An Overview of Methodologies for Moving Object Detection

Kishor Dhake and S Dalu
Department of Computer Science and Engineering,
Prof Ram Meghe College of Engineering and Management, Badnera Amravati
kishorgdhake@gmail.com
Department of Electronics and Telecommunication Engineering, Government Polytechnic, Yavatmal
surendradalu@yahoo.co.in

Abstract— ***Detection of moving objects or change detection has long been the standpoint for surveillance applications. In addition, this step is the building block for high level algorithms such as tracking. In this paper various methodologies for moving object detection are reviewed. This paper gives the comparison of various methodologies namely deep convolution networks, four-frames differencing, tracking size estimation, SVS for UAVs, moving direction identification using depth image, visual attribute classification using feature selection based convolution neural network, and unusual event detection for moving object detection.***

Index Terms—**convolution networks, neural networks, visual attribute classification, moving object detection, size estimation.**

I.INTRODUCTION

According to natural human perceptions, moving objects movement direction is more important than object detection. When a person walks, s/he concerns about moving people and other moving objects towards him/her rather than detecting and identifying what is situated around. Detecting moving object fast and accurately from airborne videos is one challenging task in computer vision. Unlike applications with fixed cameras, it is required to handle more tough problems for the aerial surveillance, such as camera motion and poor video quality caused by atmospheric turbulence. To solve these problems, much work has been done for moving object detection from airborne videos. These methods can be mainly classified into three categories.

The first category of methods is based on dense optical flow analysis. Since computing dense optical flow is time-consuming, the optical flow based methods are hard to be used for real-time applications without hardware acceleration.

The second category of methods is based on frame difference, which is relatively efficient. The main drawback is their poor ability at extracting all the relevant moving pixels, leaving holes in the foreground objects. Another drawback is that they usually detect a counterpart (often called ghost) from the background regions. Though much work has been done to deal with the holes and ghost it is still difficult to overcome their inherent shortcomings.

The third category of methods is based on background subtraction. On account of their good performance, they have been successfully applied to moving object detection from surveillance videos captured by fixed cameras.

II. Target Detection and Recognition with Deep Convolution Networks

The biggest shortcoming of traditional visual monitoring system is that it can only provide for video capture, data storage and video replay. Therefore, facing the widespread application of intelligent video surveillance system (IVSS) at point of intelligence gathering, crime prevention, protection of person or investigation of crime, etc. It's essential to pay more attention to the emerging research in the field of computer vision in order to provide effective services of intelligent video surveillance. Achieving accurate recognition in real time is a challenge due to multi-objects in complicated background. Previous works using supervised learning such as Support Vector Machine (SVM), Adaboost and unsupervised learning such as Latent Dirichlet Allocation (LDA), Principal Component Analysis (PCA), are not suitable in the real-time object detection and recognition since they can only finish recognition with a low mean average precision (mAP).

To build the IVSS [1], the first step is to find an efficient way to detect the objects (the low level). R-CNN firstly combines the bottom-up region proposals and convolution neural networks (CNNs) for localising objects and extracting the proposal features. Thanks to the CNN shared parameters and the low-dimensional feature vectors computed by the CNN, the run-time is quite satisfying as compared to the traditional ways. The second step is target recognition after finishing detection (the middle level). RCNN uses class-specific linear SVMs with the extracted proposal features for classification.

The SPP-net uses the spatial pyramid pooling layer instead of the pooling layer for the network. After that, we can load any size of images of the VISS. It sounds like that we could use the SPP-net for our DIVS, but the key issues of the IVSS is that we want to get a quick reaction without losing accuracy (high mAP on classification).

As mentioned above, the bottleneck of the object detection networks like SPP-net and R-CNN is the region proposal computations. The recent have provided several deep networks ways for detecting classes. For sharing computation with the fast R-CNN, Shaoqing Ren et al. designs the region proposes networks (RPNs) for real-time object detection in the Faster R-CNN based on the recent success of region propose methods.

In the most cases to build the unified network, stochastic gradient descent and back-propagation are used for the RPN training, the RPN then provides the proposals for training Fast R-CNN. After one-iteration, Fast R-CNN has tuned the detector network which will initialize RPN for the next training. In DIVS, at first the Faster R-CNN pre-trained models are used for initializing model, then the DIVS will generate images as further training and validation sets. The above mentioned methods are used for training proposed model.

