RESEARCH ARTICLE OPEN ACCESS

Object Detection in Video Streaming Using Deep Learning

Shaba Irram [1], Sheikh Fahad Ahmad [2]
Department of Computer Science and Engineering
Integral University, Lucknow, U.P
India

ABSTRACT
When we see any image, it is the work of our brain to instantly recognize the objects contained in it. But for machine, object identification requires data training and more time. This computer vision field has become easier with the technological advancement in hardware resources based on deep learning. When we capture a video steam of any moving object using video camera then many difficulties and challenges will arises. Nowadays detecting objects in video streaming is adopted in the security services to monitor the sensitive areas including Highways what does various public places and banks. Detecting objects in identifying other activities in the surrounding is an important part of a machine to interact with human in a very easy manner and this is the capability of a machine who identify target. Our work is to focus on achieving more accuracy rate in object detection in videos. To achieve the stars many approaches has been analysed to detect objects in videos.
Keywords :- Video streaming, deep learning, moving object detection

I. INTRODUCTION

Object recognition become crucial when we move towards complete image understanding. Deep learning method has achieved success in various computer vision task for image classification to object detection. Deep learning is a popular computer vision a source of research has been used to explore deep learning for resolving face detection task. Object tracking is the process in which moving object is detected in each stream of a video by using a camera. The first step towards detecting the object is tracking in the frames after this step the detected object will be divided as birds, humans, vehicle and soon. It is challenging in image processing approach for cracking the objects using video sequences. The tracking procedure determines the orientation of object over the time as object moves throughout the scene. A reliable tracking method is used which meet the real time restrictions and are challenging and complex with respect to the change of object movement, scale and appearance.

A.Dimou et al. [3] proposed methodology to enhance the productivity of CCTV footage to detect multiple objects. Different data sets with motion blur are researched to train the detectors. The performance to detect blurred content and original content was strong. Furthermore, a novel methodology during sharp PTZ(Pan, Tilt and Zoom) operations is proposed to dynamically raise the features of detector. The spatial transformation of the targets is practiced to train the RNN to prefigure the integrated camera features in each frame. Researchers have demonstrated that real time scaling significantly enhances the functioning of the object detector compared to static scale operations.

L.Wu et al. [16] offers embedded video surveillance based program that supported moving objects detection. With assistance of the processing of Servfox streaming media on ARM-Linux based framework, it made possible to get the videos and access of USB camera. This research paper follows the updated method of the background model of inter-frame difference to find the active targets. The experiment indicates that the images which are at the back can be found with accuracy. So far, the important algorithms for detecting movable targets are background difference method and inter-frame difference method [17,19]. The background difference technique can find moving objects information comparatively all over, though instead hard in finding and upgrading the background model. But the inter frame difference technique has average object speed. The process of video monitoring system is digital, united and smart [18].

L.H.Jadhav et al. [4] introduces a way for detection of unnoticed or removed targets in the video system. For detection of non moving target, hybrid model is

used. Once nonmoving target is found its features are carried out to classify the target. Features extracted are height, width, size, color and time. For accuracy a lot of parameters like human, luggage relative, human attitudes can be applied.

K.Tahboub et al. [9] proposed a convolutional neural network with two-stage quality-adaptive to cover the challenges of altering data rate of a video. Video compression is defined on a state of the art detection of person on foot and looked into compressed images for training. Compression of video can bring out poor quality that affects accuracy rate of video processing. In this research paper, we analyze the issue of a changing object rate of video and study how it involves the functioning of video processing, particularly detection of pedestrian, using a convolutional neural network with two-stage quality-adaptive system.

A.Ucar et al . [20] proposed a methodology for Self-directed driving which involves correct identification of targets in actual driving environments. In this paper, we offer a Convolutional Neural Network (CNNs) with new hybrid Local Multiple system (LM-CNNSVM) and Support Vector Machines (SVMs) of their strong characteristic extraction capability and sturdy classification feature, respectively. In the suggested system, we separate the entire image into local domains and use multiple CNNs to gain local target characteristics. Secondly, we take featured characteristics by applying Principal Component Analysis. Then we bring into multiple SVMs practicing both real and functional risk reduction rather than applying a direct CNN to raise the generalization power of the classifier system. At last, we mix SVM outputs. Additionally, we use the pre-trained AlexNet and a new CNN framework. We execute object identification and person detection experiments on the Caltech-101 and the Caltech Pedestrian datasets.

