Histogram-based adaptive learning for background modelling: moving object detection in video surveillance

Lavanya Sharma*
Govind Ballabh Pant Engineering College,
Pauri, Uttarakhand, Pin-246194, India
Email: shm.lavanya@gmail.com
*Corresponding author

Dileep Kumar Yadav
School of Computer and Systems Sciences,
Jawaharlal Nehru University,
New Delhi, Pin-110067, India
Email: dileep252000@gmail.com

Abstract: The detection of moving object in the presence of complex or cluttered background is a very critical challenge. The moving object may be a person, patient, vehicle, animal or any tissue inside body in medical domain. In this context, this work has proposed a robust background subtraction method for resolving illumination variation and motion-based problems. Initially, this work has developed a background modelling method using initial few frames in training stage. In testing stage, a foreground modelling method is investigated that is able to detect moving object in video frames. In testing stage, this work classify moving pixel with a suitable threshold and update the background using appropriate learning rate. The learning rate is updated through histogram of classified resulting frame and background model. Finally, morphological filters and image processing techniques are applied to improve the detection quality. The employed method also demonstrates how it can be improved using adaptive learning rate-based controlling scheme and the incorporation of feedback-based model updating scheme. It clearly depicts strength of proposed method in handling illumination variation problems and also eliminating moving environmental effects. This method presents significant performance in comparison with considered state-of-the-art methods.

Keywords: histogram; adaptive learning rate; background modelling; background subtraction; moving object detection; healthcare; cloud; IoT; video surveillance.


Biographical notes: Lavanya Sharma is pursuing her PhD from GBPEC Pauri, Garhwal, Uttarakhand. She did her MTech in Computer Science from MDU Rohtak, Haryana. She has done MSc in Information Technology after completing PGDCA. She also did her BSc from University of Delhi, India. She has published many international conference, and also papers in reputed
1 Introduction

In the area of computer vision, investigating a robust method for motion-based object detection is a very challenging job. The increasing need of consumer electronic products and technological development, many cameras are installed worldwide to provide security surveillances. The growth has a huge augmentation of data that is very difficult to store or handle manually (Zhou et al., 2013a; Reddy et al., 2013).

In most of cases, moving object may move intermittently (for example, moving vehicle at traffic light). Sometimes uncasting background regions may exhibit cluttered nature (e.g., flowing water, moving tree leaves, flags, etc.) (Bouwmans et al., 2014; Mittal et al., 2009; Haines and Xiang, 2014; Barnich and Van Droogenbroeck, 2011). It is a two-class problem. Some useful parameters are required that control adaptation rate and background model. It is again a very challenging problem to adjust parameters for particular problem like illumination variation, dynamic background scene, and camouflage or altogether presented in a scene (Barnich and Van Droogenbroeck, 2011; St-Charles et al., 2015; Yadav and Singh, 2015b).

1.1 Future of object detection in healthcare surveillance and other applications

Recent paradigm shift is being perceived in medical science, where the focus is shifting away gradually from operation to real data using cloud and IoT to observe the outstation patient living in remote areas. It is highly affecting the decision-making process of patient observation. In literature, various cloud services have been suggested for handling BigData (Sharma, 2016; Sharma et al., 2015b) or data generated through IoT. The cloud-based IoT also enables centralised monitoring of patient, working staff, and other activities and manage all the operations taking place in any healthcare centre or hospital. It also detects and tracks the behaviour of patient during day and night. With the help of thermal imaging (Yadav and Singh, 2015a), the proposed work is applicable to detect and track hospitalised patient’s behaviour during night time even when light is turned off. Such systems may bring exceptional operational efficiency to manage day-to-day clinical practices in hospitals and surgical centres. Nowadays, researchers and software industries
are focusing on such solutions around IoT that has the ability to provide better control over the operational process as well as reducing time for better care. The reliability, safety, travelling-time and cost efficiency are the major key-features achieved with IoT and cloud services (Ma et al., 2012; Sharma, 2015; Sharma et al., 2015a) in the healthcare sector. In addition, IoT and cloud can easily integrate data from consumer or client devices such as fitness band into hospital systems or healthcare chip in patients for gathering more data. Based on such data, doctors are able to deliver better care as well as services. The IoT in healthcare market provides various devices in market some of them consists of Wearable External Medical Devices, Implanted Medical Devices, and Stationary Medical Devices.

