WCRG Global Research
WORLD COMET RESEARCH GROUP

International Journal of Modern Innovations and Emerging Trends
Vol. 5, No. 2, 2022
Doi: 10.67228/30715628/IJMIET-2022PII7F9Q
PP. 01-13

Original Article

Computer Vision Techniques for Automated Surveillance Systems

Ajay Krishnan

Technical Lead, Tech Mahindra, India.

Received: 07-03-2022 Revised: 07-26-2022 Accepted: 08-03-2022 Published: 08-05-2022

ABSTRACT

Modern security systems have raised a new system of a need to integrate automated surveillance systems as part of their security infrastructure because of the sudden increase in urbanization, civil safety issues and the necessity to have a smart monitoring system. The conventional surveillance systems are very dependent on human operator hence constraints include fatigue, delay in response and subjectivity. Computer vision as a branch of artificial intelligence will allow machines to read and understand visual information automatically, and thus change the traditional surveillance into its intelligent and active form. This paper gives an extensive research of the computer vision approaches in automated surveillance systems. It also explores how the classical approaches to image processing have been transformed to deep learning based methods such as their application in object detection and tracking, activity recognition, anomaly detection and facial recognition. System architectures, data acquisition pipelines, feature extraction methods, model training strategies and performance evaluation metrics are also discussed in the paper. Moreover, the issues like occlusion, change of illumination, scalability, privacy, real-time processing are examined. The effectiveness of the modern computer vision methods is discussed with references to the experimental results of the representative surveillance scenarios. Lastly, the paper provides the future research direction, such as edge-AI surveillance, multimodal fusion, and explainable computer vision, which is important to the next-generation intelligent surveillance systems.

KEYWORDS

Computer Vision, Automated Surveillance, Object Detection, Video Analytics, Deep Learning, Intelligent Security Systems.

WCRG
Global Research
ISSN: 3071-5628
PP: 01-13.

1. INTRODUCTION

1.1. Background

The fast spread of surveillance systems in smart cities, transport centers, factories, and areas of gathering people caused the constant production of colossal amounts of video material. As CCTV cameras and smart sensing devices keep being installed in large-scale settings, the ability to monitor such settings has become a vital need to ensure the safety of people and assets, as well as efficient utilization of the urban areas. Nevertheless, the traditional surveillance methods are mostly dependent on human element to monitor video feeds, which is time consuming and very inefficient. Manual surveillance is liable to exhaustion, lapses in attention, and subjective decision making such that it becomes ascertaining to reliably spot any suspicious act as well as counteracting an impending threat. The fact that the economic cost of having individuals round the clock to monitor is also an issue adds to the fact that the conventional surveillance uses are also scaleable. These complications have fueled the implementation of automated monitoring systems that are operated by computer vision methods. Computer vision allows machines to read and process visual data by obtaining valuable semantic object information about the images and videos. Computer vision systems are able to automatically detect events of interest and signal in time without the active supervision of a human operator through techniques like object detection, tracking, activity recognition and anomaly detection. This automation can better detect the threat in real time, it helps to monitor the behavior, and the general situational awareness is promoted. Due to it, computer vision-enhanced surveillance systems can be discussed as an effective, easily scalable, and reliable substitute of the old fashioned methods of monitoring due to which they may be called one of the crucial elements of the modern intelligent security systems.

1.2. Role of Computer Vision in Surveillance

Computer vision is a revolutionary aspect in the contemporary surveillance system because it allows computerized interpretation and comprehension of visual imagery. Unlike the old-fashioned surveillance, which only captures the video footage, the computer vision-based surveillance actively investigates the scene in order to obtain meaningful information and enhance the efficiency, accuracy and responsiveness. Computer vision in surveillance may be perceived in regard to the following main aspects:

