Moving Object Localization in Video Sequences under Static and Dynamic Background Conditions

Satrughan Kumar[1]*   **Jigyendra Sen Yadav[2]**   **Kumar Manoj[1]**   **Subramaniam Rajasekaran[1]**

[1]Madanapalle Institute of Technology & Science, Madanapalle, Andhra Pradesh, India
[2]National Institute of Technology Bhopal, Madhya Pradesh, India
* Corresponding author's Email: satrughankumar@gmail.com

Abstract: Background subtraction is one of the most reliable approach to localize the moving object under static camera arrangement. As seen, the moving object detection is a preliminary task in many vision applications such as video analysis, object tracking and activity analysis. However, the quasi-stationary pixels, aperture effect, ghost trail and varying illumination are still an annoying factors in the extraction procedures of the actual moving object in video. To alleviate the above problems, a video segmentation method is proposed that utilizes background subtraction and the 3-class fuzzy c-means clustering algorithm for extracting the relevant moving pixels. The proposed algorithm modifies learning parameters of adaptive filters to adapt the changes in the background. Afterwards, it incorporates the Markovian framework in which the initial motion field provides a prior information to regularize the segmentation process. The method achieves better visual and quantitative performance than other well-known background subtraction methods reported in this paper.

Keywords: Background modelling, Background subtraction, Binary labelling, Motion detection, 3 class fuzzy c-means clustering.

1. Introduction

The motion detection algorithms are coupled to many applications in diverse disciplines such as behavior recognition, video surveillance, remote sensing, border and shopping mall monitoring etc. The aim is to perceive non-stationary pixels that differ significantly from the stationary or quasi-stationary background pixels with respect to time[1-2]. Many object detection techniques incorporate object's colour, texture, shape in order to localize it in the scene. However, features extraction approach cannot be applied commonly to the objects in the videos. These features will change as the object changes its attributes. Moreover, when the object and background have similar colour and texture, the extraction procedures face difficulties. As seen, most of the surveillance camera are fixed at monitoring sites, therefore the background subtraction method can be the most suitable option for the detection of moving pixels under such condition. The background subtraction approach represents the object in the form of blobs which is more suitable representation than corner or edges. The background subtraction offers several advantages such as the computational complexity is less as compared to other motion detection methods and avoids prior information about the scene. Moreover, the background initialization and reconstruction are easy and effective to facilitate the initial estimate of motion vectors. But, in real-time application, the local motion and diversion of global parameters destabilize the reference background and lower the object detection capability [3 - 4]. The video scene can have some quasi-stationary pixels due to wavering of water, moving of tree or curtains etc. in the background.

Several state-of-the-art motion detection methods [5-7] such as, statistical method [8], optical flow [9], and temporal difference [10, 11] have proposed to recognize the moving pixels. Among those, the statistical method fails in handling

moving objects of large area and it is less resistant to hollow space and streak phenomenon. The optical flow is an excellent approach at low-level computations like direction, velocity, etc. But, in the real time scenario, optical flow is computationally complex and suffers due to aperture effect. The temporal difference method creates holes inside moving entities, affects aperture and causes serious problem in segmentation.

It can be observed that along with local and global disturbance, the aperture distortion, ghost trail and hollow space are also a great threat in motion detection task. The proposed work localizes the moving object in video sequences using the background subtraction method. The background initialization and updating scheme in this work remove significantly the false local motion and handle motion detection task in spatiotemporally varying nature of the scene [5, 12]. The proposed method relies on the integration of background modelling with Markovian framework [13] and 3-class fuzzy c-mean clustering algorithm [14, 15], which in turns provided a good contour with an acceptable segmentation of moving blob.

The other sections of the paper are organized as follows: Section 2 describes some effective background subtraction algorithms with their pros and cons. It also illustrates the fuzzy scheme to quantify the moving pixels in the scene. In section 3, the background modelling and motion detection approach are described along with binary labelling of foreground pixels. Experimental results are explained in section 4. Our concluding remarks are discussed in section 5.

2. Related work

This section provides a brief review of some popular motion detection methods based on the technique of their background modelling and updating scheme. In [8], authors suggest a simple background subtraction technique which does not include background maintenance scheme and depends on a predefined threshold. As a result, simple background subtraction is unreliable in motion detection task under varying illumination and dynamic scene.

Running average is an adaptive approach to background modelling that solved many problems in simple background subtraction [16]. It allows the adaptation to the temporal changes in scene by blending the incoming frame with reference background. However, it is not optimized due to its dependency on a fixed learning rate. In running average, the larger value of learning coefficient

misses most parts of the target, while the smaller learning coefficient puts down a trail behind the target. The trail generates due to incorrect foreground pixels. Though, running average needs less memory and time complexity, but it shows sluggish behavior against ghost image and aperture problem.

