org/10.1016/j.neucom.2019.11.087

58. Hu, Y. Design and Implementation of Abnormal Behavior Detection Based on Deep Intelligent Analysis Algorithms in Massive Video Surveillance. Journal of Grid Computing, 2020, 2(18), 227-237. https://doi.org/10.1007/s10723-020-09506-2

59. Hu, Z., Zhang, L., Li, S., Sun, D. Parallel Spatial-Temporal Convolutional Neural Networks for Anomaly Detection and Location in Crowded Scenes. Journal of Visual Communication and Image Representation, 2020, 67, 765-771. https://doi.org/10.1016/j.jvcir.2020.102765

60. Hyunjong, Park, Jongyoun, N., Ham, B. Learning Memory-Guided Normality for Anomaly Detection. arXiv: Computer Vision and Pattern Recognition, 2020, 1-10. https://doi.org/10.1109/CVPR42600.2020.01438

61. Ichihara, K., Takeuchi, M., Katto, J. Accuracy Evaluations of Video Anomaly Detection Using Human Pose Estimation. IEEE International Conference on Consumer Electronics, 2020, 1-9. https://doi.org/10.1109/ICCE46568.2020.9043128

62. Murali, S., Govindan, V. K., Kalady, S. A Survey on Shadow Detection Techniques in a Single Image. Information Technology and Control, 2018, 47(1), 75-92. https://doi.org/10.5755/j01.itc.47.1.15012

63. Jardim, E., Thomaz, L.A., Silva, E. A. B. D., Netto, S. L. Domaintransformable Sparse Representation for Anomaly Detection in Moving-Camera Videos. IEEE Transactions on Image Processing, 2020, 29, 1329-1343. https://doi.org/10.1109/TIP.2019.2940686

64. Jiang, Z., Lin, Z., Davis, L. S. Learning a Discriminative Dictionary for Sparse Coding via Label Consistent K-svd. In: IEEE Conference on Computer Vision and Pattern Recognition, 2011, 1-8. https://doi.org/10.1109/CVPR.2011.5995354

65. Keval, D., Yilmaz, Y. Any-Shot Sequential Anomaly Detection in Surveillance Videos. 2020 IEEE CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, 4037-4042. https://doi.org/10.1109/CVPRW50498.2020.00475

66. Ke, S., Thuc, H. L. U., Lee, Y., Hwang, J., Yoo, J., Choi, K. A Review on Video-Based Human Activity Recognition. The First Computers, 2013, 2(2), 88-101. https://doi.org/10.3390/computers2020088

67. Keval, D., Yasin, Y. Continual Learning for Anomaly Detection in Surveillance Videos. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, 1-10. https://doi.org/10.1109/CVPRW50498.2020.00135

68. Khan, S. S., Nogas, J., Mihailidis, A. Spatio-Temporal Adversarial Learning for Detecting Unseen Falls. Pattern Anal Applic, 2020. https://doi.org/10.1007/s10044-020-00901-9

69. Kingma, D. P., Welling, M. Stochastic Gradient VB and the Variational Auto-Encoder. Journal of Beijing Administrative College, 2013. https://doi.org/pdf/1312.6114v7

70. Kinoshita, A., Takasu, A., Adachi, J. Real-Time Traffic Incident Detection Using a Probabilistic Topic Model. Information Systems, 2015, 54, 169-188. https://doi.org/10.1016/j.is.2015.07.002

71. Kiran, B. R., Thomas, D. M., Parakkal, R. An Overview of Deep Learning Based Methods for Unsupervised and Semi-Supervised Anomaly Detection in Videos. Journal of Imaging, 2018, 4(2), 36-47. https://doi.org/10.3390/jimaging4020036

72. Kittler, J., Christmas, W. J., De Campos, T., Windridge, D., Yan, F., Illingworth, J., Osman, M. Domain Anomaly Detection in Machine Perception: A System Architecture and Taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(5), 845-859. https://doi.org/10.1109/TPAMI.2013.209