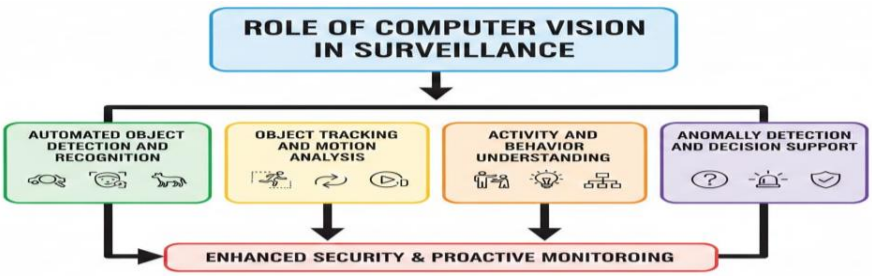

Fig 1 - Role of Computer Vision in Surveillance

ISSN: 3071-5628
PP: 01-13.

1.2.1. Automated Object Detection and Recognition

Automatic identification and registration of objects including people, vehicles, and items left unattended is one of the major functions of the computer vision in the surveillance process. Computer vision models are able to detect objects of interest within a real-time video by boosting and match the visual patterns in video frames. Such a feature makes it less intricate on manual observation and allows round-the-clock surveillance over extensive regions. With correct object recognition, superior surveillance processes (access control and intrusion detection) are based.

1.2.2. Object Tracking and Motion Analysis

Computer vision makes possible a credible tracing of the identified objects on successive video frames. Monitoring enables the system to detect the movement paths, velocity and direction which is crucial in deciphering movement behaviour. Motion analysis assists with determining the presence of some abnormal movement pattern, e.g., sudden running, loitering, and people who have gained access into restricted areas illegally. This situational awareness is supported by this continuous tracking and aids in proactive security action.

1.2.3. Activity and Behavior Understanding

Other than object detection, computer vision is very important in identifying activities and the process of understanding human behavior. Surveillance systems are able to identify normal and suspicious activity by modeling spatial and temporal characteristics using video sequences. This is highly useful in crowdy or complicated surroundings where it is not easy to personally observe. Activity recognition helps systems identify the occurrence of such events as fights, falls, and vandalism, allowing them to intervene in time.

1.2.4. Anomaly Detection and Decision Support

The computer vision systems also play a recurrent role in identifying the odd labels by picking up the rules that a normal behavior entails and discovering the variation in these norms. This type of anomaly detection helps in the rapid detection of threats and also minimizes false alarms. Moreover, the insights of computer vision help in decision-making as they produce alerts, summaries, and actionable information to the security personnel. In general, computer vision allows improving surveillance systems and turning them into intelligent, scalable, and responsive in various environments.

1.3. Evolution of Surveillance Technologies

The development of the surveillance technologies testifies to the major progress in the computing practices, sensing devices, and artificial intelligence. Simple motion detection algorithms, which included frame differencing and simple background subtraction, were the main basis of early surveillance systems. To find moving objects, these methods identified changes between the successive frames and were computationally cheap and could be used because of the limitations of early hardware. But these methods were very susceptible to noise, change in lighting and movement of cameras which in most cases gave them false alarms and inaccurate performance in real life scenarios. Their poor coverage of scene context reduced surveillance to the simplistic motion

ISSN: 3071-5628
PP: 01-13.

detection with no semantics. As machine learning developed and more computational resources became accessible, data-driven models started to be represented by surveillance systems so that the detection and classification can be enhanced. The traditional machine learning algorithms, along with handcrafted features, helped the systems to identify particular objects, as well as activities, in a more accurate way compared to rule-driven approaches. Nevertheless, the lack of manual feature engineering and poor generalization, made it difficult to deal with complicated and dynamic scenes. With the development of deep learning, a significant turning point in the surveillance technology has been made. The Convolutional Neural Networks (CNNs) allowed the automatic learning of the hierarchical spatial features that greatly enhanced the accuracy of object tracking, recognition and detection. These models exhibited good performance in different lighting conditions, perspective and the background intricacies. Subsequently, temporal modeling was proposed by implementing Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks which enabled surveillance systems to analyze time-based data and identify activities. More recently, transformer-based architectures have been sold as they are capable of modelling long-range dependencies and global context in a more efficient way. The methodologies provide strength and versatility as they utilize complex spatiotemporal associations in video records. Consequently, the contemporary surveillance systems can analyze intelligent behavior, detect anomaly and make real-time decisions, which is a major development of the primitive motion-based solutions to the elaborate AI-enhanced surveillance solutions.