Caffe [2] is used as deep learning framework, and it is an open source project which is developed by Berkeley Vision and Learning Center. Recently much other deep learning architecture have joined the open source communities such as TesnsorFlow of the Google Brain team, CNTK of Microsoft Research, MXNet and so on. The reason Caffe chosen is: (i) The expressive architecture encourages the DIVS application, it is very convenient to design and train model; (ii) The extensible code fosters the framework tracks the state-of-the-art; (iii) The most important reason is about the speed. DIVS is tested with a NVIDIA K40 GPU on Ubuntu.

Fine-tuning usually uses an existing deep neural architecture to repeat training from the learned model weights. Since the model has been trained well by the Image Networks for object classification, it has been determined to use the pre-trained architecture and then fine-tune it to build our DIVS. In video surveillance system, there are only seven main classes (person, bicycle, tricycle, motorcycle, car, truck, tractor) for recognition on the road and each class have about 7000 images for training purpose. The pre-trained weights based on the ImageNet images will be loaded into model and then fine-tuning

will be implemented.

As mainly 7 classes are predicted, firstly need to change the layer in the pre-trained model. Secondly, also the overall learning rate is decreased in the solver prototxt of the caffe framework. The reason is that it is needed to change the rest of the model slowly with the new data, but let the new layer learn fast. Thirdly, the learning rate has to go down faster, step size is set in the solver to a lower value instead of training from scratch.

III. FOUR-FRAMES DIFFERENCING

In most cases, surveillance systems rely on stationary cameras. Therefore, background subtraction techniques perform well, because the background model is fixe and any change from this model is easily detected as foreground (moving objects). In contrast, detection of moving objects in non-stationary cameras, as in WAMI [3], is more difficult because of camera jitter as the aircraft is hovering over a region of interest. Besides, there is parallax effect problem i.e., buildings and trees are moving which results in lots of false positives. A ghost is "a set of connected points, detected as in motion but not corresponding to any real moving object". In fact, three-frames differencing technique was proposed to mitigate this explicit handling of ghosting, although, it is identified that even three-frames differencing technique suffers from the ghost problem.

Three-frames differencing technique is limited because it does not contain enough temporal information to overcome completely ghosting problem. Then, a moving-objects detection algorithm is proposed that is based on four-frames differencing. These four frames represent the temporal information needed to avoid the ghost problem, which three-frames differencing technique suffers from. In addition, the proposed algorithm is pixel-based. Hence, it supports parallel processing which is beneficial for efficiency.

The simplest method is to use the temporal median of the last n frames as the background model. The main disadvantage of this method is its need for a buffer of recent pixel values of the last n frames. In addition, it does not provide any deviation measure. Generally, in an outdoor scene it is difficult to find a single value for the same pixel over time and many examples could be drawn from tree leaves, snowing, raining, or sea waves. Therefore a single valued background model is inadequate. This motivated Stauffer and Grimson to use a multi-valued background model. This model uses the last n background values. This model is also quite complex because it uses spatial correlation and has a high memory requirement.

Moving objects detection can be performed using simply two-frames differencing with explicit handling of ghosting, however, this method suffers from false negatives (missed detections). Before detecting moving objects, method should compensate for camera motion (video stabilization). However, in surveillance applications, it starves for an online video stabilization algorithm i.e., no need for incoming frames to stabilize the current frame. After compensating for camera motion, detection of moving objects was performed. Actually, the limitation of three frames differencing technique comes from the typical scenario in highways (vehicles have the same velocity).

IV. TRACKING AND SIZE ESTIMATION

Detecting and tracking moving objects in videos is a challenging problem in Computer Vision and Robotics. In its simplest form tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. In other words, a tracker programme assigns consistent labels to the tracked objects in different frames of a video. Along with detection and tracking, classification of objects can lead to a better understanding of the underlying scene. After detecting and tracking objects, some features of the objects like color, shape, size, etc. can be used to

compare, classify the objects. One important feature among these features is the size of an object in motion.

