The following publication Y. -Q. Chen, Z. -L. Sun and K. -M. Lam, "An Effective Subsuperpixel-Based Approach for Background Subtraction," in IEEE Transactions on Industrial Electronics, vol. 67, no. 1, pp. 601-609, Jan. 2020 is available at https://doi.org/10.1109/TIE.2019.2893824.

An Effective Sub-Superpixel-Based Approach for Background Subtraction

***Abstract*—How to achieve competitive accuracy and less computation time simultaneously for background estimation is still an intractable task. In this paper, an effective background subtraction approach for video sequences is proposed based on a sub-superpixel model. In our algorithm, the superpixels of the first frame are constructed using a simple linear iterative clustering method. After transforming the frame from a colour format to gray level, the initial superpixels are divided into *K* smaller units, i.e. sub-superpixels, via the *k*-means clustering algorithm. The background model is then initialized by representing each sub-superpixel as a multidimensional feature vector. For the subsequent frames, moving objects are detected by the sub-superpixel representation and a weighting measure. In order to deal with ghost artifacts, a background model updating strategy is devised, based on the number of pixels represented by each cluster center. As each superpixel is refined via the sub-superpixel representation, the proposed method is more efficient and achieves a competitive accuracy for background subtraction. Experimental results demonstrate the effectiveness of the proposed method.**

***Index Terms*—Background subtraction, *k*-means clustering, superpixel.**

I. Introduction

THE task of background subtraction is to divide an observed image into two complementary sets of pixels, i.e. the foreground object and the background that does not contain any object of interest. For analyzing an image or a video, the computational efficiency can be improved significantly by dealing with only the foreground object detected via background subtraction [1], [2]. In most background-estimation methods, the static part is generally assumed to be the background, and the rest is considered as the foreground object. Nevertheless, background subtraction remains a challenging research topic nowadays, because of many unfavorable factors in real applications, such as illumination variations, noise, ghost artifacts, etc. [3], [4], [5].

So far, many effective works have been reported for background estimation. A complete survey was provided in [6] for the traditional and recent approaches. In [6], the various algorithms were categorized in terms of the mathematical models and the critical situations to handle. Gaussian model is a relatively simple and widely used statistical model. In the Gaussian model, the pixel intensity values of the consecutive frames can be modeled by a Gaussian distribution. As the most common approach, Gaussian mixture model (GMM) was proposed in [7] to deal with the dynamic backgrounds. Subsequently, many improvements of the GMM were developed to be more robust and adaptive to the critical situations [8].

Nevertheless, for the background and foreground, the probability density functions (pdfs) are likely to vary from image to image, and will not have a known parametric form. By estimating the pdf directly from the data, a nonparametric kernel density estimation approach (KDE) [9] was proposed to build the statistical representations of the background and the foreground. Instead of using an explicit pixel model, each background pixel is modeled with a set of closest samples in the VIsual Background Extractor (ViBe) [10]. In [11], each background pixel is assigned with a series of key color values stored in a codebook. A multiscale spatial-temporal background model was proposed in [12] to detect motion in low contrast dynamic scenes. In [13], a robust background model for object detection, namely co-occurrence probability-based pixel pairs (CP3), was proposed to deal with the illumination variation and the burst motion.

In [14], a thorough review was made for the recent developments of a class of approaches, which can be represented by a decomposition framework of low-rank plus additive matrices. Compared to the mono or the trichromatic images, the multispectral images can provide a more elaborate spectral-spatial and temporal model for a more accurate segmentation. In order to alleviate the heavy computation cost and complexity, an online stochastic tensor decomposition approach was proposed for background subtraction in multispectral video sequences [15]. In [16], to ensure continuity on spatial and temporal manifolds, a background-foreground model was proposed by incorporating the spatial and temporal sparse subspace clustering into the robust principal component analysis (RPCA) framework [17].

Instead of low level or hand-crafted features, a background subtraction algorithm was proposed based on spatial features learned with convolutional neural networks [18]. In [19], a detailed analysis was performed for deep neural network-based background subtraction approach with respect to the feature maps, the important filters for the detection accuracy, and the operations to suppress false positives from dynamic backgrounds. By using the output features from different layers of the network, a multiscale fully convolutional network architecture was proposed for infrared foreground object detection [20].

In some real-time applications, the computational efficiency is a major concern for background-subtraction algorithms. In this paper, an effective background subtraction approach is proposed, based on the sub-superpixel model. Instead of a whole superpixel, each initialized superpixel is divided into K smaller units, i.e. sub-superpixels, via the k-means clustering algorithm. A multidimensional feature vector is then used to represent each superpixel. In order to deal with the ghost artifact, a background model updating strategy is devised, based on the number of pixels represented by each cluster center. As the superpixel is refined and its representation is

simplified, the proposed approach can achieve a competitive accuracy and require less computation time for background subtraction.

The remainder of the paper is organized as follows. A brief review is given in Section II for the various superpixel-based approaches. In Section III, we present our proposed background-subtraction algorithm. Experimental results and related discussions are given in Section IV, and the concluding remarks are presented in Section V.

II. Related Works

During the past few years, many research studies have been carried out on background substraction. In this study, we propose a superpixel based model for background substraction. Therefore, in this Section, we make a brief review to some superpixel based background substraction methods.

The size of the picture element to model the background can either be a pixel, a block, or a region [21]. As a typical region-level representation, superpixels, i.e. clusters of similar pixels, have been introduced for background estimation. Compared to the pixel-based methods, the use of superpixels can improve the accuracy of background estimation, and can significantly reduce the computational complexity and the memory requirement.

