CENTROID BASED MULTIPLE OBJECT TRACKING WITH SHADOW REMOVAL AND OCCLUSION HANDLING

Himani S. Parekh and Nancy A. Lalluwadia
Department of Computer Engineering,
C.G.P.I.T., Maliba Campus, Bardoli, India

ABSTRACT

This paper gives complete guidelines for authors submitting papers for the Journal. Object motion detection technology deals with how to extract moving object from videos and eliminate the background and noise as much as possible. Extracted object may be classified in various categories such as humans, vehicles, birds, floating clouds, swaying tree and other moving objects and object tracking is to monitor and objects spatial and temporal changes during a video sequence, including its presence, position, size, shape, etc. In this paper Median filter is used for the background modeling and for integrating new information into current background, Kalman filter is used. The complete and correct object is detected by removing the shadow using the Normalized Cross Covariance method. Object is classified using RMI motion image. Centroid based tracking is essential for modeling the path and occlusion is handled efficiently. The results are taken for various datasets of different scenarios.

KEYWORDS: *Object Tracking, Shadow Removal, Normalized Cross Covariance, RMI motion image, Occlusion*

I. INTRODUCTION

Object tracking means to track an object over a sequence of images. Video tracking is to associate target objects in consecutive video frames. It is the method of segmenting a region on interest from a video scene and keeping track of its motion, positioning, occlusion etc. in order to extract worthwhile evidence. Videos are actually sequences of images, each of which called a frame, displayed in fast enough frequency so that human eyes can percept the continuity of its content. It is obvious that all image processing techniques can be applied to individual frames. Besides, the contents of two consecutive frames are usually closely related [1].

The identification of regions of interest is typically the first step in many computer vision applications including event detection, visual surveillance, and robotics. A general object detection algorithm may be desirable, but it is extremely difficult to properly handle unknown objects or objects with significant variations in color, shape and texture. Therefore, many practical computer vision Systems assume a fixed camera environment, which makes the object detection process much more straightforward. An image, usually from a video sequence, is divided into two complimentary sets of pixels. The first set contains the pixels which correspond to foreground objects while the second and complimentary set contains the background pixels. This output or result is often represented as a binary image or as a mask. It is difficult to specify an absolute standard with respect to what should be identified as foreground and what should be marked as background because this definition is somewhat application specific.

Object can be tracked using various methods such as point based tracking, kernel based tracking and Silhouette Based Tracking. In Point Tracking Approach, moving objects are represented by their feature points during tracking. Point tracking [10] is a complex problem particularly in the incidence of occlusions, false detections of object. Recognition can be done relatively simple, by thresholding, at of identification of these points. The difficulty of this approach is that it can only track very small objects. Kernel Based Tracking algorithms diverge in terms of the presence representation used, the number of objects tracked, and the method used for approximation the object motion. In real-time, illustration of object using geometric shape is common. But one of the restrictions is that parts of the objects may be left outside of the defined shape while portions of the background may exist inside. This can be detected in rigid and non-rigid objects .They are large tracking techniques based on representation of object, object features ,appearance and shape of the object. In Silhouette Based Tracking Approach, some object will have complex shape such as hand, fingers, shoulders that cannot be well defined by simple geometric shapes. Silhouette based methods [9] afford an accurate shape description for the objects. The aim of a silhouette-based object tracking is to find the object region in every frame by means of an object model generated by the previous frames. This approach is capable of handling of large variety of object shapes easily, handling Occlusion and dealing with object split and merge.

In this paper, a methodology is introduced which perform background subtraction that combines statistical assumptions of moving objects using the previous frames in the dynamically varying noisy situation. Reference image is frequently updated in order to achieve reliability of the motion detection. The new incoming information is integrated into the current background image using a Kalman filter. Shadow is suppressed to reduce the false alarm as shadow can be detected as another object. Object is tracked and path of an object is modeled by centroid based method. In this paper, the next section contains the implementation of different phases, section 3 includes the experiment results, section 4 contains performance analysis followed by conclusion.

