Review: Video Analytics Technologies Available for Surveillance Systems

Utkarsha Mokashi
Department of Computer Engineering
KJSIEIT, University of Mumbai
Mumbai, India
utkarsha.mokashi@somaiya.edu

Aarush Dimri
Department of Computer Engineering
KJSIEIT, University of Mumbai
Mumbai, India
aarush.dimri@somaiya.edu

Hardee Khambhla
Department of Computer Engineering
KJSIEIT, University of Mumbai
Mumbai, India
hardee.khambhla@somaiya.edu

Pradnya Bhangale
Department of Computer Engineering
KJSIEIT, University of Mumbai
Mumbai, India
pyb@somaiya.edu

2022 5th International Conference on Advances in Science and Technology (ICAST) | 978-1-6654-9263-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAST55766.2022.10039583

***Abstract*—Smart Surveillance Systems are becoming an important aspect of our lives, reducing man labour and additionally increasing the accuracy of detection by reducing false positives. Specifically for an ATM, Surveillance system is very crucial because of the transactions happening being sensitive along with that drop-box containing confidential documents like cheques and bank forms. Hence, there is a need to develop a fool-proof system which can handle a lot of load and perform various surveillance tasks. Moreover, the systems also need to have network security to protect the data from being illegally traced and changed. In this paper, we will be reviewing and comparing various smart surveillance system methods which involve various technologies.**

Keywords— Surveillance, monitoring, video analytics, azure video indexer, CCTV, IoT, machine learning

I. INTRODUCTION

Video analytics which is also known as video content analysis uses algorithms to process videos to perform tasks like identifying moving objects, sentiment analysis etc. In the Surveillance system, video analytics automates the job of watching hours of CCTV footage to find threats or risks. It helps reduce the load on security personnel during long shifts to stay alert by automating their task or at least most of their task.

Surveillance is the monitoring of activities, people and behaviours to check if there are any thefts and malpractices taking place, or even if some risks are present. Surveillance system makes use of cameras, monitors and various other devices.

Video cameras are used to monitor high-security areas. The surveillance of the video footage is done by a supervisor who monitors the activities from the video footage acquired and detects unusual and suspicious activities. Any individual with malicious intentions is identified and if need be, an alert is generated and action is taken.

Our paper aims to compare various video analytics technologies which can be used for surveillance systems and map the advantages and limitations of each of them in order to come up with a methodology for a better design of a surveillance system.

II. RELATED WORK

[1] relates to an architectural approach to deep learning that makes use of Convolutional Neural Networks (CNN), which are effective in extracting features and as compared to its predecessors it does this without requiring human support. In [3] an approach of LSTM tesseract-based optical character recognition algorithm is used for car plate detection and the Resnet-34 model is used for facial recognition.

[2] describes a VFC approach on an edge computing network. The proximity to the data processing is a benefit of edge computing, but there are also safety and data protection concerns. This method is used to manage real-time heavy-load AI applications, such as smart surveillance systems. [4] is regarding a surveillance system in tunnels to detect fires, people out of the car and if the car has stopped or is driving in the wrong direction. This paper mentions using Faster RCNN and Support Vector Machine (SVM). Faster RCNN uses RPN along with Fast RCNN leading to a faster computation time.

In [5], Raspberry Pi used to create small networks with cameras. It is a compact model and doesn't need large cabling or routing devices. It is a compact model and doesn't need large cabling or routing devices. [6] goes over various technologies that can be used to build a small scale but robust Ad-hoc network in areas of less reachability of the internet. These technologies include Zigbee, Bluetooth and Wireless USB. The main focus of this paper is on security of such systems and resource management, and behaviour of the network in case of power failure. Its advantage is that it can be used in remote locations with less accessibility since it forms an internal network without actually using routers or the internet. It will keep the network up, even if one of the sensors dies. [8] describes a methodology for anomaly detection which is based on a global object map. This paper is very well represented with a lot of graphs and statistical data, making it easy to comprehend and understand. The normal accuracy for caution as well as warning is 99.99%, which is impressive.

