INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

OBJECT TRACKING USING KALMAN FILTER & TCM A PROPOSED SCHEME

Ankita Taneja[1], Navneet Verma[2]

[1]M.Tech Scholar
Geeta Engineering College, Naultha, Panipat,
Haryana (132103), India
ankiii.frnds@gmail.com
[2] Assistant Prof. in CSE Deptt.
Geeta Engineering College, Naultha, Panipat
Haryana (132103), India
navneet.cse@geeta.edu.in

Abstract: *Intelligent machines require basic information such as moving-object detection from videos in order to deduce higher-level semantic information. Object detection is known to be very difficult and challenging task. In this paper, we propose methodology that combines Kalman filter & TCM to find threshold value for background subtraction so as to achieve the best results. The methodology is computationally inexpensive and is resilient to noise, illumination changes, dynamic background and low frame rate. Experimental results show that performance of the proposed approach is higher than those of state-of-the-art approaches.*
Keywords: *Texton Co-occurrence Matrix (TCM), Background Subtraction (BS), Video surveillance, Kalman Filter.*

1. INTRODUCTION

Video surveillance systems have long been in use to monitor the security sensitive areas. Moving object detection is the basic step for further analysis of video. It handles segmentation of moving objects from stationary background objects. Commonly used techniques for object detection are background subtraction, statistical models, temporal differencing and optical flow. Due to dynamic environmental conditions such as illumination changes, shadows and waving tree branches in the wind object segmentation is a difficult and significant problem that needs to be handled well for a robust visual surveillance system. Object classification step categorizes detected objects into predefined classes such as human, vehicle, animal, clutter, etc. It is necessary to distinguish objects from each other in order to track and analyze their actions reliably. Currently, there are two major approaches towards moving object classification, which are shape-based and motion-based methods. Shape-based methods make use of the object's 2D spatial information whereas motion-based methods use temporal tracked features of objects for the classification solution. Background subtraction is a general term for a process which aims to segment moving foreground objects from a relatively stationary background. There is an important distinction between the background modeling and background detection stages, which comprise the whole subtraction process. As illustrated in Figure 1, these two stages are often interrelated and sometimes overlapping. The modeling stage creates and maintains a model of the background scene. The detection process is responsible for segmenting the current image into moving (foreground) and stationary (background) regions based on the current background model. The resulting detection masks may then be fed back into the modeling process in order to avoid corruption of the background model by foreground object.
Our motivation in studying this problem is to create a visual surveillance system with real-time moving object detection, classification and tracking and activity analysis capabilities. The presented system handles all of the above methods except activity recognition which will likely be the future step of our research.

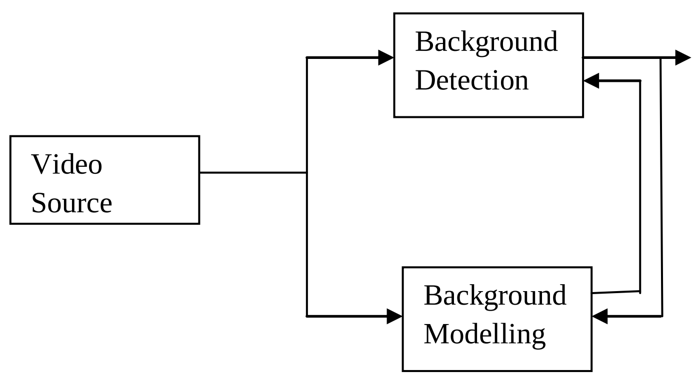


Figure 1: A video source produces images I.
The modeling stage uses previous video frames and detection results to maintain background model B. The detection stage compares the current frame to the current model to produce a detection mask E.