II. RELATED WORK

In paper [2], the proposed algorithm has been used to perform background subtraction from moving vehicles in traffic video sequences that combines statistical assumptions of moving objects using the previous frames in the dynamically varying noisy situation. Background image is frequently updated using a binary moving objects hypothesis mask. Then, the new incoming information is integrated into the current background image using a Kalman filter. In paper[1] Abhishek Kumar Chauhan, Prashant Krishan, have proposed new tracking method that combines gaussian mixture model (GMM) and optical flow approach for object tracking. Morphological and median filters are used to remove noise. In paper[4] S. Saravanakumar, A.Vadivel, C.G Saneem Ahmed have developed Multiple human object tracking approach based on motion Estimation and detection, background subtraction, shadow Removal and occlusion detection In the approach, Morphological operations are used for identifying and removed the shadow. The occlusion is one of the most common events in Object tracking and object centroid of each object is used for detecting the occlusion and identifying each object separately.

III. IMPLEMENTATION

For tracking an object, first step is to detect the object. The detected object can be tracked easily. For implementing object detection following steps have to be followed which includes:

2.1. Object Detection

2.1.1. Read Frames from Video file

To process the video, the first step is to convert the video into the number of frames. For generating frames from video sequence, multimedia reader object has to be created which can read video data from a multimedia file. It supports avi, wmv, mpg, asf etc formats.

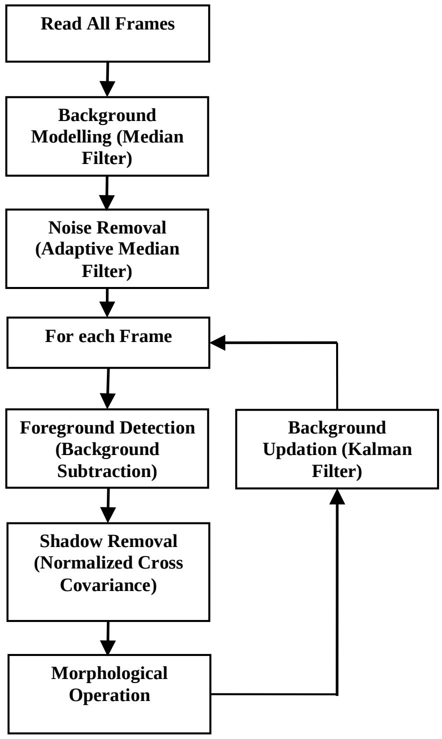


Figure 1 Step for Object Detection

2.1.2. Background modeling using Median Filter

The median filter is a sliding-window spatial filter. It replaces the value of the center pixel with the median of the intensity values in the neighborhood of that pixel. Background modeling or the initial reference frame for background can be obtained by applying median filter. The background estimate is defined to be the median at each pixel location of all the frames in the buffer. The assumption is that the pixel stays in the background for more than half of the frames in the buffer.

2.1.3. Pre-processing using Adaptive Median Filter

After determining the reference model, Adaptive Median Filter is applied to model for removing impulse noise. The Adaptive Median Filter performs spatial processing to determine which pixels in an image have been affected by impulse noise. The Adaptive Median Filter classifies pixels as noise by comparing each pixel in the image to its surrounding neighbor pixels. A pixel that is different from a majority of its neighbors, as well as being not structurally aligned with those pixels to which it is similar, is labeled as impulse noise. These noise pixels are then replaced by the median pixel value of the pixels in the neighborhood that have passed the noise labeling test.

2.1.4. Background updating using Kalman Filter

For reliable object detection, it is necessary to update the reference image frequently. The basic idea in background adaptation is to integrate the new incoming information into the current background image using a Kalman filter:

$$\mathrm{B(t+1) = B(t) + [a1 * 1 - Mt + a2 * Mt] * Dt}$$

Where, Bt represents the background model at time t and Dt is the difference between the present frame and the background model, and Mt is the binary moving objects hypothesis mask. The gain a1 and a2 are based on an estimate of the rate of change of the background. The larger it is, the faster new changes in the scene are updated to the background frame. Here, the values of a1 and a2 are taken 0.1 and 0.01 respectively.

2.1.5. Foreground Detection

Firstly, the original frame is subtracted from the reference model so that the background is subtracted.

$$Ft\,(i,j) = It\,(i,j) - Bt\,(i,j)$$

Foreground can be detected by applying segmentation. Thresholding is the simplest segmentation method. The pixels are partitioned depending on their intensity values.

$$Ft\,(i,j) = 1, \text{ if } |Ft\,(i,j)| > Tt$$
$$0, \text{ otherwise}$$

2.1.6. Shadow Removal using Normalized Cross Covariance

In foreground detection step, all the moving points of both objects and shadows are detected at the same time. So shadow is also detected as another object that leads to false detection. Moreover, the object shape is falsified that affects both the classification and the assessment of moving object position.