In [22], a simple linear iterative clustering algorithm (SLIC) was proposed by using the k-means clustering to generate superpixels in the labxy color-image plane space. Due to its simplicity and efficiency, subsequently, SLIC has been adopted or improved in other background substraction approaches. Two multi-scale background subtraction methods were proposed by varying the size and zone of superpixels [23] or by random sampling [24] during the superpixel segmentation process of SLIC. In [25], the subpixels were first extracted via the SLIC method, and the corresponding pixel means were computed to be used as the input observation matrix of RPCA for moving object detection.

Instead of the original video sequence, the static and dynamic superpixels were first separately obtained by the entropy rate superpixel segmentation method. Then, a online max-norm based matrix decomposition was employed on each segmented superpixel to separate the low rank and initial outliers support [26]. In [27], a hybrid background subtraction algorithm was proposed by using hierarchical superpixel segmentation, spanning trees and optical flow. In [28], a superpixel-based video-object segmentation algorithm was proposed by using perceptual organization and location prior.

In [29], an efficient and effective superpixel-based algorithm, named SuperBE, was proposed for background estimation. For SuperBE, the RGB mean and colour covariance matrices are used as the discriminative features in the background model. The experimental results demonstrated that SuperBE performed favorably against several state-of-the-art algorithms. Nevertheless, some useful information may be lost when a superpixel is used to represent a group of pixels.

As superBE is the most related work, Table I shows a comparative table on the different main features between SuperBE and our proposed method.

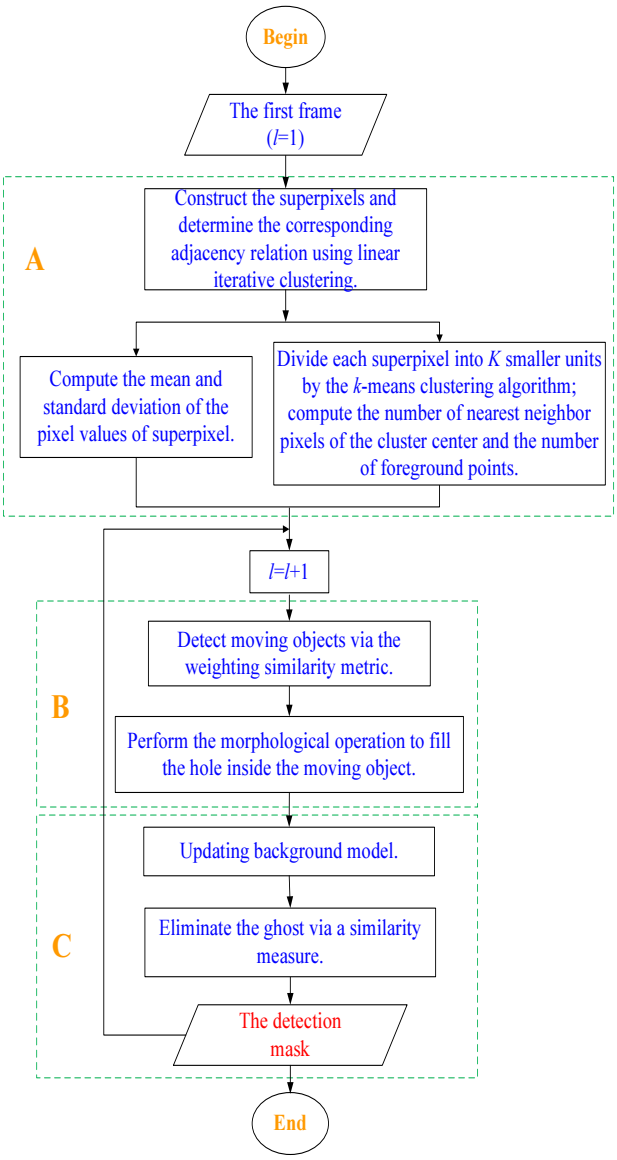

Fig. 1. Flowchart of the sub-superpixel-based approach to background subtraction. Part A: background model initialization, Part B: moving object detection, Part C: background model updating.

III. Methodology

Figure 1 shows the flowchart of the sub-subpixel-based background-subtraction method. There are three main components in the proposed method: background model initialization, moving object detection, and background model updating. A detailed description of these three parts is presented in the following subsections.

A. Background Model Initialization

The process of background model initialization is shown in the part A of Fig. 1. For the first ($l = 1$) frame, superpixels are initially constructed by using the linear iterative clustering algorithm [22]. The adjacency relations are then determined for these superpixels. The segmentation result of the first

TABLE I
A comparative table on the different main features between SuperBE and our proposed method.

Methods	Main feature of super-pixel representation	Dimension of super-pixel representation	Foreground detection method	Background model update	Ghost artifacts processing
SuperBE	Mean and covariance matrix for the pixel intensity of RGB channels.	30*(3+9)=360	The difference of mean and covariance matrix between the superpixel and background model.	Substitute the mean and covariance matrix of the background model by that of the current superpixel with a given probability.	No special measure to deal with the ghost artifacts.
SBS	(a) Cluster centers of super-pixel, the pixel number belonging to each cluster; (b) The number of outliers; (c) The variance of pixels within super-pixel.	10*(1+1)+2=22	(a) The difference between the pixel and background model; (b) The weighting metric in (6).	Update the superpixel feature representation	A background model updating strategy to deal with the ghost artifacts

frame, represented by the superpixels, is stored and used for all the other subsequent frames.