In simple statistical difference, moving pixels are separated out from video scene through the standard deviation of the difference image [8]. Though, it reacts quickly with change in illumination, but could not handle extraction task under the drastic illumination change.

In [17], author proposed a Σ-Δ (sigma-delta) background estimation that utilized difference image and time variance to compute the motion mask. The adapting rule is based on signum function that limits the changes in background model against the dynamic condition. The Σ-Δ method loses its lavishness to the constraints when either multiple objects are present in scene or moving objects exhibit different speed. In order to handle multiple objects, authors suggest a multiple background Σ-Δ estimation method in [11]. It adapts the registered background model by adjusting the weight and confidence coefficient through the variance.

In other different methods, a mixture of Gaussian kernel, mean-shift estimator and kernel density estimator are utilized for modelling the multimodal components in the unstable background. However, these methods impose trade-off between real time implementation and resolution [18 - 20]. In Gaussian mixture model (GMM) [18], background pixels are modelled by using a mixture of 'k' Gaussians distribution. GMM detects the foreground pixel by fitting it value into one of the modelled Gaussian distributions. The Gaussian parameters are updated for the matched pixels between current and background frame. The GMM method fails to handle the object in similar contrast region due to equal probability distribution between foreground and background pixels. The different variant and invariant states of background modelling schemes are also represented by Markov random field (MRF) at the cost of time and memory complexity to deal with the sudden or gradual illumination [21, 22].

In [6], edges extracted via second derivative in gradient direction are utilized to acquire the spatial consistency. However, this method updates the background model according to traditional background updating scheme. The second-order derivative requires high time during its computation.

In [24], authors registered the background pixels by collecting the blocks of background

candidate with suitable probability density function. It used here weighted variance and mean to threshold out the moving object. The method faced complexity in detection and lost the true positives pixels near those regions where the variance between foreground and background were equal. It applied traditional adaptive filter for the updating of background model.

The background model constructed in [4], utilized the circular shift method on the neighbourhood of each pixel. It evaluated the moving object by utilizing the background subtraction and graph cut techniques. Since, the modelling of background pixels relies on neighbourhood size of region, therefore the model has high computation complexity.

In [12], author utilized a pixel-wise non-parametric background subtraction technique is proposed to detect the object under spatio-temporal background. The background updating scheme is based on the random selection of pixel from the neighbour of previous detected low efficacy region. The decision between high and low efficacy region has been done empirically. Therefore, the region which have low variation in intensity level of background and foreground pixel can give erroneous true positive and false positive pixels.

A block based method is discussed in [19] where the intensity value of the registered pixels in the background model is verified through modified KDE (Kernel density estimation) rule. The centre of bin and the mean of the block are adopted to extract the moving blob in the scene. The method has time and memory complexity. It also suffered due to unresolved camouflage problem and low frame processing rate.

In [15], authors used temporal information and adopted a post-processing scheme based on local fuzzy thresholding for the detection of moving pixels. The performance degraded when encountered with poor contrast between background and foreground. Fuzzy c-means (FCM) clustering is an effective technique to detect the value of threshold due to its convergence and low complexity. In FCM, the threshold is computed with the help of the maximum with the smallest centre and the minimum with the middle centre. It proved effective over OSTU's thresholding method during the foreground and background separation [14, 15].

In another block and region based approach, the direction of pedestrian is estimated on poor resolution video. It used hidden Markov model to handle partial occlusion and segmentation error by employing intra and inter fusion between the frames [25]. For the betterment of results, the method [25]

requires few iteration of entire process which in turns increases the time complexity of the algorithm.

An automatic approach is proposed in [5], which detects the target on the foreground through alarm trigger module. The method updates background pixels traditionally and it needs little computational cost due to block-wise entropy evaluation.

The existing literatures reveal that the regional as well as pixel-wise computations are necessary to provide the spatial consistency on the foreground mask. In addition to that, handling the learning rate is a primary concern in adaptive background model. As seen, the fixed and low learning rate in background updating scheme can produce multiple copy of the moving object on foreground, while the fixed and high learning rate can result hole inside the moving object. Therefore, the learning rate should be low for the pixels which show temporally less variation in intensity and vice versa. Moreover, a suitable thresholding technique is the primary requirement for reliable motion mask and regularization in the segmentation process.

3. Proposed method

This section describes the proposed motion detection technique under static camera arrangement. The key contributions of this work are described as follows:

(a) This method utilizes varying learning rate which provide better updating in the background model under dynamic changes.

(b) This work uses Fuzzy c- means clustering to generate an initial motion field which in turn helped in the final detection of actual moving blob.