73. Ko, T. A Survey on Behavior Analysis in Video Surveillance for Homeland Security Applications. In Proceedings of 37th IEEE Applied Imagery Pattern Recognition, 2008, 1-8. https://doi.org/10.5772/15302

74. Galan-Hernandez, J. C., Alarcon-Aquino, V., Ramirez-Cortes, J. M. Region-of-Interest Coding Based on Fovea and Hierarchical Trees. Information Technology and Control, 2013, 42(4), 343-352. https://doi.org/10.5755/j01.itc.42.4.3076

75. Lavee, G., Rivlin, E., Rudzsky, M. Understanding Video Events: A Survey of Methods for Automatic Interpretation OF Semantic Occurrences in Video. IEEE Transactions on Systems, 2009, 39(5), 489-504. https://doi.org/10.1109/TSMCC.2009.2023380

76. Lee, S., Kim, H.G., Ro, Y. M. Bman: Bidirectional Multi-Scale Aggregation Networks for Abnormal Event Detection. IEEE Transactions on Image Processing, 2020, 29, 2395-2408. https://doi.org/10.1109/TIP.2019.2948286

77. Leibe, B. Combined Object Categorization and Segmentation with an Implicit Shape Model. In: Proceedings of ECCV Workshop on Statistical Learning in Computer Vision, 2004, 17-32. https://doi.org/10.1007/11957959_26

78. Li, C., Han, Z., Ye, Q., Jiao, J. Visual Abnormal Behavior Detection Based on Trajectory Sparse Reconstruction Analysis. Neurocomputing, 2013, 119, 94-100. https://doi.org/10.1016/j.neucom.2012.03.040

79. Li, W., Mahadevan, V., Vasconcelos, N. Anomaly Detection and Localization in Crowded Scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(1), 18-32. https://doi.org/10.1109/TPAMI.2013.111

80. Li, Y., Xia, R., Huang, Q., Xie, W., Li, X. Survey of Spatiotemporal Interest Point Detection Algorithms in Video. IEEE Access, 2017, 5(5), 10323-10331. https://doi.org/10.1109/ACCESS.2017.2712789

81. Lin, H., Deng, J. D., Woodford, B. J., Shahi, A. Online Weighted Clustering for Real-Time Abnormal Event Detection in Video Surveillance. In: Proceedings of the 24th ACM International Conference on Multimedia, 2016, 536-540. https://doi.org/10.1145/2964284.2967279

82. Liu, W., Luo, W., Lian, D., Gao, S. Future Frame Prediction for Anomaly Detection - A New Baseline. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018. https://doi.org/10.1109/CVPR.2018.00684

83. Lotter, W., Kreiman, G., Cox, D. D. Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning. arXiv: Learning, 2016.

84. Lu, C., Shi, J., Jia, J. Abnormal Event Detection at 150 fps in Matlab. IEEE 2013 IEEE International Confer-

ence on Computer Vision (ICCV), 2013, 2720-2727. https://doi.org/10.1109/ICCV.2013.338

85. Luo, W., Liu, W., Gao, S. Remembering History with Convolutional LSTM for Anomaly Detection. In: 2017 IEEE International Conference on Multimedia and Expo (ICME), 2017. https://doi.org/10.1109/ICME.2017.8019325

86. Luo, W., Liu, W., Gao, S. A Revisit of Sparse Coding Based Anomaly Detection in Stacked Rnn Framework. In: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, 341-349. https://doi.org/10.1109/ICCV.2017.45

87. Ma, C., Miao, Z., Li, M., Song, S., Yang, M. H. Detecting Anomalous Trajectories via Recurrent Neural Networks. In: Asian Conference on Computer Vision, 2018, 370-382. https://doi.org/10.1007/978-3-030-20870-7_23

88. Mahadevan, V., Li, W.X., Bhalodia, V., Vasconcelos, N. Anomaly Detection in Crowded Scenes. In: Computer Vision and Pattern Recognition, 2010, 1, Pages: 1975-1981. https://doi.org/10.1109/CVPR.2010.5539872

89. Mathieu, M., Couprie, C., Lecun, Y. Deep Multi-Scale Video Prediction Beyond Mean Square Error. 4th International Conference on Learning Representations, 2015.

