

Received June 10, 2020, accepted July 6, 2020, date of publication July 15, 2020, date of current version July 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009104

Background Subtraction Using an Adaptive Local Median Texture Feature in Illumination Changes Urban Traffic Scenes

YUNSHENG ZHANG^{1,2,3}, WEIBO ZHENG², KAIJUN LENG¹, AND HAO LI³

¹Research Center of Hubei Logistics Development, Hubei University of Economics, Wuhan 430205, China

²College of Management, Xi'an Jiaotong University, Xi'an 710049, China

³Xi'an Key Laboratory of IoT Application Engineering, College of Information Engineering, Xian University, Xian 710065, China

Corresponding authors: Kaijun Leng (lengkaijun@hbue.edu.cn) and Hao Li (lihao82@126.com)


This work was supported in part by the China Postdoctoral Science Foundation under Grant 2018M631169 and Shaanxi Province Postdoctoral Science Foundation under Grant 2018BSHEDZZ83, in part by the Ministry of Education of China of Humanities and Social Sciences for Young Researchers Project under Grant 18YJCZH253, in part by the Hubei Provincial Natural Science Foundation under Grant 2019CFB580 and in part by the Social science fund of Shaanxi Province under Grant 2019S048.

ABSTRACT Background subtraction is commonly employed in foreground object detection in urban traffic scenes. Most of the current color or texture feature-based background subtraction models are easily contaminated by sudden and gradual illumination variations in urban traffic scenes. To resolve this deficiency, an adaptive local median texture feature, which extracts the adaptive distance threshold employing the median information in a predefined local region of a pixel and Weber's law, is introduced. In addition, a sample consensus-based model that evolved from portable visual background extractor is proposed using an adaptive local median texture feature. Then, the foreground is labeled by comparing the input video frames feature with the model. Moreover, to adapt the dynamic background, the random update scheme is used to update the model. Extensive experimental results on the public Change Detection data set of 2014 (CDnet2014) and the real-world urban traffic videos demonstrate that our background subtraction method is superior to the other state-of-the-art texture-feature-based methods. The qualitative and quantitative results show the encouraging efficiency of the proposed technique to deal with sudden and gradual illumination variations in real-world urban traffic scenes.

INDEX TERMS Background modeling, illumination variations, local median texture feature, urban traffic scenes.

I. INTRODUCTION

Accurately and reliably segmenting foreground objects from the increasing number of video-based urban traffic scenes is the first key step for surveillance applications, developing intelligent transportation systems (ITS) and high-level vision understanding. Bottom-up that first detects and classifies parts of an object using features such as Histogram of Oriented Gradients (HOG), Haar-like features and Local Binary Pattern (LBP) and top-down which pixels are grouped into objects early during the processing using background subtraction method are typically used for foreground objects recognition [1]. In recent years, the background subtraction method, which is the comparison of observed image

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

sequences with the constructed background image, has drawn increasing research attention and is widely used to segment foreground objects. A wide variety of background subtraction methods have been proposed in [2], and most of these methods involve three basic aspects: background modeling, difference comparison and image labeling. Many background subtraction methods have been introduced for foreground segmenting. However, these models still face various challenges: bad weather, dynamic textures in the background, lighting variations and so on. In particular, various illumination changes, which frequently occur in urban traffic scenes, increase the difficulty of vehicles detection. For example, as the sun moves across the sky it provides a light source that varies during the day, which may lead to the incorrect foreground detection under gradual illumination changes scenes; frequent sudden changes, such as the

headlights and taillights of passing vehicles and sudden illumination changed from electronic billboard screens will lead to various false-positives. Many background models treat the illumination change problem by employing model updating, spatio-temporal intensity information or illumination-invariant features [3].

Recently, parametric models have been increasingly introduced to deal with illumination change due to their ability to update background models using an adaptive learning rate. To address the limitations of the single Gaussian model [4], Stauffer and Grimson proposed a more advanced multi-distribution Gaussian Mixture Model (GMM) [5] that can tackle multivariate real-world and illumination changing situations and leads to reliable results using learning rate online updating. GMM has obtained widespread popularity and inspired many researchers to improve the update scheme and parameters of GMM under varying illumination environments. For instance, Zivkovic [6] introduced the recursive updating scheme, which is effective to cope with illumination changing situations to improve GMM. Similarly, to effectively deal with illumination changes, White and Shah [7] exploited the particle swarm optimization to obtain the parameters of GMM. The optimal value of GMM and its improvements cannot be set exactly because illumination changes are hardly estimated in complex outdoor environments. Then, two or multiple GMM-based approaches were introduced to manage illumination changes. A two-layer GMM to represent the background at different lighting conditions was introduced, and a joint posterior function of background state and segmentation were simultaneously optimized using a nested two-layer optimization [8]. An illumination evaluation is used to analyze illumination changes and determine light background and dark background candidates [9]. Two background models with different adaption rates were utilized to address the updating of the model in sudden illumination changes [10]. Recently, an illumination change model, a chromaticity difference model and a brightness ratio model were developed to deal with fast illumination changes in a visual surveillance system [11]. However, the evaluation of illumination changes and selection of the corresponding learning rate or model is difficult in the multiple model method.

In a low-rank based model, the background model can be represented as a low-rank matrix while foreground objects are detected as outliers to handle illumination changes in successive frames. For example, robust principal component analysis (RPCA) is applied to the background model [12], where varying illumination can be successfully approximated using the corresponding low-rank subspace and moving objects consist of correlated sparse components. To address the computation cost issue, a fast background/foreground separation algorithm where the low-rank constraint is solved using the matrix factorization method was proposed in [13]. To overcome the background initialization challenge, a spatiotemporal low-rank modeling method was developed to estimate a robust background model by dynamic video clips

in [14]. A novel background subtraction method with multiscale structured low-rank and sparse factorization, which explore the structured smoothness with both appearance consistency and spatial compactness, was introduced in [15]. The low-rank based model has the remarkable improvement employing the implicit integration of illumination changes into low-rank space. However, low-rank based approaches need additional memory space for batch-based processing. Although improved RPCA with optimization techniques can improve the processing speed, it is also time-consuming.

Some researchers have attempted to apply a deep neural network (DNN) to maintain the background model under illumination changes. A triple multitask generative adversarial network [16], which models the semantic relationship between dark and bright images and fuses features of images with varying illumination, was proposed to extract the foreground in continuously varying illumination sequences. A novel background subtraction that uses a deep convolutional neural network (CNN) to handle various video scenes was introduced in [17]. To grapple with missing temporal information in complex situations, a 3D convolutional neural (3D-CNN) with long short-term memory (LSTM) is proposed using fully convolutional networking, 3D transpose convolution, and residual feature flows in [18]. To handle illumination variations and dynamic backgrounds, encoder-decoder fully convolutional neural network architecture is applied and trained to automatically fuse different background methods in [19]. Many researchers are now starting to pay attention to deep learning-based models for illumination changes.

