OPEN ACCESS

International Journal Of Modern Engineering Research (IJMER)

Motion Detection for Surveillance Systems: A Survey

[1]Dheeraj N Bhat, [2]Aswin Gopinathan, [3]Bharath CS, [4]Aishwarya V, [5]Dr. GC Bhanuprakash

[1,2,3,4] Department of Computer Science, Sir M Visvesvaraya Institute of Technology, Bengaluru, India
[5]Head of Department, Department of Computer Science, Sir M Visvesvaraya Institute of Technology, Bengaluru, India

Abstract: *Nowadays, as the terror incidents are reaching a broad bulging peak, the demands on intuitive real-time surveillance systems with one or multiple technologies are rapidly increasing. Commercial spaces, warehouses, universities, and hospitals have had huge impacts and necessitate an embedded mechanism that has the ability to alert and record beside live video of the intruder. The main objective of this study is to improve motion detection timely, based on the low computing systems such as IoT that approximates a desktop or a laptop in terms of functionality and which allows it to be monitored by the owner remotely. This survey paper presents a comprehensive study of the state-of-the-art models and advancement in research that has taken place in the application of motion detection for surveillance systems. Our study shows that even though traditional statistical methods have proven to be effective for motion detection, but with the availability of a huge amount of data in today's world and the recent progress in deep learning has paved a way for highly accurate motion detection systems for surveillance.*

Keywords: *Motion Detection, Surveillance Systems, Computer vision, IOT, Systems Security, Intruder Detection, Automated Security,*

Date of Submission: 16-01-2021 Date of acceptance: 27-01-2021

I. INTRODUCTION

With the proliferation of computer vision techniques, it has become easier to monitor large surveillant environments. Typically, security personnel had to stay and monitor the surveillance cameras and alert the authority if an intruder was detected. But now in the big picture, we are considering surveillant environments to be terrains where installing physical cameras is near to impossible. One example of such terrain could be the international borders where surveillance is tough due to border tensions. In such cases, an efficient and robust system is required that can detect infiltration and alert the authorities immediately. The demand for an efficient surveillance system is increasing day by day. From a broader perspective, this up-to-the-minute system is very much required in every place where security can be a concern. Moving object detection is one of the major features of such systems, and when we couple it with a pose-estimation model to detect unusual movement of people in the frames, it leads to a futuristic system that can be made efficient to support any terrain or environment.

The operational perspective of any motion detection model can be visualized using the Flow-chart mentioned in Fig 1. Raw input video frames from the surveillance system are pre-processed to clean the video frames to remove any irregularities present in the input. Moving objects are then detected using any image processing methods like background subtraction, frame differencing, temporal differencing, statistical methods or machine learning approaches like artificial neural networks. The post-processing methods can then be applied to provide a motion masked output.

In this survey paper, we have explored various works done in the domain of surveillance using motion detection. We have listed down our findings in sequential order under the Literature survey section.

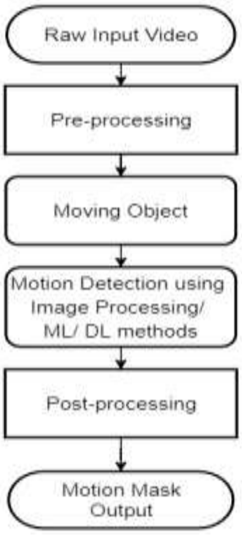

Figure 1: General Operational Framework of Motion Detection System

II. LITERATURE SURVEY

(S.Vishal and G.Prashanth [1]) have discussed in their paper about a remotely monitorable smart real-time surveillance system having an alerting mechanism to detect any kind of incursion. This system utilizes a single Raspberry PI board, an inexpensive IP camera along with a Passive Infrared Sensor to screen the network movement in the vicinity to identify intruders. This approach without using an RFID module makes it economical to be established in any kind of environment. The wireless surveillance system is compatible to be administered by any kind of workstation for the Passive IR Sensor is used to detect the motion in the vicinity of the area whose sensitivity typically ranges from 0 – 20 feet (i.e six meters) with a view of 110 x 60 degrees. It possesses a rechargeable Lithium-Ion battery offering reliable performance and less maintenance which assists during the shutdown of the main electric supply. Several software modules like Image capturing, processing, transmission, and sensor sampling are installed which helps in recognition of the face and attributes of the person for intruder detection. In the case of intruder detection, the system sends the captured image via email and alert message via text message to the user. Even though the system is affordable as compared to other conventional systems there isn't enough security and encryption which can make the system vulnerable to hackers. This surveillance system can be of some use indoors everyday rather than outdoors, say just when you are at work you will instantly be able to keep track of and get notifications if at all any activity takes place at your home. There is a high chance of false alarms and threats that can be predicted by this Surveillance system in an outside environment. The set-up when installed in an outside environment can create potential false alerts by detecting other animals and birds as intruders. This concludes that the system fails in an outdoor condition and is preferably suitable for interior surroundings.