(c) The Markovian framework maintain optimum shape of the moving object. Therefore, results can be used for higher level computer vision task.

(d) The output of this method is free from artificial trail, aperture effect, background cluttered and over segmentation error.

Initially, the framework illustrates a background model and its updating module, afterwards it explains the object extraction module.

3.1 Background model generation and updating phase

Initially, a reference background is computed using modified moving average of 'M' frames of video. These 'M' frames do not contain moving objects [5]. The modified moving average reduces memory consumption in the process. Therefore, the initial background is computed as follows:

$$B_t(x,y) = B_{t-1}(x,y) + \frac{1}{t}(I_t(x,y) - B_{t-1}(x,y)) \tag{1}$$

Where, 't' is the number of frames from 1 to M and M is set to 50. '$B_t(x,y)$' contains pixels of current background frame, '$B_{t-1}(x,y)$' is past background, while current frame is represented by '$I_t(x,y)$'.

Instead of updating the background model each instant of time, the proposed method updates it after five frames. To do so, the incoming video frames are divided into a group of small segment. Each segment composes of 'r' number of successive frames of a video sequence. The value of 'r' is taken as '5' in this experiment. Initially, the algorithm retrieves two consecutive (i-1)th and ith segments simultaneously. The first two group should represent background frames that should not have any moving object. After that, the mean of consecutive segments (i-1)th and ith are computed as follows:

$$\mu^{i-1} = \left(\frac{1}{r}\right)\sum_{t=0}^{r-1} I_{t-r}^{i-1}(x,y) \tag{2}$$

In a similar manner, the mean is computed for ith subset in the given equation as follows:

$$\mu^{i} = \left(\frac{1}{r}\right)\sum_{t=0}^{r-1} I_{t+r}^{i}(x,y) \tag{3}$$

Now the local background '$B^{i-1}(x,y)$' for (i-1)th (first segment) is estimated as follows:

$$\Omega^{i-1}(x,y) = (\mu^{i}(x,y) - \mu^{i-1}(x,y) \tag{4}$$

$$B^{i-1}(x,y) = \begin{cases} if\ |\Omega^{i-1}(x,y)| \geq \lambda \\ B_t(x,y) + sgn(\Omega^{i-1}(x,y) \\ else \\ B^{t}(x,y) \end{cases} \tag{5}$$

'λ' is set to 0.0001 in this experiment.

The process is executed for all successive pair (i-1)th and ith segments for entire video sequence. The background $B^{i-1}(x,y)$ generated through Eq. (5) is subtracted from the current frame and an initial motion field is estimated. The background $B^{i-1}(x,y)$ is updated through running average inside the (i-1)th subset, but with varying learning coefficient. The local background $B^{i-1}(x,y)$ is updated as:

$$B_t^{i-1}(x,y) = \left(1 - \frac{\beta}{j}\right) * B^{i-1}(x,y) + \frac{\beta}{j} * (K_1) \tag{6}$$

The variable $K_1 = (I_t^{i-1}(x,y) - B^{i-1}(x,y))$ is substituted in Eq. (6). Whereas 'j' is varied from 1 to r. The value of β is 0.03 for static sequence, while 0.06 for

dynamic and illumination case. The variable learning coefficient reduces temporal noise and adapts intensity variation along the temporal line. Moreover, the reference background '$B_t(x,y)$' is updated as follows:

$$B_t(x,y) = (1-\alpha)B_t(x,y) + \alpha\left(B_r^{i-1}(x,y) - B_t(x,y)\right) \tag{7}$$

Where '$B_r^{i-1}(x,y)$' is the rth updated background by Eq. (6). The value of 'α' is 0.05 for static and 0.005 for a dynamic sequence in this experiment. The initial motion field '$D^{i-1}(x,y)$' is computed by subtracting the 'r' frames of first subset from the local background '$B^{i-1}(x,y)$' as follows:

$$D^{i-1}(x,y) = (I_t^{i-1}(x,y) - B_t^{i-1}(x,y)) \tag{8}$$

3.2 Object extraction module

Since, the initial motion field may contain some erroneous false moving and noisy pixels. Therefore, the restoration of actual pixels of the target is done under a Markovian framework. As observed, the neighboring pixels provide spatial consistency and are highly correlated to each other [13, 22, 23]. The framework assumes that neighboring pixels tend to have equal intensity in an image and the intensity may distort independently with some probability. By the choice of neighboring elements and smoothing parameter, a suitable estimation for labeling the underlying scene (true image) is optimized using iterated conditional mode [13, 23].