In contrast to parametric models, low-rank based models and deep neural network-based models, some researchers have introduced an illumination-invariant feature to represent the underlying structure of the model under illumination changing environments. Kim *et al.* [20] introduced a new illumination-invariant feature (IIF) using the coefficients of the singular value decomposition (SVD) and provided that the corresponding feature is useful for modeling the background under diverse lighting conditions without any preprocessing tasks. Then, Kim proposed edges of residuals features to detect moving objects under varying illumination conditions in [21]. Local-texture feature matching with a pattern or multiple pixels is more accurate than pixel-pixel matching. The local binary pattern (LBP) [22], which is one of the more widely used and computationally simplicity descriptors to overcome illumination changes, checks the relative difference between spatially neighboring pixels to construct an illumination-invariant feature background model. Compared with Histogram of Oriented Gradients (HOG) and Haar-like features, LBP features perform better with a higher detection rate in traffic scenes [41], and its spatiotemporal information can be used to deal with illumination changes. To resolve frequent illumination changes in outdoor scenes, the LBP are computed based on the frame difference result to efficiently smooth unexpected noise and they preserve the boundaries of the moving objects using an edge-aware filtering technique

in [23]. Goyal and Singha [24] improved an LBP-based background to manage illumination variations using an adaptive learning rate. Zeng *et al.* [25] take advantage of local texture features, which are represented by an extended scale invariant LBP and color intensities, to achieve good tolerance against illumination variations. The LBP and its improved have indicated to be powerful local image descriptors, but the LBP operator is not robust to local image noise when neighboring pixels are similar. To eliminate most of the effects of changing illumination and noise, Tan and Triggs [26] introduced the local ternary pattern (LTP) operator, which is more discriminant and less sensitive to noise in uniform regions. Then, the scale invariant local ternary patterns (SILTP) [27] and 3D local spatiotemporal ternary patterns (3D-LSStTP) [28] were applied to the background model. An SILTP that improved the performance of the LTP was combined with the pattern kernel density estimation technique to model the probability distribution of local patterns under varying illumination situations. Inspired by human cognitive vision, 3D-LSStTP was proposed for employing moving object detection and collecting multi-directional spatio-temporal information from three consecutive frames, and then 3D-LSStTP-based model were constructed by using texture and color features in varying illumination scenes. Chan K.L. [29] proposed a perceptual-based LTP feature to adapt to abrupt illumination changes and background motions. A new local binary similarity pattern (LBSP) [30] feature was proposed to handle the sensitivity of illumination changes. Our previous work [31] introduced an adaptive local texture feature (ALTF) to deal with sudden and gradual illumination changes using Weber’s law and a sample consensus scheme. To address the illumination variation issue, a new spatial feature descriptor, which extracts the prominent directional information in the local neighborhood of a pixel, was introduced in [32]. The LBP and its improved descriptors are computational simplicity and can effectively manage illumination changes. However, all the LBP-based binary feature descriptors (e.g., LBP, LTP, LBSP, SILTP and ALTF) are computed based on the intensity value of the center pixel region, in which the value of the center may be a noise point in complex urban traffic scenes. Moreover, some texture features require parameter control (e.g. SILTP and LBSP), which may fail to generate reliable code in local region.

Due to the power of illumination-invariant features, many researchers are now starting to give considerable attention to it, and the development of a reliable background model with illumination-invariant features remains hot issues in complex urban traffic scenes. For example, the LBP, LTP, SILTP, LBSP, XCS-LBP, spatio-temporal local binary patterns (STLBP) and center symmetric spatio-temporal local binary patterns (CS-STLTP) [33] were proposed to deal with illumination changes and local noise. Numerous of illumination-invariant features are computed based on the center pixel, which may be an outlines point, and the performance of these features or approaches are

unsatisfactory for resolving sudden and gradual illumination variations in urban traffic scenes. To resolve those problems and increase the robustness of the feature descriptor, inspired by median LTP [34] and the perception-inspired confidence interval [35], we introduced a novel adaptive local median texture (ALMT) feature that utilizes the local median intensity value to replace the center value and an adaptive parameter to replace the fixed threshold to deal with illumination variations in urban traffic scenes. To efficiently address the deficiencies of BS-based methods that are easily contaminated by sudden and gradual illumination changes in urban traffic scenes, we combined the illumination-invariant features of the adaptive local median texture (ALMT) feature and the non-parametric sample consensus technique to introduce an adaptive local median texture feature background model (ALMTFM) to manage illumination changes.

The overall proposed adaptive local median texture feature background model (ALMTFM) is depicted in Fig. 1. First, the method obtains the median in a predefined $N \times N$ local region and extracts the adaptive parameter threshold using Weber’s law for the adaptive local median texture feature (ALMT) representation. Second, ALMTFM is derived from the calculated ALMT feature and nonparametric sample consensus scheme [36]. Then, the background model samples and the coming video images are compared using the ALMT feature to label the foreground pixels, and the model is updated employing a random update strategy. Finally, the results obtained from experiments on real world urban traffic scenes and Change Detection data sets from 2014 show that the ALMTFM performs better than numerous state-of-the-art texture-based background models.

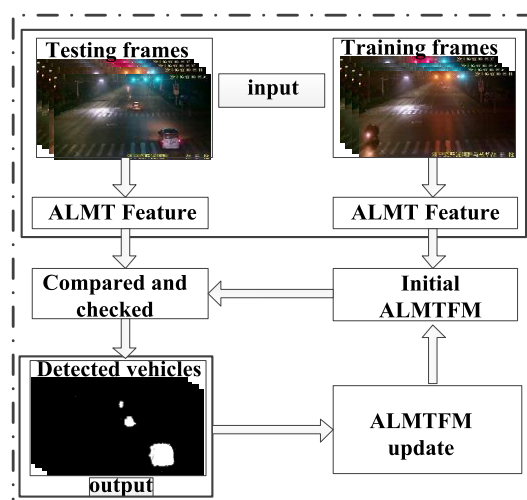


FIGURE 1. Overview of the proposed algorithm using ALMT feature.

The main contributions of this paper are summarized as follows: first a novel perceivable adaptive local median texture feature is introduced to handle illumination changes. Second, a novel model is derived from the ALMT and sample consensus technique. Finally, experimental results on

real-world urban traffic videos and the Change Detection dataset of 2014 obtain excellent performance in detection of vehicles in illumination changing urban traffic scenes.

The rest of this paper is structured as follows. In the next section, the ALMT feature is explained in detail. Section 3 provides the adaptive local median texture feature background model (ALMTFM) to detect foreground objects using adaptive local median texture (ALMT) feature. The experimental results and reports of the ALMTFM compared with state-of-the-art models are presented for urban traffic environments in Section 4. Conclusion and future work are discussed in Section 5.

II. ADAPTIVE LOCAL MEDIAN TEXTURE FEATURE

A novel perception-based adaptive local median texture feature that can be employed effectively to characterize sudden and gradual illumination changes of urban traffic scenes is introduced by the perceivable distance threshold. The adaptive local median texture (ALMT) feature is computed on a predefined local block of size 5×5 pixels, and its pattern is similar to LBSP. An example block of 5×5 and the corresponding pattern to calculate the ALMT feature are shown in Fig.2, where C is the center pixel, N denotes the corresponding neighbor pixels of C in the pattern and M denotes the median of all N and C pixels in the pattern.

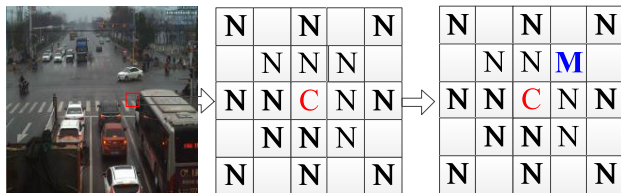


FIGURE 2. A 5×5 block pattern with center pixel C, neighbor pixels N and median pixel M are used to calculate ALMT.

