Segmentation of Moving Objects using Background Subtraction Method in Complex Environments

Satrughan KUMAR, Jigyendra Sen YADAV

Department of Electronics and Communication, MANIT, Bhopal (462003), India

satrughankumar@gmail.com, jsyadav74@gmail.com

Manuscript received June 26, 2015

Abstract. *Background subtraction is an extensively used approach to localize the moving object in a video sequence. However, detecting an object under the spatiotemporal behavior of background such as rippling of water, moving curtain and illumination change or low resolution is not a straightforward task. To deal with the above-mentioned problem, we address a background maintenance scheme based on the updating of background pixels by estimating the current spatial variance along the temporal line. The work is focused to immune the variation of local motion in the background. Finally, the most suitable label assignment to the motion field is estimated and optimized by using iterated conditional mode (ICM) under a Markovian frame- work. Performance evaluation and comparisons with the other well-known background subtraction methods show that the proposed method is unaffected by the problem of aperture distortion, ghost image, and high frequency noise.*

Keywords

Background subtraction, background modeling, initial motion field, morphology

1. Introduction

Moving object segmentation in video frames is the most significant step in many computer vision applications including human activity analysis, traffic monitoring, and video surveillance [1]. However, the complexities to identify suspicious activities of people at social places and endangered object at shopping interacts, airports, banks, have become a matter of concern and motivated others toward the development of precise and robust surveillance systems [2].

In short, motion detection is a way to determine the magnitude of point or group of points in two or more consecutive images of a video sequence, which are non-stationary. In compulsion, object segmentation and motion perception in a video frames are a prerequisite of many post-processing steps such as target classification, behavior recognition, monitoring [3], [4]. Some of the existing methods for motion detection are optical flow [5], frame difference [6], statistical method [7] and background subtraction [8–11]. The frame difference method is robust and has a strong adaptability in varying environment along with less computation time and complexity. However, it creates holes inside the target due to incomplete generation of relevant pixel on the fore-ground mask. Optical flow is a reliable approach for local motion speculation, but it demands hardware in real time putting into use. On the other hand, background subtraction is a simplistic way to localize the target in a scene without the any prior information about the scene. Although, the background subtraction method is inexpensive with respect to memory requirement and computational time, yet it faces a few difficulties to contend with accuracy under spatiotemporally behavior of the background object.

Traditional background subtraction schemes such as AMF (Approximated Median Filter) [12], Kalman filter [13], and single Gaussian filter [14] reflect some irrelevant pixels on the foreground due to lack of correlation between the spatial and temporal constraints in their background maintenance schemes. Nevertheless, adjusting the learning rate to background pixels is another potential problem in background maintenance [15]. The adaptive algorithms based on fast learning rate quickly absorb the environmental noise and contravene the generations of entire relevant pixels of the target. However, the algorithms based on low learning rate are less robust against a slow moving object and show the ways to generate multiple images or ghost on the foreground image [16]. Furthermore, these algorithms do not integrate any data validation techniques that exploit the inter-pixel relationship to reduce the misclassification on the foreground mask.

In this paper, we focus to enhance the robustness of the background subtraction method under static and dynamic conditions of background [17]. Initially, the spatial and temporal constraints are mapped to exclude the impulsive effect of the registered background model. A region level processing is conducted to assign the proper label to the moving object on the foreground image. The rest of the paper is organized as follows: Section 2 presents the essential related work concerning background modeling and

foreground validation. Section 3 explains the proposed background model and foreground validation scheme. Experimental results are explored in Section 4. Concluding remarks are given in Section 5.

2. Related Work

In this section, we present the overview of some of the well-known background subtraction methods together with their updating scheme and foreground validation task. Background subtraction methods differ in the procedure employed to update the reference background during the motion detection task.

In [14], author suggested a running average method (RA) using a first order recursive filter to update the background model by integrating the new incoming frame. Even though, it is adaptive and requires less memory, but sensitive to ghost effect and environmental noise. It is heavily dependent on fixed learning rate and its effect due to either fast or slow learning rate is discussed in the introductory part of this paper. Temporal difference is very robust to environmental noise, but it does not generate the entire relevant pixels on the foreground [18]. In [19], authors suggest to integrate the edges extracted through SDGD filter in order to complement the missing pixels on the foreground. However, background image updates according to traditional scheme and it requires high computational time due to the characteristic of second-order derivatives.

A simple statistical difference (SSD) model is proposed in [7]. It allows adaptation to the environment changes in background model through natural variation in the current frame and less dependent on a threshold value. An absolute difference image is computed by subtracting the current frame from the mean of background frame in order to achieve the segmented region on the foreground. However, it is sensitive to environmental noise due to lightening and initial start-up time. In [20], author proposed a Σ-Δ filter (SDE) to estimate the temporal statistics for each pixel of image. However, it adapts always the temporal changes in the background model by either increasing or decreasing its pixel intensity to unity. The adaptation criterion is independent upon the difference image. Through the comparison between time variance and difference image, it detects the foreground pixel. As Σ-Δ method responds to signals with absolute time variance less than unity, which is insufficient to detect multiple objects.

Further enhancement in Σ-Δ filter is suggested in multiple Σ-Δ method (MSDE) [11]. It computes a set of 'k' backgrounds instead of a single background. Each background has its weight and confidence coefficient, which vary according to the time variance. It estimates each pixel in a background model by taking the value from a set of 'k' backgrounds. Then, it compares each pixel of background model to the current input pixel in order to determine the foreground mask. An automatic motion detection algorithm is proposed in [8], which detects the moving object using alarm trigger module. However, it adapts the environmental changes in background model using traditional approach and it requires computational cost due to alarm trigger chain.

In [21], authors suggest to model each pixel of background using mixtures of Gaussians to deal the complex scene. In order to detect the foreground pixel, it compares each input pixel to each Gaussian distribution. The associated kernels of matched pixels are updated in background model. However, (Gaussian mixture model) GMM requires computational cost due to handling the associated kernels with it. It also fails to handle the foreground and background pixels, those have identical probability distribution function (camouflage effect). It is also less effective against aperture distortion. In [22], authors propose to train the background model with two mixture of Gaussian model where Gaussian kernels have identical parameters but different learning rate. Moving pixels are classified with the help of a finite-state machine. However, in case of slow motion, an operator could interactively maintain the system and estimated prior required for the input in a finite state machine. In [23], a prior knowledge, which includes spatial and temporal coherence, is fused with the cues provided by background subtraction scheme through MRF framework. Although the MRF based scheme achieves better segmentation, yet it is not applicable in real time operation and larger displacement in object motion.

