IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 24, 2019, accepted August 30, 2019, date of publication September 5, 2019, date of current version September 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939672

Dynamic ARMA-Based Background Subtraction for Moving Objects Detection

JIAN LI, ZHONG-MING PAN, ZHUO-HANG ZHANG, AND HENG ZHANG
College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410003, China
Corresponding author: Jian Li (lijiannudt@163.com)

This work was supported by the National Natural Science Foundation of China under Grant 91320202.

ABSTRACT Background subtraction is a prevailing method for moving object detection in videos with stationary backgrounds. However, accurate and real-time moving object detection is challenging in the presence of complex dynamic scenes. This paper presents a novel technique for background subtraction based on the dynamic autoregressive moving average (ARMA) model. Specifically, we utilize the temporal and spatial correlation of images in a video sequence to model each pixel to accurately model the background image's dynamic characteristics. In addition, we apply an adaptive least mean square (LMS) scheme to update the parameters of the background model to offset the dramatically dynamic characteristic of the background. The proposed algorithm is evaluated on two publicly available benchmark datasets with complex dynamic backgrounds. The experimental results show that this technique is robust and effective for background subtraction in complex dynamic backgrounds and is a promising moving object detection scheme for real-time visual surveillance.

INDEX TERMS Background subtraction, image segmentation, ARMA model, moving object detection, adaptive LMS, real-time visual surveillance.

I. INTRODUCTION

The extensive development of intelligent visual surveillance systems demands that systems possess powerful and real-time image processing ability to identify, locate and track moving objects. Moving object detection is a prerequisite to successfully implement the functions of such systems. Background subtraction, an essential technique in moving object detection, has been widely studied in various situations, such as action recognition [1], target tracking [2] and traffic monitoring [3]. Generally, a background subtraction scheme consists of the following four aspects: (1) background modeling, which constructs a model to represent the background; (2) background initializing, which initializes all the parameters of the model; (3) background updating, which updates the model to adapt to the dynamic scenes; and (4) foreground detecting, which involves the reasonable classification of all pixels. In recent years, a large variety of techniques developed for background subtraction [4], [5] have achieved promising detection results.

The associate editor coordinating the review of this manuscript and approving it for publication was Haimiao Hu.

However, several remaining technical difficulties make it difficult to detect moving objects effectively and robustly: (1) dynamic backgrounds, e.g., waving trees, illumination changes, swaying curtains, rippling water, spouting fountains; (2) camera displacements, e.g., camera jitter and tripod motion; (3) camouflage, where foreground targets have similar color or texture with the background; (4) extremely bad weather, e.g., heavy snow, rainstorms and dense fog; and (5) heavy computational cost to obtain an ideal result, thus greatly influencing the real-time performance. Moreover, most recently proposed algorithms based on deep learning increasingly rely on expensive hardware resources due to the demanding training process [6], [7]. Such methods are not practical for visual surveillance with limited computing resources and strong real-time demands. Inspired by pixel-based and region-based algorithms, we propose a simple but robust moving object detection scheme based on a dynamic ARMA model to overcome these difficulties.

In this paper, a scheme called DARMABS, which incorporates the idea of pixel-based and region-based approaches by utilizing the temporal and spatial correlation of the pixels, is developed to model the background image's dynamic characteristics. The basis of this scheme is that we consider

a particular pixel value of the video sequence over time as a time series, and the ARMA model possesses powerful modeling capability for time series, which greatly reduces the computational burden to obtain a robust background model that can describe the background image's dynamic characteristics. The parameters of the proposed ARMA scheme are constantly updated to adapt to the dynamic background, which distinguishes the model from the conventional ARMA modeling process. ARMA has been extensively used to model background noise, such as ocean clutter in radar [8], [9]. Meanwhile, several proposed methods resembling ARMA adopt temporal information of the pixel values to model the background [10]–[12]. However, the adoption of ARMA in background modeling by integrating temporal and spatial information has not been reported in the literature.

The main contributions of this paper are as follows: (1) The dynamic ARMA process is initially utilized to model the background. (2) The temporal and spatial information of the pixels is combined. (3) The model is robust to the complex dynamic elements in the background. (4) The real-time performance of moving object detection is enhanced on the premise of great detection results.

The rest of this paper is organized as follows. A review of related work on background subtraction is presented in Section II. Section III details the proposed DARMABS scheme, including building the pixel model, initializing the background model, updating the background model and classifying the current pixel point. Section IV discusses the experimental results and performance analysis. Finally, conclusions are presented in Section V.

II. RELATED WORK

Generally, background subtraction methods can be based on pixels [13]–[20], feature [21]–[23], regions [24], [25], frames [26]–[28], superpixels [29]–[31], or deep learning [32]–[34].

Pixel-based methods model each pixel to obtain a robust background. In early work, a single Gaussian model [13] or a codebook [14] was the most popular way to segment moving objects. Then, the classic Gaussian mixture model (GMM) algorithm was proposed in [15]. The algorithm utilizes a mixture of Gaussian distributions to model the background; thus, the GMM can adapt to dynamic backgrounds. As a further development, various extended versions of the GMM have been proposed in [16]–[18], and more adaptive methods have been proposed to improve the segmentation performance. In [19], Droogenbroeck and Barnich introduced a universal background subtraction algorithm called ViBe for complex dynamic backgrounds that takes the pixel values from previous frames of neighboring pixels or the same position as a pixel sample set. Then, ViBe compares the current pixel value with the set to classify the pixel and updates the set by randomly choosing the pixel to be substituted. An improved derivative of ViBe called PBAS was proposed in [20] as a nonparametric background subtraction paradigm that extends two parameters, namely, the decision threshold and learning parameters, to dynamic per-pixel state variables to estimate the background dynamics. These methods are the most popular due to their simple, efficient and high-speed implementations, but they are too sensitive to distinguish background and foreground pixels as the effects of irregular background changes.

Feature-level background modeling uses local textures around a pixel to neutralize the complex variations in the background. A local binary pattern (LBP)-based method for constructing the background model was proposed in [21]. The method adopts a texture-based feature, namely, the LBP, which can accurately model the background. Since the method extracts region-based textures, the moving background pixels can be labeled as foreground. Zhang *et al.* proposed a feature of spatiotemporal LBP (STLBP) [22], a derivative of LBP, which performs well in natural dynamic movement and is fast to compute online. The SuBSENSE algorithm incorporates local binary similarity patterns descriptors to model the background [23] and is robust to noise and variation in the background. These feature-based methods produce good segmentation results compared to those of pixel-based methods but are not sufficiently stable for complex frequent variations in the background. Region-based background modeling uses spatial correlation to alleviate the complex variations in the background via region-level foreground shape models [24] or background models [25]. Frame-based background modeling methods, such as robust principal component analysis [26]–[28], are an alternative to pixel-level and region-level modeling. However, these methods are not practical for real-time surveillance because they rely on offline or batch processing. Moreover, frame-based methods consume substantial memory and entail higher computational cost than traditional methods.