Each pixel of the block pattern is compared with the median of all N and C in the pattern. Assuming $I(x)$ is a pixel of image at a given location x and the size of predefined block of $I(x)$ is $n \times n$, the adaptive local median texture feature $ALMT(x)$ operator binary value is computed according to the following equations:

$$ALMT(x) = \sum_{p=0}^{P-1} d_T(I_{x,p}, M) \cdot 2^p, \quad (1)$$

$$d_T(I_{x,p}, M) = \begin{cases} 1, & \text{if } |I_{x,p} - M| \leq T, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

$$M = \text{median} \{I_{x,p}, I(x) | p \in P\} \quad (3)$$

where $I_{x,p}$ is the intensity of the p th pixel neighbor of $I(x)$ on predefined block pattern P , M is the median of all N and C in block, and T is the adaptive threshold.

The ideal T should be small for darker situations and larger for brighter environments. According to Weber's law, which states that the initial background stimulus intensity I is

linearly proportional to the Just-Noticeable Difference ΔI , and the adaptive distance threshold T is provided to increase the robustness of the texture feature. The relationship of ΔI and I can be expressed as the following equation:

$$\Delta I / I = c, \quad (4)$$

Here, the ratio c is a constant, and ΔI is small in dark and large in bright environments. Because human visual system (HVS) perceptual characteristics are in direct proportion to background illumination, the adaptive threshold T depends on the perceptual characteristics of the median sample intensity, and T should be large for a brighter median sample and low for a darker median sample. Mapping to Weber's law in the ALMT, the median sample M is regarded as initial intensity I and T can be regarded as Just-Noticeable Difference ΔI . The difference between neighboring pixel N and median pixel M is the intensity change. Therefore, we set

$$T / M = c, \quad (5)$$

where M is the median pixel, c can be inferred using the peak signal-to-noise-ratio (PSNR) measure similar to [37] and $c = 0.11$. The relationship of the median sample and its adaptive distance threshold is as follows:

$$T = 0.11M, \quad (6)$$

Moreover, in the extremely dark or bright regions, the Weber's law may fail to precisely describe the linear relationship between perceptible value changes of HVS and the median sample. To deal with too high or too low median samples, the adaptive distance threshold is cut off at the upper bound T_u or lower bound T_l . In the urban traffic environment, the range $[T_l, T_u] = [255 \times 0.1, 255 \times 0.9]$ is practically set to satisfy the linear relationship of Weber's law. That is, the adaptive distance threshold is set according to the following equation:

$$T = c \min \{ \max [M, T_l], T_u \}, \quad (7)$$

A simple example of the process for calculating the ALMT feature is illustrated in Fig. 3, which shows the predefined block of 5×5 pixels and the corresponding pattern in the block. The median sample value M is 98, and the corresponding adaptive threshold of 98 is 10.7 using equation (6). Finally, we obtain the binary string to be 01011000110101011, and the length is 17 bits.

The comparison encoding of the adaptive local texture feature (ALTF), LBSP and adaptive local median texture (ALMT) are depicted in Fig.4 to deal with illumination change noise. The ALTF and LBSP were introduced to efficiently cope with illumination changes in previous work. (a) and (b) of Fig.4 are contrast encoding examples of ALTF, LBSP and ALMT at the same pixel location of different grayscale frames under different illumination conditions. The center pixel value of (a) is 78 at one moment, and the (b) at the same position is 224 at the moment of local illumination change. As illustrated in Fig.4 (b), the center pixel value may

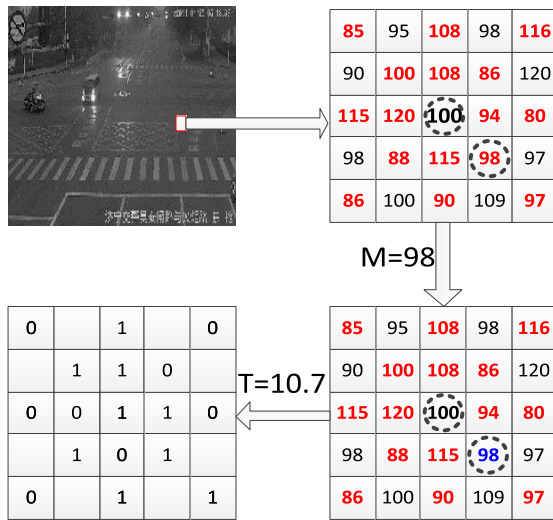


FIGURE 3. Example of ALMT feature calculating.

be a point of noise caused by the headlights of one vehicle. The result of the binary string using the ALTF and LBSP is nearly identical with the ALMT in Fig.4(a), and the ALMT is more precise. Nevertheless, the contrast encoding results are different in Fig.4(b). The LBSP operator is not robust to local image noise when noise is truly on the center of the predefined block. The ALTF and ALMT are more robust to handle local image noise, while the ALMT can improve the precise performance of the ALTF.

III. TEXTURE-BASED BACKGROUND SUBTRACTION MODELING

In this section, adaptive local median texture (ALMT) feature and sample consensus scheme are employed to construct a novel vehicle detection background subtraction model in urban traffic scenes with illumination changes adaptive local median texture feature background model (ALMTFM). The detailed description of the framework for the ALMTFM can be divided into the background model representation, background model initialization, foreground object detection and background updating.

A. BACKGROUND MODELING AND INITIALIZATION

Background subtraction models are the first step in vehicle detection, and the ideal algorithm may improve excellent performance to deal with complex environments and sudden or gradual illumination changes in urban traffic scenes. A sample consensus background approach that is pixel-based and light weight is proposed using the ALMT. For each pixel $p(x)$, the sample-based background model $B(x)$ contains an array of N recently observed ALMT features and is constructed at location x as follows:

$$B(x) = \{b_1(x), b_2(x), \dots, b_{N-1}(x), b_N(x)\}, \quad (8)$$

where $b_j(x), j = [1, N]$ is the recently observed ALMT features background samples.

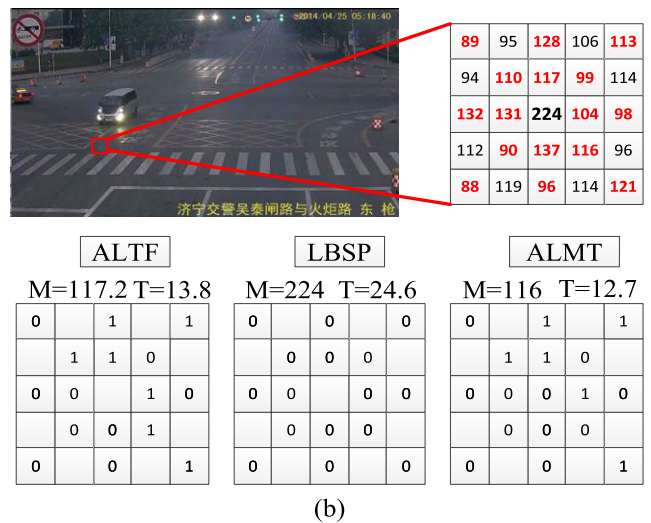
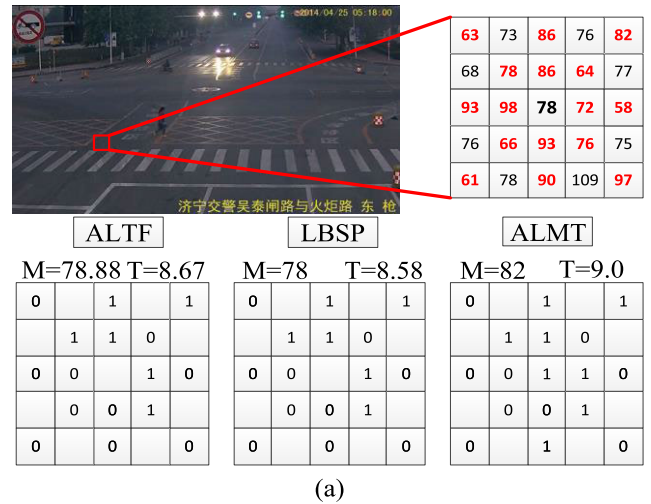


