International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRSR21869

Anomaly Detection for Video Surveillance

Jagruti Tatiya, Riya Makhija, Mrunmay Pathe, Sarika Late, Prof. Mrunal Pathak
AISSMS Institute of Information Technology, Pune, Maharashtra, India

Article Info
Volume8, Issue 3
Page Number:82-89

Publication Issue
May-June-2021

Article History
Accepted :01May2021
Published :05May2021

ABSTRACT

Anomaly Detection is system which identifies inappropriate human behavior. One of the major problems in computer vision is identifying inappropriate human behavior. It is crucial as activity detection can help many numbers of applications. It can benefit applications like image monitoring, sign language recognization, object pursue and many more. Many alternatives are there such as low-cost depth sensors, but they do have some drawbacks such as limited indoor use also with lower resolution and clamorous depth information from deep images, it becomes nearly impossible to assess human poses.In order to resolve the above issues, the proposed system plans to utilize neural networks. One of the major research area is to recognize suspicious human behavior in video monitoring, in the field of computer vision. Several surveillance cameras are situated at places like airports, banks, bus station, malls, railway station, colleges, schools, etc to detect suspicious activities such as murder, heist, accidents, etc. It is a tedious job to detect and monitor these activities in crowded places, to trace real time human behavior and classify it into ordinary and unexpected scenarios the system needs to have a smart video surveillance. The experimental results show that the proposed methodology could assuredly detect the unexpected events in the video.

Keywords : Video Surveillance, Anomaly detection, Image Processing, CNN, Machine Learning.

I. INTRODUCTION

With the increasing demand for security, surveillance cameras have been widely deployed as the infrastructure for video analysis [11]. Surveillance cameras are being used in public places e.g., streets, intersections, banks, shopping malls, etc. to increase public safety and it only store records of CCTV footages. One major challenge faced by surveillance video analysis [11] is detecting abnormal, which requires exhausting human efforts. However, the monitoring capability of law enforcement agencies has not kept pace. The result is that there is a glaring deficiency in the utilization of surveillance cameras and an unworkable ratio of cameras to human monitors. One critical task in video surveillance is detecting anomalous events such as traffic accidents, crimes, or illegal activities.

The proposed system tends to create an application for the identification of anomaly in public places in real time. The proposed system can be used for surveillance in areas such as malls, airports, train stations, etc. where there is a chance of theft or shooting. The proposed system will use deep learning and neural networks [1] to train our system. This model will then be implemented as a desktop app that will take real-time [6,8] CCTV footage as input and send a warning to the administrator when a suspicious activity is detected. Real life implementations range from gaming to AR/VR, to health care and gesture recognition. Compared to the image data domain, very little work is being done to apply CNNs to the video classification. This is because a video is more complex than a picture because it has a different dimension-temporal.

Unsupervised learning [1] exploits the temporal dependency between frames and has proved to be effective in video analysis. Some anomaly activity approaches use CPU instead of GPU so that anomaly activity can run on low-cost hardware like embedded systems and cell phones. Low-cost depth sensors are another new technology for computer vision [2,4,9]. They are present in game consoles including the Kinect for Xbox 360. They are motion sensors that allow the user to communicate with a console without a game controller, only by hand gestures. These are RGB-D sensors that provide depth information through standardized lighting technology. However, these sensors are limited to indoor use and their low resolution and noisy depth information make it difficult to estimate human anomaly in depth images.

II. LITERATURE SURVEY

Anomaly detection in high-dimensional data is becoming a fundamental research field with a wide range of real-world applications. As a result, a significant amount of research has been dedicated to resolving this problem. Nonetheless, the majority of existing research concentrate on particular aspects of anomaly detection or high dimensionality. As an example, A project reviewed various anomaly detection techniques in order to provide a basic understanding of the various approaches to anomaly detection. Another study looks at many real-world applications of graph-based [3] anomaly detection and concludes with unsolved problems in the field. A survey of various anomaly detection techniques for various applications was conducted. The survey compared the advantages and disadvantages of various detection techniques. A new survey of anomaly detection systems and hybrid intrusion detection systems [1] was conducted to recognise open issues and challenges.