(R Cucchiara, C. Grana, A Parti and R. Vezzani [2]) have mentioned in their paper about developing an integrated system that consists of a moving object detection model capable of disposing of shadows, a tracking module capable of handling large occlusions, and a poster detector. The main problems encountered in house-surveillance systems are Shadows, Occlusions, and Target's Posture. The system uses a distributed network of cameras to track people. A Probabilistic approach is used for the tracking purpose. In this the appearance and probability of occlusion are computed for camera A, then warped in the next camera B's view by positioning the camera to disambiguate the occlusions. With the proliferation of new prototypes of domotic technology, we have explicit and implicit techniques for occupant-to-house communications. Manual triggering happens in an explicit approach whereas automatic detection and triggering take place in an implicit technique. There are various factors like weather conditions, dynamic movement of target objects, shadows, jittering, the shift in camera frames at different angles, self-occlusion that can affect surveillance. All of the mentioned aspects need to be taken into consideration while developing a robust surveillance system.

(Omar EL Harrouss, Driss Moujahid and Hamid Tair [3]) have examined different kinds of approaches to motion detection, namely time difference, analysis of optical flow, and background subtraction. In this paper, the authors give insights on how the latter approach is far better than the former two approaches for the detection of objects in motion in a video. The course of action for using a background subtraction method to detect objects in motion consists of the following steps. A model of the background scene is built, then the procedure involves calculating the difference between the images in sequence and background image. As a final step, a threshold is applied to detect if the given pixel in the image belongs to the background or the object in motion. The pixels belonging to the object of interest are retained and the rest which is part of the background are masked. In this procedure, background maintenance needs to be accomplished to overcome the illumination, temporal variations of the background. This is achieved by updating the background images frequently. Algorithmic steps namely background initialization, foreground detection, and background maintenance are crucial for building a robust motion detection model. For the purpose of building the background, the median value of the sequence of the first 100 images from the video is taken. The difference between subsequent images is calculated for detecting the foreground. As mentioned earlier, a dynamic threshold value is computed to classify the pixel into the background and foreground. If the threshold value is nearly equal to 1, the color space mask is treated as 255, else 0. The binary mask of R, G, B color space is computed with the use of the OR operator. The background is updated by the implementation of the fuzzy theory which helps in the easy classification of background and foreground. This method achieves significant accuracy over other background evaluation techniques like SD and Motion Detection with Pyramid Structure (MDPS).

(Maryam Sultana, Arif Mahmood and Soon Ki Jung [4]) have explored GAN based approach for moving object detection. An unsupervised Generative Adversarial Network (GAN) based object detection algorithm is proposed to detect objects in motion in a highly complex scenario. These find their application not only in surveillance but also in object tracking, autonomous vehicles, human activity analysis, etc. Presently there are Moving Object Detection (MOD) algorithms like GT, SRPCA, GRASTA which have some drawbacks in various kinds of dynamic conditional environments. Traditional methods for MOD comprise Robust Principal Component Analysis (RPCA) and Gaussian Mixture Model (GMM) whereas deep learning methods include training the image data to map the inputs to their respective classes, GANs, etc. A novel algorithm called MOD_GAN is proposed for MOD which utilizes scene-specific GANs for unsupervised model training and backpropagation algorithm for testing the images generated for loss minimization. For loss minimization, a noise sample is searched for reducing the discriminator feature space loss and adversarial loss. The GAN consists of two neural networks generator and discriminator to play a minimax game where the generator tries to produce images similar to original images and the discriminator tries to distinguish the images produced by the generator as real and unreal images thus minimizing the losses. The major part of the work in this paper is focused on developing the loss function for the GAN which could learn fine-grained details from the input images and generate realistic images to train in an unsupervised way. The generator contains four convolutional layers to generate a 64x64x3 image. The discriminator neural network contains a similar architecture but in reverse order of generator with two-class softmax to distinguish real and fake images generated by the generator. A novel dataset was proposed by the authors called Illumination Changes Dawn till Dusk (ICDD) which has a real outdoor environment with different lighting conditions and dynamic background having 20000 and 1265 video frames for training and testing respectively. The average F-score on various illumination conditions outperformed other existing methods except for the camera jitter condition where it fails to perform better than other methods because there is no clean background and hence not much information can be extracted from the images.