For the i^{th} $(i = 1, \cdots, N)$ superpixel of the first $(l = 1)$ frame, the mean μ_{1i} and the standard deviation σ_{1i} of its pixel values can be computed as follows:

$$\mu_{1i} = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}, \tag{1}$$

and

$$\sigma_{1i} = \sqrt{\frac{1}{N_i} \sum_{j=1}^{N_i} (x_{ij} - \mu_{1i})^2}, \tag{2}$$

where x_{ij} is the j^{th} pixel in the i^{th} superpixel, and N_i is the corresponding number of pixels in the superpixel. Furthermore, each superpixel is further divided into K smaller units, i.e. sub-superpixels, via the *k*-means clustering algorithm. Given a threshold δ_1, two feature metrics are defined, which are the number of nearest-neighbor pixels (b_{1i}^k) of the cluster center c_{1i}^k $(k = 1, \cdots, K)$ and the number of foreground points (n_{1i}). For all the cluster centers c_{1i}^k, we have

$$b_{1i}^k = b_{1i}^k + 1, if \left| x_{ij} - c_{1i}^k \right| \leq \delta_1, k = 1, \cdots, K. \tag{3}$$

For all the pixels in the i^{th} superpixel, the number of foreground points n_{1i} is given as follows:

$$n_{1i} = n_{1i}+1, if \left| x_{ij} - c_{1i}^k \right| > \delta_1, k = 1, \cdots, K, j = 1, \cdots, N_i \tag{4}$$

B. Moving Object Detection

The part B of Fig. 1 shows the main steps of moving object detection. For the second $(l = 2)$ frame, the superpixels can be directly obtained by using the segmentation results from the first frame. Similarly, we compute b_{2i}^k and n_{2i}. Moreover, a mask matrix P_{li} is generated to record the estimation results of the i^{th} superpixel in the l^{th} frame. Initially, the elements $p_{lij}, (i = 1, \cdots, N; j = 1, \cdots, N_i)$ of P_{li} are all set to be zeros. If the j^{th} pixel in the i^{th} superpixel is judged as a foreground point according to (4), then the element p_{lij} is assigned to be 255.

Furthermore, we need to determine if there is a moving object contained inside or as a part of a superpixel in the second frame. Considering b_{1i}^k and b_{2i}^k in the form of a histogram, the similarity measure d_b can be computed as follows:

$$d_b = \frac{1}{N_i} \sum_{k=0}^{K} \min(b_{2i}^k, b_{1i}^k). \tag{5}$$

Given a threshold value δ_2 , there is a moving object in the superpixel, or the current background cannot represent the superpixel accurately, when $d_b < \delta_2$. Then, we need recalculate σ_{2i} according to (1) and (2).

Moreover, considering σ and n, a weighting metric is defined as follows:

$$D_i = \frac{1}{d_b}(\lambda_1 |\sigma_{2i} - \sigma_{1i}| + (1 - \lambda_1)|n_{2i} - n_{1i}|). \tag{6}$$

Given a threshold δ_3, if $D_i > \delta_3$, the foreground points are kept up for this superpixel. Otherwise, the background model corresponding to the superpixel needs to be updated. Furthermore, all the p_{ij} of a superpixel, which are not marked as foreground, are set as zero. Moreover, the morphological close and open operations are applied to fill the holes inside the moving objects.

C. Background Model Updating

The background model updating process is shown in the part C of Fig. 1. For the superpixels to be updated, the cluster centers of the first frame are randomly substituted with those of the second frame. Given the new cluster centers c_{1i}^k, we need to calculate σ_{1i}, b_{1i}^k and n_{1i} again.

In addition, an effective updating strategy is devised for the similarity measure d_b, to deal with "ghost" artifacts. If a superpixel is marked as foreground for M consecutive frames, the similarity measure d_b can be computed as follows:

$$d_b = \frac{1}{N_i} \sum_{k=0}^{K} \min(b_{1i}^k, b_{1it}^k \times \frac{N_i}{N_{it}}) \tag{7}$$

where b_{1it}^k and N_{it} are the corresponding b_{1i}^k and N_i of the t^{th} neighbor superpixel. Then, the weight metric D_i is

TABLE II
The definition of seven performance metrics and the corresponding description.

Metrics	Equation	Description
Recall	$\mathrm{Re} = \frac{TP}{TP+FN}$	Foreground accuracy
Specificity	$Sp = \frac{TN}{TN+FP}$	Background accuracy
False positive rate	$\mathrm{FPR} = \frac{FP}{TN+FP}$	Foreground error rate
False negative rate	$\mathrm{FNR} = \frac{FN}{TP+FN}$	Background error rate
Overall error rate	$\mathrm{PWC} = \frac{100*(FN+FP)}{TP+FN+FP+TN}$	Overall error rate
Precision	$\mathrm{Pr} = \frac{TP}{TP+FP}$	Overall accuracy measure
F-Measure	$FM = \frac{2*\mathrm{Pr}*\mathrm{Re}}{\mathrm{Pr}+\mathrm{Re}}$	Overall accuracy measure

computed by (6). If $D_i < \delta_3$, the superpixel is judged as a ghost superpixel. The background model of superpixel is then initiated again to eliminate the ghost.

Similarly, the subsequent frames ($l = 3, \cdots$) are proceed with the steps as described in Sections III-A-III-C.

IV. Experimental Results

A. Experimental data and set-up

The performance of the proposed sub-superpixel-based approach for background subtraction, denoted as SBS, is evaluated on three datasets, i.e., the 2012 IEEE Change Detection Workshop (CDW-2012) [30], the SBI dataset [31], and the Carnegie Mellon Test Images Sequences (CMTIS) [32].

Given True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN), Table II shows the definition of seven performance metrics and the corresponding descriptions [30], which are adopted in our experiments to evaluate the background-subtraction performance of various algorithms.