To indentify the shadows and avoid misclassification, Normalized Cross Covariance is used. For any pixel (x, y) in a shadowed region, the Normalized Cross Covariance (NCC) in a neighboring region B(x, y) is found and the shadow can be detected using equation given below,

$$NCC(x, y) \geq Lncc$$

Where, Lncc is fixed threshold. If Lncc is low, several foreground pixels corresponding to moving objects may be misclassified as shadows. On the other hand, selecting a larger value for Lncc results in less false positives, but pixels related to actual shadows may not be detected [4].

NCC for a pixel at position (i, j) is given by

$$\mathrm{NCC(i,j)} = \frac{\mathrm{ER(i,j)}}{\mathrm{E_{B(i,j)}E_{Tij}}}$$

Where,

$$\mathrm{ER(i,j)} = \sum_{n=-N}^{N}\sum_{m=-N}^{N} \mathrm{B(i+n, j+m)\,Tij(n,m)}$$

$$\mathrm{E_B(i,j)} = \sqrt{\sum_{n=-N}^{N}\sum_{m=-N}^{N} \mathrm{B(i+n, j+m)}^2}$$

$$\mathrm{E_{Tij}} = \sqrt{\sum_{n=-N}^{N}\sum_{m=-N}^{N} \mathrm{Tij(n,m)}^2}$$

B(i, j) is the background image and $Tij(n,m)$ is the template on current image.

2.1.7. Extraction of connected components

Binary image after segmentation may yields partial object. The reason behind this problem is that if an object contains any dark portion, then that portion is considered as background and removed during thresholding. So for complete object series of morphological operations such as erosion, dilation has to be performed.

2.2. Object Classification

Object can be classified according to its shape, color, motion, texture etc. Motion based classification calculates the motion of an object from video sequences.

Motion-based recognition approach using a specific feature vector called Recurrent Motion Image (RMI) [15] is to classify moving objects into predefined categories, namely single person, group of persons and vehicle. Moving objects detected from image sequences are classified based on their periodic motion patterns captured with the RMI. A specific feature vector called RMI can be used to estimate the repetitive motion behavior of moving objects yielding different RMI signature for different object's motion behavior. RMI will have high values at pixels where motion occurred repetitively and low values at pixels where little or no motion occurred.

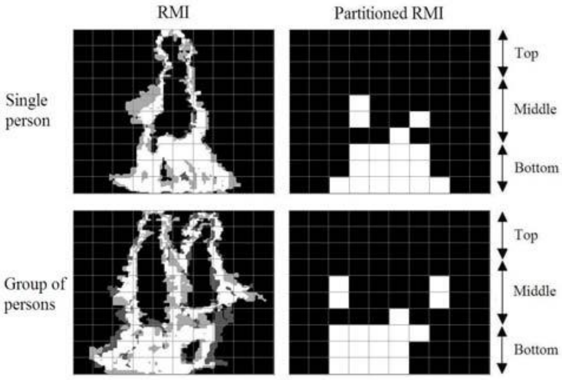


Figure 2 RMI of a single person and a group of persons [16]

RMI is computed with the following equations [16] to determine the areas of moving object's silhouette undergoing repetitive changes:

$$DS_a(x, y, t) = S_a(x, y, t-1) \oplus S_a(x, y, t)$$

$$RMI_a = \sum_{k=0}^{T} DS_a(x, y, t-k)$$

Sa is a binary silhouette for object a at frame t, and DSa is a binary image indicating areas of motion for object a between frame t and t-1. RMIa is the RMI for object a calculated over T frames. Subsequently, the RMI is partitioned into N equal-sized blocks in order to compute the average recurrence for each block. Blocks with average recurrence value greater than a threshold τRMI are set to 1 (white) and vice versa. Hence, white blocks indicate image regions with high motion recurrence whereas black blocks indicate the areas with insignificant or no motion recurrence.