For each pixel of '$D^{i-1}(x,y)$' we associate a set of 4-connected first order neighborhood pixels which shares a side with the given central pixel. The labels obtained through the initial motion field have been considered as Gibbs prior associated within Ishin model under a Markovian framework [13, 22, 23]. A fixed provisional known or a prior 'J_u' or '$D^{i-1}(x,y)$' to re-estimate the current estimated value '$\hat{J}_u$' at pixel $u \rightarrow (x,y)$ in order to get true image J* . The true image is defined here as the actual moving object on foreground i,e . '$D(x,y)$'. This estimation process is accomplished by maximizing the given argument, i.e.

$$argmaxP\left(\frac{J_u}{O_u}, \hat{J}_{s\setminus u}\right) \propto f\left(\frac{O_u}{J_u}\right)P_u\left(\frac{J_u}{\hat{J}_{\delta u}}\right) \tag{9}$$

Where, 's/u' consists of all neighbours of the pixel at 'u'. The ∂u are those neighbouring pixels at 'u' that is defined by a first order neighbourhood system. 'O_u' represents the observed moving pixel at 'u'.

The term 'U(J)' stands for potential associated with neighbourhood configuration that follows the Gibbs sampling with the Ishin model [13, 23], which is given as:

$$U(J) = -\alpha_1 \sum_u J_u - \beta_1 \sum_u \tilde{v}(J_u) \qquad (11)$$

'α_1' controls the biasing of pixels. In this experiment, the value of α_1 is set to '0'. The '$\tilde{v}(J_u)$' is all the neighbours of 'u' which have label 'J_u'. 'β_1' is experimentally set to 0.0001. In the above Eq.11,the term 'β_1' is added for every neighbour in the current estimation that doesn't match '$\tilde{v}(J_u)$'. If the difference image consists of two labels 'l_1' and 'l_2',which implies $l_1, l_2 \epsilon J_u$.

Since, the initial estimate of actual moving pixels is an important factor in the Markovian framework, therefore the 3-class fuzzy c-means clustering algorithm provided the better choice for the gathering this information. The threshold selection using 3-class fuzzy c-mean clustering technique caters well in intensity variation and also is robust to noise and interclass dependencies [14]. It is assumed that the initial motion field consists of three categories of intensity. The first category belongs to moving pixels, the second category contains the pixels that come out due to quasi-stationary behavior of background, and the third class belongs to the pixels those have equal intensity between current frame and background frame. Therefore, the 3-class fuzzy c-mean clustering is applied here to threshold out the initial estimate of correct moving pixels in the scene. The 3-class fuzzy c-means clustering is applied on the difference image that resulted three clusters with the three centers. The first threshold 'T1' is calculated by averaging the data belonging to the large center and the middle center, while the second value 'T2' is calculated by averaging the data of middle center and small center value [14, 15]. Using the value of T1 and T2, the initial motion vector is estimated. The binary motion mask is evaluated as follows:

$$D(x,y) = \begin{cases} 1 \ if \ D^{i-1}\left(p\left(\frac{O_u}{J_u=l1}\right) > D^{i-1}(p(\frac{O_u}{J_u=l2})\right) \\ -1 \ if \ D^{i-1}(p\left(\frac{O_u}{J_u=l1}\right) < D^{i-1}(p(\frac{O_u}{J_u=l2})) \end{cases} \quad (11)$$

A morphological [11] is carried out on 'D(x,y)' to remove unwanted pixels from the foreground image. The morphological process is defined by:

$$g(x,y) = (D(x,y) \ominus SE) \oplus SE \qquad (12)$$

Where 'SE' represents the structuring element, Θ and $\oplus$ perform erosion and dilation respectively.

4. Results and analysis

This section explains the experimental results, which are achieved on the video sequences available in Change detection and I2R datasets. The databases contain static and dynamic video sequences.

In order to judge the authenticity of the proposed background model, the variation of intensity near static and dynamic region in the consecutive frames of video sequences are observed. To do so, the intensity of MR video sequence at the location '$I_t(10,20)$' and '$I_t(57,68)$' is observed. The MR sequence is taken from I2R dataset, in which the background of the frames is non-stationary due to moving curtain. The intensity at '$I_t(10,20)$' is stationary throughout the MR sequence, while the intensity at '$I_t(57,68)$' trembles due to the moving curtain as well as the appearance of a moving person. The actual intensity at '$I_t(10,20)$' and '$I_t(57,68)$' are shown in Fig. 1.

However, Fig. 2 shows that the proposed method generates a stable signal against the background intensity variations which is necessary to distinguish moving objects from background clutter. The moving object mask is shown in Fig. 3. Fig. 3 shows the vriation of intensity in difference image. It can be observed that the intensity level between the background and object is sufficiently deviated and can be easily percieved.