A small amount of research has been done that explicitly or indirectly connects anomaly detection and high dimensionality problems. A thorough examination of specialised algorithms for anomaly detection in high-dimensional numerical data, as well as key aspects of the dimensionality curse. Different subspace clustering algorithms for high-dimensional data are discussed, as well as some possible applications for the algorithms.

Several new methods suggest CNNs to address particular imaging problems directly. These architectures imitate instances of the ODP system [9], but their reasons for design are vastly different. For deblurring, a network was proposed that uses a single, fixed deconvolution step followed by a learned CNN, similar to a prior step in ODP [9] but with a different initial iterate. For deblurring, a network was suggested that uses a single, studied deconvolution stage followed by a CNN, similar to a one-iteration ODP [9] network. A CNN for MRI has been suggested, whose performance is averaged with observations in k-space, similar to a one-iteration ODP network but without the prior and data steps being jointly learned. We develop these deep models by using the ODP framework to recognise the link to traditional optimization-based approaches and to build more efficient multi-iteration architectures.

Anomaly detection systems [1], a subset of intrusion detection systems model normal system/network operation, making them extremely effective at detecting and foiling both known and unknown or "zero day" attacks. Anomaly detection systems, while appealing on paper, face several technological challenges before being widely adopted. Among them are high false alarm rates [5], failure to scale to gigabit speeds, and other problems.

The definition of an anomaly is always depending on what context is of interest. A video event is considered as an anomaly if it is not very likely to occur in video [6]. In 2015, For describing such unusual events i.e., anomaly in complex scenes Real-time anomaly detection and localization method [6] was developed, where each video is defined as a set of non-overlapping cubic patches and is described using two local and global descriptors. These descriptors capture video properties in a variety of ways. By incorporating simple and cost-effective Gaussian classifiers [6], the system can distinguish normal activities and anomalies in videos. The local and global features are based on structure similarity between adjacent patches and the features learned in an unsupervised way, using a sparse autoencoder [3,7,8]. The experiments confirm that the system can reliably detect and localize anomalies as soon as they happen in a video.

Another research presents, an unsupervised dynamic sparse coding approach for detecting unusual events in videos where, system accepts video sequence as an input then the proposed method uses a sliding [5] window along both the spatial and temporal axes to define an event. The task of unusual event detection is therefore formulated as detecting unusual group of cuboids residing in the same sliding window. A dictionary is first learnt from the video using sparse coding and later updated in an online fashion as more data become available. Given the learned dictionary, a reconstruction weight vector is learned for each query event and a normality measure [5] is computed from the reconstruction vectors. The proposed algorithm only needs to scan through the video once, and online updating of the learned dictionary makes the algorithm capable of handling concept drift in the video sequence. Finally, using sparse coding enables the algorithm to robustly discriminate between truly unusual events and noisy usual events.

In 2017, Weixin Luo, Wen Liu, recommended a Temporally coherent Sparse Coding (TSC) [7] in one of his paper, inspired by the ability to detect sparse coding-based anomaly, where the proposed system implement identical neighbouring frames with similar coefficients of reconstruction. Then the proposed system maps the TSC with a special kind of stacked Recurrent Neural Network (sRNN). The nontrivial selection of hyper-parameters to TSC can be avoided 1by taking advantage of sRNN in learning all parameters simultaneously, while the reconstruction coefficients [7] can be inferred within a forward passage with a shallow sRNN, which decreases the computational cost of learning sparse coefficients.

