Slow and Fast Moving Object Detection under Illumination Variation Condition

S. Puhan[1], D. K. Rout[2], N. K. Kamila[1,*]

[1]Dept. of Computer Science & Engineering, C.V.Raman College of Engineering, Bhubaneswar, 752054, India
[2]Image Analysis & Computer Vision Lab, Dept. of Electronics & Telecommunication Engg, C.V.Raman College of Engineering, Bhubaneswar, 752054, India

Abstract   In this paper, the problem of video object detection has been addressed under illumination variation condition. It is assumed that the camera is fixed, hence the problem is formulated in a fixed background moving object framework. Various works has been done in this area, starting from simple deterministic methods to complex stochastic models. Complex stochastic models give better results but at a cost of the time complexity. Hence, simpler deterministic methods are employed to overcome this issue. In this work, the case of unavailability of reference frame is considered. To solve this problem a novel reference frame extraction method is proposed, which takes a little time to extract the reference frame from the video sequences and once the reference frame is obtained then the two-frame differencing method could be used to get better results. But this approach fails when the video object is moving very slow. Improvement in the results can be achieved by using the proposed color based three frame differencing approach. In this technique, the three frame differencing algorithm is applied to each of the RGB planes and then the results have been segregated to get a better result. A method has been proposed which computes the Inter-frame similarity between the three consecutive frames to further improve the results. This result is then fused with moving edge estimate of the particular video sequence. The inter-frame similarity gives a rough estimate of the object and the moving edge estimate provides the object boundary. It can be observed that the proposed method produces very good result in terms of miss classification error.

Keywords   Background Subtraction, Inter-Frame Similarity, (CB3FD) Color Based Three Frame Differencing

1. Introduction

Detection of moving objects from video sequences is a key issue for visual surveillance system, with the purpose of subtracting interesting target area and locating the moving objects from the video flows. It is important to segment the motion areas effectively for recognition, tracking as well as behavior understanding. Video segmentation refers to the identification of regions in a frame of video that are homogeneous in some sense. Different features and homogeneity criteria generally lead to different segmentation of same data; for example, color segmentation, texture segmentation, and motion segmentation usually result in segmentation maps. Furthermore, there is no guarantee that any of the resulting segmentation will semantically meaningful, since semantically meaningful region may have multiple colors, multiple textures, or multiple motions.

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 common approach to identify the moving objects. Many background subtraction methods exist in literature[1]. Y. Ivanov, et.al[2] described a new method of fast background subtraction based upon disparity verification that is invariant to run-time changes in illumination. Using two or more cameras, the method takes the off-line construction of disparity fields mapping the primary background image to each of the additional reference background images. At run-time, segmentation is performed by checking color intensity values at corresponding pixels. If more than two cameras are available, more robust segmentation can be achieved and in particular, the occlusion shadows can be generally eliminated as well. A. Cavallaro et al[3] 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. Rout et al[4] have proposed 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. Jiglan Li[5] has proposed a novel background subtraction method for detecting

foreground objects in a dynamic scene based upon the histogram distribution. Yu Wu et al.[6] have proposed a novel layered background modeling method for objects detection, where every block on the first layer is modeled via texture based on local binary pattern (LBP) operators. Then the modeling granularity is deflated onto the second layer to model via codebook. Layered match is done from top down when a new video frame enters. M. Kim et al.[7] have proposed a genetic algorithm based segmentation method that can automatically extract and track moving objects. The method consists of the spatial and temporal segmentation. The spatial segmentation divides each frame into regions with accurate boundaries and the temporal segmentation divides each frame into background and foreground areas. The spatial and temporal segmentation results are then combined for object extraction. Even though these methods detect moving objects well, still the object detected always suffer from the uncovered background situations, still object situation, light changing, shadow, residue background problem and noise. Spagnolo et al.[8] have proposed a method of moving object segmentation using a reliable foreground segmentation algorithm. The algorithm combines temporal segmented result with a reference background frame. In order to handle the dynamism of the scene caused due to illumination variation they have given a new approach for background modification. Even if this method detects moving objects well, it is not able to eliminate shadows especially when they are highly contrasted on the background. Lu wang et al.[9] 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. Xiaofeng et al.[10] have proposed a method where the background subtraction result is fused with an improved three frame differencing result. The adaptive background model is build by Gaussian model for each pixel in the image sequences and combining with temporal differencing method to update the selective background and simultaneously use background subtraction method to extract movement areas from the background model. They have used an adaptive threshold to take care of the illumination variation. In all most cases, either the reference frame is available or it is estimated by some modeling. Xiaowei et al.[11] have used a unified framework named detecting contiguous outliers in the low-rank representation to overcome the above challenges. Xiaofeng et al[10] deals with a case where reference frame is not available. They have used gray level video sequence instead of color sequences. The color information is a very important feature used for proper segmentation hence we have proposed a

