Computer Modeling in Engineering & Sciences
DOI:10.32604/cmes.2020.09397
Article

An Improved Non-Parametric Method for Multiple Moving Objects Detection in the Markov Random Field

Qin Wan[1,2,*], Xiaolin Zhu[1], Yueping Xiao[1], Jine Yan[1], Guoquan Chen[1] and Mingui Sun[3]

[1]College of Electric and Information Engineering, Hunan Institute of Engineering, Xiangtan, 411104, China
[2]National Engineering Research Laboratory for Robot Vision Perception and Control, Hunan University, Changsha, China
[3]The Laboratory for Computational Neuroscience, University of Pittsburgh, Pittsburgh, PA 15260, USA
*Corresponding Author: Qin Wan. Email: wanqin_10@126.com
Received: 10 December 2019; Accepted: 01 April 2020

Abstract: Detecting moving objects in the stationary background is an important problem in visual surveillance systems. However, the traditional background subtraction method fails when the background is not completely stationary and involves certain dynamic changes. In this paper, according to the basic steps of the background subtraction method, a novel non-parametric moving object detection method is proposed based on an improved ant colony algorithm by using the Markov random field. Concretely, the contributions are as follows: 1) A new non-parametric strategy is utilized to model the background, based on an improved kernel density estimation; this approach uses an adaptive bandwidth, and the fused features combine the colours, gradients and positions. 2) A Markov random field method based on this adaptive background model via the constraint of the spatial context is proposed to extract objects. 3) The posterior function is maximized efficiently by using an improved ant colony system algorithm. Extensive experiments show that the proposed method demonstrates a better performance than many existing state-of-the-art methods.

Keywords: Object detection; non-parametric method; markov random field

1 Introduction

Moving object detection has been widely applied in fields such as video surveillance [1], human-machine interaction [2], and autonomous navigation of robots [3] in the last two decades. Background subtraction is a typical method used to detect moving objects in visual surveillance systems, and it is considered as the first step in the detection algorithms for multiple moving objects [2,4,5]. For a relatively static background, background subtraction is a simple and effective method for motion segmentation. However, a stationary camera does not reflect a completely stationary scene in the real world, and dynamic scenes containing lights, rains, swaying trees, and fountains change gradually. Therefore, the background subtraction method should be improved to adapt to such scenes.

To apply the background subtraction technique, building the background model is a key step to describe the outdoor scenes. Two types of methods can be used for building a background model: parametric and non-parametric models. The parametric background model usually builds the scene as a particular distribution.

However, the background varies gradually, and thus, it cannot be modelled appropriately using the parametric background model. Wren et al. [6] proposed a single Gaussian background modelling method and assumed that the pixel values of each pixel in the background obey the Gaussian distribution with the change in time. This method exhibits excellent performance in certain scenarios; however, it cannot adapt to outdoor dynamic scenes. Stauffer et al. [7] proposed a mixture Gaussian background modelling method. This algorithm establishes multiple Gaussian distribution models and updates the relevant parameters to achieve real-time fitting of the multimodal distribution in complex scenes simultaneously. However, the detection result of this method is determined by the selection of the parameters, and the robustness of the algorithm needs to be improved. As a result, the parametric background model approach fails to obtain sensitive detection results when the scene contains gradual changes and high-frequency variations.

Another type of background model is the non-parametric background model, which does not build the scenes as having a particular distribution, and this non-parametric model can describe the scenes robustly. Although the non-parametric models appear to be a reasonable choice for background modelling, it is difficult to choose the suitable bandwidth of the kernel density estimation (KDE) [8], and the approach is usually too costly to perform in real time. Long-term and short-term KDE methods are used to obtain the background and foreground models in reference [8], respectively. However, because it is difficult to define the appropriate bandwidth of the KDE, such KDE methods are also time-costly algorithms.

In the past two decades, the background subtraction approach has been widely used for object detection and segmentation. However, the existing background subtraction methods still face the following disadvantages.

1. The parametric background modelling method usually requires that the pixels in the background obey the underlying distribution model, and this method thus cannot adapt to arbitrarily complex data distributions, thereby requiring explicit estimation parameters.
2. The non-parametric method in the background subtraction approach also has some drawbacks. The dynamic features for the background modelling are difficult to be selected as bandwidths. In addition, in the object extraction, the objects are detected considering the threshold, which is not accurate.

To address these problems, our method is inspired by the statistical probability model methods and multimodal features. The fusion of multiple features helps model the visual scenes. A novel Markov random field framework is proposed to seek the optimal labels of the image pixels. An improved ant colony system algorithm is employed to maximize the posterior function. In particular, Fig. 1 shows the outline architecture of the proposed method, and the distinctive features of this method are as follows:

1. For the background modelling, an adaptive non-parametric method with KDE is proposed, which has variable bandwidths. The colours and gradients are combined with the positions and utilized as the features in a higher-dimensional space to model the visual scenes, which is also used for describing the foreground model.
2. For extracting the objects, we consider that the label of a pixel is associated with its neighbourhoods during the decision step regarding the labels of the pixels. In addition, because the Markov random field method is a process seeking the optimal labels of the image pixels, a Markov random field framework is computed to make decisions by using the foreground and background models.
3. The optimal resolution is acquired by using an improved ant colony system algorithm, and the method is implemented for the ACS in the Markov random field framework.

This reminder of this paper is organized as follows: Section II provides a brief description of the literature pertaining to the background subtraction technique. Section III describes the modelling of the

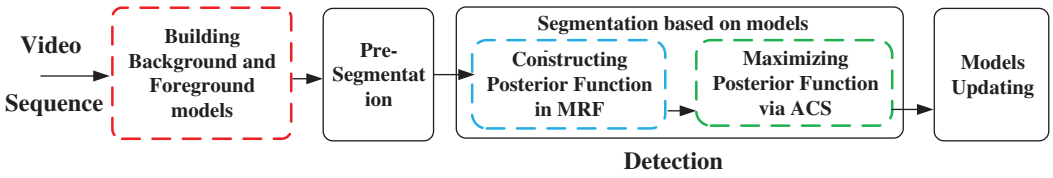

Figure 1: Architectural outline of the proposed method

background via non-parametric methods. Section IV presents the Markov random field algorithm. In Section V, the improved ant colony method is proposed for optimizing the posterior function. Section VI describes the experiments performed on different sequences, which demonstrates the performance of the proposed method compared to that of several traditional and new methods. Finally, Section VII provides the conclusions and scope for future work.

2 Related Works

The background subtraction and foreground extraction techniques are the methods most commonly used to detect moving objects in video sequences [9]. The simplest approach to detect moving objects involves using the inter-frame difference method [10]. This approach mainly subtracts two image pixel values in two adjacent frames or several frames in the video stream and extracts the motion region in the image by manual thresholding. However, this technique can only be used for static cameras and it is extremely sensitive to changes in the dynamic scenes. Therefore, as reported in the related literature, the background subtraction method has been improved to adapt to complex dynamic scenes.

In the context of this problem, the existing models based on background subtraction may be classified as being based on either the parametric or non-parametric background models. The parametric background model usually builds the scene as a particular distribution such as a multiple Gaussian distribution [11], Gaussian mixture distribution [12] hidden Markov approach [13] or other probabilistic models [14,15,16] for foreground detection. The single Gaussian distribution [17] background model is suitable for single-modal scenes, in which the model is established considering a single Gaussian distribution for the colour distribution of each image point. To solve the problem of random image noises, the Gaussian mixture distribution model is proposed for multiple modes, and it can obtain satisfactory results when the scenes are slightly dynamic. However, if the background varies gradually, it cannot be modelled appropriately by using the parametric background model. An accurate background model is generated using the approach reported in [1,18] even when unclear or blurred frames are present. The region-based MoG (RMoG) [19] considers the pixels near the objects to model the background. Generally, the method of the parametric background model fails when the scene contains gradual changes, because the variations in the scenes cause high frequency variations.

To solve the various problems of the parametric background model in dynamic complex scenes, non-parametric methods were developed as an alternative approach to model the backgrounds. The most attractive non-parametric method for modelling the distributions of the background is the kernel density estimation (KDE) technique [8], which performs the processes using Gaussian kernels to improve the building of the background model. This method can be improved to adapt to complex scenes. The Codebook algorithm [20] uses the three-dimensional colour model to calculate the matching degree between the current pixel and the corresponding Codebook model, which considerably reduces the computational time. However, these algorithms cannot update and detect the number of moving targets when the background changes. To improve the sensitivity and adaptability of the algorithm in complex scenes, SuBSENSE [21] combines the feedback mechanism of the ViBe [22] method, and the pixel-based adaptive segmenter (PBAS) [23] is used to perform a pixel-level segmentation based on the non-parametric statistical distribution model. However, the algorithm has a high complexity and high memory

usage. Background and foreground models were formulated in [8] by using a developed KDE method in the long term and short term, respectively. However, because it is difficult to define the appropriate bandwidth of the KDE, KDE approaches are also time-costly algorithms. The background and foreground were both modelled in [2] to reduce the computational complexity. In addition, Zhang [24] proposed the setting of a threshold for each pixel according to the dynamic nature of the pixel, based on the KDE adaptively. In [25], the differences between the Gaussian and non-parametric method were used for object detection.

