
Adaptive background modelling technique for moving object detection in video under dynamic environment

Dileep Kumar Yadav* and Karan Singh

School of Computer and Systems Sciences,
Jawaharlal Nehru University,
New Delhi, India
Email: dileep252000@gmail.com
Email: karan@jnu.ac.in
*Corresponding author

Abstract: This work proposes a novel method for detection of motion based object having dynamic scenario in the background. The suggested scheme has a strong potential for real-time applications especially for rafting, river, sea-beach, swimming pools, ponds, etc. Apart from these, this work is very beneficial for surveillance of border, tunnel, traffic in the sea, forest, restricted zones, deep zones, etc. This work develops a statistical p based background subtraction method and implemented in three stages. In the first stage, a background model is developed using few initial frames. In the second stage, this work classifies the foreground using the difference frame and the appropriate threshold value. An automatic threshold value is generated at run-time and updated iteratively. It also reduces the problem of using a constant threshold. In the third stage, morphological filters and connected component based region filtering technique is applied to enhance the detection quality. The extensive experimental result shows more accurate results of proposed method. It also demonstrates better performance against considered state-of-the-art methods.

Keywords: cluttered background; adaptive modelling; background subtraction; outliers; moving object segmentation; visual surveillance.

Reference to this paper should be made as follows: Yadav, D.K. and Singh, K. (2019) 'Adaptive background modelling technique for moving object detection in video under dynamic environment', *Int. J. Spatio-Temporal Data Science*, Vol. 1, No. 1, pp.4–21.

Biographical notes: Dileep Kumar Yadav received his Engineering (BTech in Computer Science and Engineering) from the Uttar Pradesh Technical University, Lucknow, UP, India in 2006 and Master's degree (MTech in Computer Science and Technology) from the School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India in 2011. He has completed his PhD (Computer Science and Technology) from the School of Computer and Systems Sciences, Jawaharlal Nehru University New Delhi, India. This work belongs to his PhD work. His primary research interests are in image processing and computer vision. He is also a Sun Certified Java Programmer for Platform 1.5 (SCJP 1.5). He has five years of working experience in industry as well as academia. He has published many research papers (in journals of repute, international and national conferences).

Karan Singh received his Engineering (Computer Science and Engineering) from the Kamala Nehru Institute of Technology, Sultanpur, UP, India and the MTech (Computer Science and Engineering) from the Motilal Nehru National Institute of Technology UP, India. He holds a PhD (Computer Science and Engineering) from the MNNIT Allahabad deemed university. Currently, he is working in the School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi. His primary research interests are in computer network and computer network security. He is a reviewer of IEEE and Elsevier conferences and reviewer of international journals. He has published many research papers (eight journals, nine IEEE conferences, two Elsevier conferences, 15 national, international conferences) and two journals are accepted. He has organised various workshop, conference and training programs. Recently, he also worked as a General Chair of Qshine 2013. He is a professional member of IEEE, ACM, CSTA, CSI, IACSIT, ICST, IAENG, ACEEE, ISOC and IEEE Computer Society.

1 Introduction

Today, the computer vision paradigm is moving in the direction of the video surveillance system to obtain appropriate, robust and affordable services. These services deal with real-time challenges of object detection and tracking. Due to the growing requirement of the consumer electronic products, the detection of moving object is highly recommended for visual surveillance system (Haines and Xiang, 2014; Dordevic et al., 2015). It also strengthens the vision based surveillance system. Various application areas such as traffic monitoring, traffic control, analysis of suspicious motion, abandoned object detection and identification, medical diagnosis and many indoor outdoor areas can be upgraded (Dordevic et al., 2015; Toyama et al., 1999; Huang et al., 2015). However, the proposed method is very sensitive to detect small variations. The existing methods perform well for the static background and also handle gradual illumination variation. So, it fails to handle the dynamic background. Due to this reason, these methods assume that the state of a pixel is changed due to the noise or unstructured patterns (Liu et al., 2014) of motion in the background scene. This work is directly motivated by various methods available in the literature and existing challenges of the background that can take advantage of multiple complementary cues to make more accurate and robust result. In literature, various methods have been investigated and implemented in last decade (Lin et al., 2014; Jung, 2009).

The more critical and common challenges are briefly explored as given below.

- *Complex dynamic background*: the background scene is not always static but, sometime consists of the dynamic behaviour due to the presence of:
 - a random motion of the water such as rain, rippling or flow of water, etc.
 - b High motion in the background scene like waving tree leaves, moving flag or banner, etc.
- *Camouflage*: indistinct object having surrounding background with the similar appearance.

- *Lighting and illumination variation*: sudden or gradual change due to the light switch being toggled, change in the illumination caused by the environment such as the sun was covered by the sky, etc.

The aim of the proposed work is to identify and detect the presence of the moving object. It also avoids the detection of noisy or unwanted pixels by upgrading the detection quality that overcomes aforementioned challenges. These pixels may be described as outliers and must be removed. This method also indicates how an algorithm can be further enriched by introducing means of a mean value based governing mechanism along with an automatic feedback strategy for the background model and threshold updation scheme.