modification to the existing Xiaofeng method to work with color information. The results are further improvised by our proposed fusion model based method.

Rest of the paper is organized as follows: Section II describes the extraction of reference frame approach. Section III describes the color based three frame differencing algorithm which is a modification of the Xiaofeng method. Section IV describes the proposed inter- frame similarity approach as well as it explains the proposed fusion model which takes in to account the out put of the inter-frame similarity and the moving edge estimate to produce excellent results. Finally section V summarizes the whole work.

2. Extraction of the Reference Frame

In case of real world scenario availability of the reference frame is not always possible. In such cases it is really very difficult to detect the object of interest. Thus the process of detection can be made easier if the reference frame can be extracted by the available video sequences. In order to deal with such cases we have devised an algorithm which first extracts the reference frame by a simple comparison between the video sequences and then the estimated reference frame is used to detect the video object in the rest of the sequences by the absolute difference method. In practice, obtaining a reference image with only stationary elements is not always possible, and building a reference from a set of images containing one or more moving objects becomes necessary. This necessity applies particularly to situations describing busy scenes or in cases where frequent updating is required. One procedure for generating a reference image is as follows. Consider the first image in a sequence to be the reference image. When a non stationary component has moved completely out of its position in the reference frame, the corresponding background in the present frame can be duplicated in the location originally occupied by the object in the reference frame. When all moving objects have moved completely out of their original positions, a reference image containing only stationary components will have been created. Object displacement can be established by monitoring the changes in the positive accumulated difference image. A difference image between two images taken at times t_i and t_j may be defined as

$$d_{i,j}(x, y, z) = f_i(x, y, z) - f_j(x, y, z) \quad (1)$$

If the $d_{i,j}(x,y,z) < Threshold(T)$, then the pixel in $f_i(x,y,z)$ is to be replaced by the pixel in $f_j(x,y,z)$, otherwise $f_i(x,y,z)$ is to be kept as it is. The $f_i(x,y,z)$ is now the extracted reference frame.

Once the reference frame is extracted then the two frame differencing approach is used to extract the object from the video sequence. The extracted reference frame is then used along with the subsequent frames to get the change difference mask using the following equation.

$$d_{i,j}(x,y,z) = \begin{cases} 1 & if \quad |f(x,y,z,t_p) - f(x,y,z,t_q)| > T \\ 0 & otherwise \end{cases} \quad (2)$$

The difference image is thresholded using a global optimum adaptive approach proposed by Otsu[12]. Where T is given by:

$$T = \frac{\mu_1 + \mu_2}{2} + \frac{\sigma^2}{\mu_1 - \mu_2} \ln\left(\frac{P_2}{P_1}\right) \quad (3)$$

where, μ_1, μ_2 are the mean and σ is the standard deviation of the object class and background class distribution. P_1, P_2 are the probability that a pixel belongs to background and object class respectively.

Filtering

The Change Detection Mask (CDM) is the image which shows the moving object area by white region and the non-moving or background area by black region. This is obtained by binarization of the background subtraction result. Many times the CDM obtained contains a lot of noises, due to the residual effect present in the extracted reference frame. The residual part left in the image after the extraction operation may create small cavities inside the object region and background regions. Thus a median filtering or morphological filtering approach may solve the problem upto a great extend.