2. LITERATURE SURVEY

2.1. Traditional Image Processing Approaches

The initial automated surveillance systems were mainly the traditional image processing systems which were founded on human made rules and statistical assumptions. Background subtraction, optical flow estimation, edge detection and morphological operations were often utilized to isolate any moving object and still backgrounds. One of these, the Gaussian Mixture Models (GMM), saw widespread use as a background modelling method because it can model dynamic environments like waving trees or slow changing scenes (tying an apron) with varying light. These were computationally efficient as well as fit well in real time operations on modest hardware. Nevertheless, they deteriorated in challenging conditions such as shadows, occlusions, abrupt changes in the lighting and camera jittering. These methods were not able to effectively deal with the real-world surveillance because they lacked flexibility and semantic interpretation.

2.2. Machine Learning–Based Methods

With the development of machine learning, a paradigm shift in the use of rules in surveillance was brought to the end of data-driven solutions. They used the Support Vector Machines (SVM), k-nearest neighbors (k-NN) and random forests as supervised learning algorithms to classify objects and detect abnormal events according to the learned decision boundaries. These models made intensive use of hand-engineered features, such as Histogram of Oriented Gradients (HOG) to encode a shape and Scale-Invariant Feature Transform (SIFT) to gather salient points which are scale and rotation-invariant. Machine learning-based systems proved to have a higher level of robustness and generalization as compared to traditional methods. Nevertheless, they continued to be

ISSN: 3071-5628
PP: 01-13.

constrained by the quality of feature extraction, noise-sensitivity and inability to reason out complicated spatial and temporal relationships in video information.

2.3. Deep Learning in Surveillance

Deep learning has completely transformed the automated surveillance because it allows the learners to operate in hierarchies and specifically, with raw data. Convolutional Neural Networks (CNNs) contributed greatly to object detection and recognition because it detects spatial features at multiple levels of abstraction automatically. Sophisticated detection systems were remarkable when it came to detecting pedestrians, vehicles, and other objects of interest in a cluttered set up. Besides, Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks, were included so that they could capture temporal relationships across video frames. It was this integration that allowed the activity recognition, event prediction, and anomaly detection to be more effectively performed. Even though they are effective, the deep learning models may require a large amount of computation power and large amount of training data which is usually not easily available.

2.4. Video Analytics and Behavior Recognition

High-level video analytics and understanding of human behavior has been turning into the main theme of latest surveillance studies as opposed to mere object detection. Action and behavior recognition algorithms focus on interpreting the communication, motion, and intention through STA sequences of video streams. The methods, including 3D Convolutional Neural Networks (3D CNNs), are able to consider both space and time glassily, and attention-based models enable them to concentrate on the most informative regions and frames. The strategies have presented encouraging outcomes in identification of suspect actions like loitering, fighting or intrusion. But the identification of complex human behaviors in unstructured settings is not an easy task based on an individual diversity in perspective, inconsistency in the view, and individuality in individuals.

2.5. Limitations in Existing Research

Even after the improvement of vast technology, surveillance systems applied nowadays have certain limitations. Large-scale annotated data are required by many approaches based on deep learning, and are expensive and time-consuming to acquire, especially in the case of rare or abnormal events. Its intensive computational and memory needs are also limiting it on the resource-limited devices like edge cameras.

In addition, issues of privacy intrusion, data safety, and bias in the algorithms are usually not considered in technical research. The unfair or inaccurate surveillance results may be caused by such problems as bias in trainings data, which is why assigning blame to the system design and its prompt and honest development should be carefully considered. The challenges are open research directions in the sphere of surveillance.

ISSN: 3071-5628
PP: 01-13.