Table 1: Comparative Analysis of Methodologies for Moving Object Detection

Title	Aim	Methodologies Used	System Name	Purpose for which it is generated	Experimentation Carried Out	Features	Experimentation/ Survey Result
Real-Time Target Detection and Recognition with Deep Convolutional Networks for Intelligent Visual Surveillance	Moving target detection and tracking, recognition, behaviors analysis are the key issues in the intelligent visual surveillance system (IVSS).	Centroid, Orientation, SIFT, color histogram, entropy, homogeneity	Not Mention	Educational applications	Tested by NVIDIA GeForce GTX 750 and NVIDIA Tesla K40	Faster R-CNN	Provided new insights about the IVSS (Intelligent visual surveillance system)
A Four-Frames Differencing Technique for Moving Objects Detection in Wide Area Surveillance	To show limitation of the widely used three-frames differencing technique	pixel-level algorithm based on four-frames differencing	Not Mention	Educational applications	compared our method to two background subtraction techniques GMM, temporal median, and the baseline threeframes differencing technique	FOUR-FRAMES DIFFERENCING	proposed a novel method to mitigate the problem of ghosting that three-frames differencing technique
Tracking and Size Estimation of Objects in Motion	Detection and tracking of moving objects in video by considering the size of an object in motion	Horn and Schunk Algorithm, Lucas and Kanade Algorith	Not mention	Educational Applications	experimented the proposed solution in Python with OpenCV using the changedetection.in data set	Motion Detection, Segmentation Size Estimation	Experimentation shows that it is better than the existing system.

		m, Gunnar Farneback Algorithm					
Object Detection and Location Estimation using SVS for UAVs	Distinct system for object detection and measuring distance to an object for an Unmanned Aerial Vehicle (UAV) utilizing stereo vision sensor	STEREO VISION SYSTEM (SVS)	Not Mention	Educational Applications	Stereo vision system mounted on a Hexacopter used for estimating the object distance from camera	-	-
Detecting Real Time Object Along with the Moving Direction for Visually Impaired People	develop a suitable and effective technique for moving object detection along with its moving direction in indoor environment.	Microsoft Kinect for image acquisition.	Not Mention	Educational Applications	Proposed novel approach of Object Moving Direction Identification using Depth Image	Distance Along Line Profile graph, Object Moving Direction Identification using Depth Image for Blind	Experimental result shows that the proposed method can successfully detect moving object along with its direction with 92% accuracy and still objects detection accuracy rate is 87%. The overall accuracy of the proposed method is 90%
Visual Attribute Classification Using Feature Selection and Convolutional Neural	To propose a visual attribute classification system based on feature selection	Berkeley Attributes of People Dataset	Not Mention	Educational applications	The Berkeley Attributes of People dataset contains 8035 images with at least a	Convolutional Neural Network, CNN applied for feature	The precision rates of the system are higher than the baseline

Network					full body of a person included	extraction	approach
Analysis of Moving Object Detection and Tracking in Video Surveillance System	To propose a review on unusual event detection in video surveillance system	Use Of Multiple Camera System	Not Mention	Educational applications	-	-	Studied various phases of moving object detection and Object tracking in video surveillance system.

This feature can be utilized while taking the decisions related to size of an object in motion. To achieve better visual quality in terms of HVS based visual quality metrics, the data is embedded into moving object. Data embedding into a moving object based on size can improve the visual quality. The size of an object in motion plays an important role in the evaluation process of visual quality i.e., the distortions resulted due to data embedding in a relatively smaller object in motion cannot be easily perceived by human observer when compared to the distortions in a larger object in motion. For instance, a tiny noise particle on a large and smoother object can be easily perceived by the human observer but the same noise particle is present on a relatively smaller and non-smoother object it is difficult perceive the noise by the human observer. The videos with the movement of the objects can be classified into three categories. The first is, the objects are moving but the observer or the camera is not moving. The second category is the objects are not moving but the camera or the observer is moving. Finally, the case of both the camera and objects are moving. In [4], the first category is considered, that is the camera is fixed and the objects are moving. Objects in motion are detected in videos and classified them based on size. Model uses Optical flow for estimating the motion moments for finding the size of each object.

