Received December 8, 2018, accepted December 30, 2018, date of publication January 7, 2019, date of current version January 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2891084

Locally Statistical Dual-Mode Background Subtraction Approach

THIEN HUYNH-THE[1], (Member, IEEE), CAM-HAO HUA[2], NGUYEN ANH TU[2], AND DONG-SEONG KIM[1], (Senior Member, IEEE)

[1]ICT Convergence Research Center, Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi 39177, South Korea
[2]Department of Computer Science and Engineering, Kyung Hee University, Yongin 446-701, South Korea

Corresponding author: Dong-Seong Kim (dskim@kumoh.ac.kr)

This work was supported in part by the Ministry of Science, ICT (MSIT), South Korea, under the Information Technology Research Center (ITRC) Support Program supervised by the Institute for Information and Communications Technology Promotion (IITP) under Grant IITP-2018-2014-1-00639, and in part by the Priority Research Centers Program through the National Research Foundation of Korea (NRF) through the Ministry of Education, Science, and Technology under Grant 2018R1A6A1A03024003.

ABSTRACT Due to the variety of background model in the real world, detecting changes in a video cannot be addressed exhaustively by a simple background subtraction method, especially with several motion detection challenges, such as dynamic background, camera jitter, intermittent object motion, and so on. In this paper, we propose an efficient background subtraction method, namely locally statistical dual-mode (LSD), for detecting moving objects in video-based surveillance systems. The method includes a local intensity pattern comparison algorithm for foreground segmentation by analyzing the homogeneity of intensity patterns of the input frame and the background model, in which the homogeneity is calculated by the mean and standard deviation of pixel intensity. Besides that, a dual-mode scheme is developed to temporally update the background model for the short- and long-term scenarios corresponding to sudden and gradual changes in the background. The advantage of this scheme is the allowance of updating the model in both pixel- and frame-wise manners simultaneously. The parameters used in both the local intensity pattern comparison algorithm and the dual-mode background model updating scheme are estimated for every input frame consecutively based on local and global statistical information of segmentation result. In experiments, the proposed LSD method is extensively evaluated on the Wallflower and CDnet2014 datasets; and remarkable performance demonstrates its preeminence to the many state-of-the-art background subtraction approaches in terms of segmentation accuracy and computational complexity.

INDEX TERMS Motion detection, background subtraction, background modeling, moving object detection, video segmentation, locally statistical dual-mode updating.

I. INTRODUCTION

Detecting moving object is a fundamental pre-processing step in numerous image processing and computer vision applications, e.g., object detection, human activity recognition, abnormal detection, video analysis, and so on [1]. In order to detect and segment moving object in a scene, background subtraction technique is widely used in real-time systems thanks to its simplicity and low computational complexity, wherein background initialization, foreground detection, and background maintenance play an important role in the general framework of background subtraction [2]. However, the detection accuracy of existing background subtraction methods is almost unfavorable under different motion detection challenges, such as dynamic background motion, camera jitter, intermittent object motion, bad weather, and mixture [3]. In other words, dealing with various realistic scenarios cannot be addressed thoroughly by a simple background subtraction technique. Therefore, a vast amount of studies have been taken to enhance the detection accuracy by improving either the segmentation algorithm or the background model updating scheme while ensuring the capability of real-time processing [4]–[6]. Especially, since the number of video-based surveillance applications increases rapidly, an efficient background subtraction method, wherein the foreground is segmented precisely with an adaptive background model updating scheme, is urgently required for the time being.

To this end, we propose a novel background subtraction method, namely Locally Statistical Dual-Mode (LSD), that is able to carry out current motion detection challenges effectively. In this method, we develop a novel local intensity pattern comparison algorithm for foreground detection and an efficient dual-mode updating scheme for background model estimation. In the beginning, a background model is initialized as a buffer containing a certain number of input frames with the size identified in advance. The buffer aims to collect and accumulate the background information during moving object detection process, hence it will be updated consecutively for every input frame coming. In our proposed LSD method, the moving objects (a.k.a. foreground) in a scene are detected based on the comparison of pattern homogeneity with the assumption that background pattern is essentially more homogeneous than foreground pattern. In fact, a pixel is classified into either foreground or background by analyzing the local homogeneity of two pixel-surrounded patterns extracted from the input frame and background model. The local homogeneity is measured by the standard deviation metric, however, for dealing with smooth texture objects, the mean metric is further used as an additional criteria of pattern comparison. The foreground is segmented following a comparison rule with two decision thresholds corresponding to the mean and standard deviation. It is noticed that there are two detection processes, i.e. called roughness and refinement detection, involved in our algorithm for improving segmentation accuracy and reduce computational complexity, respectively. For adapting to sudden and gradual changes in a scene, the background model is updated by a dual-mode scheme that responds to short- and long-term motion detection scenarios. With the short-term mode, the background model stored in the buffer is updated following a pixel-wise manner while the long-term mode update following a frame-wise manner. Two updating modes are executed simultaneously to exhaustively maintain the adaptation of the proposed model. The last step of LSD is parameter estimation, where all parameters in the local pattern comparison algorithm are estimated for every input frame based on the local and global statistical information of corresponding segmentation result. The values of parameter are updated for the next frame processing.

Compared with state-of-the-art background subtraction approaches, the proposed LSD method offers four benefits:

- Moving objects in a scene are segmented by analyzing the homogeneity of local intensity pattern for accurate detection.
- There are two detection processes, i.e. roughness and refinement for performance improvement.
- A dual-mode background model updating scheme robustly adapts to sudden and gradual changes normally occurring in various motion detection challenges.
- Segmentation accuracy is improved with the decision thresholds for each particular pixel classification using local and global statistical information.

The remainder of this paper is organized as follows. Section II summarizes state-of-the-art background subtraction approaches. Section III presents our proposed LSD method including the background model initialization, the intensity pattern comparison algorithm, the dual-model background model updating scheme, and the parameter estimation. The experiment for performance evaluation is given in Section IV, where the discussion and comparison with other methods are comprehensively provided. The conclusion and future work are offered in Section V.

II. RELATED WORKS

Most of basic background subtraction methods classify a pixel to be either the foreground class, denoted by 1-bit pixel, or the background class, denote by 0-bit pixel, based on extracting statistical information of pixel illumination. Due to the principal importance, enormous background subtraction methods mostly contribute to background modeling for performance enhancement.

The background is essentially modeled as a reference image by following techniques of running average [7], approximated median [8], and histogram over time [9], in which the foreground is extracted by differentiating input frame and background image. Despite the simplicity and ease of implementation, basic models are sensitive to sudden luminance changes. Pixel-based [10] and region-based [11] statistical information of input frames were considered in many background subtraction methods to handle the variation of luminance. As a weakness, simple statistical approaches usually fail in challenges of intermittent object motion challenge due to imbibing abandoned object to background. Several advanced statistical background modeling techniques, including Gaussian Mixture Model (GMM) [12] and its improvement versions [13]–[21], have been developed for multi-modal background scenario. Instead of modeling all pixels by only one distribution, GMM studies a pixel value by a mixture of Gaussian. A typical pixel-based improvement of GMM is recommended in [13], wherein GMM parameters are constantly learned by recursive equations to particularly select an appropriate number of components for each pixel. As an extension from the pixel-based, region-based mixture of Gaussians (RMoG) [21] models a region of pixels as a mixture distribution for the purpose of effectively handling complex dynamic texture challenge. Lately, by taking advantage of superpixel for enhancing spatial coherency, a superpixel-based hierarchical architecture of GMMs [22], denoted STSHBM, is constructed to handle repetitive and sudden changes of pixel intensity in video segmentation. Compared with other variants of GMM, STSHBM significantly improves the accuracy of foreground detection, but nevertheless, the rapid increment of computational complexity is currently its limitation. The disadvantage of GMM-based approaches is the assumption that the background region should be larger and more frequently visible than the foreground region.

To deal with the parameter estimation issue of parametric models as above mentioned, some non-parametric techniques have been introduced for background modeling, such as codebook construction [23]–[26], Kernel Density Estimation (KDE) [27]–[31]. In the most of codebook-based approaches, the background pixel intensity values are clustered to corresponding codewords, and consequently, a compressed background model has the potential of representing several frames by a set of determined codewords (a.k.a codebook). Different from GMM, KDE was proposed by Elgammal *et al.* [27] to estimate the probability density function by a histogram for updating the background model. To explicitly address multi-modal background scenario, where there exist current moving objects, short- and long-term stationary objects, Cuevas *et al.* [28] learned a background modeling architecture constructed by three parallel KDE models with different absorption rates. In other researches, this technique is extensively developed for other tasks consisted of foreground extraction [29], [30] and contour enhancement of segmentation result [31] besides background modeling. Although KDE-based approaches have the significant potential against various motion detection challenges, they consume a great deal of memory for kernel estimation due to the usage of a large number of historical observations.