The contributions of this paper are two-fold: I the proposed system propose a TSC that can be mapped to a sRNN that promotes optimization of the parameter and accelerates the prediction of anomalies. (ii) The proposed system constructs a very large dataset that is much greater in terms of both the amount of data and the variety of scenes than the description of all current datasets for anomaly detection. Extensive tests on both a toy dataset [7] and actual datasets show that our approach based on TSC and sRNN consistently outperforms existing methods, which validates our method's effectiveness.

Yong Sheen Chong and Yong Haur Tay presented an effective method for detecting anomalies in videos. Recent applications of convolutional neural networks [1,8], especially in photos, have shown promises of convolutional layers for object detection and recognition. Coevolutionary neural networks, however 1are supervised and involve labels as signals of learning. In videos that involve crowded scenes, the proposed system proposes a spatiotemporal architecture [8] for anomaly detection. Our architecture consists of two main components, one for the representation of spatial features and one for learning about the temporal evolution of spatial features. Experimental findings on benchmarks for Avenue, Subway and UCSD indicate that our method's detection accuracy is comparable at a substantial speed 2of up to 140 fps to state-of-the-art methods.

III. COMPARISON

Lavee et al. (2005) developed a framework for analysing a video for suspicious event detection [4]. In this, low level features are extracted and an event representation for several overlapping subsequences is created for use in event detection [4]. Then, newly created events are compared with a set of predefined events and classified by nearest neighbor algorithm [2].

(Willems et al. 2009; Nguyen et al. 2009), are available in the market for the fall detection, these devices are mostly electronic devices that compel to the elder people either to put it into pocket or wear it on the wrist. Normally, these wearable fall detectors have manual help button or accelerometer to detect a fall. However, these wearable fall investigators have hardly any disadvantages. One of the weaknesses for the fall detectors is that the elderly people can forget to wear devices and help buttons are useless for those people who become unconscious after falling. The modern advancements in the field of computer vision [2,4,9] have brought new solutions to overcome these drawbacks.

One of the main advantages of visual-based fall detection is that such system does not require a person to wear anything, and it is less disturbing in comparison to the wearable sensor. Moreover, computer vision [2,4,9] system provides more information on the behavior of a person compared with the normal wearable sensors. With this, visual-based home monitoring system can provide information on falls and other activities of daily living behaviors which are useful for health-care monitoring, such as mealtime, and sleep duration. A Human Fall detection image captured by an intelligent visual surveillance system can be seen in.

(Lai et al. 2012) It is well known that flame has some corner on its contour in burning stage. Due to the air and wind flow, the corner position will be in the upper half of the flame region and the position will be constantly flashing. In addition, the space experiments results have done by NASA show that due to the gravity, flame shape is not a circular, but it will always have sharp corners. The corners of each object can be acquired by the Harris corner detection algorithm. Each object is compared with the same id object which has been captured in past image. If the corner position of the same id objects in consecutive frames is different then, corner is treated as a dynamic corner otherwise it is a static corner. Flicker rate is defined as the sum of the dynamic corner counts and total corner counts. Compression is a function that calculates the shape feature in geometry. It is used as the relationship between the area and perimeter. In this, the rectangle which encloses the flame region is divided into two smaller equal rectangles.Both smaller rectangles have same perimeter.

Chua et al. (2013) proposed a visual based fall detection approach with low computational complexity for the analysis of human shape. Median filtering method has been used for the background subtraction. Human figure has been traced in three points head, physique, and legs. The bounding box of the foreground blob is divided into three portions and then centroids are calculated to draw two lines. Each line represents to the distances and orientations of the human body. Ratios of the line distance of two consecutive frames are compared and orientation difference is computed to analyse the body shape of the human. This approach has 6.7% false alarm rate [5]. This system failed in detection of two fall incidents because human body of the person was in a straight line and its ratio distance was computed only 1.

Two crouch-down activities were also detected as fall because of the sudden changes in the ratio of distances.