(Thomas Golda, Nils Murzyn, Chengchao Qu and Kristian Kroschel [5]) have researched anomaly detection using short term cyclic consistency. Anomaly detection is an important field of research whose task is to detect any outliers present within the given data. This paper brings light upon the automated analysis of surveillance cameras which can help in detecting any abnormality found in the video recordings. They focus on the static camera as these are the ones that are being widely used because of their easy installation and low affordable cost. GANs are being utilized for unsupervised detection of anomalies through the identification of normal motion patterns and distinguish them from abnormal conditions. There are anomaly detection methods based on three types -

1. Representation learning for reconstruction - Auto Encoders and Convolutional Auto Encoders
2. Predictive Modelling - Long Short Term Memory (LSTM)
3. Generative Modelling – GAN

Their hypothesis is that anomalies in surveillance systems have larger spatial differences in their heat maps and suppression of this noise plays an important part in the performance of the surveillance system. To allow for this noise suppression, a post-processing step is included which applies morphological operations. The image has a resolution of 256x256 pixels. The GAN architecture utilized consisted of a U-net generator and PatchGAN discriminator. Optical flow calculation methods likeFlowNet2, Farneback, and Brox were compared to identify the best performance with respect to anomaly detection. An extension to this was cycle-consistency which improved the performance of the system along with noise suppression as a post-processing step. For the extended version of the setup, two generators were utilized for training the system, one for transforming the sample from source to the target and another for reconstructing the translated sample to its original domain. Then the cycle-consistency loss is incorporated into the architecture which is the pixel-wise L1 distance between the input sample and its corresponding reconstruction. The results compared with other existing anomaly detection methods achieved high accuracy of 98% outperforming the others. Their results produce very few false positives due to the use of cycle-consistency loss. These models can be deployed in static cameras whereas the major issue that needs to be dealt with is adapting these techniques to non-static cameras which would be a very difficult task.

(Andrej Veľas, Milan Kutaj and Martin Ďurovec [6]) have provided detailed insights on how the change of parameters of the camera system can influence the success rate of motion detection in video-based surveillance systems. Experiments have been performed based on European technical normalities which have been explained in this article confirming the information discussed which also includes a survey being performed by surveillance camera system manufacturers. There are various parameters for CCTV systems like the frame-rate, pixel resolution which has their influence on motion detection. The article explains that video of maximum pixel resolution that the system is able to produce is used and also metadata of lower pixel resolution is captured parallelly. The CCTV system settings are independent of the capturing of video frames. The authors conducted tests on all 4 possible resolutions with frame rate differences of 5 units i.e (1fps, 5fps, 10fps, 15fps) using 9 cameras from 5 distinct camera surveillance system manufacturers. It was found that a frame rate higher than 15 had not much influence on improving motion detection abilities. They also conducted tests at distinct levels of sensitivity settings on independent devices. The most notable observation is that the best motion detection capabilities were found at the highest possible frame rate and resolution. Their research leads to the conclusion that these parameters do have a major impact on the motion detection ability of the system i.e the parameters in which the system operates have a high influence on its success. But even though these parameters have their effects, we also need to consider the working of a particular system in order to gauge its success. So, these attributes along with the way the camera communicates with the Video Management System (VMS) make the system accomplish its purpose.

(Long Liu and Jie Ding [7]) have worked on motion detection using attention temporal-spatial fusion technique and extraction of the object region based on the way the attention is formed. The motion target detection and extraction is possible in two kinds of approaches -

1. Local motion scene - This is done using a stationary camera
2. Global motion scene - This is done using a moving camera.