Superpixel-level background modeling methods, such as those proposed by Chen *et al.* [29], Fang *et al.* [30] and Giordano *et al.* [31], model the background in terms of superpixels, which results in lower memory consumption but is still computationally expensive. Recent methods based on convolutional neural networks have also been proposed for moving object detection [32]–[34]. These approaches perform well in complex dynamic scenes but require a large quantity of labeled data for training. By contrast, our proposed DARMABS background subtraction scheme requires small quantities of labeled data for training and is sufficiently robust to achieve good segmentation results.

III. THE PROPOSED DARMABS SCHEME

The procedures of the proposed scheme are as follows:

(1) Employing ARMA to build the pixel model.
(2) Initializing the background model.
(3) Adopting adaptive LMS to update the background.
(4) Using a distance measure to classify the current pixel.
(5) Updating the set of background pixel samples.

An overview of the DARMABS scheme based on a block diagram is presented in Fig. 1. We detail how the scheme works

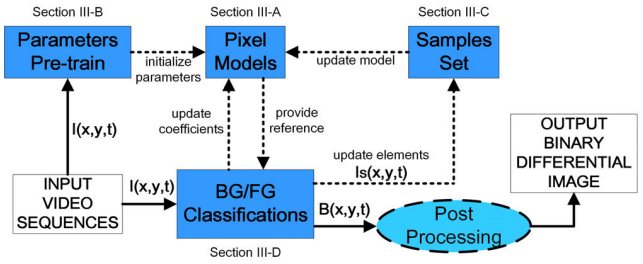


FIGURE 1. Block diagram of DARMABS. The detail of each block is described in Section III-A through Section III-D. In our scheme, the step of parameter pretraining is designed to determine the optimum initialization parameters and the postprocessing step is intended to produce better results.

in four aspects: In Section III-A, we show how to model individual pixels by using history information based on ARMA. In Section III-B, we present how the model's parameters can be initialized in an effective way. We introduce the method for updating the model's parameters via an adaptive LMS scheme in Section III-C. Finally, a strategy is proposed to perform the accurate BG/FG classification of pixels in combination with the neighborhood pixel information in Section III-D.

A. BUILDING THE PIXEL MODEL

To properly construct the background model, we analyze the modeling process and make the following definitions:

Definition 1 (Time Series of Images): We regard the pixel values of images over time as a "pixel process" [36]. Let the time series of images be $\{I(t)\}_{t=1,2,\ldots}$; then, in the t^{th} frame at position (x, y), the actual pixel value is denoted as $I(x, y, t)$, and the corresponding predicted background pixel value is represented as $I_{pred}(x, y, t)$.

Definition 2 (Set of Background Pixel Samples): At position (x, y), let the m element set of background pixel samples (*SBPS*) be $\theta = \{I_{sam}(x, y, i)|1 \leq i \leq m\}$. Note that the elements of the set are chosen from known background pixels spatially and temporally, which are constantly updating.

Definition 3 (Binary Differential Image): Let the differential image obtained from the t^{th} frame be $B(t)$. The assembly of all pixel points is $B(x, y, t)$, which denotes the segmentation result of the proposed scheme.

The central idea of our proposed scheme is to design a prediction mechanism that can obtain the actual pixel value of the background based on the historic values. Thus, the pixel model is given by

$$I_{pred}(x, y, t) = \sum_{i=1}^{p} a_i I_{samp}(x, y, i) - \sum_{j=1}^{q} b_j e_{t-j} + e_t \quad (1)$$

$$e_{t-j} = \left| I_{pred}(x, y, t-j) - I(x, y, t-j) \right| \quad (2)$$

where p and q are the orders of the process ($p \geq q$), $a_i (i = 1, 2, \ldots, p)$, $b_j (j = 1, 2, \ldots, q)$ indicate coefficients of the autoregressive and moving average parts respectively. $e_t \sim N(0, \sigma^2)$ is the deviation between the predicted pixel value and the actual value, which satisfies $E[I_{pred}(x, y, t) \cdot e_{t-j}] = 0 (j = 1, 2, \ldots, q)$.

TABLE 1. The algorithm performance comparison with different p and q values in WT the video sequence.

(p, q)	PCC	(p, q)	PCC	(p, q)	PCC
(1, 1)	64.21	(3, 1)	79.21	(5, 1)	82.26
(1, 3)	65.08	(3, 3)	82.05	(5, 3)	87.39
(1, 5)	64.33	(3, 5)	78.89	(5, 5)	85.09
(2, 2)	73.35	(4, 2)	83.28	(6, 2)	89.82
(2, 4)	74.24	(4, 4)	82.04	(6, 4)	91.13
(2, 6)	75.56	(4, 6)	84.56	(6, 6)	90.24

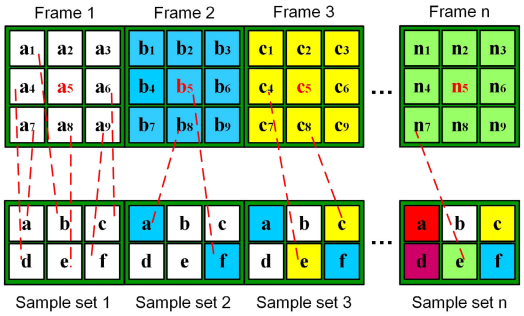


FIGURE 2. Initialization and updating scheme for each pixel's set of background pixel samples. This figure shows the updating method of the sample set over time. The area of the red number denotes the central pixel. The elements of the set are randomly chosen from the neighborhood points of the central pixel.

B. INITIALIZING THE BACKGROUND MODEL

This subsection elaborates the initialization of the background model, which includes two main aspects, namely, element selection for *SBPS* and value assignment of the model parameters.

Fig. 2 shows the initialization scheme for *SBPS*. Due to the spatial correlation between adjacent pixels, the elements of the first few frames' *SBPS* are randomly selected from the neighboring pixels to initialize the individual sets.