2. IMAGE SEGMENTATION

Segmentation techniques are divided into two basic categories: edge-based and region-based. Edge-based segmentation is primarily used to look for image discontinuities. The technique is generally applied where changes of gray-level intensity occur in the image. The assumption is that changes occur in the data at the boundary between objects of interest. In contrast, region-based segmentation is used to look for similarities between adjacent pixels. That is, pixels that possess similar attributes are grouped into unique regions. The assumption is made that each region represents one object of interest. Using gray-level intensity is the most common means of assigning similarity. Effects of uneven sample illumination,

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

shadowing, partial occlusion, clutter, noise, and subtle object-to-background changes can all contribute to errors in basic segmentation processes. They generally result in false segmentations of the background, partial segmentations of the objects of interest, clumping of objects, or inadequate segmentations. Errors in the segmentation of the data can also result in the calculation of erroneous features. Thresholding is perhaps the most common segmentation technique and is the most basic region-segmentation technique. The technique separates pixels into background and foreground (object of interest) classes based upon their similarities in gray-level intensity. To implement this technique, a threshold (T) value is chosen. Every pixel in the image is then compared to the T value. Each pixel is given a region label of "0" (background) if the pixel value is less than or equal to T or "1" (foreground) if greater than T. This form of region segmentation results in a binary image, in which each region is either white (1) or black (0). Suppose that the gray-level histogram shown in Fig 2(a) corresponds to an image, f(x; y), composed of light objects on a dark background, in such a way that object and background pixels have gray levels grouped into two dominant modes. One obvious way to extract the objects from the background is to select a threshold T that separates these modes. Then any (x; y) for which f(x; y) > T is called an object point; otherwise, the point is called a background point. Fig. 2(b) shows a slightly more general case of this approach, where three dominant modes characterize the image histogram (for example, two types of light objects on a dark background).

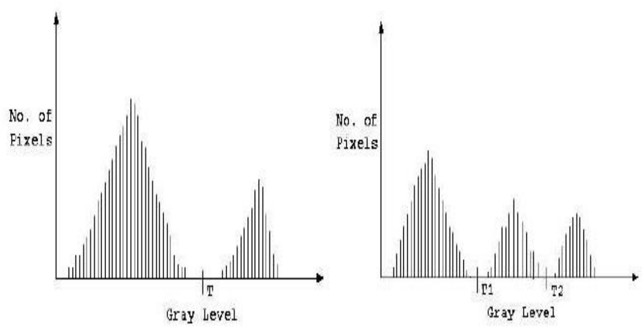


Figure 2: Gray-level histograms that can be partitioned by (a) a single threshold, and (b) multiple thresholds

Here, multilevel thresholding classifies a point (x; y) as belonging to one object class if T1 < f(x; y) T2, to the other object class if f(x; y) > T2, and to the background if f(x; y) T1. Based on the preceding discussion, thresholding may be viewed as an operation that involves tests against a function T of the form T = T[x; y; p(x; y); f(x; y)] where f(x; y) is the gray level of point (x; y) and p(x; y) denotes some local property of this point-for example, the average gray level of a neighborhood centered on (x; y). A threshold image g(x; y) is defined as

$$G(x,y) = \begin{cases} 0 - if - f(x,y), \leq T \\ 1 - if - f(x,y), > T \end{cases}$$

Thus, pixels labeled 1 (or any other convenient gray level) correspond to objects, whereas pixels labeled 0 (or any other gray level not assigned to objects) correspond to the background. When T depends only on f(x; y) (that is, only on gray-level values) the threshold is called global. If T depends on f(x; y) and p(x; y), the threshold is called local.