2 Related work

In literature, several techniques for the moving object detection using the background subtraction have been proposed in the past and some of them are briefly summarised here. To adapt to temporal changes, recursive Gaussian mixture model (GMM) based schemes by Stauffer and Grimson (1999), Lee (2005) and Haque et al. (2008) have been proposed. These schemes are based on the background subtraction and sequentially enhancement of Stauffer's work. Jung (2009) has constructed the background model using statistical descriptors where a trimmed mean based model is capable enough to minimise the noise and improve the accuracy of the pixel classification. This method generates better results. It also detects shadow pixels and handles illumination variation problem very well. But due to sudden variation in the background scene, it shows poor performance. Javed et al. (2014) investigated a background subtraction method using an online robust principal component analysis (PCA) based method to overcome computational issues. Wang et al. (2012) investigated hierarchical model for the background modelling. Su and Hu (2004) developed a GMM based statistical framework to localise moving foreground and resolve the issue of high dynamic nature of background. Yadav et al. (2014) developed a fast adaptive method that is capable enough to handle the dynamic background. Ng and Delp (2011) method handles the gradual illumination variation very well but fails to handle the sudden illumination change. This method also fails to model fast variations such as waving tree leaves or changing surface of water in the river or canal. Irani et al. (1994) proposed a powerful method that results many false positive alarm due to sudden illumination change in the environment and variation in the surface of the water. Zhou et al. (2013) investigated an improved an adaptive background modelling method that was developed using K-GMM method. Lee and Lee (2014) has developed a pixel based background subtraction method through spatial similarity and structural similarity. Pedersolib et al. (2015) investigated an unsupervised learning method for pixel extraction. The main limitation of this multi-model GMM is the requirement of pre-initialised parameters such as number of distributions for modelling, standard deviation for those distributions, mean value for those distributions, weight parameter (if required). But in reality, the accurate modelling of dynamic backgrounds cannot be possible through few (say five) Gaussian distributions. Li et al. (2017) and West (1979) investigated a method that was developed

using binary hypothesis testing. It is based on clustering of extracted features from multiple input devices which senses the object. St-Charles method (St. Charles et al., 2015) suggested a new approach using spatio temporal binary similarity descriptors that could handle local changes in the frame due to the camouflage issue. Bloisi et al. (2017) developed a modular system for the intelligent surveillance of maritime which fused on the information captured from heterogeneous sources. This work also helps environment protection issues along with dangerous threats coming from the sea such as fishing, illegal smuggling, immigration, piracy and oil spills.

Nowadays, according to the literature, the deep-learning-based background modelling methods are also acts as a powerful tool for moving object detection. Some of them are briefly summarised here. Sermnet et al. (2013) model used few new concepts for integrategration of global shape information, multi-stage features, local distinctive motif information, and a convolution sparse coding based unsupervised method that pre-train the filters at each stage. Druzhkov and Kustikova (2016) focused on various deep learning models such as auto restricted Boltzmann machines, encoders and convolutional neural networks. This paper enlightened on existing software packages, tools for deep learning based problems and also compared them. Braham and Droogenbroeck (2016) developed a background subtraction algorithm for moving object detection where spatial features are learned with convolutional neural networks. Christiansen et al. (2016) developed a background subtraction algorithm and experimented over non-static camera in the agriculture domain. This work used high level features from a deep learning network using background subtraction algorithm. Wang et al. (2016) has illustrated a method that learns the appearance of the background and the foreground. This model involves a multi-scale convolutional neural network with a cascaded architecture and depicted systematically better performance.

These works also encourage the development of intelligent video monitoring systems. So, nonparametric methods are able to handle multiple modes of the dynamic background and boost up the accelerated evolution of computer vision technology.

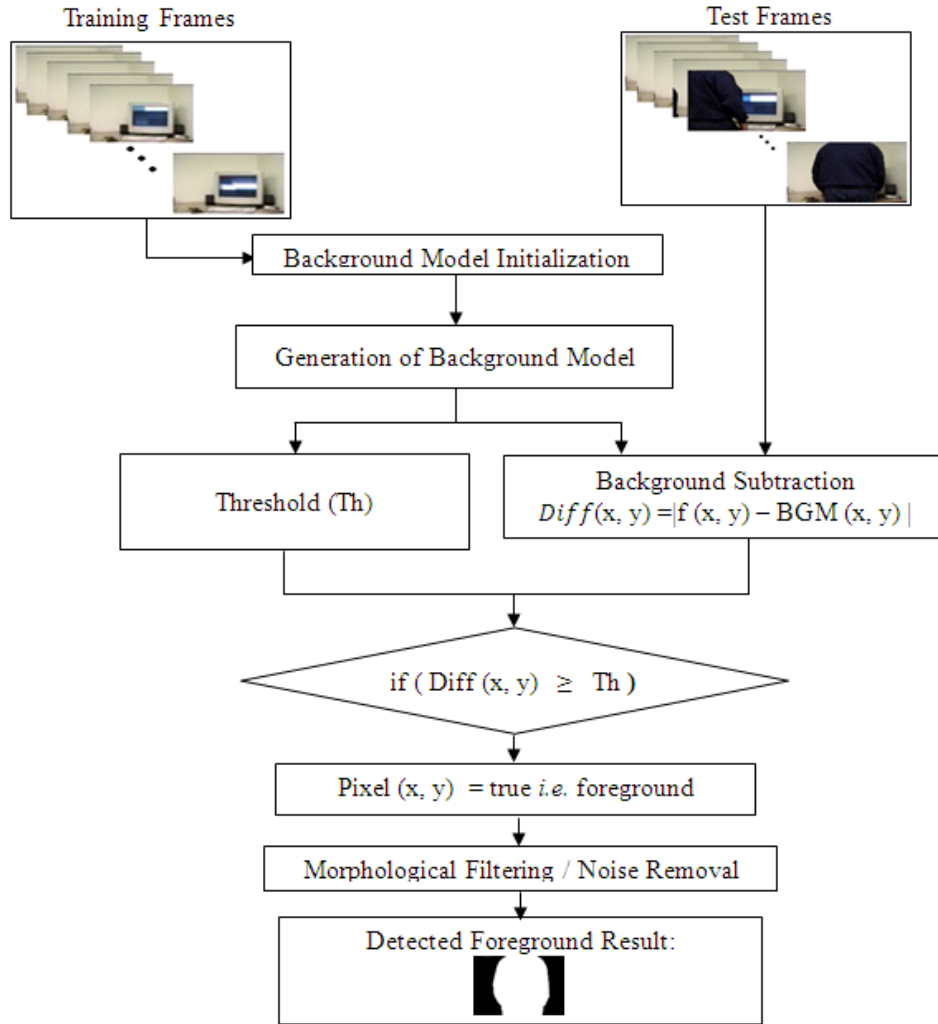
In the remainder part of this paper, the key methodology and algorithm of this work is shown in Section 3. The experimental outcomes are presented in Section 4 along with limitation of the proposed work. Finally, the conclusion and future guidelines are discussed in Section 6.