1.2 Challenging issues and problem domain

The object detection is applicable in various kind of real-time-based applications such as patient monitoring, hospitality, traffic monitoring, indoor-outdoor, shoplifting in retail, coastal surveillances, defence, border security system, police and law enforcement, night vision surveillances, etc. It can be applicable in multimedia and its applications, robotics, manufacturing, packaging, artificial intelligence and automatic computer algorithm development. In general object detection is a concept of evaluating the state of motion-based pixel (Yadav and Singh, 2015b). Various methods are available in literature for monitoring, detection, segmentation or tracking of object automatically in videos. The modern algorithms has categorised in three main assumptions:

- a background model of video frames is constructs with initial few frames
- classification of every pixel of each frame as part of foreground or background frames
- update the background model.

In real-time video surveillance applications, the background subtraction is extensively used since couple of years, yet still an open and challenging area of research due to the following challenging issues as focused in this paper are:

- Intensity variations: particularly with gradual or sudden lighting and illumination changes (such as Camouflage, ForegroundAperture, LightSwitch, Bootstrap, MovedObject).
- Dynamic backgrounds: In nature, background scene environment is not always static. Sometimes background scene is highly dynamic, e.g., spouting/rippling water, moving tree leaves or branches, heavy rain, etc.

This work focused on above mentioned challenges with the development of the background model. In testing phase, this model is updated recursively using histogram-based learning rate. The basic idea of histogram was acquired from Yao and Ahmad (2009) for background modelling in hue-saturation-value colour space. So, the proposed work effectively handles dynamic background and illumination variation problems. Some segmentation results generated by the proposed method for the challenging frame sequences are exhibited in Figure 1.
The description of challenging scenarios shown in Figure 1 can be illustrated as:

1. a randomly moving computer screen with gradual intensity variations *(first column)*
2. sudden light changing from *on* state to *off* state then again switch on *(second column)*
3. fluctuation of light and shadow around moving persons *(third column)*.

Some results generated by the proposed algorithm can be seen in Figure 1 as: row 1 and row 2 represent background frames, row 3 represents targeted foreground frame, and fourth row represents the colored segmented results of the proposed algorithm.

This paper has presented a robust algorithm that is able to handle above problems. The main objective of proposed work aims to detect meaningful, moving region and to avoid static and dynamic components of background scene from the background. This work also improves detection quality of foreground object with morphological filters and other image processing techniques which are explored later. This work also present how a system has been further improved by combined approach of statistical features and histogram-based background model. Here, the histogram is useful for evaluating effective learning rate. The histogram is made up of bins where each bin consists of a certain intensity value range. A bin simply represents the number of pixels associated to it. In reality, one-dimensional (1-D) histogram is much more easy because, in a 1-D histogram, the height of the bars represent the count, whereas the height of a two-dimensional (2-D) histogram requires the count for that histogram bucket is greater than zero, but for a
three-dimension and it is difficult to visualise. So, most of the researcher prefers 1-D histogram which act as a linear array. Hence, takes less time. In reality, processing over red-green-blue colour frame generate more accurate information but takes more processing time, but the proposed implementation over grey colour also gives good results. The running time over grey colour is very minimum as used in proposed work.

In Section 2, various methods have been summarised briefly and discussed about the background modelling and object detection in video sequences. In Section 3, this work explored the proposed method. The experimental results and analysis have discussed in Section 4. This section also demonstrates advantages and limitation of proposed work. Finally, Section 5 provides conclusion and presents an idea about future work. Before unfolding our proposed algorithm, next section reviewed some novel and existing works available in literature.

2 Related work

In literature, various methods are available for motion-based object detection, segmentation, and tracking of moving object in each videos frame. The wide range of previous approaches has been available in the comprehensive survey of Bouwmans et al. (2014), Mittal et al. (2009) and Haines and Xiang (2014). Further, complicated comparisons between different studies are presenting the fact that classification performance depends on how the classifier is trained and the test data used to evaluate accuracy.

