Journal of Electronic Imaging

JElectronicImaging.org

# Improved visual background extractor using an adaptive distance threshold

Guang Han
Jinkuan Wang
Xi Cai

SPIE. IS&T imaging.org

# Improved visual background extractor using an adaptive distance threshold

**Guang Han,[a,b] Jinkuan Wang,[b,*] and Xi Cai[b]**
[a]Northeastern University, College of Information Science and Engineering, No. 3-11 Wenhua Road, Heping District, Shenyang 110819, China
[b]Northeastern University at Qinhuangdao, School of Computer and Communication Engineering, No. 143 Taishan Road, Economic and Technological Development Zone, Qinhuangdao, Hebei 066004, China

**Abstract.** Camouflage is a challenging issue in moving object detection. Even the recent and advanced background subtraction technique, visual background extractor (ViBe), cannot effectively deal with it. To better handle camouflage according to the perception characteristics of the human visual system (HVS) in terms of minimum change of intensity under a certain background illumination, we propose an improved ViBe method using an adaptive distance threshold, named IViBe for short. Different from the original ViBe using a fixed distance threshold for background matching, our approach adaptively sets a distance threshold for each background sample based on its intensity. Through analyzing the performance of the HVS in discriminating intensity changes, we determine a reasonable ratio between the intensity of a background sample and its corresponding distance threshold. We also analyze the impacts of our adaptive threshold together with an update mechanism on detection results. Experimental results demonstrate that our method outperforms ViBe even when the foreground and background share similar intensities. Furthermore, in a scenario where foreground objects are motionless for several frames, our IViBe not only reduces the initial false negatives, but also suppresses the diffusion of misclassification caused by those false negatives serving as erroneous background seeds, and hence shows an improved performance compared to ViBe.  [DOI: 10.1117/1.JEI.23.6.063005]

Keywords: moving object detection; background subtraction; adaptive distance threshold; perception characteristics of human visual system; camouflage.

Paper 14331 received Jun. 4, 2014; revised manuscript received Sep. 1, 2014; accepted for publication Oct. 8, 2014; published online Nov. 6, 2014.

## 1 Introduction

In computer vision applications, objects of interest are often the moving foreground objects in a video sequence. Therefore, moving object detection which extracts foreground objects from the background has become a hot issue,[1–7] and has been widely applied to areas such as smart video surveillance, intelligent transportation, and human-computer interaction.

Visual background extractor[8] (ViBe) is one of the most recent and advanced techniques. In comparative evaluation,[9] ViBe produces satisfactory detection results and has been proved effective in many scenarios. For each pixel, the background model of ViBe stores a set of background samples taken in the past at the same location or in the neighborhood. Then, ViBe compares the current pixel intensity to this set of background samples using a distance threshold. Only if the new observation matches with a predefined number of background samples is this pixel classified as background, otherwise this pixel belongs to the foreground. However, ViBe uses a fixed distance threshold in the matching process; hence, it has difficulties in handling camouflaged foreground objects (intentionally or not, some objects may poorly differ from the appearance of the background, making correct classification difficult[9]). Moreover, a "spatial diffusion" update mechanism for background models aggravates the influence of misclassified camouflaged foreground pixels, and then decreases the power of ViBe in detecting still foreground objects. Camouflaged foreground objects and still foreground objects are two key reasons for false negatives in the detection results, and it is imperative and urgent to solve these two challenging issues in video surveillance.

In order to solve the aforementioned challenges, we propose an improved ViBe method using an adaptive distance threshold (hereafter IViBe for short). In light of the sensitivity of the human visual system (HVS) with regard to intensity change under certain background illumination, we set an adaptive distance threshold in the background matching process for each background sample in accordance with its intensity. Experimental evaluations validate that, because of using features of the HVS and performing background matching based on an adaptive distance threshold, IViBe has a better discriminating power concerning foreground objects with similar intensities to the background, and then effectively improves the capability of ViBe in coping with camouflaged foreground objects. Furthermore, IViBe also reduces the number of misclassified pixels which usually serve as erroneous background seeds propagating the false negatives. Experimental results show that, compared with ViBe, our IViBe allows a slower inclusion of still foreground objects into the background, and has a better performance in detecting static foreground objects.

