Identification and Detection of Abnormal Human Activities using Deep Learning Techniques

ASHISH SHARMA¹, NEERAJ VARSHNEY²

Department of Computer Engineering and Application
GLA UNIVERSITY, MATHURA
ashish.sharma@gla.ac.in

Department of Computer Engineering and Application
GLA UNIVERSITY, MATHURA
neeraj.varshney@gla.ac.in

Abstract—In recent years, it is in public 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 stimulating job in the pitch of CPU vision. There is a need for a smart surveillance system for the automatic identification of abnormal behaviour of humans for a specific-scene. Present paper stretches an overview of different machine learning methods 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 centre, Fall Detection

1. 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 reasonable besides found universally present existences, but there must be someone who must monitor the activities constantly. In numerous circumstances wherever surveillance cameras are utilized, it is mutual to invented prived observation pay able to humanoid issues such asboredom, tiredness and operator feeling exhausted as nothing new is happening to pay attention. 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 instinctive recognition of apprehensive behaviour of a person in public places such as shops, parking lots, ATM centres, airports, railway stations, entrance and corridors of buildings, etc., to identify subjects for standoff threat analysis and detection.

Identification of human activity by video stream is a stimulating job. From the past decade, human action recognition has established important consideration by the research of the CPU vision community. Analysing of human activity by presenting patterns of movement of different parts of the body; but also, explanation of the human purpose, emotions, and opinions. Human behaviour investigation and
consideration are important for numerous requests like human-computer communication, observation, 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. Present process is tracked by feature extraction such as the characteristics of human objects such as colours, shape, silhouette, body motion and poses. The human activity act appreciation segment cascades under a mid-level method surveyed by the intellectual appliances on the high-level that interprets the situation of the activities as either normal or abnormal.

2. DEFINITION OF ABNORMALITY

Giving an exact definition of abnormal behaviour 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. Unpredictability to features Typically implies those ID number of occasions that essentially veer off from standard alternately typical conduct. Significance for abnormality might differ as stated by those context, i. E., time, place, and condition. 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) centre.

For whatever particular context, there may be an idea of the thing that constitutes typical conduct technique Furthermore conversely, abnormal conduct technique. Interestingly, abnormal or surprising designs would some way or another those “interesting” things that make the consideration by mankind's spectators Furthermore need aid frequently all the exactly simple on identify [2]. Such practices would thereabouts in view they need aid unique in relation to the general examples in that connection. Accordingly, anomalies are temporal or spatial outlier occasions not adjusting to scholarly patterns [3]. Essentially, there is a critical destination to model both those presence And Progress by typical occasions with recognize the vicinity of and recognize the spatial area of the aberrance introduce in the scene. A meaning about an abnormal occasion might have been suggested by j. Varadarajanet al., over [4] as, “an activity completed during a surprising location, toward a surprising time” alternately “events that need aid Generally different On presence alternately Hosting a surprising request about occasions.”

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 centres, hospitals, military areas, fuel stations, exam canter and so on. Encounters in the irregularity recognitionmajorly comprises suitable feature extraction, addressing the variations between normal and abnormal behaviour, there are occurrence of irregular actions, background differences, camera movements, etc. Also, it is inspiring for identifying human happenings in unconstrained videos because of roughly real circumstances like changing light circumstances, deviating viewpoints, changing exploit hustles, light differences [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.

3. TRAINING AND LEARNING TECHNIQUES

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

Various investigators attempt to figure systems are able of interpret and understand human behaviours. Earlier researchers focused on supervised learning algorithms to classify various human activities. Completely managed model-based methodologies are suitable whether surprising occasions need aid well-defined and sufficient preparing specimens need aid accessible [5, 6]. The vast majority of the fill in looking into abnormal conduct technique identification took A regulated Taking in approach [7-10] [44-48], dependent upon those suspicion that there exist well-defined classes about both ordinary and abnormal conduct technique. In the managed approach, one might develop a preparing situated comprising by bizarre Furthermore typical practices to build A model; then utilize the model to arrange new conduct successions as normal or abnormal. This method is not appropriate for example of abnormal behaviour rely 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 neighbours, 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 generallyconstructs a Generative model of the ordinary or over happening conduct patterns, at that point utilization the model on arrange self-destructive considerations and conduct successions as abnormal at they need aid actually new designs from the average self-destructive considerations and conduct [11-13]. Person great sample from claiming unsupervised Taking in is grouping which Assemblies the general exercises Similarly as typical and unpredictable exercises Similarly as outlier without those need for A prepared dataset. Thus, no exceptional prerequisite Previously, feature reconnaissance should catch a greatly totally mixed bag for ordinary practices. A percentage of the worth of effort employments ensembles about classifiers, Anyway The greater part of the late Look into need cantered once aberrance identification routines utilizing incremental grouping [14-17]. These strategies by worth of effort by contrasting another example against an accumulation about groups speaking to Truly average conduct technique Furthermore classifying the new design Concerning illustration an aberrance Assuming that its separation from those closest bunch may be over the edge. To [16], the writers recommended A novel visual conduct technique demonstrating approach which is scholarly incrementally Also adaptively from a little bootstrapping training set.