Penmetsa et al. (2014) proposed an autonomous unmanned aerial vehicle visual surveillance system to detect the suspicious human activities such as slapping, punching, hitting, shooting, chain snatching and choking using pose estimation, and appearance of body parts. The system used combination of face detector and upper body detector to improve the efficiency of human detection [2]. Then, acascade filtering has been used to speed-up the face detection. Hough inclination calculator avail oneself to categorize the poses. Orientation features of the human pose is compared with the poses in the suspicious action dataset, and it is flagged with the action which matches the best. The system can detect the multiple suspicious activities such as slapping, punching, hitting, shooting, chain snatching and choking with detection accuracy 77.78, 76.67, 79.59, 73.47, 78.26% respectively. This system grows the time complication and guide us to astound as the number of people rise in the video frames.

Table 1: Pros and Cons of Anomaly Detection Techniques

Technique	Pros	Cons
Convolutional Neural Network	1)Image recognition issues require a high level of precision which is provided by CNN. 2) It detects the important features without the need for human involvement. 3)Weight sharing.	1)Inability to be spatially invariant when dealing with input data. 2) The location and orientation of an object are not encoded by CNN. 3)Lots of training data is required
Neural Network	1)Can do things that a linear programme can't. 2)When one of the neural network's elements fails, the network's parallel nature is unaffected. 3)A neural network is self-learning and does not need reprogramming. 4)It has a wide range of applications.	1)Needs a quality training to operate. 2)Since the design of a neural network differs from that of microprocessors, it must be emulated. 3)For large neural it requires high processing time.

Decision Tree	1)Easy to understand and interpret. 2)It requires a small amount data preparation. 3) Large data improves performance, and it takes less time. 4) The model is based on a white box. 5)It can validate a model using statistical tests. 6)It is capable of dealing with both numerical and categorical data.	1)Decision-tree learners produce overly complex trees that do not adequately generalise the results. 2)Under many aspects, the problem of learning an optimal decision tree is considered. to be NP complete. 3)There are several ideas that are difficult to grasp.
K -Nearest Neighbor	1)When there are few variables, it is very clear to understand. 2)Models involving non-standard data, such as text, will benefit from this feature.	1)Has huge storage requirements. 2) It is adaptive to the similarity function that has been selected. 3)It is computationally expensive technique.
Support Vector Machine	1)Some kernels have infinite dimensions, which means they can learn extremely complex concepts quickly. 2)It can handle data with a lot of dimensions with ease. 3)Have a good performance.	1) It will take a lot of memory and CPU power to complete the task. 2)In solving the constraint QP, there are several numerical stability issues. 3)Both positive and negative examples are needed. 4)It is important to have a good kernel function.

IV. ANALYSIS

Anomaly detection is the identification of rare items, events or observations which raise suspicions by differing significantly from most of the data. Machine Learning algorithm provides a wide range of applications. Neural Network, K-nearest neighbor [2], Decision Tree, Convolutional Neural Networks and Support Vector Machine are the most renowned Machine learning Algorithms. Analysis and comparison of different algorithms results that, Convolutional Neural Network (CNN) is best suited for the proposed system with accurate and efficient output.

As of CNN, various algorithms for noise removal, resize and binary conversion, and segmentation were analysed. After comparing them, Gaussian Filter [6] resulted to be suited appropriately for noise removal, Nearest-Neighbor Interpolation method for resize and binary conversion, and Thresholding method and Edge Based Segmentation for the segmentation part of the proposed system.

The proposed system will detect anomaly activity by using CNN. In this system, CNN algorithm is used to classify the dataset.