The table given below highlights the limitations in above reviewed papers:

TABLE I. HIGHLIGHTS OF IMITATIONS OF REVIEWED PAPERS

Paper Referenced	***Limitations***
Future Predicting Intelligent Camera Security System [1]	-Machine Learning models are not strong and have a 0.3715 loss. -Only alerts the supervisor, not a completely automated system. -Classifies activities as normal, malicious, and suspicious only.
Enabling Real-Time AI Edge Video Analytics [2]	The ability of the architecture to use a migration mechanism to handle node failures is not taken into consideration.
Smart Surveillance and Tracking System using Resnet and Tesseract OC [3]	Accuracy varies in real life due to many factors such as changing lighting, camera angle, weather, the number of cameras present, and other such factors.
An Application on Deep Learning for Automatic Detection of Unexpected Accidents Under Bad CCTV Monitoring Conditions in Tunnels [4]	-Due to the scarcity of Fire objects in the trained samples, false detection is greatly possible in untrained samples. -Securing a variety of pictures as well as Fire and Person items is essential to boosting the system's reliability, which is lacking in the current system that they use. -Requires 10 seconds to detect accidents -System reliability needs to be improved
Internet of Things: CCTV Monitoring by Using Raspberry Pi [5]	This would probably not be feasible because it would require setting up a microcontroller for each camera/location and the raspberry pi is not a strong enough CPU to run Deep learning algorithms.
Ad-hoc Connections of Miscellaneous sensors in a CCTV System [6]	Requires the use of multiple sensory devices, and each ad-hoc network, has to be connected to the main network to transport the data back to the server for processing
Hierarchical and Modular Surveillance Systems in ITS [7]	The paper has only considered moving traffic for its HMSS system analysis. Even though it is good at detecting speeds, it cannot retain the features of objects in detail.
Anomaly Detection Algorithm Based on Global object Map for Video Surveillance System [8]	-The implementation is done for a specific test bed position. -The rate of abnormal detection is 86.6% which can be improved
Sequential and Patch Analyses for Object Removal Video Forgery Detection and Localization [9]	Patch analysis models video sequences as both anomalous and normal patches, distinguishing them by detecting their distribution. As patch size grows, so does the false negativity rate.

Overall, the research gap observed was that there is no such surveillance system that tackles all of the requirements like sentiment analysis, tampering detection, abnormal activity, etc in a single system application. No system focuses on the network security aspect along with the AI algorithms to process the data, since cameras are endpoints in the network, they are most vulnerable and they need to be encrypted as well as the entire network needs to be secured for a secure transmission as well as secure processing. Some of the systems which we studied still focus on human handling decisions and surveillance from the camera footage, while sometimes specifically the ATM footage is not really reviewed till there is a need to check the footage because of some mishap or crime report. If humans are overlooking and doing the work a model should be doing, there is a high chance of false positives because of a variety of reasons. Furthermore, for the most part of the data, the data is stored on hardware devices, where there is a risk of the data getting deleted or lost, and it also takes up a huge memory leading to replenishing with new data and deleting the old data, a better solution and a research gap understood was that cloud storage which is a fast and efficient way to store huge amounts of data with less risk of deleting is not used by a lot of surveillance systems.

III. REVIEWING THE EXISTING TECHNOLOGIES

Let us delve deeper into the technologies that are already available to us to implement video analytics. On reviewing the technologies used in the research papers referred to by us, we have created a table for easy comparison between the existing technologies. The technologies we will be reviewing for video analytics include Deep Learning (SVM and Faster RCNN for moving vehicles), Edge Computing, Artificial Intelligence (ResNet, LSTM), Raspberry Pi, Fast Motion Imaging and Microsoft Azure's Video Indexer Tool. Each tool and technology which has been studied has some advantages as well as some limitations in its application as a video analytics tool.