For the proposed method, one problem is how to determine the weighting coefficient λ_1, and the threshold parameters δ_1, δ_2, and δ_3. As there is no a separated training procedure, traditional parameter-selection methods, such as cross validation, cannot be used to choose the optimal parameters. Generally, most background-subtraction algorithms attempt to find the optimal parameter values by trial and error. Our experiments have shown that a satisfactory performance can generally be achieved when λ_1, δ_2, δ_3, K, and M are set as 0.6, 0.8, 15, 10, and 30, respectively. The parameter δ_1 is set as 20, according to the suggested value of Vibe [10].

Take the performance metric F-Measure for example, Fig. 2 shows the experimental results when the parameters are set at different values. Except for δ_2, it can be seen that the performances are relatively good when the parameters are varied within some intervals. Figure 3 shows the performance metric F-Measure and the runtime when the parameter δ_2 is set at different values. Considering both the accuracy and the runtime, the parameter λ_2 is set at 0.8 for all videos.

B. Experimental comparisons

1) Experimental comparisons on the CDW-2012 dataset: The CDW-2012 dataset contains several challenging situations for background subtraction, such as dynamic background,

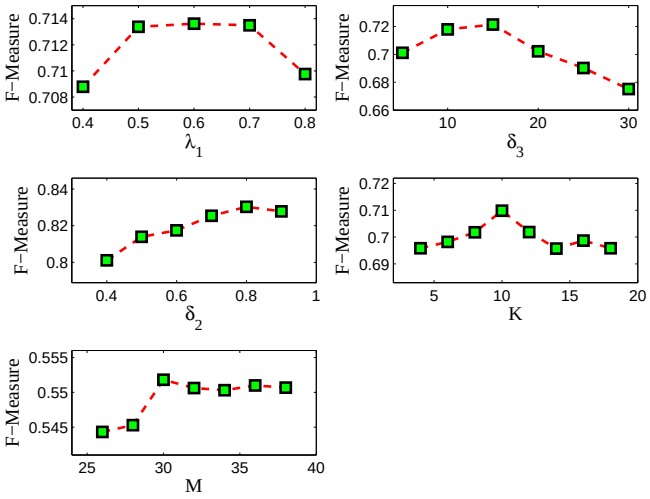

Fig. 2. The experimental results based on the performance metric F-Measure, when the parameters λ_1, δ_2, δ_3, K, and M, are set at different values.

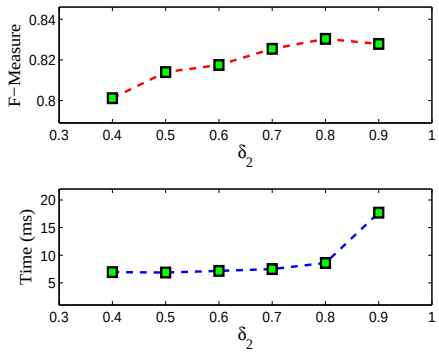

Fig. 3. The performance metric F-Measure and the runtime when the parameter δ_2 is set at different values.

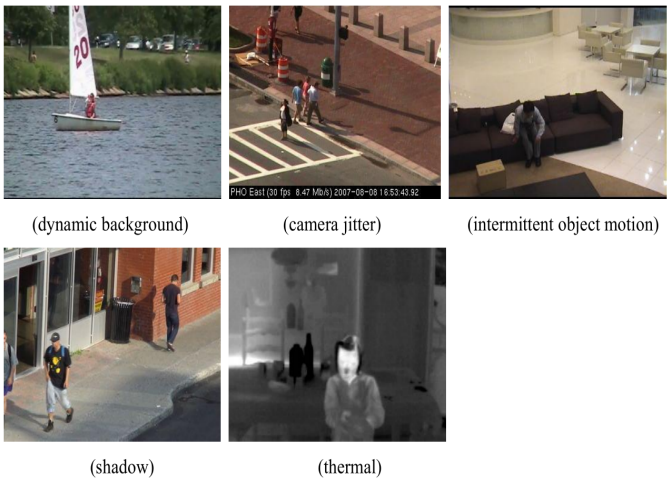

Fig. 4. Some typical frames, with the challenging situations, in the CDW-2012 dataset.

TABLE III
The experimental comparisons of the various algorithms for all categories of the CDW-2012 dataset.

Methods	Re	Sp	FPR	FNR	PWC	Pr	FM
SBS	0.6741	0.9883	0.0115	0.3258	3.0241	0.8343	**0.7197**
SuperBE	0.4515	**0.9925**	**0.0064**	0.5484	5.1092	**0.8425**	0.5356
Vibe	0.6821	0.9829	0.0170	0.3178	3.1177	0.7357	0.6683
GMM	0.7375	0.9690	0.0310	0.2624	4.0680	0.6858	0.6748
CP3	**0.7611**	0.9633	0.0366	**0.2388**	4.0041	0.6464	0.6678
Multiscale Model	0.6714	0.9773	0.0226	0.3285	3.5174	0.6818	0.6165
Euclidean distance	0.7048	0.9691	0.0308	0.2951	4.3458	0.6223	0.6111
KDE	0.6575	0.9909	0.0090	0.3424	**3.0022**	0.7341	0.6437

camera jitter, intermittent object motion, shadow and thermal signature, etc. Figure 4 shows some typical frames with the challenging situations in the dataset. There are 6 video categories in the CDW-2012 dataset, and each category includes 4 to 6 video sequences.

To evaluate the performance of the proposed method, we compare it with seven other background-subtraction algorithms, including Euclidean distance [5], GMM [7], KDE [9], Vibe [10], Multiscale Model [12], CP3 [13], and SuperBE [29]. All simulations were conducted in the C++ environment, running on an ordinary personal computer with double 3.0-GHz CPU and 4-GB memory.

Table III shows the experimental comparisons of the various algorithms for all categories of the CDW-2012 dataset. It can be seen that the proposed method is competitive with other algorithms in general. Furthermore, for the comprehensive metric FM, the performance of the proposed method outperforms all the other algorithms.