The detection-based appearances information has achieved great success in visual surveillances system, and the motion-based information is helpful in detection of moving object. In literature, various methods and algorithms have been proposed using the background subtraction technique. Among these, Gaussian mixture model (GMM)-based probabilistic model proposed by Stauffer and Grimson (1999) for motion-based object detection and extended by Lee (2005), Yadav (2014) and Haque et al. (2008) these models consists of background modelling and distribution of pixel values over time at each pixel with weighted mixture of Gaussian’s. In visual background extractor (ViBe) (Barnich and Van Droogenbroeck, 2011), parameters of GMM algorithm is determined automatically using particle swarm optimisation. A combined approach of GMM with region-based algorithm based on colour histogram and texture information has proposed. Cheng et al. (2011) have developed an improved version of ViBe named as ViBe-BMRI that detects illumination change and achieves significant in segmentation of moving object with accuracy. In research of Ghosh et al. (2015), a Dirichlet process Gaussian mixture model has employed with background subtraction that estimates distribution over each pixel. The problem of over and under fitting resolved by using Bayesian method. Some indoor application consists of sudden complex changes in background scene due to light have not resolved by this method. St-Charles et al. (2015) have proposed a new approach method high-speed parallel implementation with local adaptive sensitivity. This method works on the principle of pixel level feedback scheme that automatically adjust sensitivity and model. Dordevic et al. (2015) estimates update rate of suboptimal threshold. The main drawback of this work is that it generates lots of miss-classification when a frame obtained maximum number of filtering technique. Lee and Lee (2014) gives a spatial similarity based background subtraction method that have low complexity and focused on the problem of shadow that resolved by method of structural similarity.
Lin et al. (2014) proposed a method that incrementally updates the substances with each test frame to cope with various types of changes in scene as well as moving foreground. Ghosh et al. (2015) have investigated a robust time-frequency based approach with the support of pseudo-Wigner-Ville distribution assisted Rényi entropy for motion based vehicle detection. The background subtraction method proposed by Suo and Wang (2008) have improved the detection quality of moving object using adaptive Gaussian method based update rate. Similarly, Zhang et al. (2013) have developed a pixel intensity level migration probability map. This method effectively improves the adaptability of exponentially weighted moving average based background subtraction modelling over various surveillances applications. Panda and Meher (2016) have proposed a colour difference histogram based algorithm for background subtraction using fuzzy technique. Here, a fuzzy c-means clustering method has employed to reduce feature space by minimising the bins in histogram. Liu and Pados (2016) have developed an L1-PCA (principle component analysis) method for object detection where, this work computes the principal components under an L1-norm maximisation criterion for low-latency video surveillance applications.

3 Proposed work

The main motivation behind proposed work is the existing issues with background subtraction techniques. Main problem in background subtraction approach is how to choose a suitable learning rate and how to construct an efficient background model and how to select a threshold for better pixel’s classification. This work has introduced a histogram-based adaptive learning rate that updates the model during testing phase. The proposed work also demonstrates an outstanding improvement in the results in gradual illumination and dynamic variation. The working steps of proposed method have shown in Figure 2.

The working steps of proposed work are described as: Figure 2 clearly depicts that we have considered few initial frames (approx. 200) for background modelling. All the frames available in dataset are assumed as testing frame or we can say current frames. Here, we have applied background subtraction technique to compute the difference of current frame and background modelled frame. The threshold has generated during testing phase and updated iteratively for each frame. This work has suggested and applied a suitable threshold that is computed using average of current pixel value and mean value of background model. The learning rate at which the proposed method adapts to the changes or variations of the background. The low value of learning rate indicates slow adaptation and high value achieves quick adaptation of scene variations. This rate is updated using histogram of test frame and histogram of background model. The learning rate is used for updating the background model as presented in step 7 and step 8 of Algorithm 2. It also helps in eliminating insignificant objects generated due to variation and repetitive motion in the background such as gradual change in the intensity of light, waving tree leaves, etc. The proposed work also assumes that a pixel switches between background and foreground states infrequently. In such cases the pixel does not retain its status as background or foreground continuously. Then, it is assumed as part of background. Then, we have applied some image processing techniques to improve detection quality and avoid unusual blobs or noisy pixels. The detected object may fall in
L. Sharma and D.K. Yadav

the category of false negative alarm then such interior part can be filled using hole-filling techniques as discussed in detention in the next Section 3.3.

**Figure 2** Working steps of proposed histogram-based background subtraction method (see online version for colours)

3.1 Development of background model (training stage)

Conventionally, the first frame or simple average of initial few frames are considered for background modelling that taken from Yadav and Singh (2015b). This type of model is susceptible dynamic nature of background. Lots of solutions are available in literature, but are computationally expensive. Therefore, to resolve above mentioned problems and limitations, a very simple algorithm (refer Algorithm 1) for development of background model (Skifstad and Jain, 1989; Yadav et al., 2014) is suggested. Let, we have assume, $B_k(x)$ is the background model, and it is developed using initial $T$ frames:

$$B_{k}(x) = B_{k-1}(x) + f_k(x)/T$$

for $b = 1: T$

end

where $b$ is initial background frame from 1 to $T$

Here, $B_k(x, y)$ represent intensity of a pixel at location $(x, y)$ of a background model and $f_k(x, y)$ represents pixel’s intensity at location $(x, y)$ for $k^{th}$ frame used for background model. Generally, most of the articles have considered 20 to 50 initial frames for background modelling. So, for handling more complex background, we need more
number of training frames. Therefore, in this work, we have taken 200 (i.e., \( T = 200 \)) initial frames have taken for better modelling. Initially, for first frame \( (t = 1) \), the background model \( (B_1) \) was unavailable, i.e., we have considered a frame with zero pixel values.