A. Comparison between the existing technologies

The comparison between existing technologies is listed

TABLE II. COMPARISON BETWEEN THE EXISTING TECHNOLOGIES

Technology Used	*Methodology Used*	*Advantages and Limitations*
Deep Learning (Faster RCNN), SVM along with RCNN for moving vehicles	Faster RCNN uses a series of still images. It proposes a region proposal network (RPN) that generates varied (scales and aspect ratios) proposals. It uses anchor boxes to map and reference various objects. Faster RCNN leads to a reduction of computational time.	• Faster computational time, as RPN is also used along with Fast RCNN, training time also decreases, • Does not look after the security part and model might need a diverse dataset.
Edge Computing	It is a real time AI model which distributes the applications into many decomposed Virtual Functions. It enables low-cost and real-time surveillance analytics. Caching mechanism enables the model to process heavily loaded AI applications in the real time.	• High proximity with the data being processed limits the number of trips to the Cloud and ensures data protection and safety issues. • Data is being processed individually at each node, there is no central database
Artificial Intelligence (Resnet, LSTM)	The centralised system communicates with the cameras distributed remotely using their IP addresses and contains a detection and a recognition system (which use Resnet-34 and Tesseract OCR). On object recognition, the system stores the location and the timestamp in the database which is used for tracking the object on the map using the Object Tracking Block	• The individual persons and distinct vehicles captured are traced out on one map which is very beneficial in scenarios for large organisations such as banks. • The challenges differ in real-life. The angle of the camera influences the view and it will vary from camera to camera as there are blind spots
Raspberry Pi	A web server can be realised using a raspberry pi as it is a low-power consumption device. It is then integrated with a webcam. It works for the client's monitor through the browser which is defined according to the IP address obtained from the server.	• Modularity, Open Design, Open Source • Less powerful than edge computing, it will increase the cost as each location of the camera will need a microcontroller, raspberry pi is not a strong enough CPU to learn deep learning models.
Fast Motion History Image (MHI)	This procedure generates an object mask, a binary image, and locates the moving object in the input frame. The input frames are used to maintain the history matrix, according to the values of the incoming object masks.	• Faster way to compute behaviour recognition, reduces time by 56%. • Limited only to predefined behaviours and not applicable to unexpected and unique situations
Azure Video indexer	Azure Video Indexer is a cloud application. It uses deep search technology which enables it to use the extracted insights from one video to enhance the search algorithm on the other videos It has features such as face detection, Labels identification, Scene segmentation and Shot detection that make it favourable to be used for various use cases.	• Microsoft Azure's Video Indexer tool has multiple diverse attributes and services that can be applied to analyse a video footage and can provide deep insights. • It is not an open-source tool. It is not useful when it comes to detecting physical tampering with the cameras.
HMSS in ITS	This system is a novel method of surveillance for videos. HMSS is a 4 layered structure, the layers are: to see the world and feel the world and analyse the world along with that understanding the world. They also help to produce generic vehicle data including rate of flow, density, and average speed calculations. Additionally, this video surveillance system can record pedestrian visual data.	• The four HMSS levels are both hierarchical and flexible in structure; there is no functional overlap between them, and any level's modules can be freely chosen and combined. • It requires high resolution cameras to capture quality images of fast-moving cars. HMSS is a model developed only for ITS systems, it is not very efficient for other types of surveillance
Sequential and Patch Analysis	The methodology followed included dividing the video footage into frames and then analysing the frames at pixel level as well as video level. It follows a statistical approach towards detecting objects in video footage.	• It has lower computational complexity and provides a more robust framework than compressed low-quality videos. • The accuracy in object detection is low.

IV. PROPOSED SOLUTION

After visiting and reviewing each of the technologies for video analytics, we have proposed a method that consists of several technologies applied together to make a sustainable and efficient method of analysing video footage using AI algorithms and a way to store and manage that data.

A. Block Diagram

Figure 1 shows the proposed system for Video Analytics that maps different functions and tasks important for working of the model to the corresponding technologies.