3. LITERATURE SURVEY

The problem of video object detection under dynamic scene is considered in [1]. The dynamisms of the scene are assumed to be due to illumination variation and swaying of the trees and leaf. Many algorithms have been proposed to cope to this situation. But the major drawback in most of them is misclassified object and background area. Stochastic approaches including MRF based algorithms exist in literature but the practical implementation of such complex models remains still a challenge for the VLSI architecture designers. Thereby real-time object recognition and tracking process largely depends on simple and deterministic approaches and their accuracy. A novel approach for illumination normalization [2] is proposed by exploiting the correlation of discrete cosine transform (DCT) low-frequency coefficients to illumination variations. The input image contrast is stretched using full image histogram equalization. Then the low-frequency DCT coefficients (except first) are re-scaled to lower value to compensate the illumination variations. The first (DC) coefficient is scaled to higher value for further contrast enhancement. The experiments are performed on the Yale B database and the results show that the performance of the proposed approach is better for the images with large illumination variations. The proposed technique is computationally efficient and can easily be implemented for real time face recognition system. The several inherent characteristics of natural outdoor environmental monitoring that pose a challenge to automated background modelling and subtraction [3] is paid. Namely, foreground objects tend to, by necessity, blend into the background, and the background exhibits large variations due to non-stationary objects (moving leaves) and rapid transitions from light to shadow. These conditions present a challenge to the state of the art, which is addressed with an algorithm that exhibits comparable performance also on standard surveillance data sets. A side benefit of this approach is that it has relatively low memory requirements, does not require floating point operations, and for the most part, can run in parallel. A Mat-lab/Simulink based mode l[4] is proposed for monitoring a contact in a video surveillance sequence. For the segmentation process and correct identification of a contact in a surveillance video, the use of the Horn-Schunk optical flow algorithm. The position and the behavior of the correctly detected contact were monitored with the help of the traditional Kalman filter. After that compared the results obtained from the optical flow method with the ones obtained from the Kalman filter, and the correct functionality of the Kalman filter based tracking. A region-based method for background subtraction is proposed here [5]. It relies on color histograms, texture information, successive division of candidate rectangular image regions to model the background and detect motion. The proposed algorithm uses this principle and combines it with Gaussian Mixture background modelling to produce a new method which outperforms the classic Gaussian Mixture background subtraction method. The method has the advantages of filtering noise during image differentiation. A new approach is proposed to improve traditional background subtraction (BGS) techniques by integrating a gradient-based edge detector called a second derivative in gradient direction (SDGD) filter with the BS output [6]. The four fundamental BGS techniques, namely, frame difference (FD), approximate median (AM), running average (RA), and running Gaussian average (RGA), showed imperfect foreground pixels generated specifically at the boundary. The pixel intensity was lesser than the preset threshold value, and the blob size was smaller. The SDGD filter was introduced to enhance edge detection upon the completion of each basic BGS technique as well as to complement the missing pixels. The results proved that fusing the SDGD filter with each elementary BGS increased segmentation performance and suited post recording video applications. An intelligent machine requires basic information such as moving-object detection from videos in order to deduce higher-level semantic information. A methodology that uses a texture measure is used to detect moving objects in video. The methodology is computationally inexpensive, requires minimal parameter fine tuning and also is resilient to noise, illumination changes, dynamic background and low frame rate. Experimental results show that performance of the proposed approach [7] is higher than those of state-of-the-art approaches. A robust visual tracking is imperative to track multiple occluded objects. Kalman filter and color information tracking algorithms are implemented independently in most of the current research. The proposed method combines extended Kalman filter with past and color information for tracking multiple objects under high occlusion [8]. The proposed method is robust to background modeling technique. Object detection is done using spatio-temporal Gaussian mixture model (STGMM). Tracking consists of two steps: partially occluded object tracking and highly occluded object tracking. Tracking partially occluded objects, extended Kalman filter is exploited with past information of object, whereas for highly occluded object tracking, color information and size attributes are used.

4. PROPOSED SCHEME

The human visual system is more sensitive to contrast than absolute luminance, we can perceive the world similarly regardless of the huge changes in illumination over the day or from place to place. In visual perception of the real world, contrast is determined by the difference in the color and brightness of the object and other objects within the same field of view. The contrast of image can be categorized as: global contrast and local contrast. Global contrast measures the brightness difference between the darkest and brightest element in the entire image. The local contrast is based on the retinex theory according to which our eyes sees the difference in respect to surroundings, a color map below can prove this point.