3 Proposed method

In this section, a systematic solution for the moving object detection is investigated, no matter whether the background is stationary or dynamic. This work considers highly dynamic background. The proposed method is experimented over grey-colour that shows fast computation as compare to the coloured video. This method is also robust against dynamic challenges. The efficiency and applicability of the proposed method are shown in experimental results and analysis. Basic working of the proposed method has shown in Figure 1. Suppose, the probability of perceiving a pixel value $f_i(x, y)$ at time t is evaluated using the Gaussian distribution for two-dimension (Jung, 2009; Lee, 2005). Let us denote, $\mathbf{x} = (x, y)$ as pixel location in a frame and m represents total number of pixels in a frame. The proposed algorithm is applied over frame based processing where, each pixel

of a single frame is sequentially processed first then next frame is executed. Practically, only a single is processed at a time. It depicts low computational cost.

Figure 1 Basic working steps of the proposed method (see online version for colours)



3.1 Development of the background model

Generally, the simple average of initial few frames is computed for construction of the background model. This type of model is susceptible to dynamic nature of the background. For resolving above mentioned issue, a simple algorithm (refers *Algorithm-1*) is developed for the background model. Here, we have considered T number of initial frames $\{B_1(\mathbf{x}), B_2(\mathbf{x}) \dots B_{T-1}(\mathbf{x}), B_T(\mathbf{x})\}$ for the background model, where $100 \leq T \leq 200$, and $\mathbf{x} = (x, y)$.

Algorithm-1: Pseudo Code for proposed Background Modelling

Input: Consider T no of initial frames $\{B_1(\mathbf{x}), B_2(\mathbf{x}) \dots B_{T-1}(\mathbf{x}), B_T(\mathbf{x})\}$, where $100 \leq T \leq 200$; and $\mathbf{x} = (x, y)$

Output: Evaluate background model and threshold

- A initial background model (BGM) frame, i.e., $M_{BGM}(\mathbf{x})$
 - Initial Threshold value for pixel classification
- 1 **Read one frame at a time from initial T frames** // See background modelling as discussed in Section 2.2.
 - 2 **Convert RGB frame into Grey colour format and in 1-D array** // to optimise running time
 - 3 **Obtain the number of rows and columns from a frame**
 - 4 **Create and initialise two temporary matrices having same size as video frame**
 $M_1(\mathbf{x}) = \text{zeros}(\text{row}, \text{column}); V_1(\mathbf{x}) = \text{zeros}(\text{row}, \text{column}).$
 - 5 **Initialise the two temporary matrices with pixel values**
 $M_1(\mathbf{x}) = B_1(\mathbf{x});$ // Compute mean value of frame
 $V_1(\mathbf{x}) = M_1(\mathbf{x}) \cdot M_1(\mathbf{x});$ // Square of mean value of value
 - 6 **Development of background model**
for $k = 2$ **to** T {
 $M_k(\mathbf{x}) = M_{k-1}(\mathbf{x}) + B_k(\mathbf{x});$
 $V_k(\mathbf{x}) = V_{k-1}(\mathbf{x}) + M_1(\mathbf{x}) \cdot M_1(\mathbf{x});$
}
 - 7 **Compute modelled background frame**
 $M_{BGM}(\mathbf{x}) = M_k(\mathbf{x}) / T;$ // mean value based background modelled frame
 $V_{BGM}(\mathbf{x}) = V(\mathbf{x}) / T;$ // Normalised background frame for threshold
 - 8 **Computation of initial threshold**
 $\text{Threshold} = \text{abs}((V_{BGM}(\mathbf{x}) - M_{BGM}(\mathbf{x}) \cdot M_{BGM}(\mathbf{x}))^{1/2})$

3.2 Foreground extraction and morphological filter

The pixels belonging to the foreground may have more deviation from the mean and these pixels did not appear isolatly in the frame, but they appear in blobs. For the first test frame, classify the pixel with a suitable threshold generated during background modelling. For each successive test (current) frame, the threshold is updated adaptively. During testing phase, some pixels are misclassified as false positive or false negative. Due to false negatives, the problem of holes in the interior region of valid object may arise. So, morphological filters like dilation and erosion are applied to overcome such issues. The binary morphology is used to remove unwanted regions which were classified as the background. Here, a squared shape structure element (with size 3×3) is performed well in morphological filtering operation. Due to tree leaves/branches, spouting or rippling water, the unrequired blobs are detected. Thus, these blobs have been removed using a threshold based connected component and blob labelling scheme. This threshold represented in terms of number of pixels. Here, results of highway frame sequence in binary and coloured segments have shown in Figure 2. The idea of segmentation of the

moving object is inherited from I2R Dataset (http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html).

In the given algorithms, we have used N test frames $\{F_1(\mathbf{x}), F_2(\mathbf{x}) \dots F_N(\mathbf{x})\}$, where $1 \leq N \leq \text{Total no of frames}$; and $\mathbf{x} = (x, y)$. These N frames are processed during testing phase. Here, the threshold is updated in terms of variance where the variance is represented in terms of expectation of the squared deviation of a pixel value and its mean value. It signifies how far a set of random variable (i.e., pixel) is spread out from its mean.