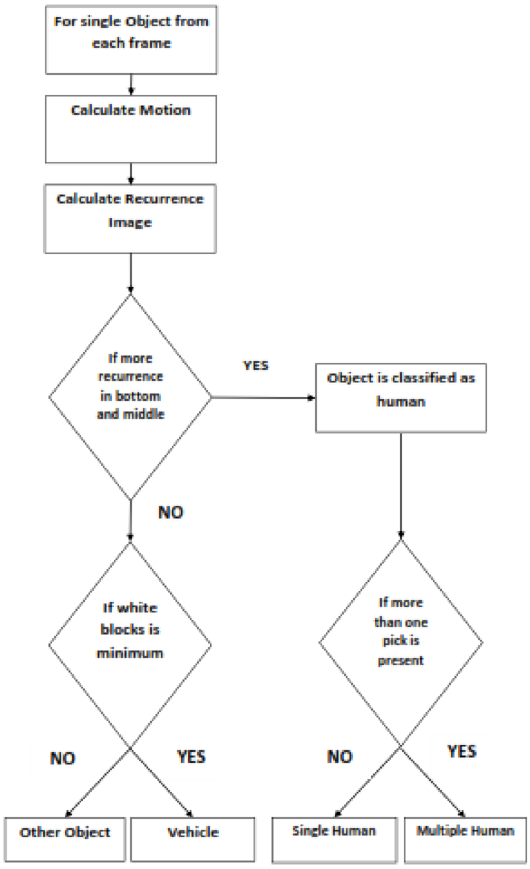


Figure 3 Steps for Object Classification

Following steps [16] are required for classify an object from RMI image:

1) If there are white blocks at the middle and bottom sections of a partitioned RMI, the black area within the respective RMI is observed. If the black area has a vertical major axis, the corresponding object is classified as human, which will then be further categorized as a single person or group of persons.
2) There are two cues to differentiate between single person and group of persons.
   i. Multiple peak points in a silhouette indicate more than one headcount, therefore representing a group of persons.
   ii. Normalized area of recurrence response at the top section of RMI being greater than that of a single person, due to presence of multiple heads.

   If either one of the aforementioned criteria is satisfied, the object is classified as a group of persons, whereas if none of them are satisfied, the object is classified as a single person.
3) If there are no white blocks, which means no recurrent motion, the object is classified as a vehicle.

2.3. Object Tracking

The goal of tracking is to establish correspondences between objects across frames. Blobs obtained from the detection section are tracked using region correspondence. Parameters such as centroid, bounding box, size, velocity and change in size of each blob are extracted. Correspondences between regions in previous frame and current frame are established using the minimum cost criteria to update the status of each object over the frames.

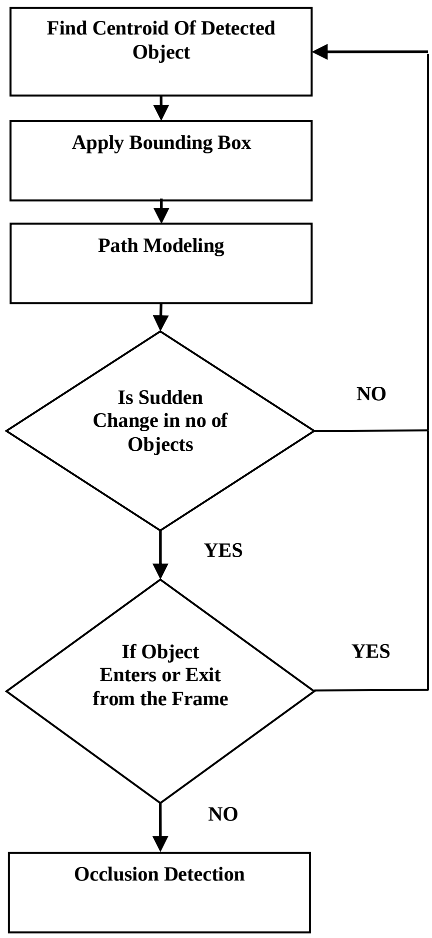


Figure 4 Steps for Object Tracking

1) Apply Bounding Box to detected Object. Bounding Box is the smallest rectangle containing the region, a 1-by-Q *2 vector, where Q is the number of image dimensions.
2) Calculate the centroid of an object bounded in bounding box. Centroid is the 1-by-Q vector that specifies the center of mass of the region.
3) Store all previous centroid of an object to show the path from whole video.

Occlusion: while two moving objects coming closer to each other, the background subtracted frame shows it as a single object. This situation is called as occlusion and will create problem while tracking two objects. For detecting the occlusion, the number of objects in all frames is obtained.

[1]. If sudden decrease in number of object is found, then two possibilities are present.

a. Object exit from the frame.
b. Occlusion of two or more objects

[2]. If number of object is sudden increases, then possibilities are

c. New object enters in the frame.
d. Occluded object is separated from other object.

