Video Object Detection in Dynamic Scene using Inter-Frame Correlation based Histogram Approach

Deepak Kumar Rout
Image Analysis & Computer Vision Lab
Dept. of Electronics & Telecommunication Engg.
C. V. Raman College of Engineering,
Bhubaneswar, India

Sharmistha Puhan
Dept. of Computer Science & Engg.
C. V. Raman College of Engineering,
Bhubaneswar, India

ABSTRACT
In this paper the problem of video object detection under dynamic scene is considered. 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. But many a times due to failure of the detection algorithms the efficacy of the hardware remains poor. In previous version of this work a supervised approach to improvised the correct classification of the object and background regions has been proposed. Although the results obtained were as per expectation but the model parameters estimation; such as the threshold selection process was manually done. In order to make it adaptive to the scene, in this paper a classification algorithm has been used which takes the histogram of correlation matrix into account and classify the object. The segmentation of the correlation plane is done by a threshold. This threshold selection is made adaptive to the video sequence considered. This segmented plane along with the moving edge image is then taken into consideration to improvise the correct classification of the moving object in the video. It is observed that the proposed algorithm yields quite manageable results in terms of correct classification.

General Terms
Object detection, dynamic scene, illumination variation, histogram, adaptive threshold selection, background subtraction, spatio-temporal framework.

Keywords
Inter-frame correlation, Correlation distribution, Correlation threshold.

1. INTRODUCTION
Video conveys better information than the image due to an extra dimension called time. It is a general idea that if an object is changing its position with respect to a point in the space, then it is considered to be moving. Rest scene is said to be the background. The movements of the object can be properly analyzed if object is detected accurately. But the difficulty in identifying the object increases since in real world scenario, getting a static scene with respect to the object of concern is not always feasible. In a synthetic environment the fixedness of the background may be achieved, otherwise the background is always dynamic. The dynamic behavior of the background may be due to various reasons, such as variation of the illumination, slight movement of the objects other than the object of interest due to wind, swaying of the leaf, branches and trunk of a tree, swaying of small plants, ripples in a water body, movement of the camera itself etc. A lot of work has been carried out starting from simple techniques such as frame differencing and adaptive median filtering, to more sophisticated probabilistic modeling techniques. Background subtraction is a commonly used class of techniques for segmenting objects of interest in a scene for applications such as surveillance. Many background subtraction methods exist in literature [1]. Although these methods are simpler and have less computational complexity, but when the reference frame is not available many of them fails to yield good result. The problem becomes more difficult when the video sequences suffer from uneven lightening and illumination variation. Cavallaro et.al.[2] have proposed a color edge based detection scheme for object detection. Specifically the color edge detection scheme has been applied to the difference between the current and a reference image. This scheme is claimed to be robust under illumination variation. Jiglan Li[3] has proposed a novel background subtraction method for detecting foreground objects in a dynamic scene based upon the histogram distribution. Lu wang et al.[4] have proposed a new method that consistently performs well under different illumination conditions, including indoor, outdoor, sunny, rainy and dim cases. This method uses three thresholds to accurately classify pixels as foreground or background. These thresholds are adaptively determined by considering the distributions of differences between the input and background images and are used to generate three boundary set that represents the boundaries of the moving objects. Ivanov et al.[6] have proposed a new method of fast background subtraction based upon disparity verification that is invariant to run-time changes in illumination. The algorithm is easily implemented in real-time on conventional hardware because no disparity search is performed at run time. This method uses two or more cameras, which requires the off-line construction of disparity fields mapping the primary background image to each of the additional difference background image by assuming the background is fixed. Segmentation is performed at run time by checking color intensity values at corresponding pixels. With more than two cameras, the method gives more robust segmentation and also the occlusion shadows are eliminated. Kim et al. [7] have proposed an image segmentation method for separating moving objects from the background in image sequences. The method utilizes the spatio-temporal information for localization of moving objects in the image sequence by taking two consecutive image frames in the temporal direction and then comparing the two variance estimates to get the change detection mask which indicates moving areas and nonmoving areas. Deng et al.[8] have proposed a method for unsupervised segmentation of color-texture regions in images and videos. This method shows the

