Detection of Abnormal Human Activities in Surveillance Video - A Survey

Divya U H¹, Josephine Prem Kumar²

¹ Assistant Professor, Department of CSE, East Point College of Engineering and Technology, Bangalore, Karnataka, India
divya.uh@eastpoint.ac.in

² Professor, Department of CSE, Cambridge Institution of Technology, Bangalore, Karnataka, India
josephine.cse@cambridge.edu.in

Abstract—In recent years, it is common to use the surveillance cameras for continuous monitoring of public and private spaces because of increasing crime. Most current surveillance systems need a human operator to constantly watch them and are ineffective as the amount of video data is increasing day by day. Surveillance cameras will be more useful tools if instead of passively recording, they generate warnings or real-time actions when unusual activity is detected. But recognizing and classifying human activity as normal or abnormal from a live video stream is a challenging task in the field of computer vision. There is a need for a smart surveillance system for the automatic identification of abnormal behavior of humans for a specific scene. This paper gives an overview of different machine learning techniques used in recent years to develop such a model. It also gives an exposure to the recent works in the field of anomaly detection in surveillance video and its applications.

Keywords—Video Surveillance, Abnormal Activity, Machine Learning, Deep Learning, ATM center, Fall Detection

I. INTRODUCTION

With the increase in crime all over the world, the use of visual surveillance and cameras for security applications is continuously growing, and it has become part of the modern era. Video surveillance is done by installing CCTV(Closed-Circuit Television), at places to be secured. Surveillance cameras are inexpensive and found everywhere these days, but there must be someone who must monitor the activities constantly. In many situations where surveillance cameras are used, it is common to find poor monitoring due to human factors like boredom, tiredness and operator feeling exhausted as nothing new is happening to pay attention. Although video footage capturing devices are easily affordable and popular in today’s world, available human resources to monitor and analyze the footage are quite limited. Despite the effort to keep the places under surveillance 24/7, most of the time it is not possible to stop the crime in that instant. A surveillance camera can be a more useful tool if instead of passively recording the activities, it can be used to detect events that need special attention of the operator on time. There is an increasing demand for automatic detection of suspicious behavior of a person in public places such as shops, parking lots, ATM centers, airports, railway stations, entrance and corridors of buildings, etc., to identify subjects for standoff threat analysis and detection.

Recognizing human activities from a video stream is a challenging task. From the past decade, human action recognition has received significant attention from the researchers of the computer vision community. Analyzing a human action is not only presenting patterns of movement of different parts of the body; but also a description of the human intention, emotions, and thoughts. Human behavior analysis and understanding are essential for many applications such as human-computer interaction, surveillance, sports, elderly-health care, training, entertainment, and so on.
In general, human activity recognition systems follow a hierarchical approach[1]. At lower-level, human objects are segmented from the video frame. This process is followed by feature extraction such as the characteristics of human objects such as colors, shape, silhouette, body motion and poses. The human activity action recognition module falls under a mid-level approach followed by the reasoning engines on the high-level that interprets the context of the actions as either normal or abnormal.

II. DEFINITION OF ABNORMALITY

Giving an exact definition of abnormal behavior is hard for several reasons. It mainly depends on understanding what is normal. This can vary from one application to another as well as one activity to another. Even for the human cognitive system, it is difficult to analyze complex scenes. Abnormality in videos usually means the identification of events that significantly deviate from regular or normal behavior. Meaning of abnormality may vary according to the context, i.e., time, place, and circumstances. And it also depends on the activity of a person individually, in a group, crowd, and with objects. For example, bending down and searching for something is normal in a supermarket, but it is considered to be abnormal inside the Automatic Teller Machine (ATM) center.