The authors focus on the latter approach i.e motion detection using global motion scenes. The optical flow of the motion is used for the computation of the motion vectors whose vector fields have sparse and soiled characteristics due to the objects in motion. The vector fields are pre-processed by superposing and applying filters for the frames. The temporal-spatial neighborhood contrast of the motion vector has a direct influence on the motion attention factors. Initially, motion estimation is done using the current, previous, and the next frames which are then pre-processed as mentioned above. The pre-processed motion vector helps in detecting the object region based on the attention equation. This helps in the precise extraction of the target object. The experiments conducted by the authors on 'Stefan' and 'Foreman' (CIF) sequence frames proved that it has good accuracy and performance in real-time detection.

(Muhammad Kamal Hossen and Sabrina Hoque Tuli [8]) have worked on real-time motion detection. Recording the events through surveillance cameras and then playing the recorded tapes to find the events is very tedious. Real-time motion detection is the need of the hour. The authors of this paper propose an efficient technique for real-time motion detection and estimation using the optical flow from the video frames. The researchers use an edge-based approach rather than background subtraction or texture-based boundary evaluation which has its intricacies and complexities associated with it. The hypothesis is that edges are very robust against the noise and change in illumination. The algorithm initially gets the current and the previous frames from the system and computes the frame difference map between the two using the Sobel Edge Detector. This helps in checking if any moving edge exists or not. If a moving edge is found, it sends an alert mail using SMTP configuration. The motion and non-motion frames are saved as separate video files. The relative motion between the frames is calculated using the optical flow estimation method. The traditional Horn-Schunk algorithm is used for optical flow estimation. It identifies the image velocity or motion vector using the Spatial-

Temporal Gradient technique. Gaussian smoothing is applied on each video frame to remove noises that may exist. Motion estimation is done in the following way -

1. Obtain n^{th} and $(n-1)^{th}$ frame from the video
2. Apply gaussian smoothing on the respective frames
3. Estimate spatio-temporal derivatives
4. Compute local averages of motion vector
5. Minimize errors
6. Estimate the motion vector

Experimental results mention that the system performance is very significant in both indoor and outdoor environments during different weather conditions. The Mean Velocity Error and Mean Angular Error over 100 iterations is low and has a reasonable computational cost.

(Gan Sun, Yang Cong, Lichen Wang, Zhengming Ding and Yun Fu [9]) have worked on motion segmentation. Online multi-agent motion segmentation faces some challenges namely Segmentation Space consistency, Distribution Disparity, and Distribution Shift. The model proposed consists of an auto encoder-decoder framework where the encoder projects the motion sequences into common motion-aware space from multiple agents performing different tasks and the decoder reconstructs the obtained data using an additional temporal regularizer using subspace clustering, multi-task clustering, and temporal data clustering techniques. This auto encoder-decoder framework helps in learning common motion characteristics from multiple tasks. The authors explain the solution to their unsupervised segmentation task in detail with respective mathematical equations. Experiments were performed on Keck, MAD, and custom human motion (Electromyography - Vision) datasets to compare the performances against the state-of-the-art models in an offline single task, offline multi-task, and real time multi-task scenarios. The model obtains the best performance amongst all others Keck and EV Action dataset. The average performance over ten random runs, has an improvement of 8.24% and 0.35% in terms of Clustering Accuracy (ACC) and Normalized Mutual Information (NMI) for the Keck dataset, similarly 12.01% and 1.21% improvement for EV-Action dataset respectively. It has an improvement of 6.77% in ACC and 0.64% decrease in NMI on the MAD dataset. The time consumption of the model is very less when compared with other models that perform offline tasks and it achieves similar performance with other online methods. Results indicate that the hyperparameters used to play a significant role in the ACC and NMI performance of the model.

