

Research Article

Top-Down and Bottom-Up Cues Based Moving Object Detection for Varied Background Video Sequences

Chirag I. Patel,¹ Sanjay Garg,¹ Tanish Zaveri,² and Asim Banerjee³

¹ Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad 382481, India

² Department of Electronics & Communication Engineering with Institute of Technology, Nirma University, Ahmedabad 382481, India

³ DA-IICT, Gandhinagar 382007, India

Correspondence should be addressed to Chirag I. Patel; chirag453@gmail.com

Received 10 June 2014; Revised 13 October 2014; Accepted 20 October 2014; Published 16 November 2014

Academic Editor: Deepu Rajan

Copyright © 2014 Chirag I. Patel et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Moving object detection is a crucial and critical task for any surveillance system. Conventionally, a moving object detection task is performed on the basis of consecutive frame difference or background models which are based on some mathematical aspects or probabilistic approaches. But, these approaches are based on some initial conditions and short amount of time is needed to learn all these models. Also, the bottleneck in all these previous approaches is that they require neat and clean background or need to create a background first by using some approaches and that it is essential to update them regularly to cope with the illuminating changes. In this paper, moving object detection is executed using visual attention where there is no need for background formulation and updates as it is background independent. Many bottom-up approaches and one combination of bottom-up and top-down approaches are proposed in the present paper. The proposed approaches seem more efficient due to inessential requirement of learning background model and due to being independent of previous video frames. Results indicate that the proposed approach works even against slight movements in the background and in various outdoor conditions.

1. Introduction

The process of detecting a moving object has a great significance for computer vision field like video surveillance, automatic target detection and tracking, and so forth. It is also very beginning and important step in the automation of surveillance system as it is bringing apart moving object from background. Objects in video frame or scene are specified as the moving objects that concern and hold the maximum attention at that moment. Capturing the same scene at various times and detecting the changes in them has a large amount of applications in diverse fields. Real time applications such as robotics visions and surveillance would use series of frames as excitation and the problem of object detection can be considered for such systems as moving objects detection.

To implement any approach and to make it useful, two factors should be considered: the computational cost and efficiency. However, most of the object detection approaches

are expensive as far as its computational cost is concerned. Static image sequence is used for reducing complexity of algorithm but applying the same on each image frame of video would not be a proper solution for object detection. An approach based on the above conditions uses the most efficient algorithms for static images which would be slower than that of required moving object detection techniques for real time analysis. Therefore, it is very essential that the proposed approach should be efficient and reliable. A bunch of thoughts comes to mind when it is about visual attention models of human. One type of approaches encourages the importance of bottom-up approach to drive out attention. These types of approaches are neither task dependent nor based on memory. The claim is that our approach is more likely to focus on objects which stand out from their background. One of the most widely used models for bottom-up saliency is the Itti-Koch model [1].

On the other hand, someone may point out the relevance of various top-down aspects of the stimuli to drive our

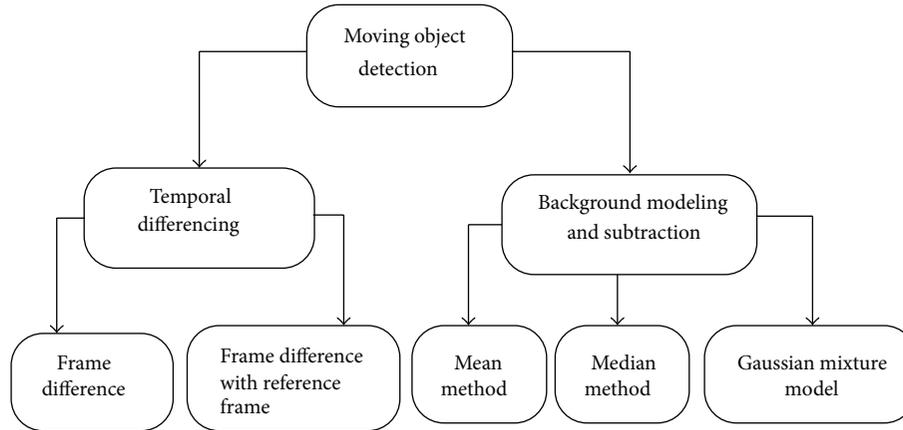


FIGURE 1: Classification of moving object detection methods.

attention. These task dependent factors include context and feature based cues amongst others. For instance, if we are given the task of searching for a pedestrian in an image, we are likely to focus on the road despite there being salient objects in the sky. Also, while searching for a car, we would call upon our knowledge of what a car looks like and would drive our attention to objects similar to it rather than objects similar to a pole. In the proposed work, the context based and feature based cues have only been considered amongst the top-down influences which drive our attention.

2. Backgrounds, Introductory and Interrelated Work

Various types of approaches have already been proposed to resolve the problem of detecting the objects from the given set of images. As shown in Figure 1, these methods of detection can be divided into two predictable classes: (1) temporal differencing and (2) background modeling and subtraction.

In temporal differencing, a scene is captured by a static camera and moving object is approximated by difference between two consecutive video frames. When a change is found to be high, the pixel is identified as a moving background. However, the real time application like surveillance does not use temporal differencing due to its poor performance.

In background modeling, the background is designed by using mathematical models or probability theorems and it is subtracted from the current frame. By comparing the current frame with a valid model of scene background, the background subtraction segments the changed regions corresponding to the moving objects, referred to as foreground. However, the automatic foreground extraction is still a severe challenge to the current state of the art for an arbitrary scene, for example, nonstatic and cluttered background. Recently the background model is estimated from the video sequence and updated accordingly and the foreground is differentiated from the background. In general,

moving objects can be detected by the background model and subtraction approaches [2, 3]. There are methods like Gaussian mixture model [4], mean model, standard deviation model, and median model [5, 6] by which background model is estimated.

Approach for building model for human body and tracking of the moving object is represented by Mikić et al. in [7]. When images from the video are acquired, object or human is detected and segmented from that image sequence. Boundary of the human body is generated from synchronized, segmented multiple video streams. Locating different parts of a body is done by template fitting and growing procedures. Initialization of this model is done by Bayesian network that integrates prerequired knowledge of the human body properties. Finally, tracking procedure executes by Kalman filter which is used to approximate position of the model.

Simple background approaches like frame differencing are not invariant to illumination changes, and hence running average or moving average method [8] is used due to the required updates of the background model with time. Even though running average has properties like low computational complexity and high memory compactness. Performance may be improved if fuzzy approach is integrated with it. In fuzzy background subtraction, saturating linear function is used to estimate pixels of the moving object instead of hard limiter function. Another background removal technique is proposed in [9] based on fuzzy theory to make algorithm illumination invariant and also invariant to shadows of colour features.

When background model is estimated by the mixture of Gaussian model [10], it fails against variation in illumination and small irregular movements in camera (camera jitter). Solution to this problem is to integrate fuzzy theory with Gaussian mixture model. Type 2 fuzzy mixture of Gaussian model [11] is the technique which overcomes traditional problems of Gaussian mixture model. Robust approach for background subtraction is defined by Cheung and Kamath [12]. An object is first detected as blob by slow adopting algorithm (Kalman filter). Validation of the object is performed

by integrating blob detection and simple frame difference. Object histogram is calculated for usage of extending object boundary. Technique is performed better under occlusion for better object localization. But having limitations like use of boundary of ellipse for modeling object like human may not be suitable because human is a complex object.

Modeling of background is also done based on analyzing colour value of pixel [13]. Colour value of each and every pixel of video frames is analyzed for a short amount of time. The most frequent colour is assigned to that pixel. The algorithm can also handle small disturbance in background due to the environment. It is easy to implement and is fast as well. But on the other hand, the algorithm gives false detection when video frame consists of complex and congested background. One of the approaches presented in [14] also handles cluttered background and slightly moving background due to tree branches and bushes. Object is detected based on estimating probability density function using previous values of pixels. Another background model presented in [15] is based on kernel density estimation which is also used for tracking.

Background subtraction and segmentation approach is used for building 3D model of the human body [16] which can be further converted into free viewpoint video. From statistics of colour of background, the pixel object is separated. Mean and standard deviation of background image is calculated and if pixel differs in one of the colour channels, the threshold is defined for detecting foreground object. The work presented in [17] denotes how many algorithms are appraised with wide variation in preprocessing algorithmic approach. Background subtraction is also used as the preprocessing task for real time 3D motion capture [18]. In this approach, the background is formed by calculating statistical parameters like mean and standard deviation over a limited learning period.

Wren et al. [19] defined background by single Gaussian with integration of mean and covariance matrix parameters but this approach does not withstand along dynamic background. Universal background subtraction model is proposed in [20]. Here the author has given detailed description of initialization and update and how the model works. The proposed algorithm can be more specifically understood by dividing it into three tasks. Firstly, it is the classification of pixel to foreground or background. Secondly, the model can be initialized by only single frame, saving time and making it suitable for real time and short sequence. And, thirdly, it is about updating model. Presently proposed approach [21] uses background subtraction model as the preprocessing and preliminary step for moving target detection. The author represents feedback background estimation framework which in turn gives the background area.

2.1. Problems. Computationally, the previously proposed aspects are drawn out as they take a great amount of time; also building and updating background model is a tedious process. All these approaches are worth without background; however, the generated background model may not be applicable in some scenes with sudden illuminating changes. As a result, it can be stated that reducing influence of these

changes is vastly dependent on a good background model. Efficient implementations often use a reliable and inexpensive technique to find regions of interest that may be further subjected to costly computations. Inspired by the success of attention based models on static images, in the proposed work, these attentional models for object detection in videos are extended with the hope of achieving better results using cues that have not been explored by the standard techniques yet.

3. Visual Attention

3.1. Why Visual Attention? Visual attention, also called saliency, is the perceptual quality that makes an object, person, or pixel stand out as compared to its neighbours and thus it captures our attention [22]. Visual attention estimation methods can broadly be classified as biological based, purely computational, or a combination of both. In general, all methods employ a low-level approach by determining contrast of image regions relative to their surroundings, using one or more features of intensity, colour, and orientation. Extraordinarily, the efficient way to perform this job is to imitate the human describing behavior. Humans have a mechanism of visual attention that is a biological equivalent of preselecting regions for further costly computations. Humans have a well-developed cognitive mechanism which allows them to perform tasks requiring visual guidance efficiently. Motivated by this efficiency in human beings, a computational model for dynamic visual attention is proposed that would help us perform the task for object detection in videos more efficiently and implement a working model along the same lines to verify the feasibility of our proposal.

Visual attention estimation has become a valuable tool in image processing; however, the existing approaches exhibit considerable variations in methodology, and it is often difficult to attribute the improvements in the resulting quality to specific algorithm properties. Inspired by the success of visual attention based models on static images, In the proposed work, these visual attentional models are extended for object detection in videos with the hope of achieving better results using cues that standard techniques have not explored yet. Visual attention and feature based cues are used for developing a model to compute regions of interest in a video while incorporating the temporal aspects of the video in computing these cues. As compared to the previous methods that used set of trainings to make the detector learn about objects in videos, the proposed method vanishes this disadvantage. Attention based approaches have often been used with a high degree of success for the task of preselecting regions in static images. These approaches are of different degree of nature of complexity and they exploit different cues to model human visual attention. It is, therefore, important that the proposed work looks at the commonly used approaches for static images to address the question of which approach would work the best in the case of videos.

