

Abnormal Activity Recognition for Visual Surveillance

Subjects: Computer Science, Artificial Intelligence

Contributor: Lakshay Goyal , Musrrat Ali , Chandra Mani Sharma , Sanoj Kumar

Due to the ever increasing number of closed circuit television (CCTV) cameras worldwide, it is the need of the hour to automate the screening of video content. Still, the majority of video content is manually screened to detect some anomalous incidence or activity. Automatic abnormal event detection such as theft, burglary, or accidents may be helpful in many situations. However, there are significant difficulties in processing video data acquired by several cameras at a central location, such as bandwidth, latency, large computing resource needs, and so on.

visual surveillance

edge computing

activity recognition

anomaly detection

surveillance systems

artificial intelligence

1. Introduction

A CCTV-based system can be used to monitor various events at many public places. Imbibing intelligence and automation in processing video captured by these systems can be useful in many ways, ranging from traffic monitoring to vandalism detection. Prompt and timely actions can be taken as soon as an abnormal event is detected in the live video streams. Visual surveillance may encompass a number of tasks. It has applications in moving object detection [1], abandoned object detection [2], pedestrian detection [3], car make or model detection that may be helpful in accident sites and traffic violations [4], socio-cognitive behaviors of crowds [5], anomaly detection in road traffic [6], shop lifting [7], etc. Object detection has been one of the most important phases in a typical vision-based surveillance system. It is the first step in extracting the most useful pixels from a video feed. The study, presented in [1], looks at a variety of related methodologies, significant obstacles, applications, and resources, including datasets and web-sources. When video sequences are collected using IP cameras, the work provides a complete review of the moving object task suitable for a number of visual surveillance scenarios. To prevent bomb blasts from causing environmental and economic damage, automated smart visual surveillance is needed to keep a watch on the open spaces and infrastructures and to identify the items left behind in public places [2]. Commonly used approaches to identify abandoned objects are based on background segmentation for static object identification, feature extraction, object classification, and activity analysis [2]. Pedestrian detection and tracking have been an important function in traffic and road safety surveillance systems [6]. Traditional models have trouble dealing with complexity, turbulence, and the presence of a dynamic environment, but intelligent analytics and modeling can help overcome these difficult issues [3]. Protection of high rise civil engineering structures and human occupants from strong winds and earthquakes is crucial to human life, the economy, and the environment. The problem of vibration suppression of structures is an active, vast, and growing research field among

mechanical, control, and civil engineers. The design of a vibration controller with high performance for passive, semi-active, active, and hybrid control of building structures is a challenging task due to model uncertainties and external disturbances. The main objective of a structural control system is to reduce the vibration of the high rise building structures when external disturbances such as strong winds, earthquakes, or heavy dynamic loads act on them.

In previous works, researchers have developed many interesting computer-based systems and techniques for various tasks associated with visual surveillance. However, these systems are either one-node heavy systems or rely on cloud resources for analytics. It means when connecting more than one camera, the data streams are sent to a cloud server for data analytics. It requires latency and bandwidth issues apart from heavy investments. In recent times, with the advent of the internet of things (IoT) and edge computing, the focus has shifted to performing computation as close to the source as possible. The edge computing model envisages a major part of computation happening on the edge of the network, i.e., the node itself. This requirement raises many concerns for performing video analytics on the edge devices due to the limited computation resources, memory, and power availability.

2. Visual Analytics and Surveillance Systems

Understanding human behavior is essential for a variety of present and future interactions among people and smart systems and entities [5]. For instance, with prevalent CCTV-based surveillance systems, such knowledge might aid in detecting (and resolving as soon as feasible) incidents of hazardous, hostile, or just disruptive conduct in public meetings. Intense amounts of video data have prompted efforts to classify video information into categories such as human activities and complicated events. A growing body of work focuses on calculating effective local feature descriptors from spatio-temporal volumes [8]. Human activity recognition in videos is an important task in visual surveillance. One rationale behind such a classification is to detect abnormal activities in videos. Mliki et al. [9] adapted convolutional neural networks, which are generally used for classification, to identify humans. Furthermore, the categorization of human activities is performed in two ways: an immediate classification of video sequences and a complete classification of video sequences. They used the UCF-ARG dataset. One-shot learning (OSL) is becoming popular in many computer vision tasks, including action recognition. Contrary to conventional algorithms, which rely on massive datasets for training, OSL seeks to learn information about item classes with the help of one or a few training samples. The work described in [10] provides a deep learning model that can categorize and locate activities identified with the help of a single-shot detector technique employing the bounding box that has been deliberately trained to recognize common and uncommon actions for security surveillance applications.