For any specific context, there is a notion of what constitutes normal behavior and conversely, abnormal behavior. Interestingly, abnormal or unusual patterns are somehow the “interesting” things that catch the attention of human observers and are often quite easy to identify[2]. Such behaviors are so because they are different from the regular patterns in that context. Thus anomalies are temporal or spatial outlier events not conforming to learned patterns[3]. Essentially, there is an important objective to model both the appearance and dynamics of normal events to detect the presence of and identify the spatial location of the anomaly present in the scene. A definition of an abnormal event was proposed by J. Varadarajan et al., in [4] as, “an action done at an unusual location, at an unusual time” or “events that are fundamentally different in appearance or having an unusual order of events”.

Smart surveillance systems are in high demand to recognize ongoing abnormal or suspicious activities in crowded areas such as the airport, railway station, shopping malls, restaurants, streets, private spaces like houses, apartments, offices, schools and sensitive areas like ATM centers, hospitals, military areas, fuel stations, exam center and so on. Challenges in the anomaly detection majorly include appropriate feature extraction, addressing the variations between normal and abnormal behavior, there occurrence of abnormal events, background variations, camera movements, etc. Also, it is challenging to recognize human activities in unconstrained videos due to some real conditions such as, varying light conditions, divergent viewpoints, varying action speeds, light variations [1]. Researchers have long sought to bridge the gap between human and machine intelligence by developing neural networks, Bayesian, and other models of human cognition. A great variety of approaches have been proposed to develop an active surveillance video system to mitigate the crimes or prohibited activities in the private and public sectors.

III. TRAINING AND LEARNING TECHNIQUES

This section gives an overview of different machine learning techniques used to develop abnormal activity detecting model.

A. Supervised Learning

Many researchers have attempted to build systems capable of interpreting and understanding human behaviors. Earlier researchers focused on supervised learning algorithms to classify various human activities. Fully supervised model-based approaches are appropriate if unusual events are well-defined and enough training samples are available [5, 6]. Most of the work on abnormal behavior detection took a supervised learning approach [7-10], based on the assumption that there exist well-defined classes of both normal and abnormal behavior. In the supervised approach, one would construct a training set consisting of anomalous and normal behaviors to build a model;
then use the model to classify new behavior sequences as normal or abnormal. This method is not suitable when examples of abnormal behavior rarely exist. Most of the research on detecting abnormal activities in the surveillance video is based on supervised learning as it gives a promising accuracy. Some popular algorithms in supervised learning are Support Vector Machine (SVM), k-nearest neighbors, Bayesian networks, Neural Networks, and decision trees.

B. Unsupervised Learning

The unsupervised learning method works based on the principle that the frequencies of occurrence of abnormal events are comparatively less than that of normal events. The unsupervised approach generally constructs a generative model of the normal or repeatedly occurring behavior patterns, then uses the model to classify behavior sequences as abnormal when they are entirely new patterns from the typical behavior [11-13]. One good example of unsupervised learning is clustering which groups the regular activities as normal and irregular activities as outlier without the need for a trained dataset. Thus, no special requirement in video surveillance to capture an extremely wide variety of typical behaviors. Some of the work uses ensembles of classifiers, but most of the recent research has focused on anomaly detection methods using incremental clustering [14-17]. These methods generally work by comparing a new pattern against a collection of clusters representing historically typical behavior and classifying the new pattern as an anomaly if its distance from the nearest cluster is above the threshold. In [16], the authors proposed a novel visual behavior modeling approach which is learned incrementally and adaptively from a small bootstrapping training set.