(Prashant W. Patil, S. Murala [10]) have mentioned in their paper that differentiating foreground from the background very efficiently is still a major challenge in moving object segmentation tasks. There can be irregular dynamic movements in the video frames which can affect the accuracy with which the background gets separated from the foreground. Moving object segmentation tasks can be performed at region-level, pixel-level, using frame difference, saliency estimation method, using background subtraction, and deep learning method. The proposed approach consists of cascading two unpaired GAN based networks by adding motion saliency estimation between the two networks. The former generator network performs background estimation using initial video frames. Motion Saliency estimation is a pre-processing step used for capturing relevant information as it provides prominent edges for foreground objects. The latter generator network is used for foreground segmentation. The generator network uses a skip connection which helps pass the feature information to lower layers thus helping to capture as minute details as possible. Comparison of background estimation was performed using pixel-level temporal histogram and GAN based unpaired learning approach in which the latter outperformed to give accurate background in case of highly dynamic, irregular video frame scenarios. The performance of the model was analyzed using the CD-net 2014 dataset, one of the standard datasets for Moving Object Segmentation tasks. The dataset consists of video sequences of different categories like normal background (baseline), dynamic background, bad weather, thermal signatures, intermittent object motion, shadows, and camera jitter. The proposed model is compared with other existing methods like ResNet, VGG-16, and Googlenet. The proposed FgGAN outperforms the existing models by giving an average F-measure score of 0.9697 using all seven categories from the CD-net dataset. The average Percentage of Wrong Classification score for the same is 0.2653 which is very low as compared to other methods. This demonstrates that the proposed unpaired GAN based learning approach is efficient both qualitatively and quantitatively.

(Meimei Gong, Y. Shu [11]) have mentioned in their paper that Surveillance cameras are used everywhere for a wide variety of purposes. It helps us in gathering valuable information about interesting areas and helps in completing the expected goal. In computer vision, this has a vital role in intruder detection, target identification, and so on. The existing methods fail when it comes to detecting these in real-time. To improve the target prediction layer of the Single Shot Detector (SSD) Network, the basic idea implemented here is to examine the multi-scale features and integrating both the low-level detailed features with the high-level semantic features. Without disturbing the spatial resolution of the Single Shot Detector prediction layer, low-scale detailed features are integrated with high-semantic features, for joint, high recognition accuracy of Single Shot Detector network for different target scales. In this way, a feature embedded structure is designed. This is done using Multi-scale feature fusion by pixel-to-pixel correspondence. Improved inception structure is employed to the target prediction layer of SSD with multi-scale feature fusion to improve the accuracy of the human body moving target detection and positioning. To improve the accuracy of target detection of a human body in motion, this paper has come up with an improved inception structure to the target prediction layer of the SSD. In this paper the trained SSD model was pre-trained with all additional network layer parameters being initialized to "Xavier", the initial learning rate is 10^{-3} and the weights attenuated the term is 0.0005, and the momentum term is 0.9. The data amplification adopted is color space transformation and random cropping. The IMFF-SSD network uses the Inception prediction structure at each target prediction layer to predict multi-scale targets, to increase the accuracy of human motion detection. The detection accuracy of this network for human moving targets has reached around 0.956. This shows that the IMFFSSD network is capable of extracting higher semantic features of the human moving targets. Therefore the performance of the IMFF-SSD Networks can be improved using the Inception prediction structure.

(Chengyuan Liu, J. Gong, J. Zhu, J. Zhang and Y. Yan [12]) present in their paper a discussion on the use of Correlation Filter with Motion Detection Algorithms. This algorithm can be employed for motion detection for rapid and robust tracking of the shape-deformed targets. Due to rapid movement, the typical algorithms available are deficient in detecting moving targets. A staple algorithm is used and motion detection coordinates are used to average the output of CF to adjust the output. This way it is possible to obtain rough target positions. Frame differentiating is adopted to detect moving targets in a given sequence of frames. It eliminates the noise caused due to lens shaking or movement in the camera by translation followed by zooming the current image with the previous frame to predict the shaking of the lens. The result is weighted to the average to obtain the target position. Both the motion detection algorithm and CF tracking are combined to obtain an efficient tracking effect.

(Pankaj Agarwal, M. K. Singh [13]) present in their paper a cost-effective, reliable, and robust drone stocked with features for real-time video surveillance and video recording with GPS. The system is capable of sending live video streams to the control center where the data can be processed. One of the major purposes of this drone is to provide real-time information about the surveillant environment efficiently and robustly. The Camera can send video streams in 4K resolution at 60fps or even in 1080p resolution at 240fps. It is water-resistant up to a depth of 10m. The camera is fitted with an electronic image stabilization mechanism by which it can smoothen out the video footage during action of vigorous motion. The camera uses the 5GHz Wi-Fi band for connecting with smartphones, so transferring content is faster to these devices. The camera has an enhanced capability of storing individual frames over a period of a few seconds. An efficient and low-latency transmitter is used to send the controls by decoding it at the control center. On retrieving the video stream from the drone, it can be visualized on any device as long it supports the relevant features.