IV. EXPERIMENTAL RESULTS

The implemented method has been simulated using MATLAB®. All simulations have been carried out on Intel core 2 duo CPU, T6400, 2 GHz with 4 GB RAM under MATLAB® R2010b environment. To study and investigate the effectiveness of implemented algorithm different data sets are used. In the present work a method for tracking has been developed for a static background. The simulation results demonstrate the effectiveness of the method and gives reasonably accurate result.
Various video datasets [19][20][21][22] are used for input videos which include a number of video clips that are recorded acting out the different scenarios of interest.

Table 1: Video Dataset

Video	Sample Frame	Source	Frame Width	Frame Height	Time Duration
Video 1		Provided by Author: Saneem Ahmed [4]	720	576	0.35
Video 2		Caviar Dataset [19]	384	288	0.19
Video 3		Pets Dataset 2000 [21]	352	288	0.20
Video 4		Candata Dataset [22]	768	576	0.25
Video 5		Nthu [20]	352	240	0.5

Fig 5: Initial Background frame

Fig 5 shows the initial background frame for video 1 taken from [4]. Fig 6 shows the detection of complete object. The complete object is detected by subtracting the reference image from the original frame which gives the binary result followed by shadow removal technique and series of morphological operations for video 5 and 1.

Original Frame	Foreground Detection	Shadow Removal	Morphological Operation
			
			

Fig 6: Implemented Results for Object Detection for video 5 and video 1 taken from [20] and [4]

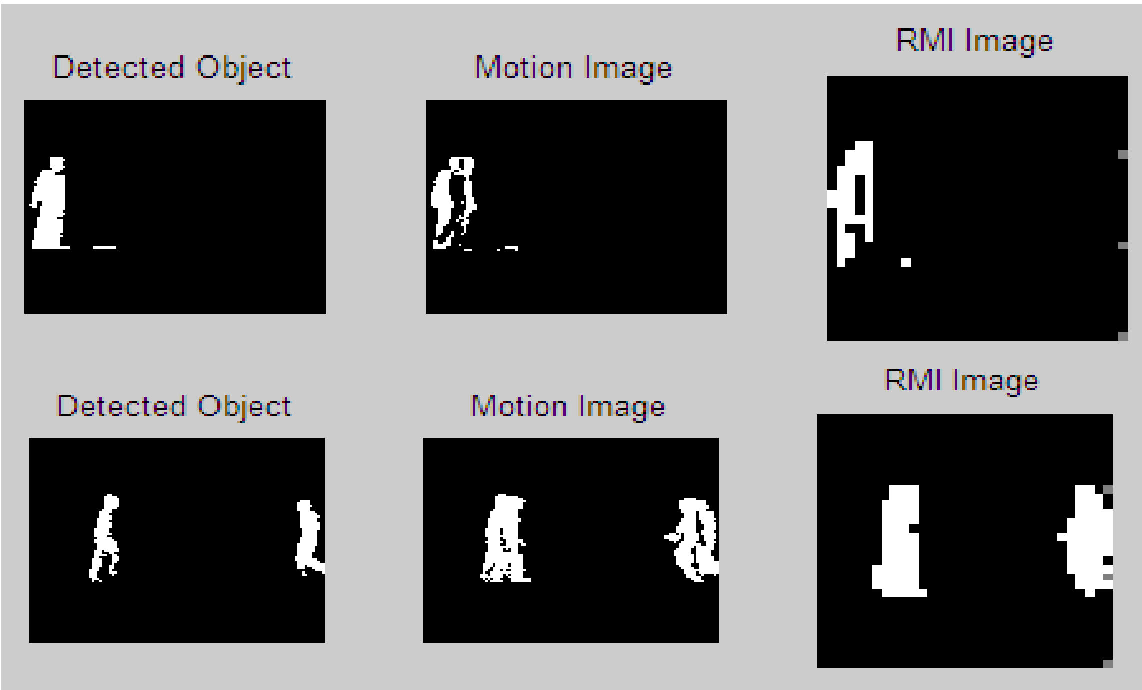


Figure 7: Implementation Results of Object Classification of Video 1[4]