Several advanced background modeling techniques have been introduced for learning multi-modal background scenarios. Recently, Visual Background Extractor (ViBE) [32], [33] initially builds a background model by accumulating several observation samples at every pixel location temporally and smoothly updates the model in decaying lifespan by a random pixel selection policy. Another contribution of ViBE is foreground detection algorithm, wherein a pixel is classified into either foreground or background based on randomly comparing to its surrounding neighbors. In spite of the less computational cost, the segmentation accuracy of foreground produced by ViBE is not impressive. Inspired by ViBE, Pixel-Based Adaptive Segmentation (PBAS) [34] adaptively estimates a particular decision threshold and learning parameter for each pixel based on the analysis of motions in the background for pixel classification. Due to the lack of an efficient illumination-object discrimination scheme, PBAS may riskily absorb motionless objects into the background class. By taking advantages of an artificial neural network, Self-Organizing Background Subtraction (SOBS) [35], [36] is capable of capturing structural background variation from observed periodic-like motion. SOBS is quite good with almost motion detection challenges, except intermittent object motion. Another advanced pixel-wise background subtraction method is Neighbor-based Intensity Correction (NIC) [37], which updates the background model, represented by a single image, over an intensity updating rule. An improvement version of NIC was presented in [38], in which foreground detection accuracy is raised by a directional feature-based mask selection scheme and a historical intensity pattern reference algorithm. Both NIC and its improvement deliver competitive accuracy, but they lack an efficient background model updating scheme. In [39], merging two foreground masks which are detected by an optical flow algorithm and GMM-based background subtraction method, is done by the Graph Cut algorithm. This method should be intensively improved due to its weakness against dynamic background motion and intermittent object motion challenges. Recently, an intelligent background updating schemes [40] has been developed for the PBAS-based background subtraction framework, in which the updating schedule of neighbor pixels is maintained by a counter to determine whether a pixel belonging to either an object class or an illumination region.

III. METHODOLOGY

This section presents our proposed background subtraction method for detecting moving objects in a scene, of which the overall workflow is illustrated in Fig. 1. Four main components, including background model initialization, intensity pattern comparison, dual-mode background model update, and parameter estimation are exhaustively described as follows.

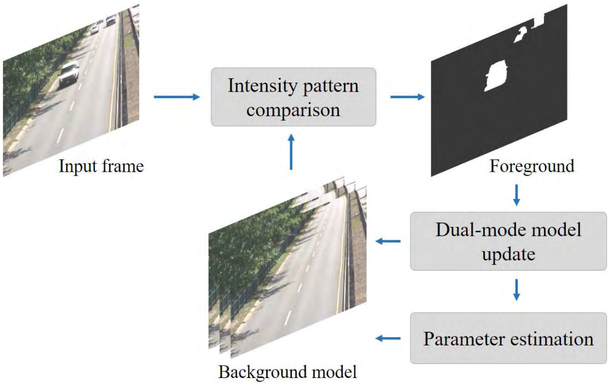

FIGURE 1. The overall workflow of the proposed background subtraction with dual-mode model updating scheme based in the locally statistical information.

A. BACKGROUND MODEL INITIALIZATION

The background model is initialized by collecting a number of input frames segmented from a video and storing in a buffer. For optimizing memory usage and computational cost of processing, these frames are transformed from RGB to gray-scale. The buffer is characterized by a size value corresponding to the number stored frames. If the buffer size is N, the method is able to detect the foreground from $(N+1)^{th}$ frame till the end of a video. Obviously, the larger the buffer size is, more memory and computational complexity the method occupies. It should be noted that the buffer is initialized at first and then consecutively updated during the foreground detection progress through the dual-mode model update scheme, where the short-term and long-term mode included in the updating scheme allow the background model to be updated at both pixel-wise and frame-wise levels to adaptively respond the variation changes of the background.

B. INTENSITY PATTERN COMPARISON

In the intensity pattern comparison component, the moving objects in a scene are automatically detected and segmented to a binary image, a.k.a. foreground image, based on the exploration of local pixel homogeneity. Concretely, by analyzing local pixel intensity patterns extracted from the input frame and the background model, the method is potential to assign a pixel to be either foreground or background class, denoted by 1-bits or 0-bits respectively. With the assumption that a local background region is generally more homogeneous than a local foreground region, and therefore it is reasonable to identify a pixel of the input frame belonging to the foreground class if its surrounding pattern gets a higher homogeneity than those of the background model at the same pixel coordinate. An example is visually illustrated in Fig. 2. To measure the homogeneity, standard deviation metric is used in this research due to its ability to represent the local dispersion of a pixel value set.

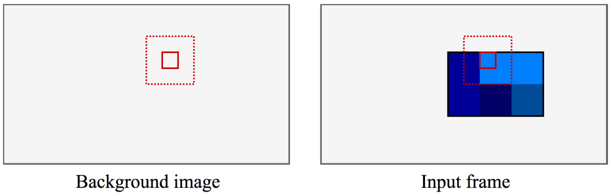

FIGURE 2. The visual illustration of local homogeneity analysis for a pixel (marked by a solid line). According to pixel coordinate, two local intensity patterns (marked by a dot line) are captured in encircling by a square mask of size 3 from the background image and the input frame. By comparing their intensity homogeneity impacts, the pixel in the input frame is classified to the foreground group because the local intensity homogeneity in the input frame is less than the homogeneity in the background image.

As the first step of the intensity pattern comparison algorithm, the difference between the input frame I and the representative image J of background model is extracted to expose arguable foreground pixels. The representative image J is defined as the median image of m frames F that are randomly selected from the N-sized buffer:

$$J\left(x,y\right) = median\left\{F_i\left(x,y\right)|i \in rand\left(m,N\right)\right\}, \qquad (1)$$

where (x, y) denotes the coordinate of a pixel. Identification of arguable foreground pixels is done by comparing the difference $|I(x, y) - J(x, y)|$ with a value, called rough threshold τ_ρ, as follows

$$D\left(x,y\right) = \begin{cases} 1; & \left|I\left(x,y\right) - J\left(x,y\right)\right| \geq \tau_\rho \\ 0; & otherwise. \end{cases} \qquad (2)$$

The binary image D is considered as a preliminary foreground mask of the input frame. Since D might contain the noise and outliers of dynamic background, camera jitter, and other challenges, all 1-bit pixels referring to foreground need to be refined with an intensity pattern comparison algorithm. Concretely, at the foreground pixel (x, y) where $D(x, y) = 1$, we extract two local intensity patterns, i.e., one of the input frame I, denoted $P_I^{(x,y)}$ and another of the representative image J, denoted $P_J^{(x,y)}$, by a square mask surrounding (x, y) as illustrated in Fig. 2. A more homogeneous pattern is equivalent to a smaller value of standard deviation. However, simply considering standard deviation is not meaningful enough to segment foreground precisely, even if two patterns can be distinguished unambiguously. Therefore, an additional mean metric is useful for detecting smooth texture objects, where the homogeneity of an object is sometimes greater than those of background. Given a pattern P consisted of n pixels $\{p_1, p_2, p_3, \ldots, p_n\}$, the mean and standard deviation are defined as follows

$$\mu = \frac{1}{n}\sum_{i=1}^{n} p_i, \qquad (3)$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(p_i - \mu\right)^2}, \qquad (4)$$

where n equals to the square mask's size squared, for instance, $n = 9$ corresponding to the mask size of 3 as the example in Fig. 2. Corresponding to $P_I^{(x,y)}$ and $P_J^{(x,y)}$, the mean values $\mu_I^{(x,y)}$ and $\mu_J^{(x,y)}$ and the standard deviation values $\sigma_I^{(x,y)}$ and $\sigma_J^{(x,y)}$, respectively, are calculated and compared together for foreground refinement. The refinement process is performed by the following rule

$$O\left(x,y\right) = \begin{cases} 1; & \Delta_\mu^{(x,y)} \geq \tau_\mu \cup \Delta_\sigma^{(x,y)} \geq \tau_\sigma \\ 0; & otherwise, \end{cases} \qquad (5)$$

where O is the segmented foreground image. Two parameters τ_μ and τ_σ refer to as the decision thresholds of mean difference and standard deviation difference, respectively. The difference of mean $\Delta_\mu^{(x,y)}$ and standard deviation $\Delta_\sigma^{(x,y)}$ are determined as follows

$$\Delta_\mu^{(x,y)} = \left|\mu_I^{(x,y)} - \mu_J^{(x,y)}\right|, \qquad (6)$$

$$\Delta_\sigma^{(x,y)} = \left|\sigma_I^{(x,y)} - \sigma_J^{(x,y)}\right|. \qquad (7)$$

It is noted that the binary image O is the final foreground mask achieved by refining all arguable pixels of the preliminary foreground D using (5). The foreground segmentation, involving rough and refinement detection, aims to improve the accuracy of moving object detection. As aforementioned, the preliminary foreground D achieved by the rough detection may have noise and outliers. By the refinement process with our proposed local pattern comparison algorithm, the noise is essentially eliminated to produce a high-quality foreground segmentation result. The binary foreground image O sometimes comprises of disconnected edges caused by the harsh changes of luminance in some background challenges. Thus, the closing morphological operation is used as a post-processing step to merge narrow disruption and long thin hollows, eliminate small holes, and fill gaps in the contour.

C. DUAL-MODE BACKGROUND MODEL UPDATE

To adaptively respond the changes of scene background over time, an efficient dual-mode updating scheme is proposed to learn the background model for the short- and long-term scenarios corresponding to the sudden and gradual changes of intensity, respectively.