Table 2: Accuracy of CNN Algorithm

Epochs	Training Accuracy	Testing Accuracy
10	94.25	92.90
20	94.23	93.93
50	95.57	94.86
100	95.72	94.47
200	94.23	91.53
500	94.59	92.27

V. CONCLUSION

Using a case study of a manual CCTV camera, the proposed framework tested and addressed Anomaly Activity detection [1,10] and the benefits of implementing Machine learning based Anomaly Activity detection. The proposed system includes a brief description of the system's architecture, design, and foundation structure, as well as a brief introduction to the system. Based on this, the proposed framework has determined that it is one of the best technologies for anomaly detection. It is more dependable, precise, and effective when used as part of the Anomaly Activity Detection process. A framework for analysing real-time CCTV footage [1,10] will assist in the development of better security with less human interference. In the area of human anomaly activity, enormous progress has been made, enabling us to better serve the countless applications that are possible with it. Furthermore, research in similar fields, such as Activity Tracking, can significantly increase its usefulness in a variety of settings.

VI. FUTURE WORK

Future work may also sort out the divergence i.e., similar color problems like black bag and black background which guide to miss detections. Future developments may be integration of intensity and depth cues in the form of 3D accretion of evidence and blockage analysis. Spatial temporal features can be enlarged to 3-dimensional space for the upgrade of stranded object detection method for various composite environments. Thresholding based future works can enhance the staging of the surveillance system by using adaptable or hysteresis thresholding approaches. Few works have been also present for forsaken object detection from the multiple views seize by multiple cameras. To integrate these multiple views to infer the information about forsaken object can also be enhanced. There is a huge scope to trace forsaken object from videos captured by operating cameras.

VII. REFERENCES

[1]. Trong-Nguyen Nguyen, Jean Meunier: “Hybrid Deep Network for Anomaly Detection” Deep Neural Networks for Video Analysis that directly adapts the patch position as the target-1908.06347v1, Aug 2019.

[2]. Naimat Ullah Khan, Wanggen Wan: “A Review of Human Pose Estimation from Single Image”-978-1-53865195-7/18/ 2018 IEEE

[3]. Trong-Nguyen Nguyen, Jean Meunier,” Anomaly Detection in Video Sequence with Appearance-Motion Correspondence” – 1908.06351v, Aug 2019.

[4]. C. Lu, J. Shi, and J. Jia, “Abnormal event detection at 150 fps in MATLAB,” in Proceedings of the IEEE international conference on computer vision, 2013, pp.2720–2727.

[5]. B. Zhao, L. Fei-Fei, and E. P. Xing, “Online detection of unusual events in videos via dynamic sparse coding,” in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011, pp. 3313– 3320.

[6]. M. Sabokrou, M. Fathy, M. Hoseini, and R. Klette, “Real-time anomaly detection and localization in crowded scenes,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2015.

[7]. W. Luo, W. Liu, and S. Gao, “A revisit of sparse coding-based anomaly detection in stacked RNN framework,” in The IEEE International Conference on Computer Vision (ICCV), Oct 2017.

[8]. Youg Shean Chong, Yong Haur Tay: “Abnormal event detection in videos using spatiotemporal autoencoder”.

[9]. University Tunku Abdul Rahman, 43000 Malysia,2017.

[10]. Steven Diamond, Vincent Sitzzmann, Felix Heide, Gordon Wetzstein: “Unrolled optimization with deep priors”, Dec 2018.

[11]. Jefferson Ryan Medel, Andreas Savakis: “Anomaly Detection in Video Using Predictive Convolutional Long Short-Term” Rochester Institute of Technology, Rochester, New York, Jan 2016

[12]. M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, “Learning temporal regularity in video sequences,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 733–742.
[13]. Z. Zhan, J.-F. Cai, D. Guo, Y. Liu, Z. Chen, and X. Qu. Fast multiclass dictionaries learning with geometrical. directions in MRI reconstruction. IEEE

Cite this article as :

Jagruti Tatiya, Riya Makhija, Mrunmay Pathe, Sarika Late, Prof. Mrunal Pathak "Anomaly Detection for Video Surveillance ", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 8 Issue 3, pp. 82-89, May-June 2021. Available at
doi : https://doi.org/10.32628/IJSRSR21869
Journal URL : https://ijsrst.com/IJSRSR21869