A review of various background subtraction methods and their updating schemes has been discussed in [24]. This review studies categorize the background representation frameworks into basic models, statistical parameters model, cluster and sparse models, artificial neural network models and fuzzy models. In our proposed work, we confine our study to basic and statistical model that provides sufficient numerical foundation to the projected background maintenance scheme. Since, the basic models use pixel-level processing that can update each pixel of initially registered background independently without any prior knowledge or cluster observation of pixels. On the other hand, the statistical modeling also proffers robustness against background motion and illumination. In [25], authors proposed fuzzy based approach focused on these two models in order to handle the dynamic background. It uses spatial and temporal constraints to enhance the performance. However, it could not handle the object when it became stationary in the scene due to traditional background maintenance scheme.

Another factor that affects the performance of foreground detection is the integration of region level processing. In [8], author proposed to integrate region level processing by evaluating block-wise entropy to estimate the initial motion field, but this affected the shape and lost the significant part of object near low entropy region. However, in our method, a regional level processing is included in data validation techniques to avoid the miss-classification between stationary and non-stationary pixels

on the fore- ground. Nevertheless, some feature and subspace learning based background modeling schemes well handled the complex videos, but at the cost of higher time and memory complexities [26]. The analysis of the existing background subtraction methods reveal that the simple schemes suffer to generate reliable moving mask on the foreground, while the complex schemes handle it at higher computational cost and complexities. In addition to that, a robust background-modeling scheme must include the temporal and spatial constraints in order to get a reliable motion mask. A regional level processing should be included in data validation techniques to avoid the misclassification between stationary and non-stationary pixels on foreground.

3. Proposed Method

In this paper, the proposed framework establishes to diminish the complexities of the background modeling for the moving object detection under static camera arrangement. It is noted that video sequences in our experiment show spatio-temporally varying behavior due to rippling water, moving curtain, changing illumination and many more. The proposed method comprises of two stages in order to alleviate these problems. The first stage provides a suitable background model followed by an updating scheme. In the second stage, a region level processing is carried out based on the assumptions that neighboring pixels tend to possess identical property and each pixel may be affected independently in an image. As a result, a set of connected component of relevant pixels is found on the foreground mask under a Markovian framework. The functional block diagram of the proposed method is shown in the Fig. 1. The steps involved in the developed method

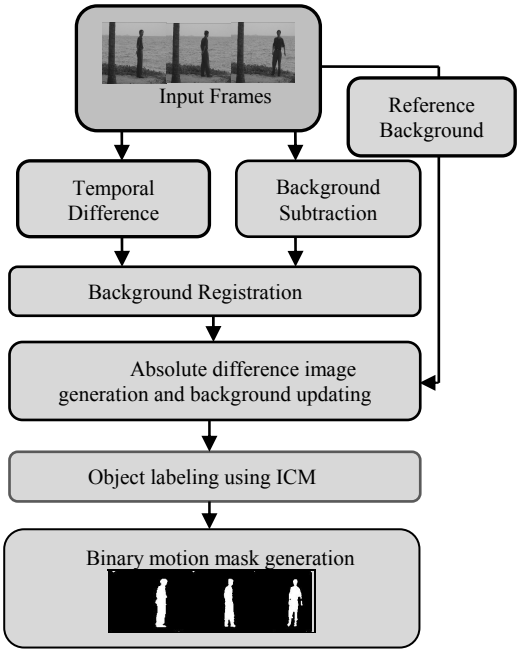


Fig. 1. Block diagram of the proposed method.

to generate an efficient background model offer several advantages over other methods reported in this paper.

- It does not require memory buffer allocation to generate the reference background model.
- This framework incorporates spatial and temporal features to characterize the registered background appearance and provides a better adaptation for the temporal changes in the environment.
- The proposed background model is updated using a selective maintenance approach based on intensity variation of difference image that reduces the aperture distortion, ghost effect and over segmentation error.
- Finally, Markovian framework provides a suitable set of connected component on the foreground mask and spatial regularization against illumination discrepancy.

Each of comprising stages is elaborated in the following subsections.

3.1 Generation of Background Model

In our implementation, we assume initial frame $I_0(x,y)$ as reference background $B_{\text{ref}}(x,y)$, which consists of no foreground object. The first stage is to compute a set of stationary pixels using frame difference and reference background image. The frame difference differentiates the stationary pixels from non-stationary pixels by using a suitable threshold. Using the difference between the current frame $I_t(x,y)$ and previous frame $I_{t-1}(x,y)$ a set of stationary pixels is selected with the aid of reference background frame $B_{\text{ref}}(x,y)$ as follows:

$$B_t^{\text{fD}}(x,y)=\begin{cases} B_{\text{ref}}(x,y), & \text{if } |I_t(x,y)-I_{t-1}(x,y)|<\tau_1 \\ B_{\text{ref}}(x,y)+\text{sgn}(I_t(x,y)-I_{t-1}(x,y)), & \text{otherwise} \end{cases} \quad (1)$$

where $B_t^{\text{fD}}(x,y)$ is a set of stationary pixels through frame difference method and τ_2 is a user defined threshold. The Signum function is defined as:

$$\text{sgn}(p)=\begin{cases} 1, & \text{if } p>0 \\ 0, & \text{if } p=0 \\ -1, & \text{if } p<0 \end{cases} \quad (2)$$

where p is the input value.

At the same time, we investigate the stationary pixels through background subtraction method that subtracts the current input frame $I_t(x,y)$ from the reference background $B_{\text{ref}}(x,y)$.

$$B_t^{\text{bG}}(x,y)=\begin{cases} B_{\text{ref}}(x,y), & \text{if } |I_t(x,y)-B_{\text{ref}}(x,y)|>\tau_2 \\ I_t(x,y), & \text{otherwise} \end{cases} \quad (3)$$

where $B_t^{\text{bG}}(x,y)$ is a set of stationary pixels through background subtraction method and τ_2 is a user defined threshold. The pixels on the current background frame are regis-

tered by averaging the stationary pixels of $B_t^{fD}(x,y)$ and $B_t^{fG}(x,y)$ which is given as:

$$B_t^{reG}(x,y)=\left(\frac{B_t^{fD}(x,y)+B_t^{bG}(x,y)}{2}\right). \quad (4)$$

The initial spatial variance is given as:

$$\sigma_i^2(x,y)=\mathrm{var}(I_0(x,y)) \quad (5)$$

where $\sigma_i^2(x,y)$ is the initial variance.