For estimating the motion between the two frames Frame i and Frame i+1 the proposed model uses the Gunnar Farneback algorithm [4]. This algorithm takes two consecutive frames of the video and estimates the flow vector or displacement vector, d of each pixel position. Flow vector or displacement vector is a 2D vector, d = (u, v) where u and v are the displacement in x and y direction respectively. This stage of the proposed model returns a flow matrix, OFi where OFi(x, y) is the flow vector corresponding to the position (x, y).

The optical flow methods have some limitations. So all the flow vectors may not be a part of moving area. It is required to differentiate between them. Moving and non moving parts are differentiated by analyzing the magnitude of the flow vectors. That is, if the magnitude of a flow vector is above a particular threshold, T then that pixel corresponding to the flow vector is a moving pixel.

Input to the size estimation stage is the segmented binary mask image, Object frame. It is needed to differentiate the objects from the segmented mask image. In the proposed solution contours are used for detecting the boundaries. To measure the accuracy of moving object segmentation there are some metrics which are calculated between the segmented image and the ground truth image. Ground truth image is the image where the moving objects are segmented manually. True Positive, TP: It is the

measure of how many moving objects are correctly segmented as moving. True Negative, TN: It is the measure of how many non moving objects are correctly segmented as non moving. False Positive, FP: It is the measure of how many non moving objects are segmented as moving. False Negative, FN: It is the measure of how many moving objects are segmented as non moving. Precision: It is the measure of among the segmented moving objects how many are really moving.

V. STEREO VISION SYSTEM

The detection of an object and obtaining distance to an object is done through stereo vision system (SVS). Stereo vision system is a reliable depth measuring technique. The study of perception of depth was started in 1584 by Leonardo da Vinci using paintings, sketches that gives a clear understanding of shading, texture and view point projection. SVS can be implemented in different ways by using a single camera. The difficulty with the single camera is the extraction of 3D information is burdensome because of uncertainties in the camera movement. Shidu Dong proposed a system with a single camera along 3-axis accelerometer in which he has obtained rotation and translation vectors in terms of the rotation angles, by which the 3D coordinate of the points on the object is obtained.

For UAVs the important factors effecting its hovering time is payload and the power. The embedded stereo vision sensor is developed with a less weight and minimal power consumption. In [5] an algorithm is proposed where the UAV is equipped with a stereo camera follows a predefined flight plan. The plan is loaded into the autopilot board to detect an object and obtaining 3D view of the scene.

There are many models available in copters like from Bicopter to Octa-copter and copter with a 18 rotors is also exists. When the number of motors increases size and cost of the copter will increase. By considering these factors Hexacopter is selected for our experiment because of its stability and more lifting power compared to Quadcopter. For mission planning and to get the status of the mission during the flight a ground control station-mission planner is used.

It is a full featured ground control station application for autopilot projects and autonomous navigation of copters. The application allows us to control the copter in different modes like alt-hold, hover, loiter and so on. We have planned a mission with different way-points and hover the Hexacopter for 40 seconds.

Zhengyou Zhang has proposed a flexible new technique for camera calibration. Where the camera observes the planar object (i.e. checker board) from different view angles and the camera or planar object motion need not to be known. The camera calibration is required for determining extrinsic and intrinsic parameters of the camera. These parameters useful in measuring focal length, lens distortion, pose estimation, location of the camera and so on. As the camera is moving instead of the object, inverse transformations are required.Ah3 in the expression (5) is not useful because in an image there is no 3rd co-ordinate. The main elements are Ah1, Ah2 and Ah4 because these are required while converting back to Cartesian. So it is not required to find all the 16 elements, to obtain the camera model. To determine camera matrix some known 3D points are selected and their relative image points (x, y) have obtained. For solving camera matrix elements least square fit method is used.

The recovery of shape from stereo is more intuitive. The image captured by left camera is shifted compared to the image captured to the right camera. The shift is nothing but the disparity, which gives an idea about the depth to an object. The proposed system is developed using two USB webcams and these webcams are mounted on the Hexacopter in such a way that their baseline can be adjustable. Raspberry pi B+ board is used for the purpose of video capture. This chip operates at 700MHz with ARM1176JZF-S architecture. The autopilot pilot board Arduino ATMEGA2560 can be easily interfaced with Raspberry pi board. MATLAB Raspberry pi support packages are installed on the chip. The MATLAB stereo camera application is used for the camera calibration.