Algorithm 2: Pseudo Code for proposed pixel classification and adaptive learning based model updation

Input: Consider N no of test frames $\{F_1(\mathbf{x}), F_2(\mathbf{x}) \dots F_N(\mathbf{x})\}$, where $1 \leq N \leq \text{Total no of frames}$; and $\mathbf{x} = (x, y)$

Output: Compute background model and threshold

- Classify state of each pixel as motion (foreground) or stationary (background)
 - Update Threshold value for pixel classification with adaptive learning
- 1 **Read one frame at a time from N test frames** // See foreground modelling and classification in Section 2.2.
 - 2 **Convert RGB frame into Grey colour** // to optimise running time
 - for* ($n = 1$ to N) {
 - $f_n(\mathbf{x}) = \text{rbg2gray}(F_n(\mathbf{x}))$ // where $f(\mathbf{x})$ is grey colour frame and $1 \leq n \leq N$;
 - 3 **Foreground Modelling and Extraction**
 - if* ($n = 1$) {
 - $BGM_n(\mathbf{x}) = M_{BGM}(\mathbf{x});$ // background model frame w. r. t. mean
 - $V_1(\mathbf{x}) = V_{BGM}(\mathbf{x});$ // initialising value for threshold updation in step-7
 - $\text{Threshold}_n = \text{Threshold};$
 - } // end if
 - else*{
 - $V_n(\mathbf{x}) = V_{n-1}(\mathbf{x});$
 - $BGM_n(\mathbf{x}) = BGM_{n-1}(\mathbf{x});$
 - }
 - 4 **Initialise a frame difference matrices**
 - $\text{diff}_n(\mathbf{x}) = |f_n(\mathbf{x}) - BGM_n(\mathbf{x})|;$
 - // $\text{Diff}_i(\mathbf{x})$ matrix consist of absolute value of difference of respective pixels from test frame and BGM frame
 - 5 **Motion based pixel classification**
 - if* ($\text{diff}_n(\mathbf{x}) \geq \text{Threshold}_n$) {
 - $FG_n(\mathbf{x}) = \text{true};$ // pixel is in motion, i.e., part of foreground
 - }
 - else* {
 - $BG_n(\mathbf{x}) = \text{false}$ // pixel is static or part of background
 - } // end if
 - 6 **Boundary Identification of Moving Object**

$$F_{boundary}(x) = f_n(x) * FG_n(x) + BGM_n(x) * BG_n(x);$$

7 Adaptive learning based updation of background model

$$BGM_{n+1}(x) = BGM_n(x) + F_{boundary}(x);$$

$$V_n(x) = V_{n-1}(x) * F_{boundary}(x) \cdot \wedge 2;$$

$$BGM_{n+1}(x) = BGM_n(x) ./n;$$

$$V_{n+1}(x) = V_{n-1}(x) ./n;$$

8 Update adaptive learning based new threshold

$$Threshold_{n+1} = | V_{n+1}(x) - BGM_{n+1}(x) * BGM_{n+1}(x) | 1/2$$

9 Convert all required 1-D array into 2-D matrices












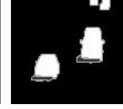






10 Display $FG_n(x)$ frame

} // end of for loop given in step-3

The difference of mean of the square and square of the mean is used as a threshold for pixel classification, i.e., $\sigma^2 = \frac{\sum x_i^2}{n} - \bar{x}$. Initially, the model learning is given in

Algorithm-1. Then again during testing phase, the background model is updated with the adaptive learning technique as shown in *Algorithm-2*. The dynamic issues are handled using the updation of the background model at run-time. Step-4 computes the difference frame and step-5 focuses on the classification of pixels. The threshold value computed from *Algorithm-1* is used as initial threshold for first current frame as shown in step-3 of *Algorithm-2*.

Figure 2 Binary and segmented result of proposed method on highway sequence (see online version for colours)

Frame No.	123	887	1170	1188	1484	1513
Original Frame						
Proposed Binary Result						
Proposed Segmented Result						

The detection quality of the proposed method is analysed through Figure 2 that presents binary and segmented results over highway frame sequence. These results indicate that the proposed method handles problem of moving tree leaves, moving shadow of tree leaves on the road. Apart from these, this method also handles the problem of illumination variation such as intensity variation on the road due to the sun light. Step-6 simply computes the boundary of the moving object. It helps in the maintenance of the background model as shown in step-7. This model is updated adaptively. Similarly, the threshold value is updated adaptively as depicted in step-8. For second frame updated

background model and updated threshold value is used for pixel classification. For each upcoming current frame these values are updated adaptively.

4 Experimental setup and analysis

This section presents the efficacy of the proposed work where the experimental work is carried out on various video sequences. All the considered frame sequences are publicly available for academic and research purpose. The considered sequences are described below.

- Microsoft's Wallflower dataset (Toyama et al., 1999): WavingTrees sequence with frame size 160×120 .
- I2R dataset (http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html): Campus, WaterSurface and fountain with frame size 160×128 .
- Change detection dataset (Goyette et al., 2012): canoe, overpass and highway sequences with frame size 320×240 and fountain02 sequence with frame size 432×288 .

In case of Campus, WavingTrees and highway sequence, it has been observed that the tree leaves were waving in the background. The Fountain and fountain02 frame sequence have water fountain in the background scene which is in the running stage. In the WaterSurface sequence, the water is spouting in the background. The canoe has flowing water and moving tree leaves in the background scene. The overpass sequence has waving tree leaves, flowing water and fountain in the background which generate more cluttered nature in the background. Due to such issues, handling the behaviour of background is very critical task. Here, the experimental work is carried out on MATLAB 2011b and Windows 8 on Intel (R) Core (TM) i5 processor with CPU having 1.70 GHz speed. The experimental results are evaluated in two categories: qualitative analysis and quantitative analysis.

4.1 Qualitative analysis

The qualitative results of all methods over various datasets are shown in Figure 3. This work has considered eight video sequences with highly dynamic background. The proposed work has tried to minimise the above mentioned issues.

The visual observation for considered sequences [row-wise: (Haines and Xiang, 2014) original frame and (Dordevic et al., 2015) ground truth] and comparative methods [row-wise: proposed, Lee and Lee (2014), Zhou et al. (2013), Jung (2009), KaKiNg (Ng and Delp, 2011), Haque et al. (2008), Stauffer and Grimson (1999) and Irani et al. (1994)] are shown in Figure 3. It shows visual results of all the methods.