90. Medel, J.R., Savakis, A.: Anomaly Detection in Video Using Predictive Convolutional Long Short-Term Memory Networks. arXiv: Computer Vision and Pattern Recognition, 2016.

91. Mehran, R., Oyama, A., Shah, M. Abnormal Crowd Behavior Detection Using Social Force Model. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009. https://doi.org/10.1109/CVPR.2009.5206641

92. Nadeem, A., Jalal, A., Kim, K. Human Actions Tracking and Recognition Based on Body Parts Detection via Artificial Neural Network. In: ICACS 2020, 2020, 1-6. https://doi.org/10.1109/ICACS47775.2020.9055951

93. Ge, H., Zhu, Z., Kang, L. Tracking Video Target via Particle Filtering on Manifold. Information Technology and Control, 2019, 48(4), 904-908. https://doi.org/10.5755/j01.itc.48.4.23939

94. Narasimhan, M., Kamath, S. S. Dynamic Video Anomaly Detection and Localization Using Sparse Denoising Autoencoders. Multimedia Tools and Applications, 2018, 77(11), 173-195. https://doi.org/10.1007/s11042-017-4940-2

95. Nayak, R., Pati, U. C., Das, S. K. Video Anomaly Detection Using Convolutional Spatiotemporal Autoencoder. In: 2020 International Conference on Contemporary Computing and Applications (IC3A), 2020, 175-180. https://doi.org/10.1109/IC3A48958.2020.233292

96. Ng, L. L., Chua, H. S. Vision-Based Activities Recognition by Trajectory Analysis for Parking Lot Surveillance. In: IEEE International Conference on Circuits and Systems, 2013, 137-142. https://doi.org/10.1109/ICCircuitsAndSystems.2012.6408305

97. Nievas, E. B., Suarez, O. D., Garc' lła, G.B., Sukthankar, R. Violence Detection in Video Using Computer Vision Techniques. In: International Conference on Computer Analysis of Images and Patterns, 2011, 332-339. https://doi.org/10.1007/978-3-642-23678-5_39

98. Nigam, S., Singh, R., Misra, A. K. A Review of Computational Approaches for Human Behavior Detection. Archives of Computational Methods in Engineering, 2018, 26, 831-863. https://doi.org/10.1007/s11831-018-9270-7

99. Afsar, P., Cortez, P., Santos, H. Automatic Visual Detection of Human Behavior: A Review from 2000 to 2014. Expert Systems with Applications, 2015, 42(20), 6935-6956. https://doi.org/10.1016/j.eswa.2015.05.023

100. Polap D., Srivastava G. Neural Image Reconstruction Using a Heuristic Validation Mechanism. Neural Computing and Applications, 2020, 1. https://doi.org/10.1007/s00521-020-05046-8

101. Parameswaran, S., Harguess, J., Barngrover, C., Shafer, S., Reese, M. Evaluation Schemes for Video and image Anomaly Detection Algorithms. Proceedings of SPIE, 2016, 9844. https://doi.org/10.1117/12.2224667

102. Polap D., Ksik, K., Winnicka, A., et al. Strengthening the Perception of the Virtual Worlds in a Virtual Reality Environment. ISA Transactions, 2020, 102, 397-406. https://doi.org/10.1016/j.isatra.2020.02.023

103. Poppe, R. A Survey on Vision-Based Human Action Recognition. Image and Vision Computing, 2010, 28(6), 976-990. https://doi.org/10.1016/j.imavis.2009.11.014