FIGURE 4. Contrast examples of calculating the ALTF, LBSP and ALMT at the same pixel location of different frames. (a) Original encoding example. (b) Encoding example with noise cause by local illumination change.

Many popular background subtraction methods employ a sequence of frames or a single frame to initialize the ideal background models. For example, the pixel-based adaptive segmentation (PBAS) model [38] initializes the background model using the first N frames, and ViBe [36] is initialized by only one frame. However, these initialization methods may fail to form the ideal initial model and lead to numerous ghosts in urban traffic scenes due to slow-moving or temporarily stopped vehicles.

To deal with slow-moving or temporarily stopped vehicles, the ideal initialization model needs to take samples from one or more traffic light cycles. According to the principle of consistency of time, assuming that we have a pixel $p(x)$, the corresponding background model $B(x)$ is initialized by a short interval selecting the ALMT features values from the feature image sequence as follows

$$B(x) = \{I_1(x), \dots, I_{1+(N-2) \times K}(x), I_{1+(N-1) \times K}(x)\}, \quad (9)$$

where N is the number of ALMT feature samples in the background model and K is the short frame interval, $I_1(x)$ is the ALMT feature sample which is selected from the first frame and $I_{1+(N-1) \times K}(x)$ is the ALMT feature sample in the $1 + (N - 1) \times K$ th frame. Therefore, the $1 + (N - 1) \times K$ th frame is at least out of the frame of one traffic light cycle, and K is variable, for example, the frame rate of real-world video sequences is 25 frames per second and one traffic light cycle is approximately 50 seconds in our real-world traffic environment, K is 40 frames at this scenes. Initialization employing no-sequence frames based on equation (9) can distinctly decrease the probability of slow-moving or temporarily stopped vehicles blending into the initial background model and to ensure accurate initial background model.

B. FOREGROUND DETECTION

After initialization of the background model, foreground detection procedure is used to generate the foreground masks. To classify a pixel $p(x)$ as background or foreground at time t , we need to calculate the times of matches between the input ALMT feature $I_t(x)$ and the ALMT feature samples $b_j(x), j = [1, N]$ in its background $B(x)$. The times of matches are presented in the following equation:

$$P(x) = \begin{cases} 1, & \text{if } H(I_t(x), b_j(x)) < R, j = [1, N] \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

$$T(x) = \sum_{M=1}^N P(x), \quad (11)$$

where R is the matching threshold of $I_t(x)$ and $b_j(x)$, $P(x)$ denotes the state of matches and $T(x)$ denotes the times of matches. $H(I_t(x), b_j(x))$ calculated the Hamming distance between $I_t(x)$ and its samples $b_j(x), j = [1, N]$ in the background model. If $H(I_t(x), b_j(x)) < R$, it denotes that $I_t(x)$ matches with $b_j(x)$ and that the possibility of background pixels will increase. The Hamming distance is adapted to measure the similarity and the XOR operator is employed to get the distance. For example, there are two ALMT in Fig.4: 011 110 00110 001 010 and 011 110 00110 001 001 and the corresponding distance is 2. If $2 < R$, a match is found and $P(x) = 1$. A small R will be very accurate to successfully classify pixels as the background, and a larger R will lead to better resistance against irrelevant change, but will make it model more difficult to find those foreground objects which are similar to the background.

After the number of matches is obtained, the label of pixel $p(x)$ is classified as foreground or background according to the following equations:

$$D(x) = \begin{cases} 1, & S(x) < Th, \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

where $D(x)$ denotes the output detection result, 1 implies that the current pixel is a foreground pixel and 0 denotes the background. Th is a fixed predefined threshold that is the minimum number of matches required for a pixel to be

detected as background. In this paper, we set $Th = 4$ to obtain a reasonable trade-off between detection results and computational complexity.

C. BACKGROUND UPDATES

Background updating is an important step of vehicle detection to adapt to scenes that change after segmenting the foreground pixels. Our proposed method uses a conservative update strategy and a random subsampling strategy, and this strategy is better than the state-of-the-art FIFO to make the background model more appropriate for the real-world urban traffic environment. Samples of the model replaced with conservative and random subsampling instead of an FIFO strategy guarantee that the “right” background representation samples can be maintained in the model. New samples may be integrated only if they are detected as background pixels. Thus, the slow-moving or temporarily stopped vehicles can be prevented from being absorbed into the background sample too fast. The main strategy of the model contains two steps: first of all, when the pixel $p(x)$ in the incoming frame is determined to be a background point, it has a $1/\theta$ ratio to replace a recorded sample which possesses the maximum Hamming distance between $p(x)$ and samples of background model in $B(x)$, where θ is a subsampling factor as described in ViBe (θ is 16 in this paper). Then, $p(x)'$, one of the neighbors of $p(x)$ in $F \times F$ region is randomly updated with the $1/\theta$ ratio to replace its sample of the background model using the ALMT feature of $p(x)$.

To deal with the ghosting problem, which is caused by the conservative update strategy, a counter is used to save the number of times one pixel has been consecutively detected as foreground. If the counter exceeds a predefined threshold, the current pixel is replaced by the recorded sample of the model at the same time. The neighbor update strategy helps the ghost region to be automatically added into the model from time to time.

IV. EXPERIMENTS

A. EXPERIMENTAL DATASETS AND MODEL COMPARISONS

To demonstrate the efficiency and robustness of our model, three comparative experiments were introduced on the real-world urban traffic scene videos and change detection challenge (CDnet2014) [39] with illumination changes. Real-world urban traffic scene video sequences provided by the traffic police detachment of Jining City in Shandong Province were recorded with traffic surveillance cameras installed at different urban traffic intersections. This dataset was captured between 7:00 A. M. and 10:00 P.M. over a one-week period and encompass a wide range of weather and illumination conditions. The CDnet2014 dataset is available online at <<http://changedetection.net/>> and it is allowed to evaluate the performance of background subtraction models. The set consists of nearly 160000 realistic scenario frames and the corresponding accurate human annotated

ground-truth is employed for performance evaluation. This is the most complete and popular dataset for background subtraction models, and all frames are grouped into 11 categories. NightVideos of dataset contains 5 kinds of night urban traffic video with illumination changes and the “fluidHighway” is relatively complex because the vehicles in different directions are in the region of interest (ROI). Three typical urban traffic videos with a number of illumination change scenes were selected from real world videos and fluidHighway of CDnet2014 to test in the following comparison experiments. The three corresponding scenes were named “the traffic scenes of strong shadows on a sunny day and waving trees (SSSW)”, “traffic-light scenes at night (TLSN)” and “the traffic scenes of night video in CDnet2014 (TSCD14)”.