C. Semi-supervised Learning

Semi-supervised learning is a period of machine learning techniques that makes use of unlabeled data for training. Here the quantity of unlabeled data is typically larger than the quantity of labelled data. Semi-supervised knowledge cascades among unsupervised learning and supervised learning. As human perception will work well with prior knowledge, various machine-learning investigator shasunla belled data, after utilization in conjunction by insignificant quantity of labelled data, to produce a remarkable improvement in classification. When a fully supervised method of classification needs a lot of labelled data for learning, there must be a capable negotiator 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] Endeavoured with consolidate the preferences for regulated
Also unsupervised Taking in for a probabilistic setting. The creator [18] recommended a structure to surprising off chance identification. Those approach might have been spurred by those perception that, it will be doubtful to acquire extensive preparing information set to surprising occasions Furthermore are unpredictable, it is conversely workable should would so to typical events, permitting the formation of a well-estimated model for common occasions. This shortage from claiming preparation datasets to surprising occasions produced those writers for [19] recommend the utilization for Bayesian adjustment techniques, which adjust an ordinary occasion model to process a few surprising occasion models done a unsupervised way by the point of anomaly detection, not much work has been done using a semi-supervised learning approach [49-51].

D. Deep Learning

Availability of GPU (Graphics Processing Unit) processors and vast datasets, the thought of deep learning is gaining popularity in the field of computer vision. Since it is exceptionally testing should distinguish mankind's exercises in unconstrained features because of A percentage true states for example, such that changing light conditions, dissimilar viewpoints, fluctuating activity speeds, light varieties [3] there will be An have of a Taking in methodology the place the offers need aid figured out how naturally diminishing those relentless human intervention, master knowledge, Furthermore determination about ideal features[21].

A Deep learning taking in model will be a machine Taking in framework actualized Toward A Deep learning neural system. Deep learning neural Networks is a standout amongst the best architectures used to perform challenging Taking in assignments. Deep learning taking in models naturally extricate Characteristics Also Fabricate a high-keyed representational of picture information. This is All the more nonexclusive on account of the transform about characteristic extraction is fully robotized. From those picture pixels, Convolutional neural system (CNN) might take visual examples specifically. On account of a feature stream, in length transient memory (LSTM) models need aid skilled of Taking in in length expression dependencies. LSTM networkscan remember things [22]. The authors of [22-30] had made use of deep learning techniques to detect abnormal behaviour in a surveillance video. A survey by S. Dargan et al., in [31] gives an overview of deep learning, the elementary and innovative architectures, techniques, and applications. Some of the key comparisons between deep learning and machine learning, provided by them are:

- Deep learning taking in takes an expansive amount about information same time machine Taking in needs A little measure from claiming information to worth of effort and land at a Decision.
- Deep learning taking in obliges fittings with high execution.
- Deep learning taking in makes new Characteristics Toward its methods Also techniques, whereas, on account of machine learning, offers need aid faultlessly Furthermore decisively perceived Eventually Tom's perusing the clients.
- The duration of the time prerequisite to train will be a great deal that's only the tip of the iceberg for Deep learning taking in over in machine Taking in.
- Those exactness rate attained by Deep learning taking in may be thick, as acceptable Concerning illustration contrasted with machine Taking in.
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 centre, crowd anomaly detection, fall detection, loitering in public places, suspicious activities in the examination hall, and supermarket.

A. Anomaly in ATM Centre

Automatic Teller Machine (ATM) centres are the most vulnerable site for criminal activities despite being under surveillance 24/7. Hence ATM centres 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 noticing irregular behaviour. Contribution to the typical included videos by surveillance camera and CNN prepared for understanding and extracting the significant structures by the surrounds of the video.

![Human Activity recognition](image)

**Figure 1. Human Activity recognition**

B. Fall Detection

Automatic detection of human fall is one of the significant 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 occurrence of numerous 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 behaviour in a crowded environment. Sometimes anomalous events may also include a person loitering about a place for unusual amounts of time. Usually, loitering human behaviour often leads to abnormal situations in bus or railway stations such as pick-pocketing, 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 signalling, 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 colour from surveillance videos. In [42] authors considered only eye gaze and head orientation information as clues to detect suspicious behaviours 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 genuinely in the classroom by comparing their faces with photos stored in the database.

4. 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 supermarket, 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.
5. 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 miss happening. Since people are aware of the existence of CCTV almost everywhere, in most situations, behaviour 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


[51] Mr. Rakesh Kumar & Prof. Diwakar Bhardwaj, as their research paper titled, "EBH-DBR Energy Balanced Hybrid Depth Based Routing Protocol for Underwater Wireless Sensor Networks", has been accepted for publication in the World Scientific Journal (SCI Indexed).