In practical environments, the sudden intensity changes of background usually occur locally due to several challenges, such as background oscillation, camera jitter, and etc. For example, some such sturdy background motions as the waving of leaves and the sparkle water can be misclassified to the foreground class. Without the knowledge of background scenarios, completely addressing these difficulties is nearly impossible due to the variety of appearance, however, their effectiveness on the segmented foreground can be mitigated by updating the background model to be more robust. Basically, sudden intensity changes can be detected based on the difference between two consecutive frames. According to the determined locations, we can update the background model stored in the buffer by a representative background image which is identified as the median image formerly. This pixel-wise background updating progress, called the short-term mode, is successively executed to quickly adapt the intensity changes of background. Insides the short-term mode, the pixel intensity of several random frames in the buffer is selectively updated. The pixel values of some samples F in the buffer will be replaced by those of the representative image J at the corresponding locations of motion pixels which is given by the difference between two consecutive input frames of I_{t-1} and I_t at the time stamp t. The short-term mode is accomplished for every input frame

$$F_i(x,y)\,|i \in rand(m,N) = J(x,y); \quad \forall(x,y)\,|\mathcal{D}(x,y) = 1, \tag{8}$$

where the difference mask $\mathcal{D}$ of two consecutive input frames is defined as follows

$$\mathcal{D}(x,y) = \begin{cases} 1; & |I_t(x,y) - I_{t-1}(x,y)| \geq \tau_\rho \\ 0; & otherwise. \end{cases} \tag{9}$$

Following (9), the background model is updated by the intensity values of J for only movement pixels, wherein $\mathcal{D}(x,y) = 1$. While remaining pixels, wherein $\mathcal{D}(x,y) = 0$, are unchanged of intensity, that means, there is no updating for them. It is noticed that only m frames in the buffer are randomly selected for the short-term mode to improve the robustness of background model. For the next input, the foreground is segmented more accurately based on the new representative image which is computed on the updated background model.

Besides locally sudden changes of background, foreground detection has to face with such several other challenges as the gradual change of illumination and intermittent object motion. Some moving objects are abandoned for a while in a scene and abruptly moved again. They can be incorporated into the background class, that conducts some misclassification, known as "ghosting" artifact (i.e., a foreground object cannot be detected during the abandonment time while a void area appears in background henceforward). In other cases, some non-stationary objects learned to the background model initially, known as "bootstrapping", will induce faulty detection in the foreground (that means, although the movement of these objects is detected, their non-stationary regions will appear hereafter). Essentially, these background scenarios are potentially discovered based on the correlation of motion pixels (known as the difference between two consecutive frames) and foreground pixels. For instance, there is no any motion detected in a scene while the abandoned object is still segmented in the foreground. To alleviate the influence of above challenges, the background model should be updated with the adaptive frequency which is estimated over the ratio of the number of motion pixels to the number of foreground pixels. This progress, called the long-term mode, updates the buffer following a frame-wise manner, i.e., some certain frames in the buffer are substituted by the current input frame I_t. In general, updating buffer is usually controlled by a factor, called the buffer sampling rate (frame per second - fps), which indicates the number of samples replaced in one second. Different from existing approaches, where the buffer sampling rate is set by a constant value, our scheme updates the buffer with an adaptive rate β (fps), which is defined by the following function

$$\beta = \frac{1}{\lfloor \alpha r^2 \rfloor}, \tag{10}$$

where α is the frame rate (fps) of a video (e.g., 30 fps for common video recording) and r referring to as the ratio of the number of motion pixels (i.e., $\mathcal{D}(x,y) = 1$) to the number of foreground pixels (i.e., $O(x,y) = 1$) is calculated as follows

$$r = \frac{\mathcal{N}_{\mathcal{D}(x,y)=1}}{\mathcal{N}_{O(x,y)=1}}. \tag{11}$$

The sampling rate β should be an integer number and constrained in the range $[0, N]$. In order to avoid null result, the value of r is constrained in the range $(0, 1]$. It should be noted that the buffer will be updated every second, therefore, $\mathcal{N}_D$ and $\mathcal{N}_O$ are summarized as the number of motion pixels and foreground pixels in one second. As an example with the frame rate $\alpha = K$ (fps), the ratio r is calculated by (10) after accumulating enough K difference masks and also K foreground images. After every K input frames, β samples in the buffer will be replaced by a corresponding number of input frames randomly. The proposed background model updating scheme is depicted in Fig. 3.

D. PARAMETER ESTIMATION

Three parameters τ_ρ, τ_μ, and τ_σ that are introduced in the intensity pattern comparison process for moving objects detection and the dual-mode background model updating scheme should be tuned carefully due to their effectiveness on the detection accuracy and model quality as well. Instead of treating them as global thresholds like most of the existing works, these values are particularly assigned to each individual pixel. These thresholds are automatically calculated for adapting to the luminance variation of gradual and sudden changes, and to be robust against various motion detection challenges.

Generally, all three thresholds are calculated over an intermediate parameter, denoted η, which represents the

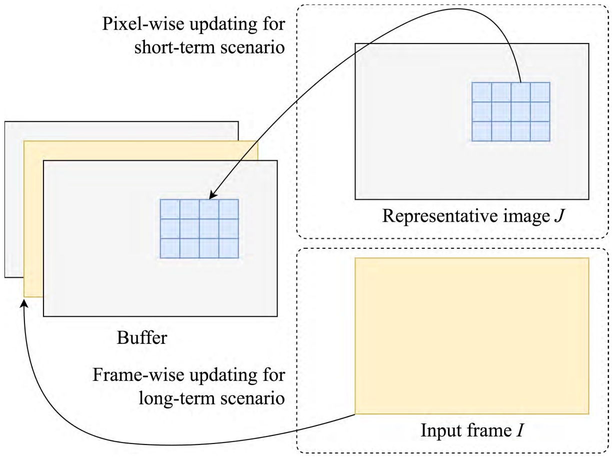

FIGURE 3. The illustration of dual-model background model updating scheme, wherein the pixel- and frame-wise update processes are for short- and long-term scenarios, respectively.

variability of a scene. In details, η is updated for processing the next input frame as follows

$$\eta_t(x,y)=\begin{cases}\min\left(1.0,\eta_{t-1}(x,y)+\delta\right); & \bar{h}_t(x,y)\geq h_0\\ \max\left(0.1,\eta_{t-1}(x,y)-\delta\right); & otherwise,\end{cases}\tag{12}$$

where $\delta = \frac{1}{2N}$ is the step size and h_0 is the threshold of normalized intensity variation. The calculation of normalized intensity variation $\bar{h}$ for all pixels is formulated as follows

$$\bar{h}_t(x,y)=\frac{h_t(x,y)}{t},\tag{13}$$

where

$$h_t(x,y)=\begin{cases}h_{t-1}(x,y)+1; & D_{\tau_\rho=20}(x,y)=1\\ h_{t-1}(x,y)+0; & otherwise,\end{cases}\tag{14}$$

where the intensity variation h is initialized to zero (i.e., $h_{t=0} = 0$) and only calculated for the case of objects detected in a scene (i.e., ignore this parameter calculation for stationary frame without moving objects detected). The rough threshold τ_ρ for processing the next frame is defined for each particular pixel based on the result of current foreground segmentation by the following function

$$\tau_\rho(x,y)=\eta_t(x,y)\,\Delta_{local}+\Delta_{global},\tag{15}$$

where Δ_{global} is the average of difference between the input frame I and the representative image J for all pixels, while Δ_{local} is the average of difference between the input frame I and the representative image J for only foreground pixels

$$\Delta_{global}=\frac{\sum\limits_{\forall(x,y)}\left|I(x,y)-J(x,y)\right|}{U},\tag{16}$$

$$\Delta_{local}=\frac{\sum\limits_{\forall(x,y)|O(x,y)=1}\left|I(x,y)-J(x,y)\right|}{\mathbf{U}},\tag{17}$$

where U is the total number of image pixels and $\mathbf{U}$ refers to as the number of foreground pixels only, wherein $O(x,y)=1$. It should be noticed that the rough threshold τ_ρ is estimated not only for segmenting the foreground but also for updating the background model following the short-term mode. The step size δ is associated with the buffer size N over the threshold τ_ρ (i.e., for a larger buffer, a smaller step size will conduct a smaller rough threshold, that leads to more pixels updated in the buffer; on the contrary, less pixels will be updated for a smaller buffer). Therefore, the step size δ should be computed based on the buffer size N to maintain the updating speed of background model appropriately. Two thresholds τ_μ and τ_σ for making decision of mean and standard deviation are estimated as follows

$$\tau_\mu(x,y)=\begin{cases}\tau_{\mu\_1}; & \Delta_\mu^{(x,y)}<\tau_{\mu\_1}\\ \tau_{\mu\_2}; & otherwise,\end{cases}$$
$$\tau_\sigma(x,y)=\begin{cases}\tau_{\sigma\_1}; & \Delta_\sigma^{(x,y)}<\tau_{\sigma\_1}\\ \tau_{\sigma\_2}; & otherwise,\end{cases}\tag{18}$$

where $\tau_{\mu\_2} = \max\left(\tau_{\mu\_1},\eta_t(x,y)\,\bar{\Delta}_\mu\right)$ and $\tau_{\sigma\_2} = \max\left(\tau_{\sigma\_1},\eta_t(x,y)\,\bar{\Delta}_\sigma\right)$, in which $\bar{\Delta}_\mu$ and $\bar{\Delta}_\sigma$ are the average of difference margin values that are greater than the lower thresholds τ_{μ_1}= 20 and τ_{σ_1}= 20 corresponding to the mean and standard deviation. In particular, $\bar{\Delta}_\mu$ and $\bar{\Delta}_\sigma$ are computed as follows

$$\bar{\Delta}_\mu=\frac{1}{n_{\Delta_\mu}}\sum_{(x,y)}\Delta_\mu^{(x,y)};\ \Delta_\mu^{(x,y)}\geq\tau_{\mu\_1},$$
$$\bar{\Delta}_\sigma=\frac{1}{n_{\Delta_\sigma}}\sum_{(x,y)}\Delta_\sigma^{(x,y)};\ \Delta_\sigma^{(x,y)}\geq\tau_{\sigma\_1},\tag{19}$$

where n_{Δ_μ} is the numbers of pixels that satisfies the condition of $\Delta_\mu^{(x,y)}\geq\tau_{\mu\_1}$ and n_{Δ_σ} stands for the number of pixels that meets the condition of $\Delta_\sigma^{(x,y)}\geq\tau_{\sigma\_1}$.