3. METHODOLOGY

3.1. System Architecture

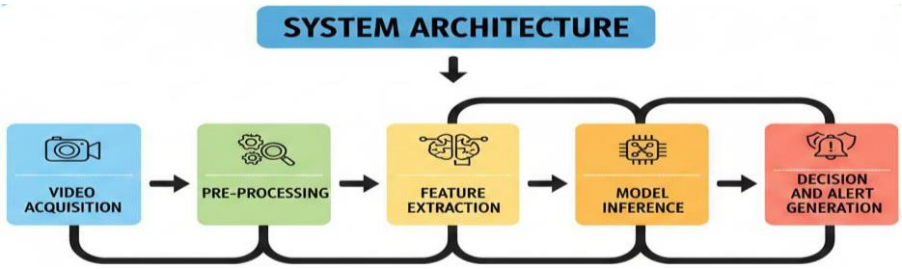

Fig 2 - System Architecture

A surveillance system that is automated consists of several interconnected modules and they collaborate collectively to observe, interpret as well as react to what occurs in a designated area of observation.

3.1.1. Video Acquisition

The video acquisition unit has the responsibility of capturing the real time video streams of the surveillance cameras or video sensors. These are fixed or mobile cameras, which are usually installed to monitor vital locations. The video data captured is the main input into the system, it should be acquired at the correct rate and resolution to be able to use it in reliable analysis.

3.1.2. Pre-processing

Pre-processing means that one improves the quality of the captured video in order to have better analysis later. Some of the common operations are noise reduction, frame resizing, normalization, background subtraction, and illumination correction. This phase is used to remove cheap data and makes sure that the data used is even and can be captured by features.

3.1.3. Feature Extraction

The feature extraction module recognizes and isolates significant data of the video frames. This can be handcrafted (e.g. due to a specific approach like edges, textures or motion vectors) or it can be automatically learned with the help of a deep learning model. Proper feature extraction is very important in the correct representation of objects and events in the scene.

3.1.4. Model Inference

Model inference is the process of using trained machine learning or deep learning models based on the extracted features. Object detection, tracking, activity recognition or anomaly detection are the tasks that this module can accomplish. The inference process is used to interpret the dynamics of the scenes and provide events of interest in real time or near real time.

3.1.5. Decision and Alert Generation

The last module analyzes the results of the inference in order to come up with a decision depending on preset rules or predefined confidence levels. In case of suspicious or abnormal

ISSN: 3071-5628
PP: 01-13.

activities, there are warnings or alerts that are sent to the security personnel by the system. The module guarantees the provision of responses on time and facilitates efficient surveillance management.

3.2. Data Acquisition and Pre-Processing

The data acquisition and pre-processing becomes a part of critical phase in an automated surveillance system since the quality of the incoming data directly determines the accuracy of the other modules of the analysis. In most cases video streams are recorded with fixed cameras or pan-tilt-zoom (PTZ) cameras that can be installed in strategic points to have a maximum coverage of the area under surveillance. With fixed cameras, there will be constant observation of given areas, whereas PTZ cameras will be flexible in that they change their orientation and crop of focus based on the object to be observed or areas of interest. To compromise between storage and computing power, these video streams are captured at suitable frame rates and resolutions. The data of the video that has been captured is normally sent or fed in real time to an analysis central processing unit or edge equipment to be further analyzed. When the raw video data is obtained, it is then processed through a number of pre-processing steps to increase robustness and consistency. To curb sensor noise and the compression artefacts which can potentially interfere with feature extraction, noise reduction methods are used (Gaussian or median filtering). Frame normalization is implemented to normalize the values of pixel intensity and minimize the effect of different illumination conditions in different scenes or time of the day. Video frames are often downsized to minimize the amount of computations necessary and yet still retaining all the critical visual data necessary to perform correct detection and recognition tasks. Also, conversion to color space, including RGB images to grayscale or HSV color space, can be useful to highlight the details of interest, including intensity or color distribution and enhance the stability of the model. All these pre-processing tasks make the video data clean, uniform, and suitable to be utilized in achieving reliable feature extraction and model inference, which eventually increases the effectiveness of the surveillance system on the whole.