The model parameters that need to be initialized are the orders p and q, the model coefficients a_i and b_j, the learning rate of the parameters $\alpha$and the threshold of classification T, τ (α, T and τ will be described in Section III-C through Section III-D). We restrict p and q to change from 0 to 6 since an ARMA model of sixth order is a good trade-off between mathematical tractability and model fidelity. Table 1 compares the algorithm performance with different p and q values for the "waving trees" video sequence. The initial values of a_i and b_j are selected based on the parameter pretraining test designed to obtain reasonable values, which selects several frames to roughly estimate the parameters. The detailed parameter setup will be elaborated in Section IV. The remaining initial parameters for α and T, τ are given as follows: α is approximately 0.01 and T, τ depends on the actual true positive (TP) and false positive (FP) values in the practical experiment. Detailed parameters will be presented in the next section.

C. UPDATING THE BACKGROUND MODEL

The model's update components include $SBPS = \{I_{sam}(x, y, 1), \ldots, I_{sam}(x, y, m)\}$ and the model parameters (e.g., a_i and b_j). The updating strategy of $SBPS$ is to select the pixel $I(x, y, t_1)$ judged as a background pixel more than a certain number of times, e.g., N, and to replace that element in the samples set, which guarantees the reliability and adaptability of the samples set. The rule is defined as follows:

$$SBPS = \begin{cases} \{I_{sam}(x, y, 1), \ldots, I(x, y, t_1), \ldots, I_{sam}(x, y, m)\} \\ \quad if\ count(I(x, y, t_1)) \geq N \\ \{I_{sam}(x, y, 1), \ldots, I_{sam}(x, y, m)\} \quad otherwise \end{cases} \tag{3}$$

where $count(\cdot)$ denotes the function that records the number of times pixel $I(x, y, t_1)$ is judged as a background pixel. Note that if $count(I(x, y, t_1)) \geq N$, the new pixel $I(x, y, t_1)$ will replace the element of $SBPS$ that has the most similar pixel value with the new pixel.

Simultaneously, to keep the background model reliable in complex dynamic scenes, we update the parameters in an online manner:

$$a_i^{(t+1)} = a_i^{(t)} + 2\alpha \cdot \varepsilon_t \cdot i(x, y, t - i) \quad (t \geq p) \tag{4}$$

$$b_j^{(t+1)} = b_j^{(t)} + 2\alpha \cdot \varepsilon_t \cdot \varepsilon_{t-j} \quad (t \geq q) \tag{5}$$

where $\alpha, i(x, y, t - i)$ and ε_t denote the learning rate, the normalized zero mean actual pixel value and the normalized deviation, respectively. Note that a larger α means that the deviation value and the pixel value strongly affect the model. ε_t and $i(x, y, t - i)$ are expressed as follows:

$$\varepsilon_{t-\mathrm{j}} = \frac{e_{t-j}}{s_t} \tag{6}$$

$$i(x, y, t - i) = \frac{I(x, y, t - i)}{s_t} \tag{7}$$

$$s_t = \sqrt{\sum_{j=0}^{q} e_{t-j}^2 + \sum_{i=1}^{p} \tilde{I}^2(x, y, t - i)} \quad (t \geq p) \tag{8}$$

$$\tilde{I}(x, y, t - i) = I(x, y, t - i) - \bar{I} \tag{9}$$

where $\bar{I}$ is the mean value of the actual pixel sequence, which is used for the de-mean procedure of normalizing the background pixel sequence. Note that the sequence contains both the elements of $SBPS$ and the pixels classified as background pixels.

D. CLASSIFYING THE CURRENT PIXEL POINT

We continue with the classification step of background subtraction based on the predicted pixel value in this subsection. To accurately classify a given pixel $I(x, y, t)$ into either foreground or background, we compare the predicted pixel value updated online with the current pixel value, as expressed in (10):

$$b(x, y, t) = \begin{cases} 1 & M(I_{pred}(x, y, t), I(x, y, t)) \geq T \\ 0 & otherwise \end{cases} \tag{10}$$

where $b(x, y, t)$ denotes the corresponding pixel point of the rudimentary segmentation result. $b(x, y, t) = 1$ denotes that the t^{th} frame's pixel at the position (x, y) belongs to the foreground, whereas $b(x, y, t) = 0$ denotes that the corresponding pixel is classified as background. T is a threshold for classification. The measure $M(\cdot, \cdot)$ used in (10), which adopts Manhattan distance, is defined as follows:

$$M(I_{pred}(x, y, t), I(x, y, t)) = \left|I_{pred}(x, y, t) - I(x, y, t)\right| \tag{11}$$

To improve the classification accuracy, we propose a scheme using the adjacent pixels based on the spatial correlation of pixels. Fig. 3 shows the pixel classification scheme. We need to consider the quantity of background pixels in the neighborhood of a pixel to ultimately determine the category of the pixel. We select an $n \times n$ neighboring pixel subblock centered on the current pixel to be classified; then, we count the number of pixels in the subblock classified as background to classify the current pixel. The rule is defined as follows:

$$B(x, y, t) = \begin{cases} 1 \quad if\ 1 - (\sum_{i=-n}^{n} \sum_{j=-n}^{n} b(x-n, y-n, t))/(2n+1)^2 \leq \tau \\ \quad (n = 1, 2, \ldots) \\ 0 \quad otherwise \end{cases} \tag{12}$$

where τ is the threshold for the pixel classification decision. Note that this rule cannot be applied to pixels on the edges of an image, which can be ignored. Algorithm 1 demonstrates the main procedure of the proposed scheme.

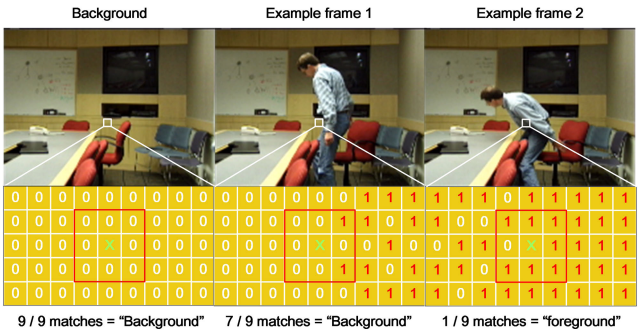


FIGURE 3. Pixel classification scheme. "1" denotes that the pixel is classified as foreground, whereas "0" denotes that the pixel is classified as background. "matches" means that the pixel matches the background model, namely, the pixel belongs to the background.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

This section introduces the datasets for moving object detection, the parameter setup and the evaluation metrics, also presents the experimental results and performance analysis.