Experimental Results

In order to justify the efficacy of the frame extraction algorithm proposed here, three video sequences are considered. Fig.-1(a) and 1(b) shows the initial frame and the frame after which the reference frame has been extracted completely, of the *Aswini* video. The extracted reference frame which is shown in Fig.-1(c) is obtained after 1.48 secs. After the extraction of the reference frame the conventional absolute difference method has been used to detect the object in the subsequent frames. Fig.-1(d) and 1(e) shows the initial frame and the frame after which the reference frame extraction is completed, respectively for the *walk* video. The reference frame has been extracted completely after 1.6secs. After that the detection of object starts with a very high degree of accuracy. Similar kind of observation can be made in Fig.-1(g), (h) and (i) which deals with the *Ivana_step* video sequence. In case of the *Ivana_step* video sequence time taken to extract the reference frame is 0.32 secs.

Figure 1. (a), (d) and (g) are the initial frames of aswini, walk and ivana_step videos and (b), (e) and (h) are the frames after which the reference frame has been completely extracted for aswini, walk and ivana_step videos. (c), (f) and (i) are the extracted reference frames of the aswini, walk and ivana_step videos respectively

In this section the problem of moving object detection has been formulated in the absence of reference frame and *a priori* knowledge of the object of concern. The reference frame extraction approach yields very good result in such conditions. In spite of the good results it is only limited to the fast moving objects and not fit for the slow moving object cases. It has another limitation that it fails to detect the video object for first few frames, i.e. unless until the reference frame is completely extracted, the object can not be detected. Thus the color based three frame differencing method has been proposed, which is a modification of the Xiaofeng et al method[10].

3. Proposed Modification to Xiaofeng *et al* Approach

Xiaofeng et al[10] suggested the pre-processing stage in their algorithm, where the color image has to be converted to gray level image. This step converts the RGB color image to single plane gray level image and there by minimizes the computational complexity. Although this produced very good results but further improvement can be achieved by the help of color information. Hence a modified algorithm based on the existing Xiaofeng algorithm has been proposed to increase the efficacy of the algorithm in terms of the misclassification error.

Background Updating

Let $f(x, y, z)_R$, $f(x, y, z)_G$ and $f(x, y, z)_B$ are the RGB components of pixel *(x, y)* respectively. The Xiaofeng method is applied to each of these three plane of each frame and then the results are segregated to get the required output. The modifications are as follows:

    1. Read three consecutive frames, say $f_{k-1}(x,y,z)$, $f_k(x,y,z)$ and $f_{k+1}(x,y,z)$.

    2. Get the gray scale difference between the two consecutive frames,

$$IZ_{k,k-1}(x,y,z)=|I_{k-1}(x,y,z) - I_k(x,y,z)|$$
$$IZ_{k+1,k}(x,y,z)=|I_k(x,y,z) - I_{k+1(x,y,z)}|$$

    3. If $IZ_{k,k-1}(x,y,z)$ and $IZ_{k+1,k}(x,y,z)<Th+(1/N)\sum|I_k(x,y,z)-I_{k+1}(x,y,z)|$

        then $B(x,y,z)=I_{k+1}(x,y,z)$

        else $M(x,y,z)=I_{k+1}(x,y,z)$

Where Th is the gray threshold, the term $(1/N)\sum|I_k(x,y,z) - I_{k+1}(x,y,z)|$ denotes image changes in overall light, so that the threshold Th can adapt to the slight changes of environmental lighting conditions. N is the number of pixels in the image, B(x,y,z) and M(x,y,z) represents separately the background and object regions.