To model dynamic backgrounds adaptively, deep learning methods [26,27,28] and RGB-D data [3,29,30] have been used for background segmentation. Convolutional networks are powerful visual models in the field of target detection and combine semantic information from a deep, coarse layer with appearance information from a shallow, fine layer to produce accurate and detailed detection results. Convolutional-network-based methods exhibit a better accuracy and robustness than those of traditional methods for detecting moving targets in complex scenes. The RGB-D data can provide the geometric position information, and the depth values can represent the distance between each pixel and the depth camera in the real world. The colour information combined with the depth position information can effectively solve the occlusion problem and improve the accuracy and robustness of the moving target segmentation. In general, establishing an accurate background or foreground model is essential for background subtraction when the background involves high-frequency variations.

In other words, detecting or extracting moving objects in a complex environment is still a challenging problem. To solve such problems, the proposed method is different from the abovementioned methods in the following three contexts: 1) The background and foreground models are established by using the non-parametric KDE method; 2) the posterior function is constructed in the MRF; 3) the posterior function is maximizing via an improved ACS.

3 Modeling the Background

For modelling the multi-variate probability distributions in dynamic scenes, the non-parametric kernel estimation can acquire the probability distribution with no fixed assumptions. Consider that at the current time t, the background set is represented as Φ_b. For the pixel $\mathbf{x}$ at time t, the estimator is defined as in [31] to obtain the probability of background:

$$\hat{p}(\mathbf{x}|\Phi_b) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\|\mathbf{B}\|^{1/2}(2\pi)^{d/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}_i)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_i)\right) \tag{1}$$

We propose a method for selecting the bandwidth B and features of the measurements in Eq. (1). The variable bandwidth B is determined by the uncertainties in the sample measurement $\mathbf{x}_i$ and the estimated measurement $\mathbf{x}$. Next, seven features of the measurements are utilized: two features for the colours, three features for the gradients and two features for the positions.

To model the background accurately, the bandwidth should reflect the variable local variances. The optimal value of B is larger in the region having sparse data and smaller in the region having dense data. Consequently, we introduce an approach of choosing the bandwidth matrix that changes with the distributions of the sample set and the estimation set. Based on the method proposed by Mittal [32], we define the bandwidth matrix as $\mathbf{B}(\mathbf{x}_i, \mathbf{x}) = \Sigma_{\mathbf{x}_i} + \Sigma_{\mathbf{x}}$. The non-parametric estimator can be described as:

$$\hat{p}(\mathbf{x}|\Phi_b) = \frac{1}{n(2\pi)^{d/2}}\sum_{i=1}^{n} \frac{1}{\|\Sigma_{\mathbf{x}_i} + \Sigma_{\mathbf{x}}\|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}_i)^T(\Sigma_{\mathbf{x}_i} + \Sigma_{\mathbf{x}})^{-1}(\mathbf{x}-\mathbf{x}_i)\right) \tag{2}$$

where $\Sigma_{\mathbf{x}_i}$ is the covariance matrix of the sample measurement $\mathbf{x}_i$, and $\Sigma_{\mathbf{x}}$ is the covariance matrix of the estimated measurement $\mathbf{x}$. However, in a practical system, since the estimated pixel $\mathbf{x}$ is only a vector in

the feature space, $\Sigma_{\mathbf{x}}$ would not be computed. Based on the mathematical explanation in [32], which proves that Eq. (2) represents the distributions of the sample measurements in the background, $\Sigma_{\mathbf{x}}$ can be computed from the set $\Phi_b^{'} = \{\Phi_b, \mathbf{x}\}$.

The selection of the features should be considered after introducing the general non-parametric kernel estimation. Here, we describe the methods for obtaining such measurements and reducing the computation.

Because colours have the advantage of being invariant to a change in the illumination, r and g, which represent the chromaticity coordinates are used. To reduce the computation, the different components are assumed independently. The uncertainties in the normalized features are

$$\Sigma_{r,g} = \begin{bmatrix} \sigma^2_{xr_c} & 0 \\ 0 & \sigma^2_{xg_c} \end{bmatrix}$$

where $\sigma^2_{xr_c}$ and $\sigma^2_{xg_c}$ are the variances in the different component using the chromaticity coordinates (r, g).

In addition to the colour information, we also use the gradients to describe the pixel features. Consequently, the uncertainty based on the gradient is

$$\Sigma_{gra} = \begin{bmatrix} \sigma^2_{gr} & 0 & 0 \\ 0 & \sigma^2_{gg} & 0 \\ 0 & 0 & \sigma^2_{gb} \end{bmatrix}$$

where σ^2_{gr}, σ^2_{gg}, and σ^2_{gb} are the gradient variances of the chromaticity coordinates (r, g, b).

The pixel positions are also useful for detecting objects. Because the position of a pixel can be stable when the pixel is labelled as a foreground or background pixel, the positions of a pixel containing x and y coordinates can also be utilized as features:

$$\Sigma_p = \begin{bmatrix} \sigma^2_{xx} & 0 \\ 0 & \sigma^2_{xy} \end{bmatrix}$$

The pixel is thus described as $\mathbf{x}_i = (\, x_{rc} \quad x_{gc} \quad x_{gr} \quad x_{gg} \quad x_{gb} \quad x_{xx} \quad x_{xy} \,)$ (i = 1, 2, ……, n.). x_{rc} and x_{gc} are the colour values of the pixel; x_{gr}, x_{gg}, and x_{gb} are the gradients; and x_{xx} and x_{xy} are the positions. Assuming the colour and gradient features are independent, the covariance $\Sigma_{\mathbf{x}}$ may be estimated:

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} \Sigma_{r,g} & 0 & 0 \\ 0 & \Sigma_{gra} & 0 \\ 0 & 0 & \Sigma_p \end{bmatrix}$$

$\Sigma_{\mathbf{x}_i}$ can also be represented on sample data as $\Sigma_{\mathbf{x}}$, which is the value for the variance in the colour channels, gradient components, and position components.

Therefore, the pixel $\mathbf{x}$ is classified using the presetting threshold T as a foreground or background pixel:

$$\hat{p}(\mathbf{x}|\Phi_b) < T \tag{3}$$

The parameter T is the presetting manual threshold, which can be used for segmentation. All the pixels are segmented as foreground or background pixels suitably in this step. Because the threshold is not robust for extracting objects, these detected results are used only as the first step to extract the object regions preliminarily, which can also reduce the computation complexity. The adaptive optimal algorithm presented in the next section is used to obtain more accurate detection results.

4 Markov Random Field

The direct threshold method is evidently not robust for a practical system. Furthermore, this method ignores the spatial context, which rarely exists among proximal sites. To solve this problem, we consider the Markov random field method to determine the pixel labels. Based on the mathematical foundation, we cast the classification problem into a Markov random field for estimating the labels of pixels.

The image label set is defined as $L = \{l_1, l_2, \ldots, l_n\}$, where l_n represents the pattern class containing the background or foreground. Subsequently, the l_n value of a pixel has two values: $l_n = 0$ (background), or $l_n = 1$ (foreground). D is a random field denoted by $D = \{d_x : x \in I, d_x \in L\}$. $d_x = l_n$ means that the label l_n is assigned to the pixel x. Let $\eta_{ij} \subset I$ represent the neighbourhood of a pixel in the location (i, j), which satisfies the condition that no pixel belongs to neighbourhood of this pixel, and the neighbourhood systems are independent. Let C be the set of cliques c associated with η_{ij}.

The maximum a posteriori (MAP) [33] can be used to define the real label of pixel $\mathbf{x}$:

$$\arg\max_{\mathbf{x}\in I} P(D|\mathbf{x}) \tag{4}$$

where $P(D|\mathbf{x})$ can be expressed using the Bayesian rule:

$$P(D|\mathbf{x}) = \frac{P(\mathbf{x}|D)P(D)}{P(\mathbf{x})} \tag{5}$$

In Eq. (5), when the image is confirmed, $P(\mathbf{x})$ is ignored which is considered to be a constant term. The optimal labels for the pixels can be computed as

$$\arg\max_{\mathbf{x}\in I} \left((P(\mathbf{x}|D)P(D)\right) \tag{6}$$

4.1 Conditional Probability

As mentioned above, the probability of the pixels belonging to foreground objects can also be described similar to in the background model via the adaptive kernel density estimation; that is, the foreground object at time t can be detected using the sample set $\Phi_f = \{f_1, f_2, \ldots, f_n\}$ of the foreground.

Using the method to estimate the foreground probability in Eq. (2), the foreground probability in the proposed system can be expressed as

$$\hat{p}(\mathbf{x}|\Phi_f) = \frac{1}{n(2\pi)^{d/2}} \sum_{i=1}^{n} \frac{1}{\|\Sigma_{\mathbf{x}_i} + \Sigma_{\mathbf{x}}\|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}_i)^T (\Sigma_{\mathbf{x}_i} + \Sigma_{\mathbf{x}})^{-1} (\mathbf{x} - \mathbf{x}_i)\right) \tag{7}$$

where $P(\mathbf{x}|D)$ is the conditional density presented based on the conditional independence assumption of $\mathbf{x}$:

$$P(\mathbf{x}|D) = \prod_{x\in I} P(\mathbf{x}|d_x) = \prod_{x\in I} \hat{p}(\mathbf{x}|\Phi_f)^{l_n} \hat{p}(\mathbf{x}|\Phi_b)^{1-l_n} \tag{8}$$

4.2 Prior Probability

The Markov random field can describe the relationships among the contexts in the image sequences. Evidently, the contexts are particularly important cues for detecting objects in the videos of complex scenes. Therefore, we propose that the method of the Markov random field is applied for object detection. The prior probability can be completely described by a Gibbs distribution:

$$P(D=d)=\frac{1}{Z}e^{-U(d)/T} \tag{9}$$

where Z is a normalized constant called the partition function, T is the temperature, and $U(d)$ is the energy function:

$$U(d)=\sum_{c\in C}V_c(d) \tag{10}$$

where V_c is a function of the cliques around the site under consideration. Some traditional forms of the MRF model exist, and the Ising model presented in [33] considers the spatial context owing to its discontinuity preserving properties:

$$P(D)\propto\exp(\sum_{i\in I}\sum_{j\in C}\lambda(l_il_j+(1-l_i)(1-l_j))) \tag{11}$$

where λ is set between 0 to 1 as a constant.