The details of our proposed Locally Statistical Dual-Mode background subtraction method is presented in Fig. 4 and summarized in the following steps:

1) Initialize the background model with a N-sized buffer.
2) Extract the segmented foreground by a local intensity pattern comparison algorithm.
   - Calculate the representative image J by (1).
   - Extract the preliminary foreground image D with the threshold τ_ρ in (2).
   - For each arguable foreground pixel in D, capture two local intensity patterns P_I and P_J.
   - According to (3) and (4), the mean μ and standard deviation σ are retrieved.
   - Obtain the difference of mean Δ_μ and standard deviation Δ_σ by (6) and (7), respectively.
   - Perform the refinement process with two decision threshold τ_μ and τ_σ by the mean of (5).
   - Finalize the foreground O with post-processing.
   - Output is the segmentation image O.
3) Update the background model with a dual-mode scheme.

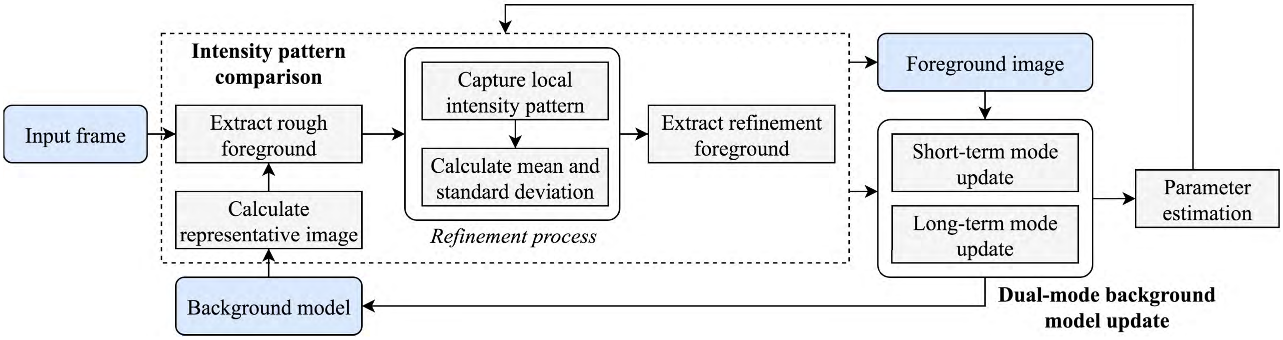
Intensity pattern comparison
Input frame
Extract rough foreground
Calculate representative image
Background model
Capture local intensity pattern
Calculate mean and standard deviation
Refinement process
Extract refinement foreground
Foreground image
Short-term mode update
Long-term mode update
Parameter estimation
Dual-mode background model update

FIGURE 4. The detail workflow of Locally Statistical Dual-Mode background subtraction method.

- For the short-term mode, extract the difference mask $\mathcal{D}$ via (9) to identify motion pixels.
- Update intensity of samples in the buffer by intensity of the representative image at corresponding pixels over (8).
- For the long-term mode, the buffer sampling rate β is determined by (10).
- Update the buffer by replacing some samples with appropriate input images.

4) Estimate parameters used in foreground detection.

- From (13), the normalized intensity variation $\overline{h}$ for all pixels is calculated via (14).
- Supported by (12), the scene variability factor η is determined.
- Retrieve the rough threshold τ_ρ (15) with Δ_{global} and Δ_{global} estimated by (16) and (17), respectively.
- Based on (19), two decision thresholds of mean τ_μ and standard deviation τ_σ are figured out over (18).

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we evaluate the proposed background subtraction method and further compare with state-of-the-art approaches in terms of foreground detection accuracy and computational complexity on several videos representing various realistic challenges.

A. EXPERIMENTAL SETUP

Our proposed Locally Statistical Dual-Mode (LSD) method is evaluated on the Wallflower dataset [41] and the ChangeDetection 2014 (CDnet2014) dataset [42]. There are seven image sequences in the Wallflower dataset for performance evaluation.

- *Camouflage*: This sequence presents the challenge of which moving objects are consolidated to the background by color or intensity.
- *Bootstrapping*: This sample is recorded in the indoor environment, where moving objects are appeared at beginning.
- *TimeOfDay*: The gradual change of illumination in the background is shown in this video.
- *LightSwitch*: This sequence depicts the sudden change of global illumination.
- *WavingTrees*: This video illustrates an dynamic background motion, where a man walks across swaying trees.
- *ForegroundAperture*: This sample represents for a motionless object which cannot distinguished from a background.
- *MovedObject*: This sequence illustrates for background displacement, where the location of a static object is changed.

In the ChangeDetection 2014 dataset, totally 12 videos selected from five challenging categories (e.g., Baseline, Dynamic Background, Camera Jitter, Intermittent Object Motion, Bad Weather) are used for detailed analysis through visual and numerical results. The remaining samples in other challenging categories (e.g., Low Framerate, Night Videos, PTZ, Shadow, Thermal, and Turbulence) are reported in overall performance evaluation.

- **Baseline**: All sequences in this category including *highway*, *office*, *pedestrians*, and *PETS2006* present a mixture of background motion, camera vibration, intermittent object motion, and shadow at a middle-level challenge. These videos are mainly used for evaluating background subtraction approaches as reference.
- **Dynamic Background**: Two sequences *canoe* and *overpass* are chosen to benchmark the method against the effect of strong background motions, such as boats on sparkle water and trees shaken by the wind.
- **Camera Jitter**: This category consists of videos captured by vibrational cameras in outdoor environment with varying jitter magnitude. Two videos *badminton* and *traffic* are selected for benchmark.
- **Intermittent Object Motion**: This category presents a common realistic scenario, known for ‘‘ghosting’’ artifact in motion detection, that means, some objects move, then stop for a while, and suddenly move again. Sometimes objects are abandoned for a long time. Two sequences *sofa* and *parking* are typical samples for such kind of challenge.
- **Bad Weather**: The poor winter weather condition with thick fog and snow storm is presented in *blizzard* and

TABLE 1. The information summarization of videos in CDnet2014 dataset used for performance evaluation.

Category/Dataset	Samples	No. frames	Img. resolution
Wallflower	camouflage	293	160 × 120
	Bootstrapping	3054	160 × 120
	TimeOfDay	5889	160 × 120
	LightSwitch	2714	160 × 120
	WavingTrees	286	160 × 120
	MovedObject	1745	160 × 120
	Fore.Aperture	2113	160 × 120
Baseline	highway	1700	320 × 240
	office	2050	360 × 240
	pedestrians	1099	360 × 240
	PETS2006	120	720 × 576
Dynamic Background	canoe	1189	320 × 240
	overpass	3000	320 × 240
Camera Jitter	badminton	1150	720 × 480
	traffic	1570	320 × 240
Intermit. Object Motion	sofa	2750	320 × 240
	parking	2500	320 × 240
Bad Weather	blizzard	7000	720 × 480
	skating	3900	540 × 360

skating sequences. This category is suitable to verify the noise reduction ability of the proposed LSD.

The information of all videos with the number of frames and image resolution is summarized in Table 1. Totally, seven common quantitative metrics, i.e. Recall (Re), Specificity (Sp), False Positive Rate (FPR), False Negative Rate (FNR), Percentage of Wrong Classifications (PWC), F-Measure (F1), and Precision (Pre) [43] are used to benchmark the foreground detection performance, of which corresponding formulations are given as below

$$Re = TP\big/(TP + FN) \tag{20}$$

$$Sp = TN\big/(TN + FP) \tag{21}$$

$$FPR = FP\big/(TN + FP) \tag{22}$$

$$FNR = FN\big/(TP + FN) \tag{23}$$

$$PWC = 100 \times \frac{FN + FP}{(TP + FN + FP + TN)} \tag{24}$$

$$Pre = TP\big/(TP + FP) \tag{25}$$

$$F1 = (2 \times Re \times Pre)\big/(Re + Pre) \tag{26}$$

where TP, TN, FP, and FN refer to as the numbers of true positives, true negatives, false positives, and false negatives, respectively. Four metrics Re, Sp, Pre, and F1 should be higher while remaining metrics are required to be smaller for a better performance. In our proposed LSD approach, the default values of hyper-parameters are set with $N = 50$, $m = 20$ in the background model initialization, and $h_0 = 0.15$, $\tau_{\mu\_1} = 20$, $\tau_{\sigma\_1} = 20$ in the parameter estimation.