The current change in spatial variance with respect to time is estimated using initial variance as follows:

$$\sigma_c^2(x,y)=\sigma_i^2(x,y)+\mathrm{sgn}(\mathrm{var}(I_t(x,y)-\sigma_i^2(x,y))) \quad (6)$$

where $\sigma_c^2(x,y)$ is the current spatial variance.

It is desirable to detect the pixels that significantly deviate from the background in order to get the moving object. Here, the approach is to categorize the background pixels on the basis of their stationary and non-stationary behavior in the consecutive frames. As compared to stationary pixels, a non-stationary pixel of background image possesses different statistical foundation. The non-stationary pixels arise due to local motion in the background image such as rippling of water, moving of curtain etc. In this regard, our proposal is to use a selective maintenance scheme that updates background model with different learning rate depending on the stationary or non-stationary background pixels. Since the running average is highly adaptive to the temporal changes in the environment. We analyze and exploit the properties of running average together with spatial variance of current input in order to update the initial background model.

The absolute difference between the current and background frame results initial motion field. Ideally, the initial motion field should contain significant magnitude of intensity of foreground pixels and zero intensity to the matched pixel. However, it is not possible ideally. The first possibility arises for erroneous detection due to the similar magnitude of intensity of foreground and background pixels that can cause holes in moving entities and increase the false-negative pixels. To minimize this error, we select a learning rate β and update those pixels of background image that satisfy the first condition of (7).

$$B_t(x,y)=\begin{cases}\beta\cdot B_{t-1}(x,y)+(1-\beta)\cdot(I_t(x,y)-B_{t-1}(x,y)),\\ \text{if}\quad 0<\left|I_t(x,y)-B_{t-1}(x,y)\right|<1\\ \alpha\cdot B_{t-1}(x,y)+(1-\alpha)\cdot(\sigma_c^2(x,y))-\sigma_i^2(x,y),\\ \text{elseif}\ \left|I_t(x,y)-B_{t-1}(x,y)\right|<\psi\sigma_c\\ \text{and}\ 1<\left|I_t(x,y)-B_t^{reG}(x,y)\right|<\tau_3\\ B_{t-1}(x,y)+\mathrm{sgn}(I_t(x,y)-B_{t-1}(x,y)),\\ \text{else}\end{cases} \quad (7)$$

In (7), $B_t(x,y)$ is the current updated background and $B_{t-1}(x,y)$ is previous background or initial reference frame. σ_i is the initial standard deviation of a reference background frame and σ_c is the current standard deviation of current frame. ψ ranges from 1 to 3 in this experiment. τ_3 is user define threshold. The value of β is taken as 0.999 for all video sequences. The value of α ranges from 0.8 to 0.99.

A recursive filter integrates the current image with the difference image to update the pixel of the background model. It updates those pixels, which range under the first condition of (7). As a result, it provides a difference in the intensity level between foreground and background pixels that initially have identical magnitude.

The second probable reason for erroneous detection can arise when the variance of pixel changes due to the local motion in the background. Concerning this problem, we emphasize to incorporate the current and initial spatial variances that are blended with the pixels of background image using a different learning rate α to update the background pixel. This time, it updates those pixels, which range under the second condition of (7). As a result, it reduces the false-positive pixels. Otherwise, the update of background model is done according to the third condition of (7). The absolute difference image between the first frame and the current background is used to compute the initial motion field. The absolute difference image is given as:

$$M_t(x,y)=\left|I_t(x,y)-B_t(x,y)\right| \quad (8)$$

where $M_t(x,y)$ is the absolute difference image.

3.2 Detection and Labeling of Foreground Object

In real-time application, the estimation of initial motion field is perturbed due to noise. Concerning to this problem, the optimum labeling of motion mask is computed using Iterated Conditional Mode (ICM) under a Markovian framework [27]. ICM is computationally efficient and provides robust smoothing to degraded image by considering the spatial correlation of neighboring pixels. ICM relies on the assumption that neighboring pixels consist of equal value of intensity and each pixel unit is corrupted independently with some probability. To estimate the foreground pixel, a Markovian framework requires prior information to the underlying scene. In this context, the labels achieved during the estimation of initial motion field or absolute difference image provide a provisional known to underlying image in this experiment. Using first order spatial neighborhood, the information regarding provisional known is provided to Gibbs prior within Ishin model [28], [29]. It is focused to update the current estimated value R at pixel $v=(x,y)$ by maximizing the given argument:

$$\arg\max P\left(R_v/O_v,\hat{R}_{s/v}\right)\propto f\left(O_v/R_v\right)P_v\left(R_v/R_{\delta v}\right) \quad (9)$$

where s/v includes the set of all neighbors of the pixel v and ∂v a small set of neighbors of pixel v defined by a first

order neighborhood system. O_v is the observed motion field at v. R_v is the current estimated value of motion field at v.

However, a posterior probability for the estimation of true image relies on minimizing the potential at cliques. This is accomplished by minimizing the given argument iteratively as follows

$$\arg\min \frac{1}{2k}\{O_v - \mu(R_v)\}^2 - \beta_1 \cdot U(R_v). \tag{10}$$

The term $U(R_v)$ is the potential at neighborhood configuration comes through following the Gibbs sampler within the Ishin model. The $U(R_v)$ is expressed as follows:

$$U(R_v) = -\alpha_1 \sum_v R_v - \beta_1 \sum_v v_1(R_v). \tag{11}$$

However, α_1 controls the biasing of negative pixel and positive pixels. The α_1 is set to '0' in this experiment. $v_1(R_v)$ is the number of neighbors of v having label R_v. β_1 is a constant, which is experimentally set to 0.0001. Generally the motion mask possesses two labels $l1$ and $l2$, such that $l1, l2 \in \{R_v\}$. The maximal likelihood of each of its labels together with the prior is expressed as:

$$P_1\left(O_v / R_v = l1\right) = \frac{1}{\sqrt{2\pi\sigma_{l1}^2}} \exp\left(\frac{-(O_v - \mu_{l1})^2}{2\sigma_{l1}^2}\right) \cdot \exp\left(\beta_1 \cdot \sum v_1(R_v)\right), \tag{12}$$

$$P_2\left(O_v / R_v = l2\right) = \frac{1}{\sqrt{2\pi\sigma_{l2}^2}} \exp\left(\frac{-(O_v - \mu_{l2})^2}{2\sigma_{l2}^2}\right) \cdot \exp\left(-\beta_1 \cdot \sum v_1(R_v)\right), \tag{13}$$

where P_1 and P_2 are the posterior probability. Using the mean μ_d and standard deviation σ_d of initial motion field, the mean parameters μ_{l1}, μ_{l2} and the variance parameters σ_{l1}^2, σ_{l2}^2 for the likelihood functions are calculated. The mean μ_d and the standard deviation σ_d are given as:

$$\mu_d = \frac{1}{w \cdot h} \sum_{x=1}^{w} \sum_{y=1}^{h} M_t(x, y), \tag{14}$$

$$\sigma_d = \left(\frac{1}{w \cdot h} \sum_{x=1}^{w} \sum_{y=1}^{h} (M_t(x, y) - \mu_d)^2 \right)^{1/2}. \tag{15}$$

The variables w and h belong to width and height of a frame respectively. The value of μ_{l1} is estimated as μ_d, while μ_{l2} ranges up to $\mu_d \pm 3\sigma_d$ for the test videos in this paper. The value of σ_{l1} is estimated as σ_d, while the ratio of σ_{l1} to σ_{l2} is taken $1.5\sigma_d$ for the static sequences. Concerning the dynamic sequences, the ratio of σ_{l2} to σ_{l1} is taken as $2.5\sigma_d$. The value of the estimated binary motion mask $D_t(x,y)$ is evaluated as follows:

$$D_t(x, y) = \begin{cases} 1, \text{ if } P_1\left(O_v / R_v = l1\right) > P_2\left(O_v / R_v = l2\right) \\ 0, \text{ if } P_1\left(O_v / R_v = l1\right) < P_2\left(O_v / R_v = l2\right) \end{cases} \tag{16}$$

To remove the unnecessary connected component from the foreground mask and filling the superfluous holes, the morphological operations are performed using structuring element [11]. In this experiment, the morphological open operation followed by close operation is performed to investigate the relevant connected components using the structuring element 'disk' shape with radius '1'.The opening operation is performed on the foreground mask as follows:

$$g(x,y) = \left(D_t(x,y) \Theta S\right) \oplus S. \tag{17}$$

Consequently, closing on image is performed as follows:

$$f(x,y) = \left(g(x,y) \oplus S\right) \Theta S \tag{18}$$

where 'S' is the structuring element, Θ operator performs erosion operation and $\oplus$ performs dilation on an image.

4. Experimental Results

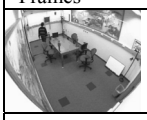
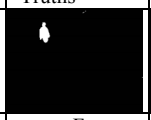
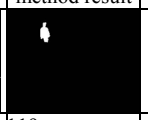
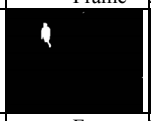
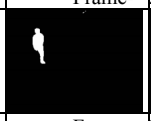
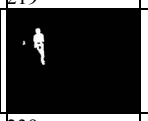
In this section, seven standard video sequences are considered to validate our results qualitatively and quantitatively. The detailed analysis of some challenging sequences is explored in this section. The primary features of these video sequences are given in Tab. 1.

Video Title	Image Size	Environment	Background Property
IR	240×320	Indoor	Static
MSA	240×320	Outdoor	Static
PET2006	240×320	Outdoor	Static
MR	128×160	Indoor	Dynamic
WS	128×160	Outdoor	Dynamic
FT	128×160	Outdoor	Dynamic
CANOE	240×320	Outdoor	Dynamic

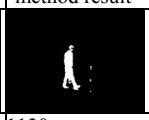
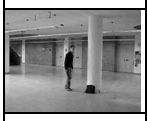
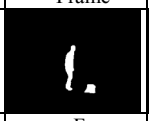
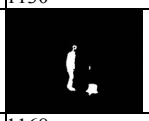
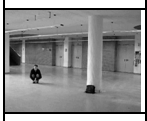
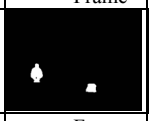
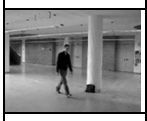
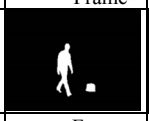
Tab. 1. Video sequences for objective evaluation.

The foreground mask may distort due to aperture effect, over-segmentation error, ghost and camouflage effect. Aperture effect is related to the problem to find the actual correspondence between the consecutive frames. False positive pixels cause over-segmentation error. A false copy of moving object generated on the foreground mask that disappears slowly with the time is called ghost effect. Camouflage effect may arise due to similar intensity between foreground and background image. In this regard, qualitative evaluation is done on various dataset against these problems.

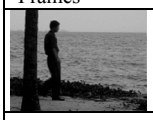
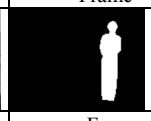
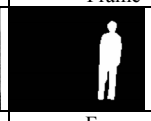
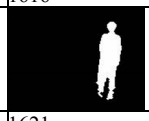
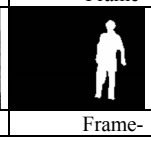
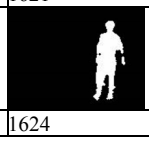
The IR and MSA video sequences consist of static background object. In IR sequence, change of illumination condition and shadows cast by object can hurdle to produce the reliable foreground mask. The radial movement of person may also affect the aperture in IR sequence. In case of MSA and PET2006 sequences, changing illumination and abandoned object in the scene may degrade the performance of binary motion mask. As shown in Fig. 2a, our proposed scheme successfully detects the moving person against the illumination and shadows. Moreover, no aperture distortion is seen in the detection results of 'IR' sequences. In other static background of MSA and PETS 2006 sequence, the proposed approach detects the person activity along with the abandoned bag continuously object.