The rest of this paper is organized as follows. In Sec. 2, we briefly explore the major background subtraction approaches. Section 3 describes our IViBe method, introduces the detailed derivation of our adaptive distance threshold,

*Address all correspondence to: Jinkuan Wang, E-mail: wjk@mail.neuq.edu.cn

and analyzes the influence of this adaptive distance threshold together with the "spatial diffusion" update mechanism on the detection results. In Sec. 4, we qualitatively and quantitatively analyze the advantages of our IViBe compared with ViBe. Finally, a conclusion is drawn in Sec. 5.

## 2 Related Work

Background subtraction[10] (BS) is an effective way of foreground segmentation for a stationary camera. In the BS methods, via comparing input video frames to their current background models, the regions corresponding to significant differences should be marked as foreground. Also, the BS techniques adapt their background models to scenario changes through online update and have a moderate computational complexity, which makes them popular methods for moving object detection.

Many BS techniques have been proposed with different kinds of background models, and several recent surveys have been devoted to this topic.[11–13] Although the last decade has witnessed numerous publications on the BS methods, according to Ref. 13, there are still many challenges not completely resolved in real scenes, such as illumination changes, dynamic backgrounds, bootstrapping, camouflage, shadows, still foreground objects, and so on. In 2014, two special issues[14,15] have just been published with new developments for dealing with these challenges.

Next, we briefly explore the major BS approaches according to the different kinds of background models they used.

### 2.1 *Parametric Models*

Gaussian mixture model (GMM) and its improved methods: GMM is a classical and probably the most widely used BS technique.[16] GMM models the temporal distribution of each pixel using a mixture of Gaussians, and many studies have proven that GMM can handle gradual illumination changes and repetitive background motion well. In Ref. 17, Lee proposed an adaptive learning rate for each Gaussian model to improve the convergence rate without affecting the stability. In Ref. 18, Zivkovic and Van Der Heijden proposed a scheme to dynamically determine the appropriate number of Gaussian models for each pixel based on observed scene dynamics to reduce processing time. In Ref. 19, Zhang et al. used a spatio-temporal Gaussian mixture model incorporating spatial information to handle complex motions of the background.

Models using other statistical distributions: recently, a mixture of symmetric alpha-stable distributions[20] and a mixture of asymmetric Gaussian distributions[21] have been employed to enhance the robustness and flexibility of mixture modeling in real scenarios, respectively. They can handle the dynamic backgrounds well. In Ref. 22, Haines and Xiang proposed a Dirichlet process Gaussian mixture model which constantly adapts its parameters to the scene in a block-based method.

### 2.2 *Nonparametric Models*

Kernel density estimation (KDE) and its improved methods: a nonparametric technique[23] was developed to estimate background probabilities at each pixel from many recent samples over time using KDE. In Ref. 24, Sheikh modeled the background using KDE over a joint domain-range representation of image pixels to sustain high levels of detection accuracy in the presence of dynamic backgrounds.

Codebook and its improved methods: the essential idea behind the codebook[25] approach is to capture long-term background motion with limited memory by using a codebook for each pixel. In Ref. 4, a multilayer codebook-based background subtraction (MCBS) model was proposed. Combining the multilayer block-based strategy and the adaptive feature extraction from blocks of various sizes, MCBS can remove most of the dynamic backgrounds and significantly increase the processing efficiency.

### 2.3 *Advanced Models*

Self-organizing background subtraction (SOBS) and its improved methods: in the 2012 IEEE change detection workshop[26] (CDW-2012), SOBS[27] and its improved method SC-SOBS[28] obtained excellent results. In Ref. 27, SOBS adopted a self-organizing neural network to build a background model, initialized its model from the first frame, and employed regional diffusion of background information in the update step. In 2012, Maddalena improved the SOBS by introducing spatial coherence into the background update procedure, which led to the SC-SOBS algorithm providing further robustness against false detections. In Ref. 29, three-dimensional self-organizing background subtraction (3D_SOBS) used spatio-temporal information to detect a stopped object. Recently, the 3DSOBS+[1] algorithm further enhanced the 3D_SOBS approach to accurately handle scenes containing dynamic backgrounds, gradual illumination changes, and shadows cast by moving objects.