104. Priyadharsini, N. K., Chitra, D. A Kernel Support Vector Machine Based Anomaly Detection Using Spatio-Temporal Motion Pattern Models in Extremely Crowded Scenes. Journal of Ambient Intelligence and Humanized Computing, 2020, 19(8). https://doi.org/10.1007/s12652-020-02000-3

105. Ramachandra, B., Jones, M. Street Scene: A New Dataset and Evaluation Protocol for Video Anomaly

Detection. arXiv: Computer Vision and Pattern Recognition, 2019, 2558-2567. https://doi.org/10.1109/WACV45572.2020.9093457

106. Rodrigues, R., Bhargava, N., Velmurugan, R., Chaudhuri, S. Multi-Timescale Trajectory Prediction for Abnormal Human Activity Detection. arXiv: Computer Vision and Pattern Recognition, 2019, 2615-2623. https://doi.org/10.1109/WACV45572.2020.9093633

107. Maskeliūnas, R., Damaševičius, R., Segal S. A Review of Internet of Things Technologies for Ambient Assisted Living Environments. Future Internet, 2019, 12(12), 1-23. https://doi.org/10.3390/fi11120259

108. Sabokrou, M., Fayyaz, M., Fathy, M., Klette, R. Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes. arXiv: Computer Vision and Pattern Recognition, 2016, 172, 88-97. https://doi.org/10.1016/j.cviu.2018.02.006

109. Sadeghi-Tehran, P., Angelov, P. A Real-Time Approach for Novelty Detection and Trajectories Analysis for Anomaly Recognition in Video Surveillance Systems. In: Evolving and Adaptive Intelligent Systems, 2012, 108-113. https://doi.org/10.1109/EAIS.2012.6232814

110. Sahoo, S. P., Ari, S. On an Algorithm for Human Action Recognition. Expert Systems with Applications, 2019, 115, 524-534. https://doi.org/10.1016/j.eswa.2018.08.014

111. Santos, F. P. D., Ribeiro, L. S. F., Ponti, M. A. Generalization of Feature Embeddings Transferred from Different Video Anomaly Detection Domains. Journal of Visual Communication and Image Representation, 2019, 60, 407-416. https://doi.org/10.1016/j.jvcir.2019.02.035

112. Valantinas, J., Kančelkis, D. Speeding-up Image Encoding Times in the Spiht Algorithm. Information Technology and Control, 2011, 40(1), 7-11.

113. Sharif, M. H., Djeraba, C. An Entropy Approach for Abnormal Activities Detection in Video Streams. Pattern Recognition, 2012, 45(7), 2543-2561. https://doi.org/10.1016/j.patcog.2011.11.023

114. Sharma, M. K., Sheet, D., Biswas, P. K. Spatiotemporal Deep Networks for Detecting Abnormality in Videos. Multimedia Tools and Applications, 2020, 79(15), 11237-11268. https://doi.org/10.1007/s11042-020-08786-w

115. She, R., Liu, S., Fan, P. Attention to the Variation of Probabilistic Events: Information Processing with Message Importance Measure. Entropy, 2019, 21(5), 439-452. https://doi.org/10.3390/e21050439

116. Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., Woo, W. C. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In: Proceedings of the 28th International Conference on Neural Information Processing Systems, 2015, 1(12), 802-810. https://doi.org/10.1007/978-3-319-21233-3_6

117. Shi, Y., Liu, Y., Zhang, Q., Yi, Y., Li, W. Saliency-Based Abnormal Event Detection in Crowded Scenes. Journal of Electronic Imaging, 2016, 25(6), 8-16. https://doi.org/10.1117/1.JEI.25.6.061608

118. Shin, W., Bu, S.J., Cho, S. B. 3d-convolutional Neural Network with Generative Adversarial Network and Autoencoder for Robust Anomaly Detection in Video Surveillance. International Journal of Neural Systems, 2020, 30(1), 34-42. https://doi.org/10.1142/S0129065720500343