Moving Object extraction

The object regions in video sequences can be obtained by thresholding the absolute differences between the pixels of background image B(x, y, z) and those of video image Ik(x, y, z). In this thesis work we adopt the discrimination method based on the relative difference which can enhance the detection effect of moving object in shadow regions.

$$\Delta I_k(x,y,z)=|I_k(x,y,z) - B(x,y,z)| \qquad (4)$$

$$DB_k(x,y,z)=$$
$$\begin{cases} 1 & \text{if } \Delta I_k(x,y,z)/B(x,y,z) \geq Th' \\ 0 & Otherwise \end{cases} \qquad (5)$$

The detection results can be further enhanced by selecting the threshold *Th'* properly, as this is one of the key component for the binarization of the CDM of $DB_k(x, y, z)$. Since the three frame differencing method has strong adaptability to variation in lighting condition while background subtraction can makeup the disadvantage of interior cavity in the moving object extracted by the traditional temporal differencing, hence we can get the moving object segment from the background images through merging the motion regions obtained by the help of equations *(4), (5)* and the following equation

$$Re_k(x,y,z)=$$
$$\begin{cases} 255 & \text{if } DB_k(x,y,z) \bigcup M(x,y,z)=1 \\ 0 & Otherwise \end{cases} \qquad (6)$$

where $Re_k(x, y, z)$ represents the final extracted moving object based on the above stated method.

Post Processing

The image obtained in the stated method, $Re_k(x, y, z)$ contains a lot of residual noises except those moving objects in the image. The moving object(s) thus obtained may also contain cavities inside. Thus the median filtering and morphological filtering like erosion and dilation of the $Re_k(x, y, z)$, will minimize such noises as well as it will further improve the small cavity regions present inside the object region, which were misclassified as background class earlier.

Experimental Results

Three video sequences are considered to verify the Color Based Three Frame Differencing (CBTFD) algorithm. In Fig.2, the *aswini* video sequence is considered and it is assumed that the reference frame is not available. The three frame differencing algorithm based Xiaofeng method results in a good result which is shown in Fig.2 (b). It can be observed here that some portion near the chest area has been missing in the change detection mask (CDM). Thus a large number of pixels get misclassified as background. By increasing the mask size in the morphological filtering operation this can be overcome but by that the size of the object will be larger than the actual size of the object. Fig.2(c) shows the CDM obtained by the modified Xiaofeng method. By introducing the color information to segment the moving object area the misclassification error get reduced to a greater extent. Fig.3(a) shows the original *walk* video sequence. Fig.3(b) and Fig.3(c) show the CDM obtained by the Xiaofeng method and color based three frame differencing method respectively. In Fig.3(b), the size of the object obtained by Xiaofeng method is quite larger than the original size which is overcome by the color based three

frame differencing method. The similar kind of observations can be made from Fig.4(b) and Fig.4(c) respectively. The values of the *Th* and *Th'* used in the experiments are given in the table-1.

Figure 2(a). Original frames of aswini video sequence

Figure 2(b). CDM obtained by Xiaofeng et al method

Figure 2(c). CDM obtained by the proposed CB3FD method

Figure 3(a). Original frames of walk video sequence

Figure 3(b). CDM obtained by Xiaofeng et al method

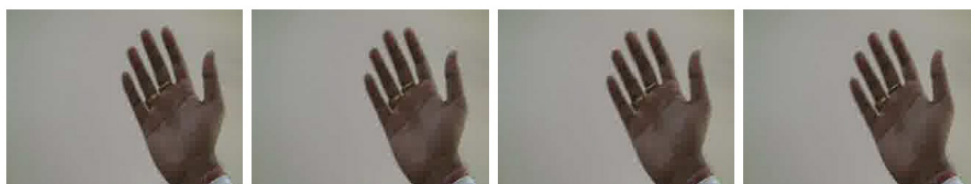
Figure 3(c). CDM obtained by the proposed CB3FD method