The proposed threshold has computed by averaging of summation of mean value \( (\mu_B) \) of \( B_k(x, y) \) and current pixel \( f(x, y) \) as shown in equation (2) (also see, step 6 of Algorithm 1):

\[
T = \left[ f_k(x, y) + \mu_B(x, y) \right]/2
\]

With this threshold as computed in equation (2), a pixel can be broadly classified as a moving foreground object or background object. The adaptive threshold is used for removal of non-moving pixels from the scene and better classification of pixels. The idea of proposing a new threshold has inherited from Skifstad and Jain (1989) and Yadav et al. (2014).

**Algorithm 1** Background modelling algorithm

| Input: Consider \( T \) no of initial frames \( \{f_1(x), f_2(x) \ldots f_{T-1}(x), f_T(x)\} \), where \( T = 200 \); and \( x = (x, y) \) |
| Output: Development of background model and threshold |

- A initial background model (BGM) frame, i.e., \( B_1(x) \)
- 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
3. Obtain the number of rows and columns from a frame
4. Create and initialise variables and temporary matrices having same size as video frame
   \( \alpha = 0.01; \) //Initial Learning rate
   \( B_0(x) = \text{zeros (row, column)}; \)
5. Development of background model
   \( \text{for } k = 1 \text{ to } T \{ \)
   \( B_k(x) = B_{k-1}(x) + f_k(x) / T; \)
   \( \} \) // here, \( B_k \) is modelled background frame
6. Compute initial threshold
   - Read first frame, i.e., \( f_1(x) \) // mean value-based background modelled frame
   - Threshold = \( \text{double } \{[f_1(x) + \text{meanValue } (B_0(x))] \times 1/2; \}

### 3.2 Background maintenance and foreground modelling (testing stage)

In previous section, the background model has been constructed using initial few frames. The background modelling has shown in Algorithm 1. In Algorithm 1, an initial threshold and background model have been developed. In this section, we are discussing about the maintenance of background model, learning rate, and threshold. The moving object’s pixels have more deviation from the mean. Instead of appear isolated in the frame,
motion-based pixels appear in blobs. For the first test frame, this work classify such pixels with a threshold generated during background modelling and for each successive test frame it is updated as shown in Algorithm 2.

To overcome issues of gradual illumination variation and motion in background, the proposed background model is updated at run time for each forthcoming test frame. The proposed learning rate $\alpha$ is used to resolve dynamic issues available in scene through the updation of background model. The adaptive learning rate is defined as:

$$\alpha = \frac{1}{2} \left\{ \text{mean} \left( 1 - \alpha \right) / \text{histSum}_n(x) + \alpha \right\}$$  \hspace{1cm} (3)

$$\text{histSum}_n(x) = \text{histFG}_n\left( \text{FG}(x) \right) + \text{histBG}_n\left( \text{BGM}_n(x) \right)$$  \hspace{1cm} (4)