119. Shuqiang, G., Dongxue, L., Lili, Y. Anomaly Detection in Crowded Scenes Based on Group Motion Features. Journal of Internet Technology, 2020, 21(3), 871-879. https://doi.org/10.3966/160792642020052103024

120. Singh, D., Mohan, C. K. Graph Formulation of Video Activities for Abnormal Activity Recognition. Pattern Recognition, 2017, 65, 265-272. https://doi.org/10.1016/j.patcog.2017.01.001

121. Singh, K., Rajora, S., Vishwakarma, D. K., Tripathi, G., Kumar, S., Walia, G. S. Crowd Anomaly Detection Using Aggregation of Ensembles of Fine-Tuned Convnets. Neurocomputing, 2020, 371, 188-198. https://doi.org/10.1016/j.neucom.2019.08.059

122. Sivic, Z. Video Google: A Text Retrieval Approach to Object Matching in Videos. In: IEEE International Conference on Computer Vision, 2003, 2, 1470-1477. https://doi.org/10.1109/ICCV.2003.1238663

123. Slavic, G., Campo, D., Baydoun, M., Marin, P., Martin, D., Marcenaro, L., Regazzoni, C. S. Anomaly Detection in Video Data Based on Probabilistic Latent Space Models. 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS). 2020, 1-8. https://doi.org/10.1109/EAIS48028.2020.9122766

124. Song, H., Sun, C., Wu, X., Chen, M., Jia, Y. Learning Normal Patterns via Adversarial Attention-Based Autoencoder for Abnormal Event Detection in Videos. IEEE Transactions on Multimedia, 2019, 1-12. https://doi.org/10.1109/TMM.2019.2950530

125. Srivastava, N., Mansimov, E., Salakhutdinov, R. Unsupervised Learning of Video Representations Using LSTMS. In: Proceedings of the 32nd International Conference on International Conference on Machine Learning, 2015, 37(10), 843-852.

126. Stauffer, C., Grimson, W. E. L. Learning Patterns of Activity Using Real-Time Tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8), 747-757. https://doi.org/10.1109/34.868677

127. Sultani, W., Chen, C., Shah, M. Real-World Anomaly Detection in Surveillance Videos. arXiv: Computer Vision and Pattern Recognition, 2018. https://doi.org/10.1109/CVPR.2018.00678

128. Tang, Y., Zhao, L., Zhang, S., Gong, C., Li, G., Yang, J. Integrating Prediction and Reconstruction for Anomaly Detection. Pattern Recognition Letters, 2020, 129, 123-130. https://doi.org/10.1016/j.patrec.2019.11.024

129. Tsakanikas, V., Dagiuklas, T. Video Surveillance Systems: Current Status and Future Trends. Computers and Electrical Engineering, 2017, 70, 736-753. https://doi.org/10.1016/j.compeleceng.2017.11.011

130. Valera, M., Velastin, S. A. Intelligent Distributed Surveillance Systems: A Review. IEE Proceedings - Vision, Image, and Signal Processing, 2005, 152(2), 192-204. https://doi.org/10.1049/ip-vis:20041147

131. Van Amersfoort, J., Kannan, A., Ranzato, M., Szlam, A., Tran, D., Chintala, S. Transformation-Based Models of Video Sequences. arXiv: Learning, 2017, 1-11.

132. Varadarajan, J., Odobez, J. Topic models for scene analysis and abnormality detection. IEEE 12th Int. Conf. Comput. Vision Workshops, 2009, 1338-1345. https://doi.org/10.1109/ICCVW.2009.5457456

133. Varadarajan, J., Subramanian, R., Ahuja, N., Moulin, P., Odobez, J.M. Active online anomaly detection using dirichlet process mixture model and gaussian process classification. In: Applications of Computer Vision, 2017, 615-623. https://doi.org/10.1109/WACV.2017.74

134. Villegas, R., Yang, J., Hong, S., Lin, X., Lee, H. Decomposing motion and content for natural video sequence prediction. arXiv: Computer Vision and Pattern Recognition, 2017.