(Gayathri Devi Ramaraj, S. Venkatakrishnan, G. Balasubramanian and S. Sridhar [14]) have mentioned in their paper that CCTVs tend to play a major role in surveillance and monitoring of surveillant environments. But it fails when it comes to solving crime due to various factors. This paper discusses various enhancements that can be brought about in the surveillance system to revolutionize this domain. Keeping the framework of a quadcopter with autonomous controls and auto-tracking capability as a benchmark, the drone uses an image enhancement algorithm of Probability Hypothesis Density (PHD) filtering using a Markov Chain Monte Carlo (MCMC) implementation. They consider a swarm of quadcopters that uses an Energy-Efficient Coverage Path Planning (EECPP) algorithm. There are some strings attached when it comes to emergency rescue. This means that there are some prerequisites to be taken into consideration before the attempt. One such prerequisite is to maximize the flight duration of the system. With a typical flight time of 25-30 minutes, the drone path to be taken should be optimized such that it efficiently follows the person in command of the rescue. The approach suggested in this paper is to use a dynamic programming approach to maximize the reach of the camera's view for an efficient route planner technique. The drone has to think from the shoes of the rescuer to efficiently sync with their movements. Three algorithms are pre-trained in the drone to achieve this benchmark:

1. Localizer bee
2. Verifier Bee
3. Precise Verifier Bee

These three algorithms are used to efficiently plan the path of movement. The basic idea of this paper is efficient route planning for an escape and also human detection in difficult terrains. Humans or pedestrians in a particular location are identified and marked for selection via the video processing unit on the drone. To detect a human in the frame, it proposes a method to filter the background out of the frame and to detect only the human in the frame.

(Abirami Srinivasan, A. Vijayakumar and A. Makur [15]) have mentioned in their paper that surveillance systems are becoming a vital part of our life. Not only for military purposes, but it is also used at all the places where security is a major concern. Unauthorized personnel access is a major issue prevailing in society and surveillance systems aim to counter that. Drones are becoming a vital solution to curb the issue of surveillance in large and difficult areas. They are capable of detecting the presence of intruders in unauthorized locations and also have an enhanced capability to track them and take appropriate actions like alerting the concerned authorities. The paper strives to search for proliferation in the hardware sector to make this a complete success. The paper implements a compressive sensing-based scheme for intruder detection on the AR Parrot Drone. The proliferation of technology is playing a major role in this Intruder detection scheme as newer technologies are becoming capable of performing this action with utmost efficiency. But still, there are limitations to these advancements as well. For example, imaging radars incur a lot of overhead, due to which the results tend to be less accurate. Surveillance systems on the other hand incorporate HD cameras but tend to be stationary and can easily be bypassed by an intruder. The most challenging solution is the design of a sensing matrix at the encoder side that could provide the capability of hardware friendly and robust reconstruction of the results at the decoder side. In this paper, they have incorporated the use of Gaussian random matrix to reduce the use of large-size data. However, a small-sized ROI can be used which is fed to the Gaussian random matrix. The Gaussian random matrix for the required ROI is stored by default in the system for easy retrieval. In case an intruder is detected, a structured random matrix (SRM) is incorporated to handle large data efficiently without any delay from the decoder side.

III. CONCLUSION

Significant efforts have been made to develop state-of-the-art models for surveillance. Numerous algorithms have emerged in the field of image processing, machine learning, and deep learning techniques to enhance the surveillance mechanisms thereby reducing the data and computation costs. Frame differencing, background subtraction, and statistical methods have played a major role in the development of motion detection systems. Recently advancements in deep learning, especially GANs have replaced these traditional methods. The capability of GANs to provide highly accurate motion masks has made it possible to outperform all the other traditional methods. Optical flow estimation is also being studied by many researchers as it provides significant results in velocity and direction estimation. In this survey paper, we have presented a comprehensive study of state-of-the art models and advancement in research that has taken place in the application of motion detection for surveillance systems. There are many interesting challenges like motion detection in a complex background scenario, abnormal events, occlusion of objects, effective tracking during various intensity changes that keep this domain open for extensive research.