Algorithm 1 Background Subtraction Using DARMABS

Input: Video sequences $I = \{I_1, I_2, \ldots, I_N\}$
Output: Binary differential sequences $i = \{i_1, i_2, \ldots, i_N\}$

```
for k ← 1 to number of frames do
  for i ← 1 to height of frame do
    for j ← 1 to width of frame do
      if (k == 1) then
        Initialize the background model for each pixel.
      else
        Calculate the predicted pixel value as (1).
        Calculate the deviation value as (2).
        Compute distance M of the above values as (11).
        if (M ≥ T && "matches" ≤ τ) then
          The pixel belongs to the foreground.
        else
          The pixel belongs to the background.
          Count the number of times n of the pixel value.
          if (n ≥ N) then
            Update the set of background pixel samples.
          end if
        end if
        Update the background model as (3)-(9).
      end if
    end for
  end for
end for
```

A. EXPERIMENTAL DATASETS AND SETUP

To assess the reliability and accuracy of the proposed DARMABS scheme, simulations have been conducted on the CDnet2014 dataset [35] and the Wallflower dataset [36]. Table 2 depicts a comprehensive overview of the two datasets.

1) CDNET2014 DATASET

The CDnet2014 dataset, a popular and publicly available benchmark dataset for change detection, comprises a wide range of video sequences with various detection challenges, which are categorized into eleven types: bad weather, baseline, camera jitter, intermittent object detection, low framerate, night videos, pan-tilt-zoom, pedestrian detection, shadow, thermal, and turbulence. Furthermore, the dataset provides the ground truth for each frames, which enables an accurate performance comparison of diverse algorithms for moving object detection.

2) WALLFLOWER DATASET

The Wallflower dataset, which is composed of 7 video sequences with a total of 16158 frames, is also a popular dataset that provides various challenging frames with complex dynamic backgrounds. The canonical dataset is grouped into seven categories: bootstrapping, camouflage, foreground aperture, light switch, moved object, time of day, and waving trees. One shortcoming of the dataset is that the ground truth is not available for all frames.

TABLE 2. An overview of the Wallflower and CDnet2014 datasets.

Dataset	Category	Videos	Total Frames
CDnet2014	bad weather	4	20900
	baseline	4	6049
	camera jitter	4	6420
	pedestrian detection	10	26248
	intermittent object detection	6	18650
	low framerate	4	9400
	night videos	6	16609
	pan-tilt-zoom	4	8630
	shadow	6	16949
	thermal	5	21100
	turbulence	4	15700
Wallflower	bootstrapping	1	3055
	camouflage	1	353
	foreground aperture	1	2113
	light switch	1	2715
	moved object	1	1745
	time of day	1	5890
	waving trees	1	287

To qualitatively validate the robustness of the proposed algorithm, we select the following representative videos from the above datasets that cover a large number of challenging surroundings, as shown in Table 3: (1) "*waving trees*" (WT) depicts a scenario with trees shaken by the wind; (2) "*time of day*" (TOD) shows a gradual illumination change; (3) "*badminton*" (BAD) is recorded in an outdoor environment with camera jitter; (4) "*camouflage*" (CAM) shows a scenario where the color of an object is similar to the background color; (5) "*skating*" (SKT) presents the challenge of poor winter weather conditions; (6) "*pets2006*" (PET) presents a scenario known for intermittent object motion; and (7) "*library*" (LIB) is a series of infrared thermal images. These sequences are taken in environments with complex dynamic backgrounds, which are the most suitable for testing the generalization ability and adaptability of the proposed algorithm.

We set the order of the autoregressive part $p = 6$, the order of the moving average part $q = 4$, the coefficient of the autoregressive part $a_i = 1$, the coefficient of the moving average part $b_j = 0.8$, the learning rate $\alpha = 0.01$, the classification threshold $T = 10$, the decision threshold $\tau = 4/9$, and the threshold for a pixel being selected as *SBPS* $N = 5$. The detailed parameter settings are shown in Table 4.

B. EVALUATION METRICS

To quantitatively evaluate the proposed algorithm, we select seven metrics to compare the performance of different background subtraction algorithms, namely, *Recall*, *Precision*, *F-Measure*, *False Positive Rate* (*FPR*), *False Negative Rate* (*FNR*), and *Percentage of Correct Classification* (*PCC*), which are defined as follows [35]:

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

TABLE 3. The selected video sequences for moving object detection from the Wallflower and CDnet2014 datasets.

Dataset	Category	Video	Total frames	Size	Characteristics
Wallflower	N/A	waving trees	287	160×120 pixels	dynamic background
Wallflower	N/A	time of day	5890	160×120 pixels	illumination change
CDnet2014	camera jitter	badminton	1150	720×480 pixels	camera displacements
Wallflower	N/A	camouflage	353	160×120 pixels	similar background color or texture
CDnet2014	bad weather	skating	3900	540×360 pixels	bad weather
CDnet2014	pedestrian detection	pets2006	1200	720×576 pixels	intermittent object motion
CDnet2014	thermal	library	4900	320×240 pixels	infrared thermal images

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

$$F - Measure = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{15}$$

$$Specificity = \frac{TN}{TN + FP} \tag{16}$$

$$FPR = \frac{FP}{TN + FP} \tag{17}$$

$$FNR = \frac{FN}{TP + FN} \tag{18}$$

$$PCC = 100 \times \frac{TP + TN}{TP + FN + FP + TN} \tag{19}$$

where *TP* denotes the number of correctly classified foreground pixels correctly classified. *TN* denotes true negatives, that is, the number of correctly classified background pixels. *FP* denotes false positives, that is, the number of background pixels mistakenly labeled as foreground pixels. *FN* denotes false negatives, that is, the number of foreground pixels mistakenly labeled as background pixels. *Recall* denotes the proportion all foreground pixels that are correctly detected. *Precision* denotes the percentage of foreground pixels correctly detected among all pixels marked as foreground. *F-Measure* is a combined metric used to evaluate the segmentation results. *Specificity* denotes the proportion of pixels correctly detected among all the background pixels. *FPR* denotes the percentage of pixels mistakenly detected among all the background pixels. *FNR* denotes the proportion of pixels mistakenly detected among all the foreground pixels. *PCC* denotes the percentage of pixels correctly classified among all pixels. Larger *Precision*, *Specificity*, *Recall*, *PCC*, and*F-Measure* and smaller *FNR* and *FPR* indicate better segmentation results.