Figure 4(a). Original frames of hand video sequence

Figure 4(b). CDM obtained by Xiaofeng et al method

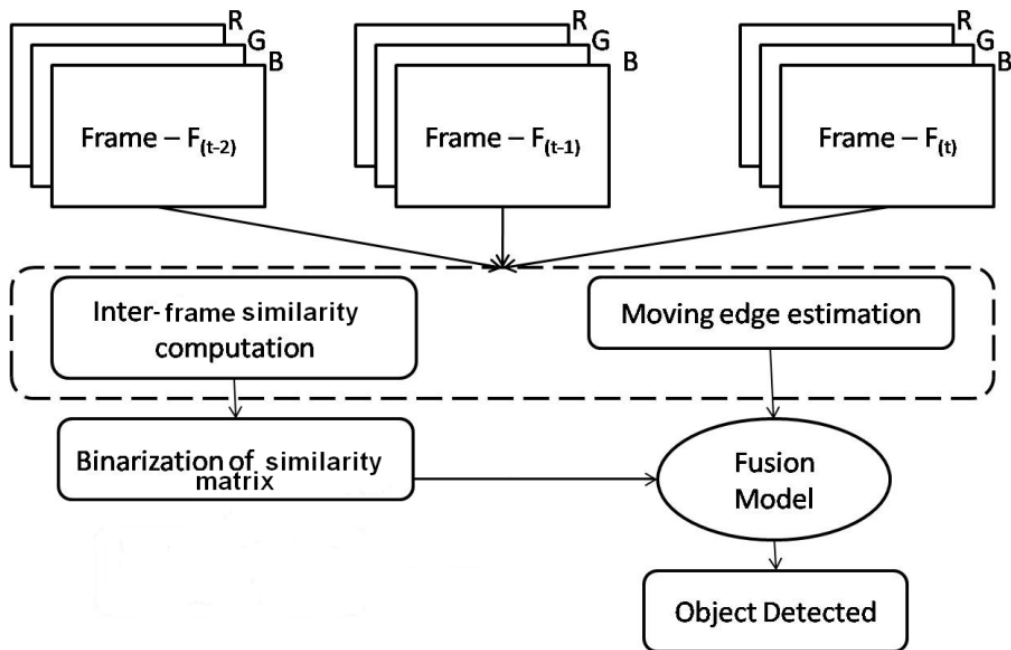**Figure 4(c).** CDM obtained by the proposed CB3FD method

Figure 5. Overview of the proposed algorithm

In this part of the paper it is assumed that the reference frame as well as any *a priori* knowledge regarding the object of concern is not available. In order to cope with such problem, the three frame differencing followed by the morphological or median filtering technique proposed by Xiofeng et al is used, which yields very good results. It also takes care of slight illumination variations. The Xiaofeng method is a very faster algorithm as it works only on gray level images. But; due to this gray level conversion the video sequences loose the color information present in each frame. Thus the modification to the Xiaofeng algorithm of three frame differencing has been proposed which not only gives very good detection but also retain the color information of the object of concern, in terms of the misclassification error.

4. Proposed Inter-Frame Similarity Approach

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. To make this more effective and robust, we have proposed a spatio-temporal framework based method, which takes care of the object shape, size as well as the cavities inside the object body. Morphological filtering some

times result in larger or smaller object size than the actual size of the object, which unnecessarily add to misclassification error. Temporal difference can detect the relatively changed information in successive frames. Therefore it is difficult to detect the overlaps in the moving object and always cause some cavities inside the moving entity. In addition the gray changing region detected by the temporal difference method contains the real moving object and the background changing area which occluded by moving object in the previous frame, thus causing the detected object larger than it's real and even overlaps. In order to cope with such issue, we have proposed a three frames temporal similarity method which computes the inter-frame similarity and then determines the moving edge image and then fuses them in a deterministic framework. Comparing with the traditional method, the improved method has a significant adaptability to any change in speed

of moving object and illumination variation.

The three color planes are extracted from the color video sequences. Then the inter frame similarity between the three consecutive Red, Green and Blue planes such as P_{red}, P_{green} and P_{blue} is computed. The similarity criterion considered here is the intensity value of each color plane.

