IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101258

REVIEW ON INTELLIGENT SURVILLENCE SYSTEM

Ms. Chinju Poulose[1], Ashlin Francis Pereira[2], Sushith K S[3], Sreekumar CM[4],Yadhukrishna K Suresh[5]

Assistant Professor, Department of Computer Science And Engineering, Universal Engineering College, Vallivattom , Thrissur, India[1,]

B. Tech Student, Department of Computer Science And Engineering, Universal Engineering College, Vallivattom , Thrissur, India[2,3,4,5]

Abstract: In today's modern world, security and safety are major concerns. To be economically strong, a country must provide a safe and secure environment for investors and tourists. Closed Circuit Television (CCTV) cameras, on the other hand, are used for surveillance and monitoring activities. A large number of surveillance cameras are available in various locations, but all of them only record footage. To analyse these videos, a large amount of manpower is required, which is always undesirable due to time and labour waste. As a result, having a system that detects crime in real time and alerts the user is advantageous. This paper proposes a system for automatically detecting crime from surveillance camera footage. The system has been pre-trained to detect crimes in real time and provide both remote and offline alerts, which is useful in reducing the possibility of criminals escaping and quickly arrest them. If the missing person enters the premises, the system detects him or her.

Keywords: IoT, CCTV, CNN, MBF.

I. INTRODUCTION

This paper proposes a method to detect crimes and recognise missing persons in real time from surveillance camera footage automatically and notify the user about the incident. Machine learning and image processing techniques are used for the achievement of the system. The existing system of CCTV monitoring is done by humans. Surveillance camera monitors are most frequently tasked with general camera monitoring. They sit at a desk before a bank of monitor screens and conduct on-going non-specific assessment of live video feeds. The review of a surveillance camera's archived video is also sometimes required to determine if an event of interest was recorded. In this second task, monitors are asked to search for a specific event, person, or object. Again, the human monitor is often asked to watch hours of video. The monitoring difficulty is further compounded in that many criminal activities have subtle precursors that are easily overlooked when humans are tasked with monitoring multiple cameras.

II. THEORY

A. IotT

The Internet of Things (IoT) is a network of interconnected computing devices, mechanical and digital machines, objects, animals, or people that have unique identifiers (UIDs) and the ability to transfer data over a network without the need for human-to-human or human-to-computer interaction. Because of the convergence of multiple technologies, such as real-time analytics, machine learning, commodity sensors, and embedded systems, the definition of the Internet of Things has evolved. Embedded systems, wireless sensor networks, control systems, automation (including home and building automation), and other traditional fields all contribute to enabling the Internet of Things. In the consumer market, IoT technology is most closely associated with products related to the concept of the "smart home," which encompasses devices and appliances (such as lighting fixtures, thermostats, home security systems and cameras, and other home appliances) that support one or more common ecosystems and can be controlled by devices associated with that ecosystem, such as smart phones and smart speakers.

B. CCTV

CCTV (Closed Circuit Television) is a closed system that includes video cameras, display devices (monitors), and wired or wireless data networks that allow images to be transferred from video cameras to monitors. In addition to cameras and monitors, video surveillance systems frequently include other devices such as servers, disc storage, and client computers that allow for the storage and processing of video data. Video surveillance systems can also be linked to security and

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 10, Issue 12, December 2021
DOI: 10.17148/IJARCCE.2021.101258

other information systems. Video surveillance systems are intended to ensure security at protected sites, monitor personnel activities, and keep track of manufacturing processes, among other things.

C. CNN

A convolutional neural network (CNN) is a type of artificial neural network that is specifically designed to process pixel data in image recognition and processing. CNNs are powerful image processing, artificial intelligence (AI) systems that use deep learning to perform both generative and descriptive tasks, frequently utilising machine vision, which includes image and video recognition, recommender systems, and natural language processing (NLP). A neural network is a hardware and/or software system that is designed to mimic the operation of neurons in the human brain. Traditional neural networks are not well suited to image processing and must be fed images in low-resolution chunks. CNN's "neurons" are arranged more like those of the frontal lobe, the area in humans and other animals responsible for processing visual stimuli. The layers of neurons are arranged in such a way that they cover the entire visual field, avoiding the problem of piecemeal image processing that plagues traditional neural networks. A CNN employs a system similar to a multilayer perceptron that has been optimised for low processing requirements. A CNN has three layers: an input layer, an output layer, and a hidden layer with multiple convolutional layers, pooling layers, fully connected layers, and normalisation layers. The removal of limitations and increase in efficiency for image processing results in a system that is far more effective, simpler to train for image processing, and more efficient for natural language processing.