3-D world coordinates of the point P is constructed from the disparity. The reconstruction of view depends on the disparity and stereo parameters. Point Cloud function of MATLAB is used for visualizing the scene. The detection of an object and its region properties depends on intensity/ color pixel classification, this approach is a fast detection and was able to detect a specific object i.e., object with a color different from the background. The one more approach is to detect an object on shape base analysis which is more independent of hue changes and algorithms used for this approach are more time consuming than intensity based approach.

VI. Real Time Object Detection along Moving Directions

The proposed novel approach of Object Moving Direction Identification using Depth Image for Blind [6] is a simple, affordable and realistic blind navigation support system. 'OMDIDIB' is used as a short form of Object Moving Direction Identification using Depth Image for Blind. OMDIDIB system does not require any complex algorithm or mathematical calculation. This system can differentiate between still staircase and moving escalators along with its moving direction. Any other moving object's moving direction is also detected with respect to the blind person's view point or position. Experimental result shows that the proposed method [6] can successfully detect moving object along with its direction with 92% accuracy and still objects detection accuracy rate is 87%. The overall accuracy of the proposed method is 90%.

RGB camera and depth sensor of Microsoft Kinect is used for image acquisition. Inspired by the initial setup of Kinect sensor was positioned as on a standing human's chest of average heights. Ground to device height is about 1600 mm with a vertical view range of about 5000 mm (starts from 600 mm front distance of sensing device). Distance covering range of the depth image is 800 mm to 6000 mm.

Among many types of indoor objects, still staircase, moving escalator, still or moving human and household accessories, wall, door and obstacle free front scenes are considered for this research work. Data are collected from open access data storage of internet and also captured with Microsoft Kinect version 1. A total of 600 samples depth images of 200 different front scenes with their respective RGB images are analyzed to test the new system. The resolution of captured and collected depth images and RGB images are 640×320 pixels.

A simple morphology processing erosion and dilation is used for noise reduction as processed depth images results better output than non-processed images as stated in [6]. After that, four vertical line profiles are extracted at pre-defined positions covering the image area. Short term 'DAP graph' is used for **D**istance **A**long Line **P**rofile graph [6].

VII. Feature Selection and Convolution Neural Networks for Visual Attribute Classification

In the face verification problem is reformulated as the recognition of the presence or absence of describable aspects of visual appearance. For computer vision tasks, feature expression is a critical factor that affects system performance. The problem of extracting discriminative and representative features has been profoundly researched in the past decades. Due to the powerful representational learning capabilities, CNNs have been widely applied. However, dimensions of CNN features are usually very large with many components irrelevant to the final tasks. Therefore, feature section could be exploited to remove the irrelevant or redundant features, meanwhile, improving classification performance. It includes - A CNN model to learn discriminative feature representation.

A novel feature selection method [7] to remove irrelevant or redundant features and to improve classification performance by reducing over-fitting is presented. The proposed visual attribute classification system mainly includes three modules: feature extraction, feature selection and

classification.

A CNN is applied for feature extraction, and the adopted model is similar to the networks in Fast R-CNN and R*CNN. These networks are built based on the 16-layer architecture, which have demonstrated outstanding performance in image classification and object detection. Since only the features are used before full connection layers, the last layer of our network is the region of interest (RoI) pooling layer.

In feature selection stage, the features from the RoI pooling layer are collected for further refinement. The RoI pooling layer is a kind of adaptive max pooling layer, the size (7x7) of its output feature maps are fixed whatever the size of inputs. Therefore, the size of extracted features for each sample is 7x7x512 (25088 in total). Then, feature selection is performed using proposed method. For each visual attribute classifier, the details can be described as follows.

Data collection: All the available samples in training set are divided into two classes (positive and negative) based on their labels of current attribute.

Data Processing: In order to measure the similarities of features belonging to the same class, all of the features are transformed into binary sequences using a threshold value. Since the activation function of the last convolution layer is ReLU, only values larger than 0 are able to pass to next layer. This means feature positions can be considered to be activated if their values are larger than 0, thus, the threshold value used here is 0. Then, all of the sequences from same class are accumulated together and normalized by dividing by the number of samples. In this manner a series of sequences are determined that indicate the probability of appearance for each feature position. Therefore, two probability sequences are achieved, namely, ppositive and pnegative.