Furthermore, the qualitative performance of this method on the varieties of real time video sequences is shown in Figs. 4 - 8. These sequences include spatio-temporal changes in the background, such as rippling water, moving curtains, and abandoned objects.
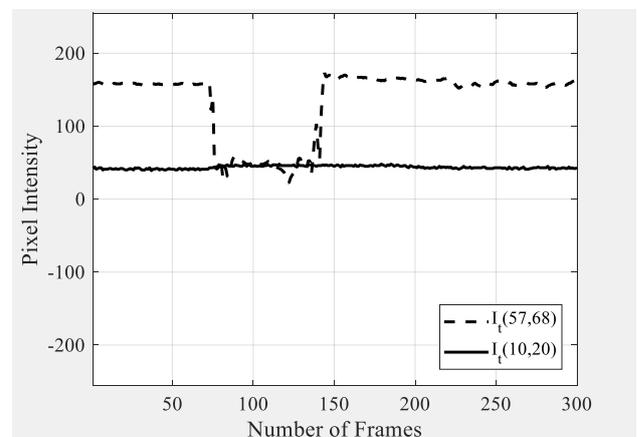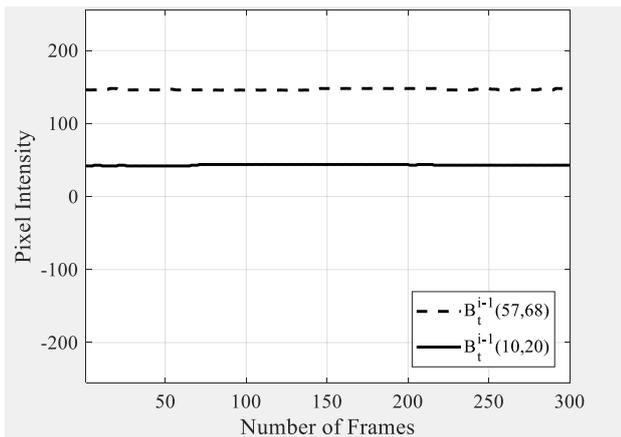

Figure. 1 Actual intensity plot of $I_t(x,y)$
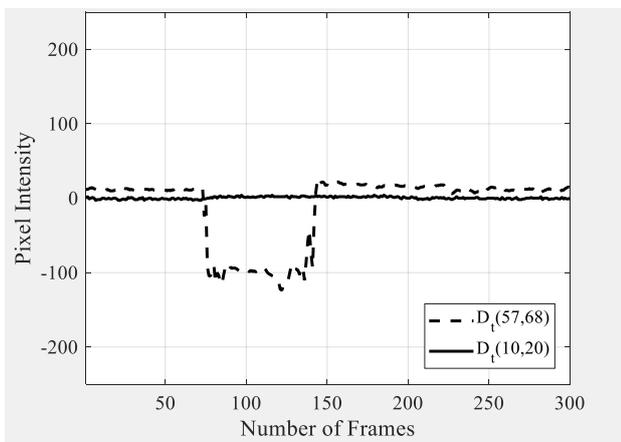
Figure. 2 Stable signal of modeled background

Figure. 3 Pixel intensity at initial motion field

As, the primary objective of this work is to project the moving object without the error due to over-segmentation, ghost image,aperture effect and hollow space [20]. The aim is also to detect only the true-positive pixels in oder to get the exact countour as well as silhouette and hanlding the object extraction procedure during the slow and stationary behaviour of moving object. Over-segmentation error arises due to the false-negative pixels on the foreground, however the aperture effect error arises because of the mis-match of the actual correspondence of object's location between consecutive frames. The hollow space cause loss of actual moving pixels inside the object due to lack of optimum learning mechanism of background, however ghost image is a false copy of image that exists during the object extraction procedure and it decays slowly with time. The annoying ghost arises because of lack of integration of spatial feature and optimum learning rate.

In Figs. 4 and 5, the videos are taken from change detection dataset. It can be seen that the proposed method distinguishes the moving pixels from the background satisfactorily. It detects the person on the foreground even though the person

became stationary in frame. However, method [11, 12, 18], and [26] lack of spatial feature integration and the property of neighbourhood, therefore one can observe in Fig. 8 that these methods fail in detection of the object under such sleeping condition. At every time step, the constant updating of background model in [11] and [17] causes hollow space in moving object which is shown in Fig. 8. In method [12], the background is updated near low efficacy region with the help of random selection of neigbourhood pixels near target pixel. However, in this perposed method, the four neighbouring pixels have been considered to verify that the target pixel is either a foreground or background pixel. Moreover, the proposed work also utlizes the clustering method to provide a prior information to the underlying image to achieve the actual background and foreground pixel. Therefore, proposed work achives better visulization of foreground image in video sequences.