D. Maximal Biclique Finding (MBF) algorithm

Maximum biclique search, which finds the biclique with the most edges in a bipartite graph, is a fundamental problem with numerous applications in fields as diverse as E-Commerce, social analysis, web services, and bioinformatics. Unfortunately, due to the difficulty of the problem in graph theory, no practical solution to the problem in large-scale real-world datasets has been proposed. Because the search objective of maximum biclique search is two-dimensional, existing techniques for maximum clique search on a general graph cannot be used.

III. RELATED WORK

The paper[1] presents a novel algorithm for detection of certain types of unusual events. The algorithm is based on multiple local monitors which collect low-level statistics. Each local monitor produces an alert if its current measurement is unusual and these alerts are integrated to a final decision regarding the existence of an unusual event. This algorithm satisfies a set of requirements that are critical for successful deployment of any large-scale surveillance system. In particular, it requires a minimal setup (taking only a few minutes) and is fully automatic afterwards. Instead of trying to track objects, proposed algorithm monitors low-level measurements in a set of fixed spatial positions. The authors present results on videos from eight cameras in five different sites. Except for the dining hall and bus terminal sites, where the authors recorded with the own camera, all videos were obtained by direct recording from site-installed surveillance cameras

The paper[2] implies, as an Active research topic in computer vision, visual surveillance in dynamic scenes attempts to detect, recognize and track certain objects from image sequences, and more generally to understand and describe object behaviour. The prerequisites for effective automatic surveillance using a single camera include the following stages: modelling of environments, detection of motion, classification of moving objects, tracking, understanding and description of behaviours, and human identification. The authors provide detailed discussions on future research directions in visual surveillance, e.g., occlusion handling, combination of two-dimensional (2-D) tracking and 3-D tracking, combination of motion analysis and biometrics, anomaly detection and behaviour prediction, behaviour understanding and nature language description, content-based retrieval of surveillance videos, fusion of information from multiple sensors, and remote surveillance.

In paper [3], Author propose a fully unsupervised dynamic sparse coding approach for detecting unusual events in videos based on online sparse reconstructablity of query signals from an atomically learned event dictionary, which forms a sparse coding bases. Based on an intuition that usual events in a video are more likely to be reconstructable from an event dictionary, whereas unusual events are not, proposed algorithm employs a principled convex optimization formulation that allows both a sparse reconstruction code, and an online dictionary to be jointly inferred and updated. Proposed algorithm is completely unsupervised, making no prior assumptions of what unusual events may look like and the settings of the cameras.

Detection of irregular visual patterns in images and in video sequences is useful for a variety of tasks. In this paper [4] ,Author address the problem of detecting irregularities in visual data, e.g., detecting suspicious behaviours in video sequences, or identifying salient patterns in images . This paper [4] show applications of this approach to identifying saliency in images and video, for detecting suspicious behaviours and for automatic visual inspection for quality assurance Detecting suspicious behaviours or unusual objects is important for surveillance and monitoring The

DOI: 10.17148/IJARCCE.2021.101258

composition process is implemented as an efficient inference algorithm in a probabilistic graphical model, which accommodates for small spatio-temporal deformations between the query and the database

In paper [5],The authors introduces novel structural assumptions on the joint distributions to account for spatial and temporal locality of anomalies. The empirical composite scoring and ranking scheme asymptotically converges to the optimal decision rule for maximizing detection power subject to false alarm constraints. Sparse decomposition for each spatio-temporal scale can be viewed as a feature vector that feeds into the local KNN block. This is because Gn(·) as described in the previous section combines statistics over local neighbourhoods of a data sample and the ranking function produces a composite score for an entire random field.

The authors of paper[6], present a unique approach for detecting and localising anomalous behaviours in crowd films based on the Social Force concept. A grid of particles is set over the image and advected with the space-time average of optical flow for this purpose. The interaction forces of moving particles are approximated using the social force model by treating them as persons. After that, the interaction force is mapped onto the image plane to obtain Force Flow for each pixel in each frame. The normal behaviour of the crowd is modelled using Force Flow's randomly selected spatio-temporal volumes. Using a bag of words methodology, the proposed method classifies frames as normal or deviant. Interaction forces are used to pinpoint anomaly regions in abnormal frames.