Wassim et al. [11] used a feature approach to detect abnormal activities in crowded scenes on the UCSD anomaly detection dataset. The first category is motion features calculated using optical flow; the second is the size of moving individuals within frames; and the third is motion magnitude. Nawaratne et al. [12] described an incremental spatiotemporal learner (ISTL) addressing some of the challenges in anomaly localization and classification in real-time surveillance applications. ISTL is the unification of fuzzy aggregation with active learning in order to continuously learn and update the distinction between an anomaly and the normality that emerges over time.

Anomaly detection using sparse encoding has shown encouraging results. Zhou et al. [13] used three joint neural architectures called “AnomalyNet” for detecting anomalies in a video stream. Human aberrant behavior can occur at various timelines and can be divided into two categories: short-term and long-term. A uniform pre-defined timescale seems insufficient to represent a variety of abnormalities that occur throughout varying time periods [4]. Therefore, a useful approach for detecting anomalous human behavior is multi-timescale trajectory prediction, as proposed in the work of Rodrigues et al. [14]. To address the issue of fewer negative examples, the technique employs an unsupervised learning method that uses the spatiotemporal autoencoder to locate and extract the negative samples, containing anomalous behaviors, from the dataset. On this foundation, a spatiotemporal convolutional neural network (CNN) with a basic structure and minimal computational complexity has been given in [15]. More atypical human activity recognition systems are proposed in [16][17][18]. Beddiar et al. [19] and Pareek et al. [20] provide surveys on vision-based human activity recognition, discussing some of the recent breakthroughs, challenges, datasets, and emerging applications of the concept.

In activity recognition [21], optical flow refers to the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between the observer and the scene [22][23]. Optical flow is often used to track and understand the movement of objects in video sequences. In the context of activity recognition, optical flow can be employed to analyze the dynamics and motion patterns of human activities. By tracking the flow of pixels between consecutive frames, it becomes possible to extract information about the direction and speed of motion, which can contribute to the recognition of various activities such as walking, running, or gestures in a video [21][22][23].

3. Edge Computing for Visual Surveillance

Edge computing is a model where computing happens locally. It is performed so as to minimize the reliance on servers. There are local computing nodes performing computation and with some storage capabilities. In traditional visual surveillance systems having a network of CCTV cameras, the video stream is first sent to a common server and from there, it is analyzed either manually or automatically. This model involves data bandwidth, privacy, and security issues due to the huge amount of data that needs to be transmitted through a network. Edge computing brings computing resources closer to the source. In the present model, an anomaly detection model runs on each individual node. Many surveillance systems have recently been proposed in the literature [24][25][26][27][28][29][30][31][32][33].

There are many small-sized embedded devices suitable for computer vision tasks, such as the Jetson Nano, Google’s Coral, and Intel’s Myriad-X vision processing unit [24]. The latest breakthrough is the VPU, developed by Intel. It focuses on the parallel processing of neural networks, having high-speed inference processing and low power consumption. They can be used in embedded systems, drones, or systems powered by external power supplies. Myriad-X is available on the market. It has been used for object classification and for an object detection system on a Raspberry Pi.

An Edge-based surveillance system can be a helpful and useful remote monitoring tool for elderly patients [26]. The work of Yang et al. [28] describes the edge-based set-up of detecting and tracking the target vehicles using unmanned aerial vehicles (UAVs). They use a CNN model for object detection and further classification. Due to the power and computational limitations of UAVs, some of the processing in the system is offloaded to a local mobile-enabled computing (MEC) server. This approach makes the overall system computationally and power consumption-wise more efficient.