Visual attention approaches are classified into two main streams: bottom-up and top-down models. Low levels features like contrast, colour, intensity, orientation, texture, and

motion can render measurement of saliency for every pixel in image. This type of model is defined as bottom-up model of saliency or visual attention. Usually, most of the bottom-up saliency approaches [1, 23] use local level features instead of global level features. In top-down approaches, saliency map is determined by global features like contextual information of the object or scene [24–27]. Context of object or scene can be defined by statistical probability distribution of it. The global information provides spatial representation and category of scene or object [28, 29]. These approaches are fast enough as compared to the bottom-up approaches due to their search mechanism and fixation.

A lot of work has been done previously on modeling human visual attention; many of them use either only bottom-up [30] or only top-down factors [31] to predict human attention. We are inspired to combine the features of top-down and bottom-up by the work shown in [32] and our area of research has been based along with the lines of [33]. Some of the more commonly used techniques are given as follows.

Hou and Zhang [34] proposed saliency detection algorithm independent of features, categories, or other forms of prior knowledge of the objects. By analyzing the log spectrum of an input image, the spectral residual of the image in spectral domain is extracted, and a faster method to construct the corresponding saliency map in spatial domain is proposed.

Harel et al. [35] had proposed new bottom-up visual saliency model, which is known as graph-based visual saliency (GBVS). It consists of two steps: first forming activation maps on certain featured channels and then normalizing them in a way which highlights conspicuously and admits combination with other maps. The model is simple and biologically plausible to the extent as it is naturally parallelized.

Itti et al. [1] have defined bottom-up approach for computing saliency map. Low level features like intensity, colour opponency (red-green, blue-yellow), and four orientations, (0, 45, 90, and 135) are used for calculating feature maps. Weights are assigned to the resultant map proportional to their differential response from neighbour. Thus, saliency map is defined by combining 42 features of maps.

Here, in this proposed approach, visual attention salience, that is, image based driver of attention approach, has been used for moving object detection for image sequence.

4. Proposed Approaches

The proposed method has been classified into two different streams for finding out moving object using visual attention of video sequence: (1) bottom-up cue based approaches and (2) combination of bottom-up and top-down cues based approach.

4.1. Bottom-Up Cue Based Approaches for Visual Attention

4.1.1. *Modified Frequency Tuned Spatial Model I.* In surveillance system, often, motion is an important cue to attract

attention [36]. The mostly salient part in the video frames sequence is considered as the moving object and hence, in surveillance it can be applied as a fundamental step. Moving objects have some salient properties and features which can be easily differentiated with the background. If an image is having a small object, the average value of gray level of that image will give a value near to the background value. In contrast to it, if the image is having an object larger in size, the average value will be near to the foreground object. This method is a modified version of the method proposed in [37]. Two modifications are done: first it is in [37], where the author has found saliency map using difference of Gaussian blurred image with size of 3×3 mask and average of an image, where the proposed saliency map of an image is calculated using difference of Gaussian low-pass (mask size 11 and σ is 5) of image and averaging filter is supplied with the size of image.

Methodology. Apply average filter on video sequence $I(x, y, t)$, having size $N \times M$ for particular time t ,

$$I_{\text{avg}} = I(x, y, t) \otimes A(N, M), \quad (1)$$

where A is averaging filter, having mask size $N \times M$, and \otimes is defined as convolution between two images.

Employ Gaussian filter on image

$$I_{\text{gaussian}} = I(x, y, t) \otimes G(h, \sigma), \quad (2)$$

where G is Gaussian low-pass filter, having mask size 11, and σ is 5. The saliency value at each pixel position (x, y) is given by

$$S(x, y) = d[I_{\text{gaussian}}(x, y), I_{\text{avg}}(x, y)], \quad (3)$$

where d is the distance between two pixels in the respective images. The first step in the proposed approach is image averaging and the second step is subtracting it from the low-pass version of the image using Gaussian low-pass filter so that it results in no spike noise and random pixels in the output image. Finally, morphological operations are performed for creating blob like structures in the resultant output image for extracting salient component of the image.

But for boosting up the speed of the approach, all the fundamental steps are applied in the frequency domain. Image average is the DC component in the frequency component which should be considered for salient object that has large portion in the image. Applying Gaussian filter, which is having low-pass characteristics that defines high cut-off frequency for the resultant salient map. High cut-off frequency delimitates edges and border of the image but to remove noise and random pixels (which are also high frequency components) as a result, blurring is enforced to the image by using Gaussian spreading so that the ringing effect can be avoided.

4.1.2. *Visual Saliency Based on Colour Image (Colour Saliency [38]).* Basically, human perception development can be better performed using HSV colour space where H stands for hue, S for saturation, and V for value [38]. Any red, green, and

blue component of a particular pixel of an image is converted to HSV component according to the following equations [39]:

$$h1 = \cos^{-1} \left(\frac{(1/2) [(R - G) + (R - B)]}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right), \quad (4)$$

where $h = h1$ if $B \leq G$; otherwise, $h = 2\pi - h1$;

$$s = \frac{\max(R, G, B) - \min(R, G, B)}{\max(R, G, B)}, \quad (5)$$

$$v = \frac{\max(R, G, B)}{255}.$$

The given colour image i is of size N by M pixels, and the pixel colour value at location (i, j) ($0 \leq i < m$, $0 \leq j < n$) is denoted as $(h(i, j), s(i, j), v(i, j))$. Compute the mean values of saturation and brightness over the whole image and denote them as m_s, m_v , respectively.

The colour saliency $sc(i, j)$ at location (i, j) ($0 \leq i < m$, $0 \leq j < n$) is given as a two-dimensional sigmoid function [40, 41] by

$$S_c(i, j) = \left(\frac{1}{c_s + \exp(-s(i, j)/m_s)} \right) \cdot \left(\frac{1}{c_v + \exp(-v(i, j)/m_v)} \right). \quad (6)$$

The parameters c_s and c_v are two constants. In general, while judging the saliency, we come across both a sensory and a subjective component for the verdict. For example, most of us can perceive brighter and purer colours more easily than duller and mixed colours; the judgment of the saliency of two hues of the same brightness and saturation can be subjective. Therefore, in the present study, saturation and brightness are chosen as the measures for saliency. Then, the colour saliency is normalized to be in the range of $[0, 1]$.

4.1.3. Gaussian Spatial Model by Coalescing Laplacian of Gaussian (LoG) Filter and Gaussian Filter. Abrupt changes in the image are detected by sharpening filter and normally Laplacian filter is employed to the same one. But as the noise is also defined by the abrupt changes in image with details of information, there is a need to a filter which can produce the resultant image where pixels are having a change in the original image without noise. Laplacian of Gaussian filter is the second derivative of a Gaussian function. The image is first smoothed by the Gaussian filter so that the noise can be eliminated and then it is enhanced with the high frequency information like edges and lines using Laplacian filter. This two-step process is called the Laplacian of Gaussian (LoG) operation [42].

The 2D LoG function that has the center located on zero and which has the Gaussian standard deviation can be expressed by the form [43]

$$\text{LoG}(x, y) = -\frac{1}{\pi\sigma^4} \left[1 - \frac{x^2 + y^2}{2\sigma^2} \right] e^{-(x^2 + y^2)/2\sigma^2}. \quad (7)$$

Methodology. Apply Gaussian filter G on image $I(x, y, t)$

$$I_{\text{gaussian}} = I(x, y, t) \otimes G(h, \sigma). \quad (8)$$

Now, calculating LoG response of the image, we get

$$I_{\text{LoG}} = I(x, y, t) \otimes \text{LoG}(h, \sigma). \quad (9)$$

Saliency map S at pixel location (x, y) can be estimated by finding out the Euclidean distance between the response of two images; that is,

$$S(x, y) = I_{\text{LoG}}(x, y) - I_{\text{gaussian}}(x, y). \quad (10)$$

4.1.4. Gradient Saliency by Utilizing Sobel Operator. The Sobel operator [43] is used for detection of rapid change in images. Gradient of an intensity of image pixels is determined using a Sobel operator. Response of Sobel filter at each pixel is an absolute magnitude response of gradient in x direction and gradient in y direction at that pixel. Computationally, the gradient calculation of an image is inexpensive and hence this is fast but also primitive.

Apply Sobel filter on image $i(x, y)$, having size of $N \times N$, with consecutive pixels having distance of h pixels,

$$I_{\text{sobel}} = I(x, y, t) * \text{sobel}(x, y, h) \quad (11)$$

with Sobel operation being performed on image using different sobel filters, one with $h = N/2$ and another with $h = N/3$:

$$I_{\text{sobel1}} = I(x, y, t) * \text{sobel}\left(x, y, \frac{N}{2}\right), \quad (12)$$

$$I_{\text{sobel2}} = I(x, y, t) * \text{sobel}\left(x, y, \frac{N}{3}\right).$$

And, after normalizing the result in range between $[0, 1]$, the combination of the results will give the saliency map S as

$$S(x, y) = I_{\text{sobel1}} - I_{\text{sobel2}}. \quad (13)$$

4.1.5. Modified Frequency Tuned Spatial Model II. As the human perception about colours and intensity can be well correlated with HSI model [44], hence, lab colour space is hereby replaced with HSI colour space in this approach. The proposed approach is almost the same as [37] but the only differences are that HSI colour model is used instead of lab colour space and gaussian filter's window size is 11×11 instead of 3×3 . The saliency map S for an image I of width N and height M pixels can be formulated according to [37]

$$S(x, y) = |I_{\mu} - I_{\text{whc}}(x, y)|, \quad (14)$$

where I_{μ} is the average or mean of two-dimensional image array and $I_{\text{whc}}(x, y)$ is the corresponding image pixel vector value in the Gaussian blurred version (using a 11×11 window size) of the original image [37]. Magnitude of the resultant deference (calculated using the Euclidean distance) is considered. Using the HSI colour space, each pixel location is an $[H, S, I]^T$ vector.

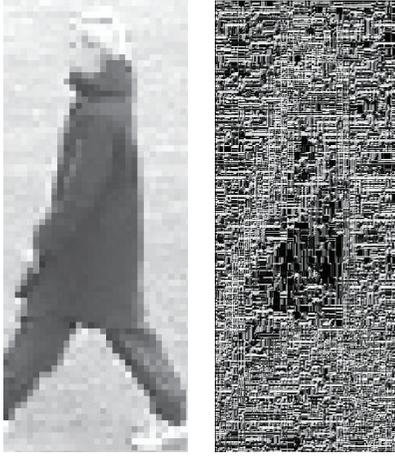


FIGURE 2: Calculating LBP of a person.