For SuperBE, the RGB mean and the colour covariance matrix of the superpixels are used as the discriminative features. SuperBE tends to have fewer false positives (FP) at the cost of more false negatives (FN) [29]. Thus, in terms of the metrics Sp, FPR, and Pr, the performance of SuperBE is better than Vibe, GMM, CP3, Multiscale Model, Euclidean distance, and KDE. For SuperBE, a highly accurate detection of true negative is achieved in background detection by using a relative high threshold value, and the RGB mean and the colour covariance matrix of the whole superpixel. Usually, the difference between the background and the current superpixel caused by the tiny object is not larger than the given threshold value. As a result, the tiny object cannot be detected. If the threshold value is decreased significantly, the whole superpixel is judged as the foreground when the tiny object is detected. As a result, SuperBE tends to have much more false negatives.

Besides the statistical features, some finer details are utilized, by using k-means clustering, in SBS. From (6), we can see that the number of foreground points n_i is used as one component of the weighting metric. When the tiny object enters into the superpixel, the feature n_i will increase significantly. As a result, the tiny object can be detected sensitively. Nevertheless, since both the whole representation and the local feature are considered in SBS, there are no much false negatives when the tiny object is detected. The performance indices Re and FNR are obviously improved with the increasing of TP. On the other hand, the performance

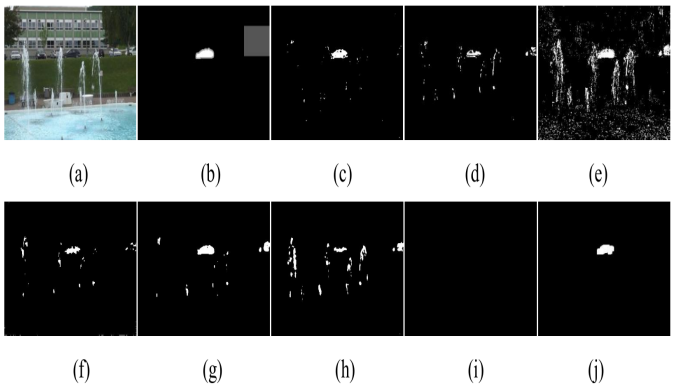

Fig. 5. Results produced by different methods: (a) The original image, (b) the ground truth, (c) CP3, (d) Vibe, (e) GMM, (f) Euclidean distance, (g) Multiscale Model, (h) KDE, (i) SuperBE, and (j) SBS.

TABLE IV
The average runtimes (TC in sec.) per frame and frames per second (FPS) of the various methods.

Methods	TC	FPS
SBS	8.59	116
SuperBE	13.31	75
GMM	27.02	37
Vibe	7.41	135
CP3	50	20
Multiscale Model	100	10
Euclidean distance	9.10	110
KDE	40	25

index Pr of SBS is only slightly lower than that of SuperBE. Moreover, the Pr of SBS and SuperBE are obviously higher than that of other methods. Therefore, the comprehensive metric FM of SBS is higher than that of SuperBE and other methods, as shown in Table III.

As shown in Fig. 5(b), the foreground is a small object in the original image. We can see from Fig. 5 that, different from other algorithms, there is no noise in the detection results based on SuperBE and SBS. Nevertheless, the small object is successfully detected by SBS, while it is missed by SuperBE.

Table IV shows the average runtimes (TC in sec.) per frame and frames per second (FPS) of the various methods. Moreover, Table V shows the average runtime ratio between the other methods and the proposed method. It can be seen from Table V that the computation time of SBS is obviously

TABLE V
The average runtime ratio between the other methods and the proposed method.

Methods	TC	FPS
SBS	1	1
SuperBE	1.5495	0.6466
GMM	3.1455	0.3190
Vibe	0.8626	1.1638
CP3	5.8201	0.1724
Multiscale Model	11.6414	0.0861
Euclidean distance	1.0594	0.9483
KDE	4.6566	0.2155

TABLE VI
The experimental comparisons of the various algorithms for the SBI dataset.

Methods	Re	Sp	FPR	FNR	PWC	Pr	FM
SBS	0.7931	0.9264	0.0735	0.1440	**11.203**	0.6650	**0.7018**
SuperBE	0.5306	0.9392	0.0607	0.3112	15.4491	**0.7597**	0.6023
Vibe	0.6112	**0.9398**	**0.0601**	0.2234	13.4579	0.7033	0.6267
GMM	0.6097	0.9215	0.0784	0.2348	15.4094	0.6606	0.5909
Euclidean distance	0.6694	0.9318	0.0681	0.1539	11.5857	0.5449	0.5603
KDE	**0.8545**	0.8708	0.1291	**0.0831**	12.8741	0.4820	0.5527

lower than that of GMM, CP3, Multiscale Model, KDE. As a similar method, the computation time of SuperBE is about 1.5 times that of SBS. The computation times of Vibe Euclidean distance, SBS are close each other. Compared to Vibe and Euclidean distance, it can be seen from Table III that SBS has a relative good comprehensive performance index FM.

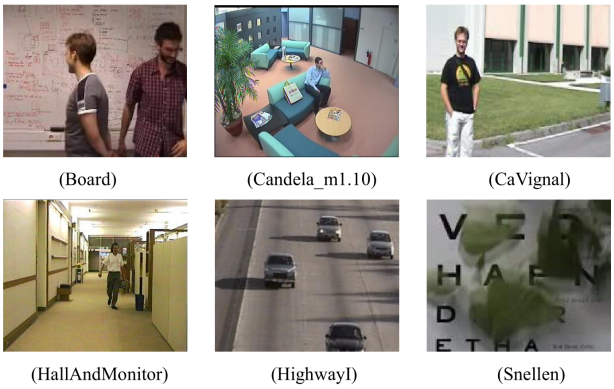

Fig. 6. Some typical frames in the SBI dataset.