Algorithm 2  \hspace{0.5cm} \text{Pixel classification and model updation}

\begin{itemize}
  \item \textbf{Input:} Consider $N$ no of test frames {$f_1(x), f_2(x), ..., f_{T-1}(x), f_T(x), f_{T+1}(x), ..., f_N(x)$}, where, $N = \text{Total no of frames}$; and $x = (x, y)$
  \item \textbf{Output:} Classify state of pixel and updation of background model and threshold
  \begin{itemize}
    \item Classify each pixel as foreground or background
    \item Update Background modelling and threshold value.
  \end{itemize}
  \begin{enumerate}
    \item \textit{Read Previously Developed Background Modelled Frame}
      \hspace{1cm} $\text{BGM}_1(x) = B_0(x)$; //BGM is Background modelled frame
    \item \textit{Foreground Modelling and Moving Pixel Classification for} \hspace{0.5cm} $n = 1 \hspace{0.5cm} \text{to} \hspace{0.5cm} N$
      \begin{itemize}
        \item Read each test frame, one frame at a time from Video Sequence
        \item Convert each frame from RGB to Grey colour format
        \hspace{1cm} $f_n(x) = \text{rgb2grey}\left( f_n(x) \right)$ // where $f(x)$ is grey colour frame and \hspace{0.5cm} $1 \leq n \leq N$;
      \end{itemize}
    \item \textit{Compute the Frame Difference Matrix}
      \hspace{1cm} \textit{// absolute difference of each pixel from test and BGM frame}
      \begin{itemize}
        \item $\text{Diff}_n(x) = \left| f_n(x) - \text{BGM}_n(x) \right|$;
      \end{itemize}
    \item \textit{Compute Threshold}
      \begin{itemize}
        \item $\text{Threshold} = \text{double} \left\{ \frac{1}{2} \times \left[ f_n(x) + \text{meanValue} \left( \text{BGM}_n(x) \right) \right] \right\}$;
        \item $\text{Threshold} = \text{Threshold} / 2$;
      \end{itemize}
    \item \textit{Motion-based Pixel Classification}
      \begin{itemize}
        \item if ($\text{Diff}_n(x) \geq \text{Threshold}$) \hspace{0.5cm} \{  \hspace{0.5cm} \begin{align*}
          \text{FG}_n(x) &= \text{true}; \hspace{0.5cm} // \hspace{0.5cm} \text{pixel is in motion, i.e., part of foreground} \\
          \text{BG}_n(x) &= \text{false}; \hspace{0.5cm} // \hspace{0.5cm} \text{pixel is part of foreground}
        \end{align*} \hspace{0.5cm} \}  \hspace{0.5cm} \textit{// end if}
      \end{itemize}
    \item \textit{Update the Learning Rate for Background Model} \hspace{0.5cm} // \hspace{0.5cm} \textit{proposed contribution}
      \begin{itemize}
        \item Compute histogram of foreground frame, i.e., $\text{histFG}_n(x) = \text{imhist} (\text{FG}_n(x))$;
        \item Compute histogram of background frame, i.e., $\text{histBG}_n(x) = \text{imhist} (\text{BG}_n(x))$;
        \item Sum of both histograms, i.e., $\text{histSum}_n(x) = \text{double} \left( \text{histFG}_n(x) + \text{histBG}_n(x) \right)$;
        \item Now update the learning rate, i.e.,
          \hspace{1cm} $\alpha = \frac{1}{2} \times \left\{ \text{mean} \left( \text{double} \left( 1 - \alpha \right) / \text{histSum}_n(x) + \alpha \right) \right\}$;
      \end{itemize}
  \end{enumerate}
\end{itemize}
Update the Background Model for next Test Frame

\[ BGM_{n+1}(x) = \alpha \cdot f_d(x) + (1 - \alpha) \cdot BGM_n(x); \]

Display \( FG_n(x) \) frame

Foreground Enhancement

where \( \text{histSum}_n(x) \) is the sum of histogram of each pixel of foreground, i.e., \( \text{histFG}_n(x) \) and histogram of respective pixel of background model, i.e., \( \text{histBG}_n(x) \). The pixel classification has described in foreground modelling that has been described in Algorithm 2 where step 5 clearly depicts the classification pixel. Step 6 shows the applicability of histogram for learning rate computation. In the histogram, the x-axis represents grey level intensities of pixels, and the y-axis represents the frequency of pixel intensities.

In many applications, the histogram also used to compute the threshold for pixel classification as it is computed using the background model. Here, histogram has been used for computing learning rate. In histogram, each bin representing a certain intensity value range of a video frame. The final value of a bin represents the number of pixels assigned to the bin. The number of bins has been usually in the order of square root of the total number of pixels of a frame. It is to be noted that for grey scale format 256 (i.e., 0–255) bins are required and use to take the histogram of histogram counts. The histogram conveys information about the distribution of the intensities of pixels in a pre-assumed number of bins (i.e., in the ranges of intensities) which are spanning from the minimum to the maximum intensity value. The use of difference between histograms has reduced the number of false errors (false alarms) due to the non-stationary background, camouflage and illumination variation.