3.3. Object Detection and Tracking

The working elements of automated surveillance systems are object detection and tracking that allow identifying and tracking such subjects as people, vehicles, and other objects of interest within a picture. Object detection pays attention to the definition and identification of single video frames with the help of visual representations like shape, texture, and movement. The contemporary surveillance systems are usually based on the deep learning-based detection model capable of recognizing the objects even in the crowded or cluttered scenarios. These detectors produce bounding boxes and confidence values that enable the system to differentiate between the relevant and non-relevant objects. Object detection should be reliable to comprehend the dynamics in the scene, which becomes the basis of the other higher-order activities, like behavior analysis and event detection. On the detection of objects, tracking algorithms then are used to equate the same objects across the consecutive frames, thus enabling temporality continuity. Object tracking allows the system to approximate motion paths, velocity, and direction, which are important in the study of the movement pattern and interaction. Some of the methods include Kalman filtering, particle filtering and data association techniques usually employed to predict the position of the objects and manage

ISSN: 3071-5628
PP: 01-13.

the temporary occlusion. In more sophisticated systems, trackers based on deep learning use appearance as a predictor to re-identify object when they come out and come back into the field of view of the camera. Good tracking makes the objects identities constant with the passage of time and in some of the situations that have objects of overlap or abrupt changes in motion. Object detection and tracking offer a complete insight of the dynamics of movement and interaction of objects in a monitored environment. This aggregate functionality underpins such applications as intrusion detection, crowd monitoring, traffic analysis and suspicious activity detection. Through proper object recognition and keeping of the object paths, surveillance systems may generate insightful information regarding the video data and provide a chance to make timely decisions and raise alarms.

3.4. Activity Recognition and Anomaly Detection

Intelligent surveillance systems are greatly reliant on activity recognition and anomalies detection in that the identification of complex human behaviors is achieved with time. Compared to object detection, where the detection conceptualizes the object in a single frame only, activity recognition in contrast examines frame sequences to learn about actions, interactions and movement patterns. The spatial aspects like motions, posture changes, and interaction patterns are obtained by processing the video information and trained through the sequential learning methods. LSTM networks and gated recurrent units (GRUs) are recurrent neural networks that are widely used to learn long-term dependencies and interrelationships between frames. These models permit the system to distinguish between routine and more complex activities including coordinated movement like walking or standing. Anomaly detection is used to detect behaviors that are highly different as compared to the learned normal patterns. The fact that abnormal events are frequently rare and hard to label many surveillance systems view unsupervised or even semi-supervised learning techniques. During such processes, models are often trained based on data depicting normal activities to have a baseline version of what is expected of them. In inference, any departure in this baseline, in cases of reconstruction errors, prediction errors or probability scores are reported as possible anomalies. Some of the examples of these anomalous behaviors are: unauthorized access, loitering, impulsive running, fighting or bizarre object contact in restricted areas. The tie-up of activity recognition to anomaly detection will increase the capability of the system to offer proactive surveillance and not passive monitoring. Moreover, because the system keeps analyzing the patterns over time, it can detect an actual security incident and provide the security staff with the necessary alert in time. Nevertheless, detection accuracy can be influenced by difficulties like difference in personality of the individual, complexity of the environment, and change of perspective. In spite of them, temporal modeling and developments in deep learning have further enhanced credibility and efficiency of activity recognition and anomaly detection in current surveillance systems.

3.5. Performance Evaluation Metrics

Measures of performance evaluation are required to determine the effectiveness, reliability and efficiency of automated surveillance systems. These measures are used to give quantitative data in order to compare the various models and then decide on whether they are suitable to be deployed in real-world situations. Accuracy is the most commonly used score and the ratio of the correctly

ISSN: 3071-5628
PP: 01-13.