REFERENCES

[1]. al, S.Vish & anth, G.Prash. (2016). Motion Detection Using IoT and Embedded System Concepts. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. 5. 7824-7829. 10.15662/IJAREEIE.2016.0510005.

[2]. R. Cucchiara, C. Grana, A. Prati and R. Vezzani, "Computer vision system for in-house video surveillance," in *IEE Proceedings - Vision, Image and Signal Processing*, vol. 152, no. 2, pp. 242-249, 8 April 2005, doi: 10.1049/ip-vis:20041215.

[3]. O. E. Harrouss, D. Moujahid and H. Tairi, "Motion detection based on the combining of the background subtraction and spatial color information," 2015 Intelligent Systems and Computer Vision (ISCV), Fez, 2015, pp. 1-4, doi: 10.1109/ISACV.2015.7105548.

[4]. M. Sultana, A. Mahmood and S. K. Jung, "Unsupervised Moving Object Detection in Complex Scenes Using Adversarial Regularizations," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2020.3006419.

[5]. Golda, Thomas & Murzyn, Nils & Qu, Chengchao & Kroschel, Kristian. (2019). What goes around comes around: Cycle-Consistency-based Short-Term Motion Prediction for Anomaly Detection using Generative Adversarial Networks. 1-8. 10.1109/AVSS.2019.8909853.

[6]. A. Vel'as, M. Kutaj and M. Ďurovec, "Influence of changing the parameters of the camera system on video-based motion detection," *2017 International Carnahan Conference on Security Technology (ICCST)*, Madrid, 2017, pp. 1-5, doi: 10.1109/CCST.2017.8167829.

[7]. L. Liu and J. Ding, "Target Detection and Extraction Based on Motion Attention Model," *2018 Chinese Automation Congress (CAC)*, Xi'an, China, 2018, pp. 1919-1922, doi: 10.1109/CAC.2018.8623780.

[8]. M. K. Hossen and S. H. Tuli, "A surveillance system based on motion detection and motion estimation using optical flow," *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, Dhaka, 2016, pp. 646-651, doi: 10.1109/ICIEV.2016.7760081.

[9]. G. Sun, Y. Cong, L. Wang, Z. Ding and Y. Fu, "Online Multi-Task Clustering for Human Motion Segmentation," *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Korea (South), 2019, pp. 970-979, doi: 10.1109/ICCVW.2019.00126.

[10]. P. Patil and S. Murala, "FgGAN: A Cascaded Unpaired Learning for Background Estimation and Foreground Segmentation," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, 2019, pp. 1770-1778, doi: 10.1109/WACV.2019.00193.

[11]. M. Gong and Y. Shu, "Real-Time Detection and Motion Recognition of Human Moving Objects Based on Deep Learning and Multi-Scale Feature Fusion in Video," in IEEE Access, vol. 8, pp. 25811-25822, 2020, doi: 10.1109/ACCESS.2020.2971283.

[12]. C. Liu, J. Gong, J. Zhu, J. Zhang and Y. Yan, "Correlation Filter With Motion Detection for Robust Tracking of Shape-Deformed Targets," in IEEE Access, vol. 8, pp. 89161-89170, 2020, doi: 10.1109/ACCESS.2020.2993777..

[13]. P. Agarwal and M. K. Singh, "A multipurpose drone for water sampling & video surveillance," *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, Gangtok, India, 2019, pp. 1-5, doi: 10.1109/ICACCP.2019.8883017.

[14]. G. Devi Ramaraj, S. Venkatakrishnan, G. Balasubramanian and S. Sridhar, "Aerial surveillance of public areas with autonomous track and follow using image processing," 2017 International Conference on Computer and Drone Applications (IConDA), Kuching, 2017, pp. 92-95, doi: 10.1109/ICONDA.2017.8270406.

[15]. A. Srinivasan, A. Vijayakumar and A. Makur, "Compressive Sensing-based On-board Intrusion Detection for a Surveillance Drone," *2019 IEEE VTS Asia Pacific Wireless Communications Symposium (APWCS)*, Singapore, 2019, pp. 1-5, doi: 10.1109/VTS-APWCS.2019.8851619.

Dheeraj N Bhat, et. al. "Motion Detection for Surveillance Systems: A Survey." *International Journal of Modern Engineering Research (IJMER)*, vol. 11(01), 2021, pp 01-08.