According to the paper[7], the volume of internet videos has been rising at an exponential rate, necessitating a significant growth in the need for easier video searching. Important visual concepts tend to emerge repeatedly across a collection of movies with the same topic, and the frequency of visual co-occurrence can be used as a proxy to measure. The sparsity of co-occurring patterns adds an additional obstacle to video co-summarization: Hundreds to thousands of shots can be seen in a single video; however, there are generally only a few shots that appear in multiple recordings. The authors offer a unique Maximal Biclique Discovering (MBF) approach to address this problem, which formulates the problem as finding complete bipartite subgraphs that maximise total visual co-occurrence within a bipartite network. The authors offer a unique Maximal Biclique Discovering (MBF) technique to address this problem, which formulates the task as finding complete bipartite subgraphs that maximise total visual co-occurrence within a bipartite graphical representation of photos and videos.

In paper[8], The authors provide an unique framework for detecting anomalies in crowded settings. Three characteristics are highlighted as crucial in the building of a localised visual representation suited for anomaly detection in such scenes: 1) joint modelling of the scene's look and dynamics, as well as the ability to detect temporal and spatial anomalies.

Outliers are labelled as anomalies in the model for typical crowd behaviour, which is based on blends of dynamic textures. Temporal abnormalities are equated to low-probability occurrences, while spatial anomalies are dealt with using discriminant saliency. A new dataset of crowded settings, consisting of 100 video sequences and five well-defined abnormality categories, is used to undertake an experimental evaluation.

The goal of this paper[9] is to follow all occurrences in a typical surveillance video feed from a car park scene in real time, and to store the accompanying pixel data in a very efficient manner. The research presented in this paper focuses on generating precise type and behaviour classifications from monitored events represented as a series of bounding boxes. The classification of TYPE and BEHAVIOR of events within the video stream is critical to achieving this goal. Two sorts of event objects are explored in this car-park scenario: Person and Vehicle. The Training dataset is partitioned into four sets of events to build the four matching HMMs - vehicle-entering, person-entering, vehicle-exiting, and person-exiting - to develop the models for these. Both models associated with each object type will share the same set of states. The behavioural analysis results are used to determine the object classification.

Human observers are unable to monitor an excessive volume of surveillance video data. A computational method must be able to isolate actions in different parts of the frame while keeping structural information about the whole picture. The approach is tested on two extremely packed real-world situations from a subway station during rush hour1, according to the authors of paper[10]. Large numbers of pedestrians move in unpredictable ways with many occlusions in both data sets. By assessing the performance of prototypical distributions, the distance threshold dKL is determined empirically. The mixing coefficient is chosen empirically, and the temporal and geographical analysis is evaluated using the best performing distance threshold. Depending on the real-world situation, the length of training data ranged between 27 and 150 observations for each example depending on the real-world data available.

This paper[11] proposes a novel method using deep spatial temporal neural networks based on deep Convolutional Neural Network (CNN) for multimedia event detection. To sufficiently take advantage of the motion and appearance information of events from videos, proposed networks contain two branches: a temporal neural network and a spatial neural network. The temporal neural network captures motion information by Recurrent Neural Networks with the mutation of gated recurrent unit. The spatial neural network catches object information by using the deep CNN, to encode the CNN features as a bag of semantics with more discriminative representations. Author propose an effective model called deep spatial temporal networks for multimedia event detection. The temporal net implemented by Recurrent Neural Networks with the MUTation of gated recurrent unit captures motion information, and the spatial net catches the object information by encoding the Convolutional Neural Network features with bag of semantics. object information by using the deep CNN, to encode the CNN features as a bag of semantics with more discriminative representations. Author propose an effective model called deep spatial temporal networks for multimedia event detection. The temporal net implemented by Recurrent

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 10, Issue 12, December 2021
DOI: 10.17148/IJARCCE.2021.101258

Neural Networks with the MUTation of gated recurrent unit captures motion information, and the spatial net catches the object information by encoding the Convolutional Neural Network features with bag of semantics.

In paper[12], The authors increasingly capture large amounts of video data, a trend that is likely to accelerate with new devices like Google Glass. Auto-Awesome works best when the event videos are short and contain only highlights It is not clear how it can handle raw personal videos typically a few minutes long. Facebook's new Look Back feature provides similar functionality, but focused on photos and your most popular posts instead of videos. These applications motivate the importance of research in automatic video editing. The authors conduct experiments on a newly created YouTube highlight dataset harvested by the system automatically. For analysis and evaluation purposes, the authors have labelled the dataset using Amazon Mechanical Turk. The authors first give details of the implementation such as feature representation, parameter setting, etc. The authors report quantitative and qualitative results on the novel YouTube highlight dataset.