Our background subtraction model was compared with several related state-of-the-art local binary texture-based methods and rigorously tested on typical urban traffic videos with illumination changes. For the purpose of fair comparison, the LBP, LTP, SILTP, LBSP and ALTF features were employed to construct a sample consensus background subtraction model. That is, all the compared models were reconstructed with a sample consensus scheme without pattern kernel density estimation or histogram approach. The ALMT feature of our model was replaced by LBP [22], LTP [26], SILTP [27], LBSP [30], LLSLSD [40], IIF [20] and ALTF [31] features to build corresponding LBP model (LBPM), LTP model (LTPM), SILTP model (SILTPM), LBSP model (LBSPM), LLSLSD model (LLSDM), IIF model (IIFM) and ALTF model (ALTFM), respectively. Moreover, all the parameters of features in the compared methods were set to the optimum values according to the original authors’ recommendations and the other parameters of consensus scheme method were set to the same value as in our model.

B. PARAMETERS SETTING

The parameters of the model may improve or weaken the performance of the background subtraction method in different environments. The matching threshold R , the minimum number of matches’ threshold Th and the number of background samples N from (8), (10) and (12) were studied in this subsection. R and Th need to be adjusted in different environments to obtain the optimal performance, which is a challenging work. Therefore, employing several experiments in real-world urban traffic scenes, we empirically set $R = Th = 4$, as in our previous work [31]. For urban traffic scenes, the threshold value of the Hamming distance $R = 4$ and the minimum match threshold value $Th = 4$ are an excellent trade-off between noise resistance and accuracy of foreground detection.

A reasonable number of samples in the background subtraction model are crucial for the balance of the detection precision, computational complexity and sensitivity in real-world urban traffic scenes. Fewer samples will increase the sensitivity of model and reduce the computational complexity, but it will decrease the precision of foreground detection. To a certain extent, increasing the number of samples

will elevate computational complexity but may not promote the precision of foreground detection. The relation between the number of samples N in the model and the corresponding performance based on real-world urban traffic video sequences is shown in Fig. 5. As we can see in this figure, the F-measure value tends to reach the plateau when N reaches the value of 35. In this paper, we determined to set $N = 35$. Although $N = 60$ will obtain a better F-measure in real-world urban traffic video sequences, a bigger N will increase memory and computational complexity, but with little F-measure improvement. Moreover, the first Th are usually sufficient to stop searching the matched sample in stable regions, so increasing N may not directly improve the performance.

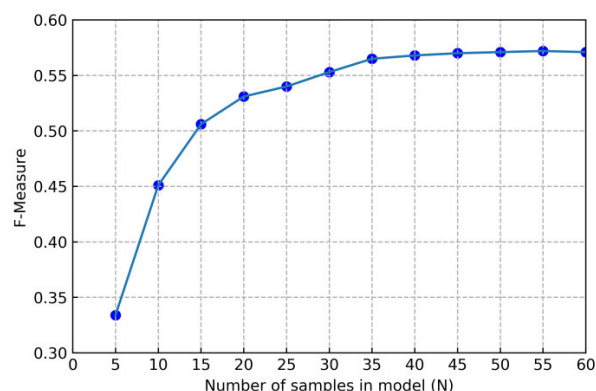


FIGURE 5. The relation between the number of samples N and the corresponding F-measure using the real-world urban traffic video sequences.

C. QUALITATIVE EVALUATIONS

Qualitative evaluation, which is provided employing visual assessment of detected binary objects’ masks for test video sequences, is a subjective measurement of foreground detection results among the different compared methods. The qualitative comparisons of the LBPM, LTPM, SILTPM, LBSPM, LLSLSDM, IIFM, ALTFM and our model under three urban traffic scenes is shown in Figs.6-8. Ground truth images of SSSW and TLSN were created in a semi-supervised way, and the corresponding images were marked with the following steps: to begin with, edge detection approach was employed to get the edges of foreground objects in the original real-world image. Then, traditional background subtraction methods (GMM and ViBe in OpenCV Library) were employed to obtain sketchy foreground. Next, an image marked by classical methods was marked a number of times based on different persons using the original image and edge image and foreground objects. Finally, the ultimate ground-truth image was constructed by averaging labeled images. Note that the foreground detection of all compared approaches is pixel-based and the experimental images of each model are original results without the application of morphological post-processing.

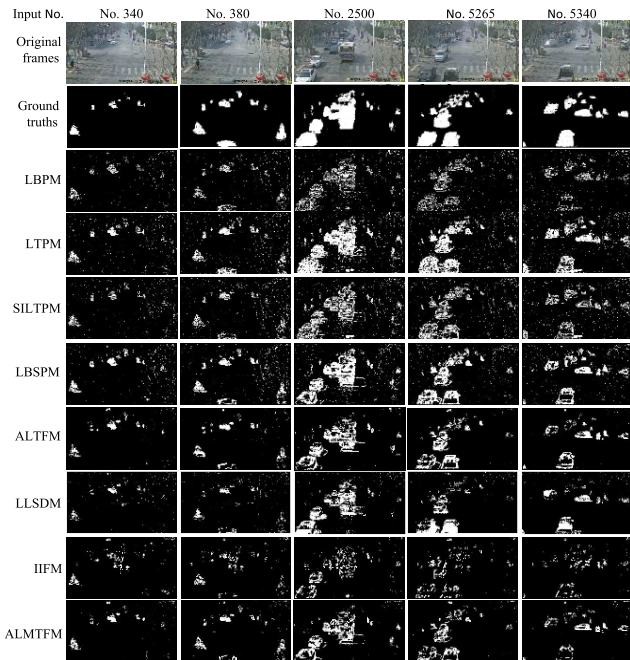


FIGURE 6. Qualitative comparative results of detection masks for sequence with strong shadows on a sunny day and waving trees.

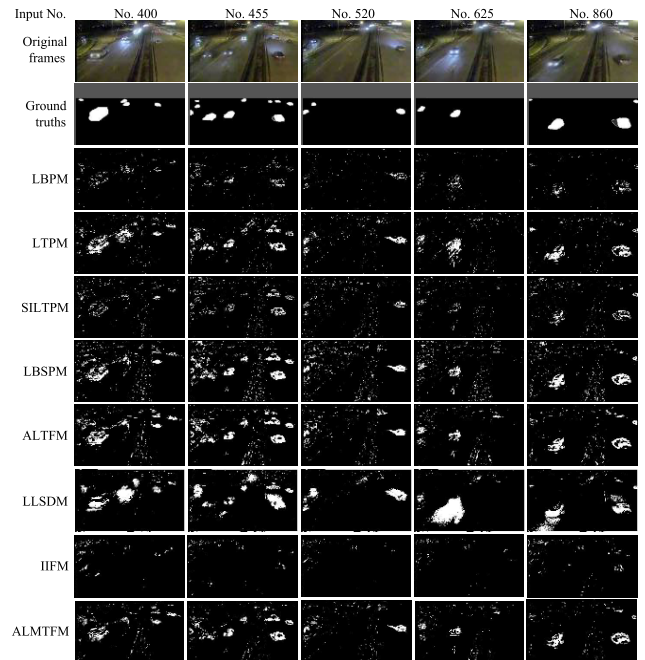


FIGURE 8. Qualitative comparative results of detection masks for traffic sequence of night video in CDnet2014.