The moving object detected through proposed algorithm is free from aperture distortion, artificial ghost trail and work well in medium changing illumination condition. The entire shape of the object is detected and the output can be used in higher level computer vision application. In Fig. 6 and Fig. 7, the pixels in background are in dynamic mode due to rippling of water and moving of curtain. As shown in Fig. 6, proposed method precisely detects the area below the knee of the person where the person tends to camouflage with the vegetation of background. The moving curtain, and color illusion between person and curtain in Fig. 7 (camouflage effect) can cause difficulties to detect the moving object in 'MR' sequence. Even under such circumstances, the proposed method excludes the impulsive noise produced by the strong edges of moving curtain. As shown in Figs. 6 and 7, the proposed background modelling scheme adapts better to the background model against enviornmental changes and differentiates the moving object from the background.

Fig. 8 shows visual comparision of proposed method with the other exisiting background subtarction method on some challenging video sequences. The experimental results show that the objects on the foreground mask are free from artificial ghost trails and aperture distortion. The proposed backgrond updating scheme adapts the spatiotemporal changes and provides a satisfactory result against moderate illumination changes and varying speed of person. The proposed background updating schemes adapts background model quickly in prespecified time or segment. However, handling the detection procedure in next segment, the recent

status of background is blended with the refrence background which inturn helps to detect object in sleep mode condition.

	F-339	F-579	F-686	F-1035
Sampled frames				
Ground truths				
Proposed method results				
Similarity	0.7700	0.8117	0.8488	0.7990
F1	0.8696	0.8961	0.9182	0.8515

Figure. 4 Output of 'PEDESTRIAN' sequences with similarity and F1

	F-599	F-674	F-766	F-1255
Sampled frames				
Ground truths				
Proposed method results				
Similarity	0.8276	0.8057	0.8175	0.7630
F1	0.9057	0.8924	0.8996	0.8656

Figure. 5 Output of 'OFFICE' sequences with similarity and F1

	F-1499	F-1616	F-1621	F-1624
Sampled frames				
Ground truths				
Proposed method				
Similarity	0.8489	0.9375	0.9050	0.8878
F1	0.9183	0.9678	0.9501	0.9406
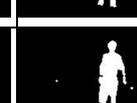
Figure. 6 Motion mask of ' WS' sequence with similarity and F1

	F-22774	F-22847	F-23857	F-23893
Sampled frames				
Ground truths				
Proposed method				
Similarity	0.9268	0.8895	0.8768	0.8528
F1	0.9620	0.9417	0.9344	0.9205

Figure. 7 Motion mask of ' MR" sequence with similarity and F1

	PEDESTRIAN	FOUNTAIN	CANOE	PET2006	TRAFFIC	SOFA	OFFICE
Sampled Frame							
Ground Truth							
SDE[17]							
MSDE[11]							
GMM[18]							
METHOD[5]							
METHOD[26]							
METHOD[12]							
PROPOSED METHOD							

Figure. 8 Performance comparison of moving mask between proposed method and other existing methods

So, the proposed learning mechanism and updating rule make this work more efficient that that proposed in [12].

The quantitative analysis of the results is evaluated in terms of several metrics, including *Recall Precision, Similarity* and *F1* [5] . The *Recall*

represents the relevant true-positive pixels on the foreground mask. *Recall* is defined as follows:

$$R=Recall=tp/(tp+fn) \qquad (13)$$

Where *tp* and *fp* belong to the true positive and

153

Table 1. Performance comparisons of accuracy metrics between proposed method and other methods.

Sequences	Evaluation parameters	Proposed Method	GMM [18]	MSDE [11]	SDE [17]	Method [26]	Method [12]	Method [5]
Office	S	0.79	0.48	0.21	0.15	0.18	0.60	0.65
	F1	0.88	0.65	0.33	0.25	0.29	0.74	0.79
	P	0.91	0.84	0.22	0.15	0.28	0.85	0.78
	R	0.86	0.53	0.83	0.83	0.39	0.66	0.80
Pedestrian	S	0.83	0.29	0.16	0.13	0.80	0.70	0.85
	F1	0.92	0.34	0.28	0.23	0.89	0.82	0.92
	P	0.93	0.62	0.13	0.13	0.94	0.83	0.93
	R	0.91	0.31	0.77	0.75	0.84	0.81	0.92
WS	S	0.88	0.53	0.54	0.35	0.72	0.59	0.76
	F1	0.93	0.69	0.69	0.51	0.83	0.74	0.86
	P	0.94	0.85	0.86	0.74	0.79	0.81	0.88
	R	0.92	0.62	0.59	0.40	0.87	0.68	0.86
MR	S	0.82	0.65	0.51	0.33	0.45	0.57	0.80
	F1	0.90	0.76	0.66	0.53	0.61	0.73	0.89
	P	0.94	0.95	0.66	0.48	0.71	0.84	0.86
	R	0.86	0.67	0.70	0.62	0.56	0.64	0.92