classifications instances to the total predictions. Although accuracy gives a global view of the system performance, it may be misleading when the situation of surveillance is in quake where most of the abnormal events are rare resulting to an imbalance between the classes. This means that more metrics are needed in order to have a more in-depth insight into model behavior. Accuracy and sensitivity are of a great essence during surveillance application testing. Precision is a measure of the fraction of all the predictions that are correctly marked positive of the total predictions, which reveals the accuracy of the system produced alerts. The high precision level also makes sure that the level of false alarm is reduced to the maximum, and this factor determines the trust in automated monitoring systems. On other hand, recall is used to measure the number of actual positive instances that are rightly tracked by the system. When the recall value is high, it means the system is effective to trigger all the events of interest to minimize chances of failure to do so. The harmonic mean of false negatives and false positives, i.e., the F1-score, gives a balanced level of the assessment in regard to the false negatives and false positives. Besides accuracy-based metrics, processing latency is an essential performance metric in an environment that has a surveillance system. Latency is defined as a time frame in which the system takes in order to process video frames and produce outputs or alerts. Real time surveillance applications need low latency to get the important responses in time. Determining the effectiveness of the detection procedure along with computer ability allows making sure that the surveillance systems are not efficient, but feasible and resilient in the absolute sense.

4. RESULTS AND DISCUSSION

4.1. Experimental Setup

The experimental environment was to test the performance of the proposed automated surveillance system in a systematic and realistic way. It used widely accepted benchmark surveillance data, which has annotated video sequences of a variety of real-world situations. Variations in lighting conditions, camera views, the complexity of the background, the degrees of the crowd density are usually included to these datasets, which allows complete evaluation of the system robustness. The experimental results have the potential to be compared with the available methods in the literature by reference to standardized benchmark datasets to be fair and reproducible in the evaluation. The experiments were conducted in different environmental conditions in order to investigate the adaptability of the system. These were conditions in the indoor and outdoor settings, daytime and nighttime settings and situations with a dynamic background like moving objects or changes in lighting. Occlusiveness in video sequences, motion blur, and change of perspective were also provided to examine the resilience of the system to typical problems found with the surveillance applications. This makes such diversity in experimental conditions useful in determining the strengths and weaknesses of the system under various operational conditions. In order to maintain uniform testing, experiments were run with a preset model parameter, input resolution or frame rate. This was done by splitting the datasets into training, validation and testing sets to ensure the absence of overfitting and also to determine the extent of generalization. Each of the experimental situations was recorded in terms of performance measures, including accuracy, precision, recall, F1-score, and processing latency. This controlled experimental design offers a solid basis to the study of

ISSN: 3071-5628
PP: 01-13.

performance in the system and the substantiation of the success of the proposed surveillance strategy under the conditions of the real operation.

4.2. Quantitative Results

Table 1: Quantitative Results

Model	Accuracy (%)	Precision (%)	Recall (%)
Traditional GMM	78.2	75.6	73.4
CNN-Based Model	91.5	89.7	90.2
CNN-LSTM Model	94.3	92.8	93.6

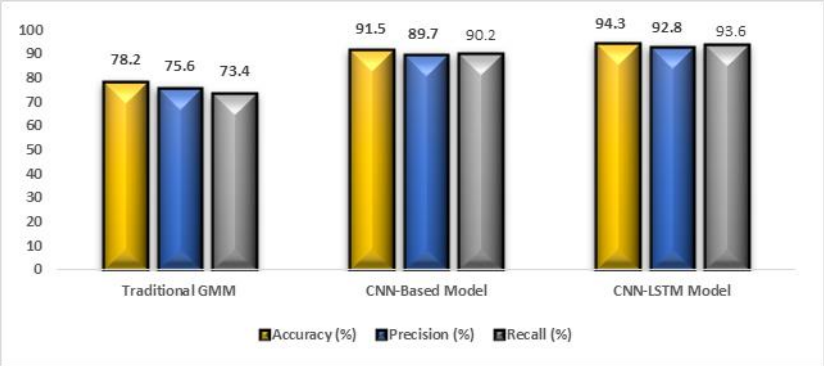

Fig 3 - Quantitative Results

4.2.1. Traditional GMM Model

The approach based on the Traditional Gaussian Mixture Model (GMM) had the accuracy of 78.2, the precision and recall of 75.6% and 73.4%, respectively. These findings mean that the GMM model can only perform simple foreground detection and motion-based analysis, and cannot work well in complex surveillance fields. The smaller recollection value indicates that, the model does not identify a large number of meaningful events, especially in bad environments like change of illumination, shadows, and moving backgrounds. This shows the natural weaknesses of the conventional method of image processing when it comes to dealing with the variability in the real world.