The edge devices have limited power and, therefore, restricted processing power. Pradeepkumar et al. [29] discuss a method to maintain the object detection accuracy of about 95% by just transmitting 5–10% of the frames captured by the edge camera. Ananthanarayanan et al. [30] propose an edge computing-based anomalous traffic detection video surveillance system that works on live video streams. Multiview activity recognition and summarization is a difficult task due to many challenges like view overlapping, inter-view correlations, and stream disparities [31]. Researchers have been trying to find innovative solutions to these problems. Combining this with edge computing can be very beneficial. Hussain et al. [31] proposed a framework to bring the task of multiview video summarization to an edge computing platform.

References

1. Sharma, L.; Lohan, N. Performance analysis of moving object detection using BGS techniques in visual surveillance. *Int. J. Spatio-Temporal Data Sci.* 2019, 1, 22–53.
2. Tripathi, R.K.; Jalal, A.S.; Agrawal, S.C. Abandoned or removed object detection from visual surveillance: A review. *Multimed. Tools Appl.* 2019, 78, 7585–7620.
3. Gawande, U.; Hajari, K.; Golhar, Y. Pedestrian detection and tracking in video surveillance system: Issues, comprehensive review, and challenges. In *Recent Trends in Computational Intelligence*; Intechopen: London, UK, 2020; pp. 1–24.
4. Gundogdu, E.; Parıldı, E.S.; Solmaz, B.; Yücesoy, V.; Koç, A. Deep learning-based fine-grained car make/model classification for visual surveillance. In *Proceedings of the Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies*, Warsaw, Poland, 11–12 September 2017; SPIE: Bellingham, WA, USA, 2017; Volume 10441, pp. 179–184.
5. Zitouni, M.S.; Sluzek, A.; Bhaskar, H. Towards understanding socio-cognitive behaviors of crowds from visual surveillance data. *Multimed. Tools Appl.* 2020, 79, 1781–1799.
6. Santhosh, K.K.; Dogra, D.P.; Roy, P.P. Anomaly detection in road traffic using visual surveillance: A survey. *ACM Comput. Surv. (CSUR)* 2020, 53, 1–26.
7. Ansari, M.A.; Singh, D.K. An expert video surveillance system to identify and mitigate shoplifting in megastores. *Multimed. Tools Appl.* 2022, 81, 22497–22525.

8. Wu, Z.; Yao, T.; Fu, Y.; Jiang, Y.G. Deep learning for video classification and captioning. In *Frontiers of Multimedia Research*; ACM: New York, NY, USA, 2017; pp. 3–29.
9. Mliki, H.; Bouhlel, F.; Hammami, M. Human activity recognition from UAV-captured video sequences. *Pattern Recognit.* 2020, **100**, 107140.
10. Sunil, A.; Sheth, M.H.; Shreyas, E. Usual and unusual human activity recognition in video using deep learning and artificial intelligence for security applications. In Proceedings of the 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India, 15–17 September 2021; IEEE: Piscataway Township, NJ, USA, 2021; pp. 1–6.
11. Wassim, A. Abnormal Activity Detection In Crowded Scenes Using Video Surveillance. In Proceedings of the Cyber-Physical Systems and Control, Sydney, Australia, 21–25 April 2020; pp. 106–118.
12. Nawaratne, R.; Alahakoon, D.; De Silva, D.; Yu, X. Spatiotemporal anomaly detection using deep learning for real-time video surveillance. *IEEE Trans. Ind. Inform.* 2019, **16**, 393–402.
13. Zhou, J.T.; Du, J.; Zhu, H.; Peng, X.; Liu, Y.; Goh, R.S.M. AnomalyNet: An anomaly detection network for video surveillance. *IEEE Trans. Inf. Forensics Secur.* 2019, **14**, 2537–2550.
14. Rodrigues, R.; Bhargava, N.; Velmurugan, R.; Chaudhuri, S. Multi-timescale trajectory prediction for abnormal human activity detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 2626–2634.
15. Fan, Z.; Yin, J.; Song, Y.; Liu, Z. Real-time and accurate abnormal behavior detection in videos. *Mach. Vis. Appl.* 2020, **31**, 72.
16. Ullah, W.; Ullah, A.; Haq, I.U.; Muhammad, K.; Sajjad, M.; Baik, S.W. CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. *Multimed. Tools Appl.* 2021, **80**, 16979–16995.
17. Singh, T.; Vishwakarma, D.K. A deeply coupled ConvNet for human activity recognition using dynamic and RGB images. *Neural Comput. Appl.* 2021, **33**, 469–485.
18. Shreyas, D.; Raksha, S.; Prasad, B. Implementation of an anomalous human activity recognition system. *SN Comput. Sci.* 2020, **1**, 168.
19. Beddiar, D.R.; Nini, B.; Sabokrou, M.; Hadid, A. Vision-based human activity recognition: A survey. *Multimed. Tools Appl.* 2020, **79**, 30509–30555.
20. Pareek, P.; Thakkar, A. A survey on video-based human action recognition: Recent updates, datasets, challenges, and applications. *Artif. Intell. Rev.* 2021, **54**, 2259–2322.
21. Kumar, S.; Kumar, S.; Raman, B.; Sukavanam, N. Human action recognition in a wide and complex environment. In Proceedings of the Real-Time Image and Video Processing 2011, San