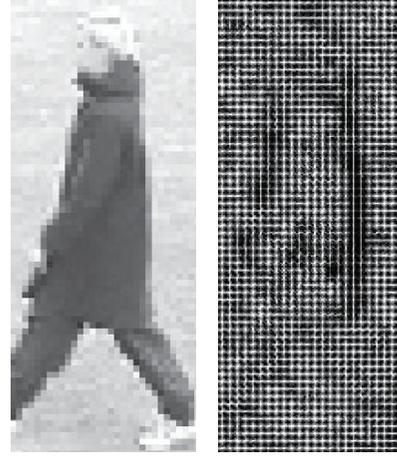


FIGURE 3: Calculating HOG of a person.

4.2. Combined Approaches (Combining the Bottom-Up and Top-Down Cues)

4.2.1. *Integrating Colour Saliency with Texture Feature (Local Binary Pattern (LBP))*. Ojala et al. [45] have proposed an approach defined as local binary pattern (LBP). Initially, the approach was used for texture analysis [46], but later on it was stated to be used for different applications. LBP is illumination invariant and has computational efficiency and is suitable for computer vision task. In LBP, neighbour of 3×3 pixels is defined for every pixel and is compared with the middle pixel. If it is higher than the middle one, it can be assigned with 0, else with 1. Thus decimal value is defined from 8 neighbourhood pixels. Later on, the extended LBP was found in which neighbourhood windows size varies [47]. Examples of the LBP calculation for person image are shown on the right of Figure 2.

In the proposed study, the saliency (color saliency) and feature (local binary pattern features) based maps are obtained for our test images. Before combining them, two maps were normalized to matrices having value between 0 and 1 with each pixel accounted for by a number. Various possible combinations were also checked for combining these two maps—weighted addition and multiplication on two sets of weights assigned to each map. This is given more elaborately in Section 7. Further, the performance of the single source models was also checked.

4.2.2. *Integrating Modified Frequency Tuned Spatial Model I with Histogram of Oriented Gradient (HOG) [48]*. Histogram of oriented gradient (HOG) [48] is feature descriptor, which is implemented for person detection. The HOG person detector uses a detection window that is 64 pixels wide by 128 pixels in height according to the structure of the person. It operates on 88 pixel cells within the detection window and these cells will be organized into overlapping blocks. Within a cell, gradient vector is computed at each pixel. In 8×8 region, 64 gradient vectors are extracted and put into 9 bin histogram. The histogram varies from 0 to 180 degrees, so

there are 20 degrees per bin. Histogram bins are filled by magnitude of gradient vector. Contribution splits between the two closest bins. The next step in computing the descriptors is to normalize the histograms. Rather than normalizing each histogram individually, the cells are first grouped into blocks and normalized based on all histograms in the block. The final descriptor is designed using this procedure as shown in Figure 3.

4.3. *Optimization Problem Formulation*. The proposed framework selects the saliency map (SM) for moving object detecting by defining a learning based framework.

The composite saliency map (SM) is defining as the parametrized linear combination of n base saliency maps (SM); for example:

$$SM(x_1, x_2) = \sum_{i=1}^n w_i SM_i(x_1, x_2), \quad (15)$$

$$SM(x_1, x_2) = \prod_{i=1}^n w_i SM_i(x_1, x_2)$$

with $w_i \geq 0 \forall i$. The nonnegativity constraint enforces Mercers condition on saliency map (SM). The proposed learning framework intended to learn optimized saliency map for moving object detection.

Therefore, we define detection performance maximization as the optimization objective. The limited amount of ground truth data may create condition of overfitting. Therefore, the optimization objective requires a regularized term to ensure the desirable moving object detection performance independent of the amount of ground truth data.

We apply the maximum entropy principle based regularizer presented by Wang et al. [49]. The regularizer is implemented by applying maximum entropy principle which assigns equal probability for function value $SM(\cdot)$ to be 0 or 1.

TABLE 1: Detail information of test videos.

Video name	Format	Length	Frames per second	Number of frames	Frame (width × height)	Image type
ATM	MPEG	29 sec	25	728	704 × 576	true colour
Cam	AVI	2 min 32 sec	29	4567	384 × 288	true colour
Campus_raw	AVI	1 min 57 sec	10	1179	352 × 288	true colour
EPC-200-day-bird	AVI	37 sec	29	1115	320 × 240	true colour
HALLWAY_A	MPEG	22 sec	30	661	704 × 576	true colour
highwayI_raw	AVI	29 sec	14	440	320 × 240	true colour
person15_walking_d1_uncomp	AVI	29 sec	25	741	160 × 120	true colour
Stair	MPEG	23 sec	30	707	704 × 576	true colour
visor_1197284015696_07_fumogeno4	AVI	1 min 20 sec	25	2024	320 × 240	true colour

Therefore, the complete optimization objective for multiple saliency map problem is defined as

$$w = \operatorname{argmax}_w F(X, X_v, w) \quad (16)$$

$$F = J(X_v, w) + \lambda V(X, v),$$

where X represents the complete dataset and X_v represents the subset of dataset examples assumed to be available with ground truth. Information for function F is evaluated for different weight parameter to represent the regularization parameter. Function $J(X_v, w)$ represents the retrieval performance of saliency maps computed over X_v and w as

$$J(X_v, w) = \operatorname{mean} \left\{ \sum_{x_i \in X_v} \Delta(y_i, \hat{y}(x_i, w)) \right\}, \quad (17)$$

where $\hat{y}(x_i, w)$ represents the set of detected results for the given input sequence $x_i \in X_v$. y_i represents the actual ground truth for the input sequence and $\Delta(y_i, \hat{y}(x_i, w))$ represents the computed detection rate for x_i . Function $V(X_v, w)$ represents regularizer term defined as the sum of variance of the detected rate for all video sequences which is computed as

$$V(X, w) = \sum_{i=1}^k \operatorname{Variance} \{SM_i(X, w)\}. \quad (18)$$

5. Video Database

Testing videos are adopted from [22, 50, 51]. Video has resolution 4CIF (common intermediate format) (704 × 576 pixels), frame rate 30 fps (frame per second), number of images 730, and video format mpeg-1. Some of the videos utilized for testing have static background, whereas others have slightly dynamic background. Details of all the videos used for testing purpose are outlined in Table 1.

6. Experimental Setup

The experiments are performed on an Intel(R) Core (TM) i5 2430M CPU 2.40 GHz with 6 GB RAM and the algorithm is implemented using MATLAB 7.8 tool.

7. Results

Videos are considered based on the variation of background and object in videos to test the stability of the proposed approach under different conditions. Change in colour, intensity, and textures in the background as well as in the object is also considered. Distinctive outdoor and indoor conditions are provided to ensure checking efficiency of the proposed approach. One of the videos is also counted for slight change in the background to check and see the stability against dynamic video. In the results, numbers define video and alphabets with the number define different image frames of a particular video. The first column in all results is the original image sequence from video, the second column is saliency map of that image sequence, and the third image consider resultant object detection of respective image sequence. Here, the saliency map has been calculated for all the proposed methods (with the gap of 10 frames) of video sequence.

7.1. Bottom-Up Cue Based Approaches for Visual Attention

7.1.1. Modified Frequency Tuned Spatial Model I. Results are shown in Figures 4, 5, 6, and 7 for 8 different types of video using combinations of averaging filter and Gaussian filter methods. All videos are having diverse backgrounds, different textures, colours, and objects. Objects are having variation from human to machines like cars and to birds as well. Even though all variations are in the background, the proposed algorithm gives an efficient and excellent result.

The first video shows that a man is going towards a car and both the objects are detected. The second video is about highway traffic and the proposed algorithm can detect all cars properly. The third one has a campus location in which some of the cars are already parked and one car is entering into campus; it is well detected too. Other videos include two



FIGURE 4: Videos (1) and (2): results using modified frequency tuned spatial model I.

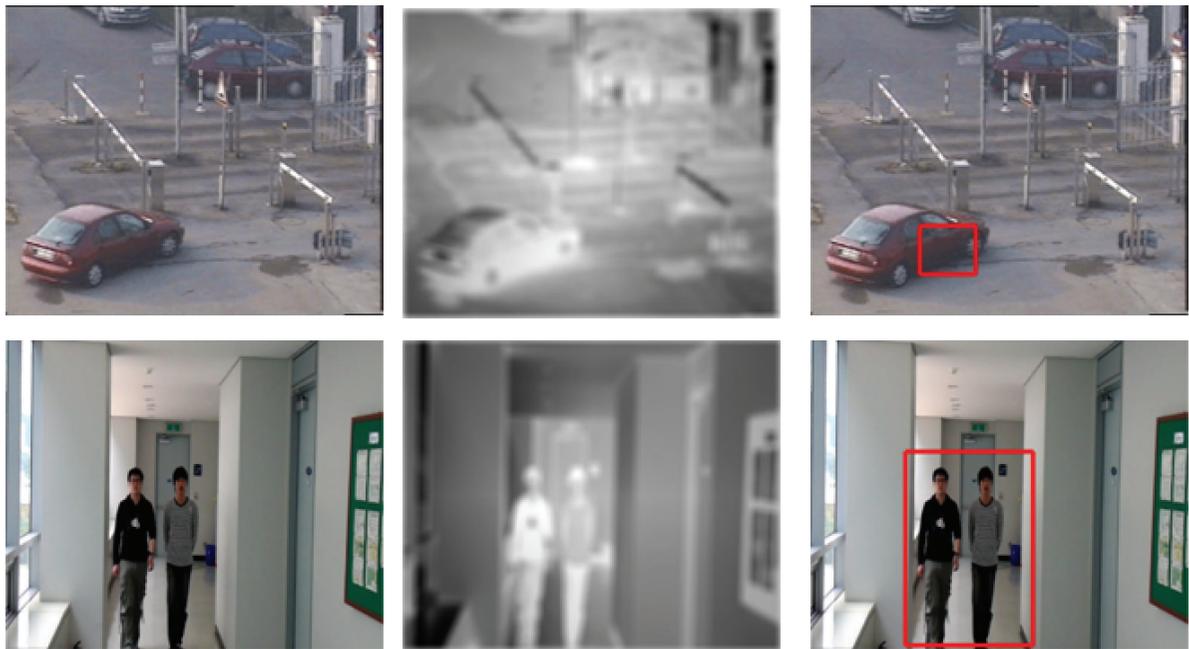


FIGURE 5: Videos (3) and (4): results using modified frequency tuned spatial model I.

persons entering into the hallway (video 4), birds in their nest (video 5), a person going far from camera in a campus area (video 6), one coming into ATM room (video 7), and one walking on road (video 8).

In video 5, where the birds are in their nest, the background is considered to be slightly moving due to tree leaves but the proposed technique is giving an appreciable result.

7.1.2. Visual Saliency Based on Colour Image (Colour Saliency [38]). Results of saliency using colour saliency approach are demonstrated in Figures 8 and 9. In the first video, a person is coming upstairs and here he is detected properly. But along with it, some false consequences have also been generated in video 1(a) as this approach is based on colour value of the video sequence and therefore, a few objects are detected



FIGURE 6: Videos (5) and (6): results using modified frequency tuned spatial model I.



FIGURE 7: Videos (7) and (8): results using modified frequency tuned spatial model I.

which are false and they are having the same color as the object.