Sampled Frames	Ground Truths	Proposed method result	Similarity F1
			0.8824 0.9375
	Frame-	119	
			0.8016 0.8878
	Frame-	211	
			0.8354 0.9103
	Frame-	219	
			0.8075 0.8935
	Frame-	239	

(a)

Sampled Frames	Ground Truths	Proposed method result	Similarity F1
			0.8525 0.9204
	Frame-	1130	
			0.8512 0.9196
	Frame-	1168	
			0.8503 0.9199
	Frame-	1232	
			0.8912 0.9425
	Frame-	1336	

(b)

Sampled Frames	Ground Truths	Proposed method result	Similarity F1
			0.8814 0.9316
	Frame-	1499	
			0.9373 0.9676
	Frame-	1616	
			0.8942 0.9441
	Frame-	1621	
			0.8900 0.9418
	Frame-	1624	

(c)

Sampled Frames	Ground Truths	Proposed method result	Similarity F1
			0.9057 0.9505
	Frame-	22774	
			0.8850 0.9390
	Frame-	22847	
			0.8834 0.9381
	Frame-	23857	
			0.8590 0.9241
	Frame-	23893	

(d)

	Frame-1165	Frame-1184	Frame-1426	Frame-1477
Sampled frames				
Ground truths				
Proposed method results				
Similarity F1	0.7435 0.8529	0.6875 0.8148	0.7099 0.8304	0.6781 0.8082

(e)

Fig. 2. Output of video sequences with Similarity and F1 (a) IR sequence, (b) MSA sequence, (c) WS sequence, (d) MR sequence, (e) FT sequence.

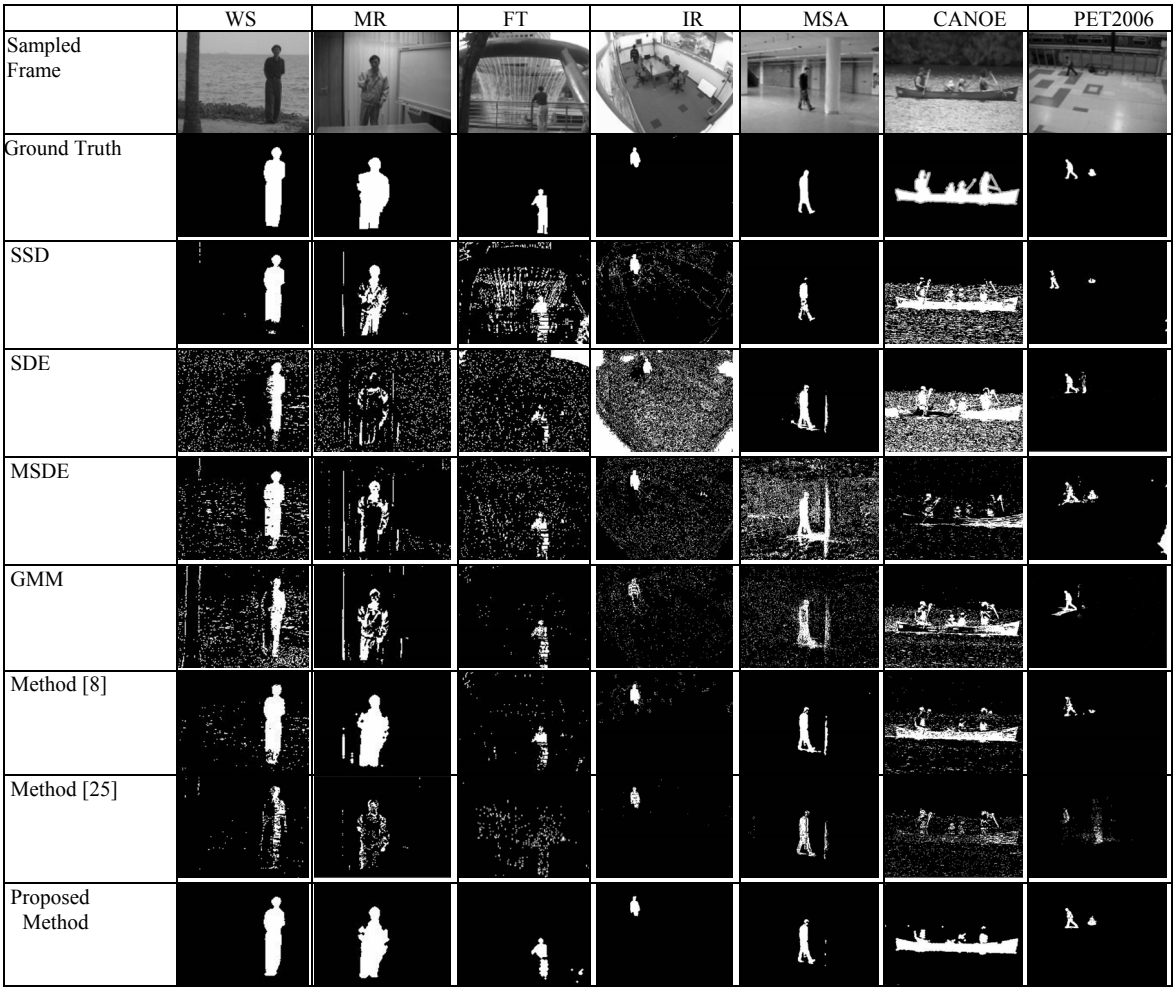


Fig. 3. Performance comparison of motion mask generated by the proposed method and other baseline methods.

in the consecutive video frames even the objects appear to sleep or motionless for some frames. The detection results of MSA sequence are shown in Fig. 2b.

The Water Sequences (WS) consist of dynamic background feature with changing illumination. This sequence suffers from high frequency noise due to rippling of water in the background. Moreover, a potential problem arises due to the identical pixel intensity of background vegetation and foreground pixel below the knee of the person (camouflage effect). Another issue arises in the WS sequence, when a person moves slowly and becomes stationary in the scene. As one can observe in Fig. 2c, the proposed method suppresses false-positives induced due to illumination changes and false negatives caused by the similarities between object and foreground.

In other dynamic background of MR sequence, waving curtain produces high frequency noise. In its consecutive images, the person's shirt tends to camouflage with color of moving curtain. Moreover, the distraction can also arise during the object segmentation in MR sequence due to slow motion and sleeping of the moving object.