The similarity is discarded and the dissimilarity in intensity between the three color planes is used as the reference foreground region. The algorithm for the computation of the crude reference foreground is:

Algorithm-1

1. Extract the color planes such as P_{red}, P_{green} and P_{blue}
2. Determine the inter-frame similarity $I_x(f^{t-1}, f^t, f^{t+1})$ between frames f^t, f^{t-1} and f^{t+1} frames using a window 'w'

$$I_x(f^{t-1}, f^t, f^{t+1})$$

$$= \frac{m[w(f^{t-1})w(f^t)w(f^{t+1})] - m[w(f^{t-1})]m[w(f^t)]m[w(f^{t+1})]}{\sqrt{\text{var}[w(f^{t-1})]\,\text{var}[w(f^t)]\,\text{var}[w(f^{t+1})]}} \qquad (7)$$

where 'm' is the mean and var is the variance computed for the respective 'w' window. 'x' can be the 'Red – plane', Green – Plain or Blue – plane.

3. The above step is repeated for the consecutive Red, Green and Blue planes of the f^t, f^{t-1} and f^{t+1} video sequences and the inter-frame similarity is computed.

4. The three inter-frame similarities obtained are then fused by logical 'OR' operator to get the final similarity matrix.

$$I(f^{t-1}, f^t, f^{t+1}) = I_R(f^{t-1}, f^t, f^{t+1}) \bigcup I_G(f^{t-1}, f^t, f^{t+1}) \bigcup I_B(f^{t-1}, f^t, f^{t+1}) \qquad (8)$$

5. Compare the elements of the final similarity matrix with a manually chosen threshold value. T_{sim}.

6. The pixels whose similarity is less than the defined threshold are marked as foreground and rest all are treated as background. i.e,

$$F_{CDM}(i,j) = \begin{cases} 255 & if \quad I(f^{t-1}, f^t, f^{t+1}) \le T_{sim} \\ 0 & otherwise \end{cases} \qquad (9)$$

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[13] is used to get the edge information in each of the consecutive video sequences. The values of the elements of the final similarity matrix generally vary within 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. More is the value of the element more is the temporal similarity and hence, less is the probability that the pixel belongs to object class. Thus, a threshold of similarity 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 conjugative frames. In this paper, we have selected this threshold manually by trial and error basis. The edge information is very vital in case of the object detection and recognition. In our method, we have computed the moving edge map by determining the differences between the edge maps of the three consecutive video sequences. The moving edge estimate is computed by

$$F_{ME}(f^{t-1}, f^t)$$

$$= |\psi(f^{t-1})^R - \psi(f^t)^R| \bigcup |\psi(f^{t-1})^G - \psi(f^t)^G| \bigcup |\psi(f^{t-1})^B - \psi(f^t)^B| \qquad (10)$$

$$= |\phi(\nabla G \times f_R^{t-1}) - \phi(\nabla G \times f_R^t)| \bigcup |\phi(\nabla G \times f_G^{t-1}) - \phi(\nabla G \times f_G^t)| \bigcup |\phi(\nabla G \times f_B^{t-1}) - \phi(\nabla G \times f_B^t)|$$

where, $\psi(f)$ is the edge map obtained by the Canny's Edge detector, which is obtained by applying gradient operation on the Gaussian convoluted image $G*f$. Then the thinning of the edges is done by applying non-maximum suppression to the gradient magnitude and linking of the non-joined edges is achieved by the help of thresholding operation with hysteresis. The results obtained are then further improvised by the fusion of the obtained object area F_{CDM}, along with the moving edge image F_{ME}, which gives a better view of the object boundary, to give a far better result of foreground region under illumination variation condition.

The Fusion algorithm which takes the F_{ME} and F_{CDM} in to account to get the final result is given below.

Algorithm-2

1. *If $F_{ME}(i,j)$ Є C_{obj} and $F_{CDM}(i,j)$ Є C_{obj} then pixel$_{(i,j)}$ Є Object*
Else
If $F_{ME}(i,j)$ Є C_{obj} and $F_{CDM}(i,j)$ Є C_{back} then pixel$_{(i,j)}$ Є Object
Else
If $F_{ME}(i,j)$ Є C_{back} and $F_{CDM}(i,j)$ Є C_{obj} then pixel$_{(i,j)}$ Є Background
Else
If $F_{ME}(i,j)$ Є C_{back} and $F_{CDM}(i,j)$ Є C_{back} then pixel$_{(i,j)}$ Є Background

where C_{obj} is the object class and C_{back} is the background class. *Pixel$_{(i,j)}$* is the corresponding pixel under consideration. The final foreground obtained is then combined with the original image sequence to get the video object plane (VOP).