2) Experimental comparisons on the SBI dataset: The SBI dataset includes 14 sequences with some negative influence factors, e.g. shadow, occlusion, camera jitter, waving leave, etc. Figure 6 shows some typical frames in the SBI dataset. Both the complete SBI dataset and the ground truth reference background images were made publicly available through the SBMI2015 website [31].

Table VI shows the experimental comparisons of the various algorithms for the SBI dataset. Nevertheless, some unclear errors are encountered when CP3 and Multiscale Model are used to process these two datasets. Thus, the experimental comparisons are only presented for six methods in Table VI. We can see that the performance metric FM of SBS is higher than that of other methods. Thus, SBS has a relative good comprehensive performance.

Figure 7 shows an example of the processed results by different methods. It can be seen that, compared to other methods, SBS can effectively deal with ghost artifacts and can detect the pedestrian completely.

3) Experimental comparisons on the CMTIS dataset: Compared to other two datasets, CMTIS is a relative small size dataset. There are only 500 raw TIF image and the corresponding manually segmented binary masks in the CMTIS dataset. The video sequences are mainly with two negative influence factors, i.e. camera jitter and waving leave. Table

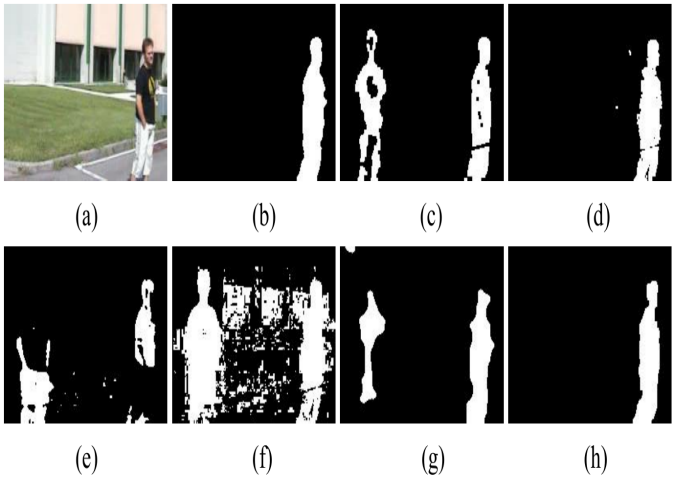

Fig. 7. Results produced by different methods: (a) The original image, (b) The ground truth, (c) Vibe, (d) GMM , (e) Euclidean distance , (f) KDE, (g) SuperBE and (h) SBS

TABLE VII
The experimental comparisons of the various algorithms for the CMTIS dataset.

Methods	Re	Sp	FPR	FNR	PWC	Pr	FM
SBS	**0.9170**	0.9975	0.0024	**0.0014**	**0.3862**	0.8672	**0.8914**
SuperBE	0.5343	**0.9994**	**0.0005**	0.0105	1.0844	**0.9552**	0.6852
Vibe	0.8950	0.9939	0.0060	0.0018	0.7804	0.7209	0.7986
GMM	0.8684	0.9891	0.0108	0.0026	1.3192	0.6156	0.7205
Euclidean distance	0.7631	0.9911	0.0088	0.0041	1.2748	0.6039	0.6743
KDE	0.8754	0.9929	0.0070	0.0022	0.9128	0.6945	0.7745

VII shows experimental comparisons of the various algorithms for the CMTIS dataset. Similar to other two datasets, we can see that the performance indices Sp, FPR, and Pr of SuperBE are higher than that of other methods by achieving a highly detection accuracy of true negatives. Nevertheless, it can be seen from Table VII that SBS still has a relative good comprehensive performance compared to other methods.

C. Related discussions

The morphological operation is a commonly used approach to fill the holes inside the moving objects. Take a frame (Fig. 8(a)) of the video "pedestrians" for example, Figs. 8(c) and 8(d) show the results without and with morphological operations, respectively. Compared to the ground truth shown in Fig. 8(b), it can be seen that the morphological operations

TABLE VIII
The experimental comparison when the morphological operation is used or not used (denoted as SBS-NMO) in the proposed method.

Datasets	Methods	Re	Sp	FPR	FNR	PWC	Pr	FM
CDW-2012	SBS	0.6741	0.9883	0.0115	0.3258	3.0241	0.8343	0.7197
	SBS-NMO	0.6483	0.9907	0.0092	0.3516	1.9420	0.7513	0.6744
SBMI	SBS	0.7931	0.9264	0.0735	0.1440	11.203	0.6650	0.7018
	SBS-NMO	0.6277	0.9700	0.0299	0.1928	10.1634	0.7462	0.6603
CMTIS	SBS	0.9170	0.9975	0.0024	0.0014	0.3862	0.8672	0.8914
	SBS-NMO	0.7458	0.9979	0.0020	0.0044	0.6437	0.8632	0.8002

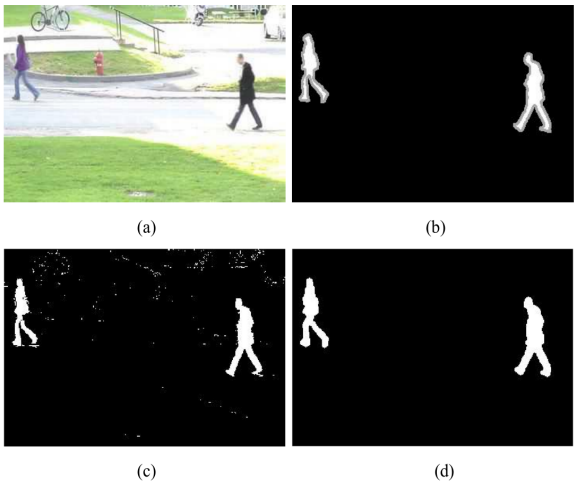

Fig. 8. A comparison of the detection results, with and without using morphological operations: (a) The original image, (b) the ground truth, (c) without morphological operations, and (d) with morphological operation.