In all these distracting cases, the background and foreground pixels are separated significantly by bounding the variance of background model through this proposed method. Our proposal eliminates the high frequency noise and false negative pixels, which arise due to non-stationary pixel on the background by the moving curtains. The detection results of 'MR' sequence are shown in Fig. 2d. The fountain sequence (FT) and CANOE sequences also consist of dynamic background feature.

The difficulties in segmentation can arise due to high frequency ripples produced by the fountain and river. As shown in Fig. 2e, this method adapts environmental changes of dynamic background and produce satisfactory results on the foreground. Figure 3 shows that the proposed method gives better performance than other background subtraction methods.

In addition to visual inspection, the performance of the proposed approach is also evaluated quantitatively on the above-mentioned video sequences with respect to their ground truths [1], [15], [17]. Quantitative evaluations with respect to the ground truth image depend on True-positive (*tp*) pixels, True-negative pixels (*tn*), False-positive pixels (*fp*), and False-negative pixels (*fn*). True-positive pixels (*tp*) are the correctly detected pixels by the algorithm of the moving object. False-positives (*fp*) concern to those pixels,

which are incorrectly detected as foreground. True negative pixels (*tn*) are correctly detected pixels that correspond to background while False-negative pixels (*fn*) correspond to the number of foreground pixels detected incorrectly as background. The relevant pixels on the binary motion mask are analyzed using *Recall* metric, which is given as:

$$Recall = tp / (tp + fn) . \quad (19)$$

The irrelevant pixels on the binary motion mask are analyzed using *Precision* metric, which is computed as:

$$Precision = tp / (tp + fp) . \quad (20)$$

An algorithm must achieve a high recall rate without sacrificing the Precision metric but, these two metric do not support the reliable measurement task. Similarity and F1 are two other parameters, which incorporate to patch up the reliable accuracy measurements in quantitative analysis. The *Similarity* and *F*1 are given as:

$$Similarity = tp / (tp + fp + fn) , \quad (21)$$

$$F1 = 2 \cdot Recall \cdot Precision / (Recall + Precision) . \quad (22)$$

Percentage of correct classification (*PCC*) is the most extensive way to assess a classifier's performance as it includes *tp, tn, fp* and *fn* parameters. The *PCC* is given as:

$$PCC = (tp + tn) / (tp + tn + fp + fn) . \quad (23)$$

We also analyze True positive rate (*TPR*) and False positive rate (*FPR*) to compare our misclassified results between foreground and background image. *TPR* is equivalent to *Recall* rate while, false positive Rate (*FPR*) is those background pixels which are misclassified as foreground. The *TPR* and *FPR* are given as:

$$TPR = tp / (tp + fn) , \quad (24)$$

$$FPR = fp / (fp + tn) . \quad (25)$$

Table 2 lists the average accuracy rates through this method along with accuracy rates that were achieved by some other existing state-of-the-art background subtraction GMM, MSDE, SDE, SSD, [8] and [25], methods reported in this paper. The accuracy rates calculated by MSDE, SDE, SSD for IR, MR and WS video sequences are taken from [8], while the rest of accuracy rates for GMM, MSDE, SSD, SDE method [8] and method [25] are calculated using the optimum parameter as given in [25], [21], [20], [11], [8], [7].

We can easily examine that the performance of the proposed method is superior to previously reported six different methods. This method achieves the higher accuracy rates of all metrics than 92.36% for WS sequence. With regard to MR, WS and MSA sequence, it is noted that this method achieves greater accuracy rates of all the metrics than 82% that reflects the significant improvement in motion detection task under circumstances with illumination discrepancy and local motion.

In WS sequence, the highest average accuracy rates secured through *F*1 and *similarity* by this method are up to 56% and 54% higher than those attained by GMM method. In FT sequence, the lowest average accuracy rates secured through *F*1 and *similarity* by this method are also up to 17% and 23% higher than those attained by GMM method.

Sequences	Evaluation	Proposed Method	Method [25]	Method [8]	GMM	MSDE	SDE	SSD
IR	Similarity	0.7989	0.6166	0.6599	0.4844	0.2120	0.1528	0.1810
	F1	0.9185	0.7550	0.7917	0.6526	0.3390	0.2574	0.2936
	Precision	0.9333	0.8491	0.7864	0.8478	0.2257	0.1594	0.2814
	Recall	0.8492	0.6803	0.8025	0.5306	0.8344	0.8328	0.3900
MSA	Similarity	0.8448	0.4446	**0.8519**	0.2916	0.1691	0.1301	0.8070
	F1	0.9159	0.6156	**0.9201**	0.3440	0.2893	0.2301	0.8930
	Precision	0.9293	0.6323	**0.9399**	0.6253	0.1360	0.1355	**0.9481**
	Recall	0.9028	0.5997	0.9201	0.3154	0.7735	0.7586	0.8440
WS	Similarity	0.9051	0.4175	0.7660	0.5344	0.5408	0.3521	0.7213
	F1	0.9501	0.5890	0.8669	0.6918	0.6977	0.5197	0.8340
	Precision	0.9787	0.9065	0.8684	0.8549	0.8665	0.7406	0.7994
	Recall	0.9236	0.4363	0.8673	0.6259	0.5938	0.4073	0.8756
MR	Similarity	0.8292	0.4809	0.8077	0.6500	0.5138	0.3374	0.4559
	F1	0.9044	0.6495	0.8929	0.7685	0.6652	0.5328	0.6185
	Precision	0.9648	0.9500	0.8665	0.9502	0.6651	0.4800	0.7141
	Recall	0.8516	0.4934	**0.9246**	0.6768	0.7041	0.6224	0.5689
FT	Similarity	0.6321	0.4173	0.6000	0.4298	0.1594	0.4053	0.3180
	F1	0.7709	0.5889	0.7431	0.6012	0.3059	0.3384	0.4808
	Precision	0.8552	**0.9064**	0.8122	0.6988	0.5176	0.5130	0.4299
	Recall	0.7067	0.4362	0.6848	0.5321	0.2281	0.2600	0.6887
PETSET	Similarity	0.7212	0.2485	0.7000	0.3592	0.4044	0.4599	0.2983
	F1	0.8378	0.3979	0.8239	0.5255	0.5407	0.6302	0.4594
	Precision	0.8014	0.5449	**0.8367**	**0.8150**	0.6848	0.6837	0.3385
	Recall	0.8781	0.3134	0.8111	0.3915	0.5100	0.5845	0.7142
CANOE	Similarity	0.6435	0.3000	0.6131	0.4601	0.2273	0.1809	0.5912
	F1	0.7831	0.4549	0.7602	0.6302	0.3707	0.2934	0.7431
	Precision	0.9793	**0.9899**	0.6921	0.6837	0.9825	0.2809	0.8122
	Recall	0.6522	0.2950	0.8432	0.3468	0.2285	0.3898	0.6842

Tab. 2. Performance comparison of average quantitative metrics.