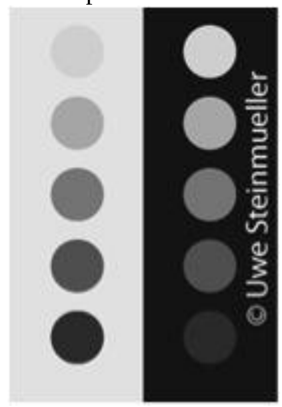


Figure 3: Color map

The circles in each row have exactly the identical brightness levels. Yet the top right circle looks a lot brighter than the one on the left. This is because our eyes see the difference to the local surrounding. The right circle looks much brighter with the dark gray background compared to a brighter background on the left. Just the opposite is true for the two circles on the bottom. For our eyes the absolute brightness is of less importance than the relative relation to other close areas. So, local contrast is very important for processing or enhancement of any image. In our work because of this human visual system local contrast map is extracted from an image and then on the basis of that a local thresholding approach will be used to convert the image onto binary format. Calculation of local contrast and then global thresholding algorithm like otsu is used and then local image edge detection is used in [9]. We have followed the same line of action but rather than using global thresholding, we use local thresholding, it removes the need of using again local edge detection algorithm like canny edge detection. Gray level co –occurrence matrix (GLCM) also called texton co- occurrence matrix (TCM) fulfills our purpose. It is a local contrast mapping method. Here basically TCM serves two purposes: make image's local contrast map, unaffected by the illumination variation of image and local edge detection. A gray-level co-occurrence matrix (GLCM) is generated by calculating how often a pixel with the intensity (gray-level) value i occurs in a specific spatial relationship to a pixel with the value j. Each element (i,j) in the resultant glcm is simply the sum of the number of times that the pixel with value i occurred in the specified spatial relationship to a pixel with value j in the input image. The number of gray levels in the image determines the size of the GLCM. GLCM of an image is computed using a displacement vector d, defined by its radius δ and orientation θ. To illustrate, the following figure shows how gray-co-matrix calculates the first three values in a GLCM. In the output GLCM, element (1,1) contains the value 1 because there is only one instance in the input image where two horizontally adjacent pixels have the values 1 and 1,

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS.....

respectively. glcm(1,2) contains the value 2 because there are two instances where two horizontally adjacent pixels have the values 1 and 2. Element (1,3) in the GLCM has the value 0 because there are no instances of two horizontally adjacent pixels with the values 1 and 3. Gray-co-matrix continues processing the input image, scanning the image for other pixel pairs (i,j) and recording the sums in the corresponding elements of the GLCM. Figure 4 shows this concept. A single GLCM matrix might not able to define all texture features of image, so multiple GLCM at different orientations are calculated. Above given example was with 0o orientation i.e. horizontally matching pairs are checked. Further it can be done at angle 45 o, 90 o, 135 o as shown in figure 5. In actual every pixel has eight neighboring pixels allowing eight choices for θ, which are 0°, 45°, 90°, 135°, 180°, 225°, 270° or 315°. However, taking into consideration the definition of GLCM, the co-occurring pairs obtained by choosing θ equal to 0° would be similar to those obtained by choosing θ equal to 180°. This concept extends to 45°, 90° and 135° as well. Hence, one has four choices to select the value of θ. In the last example matching pairs have been taken upto one distance, this constitutes the radius of GLCM. Applying large displacement value to a fine texture would yield a GLCM that does not capture detailed textural information. From the previous studies, it has been concluded that overall classification accuracies with δ = 1, 2, 4, 8 are acceptable with the best results for δ = 1 and 2. This conclusion is justified, as a pixel is more likely to be correlated to other closely located pixel than the one located far away. The dimension of a GLCM is determined by the maximum gray value of the pixel.