Two experiments are performed in this section to benchmark the performance of the proposed LSD method and further to compare with state-of-the-art background subtraction approaches as follows

- In the first experiment, we evaluate and discuss the influence of hyper-parameters on the foreground segmentation performance.
- Secondly, we benchmark the proposed background subtraction method on seven Wallflower and 12 CDnet2014 sequences, in which the visual result and segmentation accuracy are presented with detailed analysis and discussion.
- Finally, the proposed method is compared with existing background subtraction approaches, where the overall performance of the whole dataset is reported. Additionally, computational complexity is assessed in overall.

B. RESULTS AND DISCUSSION

The first experiment evaluate the influence of hyper-parameters on the foreground detection accuracy. The overall F-measure scores of 12 CDnet2014 videos are graphically plotted in Fig. 5. In the background model initialization, there are the buffer size N and the number of randomly selected frames m that should be analyzed. Particularly, we vary N and m values in the range of [10, 300] and [5, 50], respectively. When increasing the size of buffer from 10 to 30, the overall accuracy is improved significantly. However, the increment is tiny for the buffer size larger than 50 while the memory is still in growing. Compared with the buffer size N, the accuracy improvement when increasing m value is less impressive. It can be seen that the utilization of more frames from a large buffer is capable of estimating a more clean background and reducing challenges of intermittent object motion. According to the overall F-measure scores, $30 \leq N \leq 75$ and $10 \leq m \leq 50$ are recommended for good trade-off performance. Due to the close relation of two lower thresholds of mean τ_{μ_1} and standard deviation τ_{σ_1} used in parameter estimation, we investigate their effect with a same value in the range [5, 100]. Smaller or larger thresholds yield worse results of foreground segmentation (i.e., more background pixels are misclassified to the foreground class in the

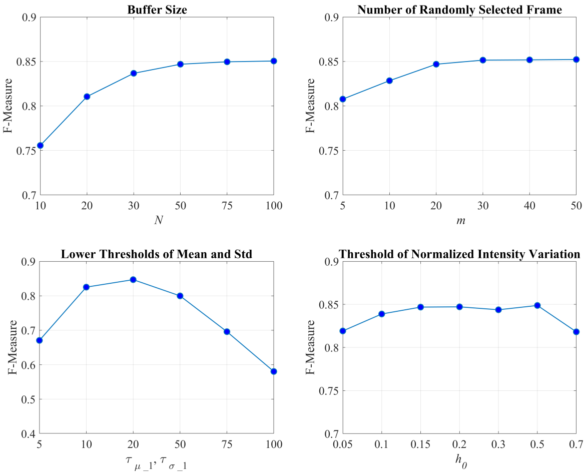

FIGURE 5. The overall F-measure scores of CDnet2014 videos for the influence evaluation of hyper-parameters: buffer size (top-left), number of randomly selected frame (top-right), lower thresholds of mean and standard deviation (bottom-left), and threshold of normalized intensity variation (bottom-right).

TABLE 2. Foreground detection accuracy of the proposed LSD method on the Wallflower and CDnet20014 datasets.

Sequences	Abbr.	Re	Sp	FPR	FNR	PWC	Pre	F1
				Wallflower				
Camouflage	CAM	0.8984	0.9891	0.0109	0.1016	6.0052	0.9898	0.9419
Bootstrapping	BOT	0.8133	0.9565	0.0435	0.1867	6.4271	0.7603	0.7859
TimeOfDay	TOD	0.7768	0.9968	0.0032	0.2232	1.9688	0.9514	0.8553
LightSwitch	LSW	0.9732	0.9811	0.0189	0.0268	2.0208	0.9106	0.9409
WavingTrees	WTE	0.9755	0.9559	0.0441	0.0245	3.8073	0.9071	0.9401
ForegroundAperture	FAP	0.7203	0.9600	0.0400	0.2797	10.1771	0.8621	0.7848
MovedObject	MOB	ND	1.0000	0.0000	ND	0.0000	ND	ND
Average		**0.8596**	**0.9732**	**0.0268**	**0.1404**	**5.0677**	**0.8969**	**0.8748**
				CDnet2014				
highway	HIG	0.9396	0.9997	0.0003	0.0604	0.0882	0.9697	0.9544
office	OFF	0.9490	0.9984	0.0016	0.0510	0.4992	0.9780	0.9633
pedestrians	PED	0.9020	0.9997	0.0003	0.0980	0.1283	0.9653	0.9325
PETS2006	PET	0.8378	0.9979	0.0021	0.1622	0.4132	0.8432	0.8405
canoe	CAN	0.7786	0.9988	0.0012	0.2214	0.8972	0.9606	0.8601
overpass	OVE	0.8206	0.9986	0.0014	0.1794	0.3763	0.8900	0.8539
badminton	BAD	0.6639	0.9923	0.0077	0.3361	1.8981	0.7533	0.7057
traffic	TRA	0.7235	0.9800	0.0200	0.2765	3.5990	0.7057	0.7145
sofa	SOF	0.7360	0.9964	0.0036	0.2640	1.4950	0.9038	0.8113
parking	PAR	0.7963	0.9737	0.0263	0.2037	4.0000	0.7175	0.7548
blizzard	BLI	0.7514	0.9998	0.0002	0.2486	0.3132	0.9740	0.8483
skating	SKA	0.8757	0.9991	0.0009	0.1243	0.5781	0.9764	0.9233
Average		**0.8145**	**0.9945**	**0.0055**	**0.1855**	**1.1905**	**0.8865**	**0.8469**

case of small lower thresholds and more moving object pixels are undetected if lower thresholds are set too large). From the result in Fig. 5, our method achieves the best accuracy at $\tau_{\mu\_1} = 20$ and $\tau_{\sigma\_1} = 20$. Compare with τ_{μ_1} and τ_{σ_1}, the threshold of normalized intensity variation is less important, that means, the overall accuracy changes insignificantly when varying the h_0 value from 0.1 to 0.5. However, due to used for estimating the rough threshold τ_ρ, a large value of h_0 may conduct a poor preliminary foreground mask which is then refined by an intensity pattern comparison algorithm. Consequently, the reasonable range of h_0 is $0.1 \leq N \leq 0.5$.

In the second experiment, the quantitative results of the Wallflower and CDnet2014 datasets are reported in Table 2 for detailed analysis. For the Wallflower dataset, the proposed method segments the foregrounds of *Camouflage*, *TimeOfDay*, and *WavingTrees* accurately with F1 results over 0.94. Besides including moving objects in a scene at the beginning of a sequence, *Bootstrapping* also represents a more complicated mixture of shadows and light reflection, that conducts foreground misclassification (i.e., more background pixels are faultily classified to the foreground class). In the case of globally gradual change of illumination depicted in *TimeOfDay*, our LSD method successfully detects a moving object, however, the homogeneity of color and illumination between the object and the background causes faulty segmentation in the foreground (that means, some pixels of the moving object are falsely segmented to the background class, that leads to high FNR metric). The visual segmentation results of Wallflower samples are presented in Fig. 6. For the CDnet2014 dataset, the proposed LSD precisely detects moving objects in middle-level challenging videos, including *highway*, *office*, and *pedestrians* of Baseline with average F1 over 0.95. Due to a quite complicated indoor environment with shadows, luminance refection, abandoned objects, and low brightness of *PETS2006*, unsuccessful object detection conducts high FNR metric. With the strongly unstable background represented in *canoe* and *overpass*, the method segments foreground fairly well by the local intensity pattern comparison algorithm in foreground refinement process. Detecting moving objects in a vibration scenario is really challenging, especially with various difference magnitude of camera jitter. From the numerical result, it can be seen that a considerable amount of pixels is incorrectly classified in *badminton* and *traffic*, e.g., background pixels are segmented as foreground pixels in *traffic* of high FPR metric and vice versa in *badminton* of high FNR metric. Intermittent object motion is one of the most popular and challenging scenarios in realistic condition because abandoned objects staying for a while can be absorbed as a part of background. Therefore, foreground segmentation sometimes fails without an adaptive updating scheme for background model. The LSD method with a dual-mode updating scheme is proficient against the intermittent object motion challenge by a quite impressive accuracy of both *sofa* and *parking*. Despite strongly influenced by incessantly falling snow, the foreground detection performance of *blizzard* and *skating* is truly remarkable. This result proves the effectiveness of combining rough detection and refinement process. Some visual segmentation results of CDnet2014 sequences performed by our proposed LSD method are shown in Fig. 7 (the last column). Additionally, the average Re, Pre, and F1 results of our proposed LSD method on several challenge categories of the CDnet2014 dataset are reported in Table 3. Besides Baseline, Bad Weather, and Dynamic Background categories, LSD segments moving objects in a scene quite precisely with Shadow, Thermal and Turbulence samples. With other specific background scenarios, LSD can successfully detects moving objects, however, the pixel-wise accuracy is unremarkable.