C. Semi-supervised Learning

Semi-supervised learning is a class of machine learning techniques that also make use of unlabeled data for training. Here the amount of unlabeled data is typically larger than the amount of labeled data. Semi-supervised learning falls between unsupervised learning and supervised learning. As human perception will work well with prior knowledge, many machine learning researchers have found that unlabeled data, when used in conjunction with a small amount of labeled data, can produce remarkable improvement in classification. When a fully supervised method of classification needs a lot of labeled data for learning, there must be a skilled agent to label which is normal and abnormal activity. The cost associated with such a process is expensive, compared with the acquisition of unlabeled data in case of unsupervised learning. Most of the time it suffers from either overfitting or underfitting. But unsupervised learning effectively works on entirely new patterns of a dataset, but results may be unpredictable. Hence, the authors of [18-20] attempted to combine the advantages of supervised and unsupervised learning in a probabilistic setting. The author [18] proposed a framework for unusual event detection. The approach was motivated by the observation that, “it is unrealistic to obtain large training data set for unusual events and are unpredictable, it is conversely possible to do so for usual events, allowing the creation of a well-estimated model of usual events”. This scarcity of training datasets for unusual events made the authors of [19] propose the use of Bayesian adaptation techniques, which adapt a usual event model to produce some unusual event models in an unsupervised manner. From the point of anomaly detection, not much work has been done using a semi-supervised learning approach.

D. Deep Learning

Availability of GPU (Graphics Processing Unit) processors and huge datasets, the concept of deep learning is gaining popularity in the field of computer vision. Since it is very challenging to recognize human activities in unconstrained videos due to some real conditions such as varying light
conditions, divergent viewpoints, varying action speeds, light variations [3] there is a need of a learning approach where the features are learned automatically reducing the laborious human intervention, expert knowledge, and selection of optimal features[21].

A deep learning model is a machine learning system implemented by a deep neural network. Deep Neural Networks is one of the best architectures used to perform difficult learning tasks. Deep Learning models automatically extract features and build a high-level representation of image data. This is more generic because the process of feature extraction is fully automated. From the image pixels, Convolutional Neural Network (CNN) can learn visual patterns directly. In the case of a video stream, long short term memory (LSTM) models are capable of learning long term dependencies. LSTM networks can remember things[22]. The authors of [22-30] had made use of deep learning techniques to detect abnormal behavior in a surveillance video. A survey by S. Dargan et al., in [31] gives an overview of deep learning, its basic and advanced architectures, techniques, and applications. Some of the key comparisons between deep learning and machine learning, provided by them are:

- Deep learning takes a large amount of data while machine learning needs a small amount of data to work and arrive at a conclusion.
- Deep learning requires hardware with very high performance.
- Deep learning creates new features by its processes and techniques, whereas, in the case of machine learning, features are accurately and precisely recognized by the users.
- The time requirement to train is much more in deep learning than in machine learning.
- The Accuracy rate achieved by deep learning is very satisfactory as compared to machine learning.

IV. ACTIVE FIELDS IN SUSPICIOUS HUMAN ACTIVITY RECOGNITION FROM VIDEO SURVEILLANCE

Some of the applications of recognizing abnormal human activity which mainly took the attention of researchers are ATM center, crowd anomaly detection, fall detection, loitering in public places, suspicious activities in the examination hall, and supermarket.

A. Anomaly in ATM Center

Automatic Teller Machine (ATM) centers are the most vulnerable site for criminal activities despite being under surveillance 24/7. Hence ATM centers are one of the most active research areas for detecting abnormal activities such as robbery, overcrowding, peeping to check the password, snatching the withdrawn money, covered face, and so on. From the review paper [32] it has been found that most of the research work done on detecting covered faces and illegal objects inside the ATM center. Only 4% of research has been done on identifying abnormal or suspicious activities. The commonly used approach is supervised learning using the SVM classification method. In [33] authors used a 3D camera like Kinect to extract skeleton data from the depth image and posture recognition was achieved using Logistic regression, a supervised learning technique to predict the class. In paper [34] the authors proposed a novel model that uses Convolutional Neural Network (CNN) with Long Short-Term Memory(LSTM) for detecting anomalous behavior. Input to the model included videos from the surveillance camera and CNN equipped for understanding and extracting the important features from the frames of the video.