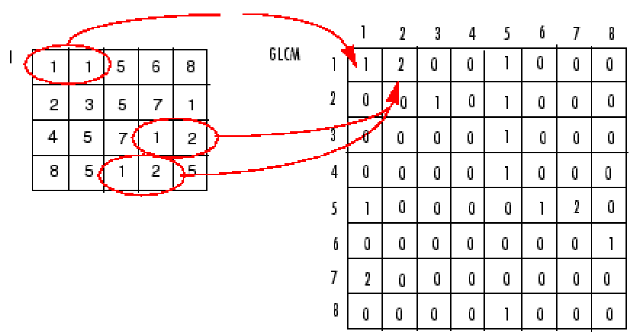


Figure 4: GLCM output of a test matrix.

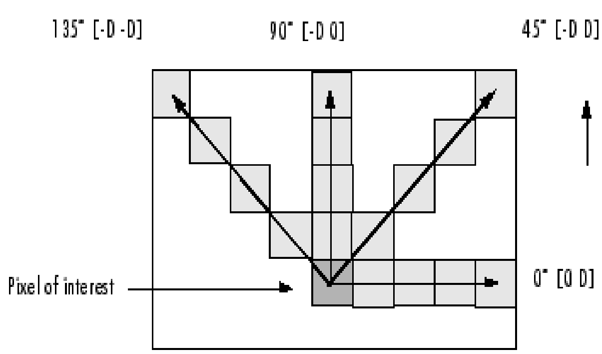


Figure 5: Multiple orientations to calculate GLCM

Following notations are used to explain the various textural features:

gij = (i, j)th entry in GLCM

gx(i) = ith entry in marginal probability matrix obtained by summing rows of gij

Ng = Number of distinct gray levels in the image

gy(i) = $\sum_{i=1}^{N_g} g(i,j)$

Contrast (con) = $\sum_i \sum_j (i-j)^2 g_{i,j}$

This statistic measures the spatial frequency of an image and is difference moment of GLCM. It is the difference between the highest and the lowest values of a contiguous set of pixels. It measures the amount of local variations present in the image. A low contrast image presents GLCM concentration term around the principal diagonal and features low spatial frequencies. From this GLCM process local contrast of image is obtained. From this we have developed the equation to calculate the threshold value. Since it will be a local threshold value so, the size of threshold matrix will be same as test image. For this formula we were inspired by work in [6]. We have done changes in that and final thresholding formula includes local mean of image and a gain factor which will act as a bias factor. This factor has the range (0-1) always. Its value will be determined experimentally. This formula is used in a window size of image and later on combined. Mathematical expression is shown below.

$$Threshold(i,j) = k(I_{mean}(i,j) + \sqrt{contrast})$$

Since the above method discussed calculates the local threshold, the operations are done in image blocks like if outcome of the algorithm is block size =55, then image matrix will be segmented into 55 by 55 blocks and thresholded.

4.1 Kalman Filter

The Kalman filter is an algorithm that uses a series of measurements observed over time, containing noise (random variations) and other inaccuracies, and produces estimates of unknown variables that tend to be more correct than those based on a single measurement alone. The algorithm works in a two-step process. In the prediction step, the Kalman filter produces estimates of the current state variables along with their uncertainties. Once the outcome of the next measurement (necessarily corrupted with some amount of error including random noise) is observed, these estimates are updated using a weighted average with more weight being given to estimates with higher certainty. Because of the algorithm's recursive nature, it can run in real time using only the present input measurements and the previously calculated state and its uncertainty matrix; no additional past information is required.

5. RESULT & DISCUSSION

The tracking algorithm has been successfully applied to video surveillance. The result shows that the algorithm has been able to track any single moving object. The algorithm has been implemented and tested on MATLAB 7.8 (2009) with operating system window 7. 10 frames at the interval of 1 sec have been set and 10 RGB frames are at the output captured

Volume 3, Issue IX, Sep 2015
ISSN 2320-6802

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS.....

by laptop's webcam. We have make an arrangement in the work such that we can either load pre captured videos or can take video in real time. Comparison of results is shown with otsu's thresholding algorithm which is a global processing thresholding method and our proposed thresholding scheme is based on local processing. Time consumed in both processes is compared. Kalman filter is used to track the moving object as discussed in previous chapter. In Kalman filter it is required to segment the moving object from background and background may consists small moving objects like tree leaves or moving fans in buildings behind etc. or lamination variation also. Initially we will discuss about the results in which small leaves movement is in the background. 15 frames are captured but here in figure 6 only 12 frames are shown because of space constraint.