Consequently, by substituting Eqs. (8) and (11) into Eq. (6), the maximum a posterior can be defined as

$$\begin{aligned}&\arg\max(\ln(P(\mathbf{x}|D)P(D))\\&=\arg\max(\sum_{i\in I}(l_i\times\ln\frac{\hat{p}(\mathbf{x}|\Phi_f)}{\hat{p}(\mathbf{x}|\Phi_b)}+\ln\hat{p}(\mathbf{x}|\Phi_b))+\sum_{i\in I}\sum_{j\in C}\lambda(l_il_j+(1-l_i)(1-l_j)))\end{aligned} \tag{12}$$

where $\ln\hat{p}(\mathbf{x}|\Phi_b)$ is a constant term, and it can be ignored.

Furthermore, we provide the explanation for Eqs. (11)–(12) and the implementation method for Eq. (12). Based on the analysis, the detailed implementation of Eq. (12) is as follows:

Step 1: Initializing the labels for pixels: Using Eq. (3), the objects are detected the first time to initialize the labels.

Step 2: In the current frame t, the energy value of pixel x is computed based on Eq. (12):

$$l_i\times\ln\frac{\hat{p}(\mathbf{x}|\Phi_f)}{\hat{p}(\mathbf{x}|\Phi_b)}+\sum_{j\in c}\lambda(l_il_j+(1-l_i)(1-l_j) \tag{13}$$

where l_j is obtained from the frame t-1. $l_i=0$ (assuming that pixel x is in the background) or $l_i=1$ (assuming that pixel x is in the foreground).

Step 3: Maximizing Eq. (12): The optimal algorithm described in Section 4 is used to maximize Eq. (12).

5 Solution Optimization by Using the Improved Ant Colony Algorithm

The traditional algorithms for optimizing such a model exhibit an inferior quality; these algorithms contain the iterated condition modes (ICM) and simulated annealing (SA). The SA algorithm involves long processes for reducing the temperature, and determining the solutions is thus computationally expensive. The solutions via the ICM are very sensitive to the initialization values. Consequently, our paper creatively adopts the ant colony optimization algorithm to solve the MAP-MRF problem in the multi-moving object detection. Recently, the ant colony algorithm has been used for many image segmentation problems [34]; however, it has not been applied for the optimal detection of objects.

In this algorithm, a colony of artificial ants search for a globally optimum solution, that is, the image pixels' labels are the "food sources" and maximizing the energy function is the target optimal function,

Algorithm 1: Ant colony optimization algorithm based on the MAP-MRF framework

Step 1: Initialize the foreground set Φ_f
(the initial foreground set is set to a null set) and background set Φ_b.
Step 2: Starting from the current frame (t-frame)
1) Calculate the background probability $\hat{p}(\mathbf{x}|\Phi_b)$.
2) Determine the threshold of and initialize the labels of the pixels.
Step 3: Calculate the foreground probability $\hat{p}(\mathbf{x}|\Phi_f)$.
Step 4: Detect the objects via the Markov random field in consecutive frames.
Step 5: Calculate the pixel initial energy value, and initialize the pixels' pheromone values.
Step 6: Optimize the posterior energy function using the improved ant colony algorithm.
Step 7: Update the background and foreground models.
Step 8: If t < N (N is the total number of frames in the video sequence), t = t + 1; if t = N, the algorithm is terminated.

such as "the shortest path" represented in a TSP problem. Next, the ants, that is, the pixels, trace a solution that assigns the labels 0 (background) or 1 (foreground) to the image pixels. To reduce the computational complexity, we remove the pixels recognized as the background in the first detection step.

In the TSP problem, an ant constructs the solution according to the "path" that has the strongest pheromone trail. Therefore, the decisional basis associates the label pixel and the labels. The pheromone trail is computed as

$$\tau_{x,l_n} = \frac{1}{N \times U_x} \tag{14}$$

where τ_{x,l_n} denotes the choosing label l_n for pixel x, and U_x is the local energy function defined as

$$U_x = l_n \times \ln\frac{\hat{p}(\mathbf{x}|\Phi_f)}{\hat{p}(\mathbf{x}|\Phi_b)} + \sum_{j\in c} \lambda(l_n l_j + (1-l_n)(1-l_j)) \tag{15}$$

where l_n denotes the label for the current pixel, which is set to be 0 or 1.

To reduce the probability of obtaining a local optimizing solution, we define the transforming probability by the following pseudo-random-proportional rule, which dynamically adjusts the assignment of the pixel labels in the search process:

$$l = \begin{cases} \arg\max\limits_{l\in L} \tau_{s,l} \;\; if \;\; q \le q_0 \\ P_{s,l} \;\; otherwise \end{cases} \tag{16}$$

where q_0 is a fixed value defined as $q_0 \in (0,1)$, and q is a random number lying between (0, 1). When $q \le q_0$, the ants assign the label l to pixel x according to the highest pheromone trail $\tau(x, l_n)$; otherwise, the choice is made with the transforming probability $P_{s,l}$ represented as

$$P_{x,l_n} = \frac{\tau_{x,l_n}}{\sum\limits_{l_n\in L} \tau_{x,l_n}} \tag{17}$$

The transforming probability is a function using only the pheromone trail because the pheromone trail not only contains the pheromone concentration via the local energy, which is the basis of constructing the

solution, but it also reflects the heuristic choice corresponding to the foreground-background ratio, which is contained in the first term of local energy Eq. (15).

The steps of the abovementioned ant colony optimization algorithm can be presented as follows:

Step 1: Initialize the pixels' pheromone values. The ants visit the proper foreground pixels in parallel, and each visiting process is treated with STEP iterations.

Step 2: The solution is computed by an ant, using Eq. (16).

Step 3: Local pheromone update occurs

$$\tau(x, l_n) = (1 - \rho) \times \tau(x, l_n) + \rho \times \tau_0 \tag{18}$$

where τ_0 is the initialized value.

Step 4: Global pheromone update occurs

$$\tau(x, l_n) = (1 - \alpha) \times \tau(x, l_n) + \alpha \times \Delta\tau(x, l_n) \tag{19}$$

where

$$\Delta\tau(x, l_n) = \begin{cases} \dfrac{1}{U(x^g)} & if\ (x, l_n) \in x^g \\ 0 & otherwise \end{cases} \tag{20}$$

The proposed multi-moving object detection approach using the ant colony optimization algorithm is based on the MAP-MRF framework, as presented in **Algorithm 1**. Fig. 2 shows the block diagram of the

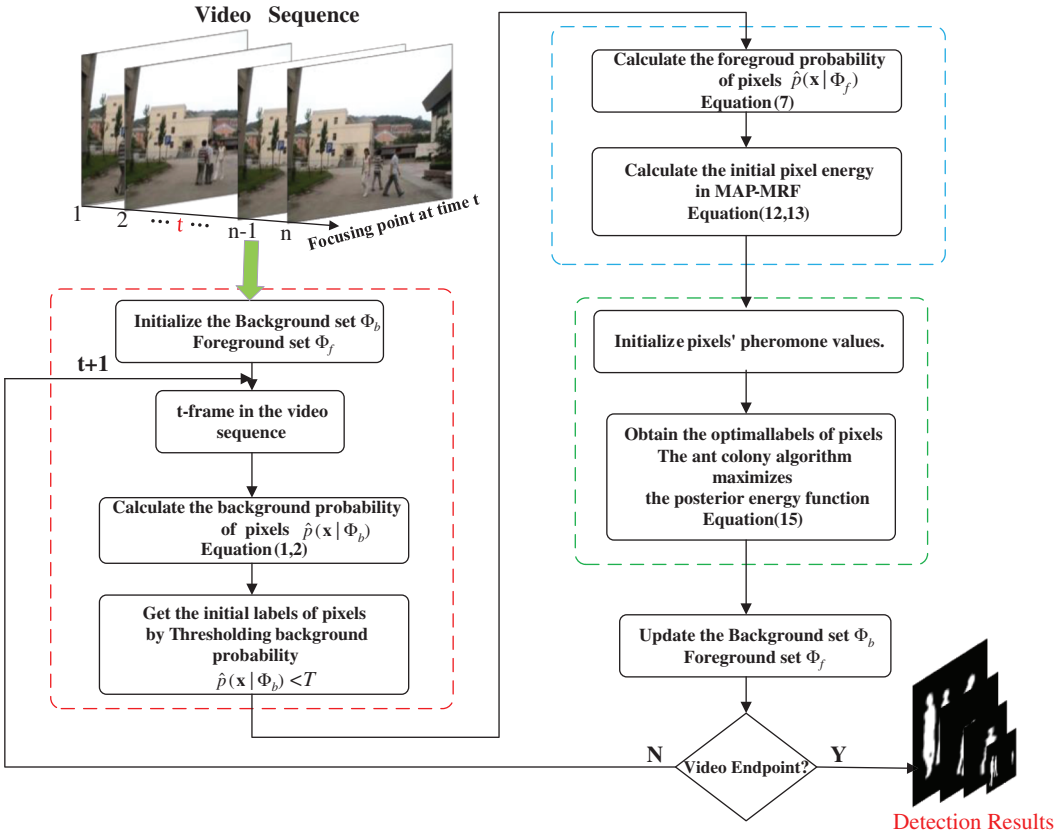

Figure 2: Flowchart of the proposed moving object detection method

proposed moving object detection approach. The pixels classified as the foreground are used to update the foreground model Φ_f, and all the pixels are used to update Φ_b to allow the consideration of the variations in the scenes.