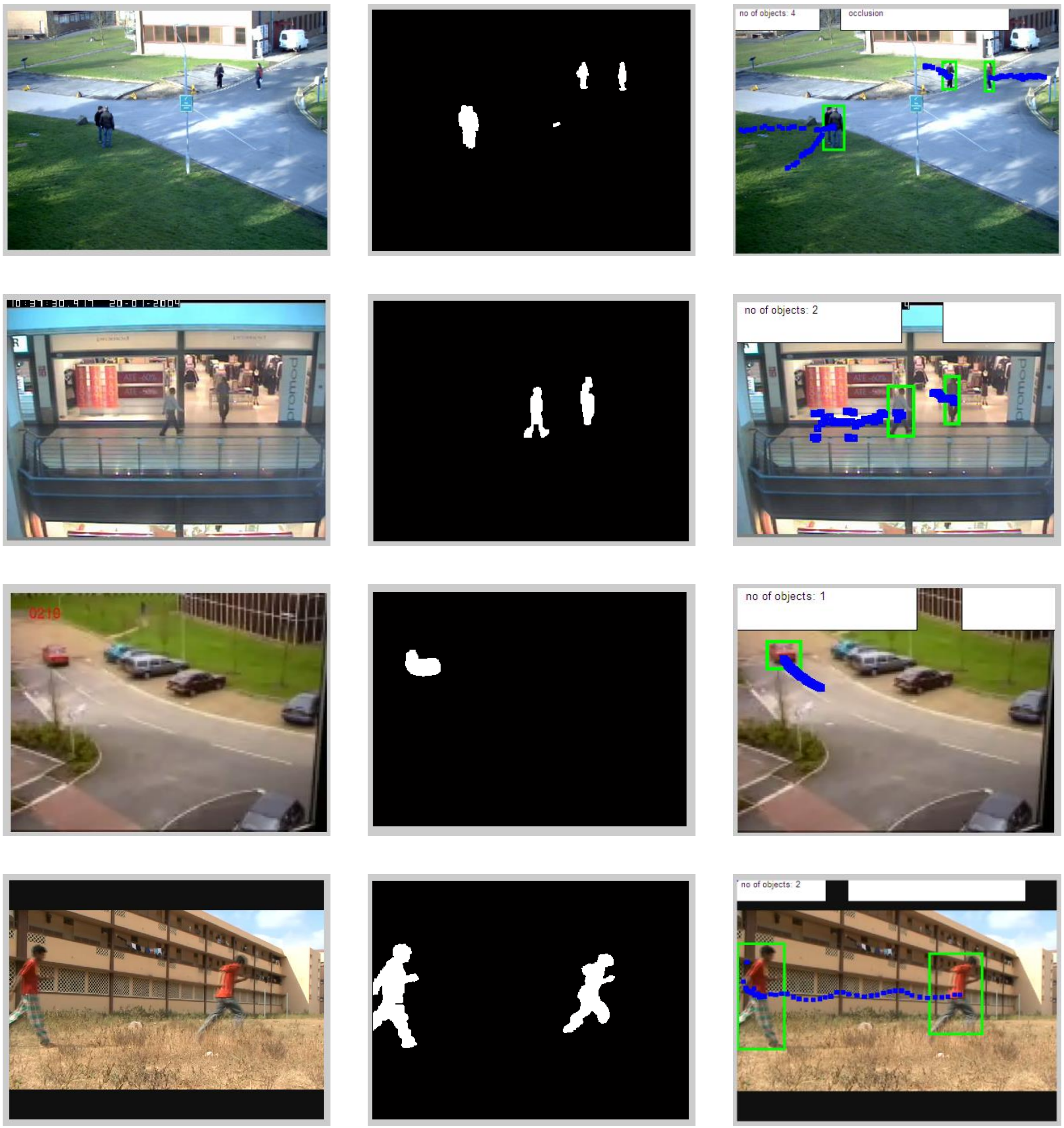


Fig 8: Implementation Results of object detection and object tracking of Video 4 [22], Video 2 [19], Video 3 [21], Video 1[4]

Fig 7 is the results for classification of single object and multiple objects for video 1 [4]. Figure includes motion image and RMI image.

Fig 8 shows the results for four different videos. Figure includes the original frame, foreground detection results and track result. The last column indicates three main conclusions: path of an object is captured properly, multiple objects are detected from single frame and occlusion is handled properly. For locating the path 5×5 red pixels are used. Path is shown from starting frame to the current frame.

V. Performance Analysis

Performance evaluations are carried out using following parameters.