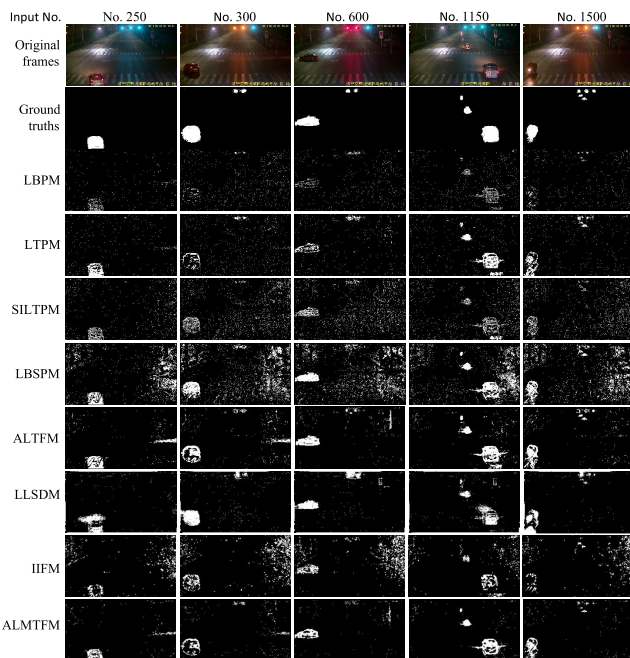


FIGURE 7. Qualitative comparative results of detection masks for sequence of in traffic-light at night.

First, the compared foreground detection results about the traffic sequence of strong shadows on sunny days and waving trees are illustrated in Fig.6, where typical samples of SSSW, corresponding ground-truth and detection results of foreground objects obtained by the LBPM, LTPM, SILTPM, LBSPM, ALTFM, LLSDM, IIFM and our adaptive local median texture feature background model (ALMTFM) method are demonstrated from the second to the tenth row,

respectively. The main challenges of these environments are the random illumination changes in nature (such as the sun reappearing from behind a cloud), the shadows of the swaying trees or foreground objects and the reflected light of vehicles and advertising boards. Five typical frames including part of the challenges were selected to illustrate the comparison of results of foreground detection from the second to the sixth column. The second column illustrates the 340-th frame in which vehicles slow down to wait for a green light and pedestrians start to pass through the intersection when the traffic light turns yellow. A few seconds later, the traffic light turning red at frame 340 is shown in the second column to describe the fact that some vehicles have stopped to wait for the green light and pedestrians are passing the intersection. As we can see, all the methods almost detect the foreground objects in the first two columns, but the LBPM, LTPM, SILTPM, LBSPM and LLSDM cannot effectively prevent a large amount of noise. In the third column at frame 2500, where the sun reappears from behind a cloud, simultaneously leading to a large amount of strong shadows, one can notice that the LBPM, LTPM, SILTPM, LBSPM and LLSDM are susceptible to shadows and cannot effectively manage strong shadows of the vehicles, whereas ALTFM and ALMTFM are visibly better and precisely acquire a large number of vehicles during the middle of the green light. In the final two columns, in which a cloud blocks the sunlight at frame 5265 and then the sun reappears 4 seconds later, leading to dazzling reflected light on the windshield at frame 5340, our approach mitigates all the challenges in a better way to obtain satisfactory results. A comparison of the results shows that the random illumination changes may produce some noise or

incorrect detection of the existing objects. Our approach is superior in the result of completely and precisely detection vehicles and can resist more noise of foreground detection in Fig.6.

Second, the comparative experiment about the traffic-light sequence at night is shown in Fig.7, where frequent sudden illumination changes are the main challenges. In this scene, the traffic light and interference of street lamps will produce strong and frequent sudden illumination variation glares and a large amount of specific image noise. In the case of the green light at frame 250, a vehicle that is passing the intersection can be partly or fully observed for all compared approaches in the second column. Two seconds later at frame 300, where the traffic light turns yellow and the vehicle in frame 250 is still in view is shown in the third column. It can be seen that the LBPM and IIMF fail to discover the vehicle and that the remaining approaches can partly or fully observe the vehicle. In the red traffic light scene at frame 600, where a vehicle is crossing the intersection, we would observe that our proposed approach can completely and precisely acquire the foreground objects with little noise. The fifth column of the figure illustrates the middle of a green light in the 1150-th frame, in which three vehicles are passing the view. It can be seen that all the approaches would partly or fully discover three vehicles and our model handles noise and the glare effect better than others. Finally, the traffic light is starting to turn yellow at frame 1500, in which two vehicles are leaving the intersection and one motorcycle is coming. The ALTFM and ALMTFM can detect the motorcycle well without much noise. Visually, the proposed approach not only observes wonderful foreground results that are closer to the ground truth but it also produce little noise in contrast with other model in Fig.7.

Finally, we present qualitative comparisons between these models using the sequence that is the traffic sequence of night video in CDnet2014 dataset in Fig.8. The main challenges of these night videos are the strong glare that is produced by vehicle headlights over the road surface, the turning on/off of high or low beam lights and the reflected light of other vehicles or electronic advertising boards. It can be seen that the LBPM, SILTPM and IIFM cannot clearly find the corresponding vehicle and LTPM, LBSPM, ALTFM, LLSDM and our approach can partly or completely detect the vehicles. Visually, our method handles the strong glare effect better than other and could obtain correct and fuller foreground vehicles without more noise in Fig.8. This demonstrates the effectiveness of the ALMTFM to manage sudden and gradual illumination variations in urban traffic scenes.

D. QUANTITATIVE EVALUATIONS

To compare the performance of different background models by quantitative evaluation, the typical evaluation metrics have been used. According to [39], the recall, precision and F-measure are employed to compare the performance of background subtraction methods in urban traffic scenes at the

pixel level, and these metrics are defined as follows:

$$Recall = \frac{TP}{total\ of\ actual\ foreground\ object\ pixels} = \frac{TP}{TP + FN}, \tag{13}$$

$$Precision = \frac{TP}{total\ of\ estimated\ foreground\ object\ pixels} = \frac{TP}{TP + FP}, \tag{14}$$

$$F - measure = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision}, \tag{15}$$

where TP stands for the total number of the true positive pixels that are correctly labeled as foreground, FN stands for the total number of false negative pixels that are incorrectly labeled as background and FP stands for the total number of false positive pixels that are incorrectly labeled as foreground. A higher metric value means a higher performance of background subtraction approaches. The F-measure, which can reconcile the accuracy measurements of *recall* and *precision* by fairly weighting their harmonic balances, is commonly employed as a good indicator of the overall performance of background subtraction approaches.