* S- Similarity, *P-Precision, *R-Recall

false-positive pixels respectively. The '*fn*' indicates the false-negative pixels. The precision give irrelevant true positive pixels and is given by:

$$P=Precision=tp/(tp+fp) \qquad (14)$$

Where *(tp+fp)* represents the total number of true-positive pixels on the foreground mask. The *F1* is given as:

$$F1=2×precison×Recall/(Recall+Precsion) \quad (15)$$

The parameters *similarity* and *F1* provide an adequate decision on the accuracy of the quantitative measurements of motion mask. The parameter similarity is given as:

$$S=Similarity=tp/(tp+fp+fn) \qquad (16)$$

The range of accuracy metrics lies between '0' to '1', where a higher value indicates better performance. The comparison results between proposed method and other existing methods (GMM [18], MSDE [11], SDE[17], method [5], method[12], method [26]) are given in Table 1. The quantitative analysis reveals that the proposed method yields higher *F1* and *similarity* values as compared to other background subtraction methods for all sequences. All metrics have a value above 82% for WS and MR sequences, which consist of local motion in the background.

The *Similarity* and F1 (Pedestrian sequence) accuracy rates achieved for the GMM method are up to 54% and 58% respectively lower than those obtained by this proposed method.

In comparison, the proposed method attains better *Similarity* and *F1* accuracy rates than method used in [26] and [12]. In [12], the random neighbouring pixel selected for the background update is good to adapt the recent change in scene but whenever the object will become stationary, the approach used in [12] will create a hollow space inside the target. Therefore, the large number of true positve pixels can be missed while in detection procedure and actual shape of the traget may devaite from its original one. However, proposed method first provides a prior information to the lattice to the target pixel through 3-class fuzzy c-mean clustering, then it applied the final classification between forground and background based on the probablistic decision. The *Similarity* and *F1* metrics obtained from the SDE method are up to 70% and 69% lower than those obtained by this proposed method respectively on 'Office' video sequence. The *Similarity* and *F1* obtained from the method [5] are up to12% and 7% lower than those obtained by this proposed method respectively for 'WS' sequence . The proposed method.

We have simulated all these algorithms using Matlab 7.1 on a desktop computer with configuration 3.2 GHz Intel CPU, 2GB RAM. The computational time complexity is computed by

summing the sole computation time and acquisition time. To process a 128×160 frame, proposed method needs 0.11s, GMM algorithm takes 0.48s, while MSDE, SDE, and method [5] need approximately 0.025 second.

5. Conclusion

This paper presents an efficient background model which utilized the concepts to adapt the changes through varying learning rate and exploited the spatio-temporal and statistical feature to minimize the background clutter. The proposed approach has been validated on static and dynamic background video sequences. The experimental results show that proposed method outperform some of the existing schemes in terms of accuracy, over-segmentation error, aperture effect, ghost error as well as computational time. A combination of Markovian framework and 3-class fuzzy c-means clustering gives promising results with a minimum iteration. In short, the information extracted by this level of segmentation is adequate for motion detection and in the various vision applications. The new work can be extended in future to solve the problem associated with complex scene, including a group of people and extravagant dynamics in background.

References

[1] R. J. Radke, S. Andra, O. Al-kofahi, and B. Roysam, "Image Change Detection Algorithms : A Systematic Survey", *IEEE Transactions on Image Processing*, Vol. 14, No. 3, pp. 294–307, 2005.

[2] C. H. Lim, E. Vats, C. S. Chan, E. Vats, and C. S. Chan, "Fuzzy human motion analysis : A review", *Pattern Recognition*, Vol. 48, No. 5, pp. 1773–1796, 2014.

[3] L. Vosters, C. Shan, and T. Gritti, "Real-time robust background subtraction under rapidly changing illumination conditions", *Image and Vision Computing*, Vol. 30, No. 12, pp. 1004–1015, 2012.

[4] J. Dou, Q. Qin, and Z. Tu, "Background subtraction based on circulant matrix", *Signal, Image and Video Processing*, Vol. 11, No. 3, pp. 407–414, 2017.

[5] S. Huang, "An Advanced Motion Detection Algorithm with Video Quality Analysis for Video", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 21, No. 1, pp. 1–14, 2011.

[6] F. Yasmin, A. Rahman, A. Hussain, W. Mimi, D. Wan, H. B. Zaman, and N. Tahir, "Enhancement of Background Subtraction Techniques Using a Second Derivative in Gradient Direction Filter", *Journal of Electrical and Computer Engineering*, Vol. 2013, No. 1, pp. 1–12, 2013.