EXPERIMENTAL RESULTS

The theoretical formulation of the inter-frame similarity based model and the algorithms proposed sounds good; but without practical analysis and simulation the actual efficacy of the algorithms can not be claimed. In order to test the robustness and efficacy of the algorithms and the framework proposed, we have considered many video sequences in indoor and outdoor environments. Special care has been taken to capture the illumination variations caused due to lighting variation and cloud movements. Among many videos three videos are considered here to prove the strength of the proposed method. These three videos are *navy, aswini* and *hand* videos, which are filmed in the lawn facing to our research lab, in the corridor and inside the lab respectively. In case of the *navy* video the illumination variation is slight and this variation is due to cloud movement. In the *aswini* video the illumination variation is due to external lighting and in case of the hand video lighting variation is achieved by varying the lighting condition in side the closed lab. In these test videos it has been assumed that the reference frames of the scene under consideration are not available. In such cases the background subtraction algorithms fail to yield accurate results.

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 video frame dimension which was 640x480. Although the window size depends upon the object shape and extent of illumination variation, in our experiment and simulations we have taken the window size to be 5x5, as it gives very good results, in accordance to the time complexity and accuracy. The thresholds used in all the algorithms are choosen manually depending upon the object body color and dimension of the object. The threshold T_{sim} varies from 0 to 1 depending upon the scene content and lighting variation. In case of the three videos shown here, the different threshold selected by trial and error basis are given in the following table.

Table 1.   Thresholds used in the considered video sequences

Video	*Th*	*Th'*	T_{sim}
aswini	43	65	0.20
navy	NC	NC	0.82
Hand	22	50	0.14
Walk	85	110	NC

NC: Not Considered

Table 2.   Computation of percentage of misclassification error (PME)

Video	Frame No.	PME CB3FD Method	PME Proposed Method
aswini	410	1.88145	0.08711
	412	1.95763	0.07877
	414	1.87342	0.08395
	416	2.01923	0.09912
hand	629	3.81768	0.06514
	630	3.98129	0.07652
	631	2.79981	0.05514
	632	3.80278	0.06190
navy	102	8.43212	1.54201
	103	8.07665	2.01652
	105	7.87630	1.89625
	106	8.20934	1.15823

The Fig-6(a) shows the original sequences of *aswini* video, which is a outdoor video. Fig 6(b) and Fig 6(c) show the CDM obtained by the modified Xiaofeng method (CB3FDM) and our inter-frame similarity based method. Fig 6(d) shows the VOP resulted by our proposed method. It can be seen here that some silhouette present in the result of Xiaofeng method where as our proposed method gives far better result. Similar kind of observation is being made from Fig 7, that shows the *navy* video sequence which is an outdoor video. The resultant CDM depicted in Fig 7(c) quite better than the

images obtained by the modified Xiaofeng method which is shown in Fig 7(b).Fig 8(a) shows the original sequences of *hand* video, which is an indoor video. Fig 8(b) and Fig 8(c) show the CDM obtained by the modified Xiaofeng method (CB3FDM) and our inter-frame similarity based method. Fig 8(d) shows the VOP obtained by our proposed method, which is far better than the Xiaofeng method. The table-2 shows the percentage of misclassification error in case of the modified Xiaofeng method (CB3FDM) and our proposed inter-frame similarity based method. Here, it is clearly observed that our proposed method decreases the PME to a good extent. Our proposed method yields such a better result

as compared to Xiofeng method because of the spatio-temporal framework which computes the inter-plane similarity, use of color information and use of moving edge information between the three consecutive video sequences. The percentage of misclassification error was calculated by comparing 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 many video sequences. Here we are providing three video results.