C. COMPARISON WITH OTHER ALGORITHMS

To assess the effectiveness of the proposed DARMABS scheme, we compared our proposed algorithm with GMM [15], ViBe [19], LBP [21], LIBS [37], PBAS [20], SuBSENSE [23] and PAWCS [38] using the above evaluation metrics.

1) QUALITATIVE PERFORMANCE

We select the following frames to make a qualitative performance comparison of the segmentation results: the 53rd frame of *waving trees*, the 1850th frame of *time of day*, the 851st frame of *badminton*, the 248th frame of *camouflage*, the 1957th frame of *skating*, the 699th frame of *pets2006*, and the 1234th frame of *library*. Fig. 4 displays the segmentation results comparison of the abovementioned background subtraction algorithms. To ensure a fair comparison, we do not perform any postprocessing of the segmentation results. In Fig. 4, the input images and ground truths are presented in the 1st column and 2nd columns, respectively. The foreground segmentation results of various algorithms are shown from the 3rd column to the 10th column, namely, GMM, ViBe, LBP, LIBS, PBAS, SuBSENSE and PAWCS.

We can clearly infer that the DARMABS algorithm performs well in detecting moving objects with high robustness and accuracy for complex dynamic background video sequences. In the sequence *waving trees*, which contains a complex dynamic background, the output of DARMABS is closest to the ground truth. In the sequence *time of day* with gradual illumination variation, our proposed algorithm performs well compared with the other methods. Due to the camera jitter, the sequence *badminton* includes repetitive motion of background objects; however, our method can accurately detect the moving objects. Most of the methods incorrectly mark foreground pixels as background pixels. The sequence*camouflage* shows a man walking towards a working computer whose coveralls resemble the background in terms of color and texture; thus, the moving object is difficult to segment perfectly. FNs are avoided in our technique. The sequence *skating* provides a scene in which snow is falling. The challenge with this sequence is the noisy and irregular background. Although the result of our method includes

TABLE 4. Detailed parameter settings of the proposed algorithm.

Parameters	Specific description	Value
m	the number of elements in *SBPS*	6
n	the size of the subblock for classification	3
p	the order of the autoregressive part	6
q	the order of the moving average part	4
T	the classification threshold	10
τ	the decision threshold	4/9
α	the learning rate of the background model	0.01
N	the threshold of a pixel being selected as *SBPS*	5
a_i	the coefficient of the autoregressive part	1
b_j	the coefficient of the moving average part	0.8
$B(x, y, t)$	the binary differential image	{0,1}

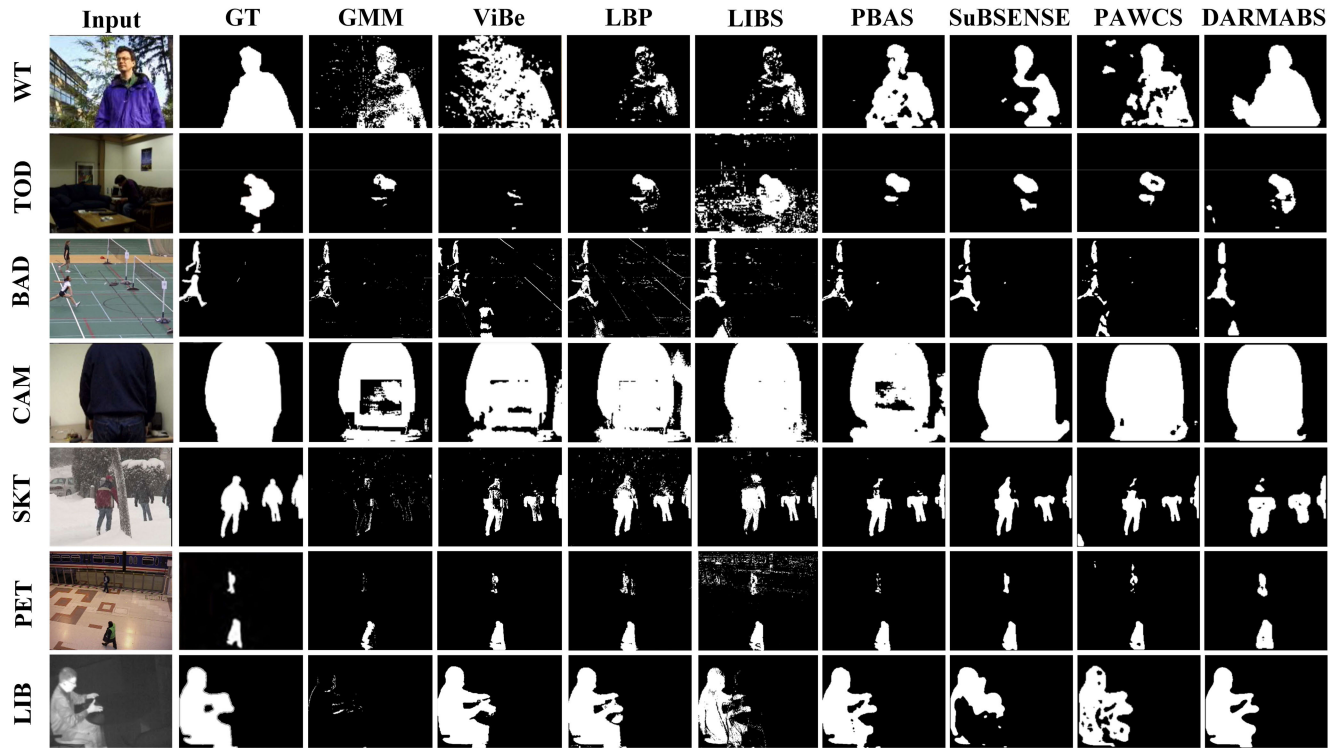


FIGURE 4. Foreground segmentation results of the video sequences: waving trees (Row 1), time of day (Row 2), badminton (Row 3), camouflage (Row 4), skating (Row 5), pets2006 (Row 6), and library (Row 7). Input images (Column 1), ground truths (Column 2), GMM [15] outputs (Column 3), ViBe [19] outputs (Column 4), LBP [21] outputs (Column 5), LIBS [37] outputs (Column 6), PBAS [20] outputs (Column 7), SuBSENSE [23] outputs (Column 8), PAWCS [38] outputs (Column 9), and DARMABS outputs (Column 10).

TABLE 5. Average results of Recall, Precision, F-Measure, Specificity, FPR, FNR and PCC on the CDnet2014 dataset.