4.1. Evaluation Parameters

In order to compare the output of the algorithm with the ground truth segmentation, a region matching procedure [17] is adopted which allows to establish a correspondence between the detected objects and the ground truth. Several cases are considered:

1) Correct Detection (CD): The detected region matches one and only one region.
2) False Alarm (FA): The detected region has no correspondence.
3) Detection Failure (DF): The ground truth region has no correspondence.

Region Matching

Object matching is performed by computing a binary correspondence matrix Ct which defines the correspondence between the active regions in a pair of images. Let us assume that we have N ground truth regions R1i and M detected regions Rj. Under these conditions Ct is an N ×M matrix, defined in [17] as follows:

$$(i,j) = \begin{cases} 0 \quad \text{if} \quad \frac{(R1_i \cap R_j)}{(R1_i \cup R_j)} & < T \\ 1 \quad \text{if} \quad \frac{(R1_i \cap R_j)}{(R1_i \cup R_j)} & > T \end{cases}$$

where T is the threshold which accounts for the overlap requirement. It is also useful to add the number of ones in each line or column, defining two auxiliary vectors as described in [17] are:

$$L(i) = \sum_{j=1}^{M} C(i,j) \qquad (12)$$

$$C(j) = \sum_{i=1}^{N} C(i,j) \qquad (13)$$

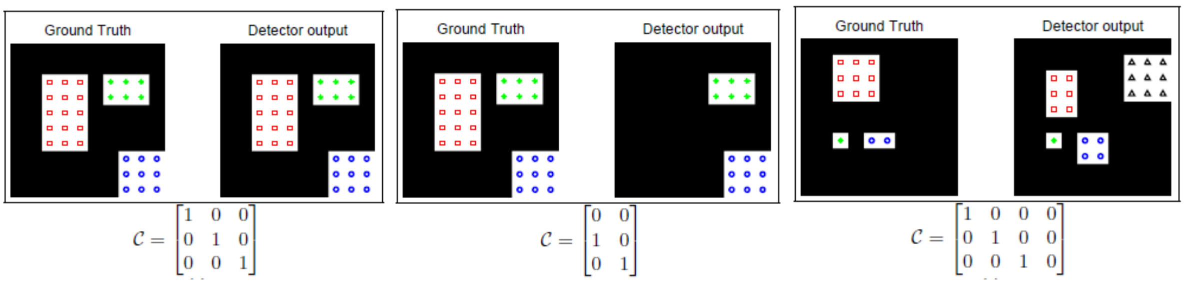
Detected regions Rj are classified according to the following rules:

Correct Detection: $\exists i : L(i) = C(j) = 1 \wedge C(i,j) = 1$

False Alarm: $\exists i : C(j) = 0$

Detection Failure: $\exists j : L(i) = 0$

FA and DF occur whenever empty columns or lines in matrix C are observed.



(a) **(b)** **(c)**

Fig 9: (a) Correct Detection, (b) Detection Failure, (c) False Alarm

4.2. Quantitative Analysis

Table 2: Time Estimation for Tracking Object (In frames/sec)

	Video 4 [22]	**Video 2[19]**	**Video 1[4]**	**Video 3[21]**	**Video 5[20]**
Median Filter	4.1276	1.5182	3.3856	1.2363	1.0756
Gaussian Filter	10.7904	3.1766	8.5265	2.9259	2.8895

Table 3: Evaluation of Object Detection

	Video 5 [20]		**Video 2 [19]**		**Video 3 [21]**		**Video 1 [4]**	
	With shadow	**Without shadow**	**With shadow**	**Without shadow**	**With shadow**	**Without shadow**	**With shadow**	**Without shadow**
Correct Detection	24	26	12	18	10	10	8	8
False Alarm	11	6	9	4	8	0	5	2
Detection Failure	5	3	4	2	0	0	1	1

Table 2 shows the comparative study for median filter and Gaussian filter for object tracking task. The table concludes that median filter takes less time as compare to Gaussian filter. Table 3 shows the results for correct detection, detection failure, false alarm for four different videos. After removing shadows, the detected complete object give better results than the object with shadows presented.