6 Experimental Results and Analysis

In this section, to evaluate the performance of the proposed multi-moving object detection algorithm, we present a set of experiments performed on the published and available CDW-2012[1] datasets and the video sequences devoted to the background/foreground segmentation employed in this study. The system configuration for the experiments is as follows: 2.80 GHz Intel(R) Core(TM) i7-8400U processor with 16 GB RAM, and the programming language is MATLAB 2018a. For a fair comparison, the values of T (Eq. (3)) and σ of the Gaussian distribution are chosen as 1.0×10^{-5} and 2, respectively, and they are fixed during our experiments.

This section is divided into three parts: 1) In the first part, we provide the qualitative analysis results pertaining to the CDW-2012 datasets with those of other existing background subtraction methods. 2) In the second part, we present the quantitative evaluation on the CDW-2012 datasets with different quantitative rules; 3) In the third part, to verify the validity of our algorithm in the real world, the algorithm is applied to the considered video sequences.

The performance of the proposed multi-moving target detection algorithm can be analysed from both qualitative and quantitative aspects. The qualitative analysis is mainly determined by the human visual perception but is subjectively influenced by the individuals. Different people draw different conclusions from different angles. Therefore, this approach is not conducive to the fairness of the algorithm evaluation, and it can only be used as an evaluation reference for the algorithm performance. The quantitative analysis is a superior evaluation of the algorithm performance through certain rules or quantitative criteria. The five criteria [35,36], that is, the precision, recall, F-measure, true positive rate (TPR), and false positive rate (FPR) are used to evaluate the detection results and ground truth:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F-measure} = \frac{2}{1/\text{Precision} + 1/\text{Recall}}$$

$$\text{TPR} = \frac{TP}{TP + FN}, \quad \text{FPR} = \frac{FP}{FP + TN}$$

where TP (true positives), TN (true negatives), FP (false positives), and FN (false negatives) denote the number of positive samples correctly detected as positive samples, number of negative samples correctly detected as negative samples, number of negative samples incorrectly detected as positive samples, and number of positive samples incorrectly detected as negative samples, respectively.

To validate the proposed method, different algorithms are used, including 1) a Gaussian mixture containing five-components [14] (5-MoG); 2) non-parametric kernel density estimator with a fixed bandwidth matrix [8] (FB-KDE); 3) probabilistic superpixel Markov random fields [37] (PSP-MRF); 4) low memory and non-parametric based background subtraction algorithm [10] (LMBS); 5) generalized fused lasso [28] (GFL) and 6) background-foreground interaction [4] (BFI). The experimental results for these methods and our method are presented in Figs. 3–10 and Tabs. 1–4. To verify the robustness and accuracy of our algorithm, we choose several complex dynamic scenes in the datasets, which makes the selected scene more similar to the real-world scenes. The performance of the approach is evaluated by

[1] http://www.changedetection.net/

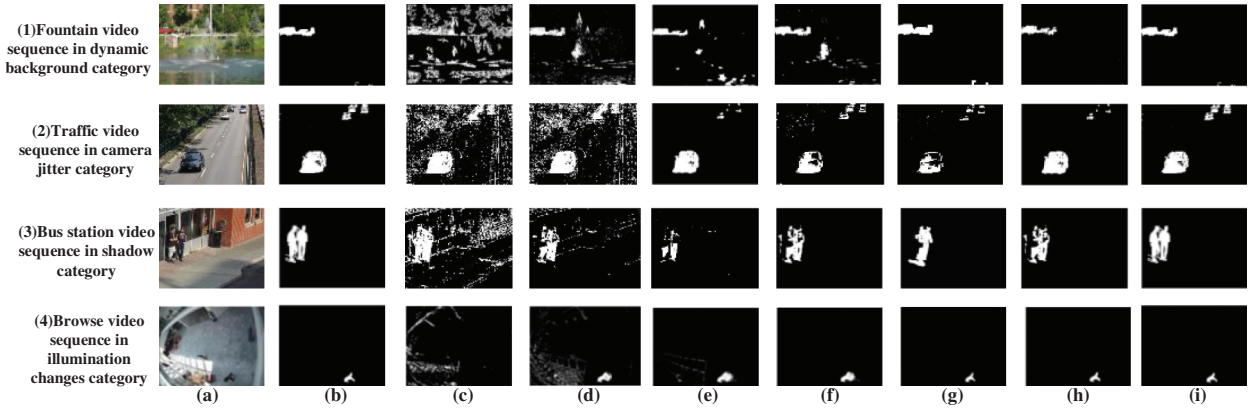

Figure 3: Moving object detection results for different video sequences: (a) original frame, (b) ground truth, (c) 5-MoG, (d) FB-KDE, (e) PSP-MRF, (f) LMBS, (g) GFL, (h) BFI, (i) proposed approach

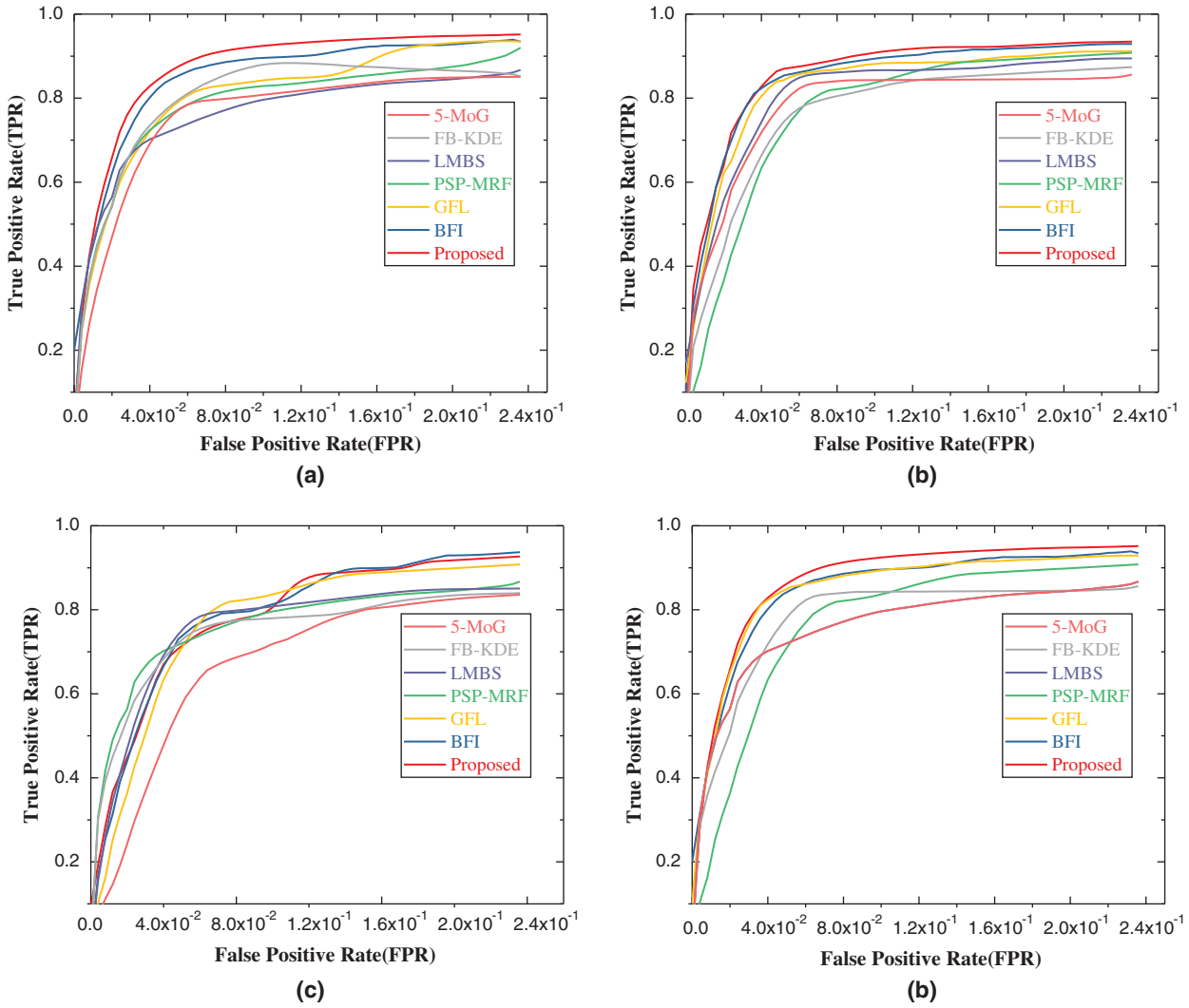

Figure 4: ROC curves under different video categories: (a) "shadow" category, (b) "camera jitter" category, (c) "dynamic background" category, (d)"illumination changes" category