Feature Selection: In this step, feature selection is performed by comparing the magnitude of the probability of each position in ppositive and pnegative. Firstly, a distance matrix can be computed based on jppositive x pnegativej. Secondly, the matrix is sorted according to its magnitude. Finally, given a desired dimension n, the original 25088 feature can be reduced to n by simply select the positions that contain top n largest values in matrix.

In classification stage, linear SVMs are introduced. Classification of SVMs are performed by constructing a hyperplane or set of hyper-planes in a high-dimensional space. With the selected features extracted from the previous stage, linear SVMs are trained to discriminate between presence or absence for each attribute.

VIII. VIDEO SURVEILLANCE SYSTEMS ANALYSIS

The Video Surveillance is used to monitor people, movement of vehicles & equipment and event of interest remotely. It comprises of several components, from video capturing to processing then analyzing to presentation with the following sequenced stages; video capture module, video stream selection, video processing and measurement and finally visualization [8]. Video processing module is related to digital image processing techniques such as object identification and tracking methods. This module processes video frames automatically and helps in detecting objects (peoples, equipments, vehicles) and event of interest for security purposes. In real time, video surveillance systems detect situations in video flow that represent a security threat and trigger an alarm accordingly.

Existing video surveillance systems takes care of capturing, storage and transmission of video to remote places but is devoid of efficient threat detection and analysis leaving these functions exclusively to human operators for manual analysis. Therefore, there is an urgent need of a surveillance system which is fast, efficient and accurate. Several methods have been proposed for object identification and tracking in video data mining literature [8]. But nearly all of these process an image or video sequentially either in spatial or frequency domain or both.

Object detection is the most primitive operation used in video processing application. Once object is detected it is then recognized or identified based on its representation. Because of different variety of objects having different colors, texture and shapes, there are numerous object detection methods researched and discussed in literature during past years. Bouwmans categorized different types of methods into traditional and recent approaches by listing nearly 15 years comprehensive research in object detection along with resources, data sets, implementation codes etc. It also lists the requirements for ideal object detection in different setup of video surveillance as preconditions. These are i) fixed cameras, ii) constant illumination and iii) static background which are never possible to be attained due to different peculiarities of indoor and outdoor scenes. Many problems identified by researchers in video surveillance are poor quality of video having lot of noise, camera jitter, slow and sudden change in illumination of scene, shadows in scene, clutter, camouflage, ghost and occlusion problem in scene. Various features considered for object detection algorithms vary in scales and types. Some features are based on pixel intensity in gray scale images, color, edge, texture. Most of the processing in object detection methods is done on pixel level. But many papers have also reported it on block level and region level.

Object detection in video is analogous in nature during finding of moving regions in the frames. In any frame every pixel is classified either as a foreground region depicting objects in motion or as a background pixel which is immovable. Statistical methods aim to understand the dynamicity of each pixel in the scene. In Gaussian filter method, intensity history of a pixel is assumed to vary with a Gaussian probability distribution function (PDF) with mean and deviation. Single Gaussian method can handle effectively background modeling from the scene where illumination is constant or gradual changing as in the case of transition of day into night and vice versa but in reality illumination may change abruptly (in case of switching a light on off in a room) which require a multimodal PDF, this necessitates background modeling with multiple Gaussian PDFs. Stauffer and Grimson proposed to model a pixel history with mixture of Gaussian (MOG) distribution. The MOG has better adaptability to complex requirement of video surveillance, but due to its complex computation it is slow and may not able to meet real-time needs.

In another method, for background modeling, every pixel creates a codebook (CB) containing quantized values named as codewords with their brightness bounds, frequency and access information. A pixel is classified as foreground or background based on color distance and brightness bound. If present value is within a threshold distance with a code word and within brightness bound then it is termed as background pixel. This information is used to eliminate the redundant code words to obtain the refined initial CB that can best represent the actual background. Codeword that occurs less than half of sampled frames are eliminated from codebooks for background modeling.