[7] G. Xue, J. Sun, and L. Song, "Background subtraction based on phase feature and distance transform", *Pattern Recognition Letters*, Vol. 33, No. 12, pp. 1601–1613, 2012.

[8] M. Oral and U. Deniz, "Centre of mass model – A novel approach to background modelling for segmentation of moving objects", *Image and Vision Computing*, Vol. 25, No. 8, pp. 1365–1376, 2007.

[9] T. K. Bruce and D. Lucas, "An iterative image registration technique with an application to stereo vision", In: *Proc. of International Joint Conference on Artificial Intelligence*, Vancouver, pp. 674–679 .1981.

[10] P. Spagnolo, T. D. Orazio, M. Leo, and A. Distante, "Moving object segmentation by background subtraction and temporal analysis", *Image and Vision Computing*, Vol. 24, No. 5, pp. 411–423, 2006.

[11] A. Manzanera and J. C. Richefeu, "A new motion detection algorithm based on Σ-Δ background estimation", *Pattern Recognition Letters*, Vol. 28, No. 3, pp. 320–328, 2007.

[12] Y. Yang, Q. Zhang, P. Wang, X. Hu, and N. Wu, "Moving Object Detection for Dynamic Background Scenes Based on Spatiotemporal Model", *Advances in Multimedia*, Vol. 2017, No. 1, pp. 1–9, 2017.

[13] J. Besag, "On the Statistical Analysis of Dirty Pictures", *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 48, No. 3, pp. 259–302, 1986.

[14] G. Xiong, X. Zhou, and L. Ji, "Automated Segmentation of Drosophila RNAi Fluorescence Cellular Images Using Deformable Models", *IEEE Transaction on Circuits and Systems*, Vol. 53, No. 11, pp. 2415–2424, 2006.

[15] N. Paul, A. Singh, A. Midya, P. Pratim, R. Debi, and P. Dogra, "Moving object detecti -on using modified temporal differencing and local fuzzy thresholding", *The Journal of Supercomputing*, Vol. 73, No. 3, pp. 1120–1139, 2016.

[16] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, and A. B. Stract, "Pfinder : Real-Time Tracking of the Human Body", *IEEE Transaction on Pattern Analysis and Mach- ine Intelligent*, Vol. 19, No. 7, pp. 780–785, 1997.

[17] A. Manzanera, J. Richefeu, A. Manzanera, and J. R. A, "A robust and computationally efficient motion detection algorithm based on sigma-delta background estimation", In: *Proc. of the*

Fourth Indian Conference on Computer Vision, Graphics & Image Processing, pp. 46–51, 2004.

[18] C. Stauffer and W. E. L. Grimson, "Learning Patterns of Activity Using Real-Time Tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 747–757, 2000.

[19] M. Piccardi, "Background subtraction techniques : a review", In: *Proc. of IEEE International Conference on Systems, Man and Cybernetics Background*, pp. 3099–3104,2004.

[20] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting Moving Objects, Ghosts , and Shadows in Video Streams", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 10, pp. 1337–1342, 2003.

[21] S. Huang, L. Fu, and P. Hsiao, "A Region-Level Motion-Based Background Modeling and Subtraction Using MRFs", In: *Proc. of the 2005 IEEE International Conference on Robotics and Automation*, pp. 79–84, 2005.

[22] V. Reddy, C. Sanderson, and B. C. Lovell, "A Low-Complexity Algorithm for Static Background Estimation from Cluttered Image Sequences in Surveillance Contexts", *EURASIP Journal on Image and Video Processing*, Vol. 2011, No. 1, pp. 1–14, 2011.

[23] S. Halim, "Modified Ising Model for Generating Binary Images", *Jurnal Informatika*, Vol. 8, No. 2, pp. 115–118, 2007.

[24] O. Elharrouss, A. Abbad, D. Moujahid, J. Riffi, and H. Tairi, "A block-based back- ground model for moving object detection", *Electronic Letters on Computer Vision and Image Analysis*, Vol. 15, No. 3, pp. 17–31, 2017.

[25] R. Raman, P. K. Sa, S. Bakshi, and B. Majhi, "Kinesiology-inspired estimation of pedest - rian walk direction for", *Future Generation Computer Systems*, *In Press*, 2017.

[26] T. Bouwmans and X. Zhang, "A Fuzzy Background Modeling Approach for Motion Detection in Dynamic Backgrounds", In: *Proc. International Conference on Multi-media and Signal Processing*, pp. 177–185, 2012.

[27] M. F. Savas and B. Erkal, "Moving Object Detection Using an Adaptive Background Modeling in Dynamic Scene", *European Journal of Engineering and Natural Sciences*, Vol. 2, No. 1, pp. 161–166, 2017.