B.Tian et al. [21] proposed a methodology for a video supported target detection technique with deep learning technique. The observation video is assembled right off the bat, from which a commented on picture database of point target with the end goal that as individuals or vehicle was worked to prepare convolutional neural net structure disconnected. With the prepared structure, an ongoing item location and ID framework is anticipated and connected. The proposed system for the most part incorporates three perspectives: video handling, object recognition and target distinguishing proof. It offers a type of video interfaces to help the extracted video clips and continuous video stream. The information based outcomes demonstrate that the proposed profound training based location system is viable for the recognizability application.

B.Hio et al. [11] proposed a methodology for background subtraction from the applied picture for moving target detection. In this paper, we recommend another moving target location approach applying profound figuring out how to achieve a hearty act even in a functioning foundation. The proposed methodology accepts visual viewpoint qualities just as movement attributes. To this end, we reason a profound learning engineering accumulated of two systems: a visual perspective system and a movement organize. The two systems are made to discover moving target powerfully to the foundation movement by applying the presence of the objective item notwithstanding the movement contrast. In the examination, it is seen that the proposed technique achieves 50fps speed in GPU and out performs cutting edge systems for various moving camera recordings.

S.Hayat et al. [13] proposed a methodology for practiced deep learning to how to multi-class target distinguishing proof and research convolutional neural network system (CNN). The convolutional neural network system is generated with standardized grade instatement and prepared with preparing group of test pictures from 9 diverse target classes in addition to test pictures utilizing generally changed dataset. All outcomes are completed in python tensor stream system. We break down and contrasted CNN outcomes and last trademark vectors extricated from various methodologies of the BOW dependent on direct L2-SVM classifier.

H. S. G. Supreeth et al. [22] proposed a methodology for Gaussian mixture framework (GMM) supported target detection, deep learning neural network system based identification and tracking of targets using correlation filter is suggested, that can deal with phony discoveries, with bettering the productivity. The calculation is anticipated to discover just vehicles and people when the working is examined utilizing “True Positive Rate” (TPR) and “False Alarm Rate” (FAR) as probabilistic measurements.

C. Li et al. [14] shows a target detector supported deep learning of small samples. As a matter of first importance, the calculation can increase preparing tests naturally by manufactured examples generator to determine the issue of few examples. Manufactured examples generator is anticipated by exchanging the objective territories in various scenes. Along these lines, profound supervised learning and thick expectation form are rehearsed in the profound convolution neural network system.

II. APPLICATIONS OF OBJECT DETECTION

Some of the applications for detecting people and vehicles are the following:

2.1 Access control in sensitive locations

When anybody is entering in some delicate areas, for example, some administration workplaces, military units then the framework could naturally store all the element of that individual like stature, facial appearance and so on from pictures taken continuously, and after that choose whether that individual can permit to enter.

2.2 Personal identification in some scene

In some scene distinguishing proof of individual explicit by a reconnaissance framework can assist the police with catching suspects. That individual can be perceived and judge naturally by the brilliant reconnaissance framework regardless of whether that individual is suspects.

2.3Target detection and alert

It is the process by which object is detected and its behaviour is analyse and determine whether it is normal or not. For alarming two ways are there. First, when any abnormal behaviour of an object is detected then a recorded announcement is generated publicly. Second, it automatically contact to the police.

2.4 Multiple cameras for video surveillance

Different cameras could be utilized to for the security of specific locale. For traffic the board, numerous cameras can enable the traffic to police find, track and catch vehicles engaged with traffic offenses.

III. CHALLENGES IN OBJECT DETECTION

The smart video surveillance can find moving target. To find the exact location of the moving target and also to reduce the excess computation for the false motion of target some challenges are faced by many approaches, they are:-

3.1 Illumination difficulties: There has ever a plausibility of presence of an alternate item or foundation with similar shading information.

3.2 Dynamic Background: Any areas of the scene may contain movement (Water wellspring, moving mists, wave of water and so on.), yet ought to be seen as foundation, as per their significance.

3.3 Occlusion

Occlusion (fractional/full) can influence the activity of figuring the foundation outline.