This paper[13] present a video summarization approach for egocentric or "wearable" camera data. Given hours of video, the proposed method produces a compact storyboard summary of the camera wearer's day. In contrast to traditional keyframe selection techniques, the resulting summary focuses on the most important objects and people with which the camera wearer interacts. The goal of video summarization is to produce a compact visual summary that encapsulates the key components of a video. Its main value is in turning hours of video into a short summary that can be interpreted by a human viewer in a matter of seconds. Automatic video summarization methods would be useful for a number of practical applications, such as analyzing surveillance data, video browsing, action recognition, or creating a visual diary.

In this paper[14]. When only limited hardware resources are available, the authors offer a discriminative video representation for event detection over a large size video dataset. The goal of this paper is to use deep Convolutional Neural Networks (CNNs) to progress event identification in situations where existing CNN toolkits can only extract frame level static descriptors. This study provides two contributions to CNN video representation inference. First, while average and maximum pooling have long been the usual ways to aggregating frame-level static information, they demonstrate that efficiency may be greatly improved by using the right encoding method. Second, as the frame descriptor, they propose utilising a set of latent concept descriptors, which adds visual information while being computationally affordable. Theintegration of the two contributions results in a new state-of-the-art performance in event detection over the largest video datasets.

In this paper[15], The author proposes a motion recognition algorithm based on a revolutionary motion feature extraction method. Auto-correlations of space–time gradients of three-dimensional motion form in a video series are used in the feature extraction method. They devised a method for extracting motion information, resulting in a motion recognition system that is both effective and fast. Unlike traditional bag-of-features, which characterises the motion sparsely, the motion is recognised in the framework of bag of-frame-features, which can sufficiently extract the motion characteristics in a computationally efficient manner. The suggested feature extraction approach is based on local auto-correlations of space–time gradients and captures the geometric properties of space–time motion shape, such as curvatures.

Real-time detection of gait events play a vital role in movement dependent control applications such as rehabilitation for lower limb amputations. It also helps in determination of spatio-temporal and kinematic parameters. Gyroscopes, inertial sensors, magnetometers and foot sensors are popular in the detection of gait events. They need to be mounted carefully, or foot should be placed specifically on foot pressure during detection. This study in paper[16] presents a framework for automated detection of gait events from conventional videography using passive markers at Robotics And Machine Analytic Laboratory (RAMAN Lab). The proposed Passive maker based Gait event detection (PMGED) algorithm automatically detects heel strike (HS) and toe-off (TO); the timing of stance and swing phase; the number of the gait cycle. Ten healthy subjects are considered to evaluate the robustness and reliability of proposed algorithm. The method is comparable when evaluated against human expert detection.

This paper[17]. describes a system that was built with two primary goals in mind. First, based on the contents of continuous video sequences, detect predetermined events in real time. Second, to aid in the discovery of certain segments of recorded video. The separation of object extraction and event detection components, as well as their bridging utilising a contents description standard, is a major aspect of this research (MPEG-7). The problem is broken down into three main phases. The first part involves extracting the objects of interest, together with a set of properties (such as position, size, and orientation), and creating an annotation file using the MPEG-7 standard; the second phase involves tracking these items through time. In order to gain some relevant information, these raw tracking data must be filtered. Finally, Finally, based on the clean tracking data, it may find events almost as they are happening

Detection of motion patterns in video data can be significantly simplified by abstracting away from pixel intensity values towards representations that explicitly and compactly capture movement across space and time. In paper[18] a novel representation that captures the spatiotemporal distributions of motion across regions of interest, called the "Direction Map," abstracts video data by assigning a two-dimensional vector, representative of local direction of motion, to quantized regions in space-time. Methods are presented for recovering direction maps from video, constructing direction map templates (defining target motion patterns of interest) and comparing templates to newly acquired video (for pattern detection and localization).

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 10, Issue 12, December 2021
DOI: 10.17148/IJARCCE.2021.101258

The Event detection method from a video surveillance has received much attention in the image processing. In this paper[19], an overview of a new approach for event detection from video surveillance system based on incremental learning is presented. In this approach, each event is modelled by a set of states, and each state is represented by a learning model containing a positive class (event) and a negative class (non-event). Experiments on real image sequences have shown encouraging results.
Speedy abnormal event detection meets the growing demand to process an enormous number of surveillance videos. Based on inherent redundancy of video structures, an efficient sparse combination learning framework is proposed in paper[20]. It achieves decent performance in the detection phase without compromising result quality. The short running time is guaranteed because the new method effectively turns the original complicated problem to one in which only a few costless small-scale least square optimization steps are involved. This method reaches high detection rates on benchmark datasets at a speed of 140-150 frames per second on average when computing on an ordinary desktop PC using MATLAB.