135. Vondrick, C., Torralba, A. Generating the future with adversarial transformers. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 2992-3000. https://doi.org/10.1109/CVPR.2017.319

136. Vu, H., Nguyen, T., Le, T., Luo, W., Phung, D. Robust anomaly detection in videos using multilevel representations. Proceedings of AAAI19-Thirty-Third AAAI conference on Artificial Intelligence, 2019, 33(01), 5216-5223. https://doi.org/10.1609/aaai.v33i01.33015216

137. Wang, T., Qiao, M., Chen, Y., Chen, J., Zhu, A., Snoussi, H. Video feature descriptor combining motion and appearance cues with length-invariant characteristics. Optik, 2018, 157, 43-54. https://doi.org/10.1016/j.ijleo.2017.11.176

138. Wang, T., Qiao, M., Deng, Y., Zhou, Y., Wang, H., Lyu, Q., Snoussi, H. Abnormal event detection based on analysis of movement information of video sequence. Optik, 2018, 152, 50-60. https://doi.org/10.1016/j.ijleo.2017.07.064

139. Wang, T., Qiao, M., Zhu, A., Niu, Y., Li, C., Snoussi, H. Abnormal Event Detection via Covariance Matrix for Optical Flow Based Feature. Multimedia Tools and Applications, 2018, 77(13), 17375-17395. https://doi.org/10.1007/s11042-017-5309-2

140. Wang, T., Qiao, M., Zhu, A., Shan, G., Snoussi, H. Abnormalevent Detection via the Analysis of Multi-Frame Optical Flow Information. Frontiers of Computer Science in China, 2020, 14(2), 304-313. https://doi.org/10.1007/s11704-018-7407-3

141. Wang, T., Snoussi, H. Detection of Abnormal Visual Events via Global Optical Flow Orientation Histogram. IEEE Transactions on Information Forensics and Security, 2014, 9(6), 988-998. https://doi.org/10.1109/TIFS.2014.2315971

142. Wang, X., Xie, W., Song, J. Learning Spatiotemporal Features with 3dcnn and Convgru for video Anomaly Detection. In: 2018 14th IEEE International Conference on Signal Processing (ICSP), 2018, 474-479. https://doi.org/10.1109/ICSP.2018.8652354

143. Wexler, Y., Shechtman, E., Irani, M. Space-Time Video Completion. Computer Vision and Pattern Recognition, 2004, 1, 120-127. https://doi.org/10.1109/CVPR.2004.1315022

144. Wiliem, A., Madasu, V. K., Boles, W., Yarlagadda, P. K. A Context-Based Approach for Detecting Suspicious Behaviours. Digital Image Computing: Techniques and Applications, 2009, 146-153. https://doi.org/10.1109/DICTA.2009.31

145. Wu, P., Liu, J., Li, M., Sun, Y., Shen, F. Fast Sparse Coding Networks for Anomaly Detection in Videos. Pattern Recognition, 2020, 107, 15-23. https://doi.org/10.1016/j.patcog.2020.107515

146. Wu, P., Liu, J., Shen, F. A Deep One-Class Neural Network for Anomalous Event Detection in Complex Scenes. IEEE Transactions on Neural Networks, 2019, 1-14. https://doi.org/10.1109/TNNLS.2019.2933554

147. Xia, L., Hu, X., Wang, J. Anomaly Detection in Traffic Surveillance with Sparse Topic Model. Journal of Central South University, 2018, 25(9), 2245-2257. https://doi.org/10.1007/s11771-018-3910-9

148. Xia, L., Li, Z. A New Method of Abnormal Behavior Detection Using LSTM Network with Temporal Attention Mechanism. The Journal of Supercomputing, 2020, 20(10). https://doi.org/10.1007/s11227-020-03391-y