7.1.3. *Gaussian Spatial Model by Coalescing Laplacian of Gaussian (LoG) Filter and Gaussian Filter.* Results are shown for 7 different videos in Figures 10, 11, and 12 and here in all these videos, the objects are detected in the right way. But there are some problems faced during detection due to increase in false positive. In video 2, the upper left corner of the image is detected which has the same colour and is

not a moving object. In video 5 and video 6 green notice boards are detected which is not required to be detected (false detection). In video 7, vehicles in the parking lot are also detected unlawfully which are objects but not the moving ones. But in this approach, in video 4 some cars already placed in the parking lot are not considered the moving objects for detection which is true.

7.1.4. *Gradient Saliency by Utilizing Sobel Operator.* In Figure 13, the results are presented using the approach of

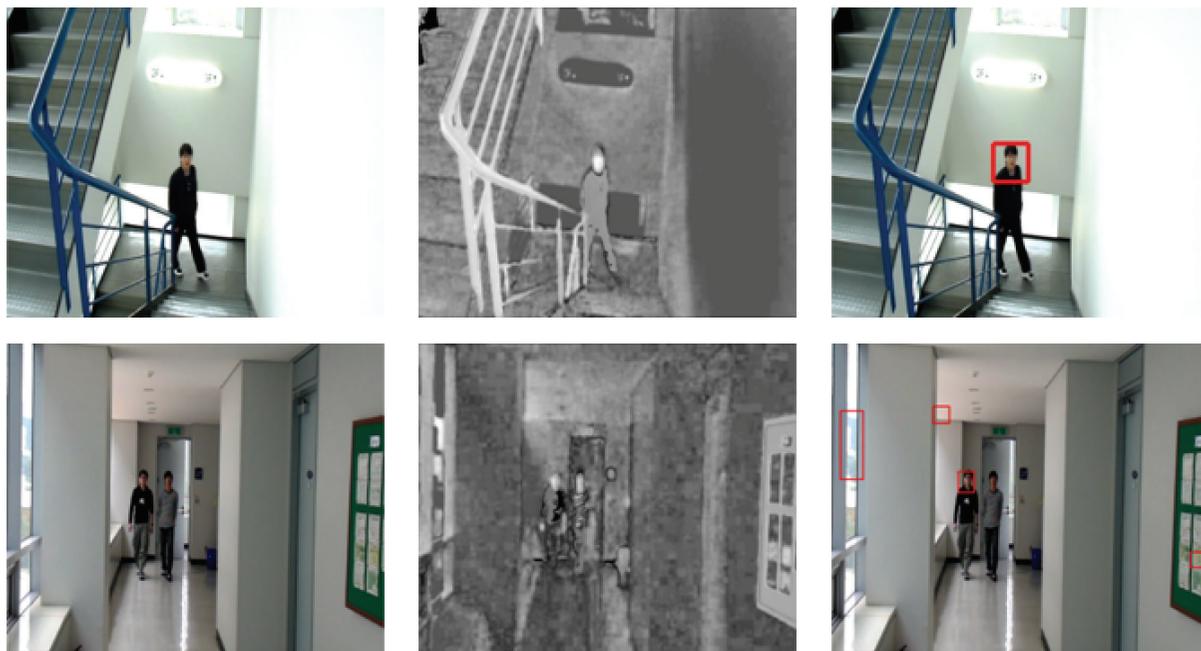


FIGURE 8: Videos (1) and (2): results using colour saliency [38].

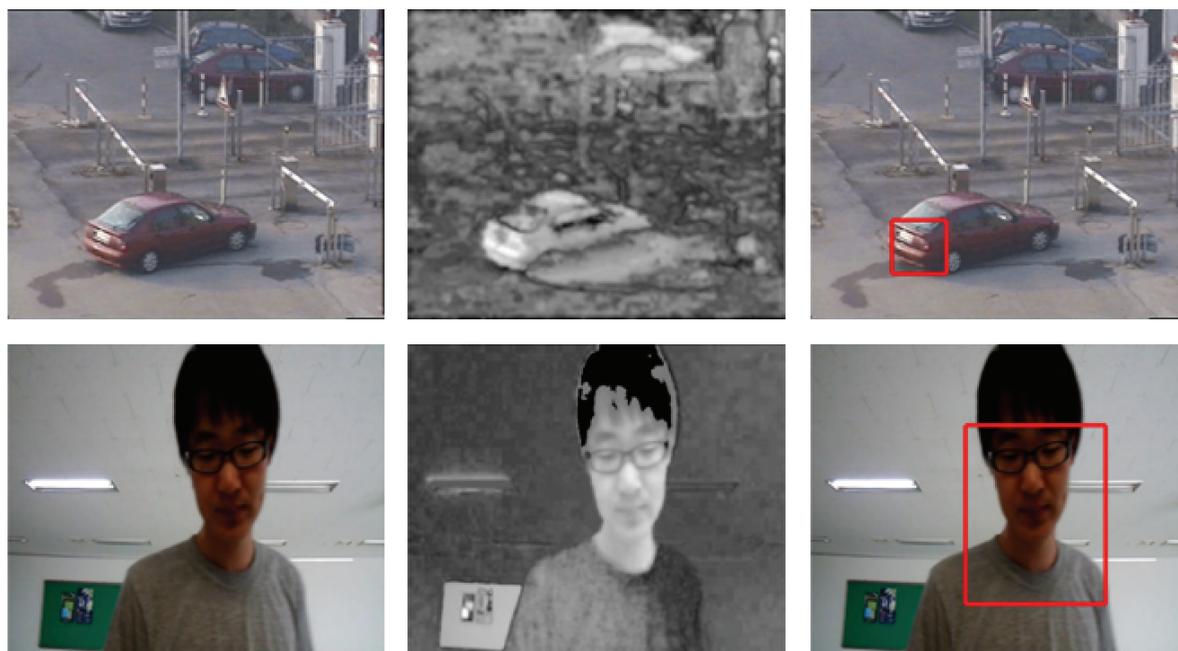


FIGURE 9: Videos (3) and (4): results using colour saliency [38].

gradient saliency using Sobel from static image for four different videos. Here, in video 1, all cars are detected correctly but due to their shadows their final detection is merged and the cars are detected to accumulate. In video 2, the fencing in stairs and the upper left corner is detected. When the object (a person here) moves towards the camera, the complete body could not be identified, whereas in video 3,

vehicles at the upper left corner should not be detected but they are detected incorrectly.

7.1.5. Modified Frequency Tuned Spatial Model II. The results shown in Figures 14 and 15 validate our proposed algorithm in detecting moving objects in video frame sequence. In video 2, 3, 5, and 6, all moving objects are detected decently. Some

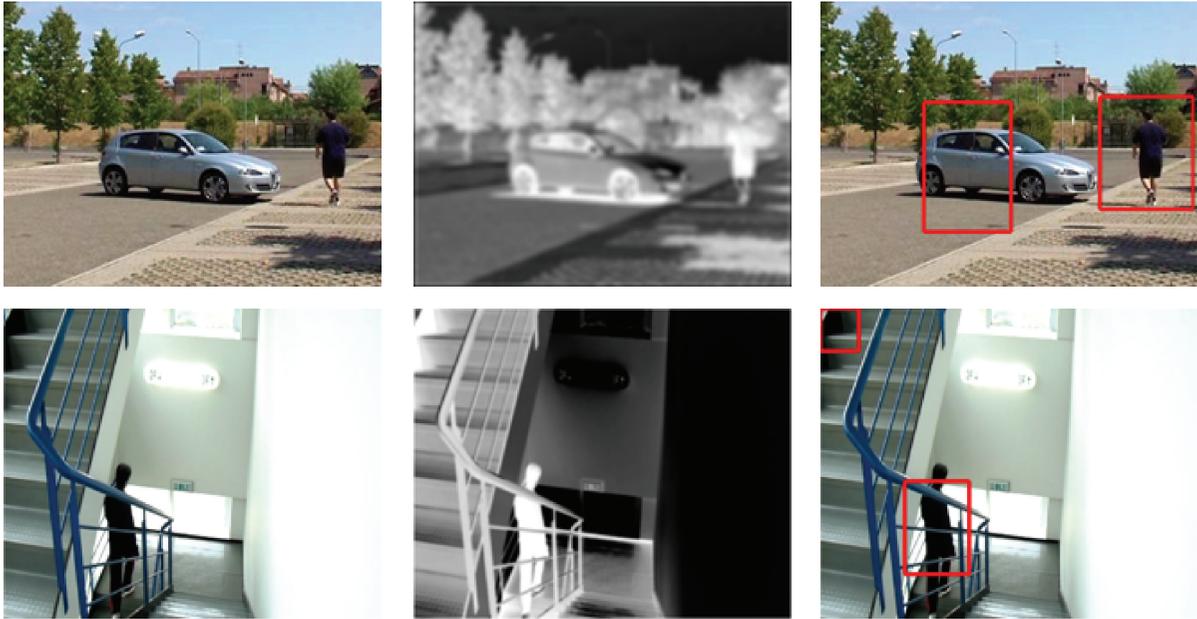


FIGURE 10: Videos (1) and (2): results using Gaussian spatial model by coalescing Laplacian of Gaussian (LoG) filter and Gaussian filter for different videos.

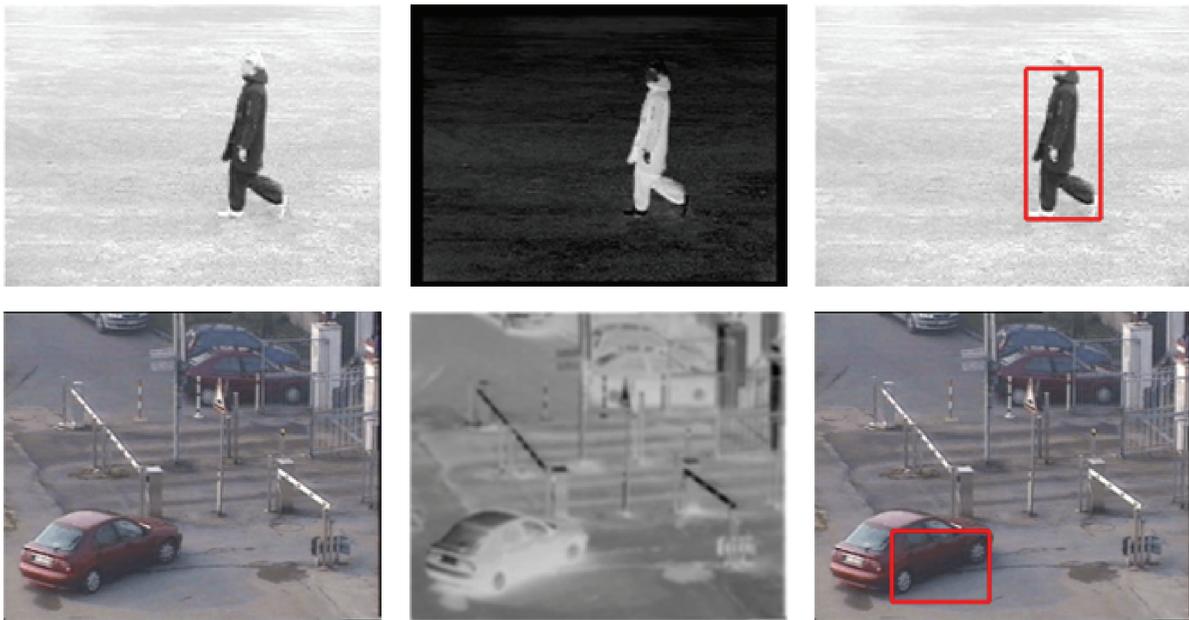


FIGURE 11: Videos (3) and (4): results using Gaussian spatial model by coalescing Laplacian of Gaussian (LoG) filter and Gaussian filter for different videos.

of the videos may define false objects like that in video 1 and video 4, where the upper left corner and the lower left corner, respectively, are wrongly detected.