B. Fall Detection

Automatic detection of human fall is one of the important research areas in computer vision. Relatively large numbers of papers were published on fall detection in the elderly home care system.
Earlier researches were mainly focused on a device-based approach, where patients were supposed to wear electronic devices like an accelerometer to detect the fall, which creates inconvenience for the subjects. Because of this reason, computer vision-based approach is gaining more attention not only for indoor home care and health care environment but also to detect a pedestrian fall in an outdoor scenario [35, 36]. The major issues in both the cases are: camera position is arbitrary and the subjects are free to move around, the presence of many covariate factors like varying view angle, illumination, and clothing. The authors of [37] have developed a recurrent neural network (RNN) with LSTM architecture that models the temporal dynamics of the 2D pose information of a fallen person.

C. Crowd Anomaly Detection

Crowd anomaly detection is an important research topic in both computer vision and video analysis. Its applications are in public transportation stations, pilgrim places, social or private events, cricket or football matches, busy streets, and markets. People involved fighting, pushing, or collapsing and crowd panicking can be assumed as abnormal activities in a crowd. A good example of a normal and abnormal activity is: crowd running in a marathon can be classified as normal, while people suddenly start running in an open market that may trigger the alarm as an emergency scenario[38]. In a review paper[39] authors proposed a general framework and pattern taxonomy for detecting abnormal behavior in a crowded environment. Sometimes anomalous events may also include a person loitering about a place for unusual amounts of time. Usually, loitering human behavior often leads to abnormal situations in bus or railway stations such as pickpocketing, snatching chains, robbery, and kidnapping in the residential area, etc. In [40] authors proposed a Markov random walk model that can robustly detect loitering individuals in any outdoor public place. Compare to crowd anomaly detection, not much research has been done in the detection of loitering in video surveillance.

D. Detecting suspicious activities in the Examination hall

Monitoring of exam hall through human invigilators is common all over the world. Even though the room is under CCTV surveillance, it is difficult to detect suspicious activities like passing incriminating material among the students, hand signaling, peeping into other paper, etc., simultaneously from multiple screens till the completion of the exam. Supervising an examination hall is a challenging task in terms of manpower. It is necessary to develop a surveillance system that can assist the educational institute in monitoring the examination hall. Automatic detection of suspicious activity involves the recognition of multiple faces, head and hand movement, gesture, and eye gaze. Supervised learning like classification is the best approach as it is easy to label certain activities as normal or abnormal. In [41] authors used a training model such as Artificial Neural Network for automated face recognition and hand detection assisted by skin color from surveillance videos. In [42] authors considered only eye gaze and head orientation information as clues to detect suspicious behaviors of the pupils. In [43] authors proposed a model based on various computer vision algorithms like Viola-Jones and related-like Feature and AdaBoost classifier algorithms, to identify the hand-contacts of students, student peeping into another answer sheet-based, and tracking of the students genuinity in the classroom by comparing their faces with photos stored in the database.

V. RESULT AND DISCUSSION

Fig. 1 gives the percentage of papers that are published in the IEEE forum from 2010 onwards in different areas of anomaly detection in surveillance video. More than 50% of the research work concentrated on classifying the set of activities as normal and abnormal using a hand-crafted training model. The survey shows that more attention needs in areas like the super-market, shops, entrance or staircase of buildings, streets, and so on. Since activities of humans are unpredictable and differ
from one scene to others, there is a large scope in developing the application-specific smart video surveillance system.

Fig. 1 Papers published in IEEE forum from 2010 onwards.

VI. CONCLUSIONS

The main intention of installing CCTV is to stop the crime or damage by detecting suspicious, or abnormal activities that are happening in the surveillance. There is a huge demand for the development of a smart surveillance system which not only reduces human involvement in monitoring but also alerts the respective authority on time from the future mishappening. Since people are aware of the existence of CCTV almost everywhere, in most situations, behavior of people involved in crimes may seem normal. But too many false alarms could also result in irritations or a loss of trust in the system. Hence, developing such a novel model with less training time and data set, with high accuracy and self-learning with time is highly in need.

REFERENCES