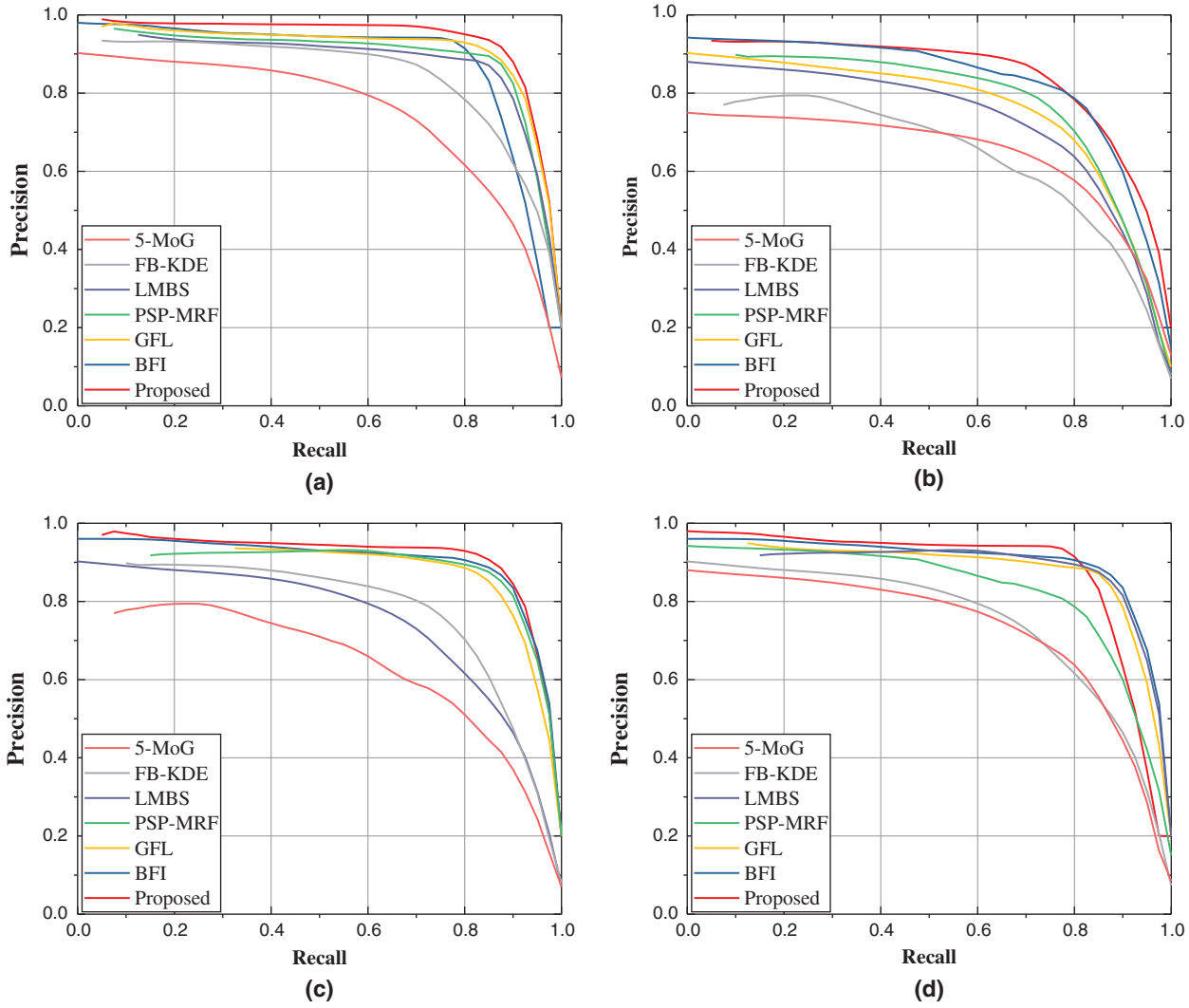

Figure 5: P-R curves for the different video categories: (a) "shadow" category, (b) "camera jitter" category, (c) "dynamic background" category, (d)"illumination changes" category

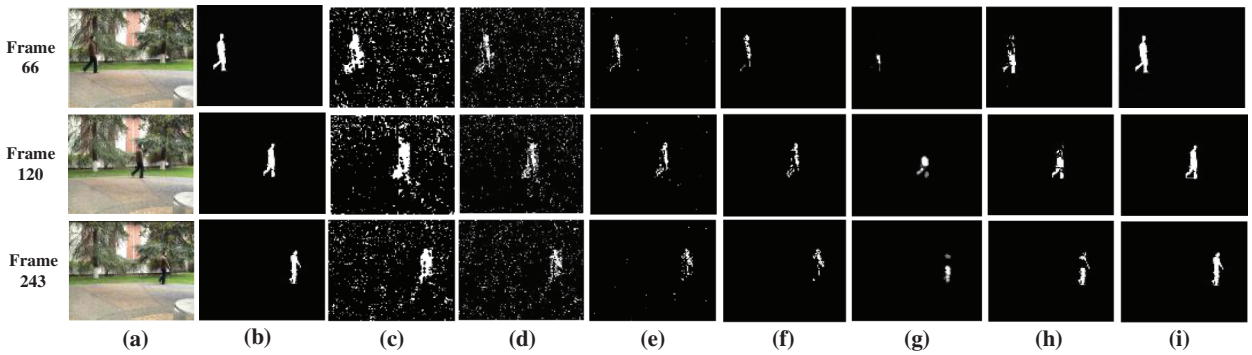

Figure 6: Moving object detection results for the video sequence Waving Tree: (a) original frame, (b) ground truth, (c) 5-MoG, (d) FB-KDE, (e) PSP-MRF, (f) LMBS, (g) GFL, (h) BFI, (i) our approach

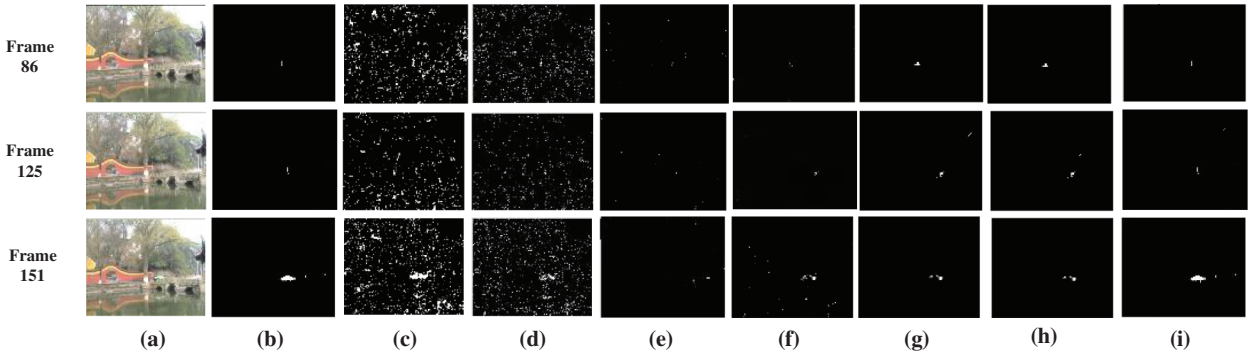

Figure 7: Moving object detection results for the video sequence Water Surface: (a) original frame, (b) ground truth, (c) 5-MoG, (d) FB-KDE, (e) PSP-MRF, (f) LMBS, (g) GFL, (h) BFI, (i) our approach

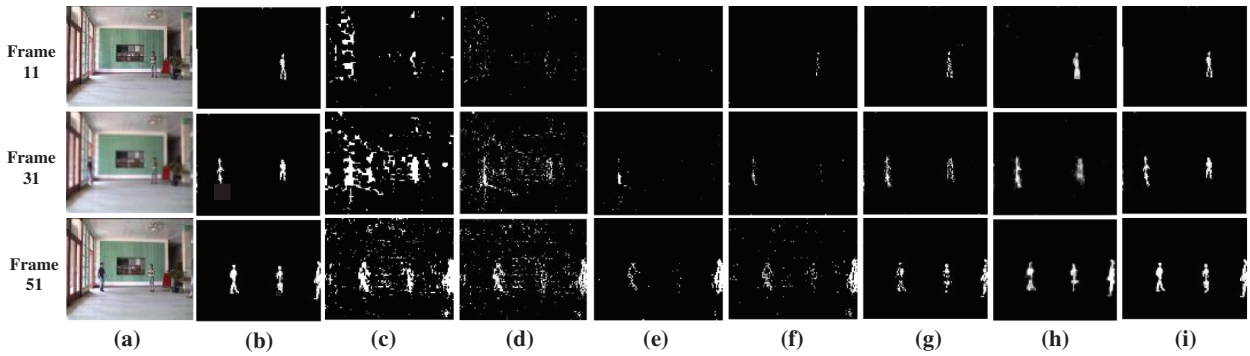

Figure 8: Moving object detection results for the video sequence Indoor Shadow: (a) original frame, (b) ground truth, (c) 5-MoG, (d) FB-KDE, (e) PSP-MRF, (f) LMBS, (g) GFL, (h) BFI, (i) our approach