robustness of the JSEG algorithm on real images and video. The method consists of two independent steps: color quantization and spatial segmentation. Trucco et al.[9] have given a survey on video tracking, the problem of following moving targets automatically over a video sequence. Xiaofeng et al.[5] proposed a novel background subtraction method for moving objects detection based on three frames differencing. Although they dealt with both slow and fast moving object detection under illumination variation, they convert the color video to gray level image sequences and then applied their technique. Since the algorithm used gray level video sequences hence it lost the color information present in the video, which otherwise could results in better segmentation results. In Rout[12] a fusion based algorithm which takes three consecutive color video sequences and use an inter-plane correlation model to obtain the rough estimate of the object and then fuse the result obtained with the moving edge plane to improvise the result has been proposed. The results obtained were very good but the difficulty that still lies in that work is the selection of the threshold value, which decides the accuracy of the correlation based segmentation. In case the scene or the object changes the threshold need to be fixed manually. Thus in this paper threshold selection process is made adaptive by modifying Otsu [13] method, so as to fit to this situation. Otsu[13] has been used here because of it's binding capability.

In this paper, an algorithm has been devised to detect moving objects in a video sequence in outdoor scene, which is an extended version of author's previous work[12][14]. The problem is addressed when the reference frame is not available and the motion of the object is fast or slow enough to be missed by temporal segmentation. An efficient correlation based three frame differencing method is proposed, which takes care of all type of cases irrespective of the availability of reference frame and irrespective of the speed of the moving object.

The rest of the paper is organized as follows:
Section-2 described the proposed background subtraction algorithm; section-3 illustrates the experimental results. The paper ends with a conclusion in section-4.

2. PROPOSED BACKGROUND SUBTRACTION ALGORITHM

Detection of a moving object in a video can be done by temporal segmentation methods. Many existing work explain about the methodologies. Although it works up to some extent but, fails to detect exact moving area in a scene. This is because of the fact that, the temporal difference can detect the relatively changed information in successive frames. If the object is moving very fast then, up to some extent it can yield good results but, if the object is moving with a slow speed, then, the temporal segmentation methods result in cavities inside the object body. Many such cases have been solved by the use of morphological operations like erosion and dilation. Such operations are basically supervised operations. Thus, each time the new object enters into the scene, the morphological filter parameters need new assignments. Morphological filtering some times result in larger or smaller object size than the actual size of the object, which unnecessarily add to misclassification error. To make this more effective and robust, a spatio-temporal framework based method has been proposed in this paper, which takes care of the object shape, size as well as the cavities inside the object body.

The correlation matrix obtained is having elements, whose value lie between 0 to 1. In Rout[12] threshold (Tcor) has been chosen manually, which was used to separate the object from background. Since, correlation is computed in a spatio-temporal framework therefore, if the co-relation is high, the probability of the corresponding pixel belonging to background class is high. As the video sequences follow some distribution, so the correlation matrix derived from the video sequences also follows some distribution. Hence, to make (Tcor) adaptive, the histogram of the correlation matrix is constructed, where the independent axis takes the correlation value(Icor) and the dependent axis takes the probability of specific range of correlation for the correlation matrix. This can be understood by the Fig 1, as shown below.

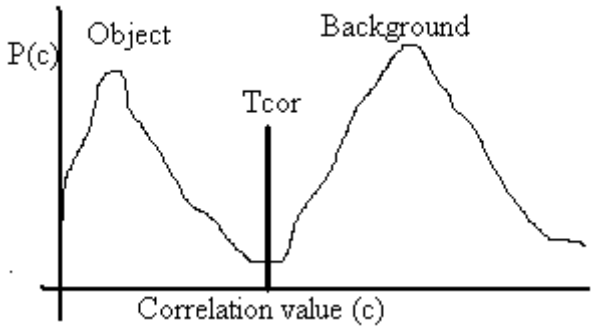

Fig 1: Histogram of the correlation matrix

Here 'C' is the random variable associated with quantized values of the actual correlation and P(c) is the probability of a particular (c) value in the correlation matrix. Since determination of the threshold largely depends upon the inter class variance of the correlation plane hence the very well known Otsu[13] method is used to find the Tcor from the correlation histogram. As per Otsu, the optimum threshold is the value (Ta) that maximizes the interclass variance and minimizes the intra-class variance is given by

$$\sigma^2(T_a) = \max_{0 \le T_{cor} \le 1} \sigma^2(T_{cor})$$

$$= \max_{0 \le T_{cor} \le 1} \frac{[m_a p_1(T_{cor}) - m(T_{cor})]^2}{P_1(T_{cor})[1 - P_1(T_{cor})]}$$