3.4 Background Clutter

Presence of foundation mess causes the errand of apportioning hard

3.5 Presence of Shadows

Shadows draw by frontal area targets every now and again confuse further working advances resultant to foundation subtraction. Covering shadows of frontal area zones for example square their partition and arrangement. Research worker have recommended diverse systems for recognition of shadows.

3.6 Video Motion

Video can be caught by shaky (for instance. Moving) cameras.

3.7 Changing weather conditions

Detection of active target turns a very hard job when videos are captured in difficult weathers.

Table 1. Overview of challenges

CHALLENGES	AUTHORS
Illumination challenges	N. K. Patil et al.[23]
Dynamic Background	B.Hio et al. [11]

Occlusion	J. Pan et al. [24]
Background Clutter	A. Zaimbashi et al. [25]
Presence of Shadows	R. Geu et al. [26]
Video Motion	R. Xu. Et al. [27]
Changing weather conditions	L. Snidaro et al. [28]

IV. VARIOUS TECHNIQUES FOR OBJECT DETECTION

Firstly, the trained dataset is created by taking different images of Objects. Then framing is done to extract frames from the videos. Here we focus on the methods that find the targets by geometric shapes and their movement is identified between consecutive frames, i.e. called frame to frame tracking. In next step object detection algorithm will be applied. Target detection means to localize targets in an input image. Then find out some prediction about the object. Object detectors are trained by assuming that all training examples are labeled. Apply algorithm to find the optimal answer. Then comparison is done with another algorithm and get the accuracy rate.

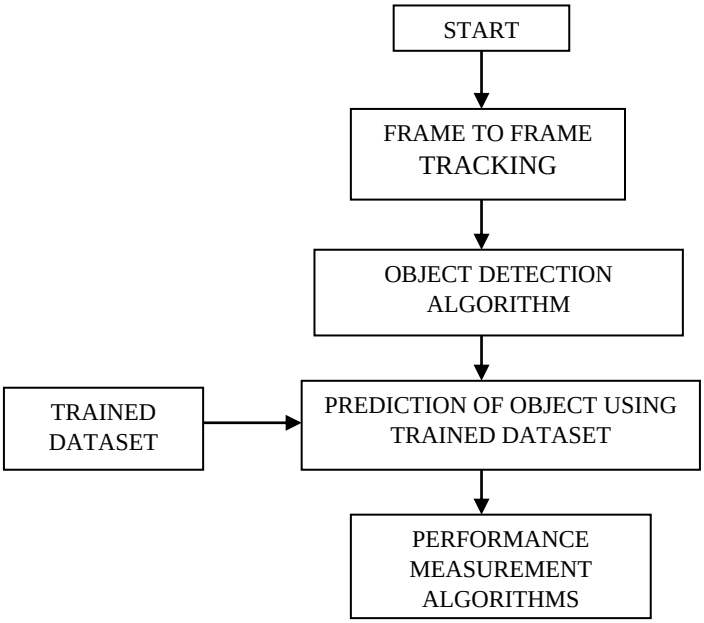

Fig. 1. Work Flow Chart

4.1 TLD Framework

It decays the long haul following employment into three sub-errands: following, learning and recognition. The tracker keeps up the item from casing to outline. The identifier sets total visual aspect* that have been seen up 'til now and amends the tracker at whatever point required. The learning ascertains identifier's missteps and updates them to keep off these slip-ups in the great beyond. TLD contains a model anticipated for long haul following of an anonymous physical article in a video stream [1].

4.2 P N Learning

This object detection technique focus on the learning attributes of the TLD demonstration. The objective of the factor is to revise the working of a physical article identifier from web handling of a video clip. The key thought of P N learning speaks to that the indicator issues are searched by two kinds of "specialists". P-master identifies just “false negatives”, Nexpert identifies just “false positives”. Both of the specialists produce flaws themselves, all things being equal, their independency empowers shared remuneration of their faults[1].

4.3 Training Data Augmentation

CCTV video recording often times include extremely obscured items inferable from poor video recording quality and quick PTZ functioning. Especially movement obscure is an extraordinary test for article identification in CCTV subject. While de-obscuring methodological analysis present good results [2], they've an extraordinary computational expense and they further undermine their visual perspective. However, in testing CCTV video recording execution drops. In this part, the after effect of information determination in the detector’s working is researched [3].