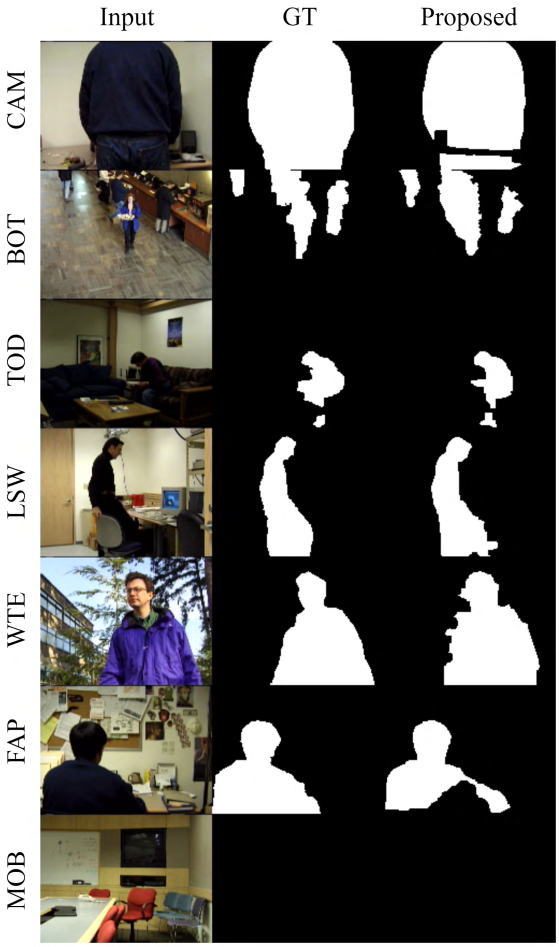

FIGURE 6. The visual segmentation results of the proposed LSD method and state-of-the-art approaches on the Wallflower dataset.

TABLE 3. Overall performance with the average of Re, Pre, and F1 metrics on several challenge categories the CDnet2014 dataset.

Category	Re	Pre	F1
Bad Weather	0.7652	0.8256	0.8004
Baseline	0.9071	0.9391	0.9226
Camera Jitter	0.6482	0.5852	0.5892
Dynamic Background	0.7890	0.7287	0.7293
Intermittent Object Motion	0.5412	0.6652	0.5748
Low Framerate	0.4026	0.7162	0.4932
Night Videos	0.8266	0.4128	0.5294
PTZ	0.5128	0.4772	0.4708
Shadow	0.8296	0.8068	0.7936
Thermal	0.7072	0.8514	0.7384
Turbulence	0.6884	0.7372	0.6846

In the last experiment, we compare our proposed LSD background subtraction method with several state-of-the-art approaches on the Wallflower dataset (i.e., Gaussian Mixture Model (GMM) [12], texture-contained Gaussian Mixture Model (TGMM) [15], Gaussian mixture shadow model (GMSM) [17], memorizing GMM (MGMM) [18], piecewise memorizing GMM (P-MGMM) [20], Lightness-Red-Green-Blue (BF-LRGB) [44]) and on the CDnet2014 dataset (i.e., Gaussian Mixture Model (GMM) [12], improved Gaussian Mixture Model (EGMM) [13], Region-based Mixture of Gaussian (RMoG) [21], Kernel Density Estimation (KDE) [27], Visual Background Subtractor (ViBE) [32], Spatially Coherent Self-Organizing Background Subtraction (SC_SOBS) [36], and Graph Cut algorithm (GraphCut) [39]).

The F-Measure-based performance comparison on the Wallflower dataset is reported in Table 4, in which the proposed method outperforms others in the most of background scenarios. In the challenges of globally sudden and gradual changes of illumination illustrated in *TimeOfDay* and *LightSwitch*, our LSD method reports the highest segmentation accuracy based on the proposed dual-mode background model updating scheme. Besides the pixel-wise strategy in the short-term mode, the long-term mode, which can update the background model quickly following the frame-wise scheme, seems to be efficient against some kinds of challenge shown in the *LightSwitch* video. For three samples *Camouflage*, *WavingTrees*, and *ForegroundAperture*, BF-LRGB is better than LSD, but the higher accuracy is insignificantly. Compared with GMM and its improvements, both BF-LRGB and LSD are considerably better in all of background challenging experiments, especially with bootstrapping, dynamic background motion, sudden and gradual illumination changes. When the global illumination of a scene (e.g., indoor environment) is changed immediately, GMM-based approaches almost get the failure of moving object detection.

TABLE 4. F-Measure comparison of the proposed LSD method and the state-of-the-art approaches on Wallflower videos.

Methods	CAM	BOT	TOD	LSW	WTE	FAP
GMM [12]	0.8033	0.4915	0.5600	0.2558	0.7086	0.3788
TGMM [15]	0.8033	0.5129	0.4998	0.2560	0.7047	0.3788
GMSM [17]	0.8465	0.6092	0.1962	0.2233	0.8225	0.3245
MGMM [18]	0.7366	0.4844	0.4185	0.2640	0.7436	0.3047
P-MGMM [20]	0.8413	0.4924	0.6749	0.6992	0.6515	0.6018
BF-LRGB [44]	**0.9750**	0.7755	0.8090	0.8398	**0.9758**	**0.8303**
LSD	0.9419	**0.7859**	**0.8553**	**0.9409**	0.9401	0.7848

We provide the F-Measure results of several methods benchmarked on CDnet2014 sequences in Table 5 for the performance comparison. Besides that, the visual comparison of foreground image is further given in Fig. 7. From the numerical results in Table 5, most of methods segment foreground of *highway* and *pedestrians* in Baseline masterfully with F1 score greater than 0.90. However, some methods get false detection of *office* and *PETS2006* due to the lack of an adaptive background model updating mechanism, e.g., GraphCut is much more fragile with intermittent object motion challenge of *office*. GMM and its improved versions, such as EGMM and RMoG, outperform GraphCut, but their accuracy cannot compete with the others. According to the running average based background model update scheme, SC_SOBS delivers superior performance over the remaining methods for the Baseline scenario, however, compared with LSD, the improvement of SC_SOBS is insignificant. In parallel, the Baseline segmentation results of SC_SOBS and LSD in Fig. 7 also approve their preeminence compared

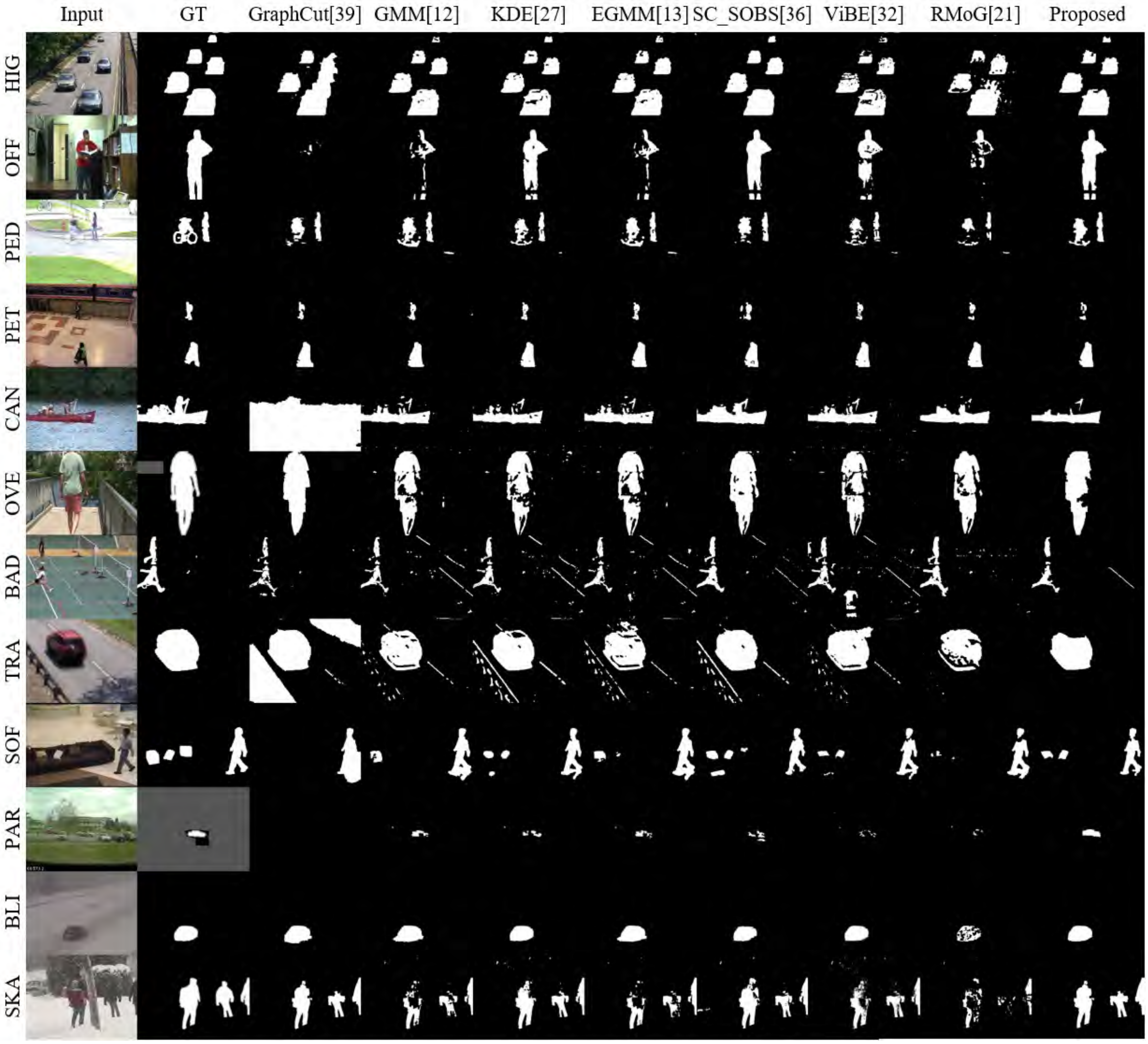

FIGURE 7. The visual segmentation results of the proposed LSD method and state-of-the-art approaches on the CDnet2014 dataset.