7.2. Combined Approach (Combining the Bottom-Up and Top-Down Cues)

7.2.1. Integrating Colour Saliency with Texture Feature (Local Binary Pattern (LBP)). Figure 16 shows the result of moving

object detection in video frames. These results demonstrate that the proposed approach can successfully distinguish the moving objects from video sequences. In video 1, the person is detected correctly but in some of the video frames, the upper left corner is detected as a moving object. In both videos 1 and 2, the person's face and upper body are located perfectly.

7.2.2. Integrating Modified Frequency Tuned Spatial Model I with Histogram of Oriented Gradient (HOG) [48]. Figures 17

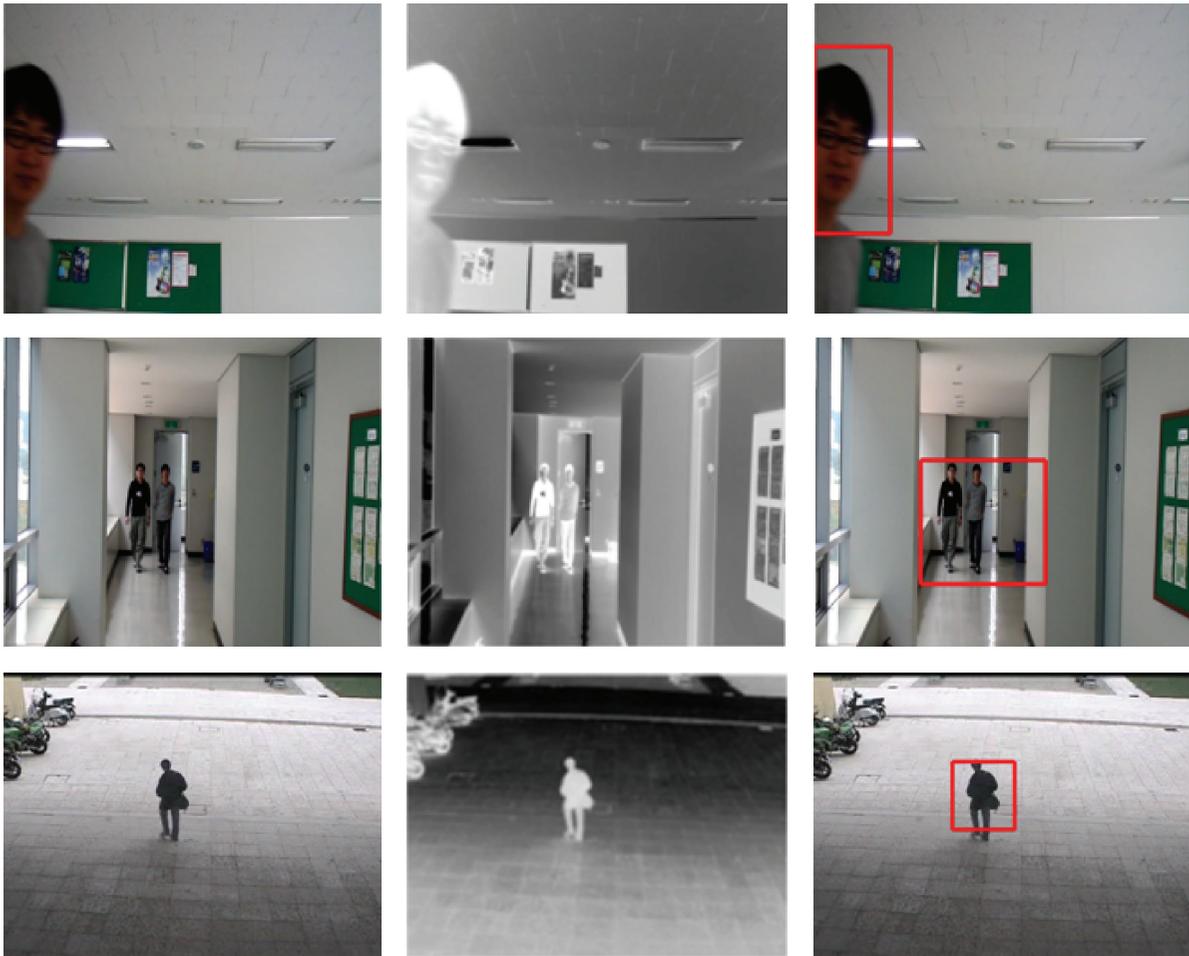


FIGURE 12: Videos (5), (6), and (7): results using Gaussian spatial model by coalescing Laplacian of Gaussian (LoG) filter and Gaussian filter for different videos.

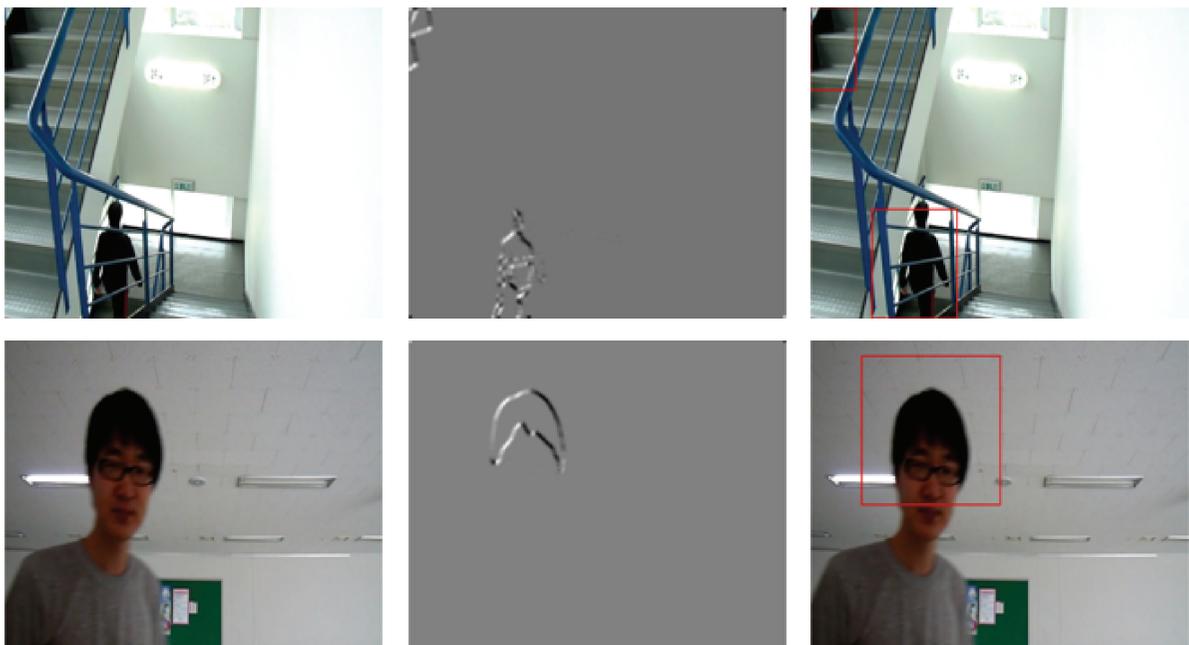


FIGURE 13: Videos (1) and (2): results using gradient saliency by utilizing Sobel operator.

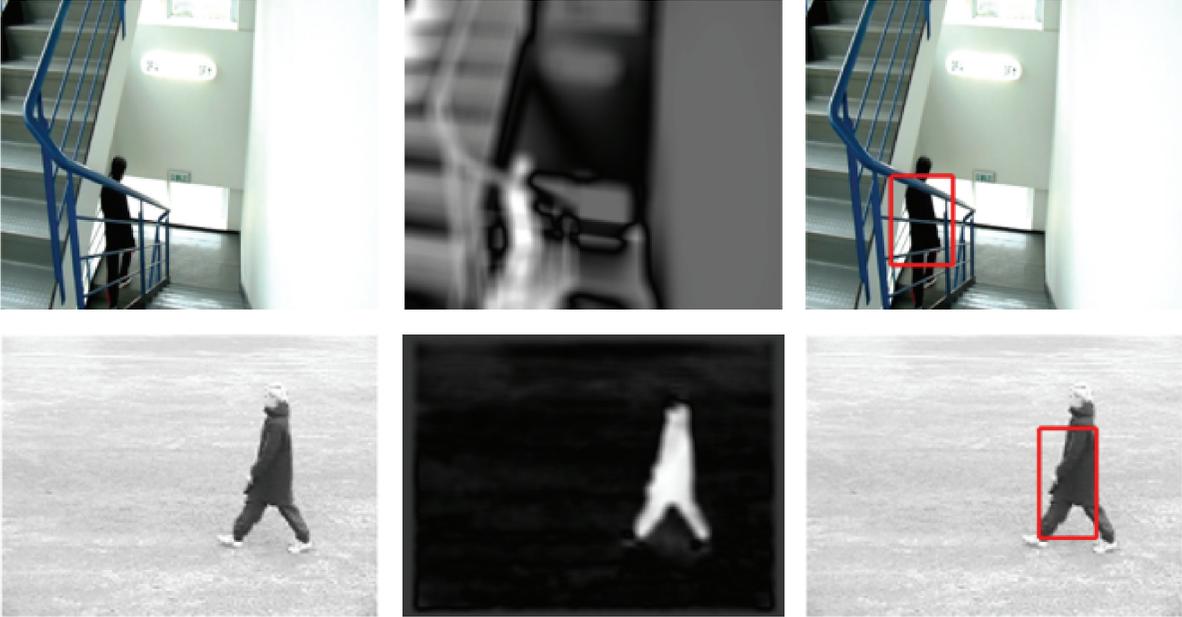


FIGURE 14: Videos (1) and (2): results of modified frequency tuned spatial model II.