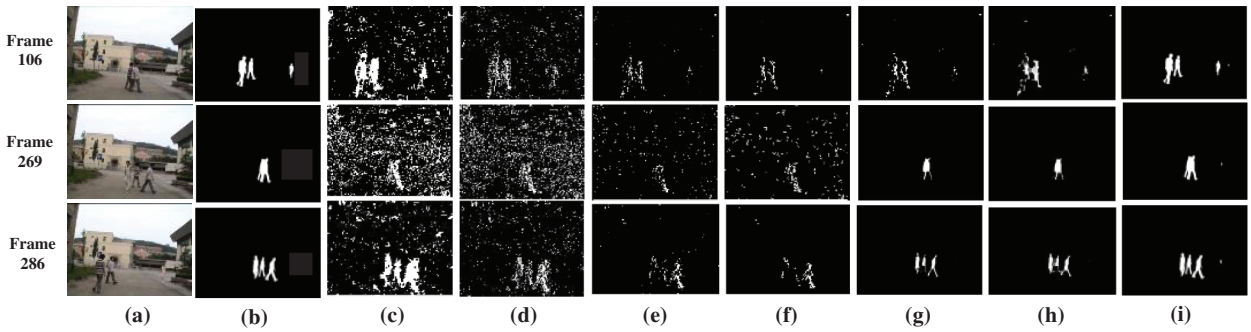

Figure 9: Moving object detection results for the video sequence Outdoor Illumination Changes: (a) original frame, (b) ground truth, (c) 5-MoG, (d) FB-KDE, (e) PSP-MRF, (f) LMBS, (g) GFL, (h) BFI, (i) our approach

some typical complex dynamic scenes of the shadow category, camera jitter category, dynamic background category and illumination changes category in the video sequences. Each category contains four or six video shots, which make the algorithm evaluation datasets richer and closer to the real world.

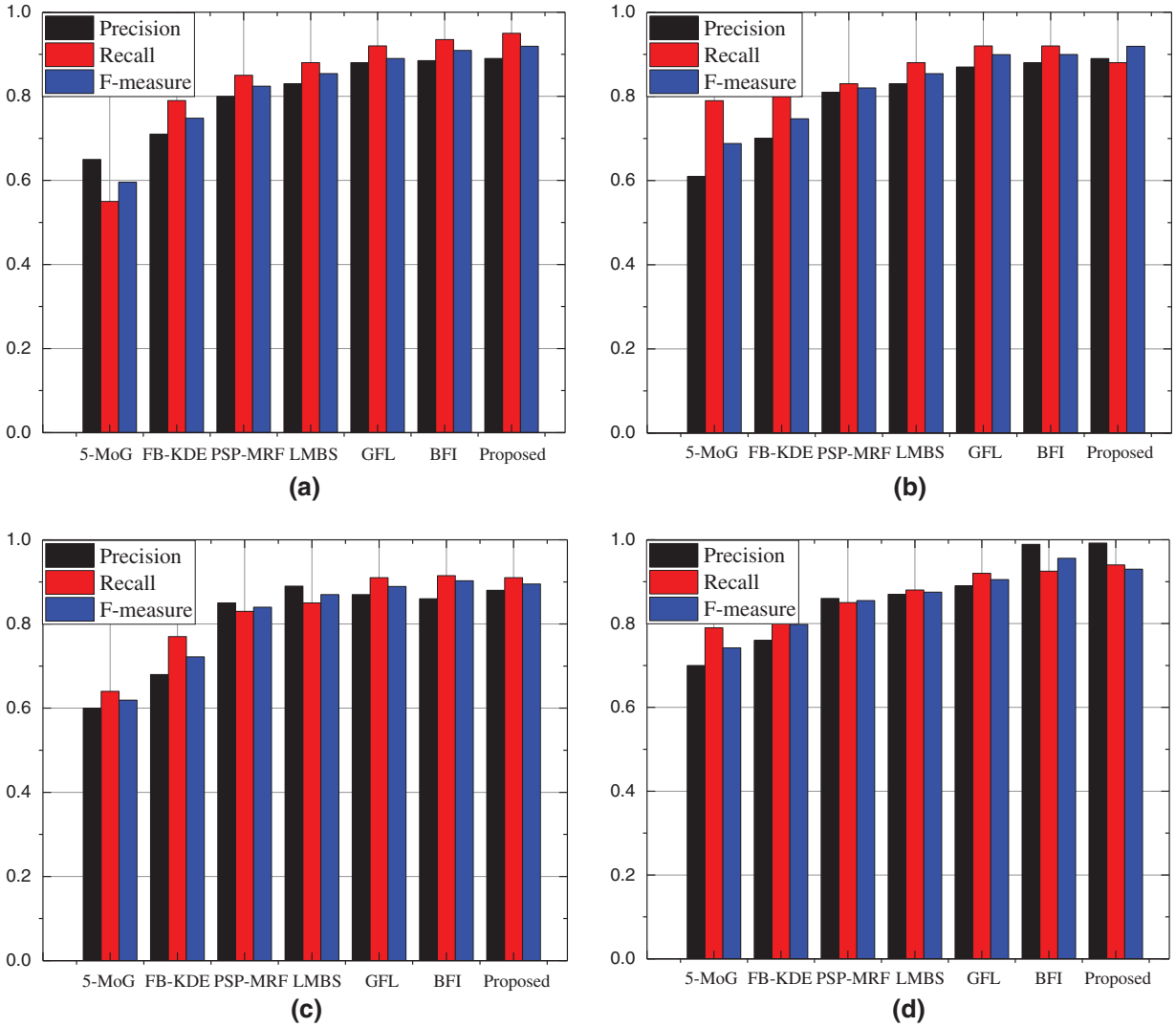

Figure 10: Histogram of the average precision, recall and F-measure values for our datasets: (a) Waving tree, (b) Water surface, (c) Indoor shadow, (d) Outdoor illumination changes

6.1 Qualitative Analysis

In recent years, multi-moving object detection algorithms have developed rapidly and performed well on various types of videos. The published and available change detection.net[2] datasets constitute a benchmark database and are popularly used for moving object detection/segmentation in dynamic scenes including sudden illumination changes, environmental conditions, background/camera motion, shadows, and camouflage effects. This dataset contains 11 video categories with 4 to 6 videos sequences in each category. Fig. 3 shows the moving object detection results of the four complex scenes in the published available datasets. The first row shows a fountain video sequence in the dynamic background category. This scene is challenging for the background model because the motion caused by the fountain in the background is very dramatic. Only the proposed method effectively tackles the challenging problems, and the 5-MoG, FB-KDE, PSP-MRF, and LMBS methods are susceptible to the dynamic background, which

[2]http://www.changedetection.net/

Table 1: Average precision, recall, F-measure and FPS values for the "Shadow" category

Method	*Precision*	*Recall*	*F-measure*	*FPS*
5-MoG [14]	0.65	0.55	0.596	15
FB-KDE [8]	0.71	0.79	0.748	86
LMBS [37]	0.80	0.85	0.824	54
PSP-MRF [10]	0.83	0.88	0.855	30
GFL [28]	0.88	0.92	0.890	18
BFI [4]	0.90	0.89	0.895	21
Proposed	**0.89**	**0.95**	**0.919**	**25**

Table 2: Average precision, recall, F-measure and FPS values values for the "Camera Jitter" category

Method	*Precision*	*Recall*	*F-measure*	*FPS*
5-MoG [14]	0.61	0.79	0.688	12
FB-KDE [8]	0.70	0.80	0.747	80
LMBS [37]	0.81	0.83	0.820	50
PSP-MRF [10]	0.83	0.88	0.854	26
GFL [28]	0.87	0.92	0.894	15
BFI [4]	0.88	0.90	0.890	19
Proposed	**0.89**	**0.88**	**0.888**	**18**

Table 3: Average precision, recall, F-measure and FPS values values for the "Dynamic Background" category

Method	*Precision*	*Recall*	*F-measure*	*FPS*
5-MoG [14]	0.60	0.64	0.619	19
FB-KDE [8]	0.68	0.77	0.722	90
LMBS [37]	0.85	0.83	0.840	66
PSP-MRF [10]	0.89	0.85	0.870	36
GFL [28]	0.87	0.91	0.889	26
BFI [4]	0.87	0.90	0.885	23
Proposed	**0.88**	**0.91**	**0.895**	**30**

causes a false detection of the detection results. The second row shows a traffic video sequence in the camera jitter category. In this scene, the camera shakes slightly and the branches shake because of the strong wind. All the methods demonstrate a satisfactory performance in terms of the detection of the moving objects with a large resolution; however, when detecting the targets with a small resolution, the contours detected by our method are clearer. The third row shows a bus station video sequence in the shadow category. This sequence involves a large shadow area. The proposed method in this case can completely detect the moving target without losing the target information. The four row shows a browse video sequence in the illumination changes category. In this scene, background changes due to outdoor sunlight and indoor lighting. The GFL, BFI and the proposed method demonstrate a satisfactory performance without losing

Table 4: Average precision, recall, F-measure and FPS values values for the "Illumination Changes" category

Method	*Precision*	*Recall*	*F-measure*	*FPS*
5-MoG [14]	0.70	0.74	0.719	23
FB-KDE [8]	0.78	0.77	0.775	107
LMBS [37]	0.86	0.85	0.855	63
PSP-MRF [10]	0.89	0.88	0.885	43
GFL [28]	0.91	0.90	0.905	22
BFI [4]	0.92	0.90	0.910	23
Proposed	**0.93**	**0.91**	**0.919**	**28**

the target information. Fig. 3 shows that the proposed algorithm exhibits a superior performance compared to that of the other six existing methods.