As presented in Fig.9, which presents a comparison of the results of recall for SSSW, TLSN and TSCD14, LBSPM outperforms all methods in real-world scenes and LLSDM outperforms all methods in CDnet2014 dataset scenes. The results of our approach are high in the ranking. It can be seen that the precision of the ALTFM and ALMTFM outperforms that of the other compared models in real-world scenes and LBSPM outperforms all other methods in CDnet2014 dataset scenes in Fig.10. In addition, as indicated in Fig10 for the TLSN sequence, the precision of our method is up to 0.545, which is higher than all the compared methods. The results of the F-measure on SSSW, TLSN and TSCD14 are indicated in Fig.11, where the F-measure results for our model are up to 0.699, 0.560 and 0.261 in the SSSW, TLSN and TSCD14 sequence, respectively. The compared results of the metrics revealed in Figs.9-11 clearly show that the proposed

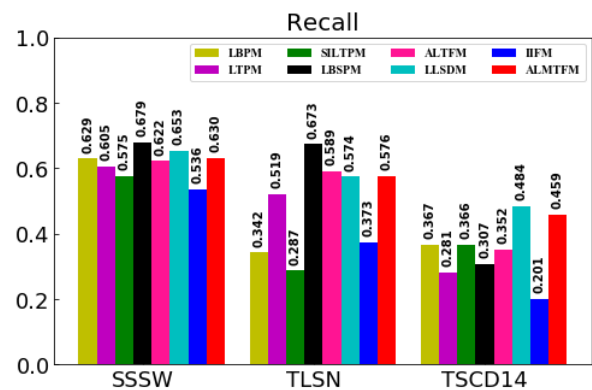


FIGURE 9. The comparison recall results of all compared for the real-world and CDnet2014 datasets.

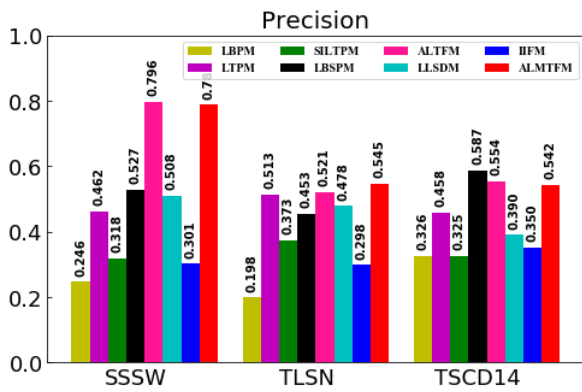


FIGURE 10. The comparison precision results of all compared for the real-world and CDnet2014 datasets.

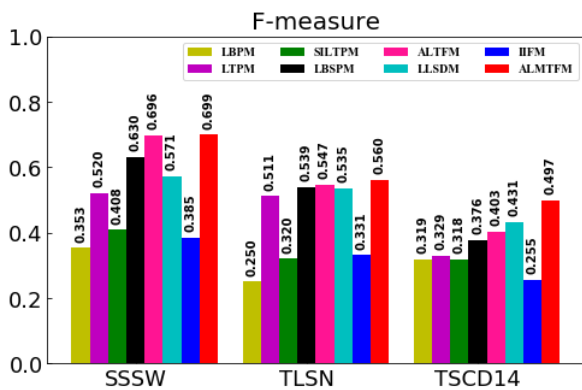


FIGURE 11. The comparison F-measure results of all compared for the real-world and CDnet2014 datasets.

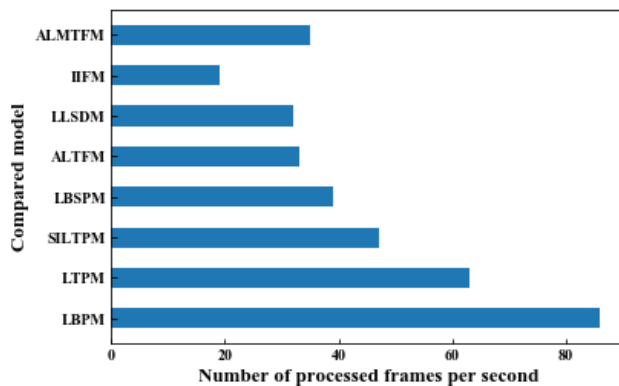


FIGURE 12. The average processing speed in terms of frames/second(fps).

model provides outstanding F-measure performance in comparison with the other excellent methods in the SSSW, TLSN and TSCD14 sequence. The reason for this is that an adaptive feature and sample consensus scheme were combined to handle sudden and gradual illumination changes in complex urban traffic scenes and performed well for illumination changes.

Finally, evaluating the computational complexity of background model is important for real-time video-based applications scenes. In this experiment, the average processing speed results of all models on three urban traffic scenes is shown in Figs.12. Full-resolution Images (1280×720 : width \times height) were sub-sampled to a resolution of 280×160 before processing in VS2012+opencv2.4.2 environment, and PC equipped with intel(R) core(TM) i5_4300 cpu@2.60GHz. LBP processed the highest number of frames per-second. Furthermore, LTPM, SILTPM, LBSPM, LLSDM and ALTFM had fairly good processing speeds. However, IIFM is not suitable for real-time applications due to its matrix decomposition operation. The experimental shown that the average process speed reach is 35 frames/s. Thus, the performance of the proposed ALMTFM is suitable for real-time surveillance systems.

V. CONCLUSION

In this study, an adaptive local median texture feature and sample consensus scheme are combined to efficiently manage the deficiencies of the background subtraction model, which is easily polluted by sudden or gradual illumination changes in complex real-world urban traffic scenes. Local median texture features extract the adaptive distance threshold employing the median information in a predefined local region of a pixel and Weber’s law. The ALMTFM is proposed to handle illumination changes and detect vehicles in real-world urban traffic scenes. The qualitative and quantitative comparisons of the results indicate that this background model achieves better performance than other methods suggested in the literature. The theoretical analysis and experiment results of the ALMTFM revealed that it is excellent for handling sudden and gradual illumination changes in real-world complex urban traffic scenes. However, the parameters of our proposed model are difficult to determine due to random illumination changes scenes and a large number of experiments are required to determine the parameters. In addition, our method cannot manage motionless or low-speed motion objects of urban traffic intersections. In future work, we will focus on eliminating the influence of these parameters, and shadows and motionless or low-speed motion problems will be discussed to improve the performance in real-world urban traffic scenes.

REFERENCES

- [1] N. Buch, S. A. Velastin, and J. Orwell, “A review of computer vision techniques for the analysis of urban traffic,” *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 920–939, Sep. 2011.
- [2] T. Bouwmans and B. Garcia-Garcia, “Background subtraction in real applications: Challenges, current models and future directions,” 2019, *arXiv:1901.03577*. [Online]. Available: <http://arxiv.org/abs/1901.03577>
- [3] W. Kim and C. Jung, “Illumination-invariant background subtraction: Comparative review, models, and prospects,” *IEEE Access*, vol. 5, pp. 8369–8384, 2017.
- [4] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, “Pfunder: Real-time tracking of the human body,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, Jul. 1997.
- [5] 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.*, Jun. 1999, pp. 246–252.