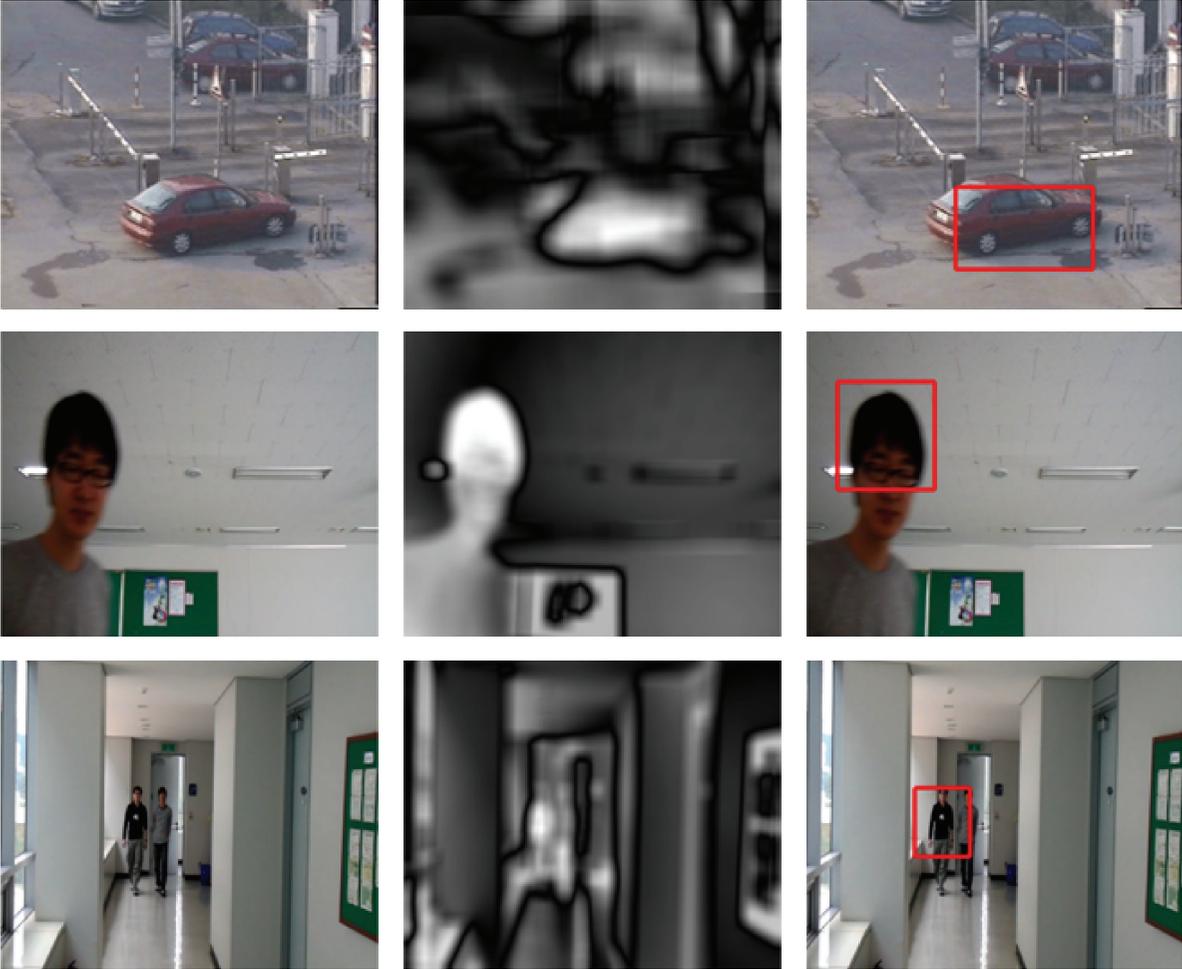


FIGURE 15: Videos (3) and (4): results of modified frequency tuned spatial model II.

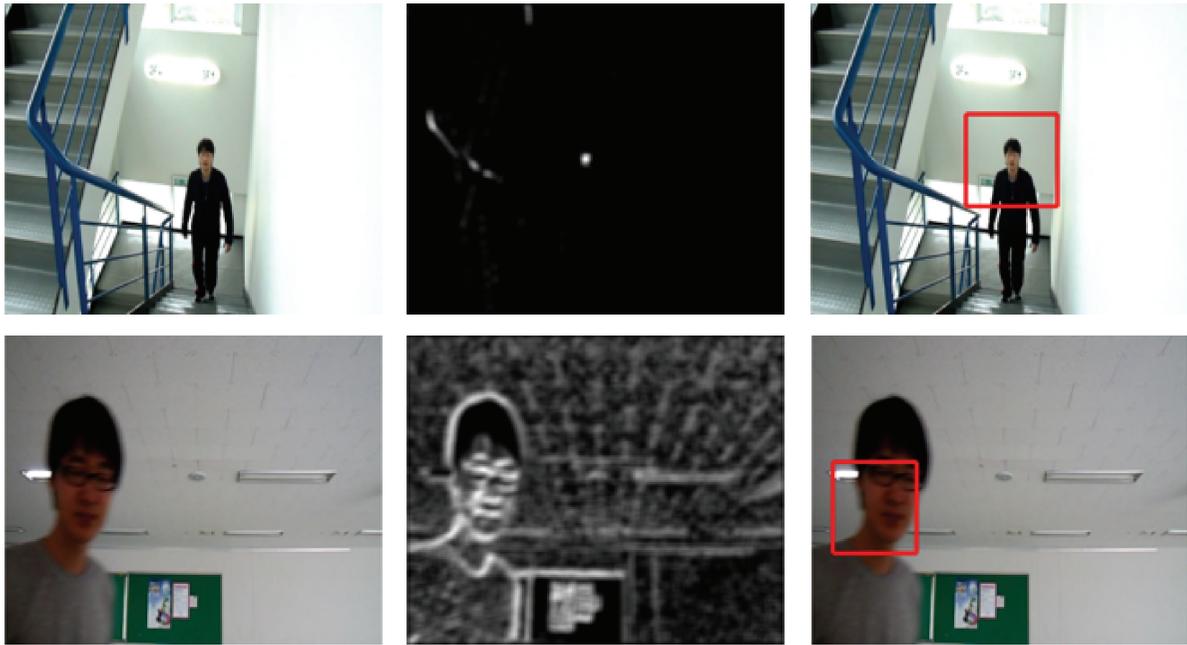


FIGURE 16: Videos (1) and (2): results using integrating colour saliency with texture feature (local binary pattern (LBP) for different videos.

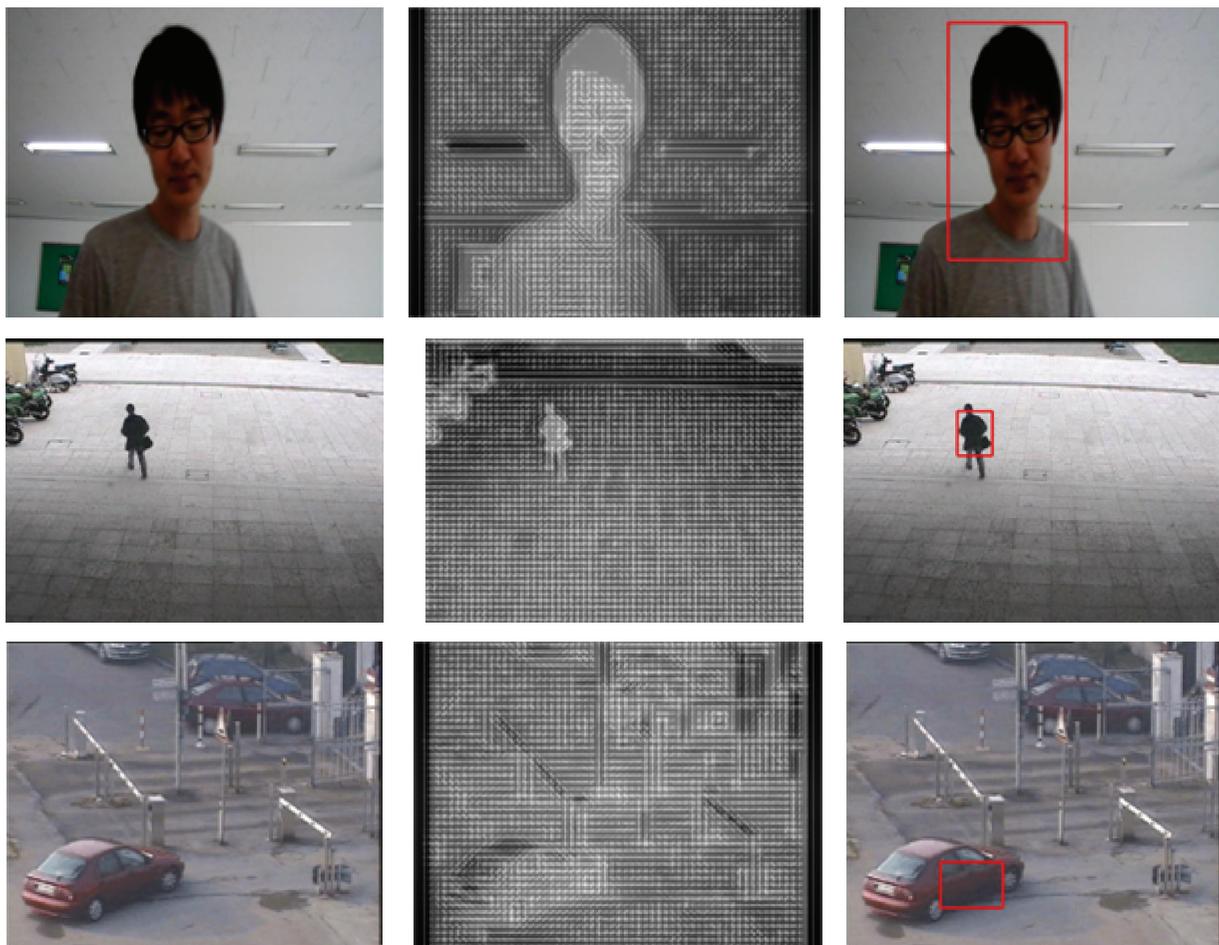


FIGURE 17: Videos (1), (2), and (3): results of integrating modified frequency tuned spatial model I with histogram of oriented gradient (HOG) [48].