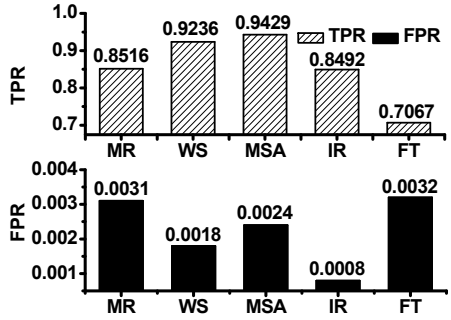


Fig. 4. TPR and FPR of video sequences.

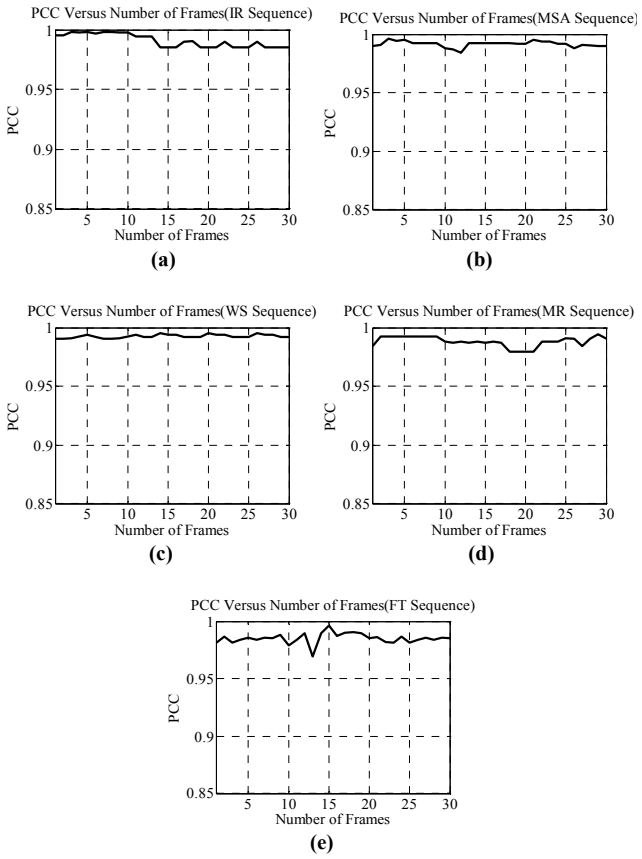


Fig. 5. PCC versus Number of Frames: (a) IR, (b) MSA, (c) WS, (d) MR, (e) FT sequence.

The quantitative analysis between *TPR* and *FPR* is shown in Fig. 4. Our method achieves lower *FPR*, which reflects that average misclassified pixels are below 0.3% by employing this background subtraction method under study.

The *PCC* is measured by taking some sampled frames of each video sequence, which are shown in Fig. 5 (a)-(e).

The average *PCC* metrics measured for WS, MSA, MR, and IR video sequences are greater than 99.12% while for FT sequence it is measured above 98%. The high value of *PCC* reflects the better segmentation and identification of foreground pixels through this method. With regard to time complexity, we perform all experiments on Matlab 7.1 using 3.2 GHz Intel CPU, 2G RAM on Window7 platform. To process a 120 × 160 frame, this method takes 0.06 sec, while GMM takes 0.48 sec. Other methods are faster than our algorithm.

5. Conclusion

In this paper, we described our contribution to characterize the background appearance by using its principle feature and statistics. A test based on the standard deviation of pixel and estimated absolute difference image is applied in order to limit the variance due to the local motion and change in illumination in the background. In addition to that, the most appropriate label assignment to the motion field has been estimated and optimized by using iterated conditional modes under a Markovian framework. Nevertheless, one can extend the work in future in regard to the problem of handling the drastic illumination changes and multiple moving objects in the scene, yet this method can extract the moving object even under circumstances with moderate illumination discrepancy and local motion. Experimental results specify that the proposed algorithm has a propensity to localize the object in the scene without over-segmentation error, aperture distortion, and ghost effect. Extensive qualitative and quantitative analysis exemplify that our method attains greater accuracy rates than some other state–of–the-art background subtraction methods previously reported in the paper.

References

[1] RADKE, R. J., ANDRA, S., AL-KOFAHI, O., et al. Image change detection algorithm: a systematic survey. *IEEE Transactions on Image Processing*, 2005, vol. 14, no. 3, p. 294–307. DOI:10.1109/TIP/2004.838698

[2] PAUL, M., HAQUE, S. M, CHAKRABORTY, S. Human detection in surveillance videos and its application - a review. *EURASIP Journal on Advances in Signal Processing*, 2013, no. 11, p. 1–25. DOI: 10.1186/1687-6180-2013-176

[3] MANDELLOS, N. A., KERAMITSOGLU, I., KIRANOUDIS, C. T. A background subtraction algorithm for detecting and tracking vehicles. *Expert System with Applications*, 2011, vol. 38, no. 3, p. 1619–1631. DOI:10.1016/j.eswa.2010.07.083

[4] ZHANG, W., ZHANG, Y., GAO, C., et al. Action recognition by joint spatial-temporal motion feature. *Journal of Applied Mathematics*, 2013, vol. 2013, 9 p. DOI: 10.1155/2013/605469

[5] LUCAS, B. D, KANADE, T. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI)*. Vancouver (British Columbia), August 1981, vol. 81, p. 674–679.