A threshold (refer Algorithm 2, step 4) has been applied to determine the pixel state if the current pixel value is close enough to its corresponding background value then pixel is considered as background pixel. If it the difference between pixels is greater than or equal to threshold than it is considered as foreground. This threshold is more optimal than global threshold as suggested by many researchers. This method update threshold and learning rate automatically. So, it also alleviates the burden of constructing and updating parameters externally. The proposed work has computed histogram for both background model and foreground frame as mentioned earlier in equations (3) and (4). The sum of histograms of both frames has been applied to evaluate an effective learning rate. This learning rate has been able to construct an effective background model for next iteration. It is also helpful for evaluating the portion of the background pixel and portion of foreground that have participated for construction of background model (refer step 7 of Algorithm 2). It makes pixel classification more effective that improves the detection quality during classification. Here, the involvement of histogram significantly reduces the inference of background.
3.3 Foreground enhancement

In the testing phase, extracted pixels can be categorised as part of moving object or part of background scene. But due to cluttered nature and illumination variation of background scene, proposed method generates false alarms. Morphological filters have reduced such kind of misclassifications. In this stage, a flat, disk-shaped structuring element with radius 2 has been applied, and holes have filled by hole-filling method of image processing. Some noisy blobs have been detected due to fast variation in the scene cause of fluctuation of light, switch on or off light, tree leaves/branches, etc. Therefore, such isolated blobs have removed (noise removal) by connected component, blob labelling and thresholding technique provided by MATLAB Tool. A region-based threshold value has been applied to consider isolated blobs as part of background. It also increases the true positive rate and reduces false positive rate.

4 Experimental setup and results analysis

This section shows the efficacy of proposed work where, the simulation has been carried out on a standard and well-known dataset recorded by single static camera, named as Microsoft’s Wallflower dataset (Toyama et al., 1999): this dataset consist of many problematic video frame sequences with frame size 160 × 120. This work performed experimental analysis over six frame sequences. In Wallflower, it is observed that the various kind of problematic issues are available like motion in the background scene caused by tree leaves, sudden/gradual variation in illumination. The video frame sequences used in this work are:

1. bootstrap (BS)
2. camouflage (CAM)
3. foreground aperture (FA)
4. waving trees (WT)
5. light switch (LS)
6. moved object (MO).

In this work, LS dataset is considered from change detection (CDNET-2012) (Goyette et al., 2012) where the illumination due to light varies from one frame to next frame.

In this work, all implementation and experimental analysis have been performed on MATLAB 2012b and Windows 8 on Intel (R) Core (TM) i5 processor with CPU 1.70 GHz speed. The experimental analysis for quality measurement has evaluated in two ways: qualitative analysis and quantitative analysis. The qualitative analysis consists of visual results of moving object whereas quantitative analysis demonstrates the analysis using various metrics.
4.1 Qualitative analysis

In various exiting methods, the initial learning rate (Stauffer and Grimson, 1999; Lee, 2005; Yadav, 2014; Haque et al., 2008) \( \alpha = 0.01 \). Similarly, the proposed work also uses same initial value of learning parameter. This value has been updated in testing phase at run time. The qualitative results of proposed work has compared with considered peer methods over various datasets. The visual observation for considered frame sequences (row-wise: 1: original frame, 2: ground truth) and peer methods [row-wise: Lee and Lee (2014), Zhou et al. (2013b), Jung (2009), Haque et al. (2008) and Stauffer and Grimson (1999)] are shown in Table 1.

4.2 Quantitative analysis

In this segment, the performance is evaluated quantitatively with the help of some parameters, let FP: false positive, FN: false negative. In literature, various metrics (Yadav, 2014; Yadav et al., 2014) are applied to evaluate the performance of various methods. In this work, the following metrics are used for analysis:

\[
\text{FP_Error} = \frac{\text{FP} \times 100}{\text{total_row} \times \text{total_column}}
\]

\[
\text{FN_Error} = \frac{\text{FN} \times 100}{\text{total_row} \times \text{total_column}}
\]

\[
\text{Total_Error} = \text{FP_Error} + \text{FN_Error}
\]

Here, information measurements quantify how a method can match its outcomes with ground-truth as shown in Table 1. The values of precision, recall and F-measure metrics (Yadav et al., 2014; Zhou et al., 2013b) are applied for performance analysis of proposed method against considered state-of-the-art methods. For all sequences, average value of precision, recall has been computed using equations (8) and (9) and then the values of F-measure is calculated by evaluating the harmonic mean of precision and recall as shown in equation (10).

\[
\text{Precision} \ (p) = \frac{\sum_p \text{TP}}{\sum_p (\text{TP} + \text{FP})}
\]

\[
\text{Recall} \ (r) = \frac{\sum_p \text{TP}}{\sum_p (\text{TP} + \text{FN})}
\]

\[
\text{F-measure} \ (F\_\text{score}) = \frac{2 \times (p \times r)}{(p + r)}
\]