Algorithm	Recall	Precision	F-Measure	Specificity	FPR	FNR	PCC
GMM [15]	0.5253	0.8376	0.6094	0.9122	0.0878	0.4747	70.74
ViBe [19]	0.8236	0.8382	0.8254	0.9101	0.0899	0.1764	86.15
LBP [21]	0.8724	0.8311	0.8461	0.8846	0.1155	0.1276	85.32
LIBS [37]	**0.8804**	0.7880	0.8278	0.8652	0.1349	**0.1196**	85.97
PBAS [20]	0.8717	**0.8948**	**0.8831**	**0.9475**	**0.0525**	0.1283	**90.27**
SuBSENSE [23]	0.8726	**0.9000**	**0.8843**	**0.9528**	**0.0472**	0.1274	88.27
PAWCS [38]	**0.8883**	0.8478	0.8664	0.9178	0.0822	**0.1117**	**89.71**
DARMABS	**0.9157**	**0.8742**	**0.8940**	**0.9316**	**0.0684**	**0.0843**	**92.08**

tiny gaps and is divided into several parts, our detection result is more reasonable than that of the other methods. In the sequence *pets2006*, there exist challenging intermittent object motions, but the segmentation results show that the proposed method is robust to the intractable scene. The sequence *library* is obtained from thermal videos and consists of infrared thermal images whose pixel values are distributed in a narrow range. The detection results indicate that our algorithm accurately segments the moving objects. These results further verify the effectiveness of the DARMABS scheme in complex dynamic scenes.

2) QUANTITATIVE PERFORMANCE

As a qualitative evaluation of the proposed method, we calculate the abovementioned evaluation metrics on the aforementioned datasets. Fig. 5 presents the detailed scores of the seven evaluation metrics on the selected video sequences. In the sequences *waving trees*, *time of day* and *camouflage*, DARMABS has the best F-Measure and PCC. DARMABS also performs well in the remaining sequences; therefore, our scheme is an effective means of robust background subtraction for moving object detection.

Table 5 and Table 6 list the average results of *Recall*, *Precision*, *F-Measure*, *Specificity*, *FPR*, *FNR* and *PCC* on the CDnet2014 dataset and Wallflower dataset, respectively. The top three of the segmentation results are emphasized in bold red, bold blue and bold green, respectively. On the CDnet2014 dataset, the proposed algorithm yields an average *Recall*, *Precision*, *F-Measure*, *Specificity*, *FPR*, *FNR* and *PCC* of 0.9157, 0.8742, 0.8940, 0.9316, 0.0684, 0.0843 and 92.08%, respectively. On the Wallflower dataset, DARMABS yields average results of 0.9229, 0.9445, 0.9333, 0.9297, 0.0703, 0.0771 and 91.12%. The *F-Measure* is the most important metric to assess the segmentation results of background subtraction algorithms. DARMABS has the best

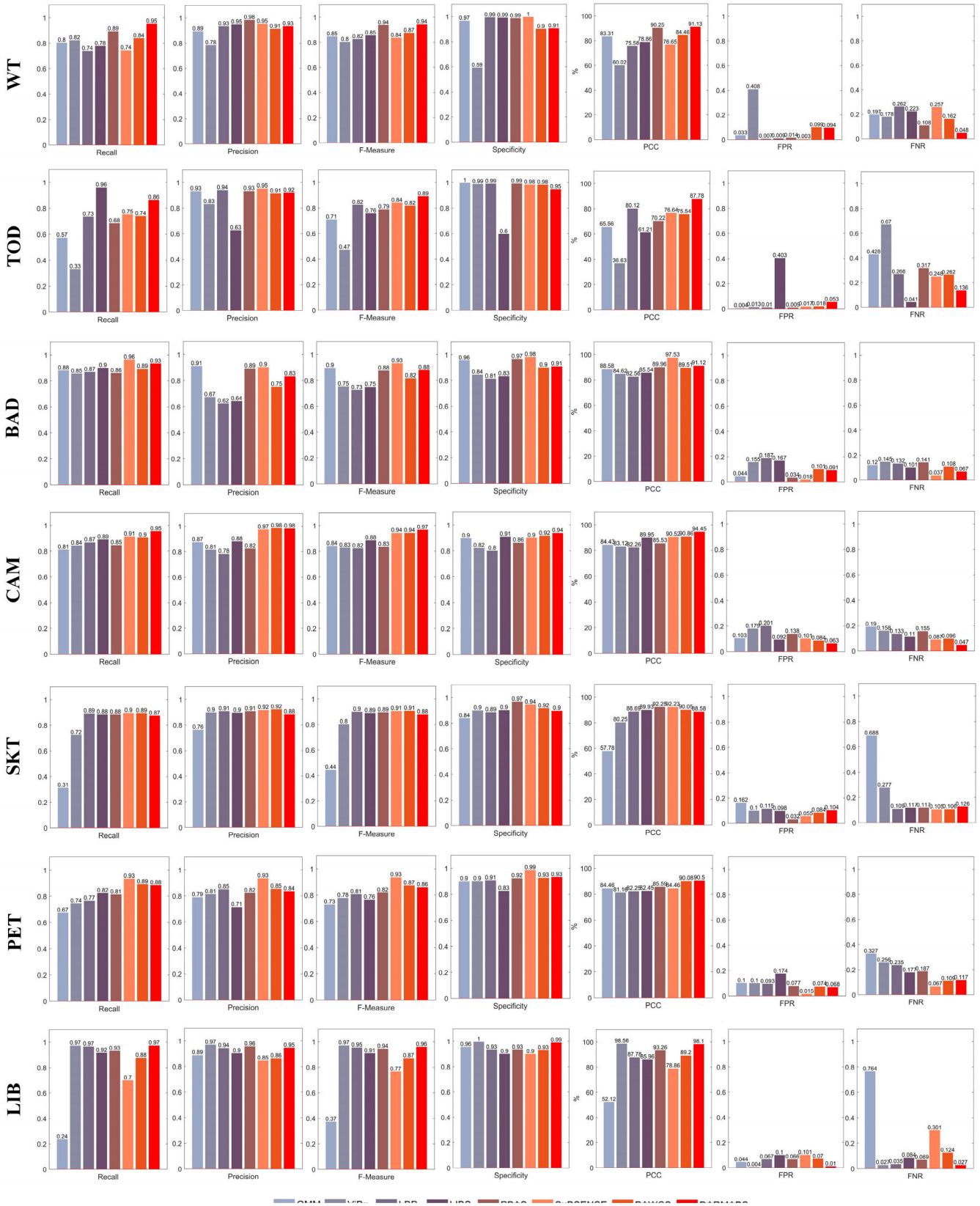


FIGURE 5. Scores of the evaluation metrics (Recall, Precision, F-Measure, Specificity, FPR, FNR and PCC) for the algorithms (GMM [15], ViBe [19], LBP [21], LIBS [37], PBAS [20], SuBSENSE [23], PAWCS [38] and DARMABS) based on the selected video sequences: waving trees (Row 1), time of day (Row 2), badminton (Row 3), camouflage (Row 4), skating (Row 5), pets2006 (Row 6), and library (Row 7).