Where, m_a is the global mean given by $\sum_{i=0}^{1} iP_i$, $P_1(T_{cor})$ is the probability that a correlation is corresponding to object class and $P_2(T_{cor})$ is the probability that a correlation is corresponding to the background class.

Then, the inter frame correlation is computed using the gray scale equivalent video sequences. The correlation can be obtained by the following formulation:

$$I^{x}_{corr_{(i,j)}} = \frac{\sum_{\text{all m,n}} \left(\mu^{X}_{(i,j)} - C^{X}_{(m,n)}\right)_{f_{t-2}} \left(\mu^{X}_{(i,j)} - C^{X}_{(m,n)}\right)_{f_{t-1}} \left(\mu^{X}_{(i,j)} - C^{X}_{(m,n)}\right)_{f_t}}{\sqrt{\sum_{m,n} \left(\mu^{X}_{(i,j)} - C^{X}_{(m,n)}\right)^2_{f_{t-2}} \sum_{m,n} \left(\mu^{X}_{(i,j)} - C^{X}_{(m,n)}\right)^2_{f_{t-1}} \sum_{m,n} \left(\mu^{X}_{(i,j)} - C^{X}_{(m,n)}\right)^2_{f_t}}}$$

where, $\mu^{X}_{(i,j)}$, is the mean intensity value of (m×n) window in the X-plane. $C^{X}_{(m,n)}$ is the intensity value of the pixel in the X-plane. After computation of the $I^{X}_{corr(i,j)}$ values a matrix of dimension equal to the dimension of the image is constructed called I^{X}_{corr}. The values of elements of I^{X}_{corr} generally vary from 0 to 1. Less is the value less is the correlation between consecutive planes and hence, more is the probability that the pixel belongs to object class. Higher is the value, more is the temporal similarity and hence, less is the probability that the pixel belongs to object class. Thus, a threshold (Tcor) for the correlation is chosen in such a way that below which it corresponds to the object class and above which it corresponds to the background. The selection of such a threshold basically depends upon the amount of overall illumination variation and the change in overall intensity between the three consecutive frames. In this paper, this threshold has been automatically selected using Otsu[13]. Thus the result obtained by this can be given as

$$F^{X}{}_{CDM}(i,j) = \begin{cases} foreground & if \quad I^{X}{}_{corr(i,j)} \le T_{cor} \\ background & otherwise \end{cases}$$

Thus $F^{R}{}_{CDM}(i,j)$, $F^{G}{}_{CDM}(i,j)$ and $F^{B}{}_{CDM}(i,j)$ are obtained here. The three segmented planes are then merged by a logical union operator to get the final segmented result.

$$F_{CDM}(i,j) = F^{R}{}_{CDM}(i,j) \bigcup F^{G}{}_{CDM}(i,j) \bigcup F^{B}{}_{CDM}(i,j)$$

The result obtained here is a crude result which contains the expected foreground region and not the exact region. In order to get the exact region the moving edge is determined. Canny's edge detector[11] is used to get the edge information in each of the consecutive video sequences.

$$I^{x}_{edge(t)} = Canny\left(I^{x}_{t}\right)$$

The moving edge P_{cdme} is estimated using two consecutive edge frames

$$I_{edge(t)} = I^{R}_{edge(t)} \bigcap I^{G}_{edge(t)} \bigcap I^{B}_{edge(t)}$$

$$P_{cdme}(i,j) = \begin{cases} I_{edge(t)}(i,j) & if \quad I_{edge(t)}(i,j) \neq I_{edge(t-1)}(i,j) \\ 0 & otherwise \end{cases}$$

Where, P_{cdme} is the moving edge image. The results are then further improvised by the fusion of the obtained object area $F_{cdm}(i,j)$ along with the moving edge image P_{cdme} which gives a better view of the object boundary, to give a far better result of foreground region under illumination variation condition. The algorithm used in Rout[12] has been used for fusion of the above planes.