Figure 6(a).   Original frames of aswini video sequence

Figure 6(b).   CDM obtained by modified Xiaofeng et al method

Figure 6(c).   CDM obtained by proposed inter- frame similarity based method

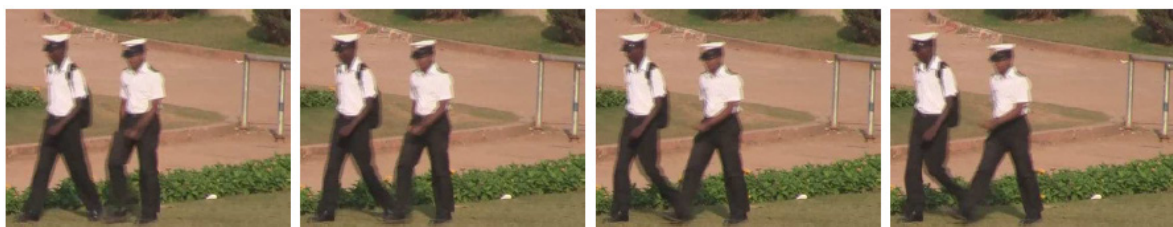**Figure 6(d).**   Final VOP obtained by proposed method

Figure 7(a).   Original frames of navy video sequence

Figure 7(b).   CDM obtained by modified Xiaofeng et al method

Figure 7(c).   CDM obtained by proposed inter- frame similarity based method

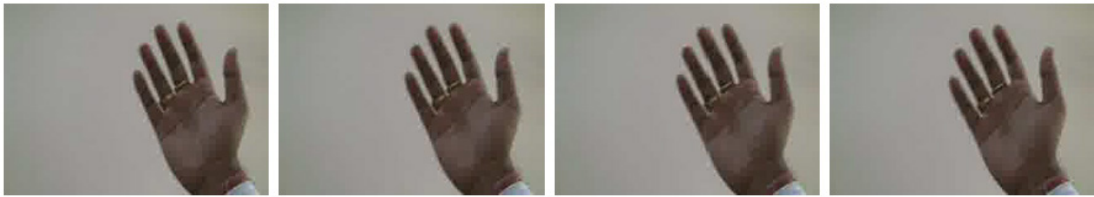**Figure 7(d).**   Final VOP obtained by proposed method

Figure 8(a).   Original frames of hand video sequence

Figure 8(b).   CDM obtained by modified Xiaofeng et al method

Figure 8(c).   CDM obtained by proposed inter- frame similarity based method

Figure 8(d).   Final VOP obtained by proposed method

5. Conclusions

In this paper, the problem of moving object detection under illumination variation has been addressed. The color based three frame difference method is implemented which is just a modification of the Xiaofeng method. Although it yields good result but the silhouettes present around the object. Thus the proposed method is used to detect the object perfectly. The inter-frame similarity is used to tackle the problem of illumination variation. Efficacy of the algorithms proposed is compared on the basis of misclassification error. The algorithm can further be modified to tackle the stop and go motion detection, shadow elimination problems. The thresholds used are all chosen manually which can be made adaptive to the video sequence to make the algorithm more efficient in different environment and situations.

REFERENCES

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

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

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

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

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

[6] Yu Wu; Delong Zeng; Hongbo Li, "Layered Video Objects Detection Based on LBP and Codebook," Education Technology and Computer Science, 2009. ETCS '09. First International Workshop on, vol.1, no., pp.207,213, 7-8 March 2009.

[7] M. Kim, J. choi, D. Kim and H. Lee, "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,Dec.1999.

[8] P. Spagnolo, T. Orazio, M. Leo, A. Distante, "Moving object segmentation by background subtraction and temporal analysis", Image and Vision Computing, Volume 24, Issue 5, pp. 411-423, May 2006.

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

[10] Lian Xiaofeng, Zhang Tao and Liu Zaiwen, "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, 2010.

[11] Xiaowei Zhou, Can Yang and Weichuan Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation", IEEE Transaction on Pattern Analysis and Machine Intelligence. Vol. PAMI-35 no.3, pp. 597-610, March 2013.

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

[13] J. F. Canny, "A computational approach to edge detection", IEEE Transaction on Pattern Analysis and Machine Intelligence. Vol. PAMI-6, pp. 679-698, Nov. 1986.