TABLE 6. Average results of Recall, Precision, F-Measure, Specificity, FPR, FNR and PCC on the Wallflower dataset.

Algorithm	Recall	Precision	F-Measure	Specificity	FPR	FNR	PCC
GMM [15]	0.7286	0.8992	0.7984	0.9531	0.0469	0.2714	77.77
ViBe [19]	0.6648	0.8089	0.7008	0.8001	0.1999	0.3352	59.92
LBP [21]	0.7797	0.8847	0.8235	0.9275	0.0725	0.2203	79.32
LIBS [37]	**0.8751**	0.8176	0.8319	0.8321	0.1679	**0.1249**	76.67
PBAS [20]	0.8069	0.9133	0.8529	**0.9464**	**0.0536**	0.1931	**82.00**
SuBSENSE [23]	0.8026	**0.9597**	**0.8726**	**0.9595**	**0.0405**	0.1974	81.27
PAWCS [38]	**0.8267**	**0.9365**	**0.8774**	**0.9328**	**0.0672**	**0.1733**	**83.62**
DARMABS	**0.9229**	**0.9445**	**0.9333**	0.9297	0.0703	**0.0771**	**91.12**

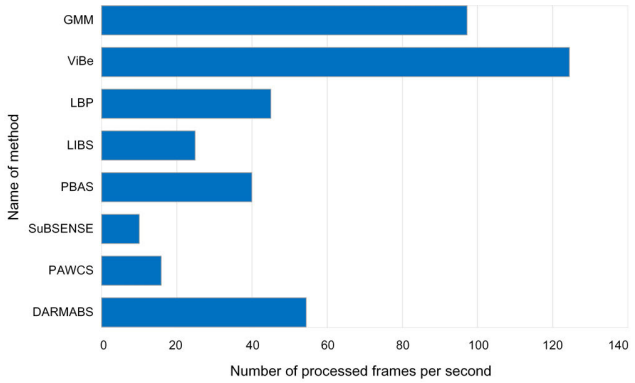


FIGURE 6. FPS of eight background subtraction algorithms for a 320 × 240 pixel image.

average *F-Measure* on both datasets, thereby demonstrating the superiority of the proposed scheme. In terms of *FNR*, *Recall* and *PCC*, DARMABS obtains the lowest, highest and highest values, respectively, which suggests that its detection results are exceedingly accurate. The scores of the remaining metrics are slightly inferior to that of SuBSENSE. However, it can be inferred from the next subsection that DARMABS performs well in terms of real-time performance compared with SuBSENSE. Thus, DARMABS shows considerably good performance in terms of the evaluation metrics.

3) RUNTIME ANALYSIS

Furthermore, to compare the time complexity of our algorithm with that of the other algorithms, we define the frames per second (*FPS*) as (20) to evaluate the processing speed. We do not compare DARMABS with the deep learning-based methods [6], [7], [33] since the latter consume substantial time and resources to train and are not a good choice for real-time visual surveillance applications. We run the proposed method on a 64-bit Windows 10 platform with an Intel Core i7-7700HQ CPU, 16 GB RAM and 2.8 GHz. The average processing speed of the proposed scheme is approximately 50 *FPS*. Fig. 6 shows the *FPS* of the background subtraction algorithms. Our algorithm is faster than LBP, LIBS, PBAS, SuBSENSE, and PAWCS and slower than GMM and ViBe. However, GMM and ViBe yield poor segmentation performance compared with our algorithm. Thus, our algorithm demonstrates great performance in processing speed and is suitable for applications requiring high real-time performance.

$$FPS = \frac{Number\ of\ processed\ frames}{Total\ computation\ time(in\ seconds)} \tag{20}$$

V. CONCLUSION

In this paper, we propose a simple but robust background subtraction technique called DARMABS for moving object detection in real-time visual surveillance systems. The basic idea is to adopt the ARMA process to model each pixel of an image and implement an adaptive LMS scheme to update the parameters to establish a robust and dynamic background model. The proposed algorithm does not require expensive hardware resources to perform the heavy computational task, in contrast to most of the recently proposed algorithms.

Various video sequences with complex dynamic backgrounds from the CDnet2014 and Wallflower datasets are used to test the generalization ability and adaptability of the proposed scheme. This technique shows good performance for moving objects detection compared to other mainstream techniques (i.e., GMM, ViBe, LBP, LIBS, PBAS, SuBSENSE and PAWCS). Based on the above experimental results, we believe that the proposed scheme can provide a robust and real-time moving object detection method for video sequences with complex dynamic backgrounds. In future work, we will further study the mixture of ARMA models to model the background to improve the detection results.

ACKNOWLEDGMENT

The authors are grateful to the reviewers and editors for their valuable feedback on our work that helped to improve the quality of this paper.

REFERENCES

[1] H. Wang, C. Yuan, W. Hu, H. Ling, W. Yang, and C. Sun, "Action recognition using nonnegative action component representation and sparse basis selection," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 570–581, Feb. 2014.

[2] A. Elnakeeb and U. Mitra, "Line constrained estimation with applications to target tracking: Exploiting sparsity and low-rank," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6488–6502, Dec. 2018.

[3] H. Lee, H. Kim, and J.-I. Kim, "Background subtraction using background sets with image- and color-space reduction," *IEEE Trans. Multimedia*, vol. 18, no. 10, pp. 2093–2103, Oct. 2016.

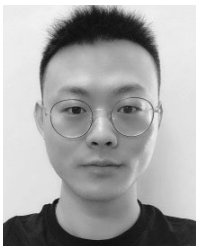
[4] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," *Comput. Sci. Rev.*, vols. 11–12, pp. 31–66, May 2014.

[5] S. Brutzer, B. Hoferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1937–1944.