The final foreground obtained is then combined with the original image sequence to get the video object plane (VOP).

3. RESULTS & DISCUSSIONS

The main objective of this work is to test the robustness of the algorithm proposed in dynamic scenes. In order to include different types of dynamism, four different types of videos are chosen under four different situations. In particular, the sequences acquired in four different outdoor contexts are: walk video, where a man is moving slowly in the scene; hand video, which has been taken outside of a room with the wall as background, aswini video, which has been taken in the laboratory corridor with varying lighting condition with the help of an lighting source besides the normal sunlight and the navy video which was taken in the outside environment. The experiments were performed on a dual core system with 3GHz processor speed and with a DRAM of capacity 2GB. The processing time is strictly dependent on the quality of moving points and on the image dimension which is 640x480. Although the window size depends upon the object shape and extent of illumination variation, in the experiments and simulations the window size is taken to be 3x3, as it gives very good result, in accordance to the time complexity and accuracy. The thresholds used in the videos were estimated using the Otsu[13] formulation as explained in section 2.

In case of the four videos shown here, the different threshold estimated is given in Table 1.

Table 1: Threshold selection

Video	***Tcor (ground truth)***	***Tcor (estimated)***
Walk	0.23	0.27
Hand	0.18	0.16
Aswini	0.17	0.19
Navy	0.20	0.19

The Figure-2(a) show the original sequences of navy video, where the dynamism in the scene is due to variation of lighting condition due to movement of cloud. Figure-2(b) shows the change detection mask resulted by the proposed method. It can be observed here that the objects of interest that are the navy cadets were detected properly. Similar kind of observation is being made from Figure-3, which shows the aswini video sequence. The resultant image shown in Figure-3(b) is quite manageable as far as the misclassification error is concerned. Similar kind of observation is also being made from Figure-4, which shows the hand video sequence. The resultant image shown in Figure-4(b) is quite manageable as far as the misclassification error is concerned. Figure-5(a) shows the walk video sequences. The dynamism in the scene is due to the swaying of the tree, leafs, branches of the tree, the bush. The proposed method yields the change detection masks as shown in Figure-5(b). The proposed method yields such a better result because of the spatio-temporal framework which computes the inter-plane correlation, the efficacy of the threshold selection algorithm and the fusion model. Use of color information and use of moving edge information between the three consecutive video sequences enhances the

efficacy of the result. The percentage of error with respect to the threshold was calculated by comparing threshold value obtained using Otsu method and the manual method. The segmented video frames obtained with their corresponding manually constructed ground truth sequences with the help of GIMP software. This was done only for a comparison and analysis purpose. The algorithm was tested with different video sequences. Here four video results have been provided.

4. CONCLUSION

In this paper, the problem of moving object detection under dynamic scene has been addressed. The dynamism in the scene is considered due to the illumination variation, swaying of the tree and leafs. The correlation based method yields good result but the silhouettes present around the object. The inter-plane correlation is used to tackle the problem of illumination variation. The algorithm is modified so as to make the threshold selection process adaptive to the video sequence to make the algorithm more efficient in different environment and situations. In order to make the algorithm adaptive to the scene and environment, Otsu's threshold selection algorithm has been used with the histogram of the correlation matrix to classify the foreground and the background efficiently. The fusion of edge map and the correlation based result improves the detection result.