149. Xu, K., Sun, T., Jiang, X. Video Anomaly Detection and Localization Based on an Adaptive Intra-Frame Classification Network. IEEE Transactions on Multimedia, 2020, 22(2), 394-406. https://doi.org/10.1109/TMM.2019.2929931

150. Xu, X., Liu, L., Zhang, L., Li, P., Chen, J. Abnormal Visual Event Detection Based on Multil'learning and Autoregressive Integrated Moving Average Model in Edge l'smart City Surveillance. Software - Practice and Experience, 2019, 50(5), 476-488. https://doi.org/10.1002/spe.2701

151. Yamanaka, Y., Iwata, T., Takahashi, H., Yamada, M., Kanai, S. Autoencoding Binary Classifiers for Supervised Anomaly Detection. arXiv: Machine Learning, 2019, 647-656. https://doi.org/10.1007/978-3-030-29911-8_50

152. Yan, M., Meng, J., Zhou, C., Tu, Z., Tan, Y., Yuan, J. Detecting Spatiotemporal Irregularities in Videos via a 3d Convolutional Autoencoder. Journal of Visual Communication and Image Representation, 2020, 67, 47-59. https://doi.org/10.1016/j.jvcir.2019.102747

153. Yan, S., Smith, J. S., Lu, W., Zhang, B. Abnormal Event Detection from Videos Using a Two-Stream Recurrent Variational Autoencoder. IEEE Transactions on Cognitive and Developmental Systems, 2020, 12(1), 30-42. https://doi.org/10.1109/TCDS.2018.2883368

154. Yong, S. C., Yong, H. T. Abnormal Event Detection in Videos Using Spatiotemporal Autoencoder. In: International Symposium on Neural Networks, 2017, 189-196. https://doi.org/10.1007/978-3-319-59081-3_23

155. Žemgulys J., Raudonis V., et al. Recognition of Basketball Referee Signals from Videos Using Histogram of Oriented Gradients (HOG) and Support Vector Machine (SVM). Procedia Computer Science, 2018, 130, 953-960. https://doi.org/10.1016/j.procs.2018.04.095

156. Zhang, T., Fan, Z., Yu, X. Anomaly Detection Based on Locality Sensitive Hashing with Genetic Algorithm. IOP Conference Series Materials Science and Engineering, 2020, 790, 67-79. https://doi.org/10.1088/1757-899X/790/1/012067

157. Zhang, X., Yang, S., Zhang, J., Zhang, W. Video Anomaly Detection and Localization Using Motion-Field Shape Description and Homogeneity Testing. Pattern Recognition, 2020, 105, 94-108. https://doi.org/10.1016/j.patcog.2020.107394

158. Zhao, F., Huang, Q., Gao, W. An Image Matching Algorithm Based on Singular Value Decomposition. Journal of Computer Research and Development, 2010, 47(1). https://doi.org/10.3724/SP.J.1016.2010.02202

159. Zhao, Y., Deng, B., Shen, C., Liu, Y., Lu, H., Hua, X. Spatio Temporal Autoencoder for Video Anomaly Detection. Proceedings of the 25th ACM International Conference on Multimedia, 2017, 1933-1941. https://doi.org/10.1145/3123266.3123451

160. Zhou, J. T., Du, J., Zhu, H., Peng, X., Liu, Y., Goh, R. S. M. Anomalynet: An Anomaly Detection Network for Video Surveillance. IEEE Transactions on Information Forensics and Security, 2019, 14(10), 2537-2550. https://doi.org/10.1109/TIFS.2019.2900907

161. Zhu, X., Liu, J., Wang, J., Li, C., Lu, H. Sparse Representation for Robust Abnormality Detection in Crowded Scenes. Pattern Recognition, 2014, 47(5), 1791-1799. https://doi.org/10.1016/j.patcog.2013.11.018