- [6] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 773–780, May 2006.
- [7] B. White and M. Shah, "Automatically tuning background subtraction parameters using particle swarm optimization," in *Proc. IEEE Multimedia Expo Int. Conf.*, Jul. 2007, pp. 1826–1829.
- [8] J. Li and Z. Miao, "Foreground segmentation for dynamic scenes with sudden illumination changes," *IET Image Process.*, vol. 6, no. 5, pp. 606–615, Jul. 2012.
- [9] F. C. Cheng, S. C. Huang, and S. J. Ruan, "Illumination-sensitive background modeling approach for accurate moving object detection," *IEEE Trans. Broadcast.*, vol. 57, no. 4, pp. 794–801, Jul. 2011.
- [10] S. Mahmoudpour and M. Kim, "Robust foreground detection in sudden illumination change," *Electron. Lett.*, vol. 52, no. 6, pp. 441–443, Mar. 2016.
- [11] J. Choi, H. J. Chang, Y. J. Yoo, and J. Y. Choi, "Robust moving object detection against fast illumination change," *Comput. Vis. Image Understand.*, vol. 116, no. 2, pp. 179–193, 2012.
- [12] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–39, 2011.
- [13] S. Wang, Y. Wang, Y. Chen, P. Pan, Z. Sun, and G. He, "Robust PCA using matrix factorization for background/foreground separation," *IEEE Access*, vol. 6, pp. 18945–18953, 2018.
- [14] S. Javed, A. Mahmood, T. Bouwmans, and S. K. Jung, "Spatiotemporal low-rank modeling for complex scene background initialization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 6, pp. 1315–1329, Jun. 2018.
- [15] A. Zheng, T. Zou, Y. Zhao, B. Jiang, J. Tang, and C. Li, "Background subtraction with multi-scale structured low-rank and sparse factorization," *Neurocomputing*, vol. 328, pp. 113–121, Feb. 2019.
- [16] D. Sakkos, E. S. L. Ho, and H. P. H. Shum, "Illumination-aware multi-task GANs for foreground segmentation," *IEEE Access*, vol. 7, pp. 10976–10986, 2019.
- [17] M. Babae, D. T. Dinh, and G. Rigoll, "A deep convolutional neural network for video sequence background subtraction," *Pattern Recognit.*, vol. 76, pp. 635–649, Apr. 2018.
- [18] T. Akilan, Q. J. Wu, A. Safaei, J. Huo, and Y. Yang, "A 3D CNN-LSTM-based image-to-image foreground segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 959–971, Mar. 2020.
- [19] D. Zeng, M. Zhu, and A. Kuijper, "Combining background subtraction algorithms with convolutional neural network," *J. Electron. Imag.*, vol. 28, no. 1, pp. 013011–013021, 2019.
- [20] W. Kim and Y. Kim, "Background subtraction using illumination-invariant structural complexity," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 634–638, May 2016.
- [21] W. Kim, "Moving object detection using edges of residuals under varying illuminations," *Multimedia Syst.*, vol. 25, no. 3, pp. 155–163, Jun. 2019.
- [22] M. Heikkilä and M. Pietikainen, "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.
- [23] W. Kim, "Background subtraction with variable illumination in outdoor scenes," *Multimedia Tools Appl.*, vol. 77, no. 15, pp. 19439–19454, Aug. 2018.
- [24] K. Goyal and J. Singh, "Texture-based self-adaptive moving object detection technique for complex scenes," *Comput. Electr. Eng.*, vol. 70, pp. 275–283, Aug. 2018.
- [25] D. Zeng, M. Zhu, F. Xu, and T. Zhou, "Extended scale invariant local binary pattern for background subtraction," *IET Image Process.*, vol. 12, no. 8, pp. 1292–1302, Aug. 2018.
- [26] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.
- [27] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikainen, and S. Z. Li, "Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1301–1306.
- [28] S. Vasamsetti, N. Mittal, B. C. Neelapu, and H. K. Sardana, "3D local spatio-temporal ternary patterns for moving object detection in complex scenes," *Cognit. Comput.*, vol. 11, no. 1, pp. 18–30, Feb. 2019.
- [29] K. L. Chan, "Segmentation of moving objects in image sequence based on perceptual similarity of local texture and photometric features," *EURASIP J. Image Video Process.*, vol. 62, no. 1, pp. 1–16, 2018.
- [30] 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.
- [31] Y. Zhang, C. Zhao, W. Shi, and K. Leng, "Vehicles detection for illumination changes urban traffic scenes employing adaptive local texture feature background model," *IET Intell. Transp. Syst.*, vol. 12, no. 10, pp. 1283–1290, Dec. 2018.
- [32] K. Roy, R. Arefin, F. Makhmudkhujayev, O. Chae, and J. Kim, "Background subtraction using dominant directional pattern," *IEEE Access*, vol. 6, pp. 39917–39926, 2018.
- [33] 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.
- [34] L. Ji, Y. Ren, X. Pu, and G. Liu, "Median local ternary patterns optimized with rotation-invariant uniform-three mapping for noisy texture classification," *Pattern Recognit.*, vol. 79, pp. 387–401, Jul. 2018.
- [35] M. Haque and M. Murshed, "Perception-inspired background subtraction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 12, pp. 2127–2140, Dec. 2013.
- [36] 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.
- [37] G. Han, J. Wang, and X. Cai, "Improved visual background extractor using an adaptive distance threshold," *J. Electron. Imag.*, vol. 23, no. 6, pp. 1–12, 2014.
- [38] 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*, Jun. 2012, pp. 38–43.
- [39] 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 (CVPRW)*, Jun. 2014, pp. 393–400.
- [40] D. Zeng, M. Zhu, T. Zhou, F. Xu, and H. Yang, "Robust Background Subtraction via the Local Similarity Statistical Descriptor," *Appl. Sci.*, vol. 7, no. 10, pp. 989–1011, 2017.
- [41] A. Arunmozhi and J. Park, "Comparison of HOG, LBP and Haar-like features for on-road vehicle detection," in *Proc. IEEE Int. Conf. Electro/Inf. Technol. (EIT)*, May 2018, pp. 362–367.



YUNSHENG ZHANG received the B.S. degree in applied mathematics from Hubei University for Nationalities, Enshi, China, in 2007, and the M.S. degree in applied mathematics from Liaoning Technical University, Fuxin, China, in 2013, and the Ph.D. degree in transportation engineering from Southeast University, Nanjing, China, in 2016. He worked as an Assistant Professor with the School of logistics and Engineering Management, Hubei University of Economics, China, from May 2017 to April 2018. He is currently a Postdoctoral Fellow with Xi'an Jiaotong University, China. His research interests include image processing, target detection, and intelligent transportation systems.



WEIBO ZHENG received the B.S. degree in industrial engineering from the Xian University of Technology, Xian, Shaanxi, China, in 2010, and the M.S. and Ph.D. degrees in industrial engineering from Xian Jiaotong University, in 2017. Since 2017, he was a Research Assistant with the School of Management, Xian Jiaotong University. His research interests include project scheduling optimization with cash flow under uncertainty environment, and using method and techniques of operations research to optimize energy management.



KAIJUN LENG received the M.S. degree in applied mathematics from the Wuhan University of Technology, in 2006, and the Ph.D. degree in management science and engineering from the Huazhong University of Science and Technology, in 2010. He is currently a Postdoctoral Researcher who works in the China Academy of Social Sciences at the moment. He is also working as a Full Professor with the School of Business Administration, Hubei University of Economics, China. His research interests include logistics and supply chain management, theory of constraints, and operational research. In recent years, he has presided and attended a number of the National Natural Science Foundation of China, the ministry of education funding in key projects, the Hubei province and abroad research project, and so on. And, he has already published certain academic research articles in many international journals and conferences with high IF index.



HAO LI received the Ph.D. degree in highway & railway engineering from Chang'an University, in 2013. He has been a Professor with the Department of Information Engineering, Xi'an University, since 2015, and the Director of the Xi'an Internet of Things Application Engineering Laboratory, since 2016. His research interests include intelligent transportation and traffic information perception.

...