Francisco, CA, USA, 1 January 2011; SPIE: Bellingham, WA, USA, 2011; Volume 7871, pp. 176–187.

- 22. Kumar, S.; Kumar, S.; Sukavanam, N.; Raman, B. Human visual system and segment-based disparity estimation. *AEU-Int. J. Electron. Commun.* 2013, 67, 372–381.
- 23. Kumar, S.; Kumar, S.; Sukavanam, N.; Raman, B. Dual tree fractional quaternion wavelet transform for disparity estimation. *ISA Trans.* 2014, 53, 547–559.
- 24. Cob-Parro, A.C.; Losada-Gutiérrez, C.; Marrón-Romera, M.; Gardel-Vicente, A.; Bravo-Muñoz, I. Smart video surveillance system based on edge computing. *Sensors* 2021, 21, 2958.
- 25. Zhang, J.; Xu, C.; Gao, Z.; Rodrigues, J.J.; de Albuquerque, V.H.C. Industrial pervasive edge computing-based intelligence IoT for surveillance saliency detection. *IEEE Trans. Ind. Inform.* 2020, 17, 5012–5020.
- 26. Rajavel, R.; Ravichandran, S.K.; Harimoorthy, K.; Nagappan, P.; Gobichettipalayam, K.R. IoT-based smart healthcare video surveillance system using edge computing. *J. Ambient. Intell. Humaniz. Comput.* 2022, 13, 3195–3207.
- 27. Ahmed, I.; Ahmad, M.; Rodrigues, J.J.; Jeon, G. Edge computing-based person detection system for top view surveillance: Using CenterNet with transfer learning. *Appl. Soft Comput.* 2021, 107, 107489.
- 28. Yang, B.; Cao, X.; Yuen, C.; Qian, L. Offloading optimization in edge computing for deep-learning-enabled target tracking by internet of UAVs. *IEEE Internet Things J.* 2020, 8, 9878–9893.
- 29. Kumar, P.P.; Pal, A.; Kant, K. Resource efficient edge computing infrastructure for video surveillance. *IEEE Trans. Sustain. Comput.* 2021, 7, 774–785.
- 30. Ananthanarayanan, G.; Bahl, P.; Bodík, P.; Chintalapudi, K.; Philipose, M.; Ravindranath, L.; Sinha, S. Real-time video analytics: The killer app for edge computing. *Computer* 2017, 50, 58–67.
- 31. Hussain, T.; Muhammad, K.; Ullah, A.; Del Ser, J.; Gandomi, A.H.; Sajjad, M.; Baik, S.W.; de Albuquerque, V.H.C. Multiview summarization and activity recognition meet edge computing in IoT environments. *IEEE Internet Things J.* 2020, 8, 9634–9644.
- 32. Aishwarya, D.; Minu, R. Edge computing based surveillance framework for real time activity recognition. *ICT Express* 2021, 7, 182–186.
- 33. Subramanian, R.R.; Vasudevan, V. A deep genetic algorithm for human activity recognition leveraging fog computing frameworks. *J. Vis. Commun. Image Represent.* 2021, 77, 103132.

Retrieved from <https://encyclopedia.pub/entry/history/show/121334>