[6] SPAGNOLO, P., D'ORAZIO, T., LEO, M., DISTANTE, A. Moving object segmentation by background subtraction and temporal analysis. *Image and Vision Computing*, 2006, vol. 24, no. 5, p. 411–423. DOI: 10.1016/j.imavis.2006.01.001

[7] ORAL, M., DENIZ, U. Centre of mass model – A novel approach to background modeling for segmentation of moving objects. *Image and Vision Computing*, 2007, vol. 25, no. 8, p. 1365–1376. ISSN: 0262-885 6. DOI:10.1016/j.imavis.2006.10.001

[8] HUANG, S.C. An advanced motion detection algorithm with video quality analysis for video surveillance systems. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, vol. 21, no. 1, p. 1–14. DOI: 10.1109/TCSVT.2010.2087812

[9] XUE, G., SUN, J., SONG, L. Background subtraction based on phase feature and distance transforms. *Pattern Recognition Letters*, 2012, vol. 33, no. 12, p. 1601–1613. DOI: 101016/j.patrec. 2012.05.009

[10] VOSTERS, L., SHAN, C., GRITTI, T. Real-time robust background subtraction under rapidly changing illumination conditions. *Image and Vision Computing*, 2012, vol. 30, no. 12, p. 1004–1015. DOI: 10.1016/j.imavis.2012.08.017

[11] MANZANERA, A., RICHEFEU, J. C. A new motion detection algorithm based on Σ–Δ background estimation. *Pattern Recognition Letters*, 2007, vol. 28, no. 3, p. 320–328. DOI: 10.1016/j.patrec.2006.04.007

[12] MCFARLANE, N. J., SCHOFIELD, C. P. Segmentation and tracking of piglets in images. *Machine Vision and Applications*, 1995, vol. 8, no. 3, p. 187–193. DOI: 10.1007/BF01215814

[13] ZHONG, J., SCLAROFF, S. Segmenting foreground objects from a dynamic textured background via a robust Kalman filter. In *Proceedings of the 9th IEEE International Conference on Computer Vision*. Nice (France), 2003, vol. 1, p. 44–50. DOI: 10.1109/ICCV.2003.1238312

[14] WREN, C. R., AZARBAYEJANI, A., DARRELL, T., PENTLAND, A. P. Pfinder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, vol. 19, no. 7, p. 780–785. DOI: 10.1109/34.598236

[15] CHEUNG, S. C. S., KAMATH, C. Robust background subtraction with foreground validation for urban traffic video. *EURASIP Journal on Advances in Signal Processing*, 2005, p. 2330–2340. DOI: 10.1155/ASP.2005.2330

[16] CUCCHIARA, R., GRANA, C., PICCARDI, M., et.al. Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, vol. 25, no. 10, p. 1337–1342. DOI: 10.1109/TPAMI2003.1233909

[17] DO, B. H., HUANG, S. C. Dynamic background modeling based on radial basis function neural networks for moving object detection. In *IEEE International Conference on Multimedia and Expo.* Barcelona (Spain), 2011, 4 p. DOI: 10.1109/ICME.2011.6012085

[18] NIKOLOV, B., KOSTOV, N. Motion detection using adaptive temporal averaging method. *Radioengineering*, 2014, vol. 23, no. 2, p. 652–658. ISSN:1210-2512

[19] RAHMAN, F. Y. A., HUSSAIN, A., ZAKI, et al. Enhancement of background subtraction techniques using a second derivative in gradient direction filter. *Journal of Electrical and Computer Engineering*, 2013, no. 21, 12 p. DOI: 10.1155/2013/598708

[20] MANZANERA, A., RICHEFEU, J. C. A robust and computationally efficient motion detection algorithm based on Σ-Δ background estimation. In *Proceedings of the Fourth Indian Conference on Computer Vision, Graphics and Image Processing.* Kolkata (India), 2004, p. 46–51. ISBN: 81-7764-7075

[21] STAUFFER, C., GRIMSON, W. E. L. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, vol. 22, no. 8, p. 747–757. DOI: 10.1109/34.868677

[22] EVANGELIO, R. H., SIKORA, T. Static object detection based on a dual background model and a finite-state machine. *EURASIP Journal on Image and Video Processing*, 2010, 11 p. DOI: 10.1155/2011/ 858502

[23] HUANG, S. S., FU, L. C., HSIAO, P. Y. Region-level motion-based background modeling and subtraction using MRFs. *IEEE Transactions on Image Processing*, 2007, vol. 16, no. 5, p. 1446–1456. DOI: 10.1109/TIP.2007.894246

[24] BOUWMANS, T. Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*, 2014, vol. 11–12, p. 31–66. DOI: 10.1016/j.cosrev.2014.04.001

[25] ZHAO, Z., BOUWMANS, T., ZHANG, X., et al. A fuzzy background modeling approach for motion detection in dynamic backgrounds. *Multimedia and Signal Processing,* 2012, Springer Berlin Heidelberg, vol. 346, p. 177–185. DOI: 10.1007/978-3-642-35286-7_23

[26] LIN, L., XU, Y., LIANG, X., LAI, J. Complex background subtraction by pursuing dynamic spatio-temporal models. *IEEE Transactions on Image Processing*, 2014, vol. 23, no. 7, p. 3191–3202. DOI: 10.1109/TIP.2014.2326776

[27] BESAG, J. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1986, vol. 48, no. 3, p. 259–302. DOI: 10.2307/2345426

[28] HALIM, S. Modified Ishin model for generating binary images. Jurnal Informatika, 2008, vol. 8, no. 2, p. 115–118, ISSN: 1411-0105

[29] REDDY, V., SANDERSON, C., LOVELL, B. C. A low-complexity algorithm for static background estimation from cluttered image sequences in surveillance contexts. *Journal on Image and Video Processing*, 2011, p. 1–14. DOI: 10.1155/2011/164956

About the Authors...

Satrughan KUMAR was born on 15th May 1982. He received his B.E (Elec. & Comm.) and M.Tech from RGTU University, Bhopal in 2006 and 2010, respectively. Currently, he is pursuing his Ph.D from MANIT Bhopal in the area of image processing and surveillance system.

Jigyendra Sen YADAV is an Associate Professor in Elec. & Comm. Engineering Dept., MANIT, Bhopal, India. Currently, he is the senior member of the International Asssociation of Computer Science and Information Technology, member of Human Right Commission to study the "Effect of Radiation from Mobile Tower". His area of interest includes optimization techniques, model order reduction, digital communication, signal & system and control system.