4.2.2. CNN-Based Model

The CNN-based model of surveillance had an impressive improvement in its performance with an accuracy of 91.5. The accuracy of 89.7% should reflect the fact that it has a high potential to detect the relevant objects and activities with high accuracy and with limited cases of false positive. Also, the recall value of 90.2 indicates that the model is an effective one in identifying most of the target events in the video content. The achievement of the better performance can be explained by the fact that the CNN is able to learn hierarchical spatial features of the video frames automatically, and provides better object representation and discrimination than handcrafted features do.

ISSN: 3071-5628
PP: 01-13.

4.2.3. CNN-LSTM Model

CNN-LSTM model achieved the best performance when compared to the other tested models with an accuracy of 94.3 per cent, a precision of 92.8 per cent and a recall of 93.6 per cent. By combining LSTM networks with CNN based feature extraction, the model would be able to extract spatial and temporal information resulting in a higher accuracy in recognizing activities and detecting events. The large value of recall is indicative of the fact that the model can always detect interested events with time whereas the large value of precision means that the model produces accurate alerts. These findings are valid to support the efficiency of temporal modeling in improving surveillance systems.

4.3. Analysis of Results

The results of the experiment makes it clear that the method of surveillance based on deep learning is a lot more effective than the traditional image processing techniques used in scope and changing environments. The conventional methods, including Gaussian Mixture Models, have statistical premises, as well as low-level visual features, which restrict their capacity to adjust to changes in the lighting process, the background motion, and the appearance of an object. As it is seen in the quantitative results, these restrictions give lower accuracy, precision and recall values and thus traditional methods cannot be highly trusted in real-life surveillance applications where there is an occlusion, a tight crowd or even some sudden change in the environment. In comparison, deep learning based models such as CNN and CNN-LSTM structures are highly performing based on the fact that they are capable of automatically acquiring rich and discriminative features within data. CNN-based models have quite a good ability to extract spatial details like shape, texture, and context of objects which leads to the area of quite a significant improvement in the status of detection. The use of the LSTM networks also adds to the performance improvement as it views temporal dependencies between frames in a video. This time modeling facilitates the system to learn more about motion patterns and activity sequences and results in the improved development of events recognition and a decrease in the number of false alarms. The fact that the accuracy and the recall rates are always on the increase confirms the strength of deep learning models in recognizing important activities and reducing the number of unnoticed ones. Nonetheless, deep learning models have a higher accuracy at the expense of greater computational complexity. CNN-LSTM models take large amounts of processing power, memory, and in some cases specialist hardware like GPUs to train and deploy. This computational complexity may be a challenge to real-time surveillance programs, especially in low resource situations or edge programs. Thus, on the one hand, the performance gains of deep learning methods are high, but on the other hand, performance and accuracy have a tradeoff. This trade-off is also something that has to be struck to come up with viable and scalable surveillance systems that are applicable on the actual practice.

4.4. Discussion on Practical Deployment

In real world applications, automated surveillance systems need to be properly considered regarding the aspects of real-time working and scaling as well as resource limitations. Although model systems built using deep learning are highly accurate and perform well, their complexity in computer processing makes it difficult to use in real-time. Hardware acceleration with Graphics

ISSN: 3071-5628
PP: 01-13.