VI. CONCLUSION

In this paper various methods object tracking namely Point Based, Kernel Based and Silhouette Based have been reviewed. Because of fast computation, median filter is used for background modeling. Kalman filter gives better result for updating background compare to other methods as it deals with the adaptive background. Implementation results show that shadow is suppressed efficiently by normalized cross covariance. RMI motion image is able to identify single human, multiple human, and vehicle properly. Centroid based method can track the multiple objects and model the path correctly. The major problem of occlusion is handled efficiently by this tracking method

REFERENCES

[1]. Abhishek Kumar Chauhan, Prashant Krishan, “Moving Object Tracking Using Gaussian Mixture Model And Optical Flow”, International Journal of Advanced Research in Computer Science and Software Engineering**,** April 2013

[2]. M.Sankari, C. Meena, “Estimation of Dynamic Background and Object Detection in Noisy Visual Surveillance”, International Journal of Advanced Computer Science and Applications, 2011, 77-83

[3]. Cheng-Laing Lai; Kai-Wei Lin, "Automatic path modeling by image processing techniques," Machine Learning and Cybernetics (ICMLC), 2010 International Conference on, vol.5, no., pp.2589, 2594, 11-14 July 2010

[4]. Saravanakumar S., Vadivel A. and Saneem Ahmed, C.G., "Multiple human object tracking using background subtraction and shadow removal techniques," Signal and Image Processing (ICSIP), 2010 International Conference on , vol., no., pp.79,84, 15-17 Dec. 2010

[5]. Ruolin Zhang, Jian Ding, “Object Tracking and Detecting Based on Adaptive Background Subtraction”, International Workshop on Information and Electronics Engineering, 2012, 1351-1355.

[6]. K.Srinivasan, K.Porkumaran, G.Sainarayanan, “Improved Background Subtraction Techniques For Security In Video Applications”

[7]. Rupali S.Rakibe, Bharati D.Patil, “Background Subtraction Algorithm Based Human Motion Detection”,International Journal of Scientific and Research Publications, May 2013

[8]. Jae-Yeong Lee; Wonpil Yu, "Visual tracking by partition-based histogram backprojection and maximum support criteria," Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on , vol., no., pp.2860,2865, 7-11 Dec. 2011

[9]. Mr. Joshan Athanesious J; Mr. Suresh P,”Implementation and Comparison of Kernel and Silhouette Based Object Tracking”, International Journal of Advanced Research in Computer Engineering & Technology, March 2013, pp 1298-1303.

[10]. J.Joshan Athanesious, P.Suresh, “Systematic Survey on Object Tracking Methods in Video”,International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) October 2012, 242-247.

[11]. Sen-Ching S. Cheung and Chandrika Kamath, “Robust techniques for background subtraction in urban traffic video”

[12]. Greg Welch, Gary Bishop**,”** An introduction to the Kalman Filter”, In University of North Carolina at Chapel Hill, Department of Computer Science. Tech. Rep. 95-041, July-2006.

[13]. Rahul Mishra, Mahesh K. Chouhan, Dr. Dhiiraj Nitnawwre, "Multiple Object Tracking by Kernel Based Centroid Method for Improve Localization", International Journal of Advanced Research in Computer Science and Software Engineering, July-2012, pp 137-140.

[14]. Hitesh A Patel, Darshak G Thakore,"Moving Object Tracking Using Kalman Filter", International Journal of Computer Science and Mobile Computing,April 2013, pg.326 – 332.

[15]. Wong Chuan Ern, Ong Teong Joo, "A New RMI Framework for Outdoor Objects Recognition"

[16]. Chuan Ern Wong, Teong Joo Ong, "An Improved Recurrent Motion Image Framework for Outdoor Objects Recognition", International MultiConference of Engineers and Computer Scientists 2009 Vol I

[17]. Jacinto Nascimento, Jorge Marques, "Performance evaluation of object detection algorithms for video surveillance"

[18]. R. Cucchiara1, C. Grana, M. Piccardi, A. Prati, "Detecting Objects, Shadows and Ghosts in Video Streams by Exploiting Color and Motion Information"

[19]. http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/

[20]. http://www.ee.nthu.edu.tw/cwlin/

[21]. ftp://ftp.pets.rdg.ac.uk/pub/PETS2000/

[22]. http://www.multitel.be/cantata/

AUTHORS

Himani S. Parekh

Assistant Professor,

Computer Engineer Dept.,

C.G.P.I.T., Maliba Campus, Bardoli, India

Nancy A. Lalluwadia

Assistant Professor,

Computer Engineer Dept.,

C.G.P.I.T., Maliba Campus, Bardoli, India