4.4 Dynamic Detector Configuration

CCTV cameras much of the time have Pan Tilt Zoom possibilities that accessed by their administrators to follow process in a view. These camera functioning contain a genuine test for item identification and following because of the

inexplicit scale suppositions caused. In this part it is anticipated to powerfully set this scale run bolstered forecasts of the followed articles' size in the succeeding edge. The first pace towards this methodology is to have a definite estimate of the recognized article's scale [3].

4.5 Foreground Blob Extraction Technique

The foreground blob extraction is made for that we structure two convenient refreshing foundations; one for small size and other for large size through modifying learning rates of foundation refreshing. After foundation displaying, foundation subtraction is executed to take out forefront objects. To draw definite closer view object shadow evacuation is utilized, shadow expelling technique wipes out the shadow pixels point that are seen as forefront at time of foundation subtraction process [4].

4.6 Pedestrian Detection using Compressed Video Sequences

Deep convolutional systems [5] (known as VGG), Fast/Faster Region-based Convolutional Neural Network system (R-CNN) [6,8] have exhibited excellent execution for substantial scale target acknowledgment. In [7], Fast/Faster R-CNN systems are broke down and embraced for person on foot detection [9].

4.7 The new hybrid Local Multiple Convolutional Neural Network System(LMCNN)

CNNs comprise of multiple layers similar to feed forward neural network. The outputs and inputs of the layers are made as a set of image arrays. CNNs can be made by various combinations of the convolutional layers, pooling layers, and fully connected layers with point-wise non linear activation functions

Input layer: Images are direct imported to input of the network.

Convolutional layer: This layer executes main workhorse functioning of CNN. [10,12]

Pooling layer: This layer is used to scale down the feature proportion.

Rectified Linear Units (ReLU) layer: This layer includes units that applying the rectifier to accomplish scale invariableness. The activation function of this layer is mathematically identified as $f(x) = \max(0,x)$ for an input x.

Fully Connected layer: After convolutional network, max pooling, and ReLU layers the fully connected layer is attached.

Loss layer: A loss function is used to calculate variance between the prediction of CNN and the real object at the final layer of CNN.

4.8 The Appearance Network(A-net) and the Motion Network (M-net)

The A-net is a convolutional neural network system that executes visual perspective based article identification. To concentrate on visual part of elements, the A-net allows just a single picture without foundation data. Along these lines, the A-net watches the visual part of moveable articles (individuals, vehicles, creatures, etc.)[11].The M-net is a system that sees development. Other than the information picture, the M-net acquires the foundation picture got from the foundation model [4]. The M-net is like the foundation driven methodology in that it utilizes the distinction among foundation and image [11].

4.9 Regularization Techniques

Information growth has far been connected for preparing CNN. In the field of PC visual sensation, it is accepted universal because of its ease to orchestrating it into exercise and utility. We executed constant information expansion with pictures while preparing the structure which may turn, rescale and downy it. A strategy is utilized in picture visual methodology territory while preparing profound learning structure. In this work we connected the dropout strategy in comparable manner [13].

4.10 Synthetic Samples Generator

In this technique first of all, object fields are took out by the preparation set, and the closer view objects are found after division of the article fields. In this way, arbitrarily forefront objects are joined to foundation pictures to produce engineered pictures [14]. This technique has the following two processes.

1) Foreground Objects Extraction: Source targets of engineered pictures are following from the item fields in the preparation set [14].

2) Background Selection and Fusion Processing: The foundation pictures that involve indistinguishable class from the closer view objects are chosen. And after that the fields in foundation pictures are substituted by fields which are chosen aimlessly from closer view objects [14].