The value of precision (positive predictive value) shows ‘how many selected pixels are relevant?’. Whereas recall (sensitivity) represents ‘how many relevant items are selected?’. Commonly, the precision and the recall are combined into a single measure, i.e., harmonic mean of precision and recall. It is known as F-measure if both are evenly weighted.

where \( S \) represents no of video frame sequence, i.e., \( S = 6 \). Mostly, the performance of detection has seen regarding value of F-measure. The value of F-measure varies within the range of 0 and 1. The highest value of F-measure depicts best result for any method over a particular frame sequence. In Table 2, relevant higher values are shown in bold with underline.
### Table 1
Qualitative analysis: row-wise (see online version for colours)

<table>
<thead>
<tr>
<th>Model</th>
<th>Microsoft's wallflower dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BS</td>
</tr>
</tbody>
</table>

### Table 2
Average value of precision, recall, F-measure for each method overall frame sequences

<table>
<thead>
<tr>
<th>Model</th>
<th>Average error</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
<th>Time (second per frame)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>6.2543</td>
<td>0.7029</td>
<td>0.6613</td>
<td>0.6708</td>
<td>0.02491</td>
</tr>
<tr>
<td>Lee and Lee (2014)</td>
<td>5.6042</td>
<td>0.6846</td>
<td>0.5990</td>
<td>0.6287</td>
<td>0.4021</td>
</tr>
<tr>
<td>Zhou et al. (2013b) (DECOLOR)</td>
<td>13.2855</td>
<td>0.4969</td>
<td>0.4504</td>
<td>0.4053</td>
<td>2.8142</td>
</tr>
<tr>
<td>Jung (2009)</td>
<td>24.0477</td>
<td>0.4962</td>
<td>0.5133</td>
<td>0.4594</td>
<td>0.5296</td>
</tr>
<tr>
<td>Haque et al. (2008)</td>
<td>18.2794</td>
<td>0.6070</td>
<td>0.4645</td>
<td>0.4576</td>
<td>0.3506</td>
</tr>
<tr>
<td>Stauffer and Grimson (1999) (GMM)</td>
<td>28.9184</td>
<td>0.3370</td>
<td>0.7493</td>
<td>4.2588</td>
<td></td>
</tr>
</tbody>
</table>
The ROC-curve predicts performance of a binary classifier that has its discrimination with a varying threshold. The performance of these methods has been validated through analysing Figure 3 and Figure 4.

**Figure 3** Precision-recall curve for proposed algorithm and state-of-the-art methods (see online version for colours)

4.3 Discussion and observations

The comparison of the proposed method and considered state-of-the-art methods have been shown in Table 1, Table 2, Figure 3 and Figure 4. The performance regarding detected results has been discussed as follows:

- Stauffer and Grimson (1999) method has even detected minor background motion as true positive because of dynamic nature or fluctuation in background through GMM. Therefore, rate of false positive alarms have been increased. Apart from it, due to pixel processing approach, this method also takes more execution time.

- Haque et al. (2008) proposed a method, simply modification of Stauffer and Grimson (1999) and it generates good foreground result, but few false positive alarms have still being detected.

- Jung (2009) has applied a pixel ratio, and mean of each pixel to classify the state of pixel. In this work, the performance of pixel classification is still lagging due to motion in the background or gradual illumination variation. In case of LightSwitch sequence, this method has generated lots of false positive alarms.

- Zhou et al. (2013b) method (DECOLOR) classified many pixels as false alarms also take much run time due to higher level of mathematical computations. This method also consumed extra space at runtime for all frames due to generation of 3D matrix.

- Lee and Lee (2014) have performed better, but its performance degrades due to the misclassification of pixels, i.e., false negatives. This proposed method has better outcomes for CAM and LS sequence. According to the experimental analysis, its performance is better than peer methods except proposed method.
The overall performance of proposed work is better against peer methods. According to Table 1, Table 2, the value of precision, recall, and F-measure is maximum.

The proposed precision-recall curve and roc-curve also demonstrates better performance as shown in Figure 3 and Figure 4.

The proposed results have been observed from Table 1, Table 2, Figure 3 and Figure 4.

The average error is less than others except (Lee and Lee, 2014) but can be acceptable that could be improved through more tuning among parameters.

According to Table 1, and Table 2, the overall performance of proposed method is better regarding precision, recall and F-measure.