Hu et al presented an overview of the developments in the field of video surveillance in dynamic scenes involving surveillance of people or vehicles. This survey identified five stages that are common for most surveillance systems. The first stage is described by pointing out the problems regarding multi-camera surveillance. The second stage was motion detection involving environment modeling, motion segmentation such as background subtraction, temporal difference and optical flow and shape based or motion based object classification.

Another survey on contemporary remote surveillance systems for public focused on the evolution of video surveillance applicable to public safety. The paper identified several future research directions such as real-time distributed architecture, intelligent cooperation between agents, addressing occlusion (objects become occluded by buildings, trees or other objects), the detection of ghosts, multi-sensor surveillance system etc.

In general, the processing framework of an automated video surveillance system includes the following stages: Video Capture Module, Video Selection Module, Video Processing Module, and

Human Machine Interface Module.

Video Capture and selection module is responsible for streaming the desired frames of video for processing into processing module. Video selection module can convert multiple streams from multiple cameras into single stream by employing a fuzzy-based selection methodology. Thin line shows control structure while thick arrow signifies data streaming. Except Video capture module, all other three modules are working under the framework of Logical Video Processing system. The algorithms employed in this system will be developed using Cellular Logic Array framework that uses pattern-directed Search and Replace (SAR) techniques and by its working it is inherently parallel, and so, it guarantees speed and precision.

Video data mining is a process which automatically extract content and discover patterns of structure of video, features of moving objects, spatial or temporal correlations of those features, objects activities, video events, etc. from vast amounts of video data. Video data mining can be classified in pattern detection, video clustering and classification and video association mining. In case of images and frame of videos, the cellular logic array of elements are image pixels which are processed by multiple digital image processing techniques such as motion/object detection, object classification, object tracking. The problem of object detection is mapped to cellular logic array by representing images or video frames as cellular automata and rules to modify these representations will be taken from various basic algorithms such as thinning, edge detection, registration, image erosion and dilation in order to process video streams.

IX. CONCLUSION

The comparison of various methodologies namely deep convolution networks, four-frames differencing, tracking size estimation, SVS for UAVs, moving direction identification using depth image, visual attribute classification using feature selection based convolution neural network, and unusual event detection for moving object detection is shown in table. The findings, used datasets, methodologies applied, experimentations carried out and future aspects in the domain of moving object detection and its applications are presented in this paper. The content of paper can be used as the guidelines for research in moving object detection from video sequences.

REFERENCES

[1] Wen Xu, Jing He, HaoLan Zhang, Bo Mao, and Jie Cao, “Real-Time Target Detection and Recognition with Deep Convolutional Networks for Intelligent Visual Surveillance,” IEEE/ACM 9th International Conference on Utility and Cloud Computing, 2016.

[2] Y. Jia. Caffe: An open source convolutional architecture for fast feature embedding, http://caffe.berkeleyvision.org/, 2018.

[3] Ahmed Abdelli and Ho-Jin Choi, “A Four-Frames Differencing Technique for Moving Objects Detection in Wide Area Surveillance,” IEEE International Conference on Big Data and Smart Computing (BigComp), 2017.

[4] SagarGujjunoori, S. SaiSatyanarayana Reddy, Gouthaman K V, “Tracking and Size Estimation of Objects in Motion”, International Conference on Machine Vision and Information Technology, 2017.

[5] Sk. Noor Kizar, G. S. R. Satyanarayana, “Object Detection and Location Estimation using SVS for UAVs,” International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), 2016

[6] Aniqua Nusrat Zereen and Sonia Corraya, "Detecting Real Time Object Along with the Moving Direction for Visually Impaired People," 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), 2016.

[7] Rongqiang Qian, Yong Yue, Frans Coeneny and Bailing Zhang, "Visual Attribute Classification Using Feature Selection and Convolutional Neural Network," IEEE 13th International Conference on Signal Processing (ICSP), 2016.

[8] Singh, Ajay Prasad, Kingshuk Srivastava, Suman Bhattacharya, "A Cellular Logic Array Based Data Mining Framework for Object Detection in Video Surveillance System," IEEE 2nd International Conference on Next Generation Computing Technologies (NGCT), 2017.