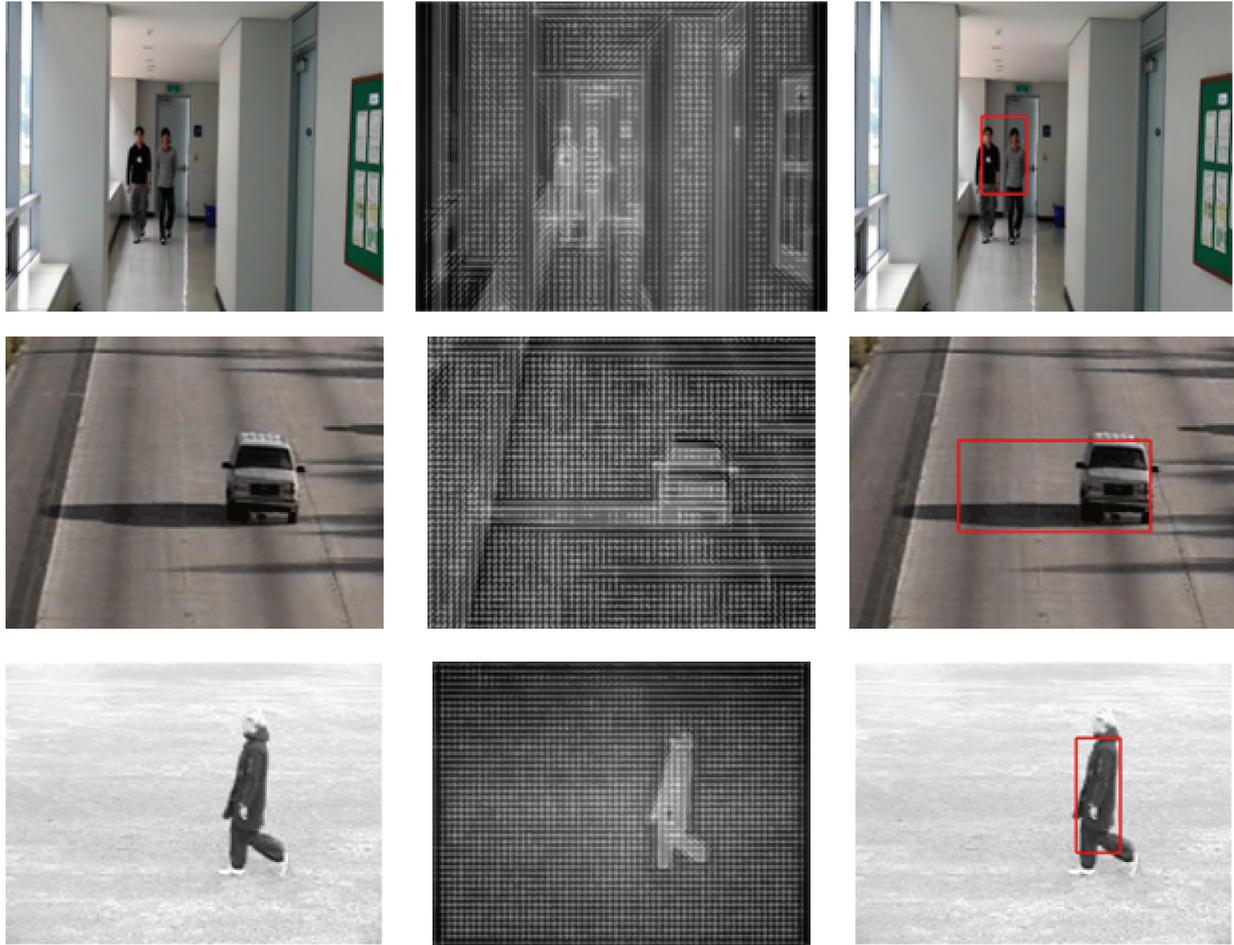


FIGURE 18: Videos (4), (5), and (6): results of integrating modified frequency tuned spatial model I with histogram of oriented gradient (HOG) [48].

TABLE 2: Comparison of time analysis of proposed approaches.

Method name	Execution time (sec)
Modified frequency tuned spatial model I	0.842
Visual saliency based on colour image (colour saliency [38])	0.670
Gaussian spatial model by coalescing Laplacian of Gaussian (LoG) filter and Gaussian filter	0.68
Gradient saliency by utilizing Sobel operator	0.620
Modified frequency tuned spatial model II	1.360
Integrating colour saliency with texture feature (local binary pattern (LBP))	0.7
Integrating modified frequency tuned spatial model I with histogram of oriented gradient (HOG)	0.849
Gaussian mixture model (GMM)	1.560
Kernel density estimation (KDE)	1.489

and 18 show the results of integrating modified frequency tuned spatial model with HOG features. From results, it can be defined that the algorithm properly detects the moving object from background. Overall saliency map after combining bottom-up cues with top down cues is shown in the middle column of all results.

8. Comparative Analysis

Comparison is done based on time, accuracy, and results of the proposed algorithm. In Table 2, a particular method is defined with execution time of that method. Maximum time is taken by saliency method using frequency tuned approach

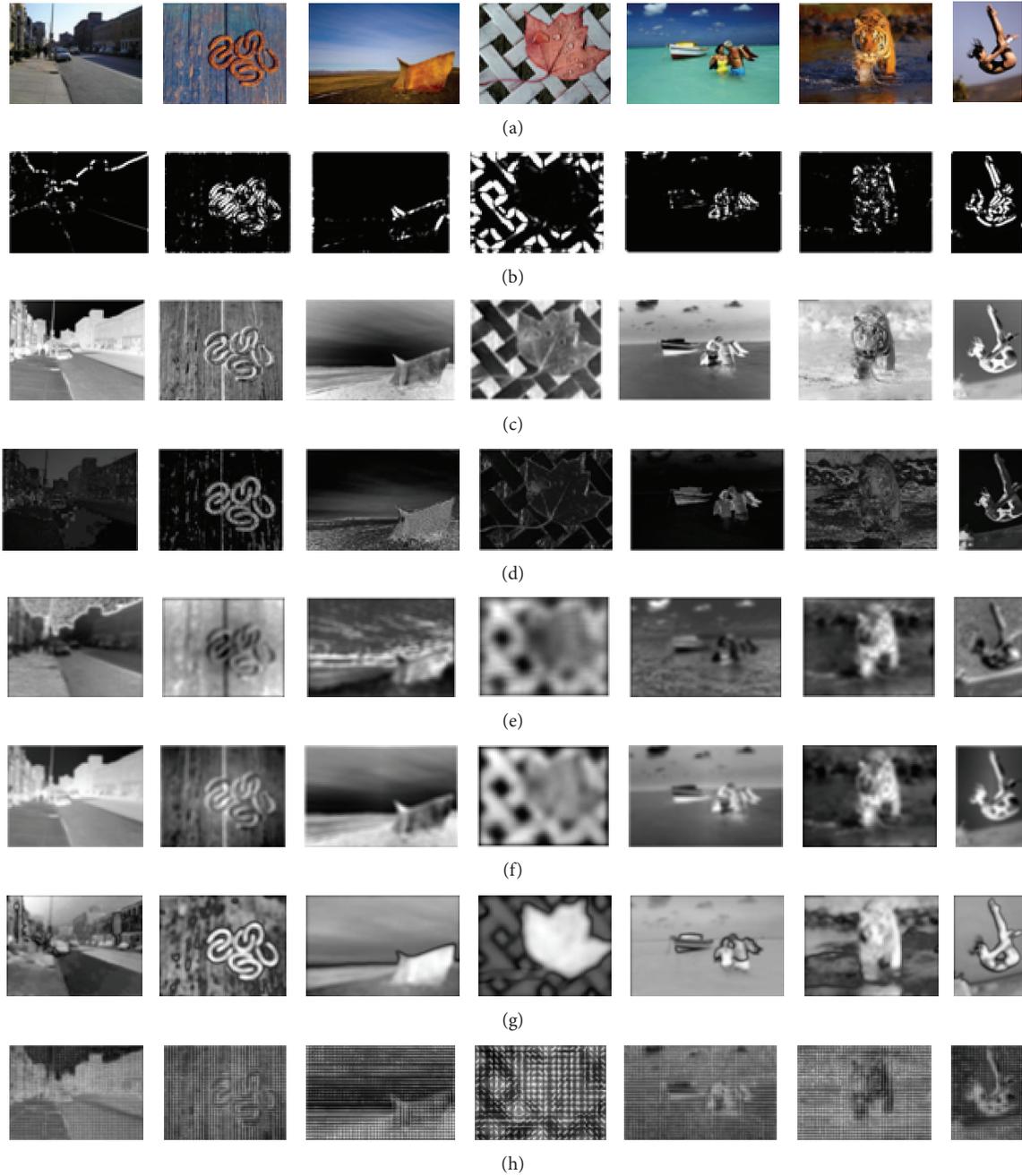


FIGURE 19: Comparative analysis for static images: (a) original image, (b) gradient saliency by utilizing Sobel operator, (c) modified frequency tuned spatial model II, (d) integrating colour saliency with texture feature (local binary pattern (LBP)), (e) modified frequency tuned spatial model I, (f) visual saliency based on colour image (colour saliency [38]), and (h) integrating modified frequency tuned spatial model I with histogram of oriented gradient (HOG) [48].

TABLE 3: Comparison of accuracy of proposed approaches.

Method 1	Method 2	Method 3	Method 4	Method 5	Method 6	Method 7
Modified frequency tuned spatial model I	Visual saliency based on colour image (colour saliency [38])	Gaussian spatial model by coalescing Laplacian of Gaussian (LoG) filter and Gaussian filter	Gradient saliency by utilizing Sobel from static image	Modified frequency tuned spatial model II	Integrating colour saliency with texture feature (local binary pattern (LBP))	Integrating modified frequency tuned spatial model I with histogram of oriented gradient (HOG)

TABLE 4: Comparison of accuracy of proposed approaches.

Video name	Total frames	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6	Method 7	GMM [10]	KDE [15]
ATM	728	651	511	597	604	540	623	674	487	503
Cam	4567	4329	4018	4111	3986	3898	4178	4415	3795	3852
Campus_raw	1179	1094	993	1021	956	997	1024	1099	885	924
EPC-200-day-bird	1115	1005	994	973	989	898	1023	1097	774	855
HALLWAY_A	661	606	545	506	469	511	595	614	363	425
highwayLraw	440	389	328	298	305	340	370	398	286	234
person15_walking_dl_uncomp	741	689	602	624	597	611	662	704	583	604
Stair	707	644	578	603	592	612	630	675	520	563
visor_1197284015696_07_fumogeno4	2024	1965	1901	1898	1830	1920	1946	1990	1719	1756
Total	12162	11435	10470	10631	10328	10327	11051	11739	9412	9716
Accuracy (%)		94.02	86.09	87.41	84.92	84.91	90.86	96.52	77.39	79.89

TABLE 5: Comparison of recall of proposed approaches.

Video name	Total Objects	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6	Method 7	GMM [10]	KDE [15]
ATM	698	0.9327	0.7321	0.8553	0.8653	0.7736	0.8926	0.9656	0.6977	0.7206
Cam	4431	0.977	0.9068	0.9278	0.8996	0.8797	0.9429	0.9964	0.8497	0.8693
Campus_raw	1104	0.9909	0.8995	0.9248	0.8659	0.9031	0.9275	0.9508	0.7506	0.7837
EPC-200-day-bird	1100	0.9136	0.9036	0.8845	0.8991	0.8164	0.93	0.9955	0.8016	0.8370
HALLWAY_A	621	0.9758	0.8776	0.8148	0.7552	0.8229	0.9581	0.9973	0.7036	0.7773
highwayl_raw	410	0.9488	0.8	0.7268	0.7439	0.8293	0.9024	0.9887	0.5845	0.6844
person15_walking_dl_uncomp	705	0.9773	0.8539	0.8851	0.8468	0.8667	0.9390	0.9707	0.6976	0.5707
stair	678	0.9499	0.8525	0.8894	0.8732	0.9027	0.9292	0.9574	0.7376	0.7986
visor_1197284015696_07_fumogeno4	1995	0.985	0.9529	0.9514	0.9173	0.9624	0.9754	0.9975	0.8617	0.8802
Overall Recall		0.9612	0.8643	0.8733	0.8518	0.8619	0.9330	0.98	0.7427	0.7691

using HSI colour model and minimum time is taken by method gradient saliency using Sobel from static image.