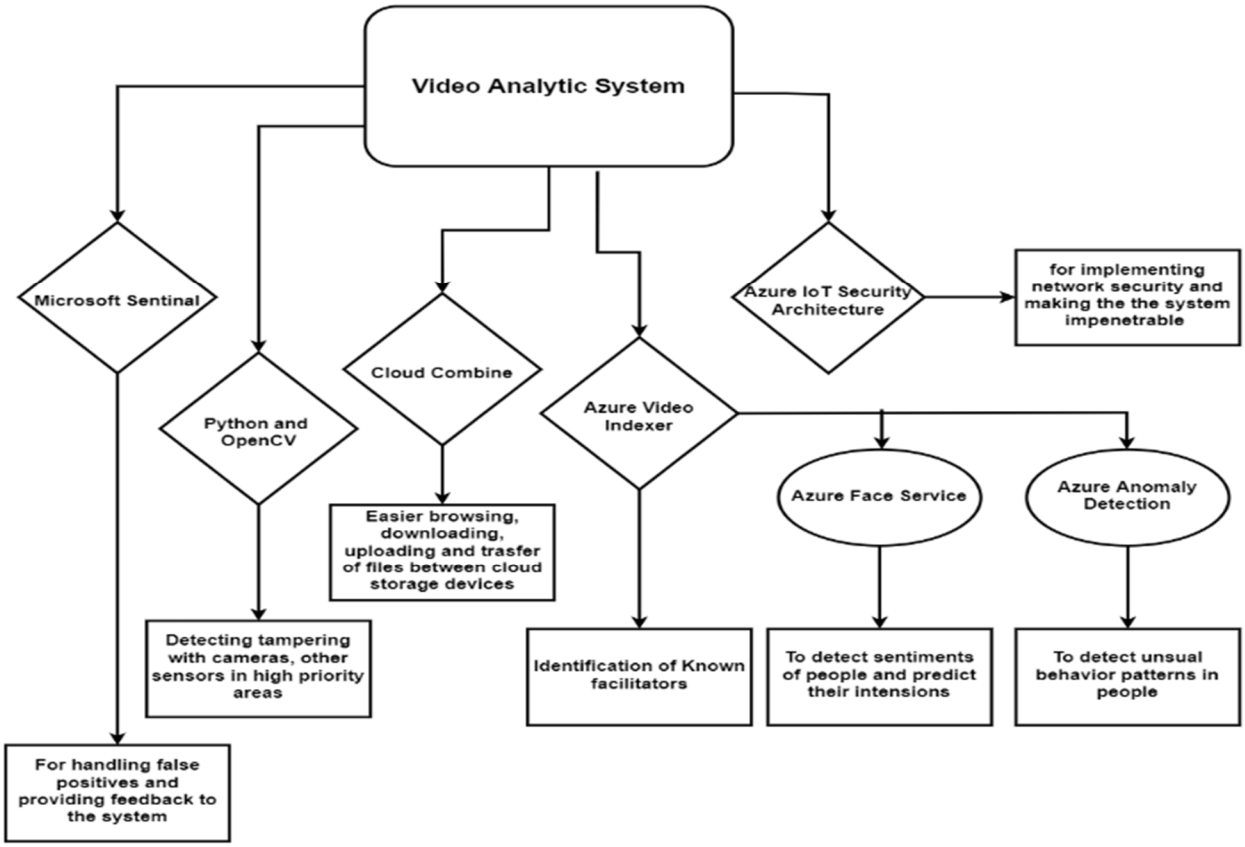

Fig. 1. Wireframe diagram

B. Functionality of the system

1) Azure Video Indexer: The insights of Azure Video Indexer can be applied to many scenarios. It is mainly used for handling 3 different aspects of video analysis which are widely used in most of the use cases
   - Azure face service: This aspect of Video indexer trains a model for a specific use case. It then recognizes faces in the video based on the trained model. Can be employed to detect sentiments of people and predict behaviour
   - Azure Anomaly detection: It uses batch validation or real-time inference to monitor and detect anomaly is a series of data
   - Identification of known facilitators, i.e., identifies visual objects and actions that are labelled before.
2) Azure IoT architecture: It is used to implement network security and make the connection between cameras and the server secure. The Azure IoT Edge technology addresses the risk in moving data and analytics to the intelligent edge module, which does the processing
3) Cloud Combine: A cloud-based solution for easy downloading, uploading and transfer of files between devices and applications. Can be used to store video footage as well as model data
4) Python lib and OpenCV: Detecting any physical tampering with cameras and other sensors in high priority areas and generating an alert
5) Microsoft Sentinel: It is a scalable, cloud-native solution used for detecting false positives and providing feedback to the system. It provides 2 effective services: SIEM: Security information and event management and SOAR

V. CONCLUSION

In the above sections, we reviewed various technologies for video analytics. We understood the advantages and limitations of each technology and understood the research gaps which can be worked upon. Working on the research gaps, we have proposed an efficient system that consists of several individually powerful technologies combined together to make a sustainable and efficient method of analysing video footage and finding useful insights from the same. The proposed model is tailored to handle various Surveillance needs an ATM might have, like tampering with cameras, abnormality detection etc. The system is said to reduce the human burden and hence help produce an effective result and alert on the spot if there is any mishap. Along with that we have also proposed to add a security feature, so the

data collected by the cameras cannot be tampered with. The backbone of our system is Microsoft Azure and the numerous tools it provides. It has a few open source tools and a cloud storage solution to efficiently store and manage data. With this system, we aim to revolutionise ATM surveillance and lead towards an automated, secure future.