6.2 Quantitative Evaluation

In addition to the quantitative analysis, we also quantitatively evaluated our method on the available datasets. The human visual perception is the best evaluator of any vision system, but it lacks in terms of the quantitative performance assessment. Hence, the aim of this part was to compare the existing methods in terms of the five criteria in different complex scenarios. The average precision, recall, F-measure and FPS values for the "shadow", "camera jitter", "dynamic background" and "illumination changes" categories are listed in Tabs. 1–4 respectively. The precision and recall values in some cases exhibit contradictory trends, and the F-measure is the weighted harmonic average of the precision and recall. When A higher F-measure corresponds to a more effective test method. From the scores in Tabs. 1 and 3, the average precision, recall and F-measure of our methods are superior to those of the other six methods. The precision and recall scores for the proposed approach exhibit an increase of 0.01 and 0.03, respectively, compared to those of the GFL method in Tab. 1, and the proposed method demonstrates the most superior performance in terms of the F-measure (0.919). In the "camera jitter" category, due to the high frequency of the scene changes, the performance of our method is slightly lower than those of the GFL [28] and BFI [4]. Compared with the precision and recall scores of the 5-MoG and FB-KDE methods listed in Tab. 2, our method exhibited significantly increased values by 0.28 and 0.09, respectively. In the "dynamic background" category, compared with the non-parametric methods (FB-KDE [8] and LMBS [37]), our method exhibited increased precision scores by 0.20 and 0.03, respectively, and the recall scores were increased by 0.14 and 0.08, respectively. We also compared our method with the probabilistic model (PSP-MRF [10]) and fast parametric-flow algorithms (GFL [28] and BFI [4]). The proposed method outperformed the three competing methods by 0.025, 0.006 and 0.010 in terms of the F-measure, respectively. From the scores in Tab. 4, the average precision, recall and F-measure of our methods are superior to those of the other six methods. Compared with the GFL [28], BFI [4] methods, our method exhibited increased precision scores by 0.01 and 0.02, respectively, and the recall scores were increased by 0.01 and 0.01, respectively.

To further examine the quantitative performance of our method, the complexity of our method, reflected by the FPS score is demonstrated in Tabs. 1–4. The FPS of our method is lower than those of the FB-KDE [8], LMBS [37], and PSP-MRF [10] methods, but our method is better than 5-MoG [14], GFL [28] and BFI [4] because our method achieves more accurate results.

For a detailed analysis, we also evaluated the effectiveness of the proposed algorithm by two typical performance evaluation methods: the ROC curve and P-R curve [38–41]. Fig. 4 displays the ROC curves

for the different video categories, corresponding to the proposed methods and other existing methods. Compared with the other competing methods, our detection algorithm demonstrated the best mean TPR scores of 0.926. The ROC curve visually shows the relationship between the FPR and TPR. The closer the curve is to the upper left, and the larger the area is under the ROC curve, the better is the performance of the proposed algorithm. Fig. 5 shows the P-R curve for the different video categories, corresponding to the proposed methods and other existing methods. Compared with the GFL methods, our methods improved the mean precision scores by 0.013, 0.011 and 0.008, and the mean recall scores by 0.020, 0.019 and 0.011. The P-R curve visually shows the relationship between the precision and recall. The closer the curve is to the upper right, the better is the robustness of the proposed algorithm. These measures indicate that the proposed method works well, exhibited superior detection performance compared to that of the other six existing methods.

6.3 Results for Our Datasets

To demonstrate the real-life application of the proposed algorithm and the consistency of the detection effect on long as well as short video sequences, the performance evaluation of the approach on our datasets is presented in this section. In terms of the time constraint, we show the detection results on four video sequences. The selected video sequences are: (1) waving tree, (2) water surface, (3) indoor shadow, and (4) outdoor illumination changes. In Figs. 6–9, the number of different frames in the scene video sequences is shown on the leftmost axis. The first column corresponds to the original image in the sequences, the second column shows the ground truth in the sequences, the third column corresponds to the results detected using the 5-MoG method, the fourth column illustrates the results derived using FB-KDE method, the fifth column shows the detection results obtained using the PSP-MRF method, the sixth column shows the results by the LMBS method, the seventh row presents the segmentation results by the GFL method, the eighth column shows the detection results obtained using the BFI method, and the last column shows the detection results obtained using our method.

The first sequence shows a site with swaying trees, rain and varying illumination, which are challenging aspects for object detection. The waving tree sequence is a long video sequence of 5 min, containing 5863 frames. Fig. 6 shows the results obtained using by the proposed algorithm compared with that of the other existing methods. Evidently, since the distribution of the natural dynamic texture is complex and variable, parametric methods such as the mixture of Gaussian approachs and non-parametric methods with the fixed bandwidth matrix handled the dynamic texture of the scene in an inferior manner. However, considering the effects of the movement of the tree leaves, rain and shadows, the proposed algorithm overcame such challenges and detected the moving person accurately.

The second sequence involved two people walking, who were extremely far from the camera across, and a car moved in the frame rapidly. The water surface sequence is a short video sequence of 25 s, containing 503 frames. Two challenges for detection existed: 1) The natural dynamic texture containing water waves, swaying trees, and shadows of the objects, and 2) the low resolution of the two people. As shown in Fig. 7, the proposed algorithm overcame the two difficulties and accurately detected the car moving across the frame and the person walking far from the camera.

The algorithm was tested on the third sequence in the presence of a global illumination and indoor shadow. The indoor shadow sequence is a short video sequence of 20 s, containing 396 frames. Three people moved in the room. The results indicate that the 5-MoG and FB-KDE methods can solve the challenge of the varying illumination to a certain extent but cannot recognize the objects in a shadow effectively. Our method deals with the varying illumination and shadow adaptively.

An outdoor surveillance scene was considered in the fourth sequence, involving clouds, bushes and global illumination changes. The outdoor illumination change sequence is a long video sequence of 4 min, containing 3396 frames. Fig. 9 also shows that our method outperforms the other methods.

The proposed scheme was successfully tested on both the long and short video sequences, as shown in Figs. 6–9. In the waving tree video sequences, the effects of the 5-MoG and FB-KDE methods was considerably affected by the swaying leaves, and the LMBS and GFL methods could not detect the moving object clearly. The proposed method overcome all the shortcomings of the above mentioned six methods and could accurately detect the moving object. Furthermore, the resolution of the moving target was incredibly low. The GFL and BFI methods overcome the influence of the dynamic scene, and the detection accuracy of the moving target is also considerably improved; however, compared to that obtained using our method, part of the information of the detecting object was missing, as shown in Fig. 7. The LMBS method could detect the target object, but the information of the target object was critically missing. The proposed method can overcome the shortcomings of the other methods and accurately track the target object, as shown in Fig. 8. Our method could segment occlude the pedestrians and was less susceptible to illumination, as shown in Fig. 9. The shortcoming of this method is that it cannot accurately identify the number of pedestrians and the time complexity is high. At the same time, the key objective of future work is to improve the speed of the proposed algorithm.

To understand the global results, we analysed the performance of these algorithms for four video sequences in terms of the average precision, recall and F-measure values. It was observed that the F-measure obtained by the proposed method increased to 0.92 on our datasets, compared to the values of 0.60, 0.75, 0.82, 0.85, 0.89 and 0.90 pertaining to the 5-MoG method, FE-KDE method, LMBS method, PSP-MRF method, GFL method and BFI method, as shown in Fig. 10(a). In the water surface dynamic scene (Fig. 10(b)), although the recall rate of the proposed detection method was slightly lower than that of the GFL method, the F-measure of the proposed method was 0.92, and that of the GFL method was only 0.89. The average precision, recall and F-measure values of the proposed method were the most stable and relatively high, as shown in Figs. 10(c) and 10(d). The detection accuracy of the proposed algorithm was 0.98, which is 0.33 higher than that of the 5-MoD method, as shown in Fig. 10(d). In general, from the result analysis discussed above, the proposed method exhibits a better performance in the dynamic and challenging scenes.

7 Conclusion

An effective algorithm named the improved non-parametric method in the Markov random field is proposed in this paper to help realize object detection in complex surveillance scenes. The proposed method has several significant contributions that are in contrast from those of existing background subtraction methods.

For modelling the background, an adaptive kernel density estimation method based on the variable bandwidth is presented to realize adaptive detection. A Markov random field framework is improved based on the model, and the optimal solution is acquired by using an improved ant colony algorithm. We present the detailed steps to implement the ant colony algorithm in the Markov random field. A series of experiments were conducted considering complex scenes, and the results indicated that the proposed method can adapt to the dynamic changes and improve the detection results. The superiority of the proposed method was established in terms of three performance evaluation measures, namely, the precision, recall and F-measure.

In the future work, for reducing the computation complexity, more accurate varying kernel density estimators can be defined for the probability estimation, and other optimization methods can be estimated for maximizing the a posterior solution. Furthermore, it may be interesting to further research the object detection in dynamic scenes by using deep learning methods and RGB-D data.

Acknowledgement: The author would like to thank the anonymous reviewers for their constructive comments.