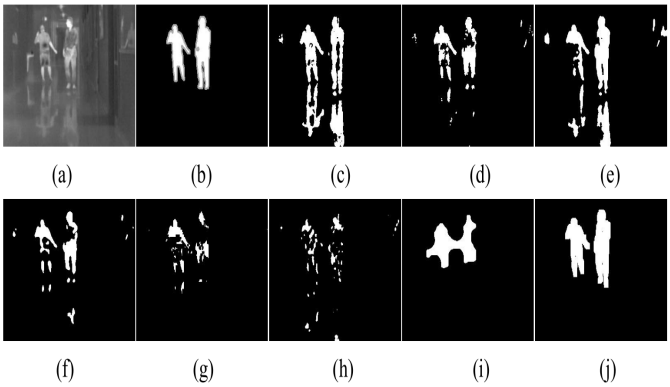

Fig. 9. A comparison of the foreground intrusion for different algorithms: (a) The original image, (b) The ground truth, (c) CP3,(d) Vibe, (e) GMM, (f) Euclidean distance, (g) Multiscale Model, (h) KDE, (i) SuperBE, and (j) SBS

can effectively eliminate noise and fill the hole inside the moving object. Moreover, Table VIII shows the experimental comparison when the morphological operation is used or not used (denoted as SBS-NMO) in the proposed method. We can see that the morphological operation indeed has an obvious effect on the background substraction performance. As a commonly used approach, the morphological operation is also used in other background subtraction approaches, e.g., Multiscale model, SuperBE, Euclidean distance, etc.

When there is no obvious difference between the foreground and the background, the foreground may intrude in the background in most algorithms. For SBS, the updating strategy for the background model is based on two levels, i.e. the superpixel level and the pixel level. Therefore, the foreground intrusion can be effectively alleviated. Take a frame of the category "thermal" for example, Fig. 9 shows the results of the various algorithms. It can be seen that there is no foreground intrusion for SBS.

Denote R, Q, and S as the average pixel number of one superpixel, the pixel number and superpixel number of one frame, respectively. The variables w and h are the parameters of the morphological operation. The operation number of four basic operations is about $24Q + (120 \sim 1800) * Q/R + 4 * w * h * Q$ for SuperBE. Denote p_1 and p_2 as the probabilities caused by the judgment conditions of (5) and (6), the operation number of SBS is about $(3 \sim 30)Q + 4 * p1 * Q + (17 \sim 44) * p1 * p2 * Q + 21Q/R + 5p1 * Q/R + 4 * w * h * Q$. Compared to SuperBE, the computational complexity of SBS is decreased by the judgment conditions. As shown in Table V, the computation time of SuperBE is about 1.5 times that of SBS.

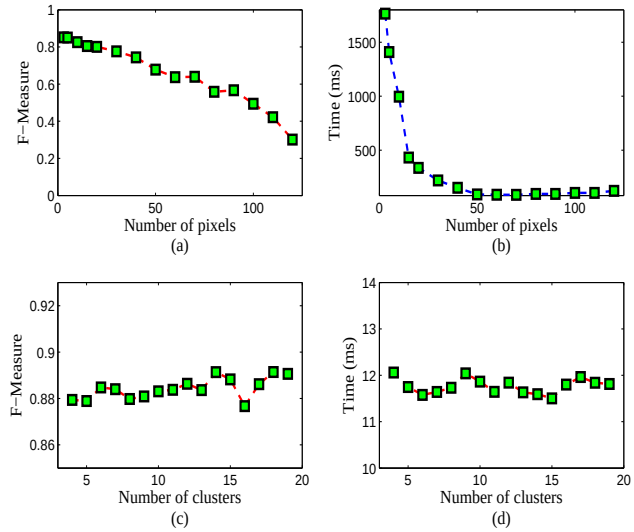

Fig. 10. The performance metric F-Measure and the runtime when different numbers of pixels in superpixel are set for SuperBE, and when different numbers of clusters are set for SBS.

Figure 10 shows the performance metric F-Measure and the runtime when different numbers of pixels in superpixel are set for SuperBE, and when different numbers of clusters are set for SBS. For SuperBE, we can see from Fig. 10 (a) and (b) that, the runtime obviously increases when the number of pixels in superpixel is smaller than 50. Nevertheless, we can see from Fig. 10 (c) and (d) that, the runtime only varies slightly when different numbers of clusters are set for SBS.

V. Conclusion

In this paper, a more effective background-subtraction approach is proposed, based on the sub-superpixel model. A one-order statistic-based representation is proposed for the sub-superpixel model. The proposed sub-superpixel representation was more effective and more robust to scene noise. Moreover, tiny objects can be detected successfully with our proposed method. The proposed background updating strategy can alleviate foreground intrusion. As a post-processing step, morphological operations have been verified to be able to effectively eliminate noise and fill the holes inside moving

objects. Experimental results have demonstrated the effectiveness of the proposed method.

REFERENCES

[1] M. I. Chacon-Murguia and S. Gonzalez-Duarte, "An adaptive neural-fuzzy approach for object detection in dynamic backgrounds for surveillance systems," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 8, pp. 3286-3298, 2012.

[2] T. Celik and H. Kusetogullari, "Solar-powered automated road surveillance system for speed violation detection," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 9, pp. 3216-3227, 2010.