[6] M. Braham and M. Van Droogenbroeck, "Deep background subtraction with scene-specific convolutional neural networks," in *Proc. IEEE Int. Conf. Syst., Signals Image Process.*, May. 2016, pp. 1–4.
[7] D. Zeng and M. Zhu, "Background subtraction using multiscale fully convolutional network," *IEEE Access*, vol. 6, pp. 16010–16021, 2018.
[8] Y. Zhang, Y. Zhang, Z. Deng, X.-P. Zhang, and H. Liu, "Sea surface target detection based on complex ARMA-GARCH processes," *Digit. Signal Process.*, vol. 70, pp. 1–13, Nov. 2017.
[9] J. P. Pascual, N. von Ellenrieder, M. Hurtado, and C. H. Muravchik, "Adaptive radar detection algorithm based on an autoregressive GARCH-2D clutter model," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3822–3832, Aug. 2014.
[10] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1778–1792, Nov. 2005.
[11] J. Zhong and S. Sclaroff, "Segmenting foreground objects from a dynamic textured background via a robust Kalman filter," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 44–50.
[12] L. Lin, Y. Xu, X. Liang, and J. Lai, "Complex background subtraction by pursuing dynamic spatio-temporal models," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3191–3202, Jul. 2014.
[13] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, Jul. 1997.
[14] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Background modeling and subtraction by codebook construction," in *Proc. Int. Conf. Image Process. (ICIP)*, Singapore, Oct. 2004, pp. 3061–3064.
[15] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Fort Collins, CO, USA, vol. 2, Jun. 1999, pp. 246–252.
[16] S. Varadarajan, P. Miller, and H. Zhou, "Spatial mixture of Gaussians for dynamic background modelling," in *Proc. 10th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Krakow, Poland, Aug. 2013, pp. 63–68.
[17] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. 17th Int. Conf. Pattern Recognit.*, vol. 2, Aug. 2004, pp. 28–31.
[18] Q. Zang and R. Klette, "Robust background subtraction and maintenance," in *Proc. Int. Conf. Pattern Recognit.*, vol. 2, Aug. 2004, pp. 90–93.
[19] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.
[20] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 38–43.
[21] M. Heikkilä and M. Pietikäinen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 657–662, Apr. 2006.
[22] S. Zhang, H. Yao, and S. Liu, "Dynamic background modeling and subtraction using spatio-temporal local binary patterns," in *Proc. ICIP*, Oct. 2008, pp. 1556–1559.
[23] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "SuBSENSE: A universal change detection method with local adaptive sensitivity," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 359–373, Jan. 2015.
[24] N. Jacobs and R. Pless, "Shape background modeling : The shape of things that came," in *Proc. IEEE Workshop Motion Video Comput.*, Feb. 2007, p. 27.
[25] S. Varadarajan, P. Miller, and H. Zhou, "Region-based mixture of Gaussians modelling for foreground detection in dynamic scenes," *Pattern Recognit.*, vol. 48, no. 11, pp. 3488–3503, Nov. 2015.
[26] Z. Gao, L.-F. Cheong, and Y.-X. Wang, "Block-sparse RPCA for salient motion detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 1975–1987, Oct. 2014.
[27] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 597–610, Mar. 2013.
[28] F. Seidel, C. Hage, and M. Kleinsteuber, "pROST: A smoothed p-norm robust online subspace tracking method for background subtraction in video," *Mach. Vis. Appl.*, vol. 25, no. 5, pp. 1227–1240, Jul. 2014.
[29] M. Chen, X. Wei, Q. Yang, Q. Li, G. Wang, and M.-H. Yang, "Spatiotemporal GMM for background subtraction with superpixel hierarchy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1518–1525, Jun. 2018.
[30] W. Fang, T. Zhang, C. Zhao, D. Soomro, R. Taj, and H. Hu, "Background subtraction based on random superpixels under multiple scales for video analytics," *IEEE Access*, vol. 6, pp. 33376–33386, 2018.
[31] D. Giordano, F. Murabito, S. Palazzo, and C. Spampinato, "Superpixel-based video object segmentation using perceptual organization and location prior," in *Proc. IEEE CVPR*, Jun. 2015, pp. 4814–4822.
[32] M. Babaee, D. T. Dinh, and G. Rigoll, "A deep convolutional neural network for video sequence background subtraction," *Pattern Recognit.*, vol. 19, no. 2, pp. 635–649, Apr. 2017.
[33] Y. Zhang, X. Li, Z. Zhang, F. Wu, and L. Zhao, "Deep learning driven blockwise moving object detection with binary scene modeling," *Neurocomputing*, vol. 168, pp. 454–463, Nov. 2015.
[34] Y. Wang, Z. Luo, and P.-M. Jodoin, "Interactive deep learning method for segmenting moving objects," *Pattern Recognit. Lett.*, vol. 96, pp. 66–75, Sep. 2016.
[35] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "CDnet 2014: An expanded change detection benchmark dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 387–394.
[36] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," in *Proc. 7th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1. Sep. 1999, pp. 255–261.
[37] K. K. Hati, P. K. Sa, and B. Majhi, "Intensity range based background subtraction for effective object detection," *IEEE Signal Process. Lett.*, vol. 20, no. 8, pp. 759–762, Aug. 2013.
[38] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "Universal background subtraction using word consensus models," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4768–4781, Oct. 2016.

JIAN LI received the B.S. degree in mechanical engineering and automation from the Harbin Institute of Technology, Harbin, China, in 2017. He is currently pursuing the M.S. degree in instrument engineering with the College of Intelligence Science and Technology, National University of Defense Technology. His research interests include image processing, computer vision, image compression, information fusion, wireless sensor networks, and pattern recognition.

ZHONG-MING PAN received the B.S., M.S., and Ph.D. degrees from the National University of Defense Technology, China, in 1982, 1985, and 2006, respectively, where he is currently a Professor with the College of Intelligence Science and Technology. His research interests include wireless sensor networks, pattern recognition, sensors, and battlefield condition monitoring.



ZHUO-HANG ZHANG received the B.S. degree in mechanical engineering from the South China University of Technology, Guangzhou, China, and the M.S. degree from Air Force Engineering University, China. He is currently pursuing the Ph.D. degree in instrumentation science and technology with the College of Intelligence Science and Technology, National University of Defense Technology. His research interests include ultra-wideband radar design and wireless sensor networks.

HENG ZHANG received the B.S. degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, and the M.S. and Ph.D. degrees from the National University of Defense Technology, China, where he is currently a Lecturer with the College of Advanced Interdisciplinary Research. His research interests include high power microwave and wireless sensor networks.