The ROC-curve as shown in Figure 4 has clearly depicts that the overall performance of the proposed work is outranking.

The proposed work handles illumination variation of light and motion of background smoothly.

The proposed work has minimum running time (0.02491 sec/frame) as given in Table 2, the minimum execution time have applied over real-time-based video analysis in vision systems and surveillance applications. In other words, the proposed work processed approximately 35 frames per second.

The overall experimental results and analysis have proven the proposed work is able to handle the problems of illumination variation and motion in the background. It is also more appropriate for real-time applications.

**Figure 4** ROC-curve analysis for proposed algorithm and state-of-the-art methods (see online version for colours)
4.4 Merits

The proposed work has some merits and demerits as given below:

- A significant advantage of the proposed work is the fact that it does not require any complex training or initialisation.
- It maintains low complexity than other peer methods and allows real-time execution, even in MATLAB environment.
- This work presents a fast and effective method for background modelling and maintenance.
- It improves the robustness of the proposed work for detection of a moving object.
- The experimental results and analysis demonstrates that problems of gradual or sudden illumination and motion in the background scene have reduced as compared to considered peer methods.
- The proposed method works on the concept of frame wise pixel processing (i.e., process one frame at a time then execute next frame). Therefore, the suggested algorithms run faster than real implementation of other state-of-the-art methods on our machine.
- The proposed histogram-based adaptive learning rate is different from existing methods, Due to this, it achieves better results.
- Simple and easy to understand.
- Fast speed causes, this method can be applicable for real-time video surveillance applications.

4.5 Usability in healthcare system

Using radio frequency identification (RFID) and electronic health records (EHRs) tags allowed healthcare administrators to leverage various kinds of data from patient wristbands to diagnostic test results. It enhances the efficiency and perfectness with cost optimality. The next level intelligence and its actionable insight gained using enabling technologies in the healthcare of asset management, cloud, IoT and Big Data (Sharma 2016, 2015; Sharma et al., 2015a, 2016). To reach such kind of new level of efficiency, perfectness and patient safety. The healthcare centres or organisations require real-time visibility and intelligence into generated data such as patient details, medications, specimens, supply inventories or inventory stock details, etc.

From clinicians that need to identify, detect, track the patients’ behaviour as well as working of internal organs or detect various kinds of tissues through ultrasound, magnetic resonance imaging (mri), gathering specimens, administer medication and monitor essential signs to pharmacists. So with the cloud, Big Data and internet of things, all above must required an accurate inventory count. Therefore, healthcare organisations gain more benefit from next level Intelligence through delivering such healthcare services.
4.6 Demerits and limitations

The proposed work also faced some limitations during experiment and depicted as:

- Dependency of threshold value on learning rate because updated threshold depends on the background model and it depends on learning rate.
- Due to the limited hardware resources, the proposed method has processed approximately 35 frames per second in grey format with existing hardware. If we have real-time-based fast processing machine, then the processing speed should be intensified.
- With limited resources this work has not been experimented on cloud environment.

5 Conclusions and future work

In this work, a simple and novel method has been developed for moving object detection. That proposed work has minimised the problem of motion and illumination variation in background scene. This work has been tested over various video sequences of Microsoft’s Wallflower dataset. This work contributed a histogram-based adaptive learning rate that has easily resolved problems generated by illumination variation and motion in background scene. The adaptive learning rate update scheme has been able to construct new background model during testing stage. The overall qualitative and quantitative results regarding F-measure, processing-time, have demonstrated better detection quality. The performance of the proposed method has been justified through precision-recall and roc-curve analysis. The proposed method achieved better performance and also compared with considered state-of-the-art methods. Due to better execution speed, the proposed method is more appropriate for real-time video surveillance applications where fixed camera will capture video frames.

The further extension of this work is to detect moving object in complex and problematic real-time scenarios where video frames will be captured online using cloud environment and directly used as an input in the implementation. In future, we will extend our work in healthcare domain using cloud, and internet of things technologies.

Acknowledgements

The first author is very grateful to her Daadu Ji (Shri. Ram Krishan Choudhary Ji) for their blessings and motivation during this research work. All this work has been carried out under his kind support and blessings. The authors are very thankful to the Microsoft’s wallflower dataset (Toyama et al., 1999) community for providing the dataset along with ground-truth. The authors also expressing their gratitude to change detection dataset (Goyett =et al., 2012) community for providing support for analysis of this work. The first author also expressing millions of thanks to principle and HOD (CSE Department), GBPEC, Pauri, Garhwal, UK for providing good environment.
References