Processing Unit (GPU) or Tensor Processing Unit (TPU) or dedicated AI accelerators may be required to answer real-time needs. These hardware platforms are much faster to model inference and support high-resolution video streams with very low delay. Also, more methods like model pruning, quantization, and knowledge distillation are typically used to compress deep learning models, decreasing their size and reducing the computational cost without much reduction in accuracy. The idea of edge computing provided an efficient solution to the issue of latency and bandwidth constraints in the mass surveillance systems. Edge devices also do data processing and inference at the source in lieu of transferring raw video streams to centralized cloud repositories. This will reduce the network congestion, minimize bandwidth use, and improve a faster response time that is important in time sensitive security applications. Edge-based deployment is also much more reliable in the system since it allows it to keep on despite network failure. Moreover, local processing of data will reduce the privacy problem as it will reduce the dissemination of sensitive video data. In spite of these issues, such advantages do not yet guarantee wide practical use because of such challenges as heterogeneity of the devices, power consumption, and complexity of maintenance. The surveillance systems should be configured in a manner that can be effective and operate well on different hardware platforms and perform the same way. There is hence a need to balance between accuracy, latency and resource utilization to ensure effective real-time usage. With an integrated approach involving model optimization methods and edge computing to infrastructure, contemporary surveillance systems are able to provide scalable, responsive, and practical solutions to fit into the reality of security practices.

5. CONCLUSION

This paper has provided an in-depth discussion of the use of computer vision in automated surveillance systems and the trends of transitioning between the traditional image processing to the advanced deep learning-based approaches. The first methods were done by hand and based on statistical models thus gave computationally efficient answers but failed to represent complex backgrounds and dynamic environments and different light conditions. The advent of machine learning made systems more adaptable to the data by allowing learning of features through data, the introduction of deep learning also brought about a miracle of breakthrough because it allows the setting up of end-to-end learning and representation of visual information robustly. Convolutional Neural Networks and sequential models made better results in object detection, tracking, activity recognition, and anomaly detection tasks, which makes deep learning the most popular paradigm of a modern surveillance system. The effectiveness of deep learning methods is also supported by the experimental test, carried out in the present work. Quantitative findings indicate that models that are developed using deep learning achieve a considerably high level of accuracy, precision, recall, and general robustness compared to those developed using traditional approaches, specifically under complex and dynamic surveillance conditions. Combination of temporal modeling enables the system to understand activities in a time sequence to minimize false alarms as well as a variation in detection. These discoveries substantiate that the nature of advanced learning architectures enables them to handle the demands of real world surveillance setting. Although these developments have been made, there are a number of challenges that are critical. One of the primary considerations is

ISSN: 3071-5628
PP: 01-13.

scalability, where implementing deep learning models in extensive surveillance systems requires a large amount of computational memories and worry-free system administration. Real-time processing demands also increase the deployment complexity particularly where resources are limited. Also, privacy protection and ethics issues, such as data security, surveillance bias and transparency of the decision-making process are not yet profusely covered and should be paid more attention. These issues need to be addressed to facilitate acceptable surveillance solutions that are quite responsible. The further work will be devoted to constructing lightweight and efficient models, which should be optimized to work with edge-AI, so that the real-time processing time and the consumption of bandwidth could be minimized. Multimodal sensor fusion, which will involve the integration of the visual information with the other information sensors such as audio, thermal, or motion sensors, will improve situational awareness and reliability of the system. Moreover, the use of explainable computer vision will enhance transparency and user trust through offering interpretable information of the model decisions. These research directions will address the enhanced effectiveness, scaling and the ethical application of intelligent surveillance systems.

REFERENCES

[1] Stauffer, C., & Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2, 246–252.

[2] Bouwmans, T. (2014). Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*, 11–12, 31–66.

[3] Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17(1–3), 185–203.

[4] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1, 886–893.

[5] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.

[6] Vapnik, V. N. (1998). *Statistical Learning Theory*. New York: Wiley.

[7] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

[8] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems (NIPS)*, 1097–1105.

[9] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 580–587.

[10] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788.

[11] Liu, W., Anguelov, D., Erhan, D., et al. (2016). SSD: Single shot multibox detector. *European Conference on Computer Vision (ECCV)*, 21–37.

[12] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.

[13] Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). Learning spatiotemporal features with 3D convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 4489–4497.

[14] Wang, X., Ma, X., Grimson, W. E. L., & Wang, S. (2009). Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3), 539–555.

[15] Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.