Funding Statement: This work was supported in part by the National Natural Science Foundation of China under Grants 61841103, 61673164, and 61602397; in part by the Natural Science Foundation of Hunan Provincial under Grants 2016JJ2041 and 2019JJ50106; in part by the Key Project of Education Department of Hunan Provincial under Grant 18B385; and in part by the Graduate Research Innovation Projects of Hunan Province under Grants CX2018B805 and CX2018B813.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Bloisi, D. D., Pennisi, A., Iocchi, L. (2017). Parallel multi-modal background modeling. *Pattern Recognition Letters, 96,* 45–54. DOI 10.1016/j.patrec.2016.10.016.
2. Berjón, D., Cuevas, C., Morán, F., García, N. (2018). Real-time nonparametric background subtraction with tracking-based foreground update. *Pattern Recognition, 74,* 156–170. DOI 10.1016/j.patcog.2017.09.009.
3. Moyà-Alcover, G., Elgammal, A., Jaume-i-Capó, A., Varona, J. (2017). Modeling depth for nonparametric foreground segmentation using RGBD devices. *Pattern Recognition Letters, 96,* 76–85. DOI 10.1016/j.patrec.2016.09.004.
4. Chen, Z., Wang, R., Zhang, Z., Wang, H., Xu, L. (2019). Background-foreground interaction for moving object detection in dynamic scenes. *Information Sciences, 483,* 65–81. DOI 10.1016/j.ins.2018.12.047.
5. Subudhi, B. N., Ghosh, S., Shiu, S. C., Ghosh, A. (2016). Statistical feature bag based background subtraction for local change detection. *Information Sciences, 366,* 31–47. DOI 10.1016/j.ins.2016.04.049.
6. Wren, C. R., Azarbayejani, A., Darrell, T., Pentland, A. P. (1997). Pfinder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7),* 780–785. DOI 10.1109/34.598236.
7. Stauffer, C., Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. *Proceedings. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), Fort Collins,* vol. 2, pp. 246–252. *CO, USA*, IEEE.
8. Elgammal, A., Harwood, D., Davis, L. (2000, June). Non-parametric model for background subtraction. *European Conference on Computer Vision,* pp. 751–767. Springer, Berlin, Heidelberg.
9. Zhu, T., Zeng, P. (2015). Background subtraction based on non-parametric model. *4th International Conference on Computer Science and Network Technology, Harbin, China,* vol. 1, pp. 1379–1382. IEEE.
10. Jain, R., Nagel, H. H. (1979). On the analysis of accumulative difference pictures from image sequences of real world scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1(2),* 206–214. DOI 10.1109/TPAMI.1979.4766907.
11. Hassan, M. A., Malik, A. S., Nicolas, W., Faye, I., Mahmood, M. T. (2014). Mixture of gaussian based background modelling for crowd tracking using multiple cameras. *5th International Conference on Intelligent and Advanced Systems, Kuala Lumpur, Malaysia,* pp. 1–4. IEEE.
12. Bouwmans, T., El Baf, F., Vachon, B. (2008). Background modeling using mixture of gaussians for foreground detection-a survey. *Recent Patents on Computer Science, 1(3),* 219–237. DOI 10.2174/2213275910801030219.
13. Rittscher, J., Kato, J., Joga, S., Blake, A. (2000). A probabilistic background model for tracking. *European Conference on Computer Vision,* pp. 336–350. Springer, Berlin, Heidelberg.
14. Stauffer, C., Grimson, W. E. L. (2000). Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8),* 747–757. DOI 10.1109/34.868677.
15. Saito, M., Kitaguchi, K. (2010). Probabilistic appearance based object modeing and its application to car recognition. *Proceedings of SICE Annual Conference, Taipei, Taiwan,* pp. 2360–2363. IEEE.

16. Zhou, X., Yang, C., Yu, W. (2012). Moving object detection by detecting contiguous outliers in the low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(3),* 597–610. DOI 10.1109/TPAMI.2012.132.
17. Huynh-The, T., Banos, O., Lee, S., Kang, B. H., Kim, E. S. et al. (2016). NIC: a robust background extraction algorithm for foreground detection in dynamic scenes. *IEEE Transactions on Circuits and Systems for Video Technology, 27(7),* 1478–1490. DOI 10.1109/TCSVT.2016.2543118.
18. Li, C., Wang, X., Zhang, L., Tang, J., Wu, H. et al. (2016). Weighted low-rank decomposition for robust grayscale-thermal foreground detection. *IEEE Transactions on Circuits and Systems for Video Technology, 27(4),* 725–738.
19. Varadarajan, S., Miller, P., Zhou, H. (2015). Region-based mixture of gaussians modelling for foreground detection in dynamic scenes. *Pattern Recognition, 48(11),* 3488–3503. DOI 10.1016/j.patcog.2015.04.016.
20. Kim, K., Chalidabhongse, T. H., Harwood, D., Davis, L. (2005). Real-time foreground-background segmentation using codebook model. *Real-Time Imaging, 11(3),* 172–185. DOI 10.1016/j.rti.2004.12.004.
21. St-Charles, P. L., Bilodeau, G. A., Bergevin, R. (2014). Subsense: a universal change detection method with local adaptive sensitivity. *IEEE Transactions on Image Processing, 24(1),* 359–373. DOI 10.1109/TIP.2014.2378053.
22. Barnich, O., Van Droogenbroeck, M. (2010). ViBe: a universal background subtraction algorithm for video sequences. *IEEE Transactions on Image Processing, 20(6),* 1709–1724. DOI 10.1109/TIP.2010.2101613.
23. Hofmann, M., Tiefenbacher, P., Rigoll, G. (2012). Background segmentation with feedback: the pixel-based adaptive segmenter. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence,* pp. 38–43, *RI, USA*. IEEE.
24. Zhang, G., Yuan, Z., Tong, Q., Zheng, M., Zhao, J. (2018). A novel framework for background subtraction and foreground detection. *Pattern Recognition, 84,* 28–38. DOI 10.1016/j.patcog.2018.07.006.
25. Lee, S., Kim, N., Jeong, K., Park, K., Paik, J. (2015). Moving object detection using unstable camera for consumer surveillance systems. *Optik—International Journal for Light and Electron Optics, 126(20),* 2436–2441. DOI 10.1016/j.ijleo.2015.06.003.
26. Cao, F., Liu, Y., Wang, D. (2018). Efficient saliency detection using convolutional neural networks with feature selection. *Information Sciences, 456,* 34–49. DOI 10.1016/j.ins.2018.05.006.
27. Shelhamer, E., Long, J., Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(*4),* 640–651.
28. Xin, B., Tian, Y., Wang, Y., Gao, W. (2015). Background subtraction via generalized fused lasso foreground modeling. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 4676–4684. *Boston, MA, USA*.
29. Trabelsi, R., Jabri, I., Smach, F., Bouallegue, A. (2017). Efficient and fast multi-modal foreground-background segmentation using RGBD data. *Pattern Recognition Letters, 97,* 13–20. DOI 10.1016/j.patrec.2017.06.022.
30. Islam, M. M., Hu, G., Liu, Q., Dan, W., Lyu, C. (2018). Correlation filter based moving object tracking with scale adaptation and online re-detection. *IEEE Access, 6,* 75244–75258. DOI 10.1109/ACCESS.2018.2883650.
31. Parzen, E. (1962). On estimation of a probability density and mode. *Annals of Mathematical Statistics, 33(3),* 1065–1076. DOI 10.1214/aoms/1177704472.
32. Mittal, A., Paragios, N. (2004). Motion-based background subtraction using adaptive kernel density estimation. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR,* vol. 2, pp. II–II. *Washington, DC, USA,* IEEE.
33. Sheikh, Y., Shah, M. (2005). Bayesian modeling of dynamic scenes for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(11),* 1778–1792. DOI 10.1109/TPAMI.2005.213.
34. Kheirinejad, S., Hasheminejad, S. M. H., Riahi, N. (2018). Max-min ant colony optimization method for edge detection exploiting a new heuristic information function. *8th International Conference on Computer and Knowledge Engineering,* pp. 12–15. *Mashhad, Iran,* IEEE.
35. Smeulders, A. W., Chu, D. M., Cucchiara, R., Calderara, S., Dehghan, A. et al. (2013). Visual tracking: an experimental survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(7),* 1442–1468.

36. Quast, K., Kaup, A. (2011). AUTO GMM-SAMT: an automatic object tracking system for video surveillance in traffic scenarios. *EURASIP Journal on Image and Video Processing, 2011,* 1–14. DOI 10.1155/2011/814285.
37. Schick, A., Bäuml, M., Stiefelhagen, R. (2012). Improving foreground segmentations with probabilistic superpixel markov random fields. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence,* pp. 27–31. *RI, USA,* IEEE.
38. Roy, K., Arefin, M. R., Makhmudkhujaev, F., Chae, O., Kim, J. (2018). Background subtraction using dominant directional pattern. *IEEE Access, 6,* 39917–39926. DOI 10.1109/ACCESS.2018.2846749.
39. Keivani, A., Tapamo, J. R., Ghayoor, F. (2017). Motion-based moving object detection and tracking using automatic K-means. *IEEE AFRICON,* pp. 32–37. IEEE.
40. Kalantar, B., Mansor, S. B., Halin, A. A., Shafri, H. Z. M., Zand, M. (2017). Multiple moving object detection from UAV videos using trajectories of matched regional adjacency graphs. *IEEE Transactions on Geoscience and Remote Sensing, 55(9),* 5198–5213. DOI 10.1109/TGRS.2017.2703621.
41. Eltantawy, A., Shehata, M. S. (2019). An accelerated sequential PCP-based method for ground-moving objects detection from aerial videos. *IEEE Transactions on Image Processing, 28(12),* 5991–6006. DOI 10.1109/TIP.2019.2923376.