IV. CONCLUSION

This system can detect crimes and suspects from CCTV footages in real time. This helps to prevent the most of the crimes and also helps to capture suspects and missing persons. It is an cost effective system that can also be installed in existing CCTV surveillance system. The missing persons are easily been found without any delay

ACKNOWLEDGMENT

We would like to take this opportunity to express our gratitude to everyone who has assisted us, directly or indirectly, in completing our work. We would like to express our heartfelt gratitude to **Dr. R Sreeraj -HOD**, Computer Science and Engineering - for his invaluable advice and encouragement. We are especially grateful to our guide and supervisor, **Ms. Chinju Poulose** -Assistant Professor, Computer Science and Engineering, for providing me with helpful suggestions and critical feedback during the preparation of this paper. We would like to express our heartfelt gratitude to the entire faculty of the Computer Science and Engineering department for their assistance and suggestions in shaping our work into what it is today. We thank our parents and friends for the mental support provided during the course of our work at the times when our energies were the lowest.

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 30(3), 555- 560 (2008)
[2]. W. Hu, T. Tan, L. Wang, and S. Maybank, “A survey on visual surveillance of object motion and behaviors,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 34, no. 3, pp. 334–352, Aug 2004.
[3]. B. Zhao, L. Fei-Fei, and E. P. Xing, “Online detection of unusual events in videos via dynamic sparse coding,” in CVPR, 2011, pp. 3313–3320.
[4]. O. Boiman and M. Irani.” Detecting irregularities in images and in video”. In ICCV, pages 462–469, 2005
[5]. Saligrama, V., and Chen, Z. 2012.”Video anomaly detection based on local statistical aggregates”. In CVPR, 2112–2119. IEEE.
[6]. R. Mehran, A. Oyama, and M. Shah. ”Abnormal crowd behavior detection using social force model”. CVPR, 2009
[7]. W.-S. Chu, Y. Song, and A. Jaimes, “Video Cosummarization: Video Summarization by Visual Cooccurrence,” in CVPR, 2015
[8]. Mahadevan, V.; Li, W.; Bhalodia, V.; and Vasconcelos, N. 2010. Anomaly detection in crowded scenes. In CVPR, 1975–1981. IEEE
[9]. P. Remagnino and G. A. Jones. ”Classifying surveillance events from attributes and behaviour”. In BMVC, 2001.
[10]. Kratz, L., and Nishino, K. 2009. “Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models”. In CVPR, 1446–1453. IEEE
[11]. V. Ramanathan, K. Tang, G. Mori, and L. Fei-Fei, “Learning Temporal Embeddings for Complex Video Analysis,” arXiv preprint arXiv:1505.00315, 2015
[12]. M. Sun, A. Farhadi, and S. Seitz, “Ranking Domain-specific Highlights by Analyzing Edited Videos,” in ECCV, 2014.
[13]. Lee, Y.J., Ghosh, J., Grauman, K. “Discovering important people and objects for egocentric video summarization”. In: CVPR (2012)
[14]. Zhong wen Xu, Yi Yang, and Alexander G Hauptmann, “A discriminative cnn video representation for event detection,” in CVPR, 2015, pp. 1798–1807
[15]. T. Kobayashi and N. Otsu, “Motion recognition using local auto-correlation of space–time gradients,” Pattern Recognition Letters, vol. 33, no. 9, pp. 1188–1195, 2012.
[16]. C. Prakash, R. Kumar and N. Mittal, ”Automated detection of human gait events from conventional videography,” 2016 International Conference on Emerging Trends in Communication Technologies (ETCT), Dehradun, 2016, pp. 1-4.
[17]. J. Silla, A. Albiol, J. M. Mossi and L. Sanchis, “Automatic video annotation and event detection for video surveillance,” 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009), London, 2009, pp. 1-5.
[18]. J. Gryn, R. Wildes, and J. Tsotsos. Detecting Motion Patterns via Direction Maps with Application to Surveillance. In IEEE Workshop on Motion and Video Computing, pages 202–209, 2005
[19]. Y. Aribi, A. Wali and A. M.Alimi, ”An intelligent system for video events detection,” 2013 9th International Conference on Information Assurance and Security (IAS), Gammarth, 2013, pp. 108-113.
[20]. C. Lu, J. Shi, and J. Jia, “Abnormal event detection at 150 fps in matlab,” in Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, 2013, pp. 2720–2727.