In Table 4, a comparative analysis is made based on accuracy of the proposed approach. Method names are defined in a respective manner as shown in Table 3. The first column specifies the name of the video used for checking accuracy of various methods and the second column defines a total number of image frames in the video. The rest of the columns represent the number of frames in which a moving object is correctly detected with respect to a particular method. Accuracy with the concerned approaches is defined in the last row of Table 4 and this shows that method 1 (frequency tuned spatial model using aggregation of averaging filter and Gaussian filter) is having efficient and satisfactory results as compared to other algorithms.

In Figure 19, the results are shown for static image using the proposed approaches and it is found that it is performing excellent on static image as well as video. Evaluation of results is performed using recall parameter as shown in Table 5.

9. Conclusion

Moving object detection is done using visual attention in indoor as well as outdoor conditions. The present paper proposes different techniques for detection of moving objects. In the proposed method, the moving objects are detected by seven different methods. Our test results show that frequency tuned spatial model using combination of averaging filter and Gaussian filter is effective in representing the moving object in a static background. Frequency tuned spatial model using combination of averaging filter and Gaussian filter method detects the moving object when there are slight variations or a dynamic environment in the background. The proposal for detecting moving objects using visual attention provides positive error rates with low grade possibility of being false. It also supports a successful identification of the moving object. Effectiveness of the proposed algorithm is proved by their results and comparison with other methods. Future enhancements can be accomplished by applying these approaches to the moving backgrounds.

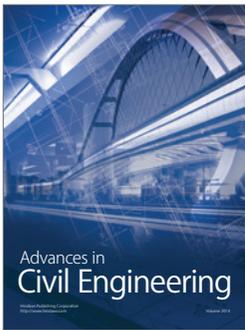
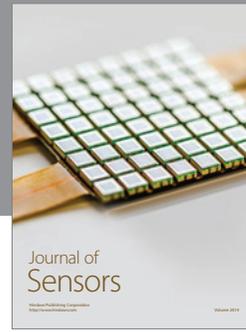
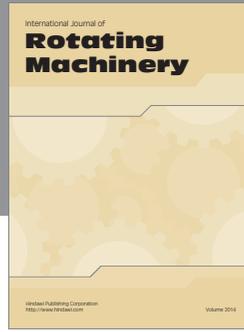
Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [2] M. Oral and U. Deniz, "Centre of mass model—a novel approach to background modelling for segmentation of moving objects," *Image and Vision Computing*, vol. 25, no. 8, pp. 1365–1376, 2007.
- [3] A. Manzanera and J. C. Richefeu, "A new motion detection algorithm based on Sigma-Delta background estimation," *Pattern Recognition Letters*, vol. 28, no. 3, pp. 320–328, 2007.
- [4] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, pp. 2246–2252, Fort Collins, Colo, USA, June 1999.
- [5] B. Coifman, D. Beymer, P. McLauchlan, and J. Malik, "A real-time computer vision system for vehicle tracking and traffic surveillance," *Transportation Research C: Emerging Technologies*, vol. 6, no. 4, pp. 271–288, 1998.
- [6] J. Kong, Y. Zheng, Y. Lu, and B. Zhang, "A novel background extraction and updating algorithm for vehicle detection and tracking," in *Proceedings of the 4th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD '07)*, vol. 3, pp. 464–468, August 2007.
- [7] I. Mikić, M. Trivedi, E. Hunter, and P. Cosman, "Human body model acquisition and tracking using voxel data," *International Journal of Computer Vision*, vol. 53, no. 3, pp. 199–223, 2003.
- [8] M. Sigari, N. Mozayani, and H. Pourreza, "Fuzzy running average and fuzzy background subtraction: concepts and application," *International Journal of Computer Science and Network Security*, vol. 8, no. 2, pp. 138–143, 2008.
- [9] F. El Baf, T. Bouwmans, and B. Vachon, "Fuzzy integral for moving object detection," in *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ '08)*, pp. 1729–1736, June 2008.
- [10] C. I. Patel and R. Patel, "Gaussian mixture model based moving object detection from video sequence," in *Proceedings of the International Conference & Workshop on Emerging Trends in Technology (ICWET '11)*, pp. 698–702, ACM, New York, NY, USA, February 2011.
- [11] F. El Baf, T. Bouwmans, and B. Vachon, "Type-2 fuzzy mixture of Gaussians model: application to background modeling," in *International Symposium on Visual Computing (ISVC'08)*, vol. 1, pp. 772–781, Las Vegas, Nev, USA, December 2008.
- [12] S.-C. S. Cheung and C. Kamath, "Robust background subtraction with foreground validation for urban traffic video," *Eurasip Journal on Applied Signal Processing*, vol. 2005, pp. 2330–2340, 2005.
- [13] N. Nihan, E. Hallenbeck, J. Zheng, and Y. Wang, "Extracting roadway background image: a mode based approach," *Journal of Transportation Research Report*, vol. 1994, pp. 82–88, 2006.
- [14] A. Elgammal, D. Harwood, and L. Davis, "Nonparametric model for background subtraction," in *Proceedings of the 6th European Conference on Computer Vision*, pp. 751–767, IEEE, Dublin, Ireland, July 2000.
- [15] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of the IEEE*, vol. 90, no. 7, pp. 1151–1163, 2002.
- [16] J. Carranza, C. Theobalt, M. A. Magnor, and H. P. Seidel, "Free view point video of human actors," *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 569–577, 2003.
- [17] S. Y. Elhabian, S. H. Ahmed, and K. M. El-Sayed, "Moving object detection in spatial domain using background removal techniques—state-of-art," *Recent Patents on Computer Science*, vol. 1, no. 1, pp. 32–54, January 2008.
- [18] T. Horprasert, I. Haritaoglu, C. Wren, D. Harwood, L. Davis, and A. Pentland, "Real-time 3d motion capture," in *Proceedings of the Workshop on Perceptual User Interfaces*, pp. 87–90, 1998.

- [19] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, 1997.
- [20] O. Barnich and M. van Droogenbroeck, "ViBe: a universal background subtraction algorithm for video sequences," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709–1724, 2011.
- [21] K. Huang, J. Zhang, S. Xu, and T. Luo, "Accurate moving target detection based on background subtraction and susan," *International Journal of Computer and Electrical Engineering*, vol. 4, no. 4, pp. 436–439, 2012.
- [22] C. Schüldt, I. Laptev, and B. Caputo, "Recognizing human actions: a local SVM approach," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 3, pp. 32–36, August 2004.
- [23] R. Rosenholtz, "A simple saliency model predicts a number of motion popout phenomena," *Vision Research*, vol. 39, no. 19, pp. 3157–3163, 1999.
- [24] A. Oliva and A. Torralba, "The role of context in object recognition," *Trends in Cognitive Sciences*, vol. 11, no. 12, pp. 520–527, 2007.
- [25] E. Blaser, G. Sperling, and Z.-L. Lu, "Measuring the amplification of attention," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 20, pp. 11681–11686, 1999.
- [26] V. Navalpakkam and L. Itti, "Modeling the influence of task on attention," *Vision Research*, vol. 45, no. 2, pp. 205–231, 2005.
- [27] S. Frintrop, G. Backer, and E. Rome, "Goal-directed search with a top-down modulated computational attention system," in *Proceedings of the Annual Meeting of the German Association for Pattern Recognition (DAGM '05)*, pp. 117–124, Wien, Austria, 2005.
- [28] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980.
- [29] W. Einhäuser, U. Rutishauser, and C. Koch, "Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli," *Journal of Vision*, vol. 8, no. 2, article 2, 2008.
- [30] J. K. Tsotsos and N. D. B. Bruce, "Saliency based on information maximization," in *Advances in Neural Information Processing Systems 18*, Y. Weiss, B. Scholkopf, and J. Platt, Eds., pp. 155–162, MIT Press, Cambridge, Mass, USA, 2006.
- [31] A. Torralba, M. S. Castelano, A. Oliva, and J. M. Henderson, "Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search," *Psychological Review*, vol. 113, no. 4, pp. 766–786, 2006.
- [32] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.
- [33] K. A. Ehinger, B. Hidalgo-Sotelo, A. Torralba, and A. Oliva, "Modelling search for people in 900 scenes: a combined source model of eye guidance," *Visual Cognition*, vol. 17, no. 6-7, pp. 945–978, 2009.
- [34] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, IEEE Computer Society, Minneapolis, Minn, USA, June 2007.
- [35] J. Harel, C. Koch, and P. Perona, "Graphbased visual saliency," in *Advances in Neural Information Processing Systems 19*, pp. 545–552, MIT Press, Cambridge, Mass, USA, 2007.
- [36] D. A. Poggel, H. Strasburger, and M. MacKeben, "Cueing attention by relative motion in the periphery of the visual field," *Perception*, vol. 36, no. 7, pp. 955–970, 2007.
- [37] R. Achantay, S. Hemamiz, F. Estraday, and S. Süssstrunky, "Frequency-tuned salient region detection," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 1597–1604, June 2009.
- [38] C. Huang, Q. Liu, and S. Yu, "Regions of interest extraction from color image based on visual saliency," *The Journal of Supercomputing*, vol. 58, no. 1, pp. 20–33, 2011.
- [39] C. Huang, Q. Liu, and S. Yu, "Automatic central object extraction from color image," in *Proceedings of the International Conference on Information Engineering and Computer Science (ICIECS '09)*, vol. 5, pp. 1–4, IEEE, Wuhan, China, December 2009.
- [40] Q. Zhou, L. Ma, M. Celenk, and D. Chelberg, "Content-based image retrieval based on ROI detection and relevance feedback," *Multimedia Tools and Applications*, vol. 27, no. 2, pp. 251–281, 2005.
- [41] C. Huang, Q. Liu, and S. Yu, "Regions of interest extraction from color image based on visual saliency," *Journal of Supercomputing*, vol. 58, no. 1, pp. 20–33, 2011.
- [42] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, Addison-Wesley Longman, Boston, Mass, USA, 1st edition, 1992.
- [43] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice-Hall, Upper Saddle River, NJ, USA, 3rd edition, 2006.
- [44] K. Ilanthodi, *Color Image Analysis for Staining Intensity Quantification—Its Application to Medical Research and Diagnostic Purposes*, Manipal University, Manipal Institute of Technology, 2012, <http://shodhganga.inflibnet.ac.in/handle/10603/5042>.
- [45] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [46] D. S. Bolme, J. R. Beveridge, M. Teixeira, and B. A. Draper, "The csu face identification evaluation system: its purpose, features and structure," in *Proceedings of the International Conference on Vision Systems (ICVS '03)*, pp. 304–311, 2003.
- [47] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [48] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 886–893, IEEE Computer Society, San Diego, Calif, USA, June 2005.
- [49] J. Wang, S. Kumar, and S.-F. Chang, "Sequential projection learning for hashing with compact codes," in *Proceedings of the International Conference on Machine Learning (ICML '10)*, pp. 1127–1134, June 2010.
- [50] H. Sohn, W. de Neve, and Y. M. Ro, "Privacy protection in video surveillance systems: analysis of subband-adaptive scrambling in JPEG XR," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 2, pp. 170–177, 2011.
- [51] R. Vezzani and R. Cucchiara, "Video surveillance online repository (ViSOR): an integrated framework," *Multimedia Tools and Applications*, vol. 50, no. 2, pp. 359–380, 2010.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