Fig 2(a): Original video sequence of navy video

Fig 2(b): The Change Detection Mask obtained by the proposed method for navy video sequences

Fig 3(a): Original video sequence of aswini video

Fig 3(b): The Change Detection Mask obtained by the proposed method for aswini video sequences

Fig 4(a): Original video sequence of hand video

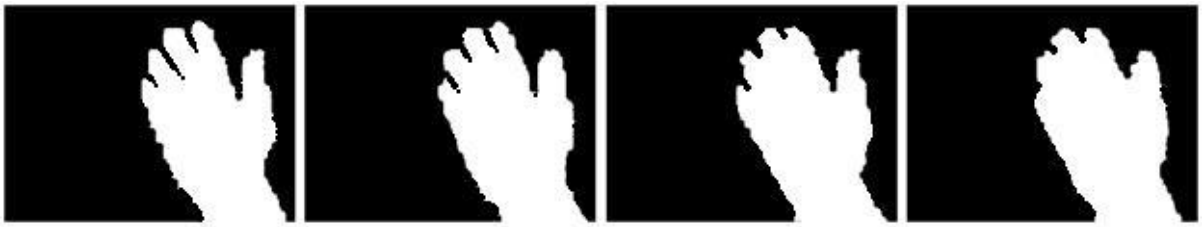

Fig 4(b): The Change Detection Mask obtained by the proposed method for hand video sequences

Fig 5(a): Original video sequence of walk video

Fig 5(b): The Change Detection Mask obtained by the proposed method for walk video sequences

5. REFERENCES

[1] S. Y. Elhabian, K. M. El-Sayed and S. H. Ahmed 2008. Moving Object Detection in Spatial Domain using Background Removal Techniques State of Art, Recent Patents on Computer Science, Vol. 1, No. 1, Bentham Science Publishers Ltd. pp.32-54.

[2] Cavallaro and T. Ebrahimi, 2001. Change Detection based on Color Edges, IEEE International Symposium on Circuits and Systems 2001, ISCAS2001, Vol. 2, pp. 141-144.

[3] Jinglan Li, 2009. Moving Object Segmentation Based on Histogram for Video Surveillance, Journal of Modern Applied Science, Vol.3, No.11.

[4] Lu Wang and N. H. C. Yung, 2010. Extraction of moving objects from their background based on multiple adaptive thresholds and boundary evaluation, IEEE Tran. on Intelligent Transportation System, vol 11, No.1, pp 40-51.

[5] Lian Xiaofeng, Zhang Tao and Liu Zaiwen, 2010. A Novel Method on Moving Objects Detection Based on Background Subtraction and Three Frames Differencing, Proc. of IEEE International Conference on Measuring Technology and Mechatronics Automation, pp 252-256.

[6] Yuri Ivanov, Aaron Bobick, John Liu, 1998. Fast Lighting Independent Background Subtraction, Proc. IEEE Workshop on Visual Survillance, Bombay-India, pp.49-55.

[7] M. Kim, J. choi, D. Kim and H. Lee, 1999. A VOP Generation Tool: Automatic Segmentation of Moving Objects in Image Sequences based on Spatio-Temporal information, IEEE Transaction on circuits and Systems for Video Technology, Vol. 9, No. 8, pp. 1216-1226.

[8] Y.Deng, B.S.Manjunath, 2001. Unsupervised Segmentation of Color-Texture Regions in Images and Video, IEEE Transactions on Pattern Analysis And Machine Intelligence vol. 23,pp. 800-810.

[9] E. Trucco and K. Plakas, 2006. Video Tracking: A Concise Survey, IEEE Journal of Oceanic Engineering, Vol. 31, No. 2, pp. 520-529.

[10] S. D .Babacan and T. N. Pappas, 2007. Spatiotemporal Algorithm for Background Subtraction, Proc. of IEEE

International Conf. on Acoustics, Speech, and Signal Processing, ICASSP 07, Hawaii, USA,pp.1065-1068.

[11] John Canny, 1986. A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-8(6):679–698.

[12] Deepak Kumar Rout, Sharmistha Puhan, 2012. A Spatio-Temporal Framework for Moving Object Detection in Outdoor Scene, Global Trends in Information Systems and Software Applications, Springer Berlin Heidelberg CCIS, Vol. 270, pp. 494-502.

[13] N. Otsu, 1979. A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man and Cybernetics, Vol. 9, No. 1,pp.62-66.

[14] Deepak Kumar Rout, Sharmistha Puhan, 2013. Video Object Detection using Inter-frame Correlation Based Background Subtraction, to be published in proceedings of IEEE RAICS.