Figure 6: Captured frames

These frames are further converted to gray scale frames to save memory requirements for the processing of three color bands of colored image, although very small losses are still but negligible. The gray scale frames are shown in figure 7.

Table 1: Time elapsed in each input video frames.

	Proposed Scheme (in sec)	Otsu's scheme (in sec)
Input video 1	14.075072	48.450924
Input video 2	18.031187	57.678786

In our work, initially background variable is initialized by making every pixel element value zero.

Then all coming frames are subtracted from the next frame from previous frame. This will constitute a frame difference matrix. A threshold value using proposed work is generated corresponding to every pixel of the frame difference matrix and compared with each pixel to get a binary frame difference mask (FDM) which will have only moving object in each frame. Background pixel considered in that particular frame will be black as FDM is binary image. Figure 8 shows the FDM by our proposed work and 9 shows by otsu's thresholding algorithm. A lot of noise in the background is visible by otsu's algorithm.

Figure 7: Gray scale frames of frames in figure 6

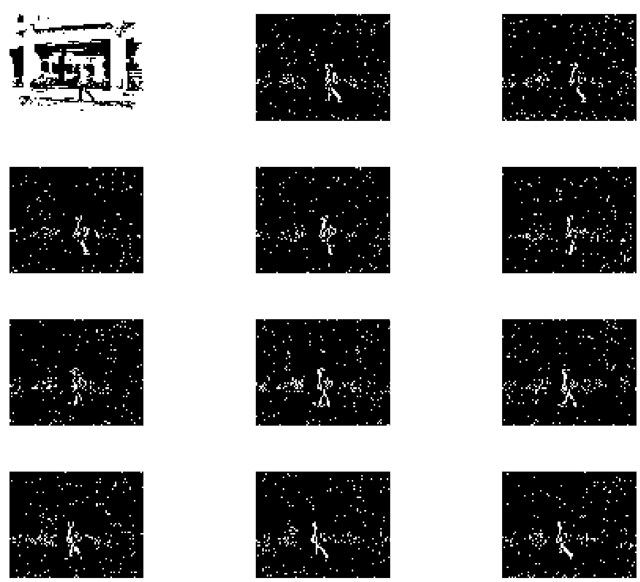


Figure 10: Time Elapsed Comparison in both cases

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

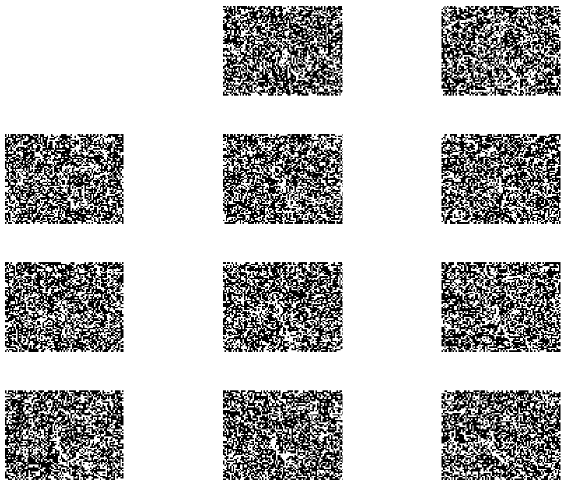
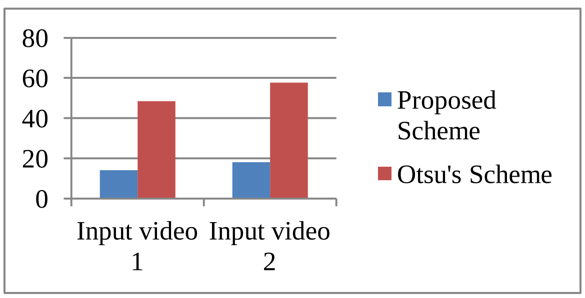