REFERENCES

[1] M. Abraham, N. Suryawanshi, N. Joseph, D. Hadsul, "Future Predicting Intelligent Camera Security System," presented at the 2021 International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, Feb. 11-12, 2021. Available: https://ieeexplore.ieee.org/document/9399597

[2] V. Tsakanikas, T. Dagiuklas, "Enabling Real-Time AI Edge Video Analytics," presented at the ICC 2021 - IEEE International Conference on Communications, Montreal, QC, Canada, Jun. 14-23, 2021. Available: https://ieeexplore.ieee.org/document/9500902

[3] C. Sonavane, P. Kulkarni, O. Podey, P. Rewane, "Smart Surveillance and Tracking System using Resnet and Tesseract-OCR," presented at the 2021 IEEE Pune Section International Conference (PuneCon), Pune, India, Dec. 16-19, 2021. Available: https://ieeexplore.ieee.org/document/9686493

[4] K. B. Lee, H. S. Shin, "An Application of a Deep Learning Algorithm for Automatic Detection of Unexpected Accidents Under Bad CCTV Monitoring Conditions in Tunnels," presented at the 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), Istanbul, Turkey, Aug. 26-28, 2019. Available: https://ieeexplore.ieee.org/document/8876906

[5] E. Rohadi, S. Adhi Suwignjo, M. Candra Pradana, A. Setiawan, I. Siradjuddin, F. Ronilaya, Amalia, R. A. Asmara, R. Ariyanto, "Internet of Things: CCTV Monitoring by Using Raspberry Pi," presented at the 2018 International Conference on Applied Science and Technology (iCAST), Manado, Indonesia, Oct. 26-27, 2018. Available: https://ieeexplore.ieee.org/document/8751612

[6] K. Kraus, R. Reda, "Ad-hoc connections of miscellaneous sensors in a CCTV system," presented at the 2008 23rd International Symposium on Computer and Information Sciences, Istanbul, Turkey, Oct. 27-29, 2008.
Available: https://ieeexplore.ieee.org/document/4717969

[7] G. Yuan, X. Zhang, Q. Yao and K. Wang, "Hierarchical and Modular Surveillance Systems in ITS," in IEEE Intelligent Systems, vol. 26, no. 5, pp. 10-15, Sept.-Oct. 2011, doi: 10.1109/MIS.2011.88.
Available: https://ieeexplore.ieee.org/document/60358

[8] H. C. Shin, J. Chang and K. Na, "Anomaly Detection Algorithm Based on Global Object Map for Video Surveillance System," 2020 20th International Conference on Control, Automation and Systems (ICCAS), 2020, pp. 793-795, doi: 10.23919/ICCAS50221.2020.9268258.
Available: https://ieeexplore.ieee.org/document/ 9268258

[9] M. Aloraini, M. Sharifzadeh and D. Schonfeld, "Sequential and Patch Analyses for Object Removal Video Forgery Detection and Localization," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 3, pp. 917-930, March 2021, doi: 10.1109/TCSVT.2020.2993004.
Available: https://ieeexplore.ieee.org/document/9089005

[10] Yi-Ching Liaw, Wei-Chih Chen, Tsung-Jen Huang, "Video Objects Behaviour Recognition Using Fast MHI Approach," presented at the 2010 Seventh International Conference on Computer Graphics, Imaging and Visualisation, Sydney, NSW, Australia, Aug. 07-10, 2010.
Available: https://ieeexplore.ieee.org/document/5576204

[11] G. Ananthanarayanan, P. Bahl, P. Bodík, K. Chintalapudi, M. Philipose, L. Ravindranath, S. Sinha, "Real-Time Video Analytics: The Killer App for Edge Computing," Computer, vol. 50, issue 10, pp. 58 - 67, Oct. 2017 [Online]. Available: https://ieeexplore.ieee.org/document/8057318

[12] Yi-Zeng Hsieh, Yu-Lin Jeng, "Development of Home Intelligent Fall Detection IoT System Based on Feedback Optical Flow Convolutional Neural Network," IEEE Access, vol. 6, pp. 6048 - 6057, Nov. 2017. [Online],
Available: https://ieeexplore.ieee.org/document/8101471