Table 2. Overview of Techniques

Techniques	Researchers
P-N Learning	Zdenek Kalal et al. [1]
Training Data Augumentation	Jian-FengCai et al. [2]
Dynamic Detector Configuration	A. Dimou et al. [3]
Foreground Blob Extraction	L. H. Jadhav et al. [4]
Compressed Video Sequences	K. Tahboub et al. [9]
LMCNN	A. Uçar et al. [20]
A –Net and M- Net	B. Heo et al. [11]
Regularization Techniques	S. Hayat et al. [13]
Synthetic Samples Generator	C. Li et al. [14]

V. EFFECIENCY OF OBJECT DETECTION METHOD

There are 3 critical kinds of execution measurements in our framework: detection-based metrics; tracking-based metrics; and perimeter violation detection metrics. The detection-based metrics are connected to gauge the working of a System Under Test (SUT) on isolated edges from video detector data. Every one of the objectives is independently inspected to check whether there's a match between the SUT and the Ground-truth (GT) framework for every video outline. The tracking-based metrics utilize the personality and the all over direction of all items independently over the test grouping and contrast the GT tracks and the SUT tracks upheld best symmetry. At that point, in view of the best matches, a few blunder rates and execution measurements, spoke to underneath, are determined. Finally, the perimeter violation detection measure depends on finding any objective as it moves into a predefined area.

1. False Positive Rate (FPR)
It is an act measurement which estimates how well the framework effectively rejects.
"FPR = FP/(FP + TN)", FP is false positives and TN is true negatives.

2.False Alarm Rate (FAR)
It is an act measurement which estimates that the recognized item is right.
"FAR = FP/(TP+FP)", FP is false positives and TP is true positives.

3.Detection Rate (DR)
It is an act measurements which estimates level of genuine focuses on that is identified.
"DR = TP/(TP+ FN)", TP is true positives and FN is false negatives.

4.False Negative Rate
"False Negative Rate = FN/(TP+FN)", FN is false negatives and TP is true positive.

5. True Negative Rate (TNR)
"TNR = TN/(TN + FP)", the true false endless supply of the true false identifications and the false positive.

6. Accuracy
"Accuracy = (TP+TN)/CGT", the aggregate of the true positives and the true negatives partitioned by the complete number of GT objects.

7. Precision
"Precision= TP/(TP + FP)", the quantity of true positives upon the aggregate of the true positives and the false positives.

8. Recall
"Recall= TP/(TP + FN)", the quantity of true positives upon the aggregate of the true positives and the false negatives.

9. ROC (Receiver Operating Characteristic) Curve
It is a graphical result that analysis the recognition rate and False Positive Rate

VI. CONCLUSION AND FUTURE WORK

Now days, object detection is in use in various fields, from security services to transportation. If the detection of an object is not accurate then the problems (like road accident) can occur. So, Detecting the objects with more accuracy rate is very important task of object detection.
From many previous research papers it is concluded that the optimal results were not obtained for

accuracy rate of object detection. In future, this literature survey can be used to increase the accuracy rate in detecting the object in the videos so that the optimal result can be achieved**.**

REFERENCES

[1] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas," Tracking-Learning-Detection", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 6, NO. 1, JANUARY 2010

[2] Jian-Feng Cai, Hui Ji, Chaoqiang Liu, and Zuowei Shen, "Framelet-based blind motion deblurring from a single image," Image Processing, IEEE Transactions on, vol. 21,no.2,pp.562–572,2012

[3] A. Dimou, P. Medentzidou, F. Á. García and P. Daras, "Multi-target detection in CCTV footage for tracking applications using deep learning techniques," 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, 2016, pp. 928-932.

[4] L. H. Jadhav and B. F. Momin, "Detection and identification of unattended/removed objects in video surveillance," 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, 2016, pp. 1770-1773

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Proceedings of the International Conference on Learning Representations (arXiv:1409.1556), May 2015, San Diego, CA

[6] R. Girshick, "Fast R-CNN," Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448, December 2015, Santiago, Chile.

[7] L. Zhang, L. Lin, X. Liang and K. He, "Is faster R-CNN doing well for pedestrian detection?" Proceedings of the IEEE European Conference on Computer Vision, pp. 443–457, October 2016, Amsterdam, Netherlands.

[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," Proceedings of the Advances in Neural Information Processing Systems Conference, pp. 91–99, December 2015, Montr´eal, Canada.

[9] K. Tahboub, D. Güera, A. R. Reibman and E. J. Delp, "Quality-adaptive deep learning for pedestrian detection," 2017 IEEE International Conference on Image Processing (ICIP), Beijing, 2017, pp. 4187-4191.

[10] Deng L and Yu D. Deep learning: methods and applications. Foundat Trend Signal Proc 2014; 7: 3–4.