Figure 9: Otsu's threshold binary background image



6. CONCLUSION & FUTURE SCOPE

In object tracking various problem come in the way, one of them is the light intensity variation which results in false tracking of object. In this work illumination variation problem in object detection is tackled. Kalman filter is used in combination to track the object. Basically the problem lies with the foreground subtraction and in Kalman filter also the smallest rectangle coordinates are predicted and updated. If there is illumination variance then false object detection may take place. That's why we develop an algorithm which is resist to illumination variation and gives rise to correct tracking of object. For this we developed a thresholding algorithm which can compensate the illumination variation. This also reduces the time consumed in operation. The difference between otsu's global thresholding and our's thresholding in kalman filter are clearly shown in above section. So our proposed method performs better but there are still some limitations of this work like if background is having noise and very much dynamic like activity recognition i.e. identifying and distinguishing between a person's orientation that may impose challenge to background subtraction and object tracking..For that purpose adaptive background subtraction might not work well. If application is real time then only tracking of single object doesn't fulfill requirement. Multiple objects need to be tracked. For example if a car is to be automated on the basis of this then it must cite all objects in the way and on the sides. So this work can be extended to track multiple moving objects. In our work, camera is placed longitudinally to object moving. If object changes its direction or angle between camera and object is different than effectiveness of this algorithm is to be checked. Moreover in kalman filter design the coefficient values are selected by hit and trial method which is not adaptable to every type of frames captured. So an optimized algorithm for these coefficients can also be developed.

REFERENCES

[1] Deepak Kumar Rout, Sharmistha Puhan, "Video Object Detection in Dynamic Scene using Inter-Frame Correlation based Histogram Approach" International Journal of Computer Applications (0975 – 8887) Volume 82 – No 17, November 2013, pp19-24

[2] Virendra P. Vishwakarma, Sujata Pandey Member IEEE, and M. N. Gupta, "A Novel Approach for Face Recognition Using DCT Coefficients Re-scaling for Illumination Normalization" 15th International Conference on Advanced Computing and Communications © 2007 IEEE.

[3] D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part III, LNCS 5304, pp. 276–289, 2008 ©Springer-Verlag Berlin Heidelberg 2008

[4]Caius SULIMAN, Cristina CRUCERU, Florin MOLDOVEANU, "Kalman Filter Based Tracking in an Video Surveillance System" 10th International Conference on Development and Application Systems, Suceava, Romania, May 27-29, 2010.

[5] Parisa Darvish Zadeh Varcheie, Michael Sills-Lavoie and Guillaume-Alexandre Bilodeau, "A Multiscale Region-Based Motion Detection and Background Subtraction Algorithm" Sensors 2010, ISSN 1424-8220, 1041-1061.

[6]Farah Yasmin Abdul Rahman, Aini Hussain, Wan Mimi Diyana Wan Zaki, Halimah Badioze Zaman, and Nooritawati Md. Tahir, "Enhancement of Background Subtraction Techniques Using a Second Derivative in Gradient Direction Filter" Hindawi Publishing Corporation Journal of Electrical and Computer Engineering Volume 2013.

[7] Pranam Janney and Glenn Geers, "A Robust Framework For Moving-Object Detection And Vehicular Traffic Density Estimation" Arxiv: 1402.0289v1 [Cs.CV] 3 Feb 2014.

[8] Malik M. Khan, Tayyab W. Awan, Intaek Kim, and Young sung Soh, "Tracking Occluded Objects Using Kalman Filter and Color Information" International Journal of Computer Theory and Engineering, Vol. 6, No. 5, October 2014.

[9] Bo Liu, Yan Lin, Guan Guan, "A Method of Multi-scale Edge Detection for Underwater Image" Journal of Information & Computational Science 10: 2 (2013) 345–354.