4.2 Quantitative analysis

The performance of the proposed method over considered sequences is evaluated in terms of false positive, false negative, true positive and true negative. The qualitative analysis is explored through error, precision, recall, F-measure, accuracy, etc., as given below.

4.2.1 Error analysis

The error analysis of this work and its peer methods is presented in Figure 4. The following metrics are used for error analysis. The error analysis of proposed, Lee and Lee (2014), Zhou et al. (2013), Jung (2009), KaKiNg (Ng and Delp, 2011), Haque et al. (2008), Stauffer and Grimson (1999), and Irani et al. (1994) methods are shown in Figure 4.

- $FP_Error = FP * 100 / total_rows * total_columns;$
- $FN_Error = FN * 100 / total_rows * total_columns;$
- $Error = FP_Error + FN_Error;$

The pixels are classified falsely whether these are classified as false positive or false negative. We divide total pixels classified as false positive with the total pixels of the frame. It is also known as false positive error. Similarly, we compute false negative error.

Figure 3 Qualitative results of all methods (see online version for colours)

Sequence/ Frame No.	I2R Dataset			Wallflower Dataset	ChangeDetection Dataset			
	Campus 372	WaterSurfa ce 499	Fountain 202	WavingTrees 247	fountain02 745	Canoe 963	Overpass 2372	Highway 1171
Original								
Ground Truth								
Proposed								
Lee [18]								
Zhou [17]								
Jung [7]								
KaKiNg [15]								
Haque [10]								
Stauffer [8]								
Irani [16]								

4.2.2 Performance analysis

The performance of the proposed work is evaluated and compared with considered state-of-the-art methods using some metrics (Yadav et al., 2014; Yadav and Singh, 2015; Yilmaz et al., 2004) as given below. Here, TP represents valid object pixel which is correctly classified as part of the foreground. The FP depicts invalid pixel which is falsely classified as part of the foreground. The TN represents invalid pixel which is correctly classified as the background. The FN is considered as valid object pixel which is falsely classified as the background. The matrices are given below.

- Precision (P) = $TP / (TP + FP)$
- Recall (R) = $TP / (TP + FN)$
- F-measure (F_1 -score) = $2 * P * R / (P + R)$
- Accuracy (η) = $(TP + TN) / (TP + FN + FP + TN)$

Precision represents the extraction of relevant pixels whereas recall represents relevant foreground pixels. So, the large value of precision and recall is preferred for better quality. The F_1 -score is computed by computing the harmonic mean of P and R. Similarly, large value of F_1 -score is considered for better detection quality. Its worst and best value is represented as zero and one, respectively.

Figure 4 Error analyses of all methods (see online version for colours)

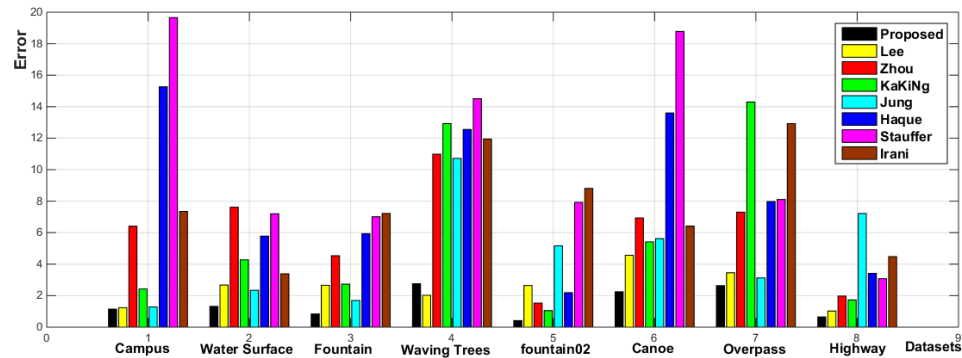


Table 1 shows the value of average results of all sequence. The accuracy refers, how closely the measured value agreed with the true value. Along with varying threshold, the receiver operating characteristic (ROC) curve computes the performance of the binary classifier. The threshold value is tuned and it provides smoothness for ROC plot. The proposed method performs better with the grey channel and takes minimum computational time. In ROC-curve, the proposed result is depicted in the black line which shows the outstanding performance of the proposed method.

Table 1 Average value of precision, recall and F-measure of each method

<i>Model</i>	<i>Precision</i>	<i>Recall</i>	<i>F₁-score</i>
Proposed	0.8482	0.9208	0.8721
Lee and Lee (2014)	0.7471	0.5890	0.6433
Zhou et al. (2013)	0.4281	0.4550	0.3950
Ka Ki Ng (Ng and Delp, 2011)	0.6019	0.1854	0.2584
Jung (2009)	0.7244	0.6555	0.6505
Haque et al. (2008)	0.3486	0.7967	0.4628
Stauffer and Grimson (1999)	0.3412	0.8280	0.4589
Irani et al. (1994)	0.3518	0.7217	0.4310

4.2.3 Area under ROC curve (AUC)

A common metric, AUC (St. Charles et al., 2015; Sermnet et al., 2013) measures the performance of the ROC curve, where, $0 \leq \text{AUC} \leq 1$. This section uses AUC value by approximating the region of the trapezoid under the curve for a function $f(x)$. The concept of AUC is inherited from St. Charles et al. (2015) and Sermnet et al. (2013).

$$\text{AUC} = \int_a^b f(x)dx \approx (b-a) \left[\frac{f(a)+f(b)}{2} \right]$$

The roc-curve of proposed, Lee and Lee (2014), Zhou et al. (2013), Jung (2009), KaKiNg (Ng and Delp, 2011), Haque et al. (2008), Stauffer and Grimson (1999), and Irani et al. (1994) methods are drawn in Figure 5. It clearly depicts the performance of all the methods. The ROC analysis is also computed in terms of AUC value that represents the area under the ROC curve. Based on the area under ROC curve, the proposed method shows maximum area under the ROC curve that demonstrates better performance of the proposed method.