TABLE 5. F-Measure comparison of the proposed LSD method and the state-of-the-art approaches on the CDnet2014 dataset.

Methods	HIG	OFF	PED	PET	CAN	OVE	BAD	TRA	SOF	PAR	BLI	SKA
GraphCut [39]	0.9033	0.3513	0.9259	0.6784	0.1194	0.8352	0.6418	0.3005	0.5578	0.2077	**0.8621**	0.9218
GMM [12]	0.9240	0.5919	0.9536	0.8286	0.8817	0.8719	0.6912	0.6636	0.6449	0.7494	0.7383	0.8778
KDE [27]	0.9353	0.9355	0.9572	0.8089	0.8822	0.8250	0.7233	0.5846	0.6466	0.3731	0.7720	0.9081
EGMM [13]	0.9038	0.6564	**0.9598**	0.8327	0.8851	0.8673	0.6669	0.6137	0.6524	0.7179	0.7585	0.8644
SC_SOBS [36]	0.9455	**0.9703**	0.9492	**0.8684**	**0.9525**	0.8838	**0.8818**	0.7062	0.6403	0.4034	0.5997	0.8956
ViBE [32]	0.8546	0.8170	0.8077	0.7065	0.7794	0.7464	0.6442	0.7384	0.5461	0.3877	0.7185	0.7616
RMoG [21]	0.8650	0.5864	0.9374	0.7505	0.9357	**0.9011**	0.7907	**0.7482**	0.5459	0.4585	0.5814	0.7921
LSD	**0.9544**	0.9633	0.9325	0.8405	0.8601	0.8539	0.7057	0.7145	**0.8113**	**0.7548**	0.8483	**0.9233**

with others. For the dynamic background motion challenge, both SC_SOBS and RMoG are the leading methods with outstanding accuracy. GraphCut conducts a poor performance with *canoe*, wherein the water surface is recognized as the foreground due to light reflection. It can be seen that the F1-based results of remaining methods and LSD are almost equivalent. For the sequences in Camera Jitter scenario, by modeling background as a flat neuronal map wherein neurons are constrained by spatial weights, SC_SOBS has the potential of dealing with vibration issue. As mentioned before, due to not being an expert in handling vibration issue, our LSD presents moderate results of *badminton* and *traffic*, at which it reaches the third place behind SC_SOBS and RMoG. However, LSD is more efficient than other methods of eliminating noise caused by camera jitter since the foreground images of LSD are cleaner than those of competitors in Fig. 7. With respect to the intermittent object motion challenge, LSD significantly outperforms the others based on the advantage of dual-mode background updating scheme. For more explanations, the buffer is updated following the long-term mode for the scenario of abandoned objects in *sofa* and *parking*. Remaining methods usually assimilate abandoned objects to the background through their updating progress, therefore, most of them cannot detect or segment those things precisely. Indeed, after abandoned objects are absorbed to the background, there are two scenarios: (i) the abandoned objects cannot be segmented to the foreground during its stay in a scene and (ii) the abandoned objects are still detected as the foreground after moving out of a scene. Notably, the second scenario is commonly known as the ''ghosting'' artifact in motion detection. For the last challenge, where moving object detection is strongly affected by bad weather conditions, concretely, dense fog and snow storm in both benchmark videos. GraphCut and LSD share the leading position of F1-based performance competition of *blizzard* and *skating*, respectively. The remarkable accuracy of GraphCut and LSD is further evidenced by the foreground images in Fig. 7. Compared to GraphCut and LSD, SC_SOBS and RMoG produce inferior performance drastically with F1 of approximately 0.6 of *blizzard*. With *skating*, the most of experimental methods accomplish the segmentation task well with F1 scores over than 0.86, except ViBE and RMoG.

The overall averaged results of Recall, Precision, and F-Measure of the whole Wallflower and CDnet2014 datasets are reported in Table 6. In general, the proposed LSD method is remarkably better than other methods on the experiments of Wallflower sequences (e.g., higher than GMM-based approaches by 0.16, 0.39, and 0.34 of Re, Pre, and F1 metrics, respectively). Compared with BF-LRGB, a very recently state-of-the-art foreground detection approach, LSD yields the higher performance of Re and F1, but worse result of Pre with a minor margin. For the CDnet2014 dataset, ViBE shows the worst overall performance. In spite of the fact that ViBE is better than KDE, GMM-based for the Pre metric, it cannot be competitive in terms of F1 metric due to its lowest Re score (i.e., high *FN* indicates that a large number of background pixels are misclassified to the foreground class). Despite achieving the high overall Re of 0.74, KDE finalizes the experiment with medium Pre and F1 results. Three methods consisted of GraphCut, GMM, EGMM, and RMoG segment foreground equally well, but not impressive. SC_SOBS shows the robustness against some challenges based on results reported in Table 5 with the best overall Re score of 0.76, however, the overall performance of F1 is just acceptable due to its fragility with such some challenges as the intermittent object motion and bad weather (particularly, detection failure sometimes occurs if a moving object stops for a while, and then is absorbed to the background). Our LSD proposed method, which involves the local intensity pattern comparison algorithm for detecting foreground and the dual-mode scheme for updating background model, achieves the best performance with greatest overall Pre and F1 scores. Additionally, it can be recognized that LSD is more stable than other benchmark approaches, that means, the method durably produces results in leading portion of such accuracy competition.

TABLE 6. Overall performance comparison with the average of Re, Pre, and F1 metrics on the CDnet2014 dataset.

Methods	Re	Pre	F1
Wallflower			
GMM [12]	0.7911	0.4338	0.5330
TGMM [15]	0.8044	0.4197	0.5259
GMSM [17]	0.6110	0.4623	0.5037
MGMM [18]	0.7176	0.4226	0.4920
P-MGMM [20]	0.5965	0.7690	0.6602
BF-LRGB [44]	0.8484	**0.9026**	0.8676
LSD	**0.8596**	0.8969	**0.8748**
CDnet2014			
GraphCut [39]	0.6297	0.6666	0.5684
GMM [12]	0.6846	0.6025	0.5707
KDE [27]	0.7375	0.5811	0.5688
EGMM [13]	0.6604	0.5973	0.5566
SC_SOBS [36]	**0.7621**	0.6091	0.5961
ViBE [32]	0.3072	0.6322	0.4134
RMoG [21]	0.5940	0.6965	0.5735
LSD	0.6925	**0.7041**	**0.6660**

TABLE 7. Overall processing speed comparison on the CDnet2014 dataset.

Methods	Environment	Processing time
GraphCut [39]	C++	~7fps for 320 × 240
GMM [12]	C++	~21fps for 720 × 480
KDE [27]	C++	~9fps for 720 × 480
EGMM [13]	C++	~49fps for 720 × 480
SC_SOBS [36]	C	~4fps for 720 × 576
ViBE [32]	C++	~39fps for 720 × 480
RMoG [21]	N/A	N/A
LSD	C++	~31fps for 720 × 480

Finally, evaluating and measuring the computational complexity of a background subtraction method in terms of processing time metric is important for real-time video-based surveillance systems. In this experiment, we summarize the processing speed results of all comparing methods (only evaluated on the CDnet2014 dataset) in Table 7. With a notebook equipped with Core i7 2.7GHz and 8GB RAM, the processing speed of our proposed LSD method reaches ~31fps for 720 × 480 video in C++ environment. It should be noted that other methods are benchmarked on different machine configurations, for example, EGMM achieves 49fps with Core i7 3.4Ghz. However, it can be observed that the proposed LSD is pretty competitive with the state-of-the-art background subtraction approaches in terms of computational complexity.

V. CONCLUSION

In this research, we have proposed a novel background subtraction method called LSD for detecting moving objects precisely in video-based surveillance systems. The method consists of background model initialization, intensity pattern comparison based foreground detection, dual-mode background update, and parameter estimation stage. Regarding foreground detection, besides a rough process, we principally recommend an efficient algorithm for a refinement process based on analyzing the homogeneity of local intensity patterns of the input frame and the background model to classify a pixel into either foreground or background class. From these two detection processes, the foreground is segmented more precisely with low cost of computation. We have also proposed a dual-mode background model updating scheme for short- and long-term motion detection scenarios. The short-term mode is operated following a pixel-wise updating manner while the long-term mode updates the model in a frame-wise manner. The dual-mode updating scheme profits the background model to be more robust against many foreground detection challenges, especially intermittent object motion. Parameters used for foreground detection of the next frame is estimated based on the locally and globally statistical information of segmented foreground. Different from existing approaches where there is only one global threshold for all pixels classification, our proposed LSD calculates local thresholds for every particular pixel which subsequently yields higher classification accuracy.

In the experiment, the proposed LSD method is benchmarked on several practical motion detection challenges of the Wallflower and CDnet2014 datasets. Outstanding experimental results have demonstrated that LSD outperforms several other existing background subtraction approaches (i.e., GMM, KDE, EGMM, ViBE, SC_SOBS, GraphCut, RMoG, and BF-LRGB) in terms of segmentation accuracy. Thanks to maintaining a reasonable processing speed over 31fps, our method is practically suitable to indoor and outdoor video-based surveillance systems as CCTV for many realistic applications, such as intelligent traffic monitoring, abnormal event detection, human activity analysis, and etc. In future, we would like to update the rough and refinement detection processes for the proficiency of eliminating dynamic background motions.