[3] R. C. Luo and C. C. Lai, "Multisensor fusion-based concurrent environment mapping and moving object detection for intelligent service robotics," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 8, pp. 4043-4051, 2014.

[4] Y. L. Chen, B. F. Wu, H. Y. Huang, and C. J. Fan, "A real-time vision system for nighttime vehicle detection and traffic surveillance," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 2030-2044, 2011.

[5] P. M. Jodoin, "Comparative study of background subtraction algorithms," *Journal of Electronic Imaging*, vol. 19, no. 3, 033003, 2010.

[6] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: an overview," *Computer Science Review*, vol. 11, pp. 31-66, 2014.

[7] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," *IEEE Conference on Computer Vision and Pattern Recognition*, 1999.

[8] T. Bouwmans, F. El-Baf, and B. Vachon, "Background modeling using mixture of gaussians for foreground detection: a survey," *Recent Patents on Computer Science*, vol. 1, no. 3, pp. 219-237, 2008.

[9] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of the IEEE*, vol. 90, no. 7, pp. 1151-1163, 2002.

[10] O. Barnich and M. V. Droogenbroeck, "ViBe: a universal background subtraction algorithm for video sequences," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709-1724, 2011.

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

[12] X. Lu, "A multiscale spatio-temporal background model for motion detection," *IEEE International Conference on Image Processing*, vol. 6, pp. 3268-3271, 2014.

[13] D. Liang, S. Kaneko, and M. Hashimoto, "Co-occurrence probability-based pixel pairs background model for robust object detection in dynamic scenes," *Pattern Recognition*, vol. 48, no. 4, pp. 1374-1390, 2015.

[14] T. Bouwmans, el al, "Decomposition into low-rank plus additive matrices for background/foreground separation: a review for a comparative evaluation with a large-scale dataset," *Computer Science Review*, vol. 23, pp. 1-71, 2017.

[15] A. Sobral, s. Javed, K. Soon, T. Bouwmans, and E. H. Zahzah, "Online stochastic tensor decomposition for background subtraction in multispectral video sequences," *Workshop on Robust Subspace Learning and Computer Vision*, 2016, pp. 946-953.

[16] S. Javed, A. Mahmood, T. Bouwmans, and S. K. Jung, "Background-foreground modeling based on spatio-temporal sparse subspace clustering," *IEEE Transactions on Image Processing*, vol. 26, no. 12, pp. 5840-5854, 2017.

[17] N. Vaswani, T. Bouwmans, and S. Javed, and P. Narayanamurthy, "Robust subspace learning: robust pca, robust subspace tracking and robust subspace recovery," *IEEE Signal Processing Magazine*, vol. 35, no. 4, pp. 32-55, 2018.

[18] M. Braham and M. V. Droogenbroeck, "Deep background subtraction with scene-specific convolutional neural networks," *International Conference on Systems, Signals and Image Processing*, 2016, pp. 113-116.

[19] T. Minematsu, A. Shimada, H. Uchiyama, and R. Taniguchi, "Analytics of deep neural network-based background subtraction," *Journal of Imaging*, vol. 4, no. 6, pp. 78, 2018.

[20] D. D. Zeng and M. Zhu, "Multiscale fully convolutional network for foreground object detection in infrared videos," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 4, pp. 617-621, 2018.

[21] T. Bouwmans, C. Silva, C. Marghes, M. S. Zitouni, H. Bhaskar, C. Frelicot, "On the role and the importance of features for background modeling and foreground detection," *Computer Science Review*, vol. 28, pp. 26-91, 2018.

[22] R. Achanta, A. Shaji, K. Smith, and A. Lucchi, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274-2282, 2012.

[23] C. Zhao, T. Zhang, Q. Huang, X. Zhang, D. Yang, Y. Qu, and S. Huang, "Background subtraction based on superpixels under multi-scale in complex scenes," *Chinese Conference on Pattern Recognition*, 2016, vol. 662, pp. 392-403.

[24] W. Fang, T. Zhang, C. Zhao, D. Soomro, R. Taj, and H. Hu, "Background subtraction based on random superpixels under multiple scales for video analytics," *IEEE Access*, vol. 6, pp. 33376-33386, 2018.

[25] S. Javed, A. Mahmood, T. Bouwmans, S. K. Jung, "Superpixels based manifold structured sparse rpca for moving object detection," *British Machine Vision Conference*, 2017.

[26] S. Javed, S. H. oh, A. Sobral, T. Bouwmans, S. K. Jung, "Background subtraction via superpixel-based online matrix decomposition with structured foreground constraints," *IEEE International Conference on Computer Vision Workshop*, 2015, pp. 930-938.

[27] M. Chen, Q. Yang, Q. Li, G. Wang, and M. Yang, "Spatiotemporal gmm for background subtraction with superpixel hierarchy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1518-1525, 2018.

[28] D. Giordano, F. Murabito, S. Palazzo, and C. Spampinato, "Superpixel-based video object segmentation using perceptual organization and location prior," *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4814-4822.

[29] T. Y. Chen, M. Biglari-Abhari, and I. K. Wang, "SuperBE: computationally light background estimation with superpixels," *Journal of Real-Time Image Processing*, vol. 11, pp. 1-17, 2018.

[30] N. Goyette, P. M. Jodoin, F. Porikli, J. Konrad, P. Ishwar, "Changedetection.net: a new change detection benchmark dataset," *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2012, pp. 1-8.

[31] L. Maddalena and A. Petrosino, "Towards benchmarking scene background initialization," *the 18th International Conference on Image Analysis and Processing*, 2015, vol. 9281, pp. 469-476.

[32] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1778-1792, 2005.