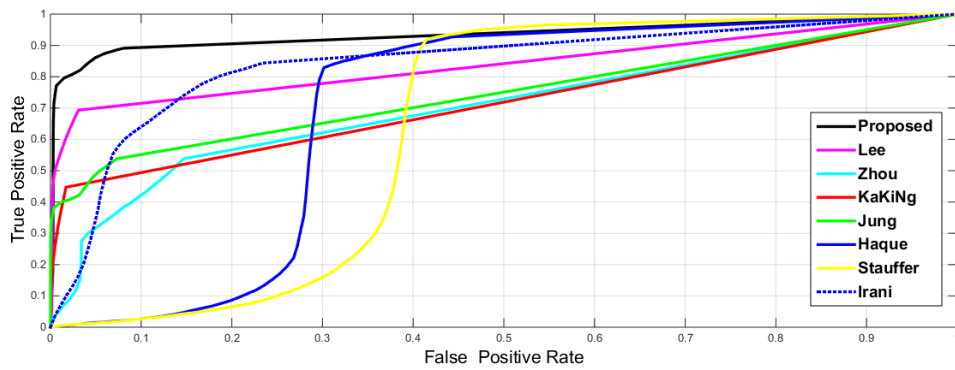
Figure 5 Analysis of roc curve for all methods (see online version for colours)

Table 2 Average analysis of AUC, accuracy and execution time for each method

<i>Model</i>	<i>AUC</i>	<i>Accuracy</i>	<i>Time (sec) per frame</i>
Proposed	0.9331	0.9837	0.1286
Lee and Lee (2014)	0.7451	0.9713	1.0824
Zhou et al. (2013)	0.5889	0.9150	22.1984
Ka Ki Ng (Ng and Delp, 2011)	0.4482	0.9338	0.1491
Jung (2009)	0.7465	0.9538	2.9834
Haque et al. (2008)	0.7606	0.8652	0.5639
Stauffer and Grimson (1999)	0.8067	0.8454	45.8124
Irani et al. (1994)	0.7747	0.8889	4.0581

The AUC value and accuracy of all methods are shown in Table 2. The AUC value and maximum accuracy demonstrate better detection quality of the proposed method. It also shows the execution time, i.e., running time of each method over considered frame sequences. This table shows the average value of each method over all sequences.

4.2.4 Jaccard similarity index

For evaluating the detection quality, we applied another performance measuring metric, i.e., Jaccard index. This metric aims to measure the similarity between the ground truth image and the detected result in the binary representation. This metric is also used to compute the overlap rate. The detected result and corresponding ground truth are represented as $FG(x)$ and $GT(x)$. Mathematically, the Jaccard index is defined as the ratio of the intersection of detected result of the frame $FG(x)$ and true region of the ground truth $GT(x)$ to the union of detected result $FG(x)$ and the ground truth area $GT(x)$. It is to be noted, $FG(x)$ and $GT(x)$ are also written as FG and GT respectively. So, the Jaccard index is computed using the following equation.

$$J(FG, GT) = \frac{|FG \cap GT|}{|FG \cup GT|}$$

This index is simply compared the retrieved region and the ground truth mask at the pixel level in the frame. In this work, we have completed the Jaccard index of each method over every frame sequence. Then we have evaluated the mean Jaccard Index which is simply the mean of all value of Jaccard Index computed over all frame sequences by a particular method. The mean value of Jaccard index is shown in Table 3.

Table 3 Jaccard similarity index for each method

<i>Methods</i>	<i>Proposed</i>	<i>Lee and Lee (2014)</i>	<i>Zhou et al. (2013)</i>	<i>Jung (2009)</i>
Jaccard index	0.7323	0.4800	0.2677	0.1776
<i>Methods</i>	<i>Ka Ki Ng (Ng and Delp, 2011)</i>	<i>Haque et al. (2008)</i>	<i>Stauffer and Grimson (1999)</i>	<i>Irani et al. (1994)</i>
Jaccard index	0.4696	0.2404	0.3221	0.3003

Table 3 represents the mean value of Jaccard index of each method. Table 3 depicts that the proposed method has maximum value as compare too the considered peer methods. It clearly indicates outranking performance of the proposed method.

4.2.5 Observations

Based on above experimental results and analysis, the following observations are obtained from Figures 2–5 and Table 1–2.

- Figure 2 clearly shows the segmented and binary form of moving vehicles from highway frame sequence. The proposed result signifies that the detection quality of running vehicles is better against peer methods and also depicted in Figure 3.
- According to Figure 4, the proposed method presents minimum error for each frame sequence as compared to considered state-of-the-art methods. The average total error is also minimum against others.
- Table 1 shows outstanding detection quality of proposed method to achieve highest average value of F-measure over all frame sequences. The proposed method also shows maximum accuracy value as shown in Table 1.
- The ROC curve for all methods is presented in Figure 5. According to the Figure 5, the overall performance of suggested method is better.
- The performance of ROC curve is also measured in terms of AUC value. The proposed method also has maximum AUC value as represented in Table 2.
- Table 2 depicts the maximum accuracy of the proposed method. It also demonstrates the minimum running speed of the proposed method over its peer methods.
- Table 3 clearly depicts the outranking performance of the proposed method as it has maximum Jaccard index value for similarity measure as compare to peer methods.

4.3 Advantages and limitations

The main advantage of proposed work:

- the proposed method is similar to traditional methods but it has fast execution speed that may allow its applicability for real-time applications
- ability to deal with challenging cluttered background scene
- low computational cost and easy to implement
- does not require any external parameter as used in various methods
- better performance against peer methods.

Apart from above contributions toward the detection quality of the proposed work, this work has some limitations which are mentioned below.