REFERENCES

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

[2] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Comput. Vis. Image Understand.*, vol. 122, pp. 4–21, May 2014.

[3] S. K. Choudhury, P. K. Sa, S. Bakshi, and B. Majhi, "An evaluation of background subtraction for object detection vis-a-vis mitigating challenging scenarios," *IEEE Access*, vol. 4, pp. 6133–6150, 2017.

[4] W. Kim and C. Jung, "Illumination-invariant background subtraction: Comparative review, models, and prospects," *IEEE Access*, vol. 5, pp. 8369–8384, 2017.

[5] W. Fang, T. Zhang, C. Zhao, D. B. 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.

[6] K. Roy, M. R. Arefin, F. Makhmudkhujaev, O. Chae, and J. Kim, "Background subtraction using dominant directional pattern," *IEEE Access*, vol. 6, pp. 39917–39926, 2018.

[7] W.-X. Kang, W.-Z. Lai, and X.-B. Meng, "An adaptive background reconstruction algorithm based on inertial filtering," *Optoelectron. Lett.*, vol. 5, no. 6, pp. 468–471, Nov. 2009.

[8] X. Li, M. Ng, and X. Yuan, "Median filtering-based methods for static background extraction from surveillance video," *Numer. Linear Algebra Appl.*, vol. 22, pp. 845–865, Oct. 2015.

[9] Y. Hoshen, C. Arora, Y. Poleg, and S. Peleg, "Efficient representation of distributions for background subtraction," in *Proc. 10th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Krakow, Poland, Aug. 2013, pp. 276–281.

[10] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1459–1472, Nov. 2004.

[11] K. O. De Beeck, I. Y.-H. Gu, L. Li, M. Viberg, and B. De Moor, "Region-based statistical background modeling for foreground object segmentation," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2006, pp. 3317–3320.

[12] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 1999, pp. 246–252.

[13] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, Cambridge, MA, USA, vol. 2, 2004, pp. 28–31.

[14] D.-S. Lee, "Effective Gaussian mixture learning for video background subtraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 827–832, May 2005.

[15] Y.-L. Tian, M. Lu, and A. Hampapur, "Robust and efficient foreground analysis for real-time video surveillance," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Diego, CA, USA, vol. 1, Jun. 2005, pp. 1182–1187.

[16] A. Shimada, D. Arita, and R. I. Taniguchi, "Dynamic control of adaptive mixture-of-Gaussians background model," in *Proc. IEEE Int. Conf. Video Signal Based Surveill. (AVSS)*, Nov. 2006, p. 5.

[17] N. Martel-Brisson and A. Zaccarin, "Learning and removing cast shadows through a multidistribution approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 7, pp. 1133–1146, Jul. 2007.

[18] Y. Qi and Y. Wang, "Memory-based Gaussian mixture modeling for moving object detection in indoor scene with sudden partial changes," in *Proc. IEEE 10th Int. Conf. Signal Process.*, Beijing, China, Oct. 2010, pp. 752–755.

[19] A. Elqursh and A. Elgammal, "Online moving camera background subtraction," in *Proc. 12th Eur. Conf. Comput. Vis. (ECCV)*, vol. 4. 2012, pp. 228–241.

[20] W. Zhao, X. D. Zhao, W. M. Liu, and X. L. Tang, "Long-term background memory based on Gaussian mixture model," in *Proc. Vis. Commun. Image Process. (VCIP)*, Kuching, Malaysia, Nov. 2013, pp. 1–5.

[21] 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.

[22] 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.

[23] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground–background segmentation using codebook model," *Real-Time Imag.*, vol. 11, no. 3, pp. 172–185, 2005.

[24] A. Ilyas, M. Scuturici, and S. Miguet, "Real time foreground-background segmentation using a modified codebook model," in *Proc. 6th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Sep. 2009, pp. 454–459.

[25] J.-M. Guo, C.-H. Hsia, Y.-F. Liu, M.-H. Shih, C.-H. Chang, and J.-Y. Wu, "Fast background subtraction based on a multilayer codebook model for moving object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1809–1821, Oct. 2013.

[26] C.-W. Lin, W.-J. Liao, C.-S. Chen, and Y.-P. Hung, "A spatiotemporal background extractor using a single-layer codebook model," in *Proc. 11th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2014, pp. 259–264.

[27] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in *Proc. 6th Eur. Conf. Comput. Vis.*, 2000, pp. 751–767.

[28] C. Cuevas, R. Martinez, D. Berjón, and N. Garcia, "Detection of stationary foreground objects using multiple nonparametric background-foreground models on a finite state machine," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1127–1142, Mar. 2017.

[29] A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun./Jul. 2004, pp. II-302–II-309.

[30] Z. Tang, Z. Miao, Y. Wan, and J. Li, "Automatic foreground extraction for images and videos," in *Proc. 17th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 2993–2996.

[31] D. Li, L. Xu, and E. Goodman, "A fast foreground object detection algorithm using kernel density estimation," in *Proc. IEEE 11th Int. Conf. Signal Process. (ICSP)*, Oct. 2012, pp. 703–707.

[32] O. Barnich and M. Van Droogenbroeck, "ViBE: A powerful random technique to estimate the background in video sequences," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Taipei, Taiwan, Apr. 2009, pp. 945–948.

[33] 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.

[34] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Providence, RI, USA, Jun. 2012, pp. 38–43.

[35] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1168–1177, Jul. 2008.

[36] L. Maddalena and A. Petrosino, "The SOBS algorithm: What are the limits?" in *Proc. IEEE Comp. Soc. Conf. Comp. Vis. Pattern Recognit. Workshops*, Providence, RI, USA, Jun. 2012, pp. 21–26.

[37] T. Huynh-The, O. Banos, S. Lee, B. H. Kang, E.-S. Kim, and T. Le-Tien, "NIC: A robust background extraction algorithm for foreground detection in dynamic scenes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 7, pp. 1478–1490, Jul. 2017.

[38] T. Huynh-The, S. Lee, and C.-H. Hua, "ADM-HIPaR: An efficient background subtraction approach," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Lecce, Italy, Aug./Sep. 2017, pp. 1–6.

[39] A. Miron and A. Badii, "Change detection based on graph cuts," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, London, U.K., 2015, pp. 273–276.

[40] Z. Zhong, B. Zhang, G. Lu, Y. Zhao, and Y. Xu, "An adaptive background modeling method for foreground segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 5, pp. 1109–1121, May 2017.

[41] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Kerkyra, Greece, vol. 1, Sep. 1999, pp. 255–261.

[42] 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*, Columbus, OH, USA, Jun. 2014, pp. 393–400.

[43] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection.net: A new change detection benchmark dataset," in *Proc. IEEE Comput. Soc. Conf. Comp. Vis. Pattern Recognit. Workshops*, Providence, RI, USA, Jun. 2012, pp. 1–8.

[44] J. D. Romero, M. J. Lado, and A. J. Méndez, "A background modeling and foreground detection algorithm using scaling coefficients defined with a color model called lightness-red-green-blue," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1243–1258, Mar. 2018.

THIEN HUYNH-THE (S'15–M'19) received the B.S. degree in electronics and telecommunication engineering and the M.Sc. degree in electronics engineering from the Ho Chi Minh City University of Technology and Education, Vietnam, in 2011 and 2013, respectively, and the Ph.D. degree in computer science and engineering from Kyung Hee University (KHU), South Korea, in 2018. He was a recipient of the Superior Thesis Prize by KHU.

He is currently a Postdoctoral Research Fellow with the ICT Convergence Research Center, Kumoh National Institute of Technology, South Korea. His current research interests include digital image processing, computer vision, and machine learning.

CAM-HAO HUA received the B.S. degree in electrical and electronic engineering from Bach Khoa University, Ho Chi Minh, Vietnam, in 2016. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Kyung Hee University, Gyeonggi, South Korea. His research interest include image processing, computer vision, and deep learning.

NGUYEN ANH TU received the B.S. degree in electrical and electronics engineering from the Ho Chi Minh City University of Technology, Vietnam, in 2010, and the Ph.D. degree in computer science and engineering from Kyung Hee University, South Korea, in 2018.

He is currently a Postdoctoral Research Fellow in data and knowledge engineering with the Department of Computer Science and Engineering, Kyung Hee University. His current research interests include computer vision, machine learning, image retrieval, and big data processing.

DONG-SEONG KIM received the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2003, where he was a full-time Researcher with ERC-ACI, Seoul National University, from 1994 to 2003. From 2003 to 2005, he was a Postdoctoral Researcher with the Wireless Network Laboratory, School of Electrical and Computer Engineering, Cornell University, NY, USA. From 2007 to 2009, he was a Visiting Professor with the Department of Computer Science, University of California at Davis, Davis, CA, USA. He is currently the Director of the KIT Convergence Research Institute and the ICT Convergence Research Center (ITRC and NRF Advanced Research Center Program) supported by the Korean Government, Kumoh National Institute of Technology. His current research interests include real-time IoT and smart platform, industrial wireless control networks, and networked embedded systems. He is a Senior Member of IEEE and ACM.