[11] B. Heo, K. Yun and J. Y. Choi, "Appearance and motion based deep learning architecture for moving object detection in moving camera," *2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, 2017, pp. 1827-1831.

[12] Christian S, Toshev A and Erhan D. Deep neural networks for object detection. In: Proceedings of advances in neural information processing systems (NIPS), Lake Tahoe, Nevada, 5–10 December 2013, pp.2553–2561. Red Hook, NY: Curran Associates Inc.

[13] S. Hayat, S. Kun, Z. Tengtao, Y. Yu, T. Tu and Y. Du, "A Deep Learning Framework Using Convolutional Neural Network for Multi-Class Object Recognition," *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, Chongqing, 2018, pp. 194-198.

[14] C. Li, Y. Zhang and Y. Qu, "Object detection based on deep learning of small samples," *2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)*, Xiamen, 2018, pp. 449-454.

[15] S. Rosi*, W. Thamba Meshach**, J.Surya Prakash**" A Survey on Object detection and Object tracking in Videos", International Journal of Scientific and Research Publications, Volume 4, Issue 11, November 2014 ISSN 2250-3153

[16] L. Wu, Z. Liu and Y. Li, "Moving objects detection based on embedded video surveillance," 2012 International Conference on Systems and Informatics (ICSAI2012), Yantai, 2012, pp. 2042-2045.

[17] ZH. J. Zhao, X.ZH. Lin, J.Y. Zhang, "The moving object detection algorithm Based on the background reconstruction," Journal of Beijing Institute of Petro-chemical Technology. Beijing, vol. 16,pp 27-29, February 2008.

[18] Z. D. Ma, ZH. B. Zhang, "Software Realization of Video Surveillance Terminal Based on ARM- Linux," Computer Measurement & Control .Beijing ,vol.19,pp.456458, February 2011.

[19] N. Paragios , R. Deriche," Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects," IEEE T PATTERN ANAL.USA, vol. 22,pp. 266-280, March 2000.

[20] A. Uçar, Y. Demir and C. Güzeliş, "Moving towards in object recognition with deep learning for autonomous driving applications," 2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA), Sinaia, 2016, pp. 1-5.

[21] B. Tian, L. Li, Y. Qu and L. Yan, "Video Object Detection for Tractability with Deep Learning Method," 2017 Fifth International Conference on Advanced Cloud and Big Data (CBD), Shanghai, 2017, pp. 397-401.

[22] H. S. G. Supreeth and C. M. Patil, "Moving object detection and tracking using deep learning neural network and correlation filter," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 1775-1780.

[23] N. K. Patil, S. Vasudha and L. R. Boregowda, "A Novel Method for Illumination Normalization for Performance Improvement of Face Recognition System," 2013 International Symposium on Electronic System Design, Singapore, 2013, pp. 148-152

[24] J. Pan, B. Hu and J. Q. Zhang, "Robust and Accurate Object Tracking Under Various Types of Occlusions," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 2, pp. 223-236, Feb. 2008

[25] A. Zaimbashi, "Target detection in clutter background: Null or whiten the clutter," 2016 24th Iranian Conference on Electrical Engineering (ICEE), Shiraz, 2016, pp. 1963-1966.

[26] R. Guo, Q. Dai and D. Hoiem, "Paired Regions for Shadow Detection and Removal," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 12, pp. 2956-2967, Dec. 2013.

[27] R. Xu, Z. Zhang, F. Qin, Z. Zhou, H. Yang and R. Li, "Research on video motion compensation method based on the analysis of airborne video camera shaking mode of continuous video frames," 2016 IEEE PES 13th International Conference on Transmission & Distribution Construction, Operation & Live-Line Maintenance (ESMO), Columbus, OH, 2016, pp. 1-4.

[28] L. Snidaro, R. Niu, P. K. Varshney and G. L. Foresti, "Automatic camera selection and fusion for outdoor surveillance under changing weather conditions," Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, 2003., Miami, FL, USA, 2003, pp. 364-369.

Authors Profile

Mrs. Shaba Irram has completed Bachelor of engineering in computer science from University of Rajasthan in year 2009. She is currently pursuing Master of Technology in computer science and engineering from Integral University, Lucknow, Uttar Pradesh, India,