- This work performs well with coloured video frames but does not generate good results in thermal video frames. In high motion or variation in background, the distribution of pixel value is too either high of too low, but in thermal environment

the distribution is low. This limit shows its applicability towards coloured video frames.

5 Importance of cloud-IoT and BigData in new era of visual surveillance

In the current era, the surveillance system is moving towards internet-of-things (IoT), cloud and BigData for observing the real-time situation through sensor based technology. According to the literature, very less number of cloud-IoT based services is available for handling the visual surveillance system. The visual surveillance system is able to protect millions of people around the globe, valuable places such as monumental and assets worldwide with innovative cloud oriented video surveillance along with internet of things video (IoTV) solutions. The IoTV platform delivers vision oriented services to end users. It also allows to store video in secure and fast way on a device (i.e., network video recorder or as a hybrid option or double redundant global cloud). The cloud oriented IoTV system enables centralised monitoring of targeted objects and observe their activities. Such surveillance area applied to healthcare domain, traffic, sea, underwater surveillance, border, restricted zones, hill-area, flood-area, fired zone, and deep zones, etc. For this reason, such IoTV based systems are very helpful for the air fighter, army, or military forces. These systems have capability to wield voluminous streaming of video data captured from various dispersed sensory devices [closed-circuit television camera (CCTV), internet protocol (IP), raspberry-pi or IoT devices]. In current era of upcoming technologies, cloud-IoT oriented services are investigated for handling the BigData. The combination of cloud-IoT on BigData enables the monitoring of moving objects, and their behavioural activities for managing operations of IoTV surveillance.

5.1 Real-time applications for visual surveillance

The real-time based requirement of various kinds of professional and commercial surveillance activities can be handled using the proposed method. The object detection can also be applicable for border, port, coastal, airport surveillance and other wide range of police applications. For example, Yahoo Labs (2010), the proposed method can also enable a coastguard to monitor large stretches of waterway and coastline for detection of fishing, smuggling, or illegal intrusion etc. Some other open problem oriented video surveillance applications are given below (Hu et al., 2013; Yilmaz et al., 2004; Yadav et al., 2015; Sharma et al., 2016).

- to identify and detect the suspect behind foliage in poor visibility or darkness
- to identify and detect the suspect like terrorist at border and generate the alertness
- the search operations can be easily accelerated in crowded areas
- to detect and track uncontrolled vehicle on highways or in forest areas
- to monitor the manufacturing and production equipment inside an industry
- to detect cattle or thief in the agriculture land in Indian village, fields or gardens, etc.
- to improve the driver assistance system in vehicle for monitoring and analysis of vehicles.

6 Conclusions and future directions

The proposed scheme is potentially useful for real-time computer vision applications. This work is more suitable for outdoor video surveillance applications. The detection results have been evaluated qualitatively and quantitatively then compared with considered state-of-the-art methods using a wide range of problematic real-time video sequences. This work is an attempt to automate the process of decision making scheme on the threshold value. The proposed work exhibits considerable robustness against noise, and dynamic nature of the background. The effectiveness of the proposed method has been proved under various challenging situations of motion in the background. The visual observation and error, precision, recall, f-measure and accuracy demonstrate the better detection quality of our method as compare to its peer methods. The maximum value of the Jaccard index demonstrates the outranking performance of the proposed work in terms of similarity measure against peer methods. Moreover, the proposed method is easy to implement, leading to low computation and fast experimental running speed.

The authors are also planned to extend the proposed work to detect, count the number of objects and track their path also. The authors will try to perform such work over a cloud based online video capturing scheme for detection and tracking of object in the real-time video surveillance system.

Acknowledgements

The authors are very grateful to Microsoft's Wallflower (Toyama et al., 1999), I2R (http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html), and ChangeDetection (Goyette et al., 2012) for providing video data in terms of frame sequences and providing support in the performance analysis of this work. In future, we will apply deep learning approaches for moving object detection or tracking.

References

- Bloisi, D.D., Previtali, F., Pennisi, A., Nardi, D. and Fiorini, M. (2017) 'Enhancing automatic maritime surveillance systems with visual information', *IEEE Transactions on Intelligent Transportation Systems*, April, Vol. 18, No. 4, pp.824–833.
- Braham, M. and Droogenbroeck, M.V. (2016) 'Deep background subtraction with scene-specific convolutional neural networks', International Conference on Systems, Signals and Image Processing, IEEE, July, pp.1–5.
- Christiansen, P., Nielsen, L.N., Steen, K.A., Jorgensen, R.N. and Karstoft, H. (2016) 'DeepAnomaly: combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field', *Sensors*, November, Vol. 16, No. 11, pp.1–21.
- Dordevic, Z.P., Graovac, S.G. and Mitrovic, S.T. (2015) 'Suboptimal threshold estimation for detection of points-like objects in radar images', *Eurasip Journal on Image and Video Processing, Springer Open Access Journal*, February, Vol. 5, pp.1–12.
- Druzhkov, P.N. and Kustikova, V.D. (2016) 'A survey of deep learning methods and software tools for image classification and object detection', *Pattern Recognition and Image Analysis*, January, Vol. 26, No. 1, pp.9–15.

- Goyette, N., Jodoin, P.M., Porikli, F. and Ishwar, P. (2012) 'Changetection.net. A new change detection benchmark database', *Proc. IEEE workshop on Change Detection at CVPR*, June, pp.1–8.
- Haines, T.S.F. and Xiang, T. (2014) 'Background subtraction with Dirichlet process mixture model', *IEEE Trans. on Pattern Analysis and Machine Intelligence*, April, Vol. 36, No. 4, pp.670–683.
- Haque, M., Murshed, M. and Paul, M. (2008) 'On stable dynamic background generation technique using Gaussian mixture models for robust object detection', *5th Int. Conference on Advanced Video and Signal Based Surveillance*, IEEE, September, pp.41–48.
- Hu, W., Zhou, X. and Li, W. et al. (2013) 'Active contour-based visual tracking by integrating colors, shapes, and motions', *IEEE Trans. on Image Processing*, May, Vol. 22, No. 5, pp.1778–1792.
- Huang, M., Chen, Y. and Ji, W. (2015) 'Accurate and robust moving-object segmentation for telepresence systems', *ACM Transactions on Intelligent Systems and Technology*, May, Article 17, Vol. 6, No. 2, pp.1–28.
- I2R Dataset [online] http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html (accessed 1–23 December 2015).
- Irani, M., Rousso, B. and Peleg, S. (1994) 'Computing occluding and transparent motions', *Int. Journal of Computer Vision*, February, Vol. 12, No. 1, pp.5–16.
- Javed, S., Oh, S.H. and Heo, J. et al. (2014) 'Robust background subtraction via online robust PCA using image decomposition', *Proc. RACS, ACM Digital Library*, USA, October, pp.105–110.
- Jung, C.R. (2009) 'Efficient background subtraction and shadow removal for monochromatic video sequences', *IEEE Transactions on Multimedia*, April, Vol. 11, No. 3, pp.571–577.
- Lee, D.S. (2005) 'Effective Gaussian mixture learning for video background subtraction', *IEEE Trans. On Pattern Analysis Machine Intelligence*, March, Vol. 27, No. 5, pp.827–832.
- Lee, S. and Lee, C. (2014) 'Low-complexity background subtraction based on spatial similarity', *Eurasip Journal on Image and Video Processing*, Springer, June, Vol. 1, No. 30, pp.1–16.
- Li, Y., Jha, D.K., Ray, A. and Wettergren, T.A. (2017) 'Information fusion of passive sensors for detection of moving targets in dynamic environments', *IEEE Transactions on Cybernetics*, January, Vol. 47, No. 1, pp.93–104.
- Lin, L., Xu, Y. and Liang, X. et al. (2014) 'Complex background subtraction by pursuing dynamic Spatio-temporal models', *IEEE Transaction on Image Processing*, Vol. 23, No. 7, pp.3191–3202.
- Liu, W., Yu, H. and Yuan, H. et al. (2014) 'Effective background modelling and subtraction approach for moving object detection', *IET Computer Vision*, April, Vol. 9, No. 1, pp.13–24.
- Ng, K.K. and Delp, E.J. (2011) 'Background subtraction using a pixel-wise adaptive learning rate for object tracking initialization', *Proceedings of SPIE Digital Library Visual Information Processing and Communication II*, January, pp.1–9.
- Pedersolib, M., González, J. and Sanfeliud, A. (2015) 'Combining where and what in change detection for unsupervised foreground learning in surveillance', *Pattern Recognition*, Elsevier, March, No. 48, pp.709–729.
- Sermnet, P., Kavukcuoglu, K., Chintala, S. and Lecun, Y. (2013) 'Pedestrian detection with unsupervised multi-stage feature learning', *IEEE conference on CVPR*, June, pp.3626–3633.
- Sharma, L., Yadav, D.K. and Singh, A. (2016) 'Fisher's linear discriminant ratio based threshold for moving human detection in thermal video', *Infrared Physics and Technology*, Elsevier, July, Vol. 78, pp.118–128 [online] <http://www.sciencedirect.com/science/article/pii/S1350449516301621>.
- St. Charles, P., Bilodeau, G. and Bergevin, R. (2015) 'SuBSENSE: a universal change detection method with local adaptive sensitivity', *IEEE Transactions on Image Processing*, January, Vol. 24, No. 1, pp.359–373.

- Stauffer, C. and Grimson, W.E.L. (1999) 'Adaptive background mixture models for real-time tracking', *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, June, Vol. 2, pp.246–252.
- Su, T. and Hu, J. (2004) 'Background removal in vision servo system using Gaussian mixture model framework', *Proc. of ICNSC*, IEEE, Taiwan, March, pp.70–75.
- Toyama, K., Krumm, J., Brumitt, B. and Meyers, B. (1999) 'Wallflower: principles and practice of background maintenance', in *7th International Conference on Computer Vision*, *IEEE Computer Society Press*, 20–27 September, Vol. 1, pp.255–261.
- Wang, Y., Liang, Y. and Zhang, L. et al. (2012) 'Adaptive spatiotemporal background modelling', *IET Computer Vision*, September, Vol. 6, No. 5, pp.451–458.
- Wang, Y., Luo, Z. and Jodoin, P.M. (2016) 'Interactive deep learning method for segmenting moving objects', *Pattern Recognition Letters*, in Press, September, pp.1–10 [online] <http://dx.doi.org/10.1016/j.patrec.2016.09.014> (accessed 22–23 April 2017).
- West, D.H.D. (1979) 'Updating mean and variance estimates: an improved method', *Communications of the ACM*, September, Vol. 22, No 9, pp.532–535.
- Yadav, D.K. and Singh, K. (2015) 'Moving object detection in real-time visual surveillance using background subtraction technique', *Journal of Information Assurance and Security*, March, Vol. 10, No. 2, pp.79–99.
- Yadav, D.K., Sharma, L. and Bharti, S.K. (2014) 'Moving object detection in real-time visual surveillance using background subtraction technique', *14th Int. Conf. on Hybrid Intelligent Systems*, IEEE, Kuwait, December, pp.79–84.
- Yahoo Labs, (2010) *AUC Definitions and Calculations*, Technical Report, March, pp.1–28.
- Yilmaz, A., Ali, X. and Shah, M. (2004) 'Contour-based object tracking with occlusion handling in video acquired using mobile cameras', *IEEE Trans. On Pattern Analysis and Machine Intelligence*, September, Vol. 26, No. 11, pp.1531–1536.
- Zhou, X., Yang, C. and Yu, W. (2013) 'Moving object detection by detecting contiguous outliers in the low-rank representation', *IEEE Trans. on Pattern Analysis and Machine Intelligence*, March, Vol. 35, No. 3, pp.597–610.