ViBe and its improved methods: in the CDW-2012, ViBe[8] and its improved method ViBe+[30] also achieved remarkable results. Barnich and Van Droogenbroeck proposed a sample-based algorithm that builds the background model by aggregating previously observed values for each pixel location. The key innovation of ViBe is introducing the random policy into the BS, which makes it the first nondeterministic BS method. In Ref. 30, Van Droogenbroeck and Barnich improved ViBe in many aspects, including an adaptive threshold. They computed the standard deviation of background samples of a pixel to define a matching threshold. The matching threshold adapts itself to statistical characteristics of background samples; however, all background samples of a pixel have the same thresholds, and one wrongly updated background sample will affect the thresholds of other background samples, which will lead to more misclassification. In Refs. 30 and 31, a new update mechanism separating "segmentation map" and "updating mask" was proposed. The "spatial diffusion" update mechanism can be inhibited in the "updating mask" to detect still foreground objects. In Ref. 32, Mould and Havlicek proposed an update mechanism in which foreground pixels can update their background models by replacing the most significant outlying samples. This update policy can improve the ability to deal with ghosts.

### 2.4 *Human Visual System-Based Models*

Visual saliency, another important concept about the HVS, has already been used in the BS methods. In Ref. 33, Liu et al. represented object saliency for moving object detection by an information saliency map calculated from spatio-temporal volumes. In Ref. 34, Mahadevan and Vasconcelos

proposed a BS algorithm based on spatio-temporal saliency using a center-surround framework, which is inspired by biological mechanisms of motion-based perceptual grouping. These methods have shown the potential of the HVS in moving object detection.

In this paper, we propose an improved BS technique which uses the characteristic of the HVS.

We introduce an adaptive distance threshold into ViBe to simulate the capacity of the HVS in perceiving noticeable intensity changes, which can discriminate camouflaged foreground objects and reduce false negatives. Together with ViBe's update policy, our method further improves the ability to detect foreground objects that are motionless for a while. Hence, IViBe improves the ability of ViBe in dealing with camouflaged and still foreground objects.

## 3 Improved ViBe Method

Our IViBe is a pixel-based BS method. When building the background model for each pixel, it does not rely on a temporal statistical distribution, but employs a universal sample-based method instead. Let $x_i$ be an arbitrary pixel in a video image, and $\boldsymbol{B}(x_i)$ be its background model containing $N$ background samples (values taken in the past in the same location or in the neighborhood):

$$\boldsymbol{B}(x_i) = \{B_1(x_i), \ldots, B_k(x_i), \ldots, B_N(x_i)\}. \tag{1}$$

The background model $\boldsymbol{B}(x_i)$ is first initialized from one single frame according to the intensities of pixel $x_i$ and its neighboring pixels, and then updated online when pixel $x_i$ is classified as background or by a "spatial diffusion" update mechanism.

The pixel $x_i$ is classified as a background pixel only if its current intensity $I(x_i)$ is closer than a certain distance threshold $R_k(x_i)$ $(1 \le k \le N)$ to at least $\#_{\min}$ of its $N$ background samples. Thus, the foreground segmentation mask is calculated as

$$F(x_i) = \begin{cases} 1, & \#\{|I(x_i) - B_k(x_i)| < R_k(x_i)\} < \#_{\min}, \\ 0, & \text{else.} \end{cases} \tag{2}$$

Here, $F(x_i) = 1$ signifies that the pixel $x_i$ is a foreground pixel, # denotes the cardinality of a set, $\#_{\min}$ is a fixed parameter indicating the minimal matching number, and $R_k(x_i)$ is an adaptive distance threshold according to the perception characteristics of the HVS.

In Sec. 3.1, we introduce our adaptive distance threshold and its derivation. Section 3.2 shows how our adaptive distance threshold together with the "spatial diffusion" update mechanism affects the detection results.

### 3.1 *Adaptive Distance Threshold*

In order to better segment foreground objects similar to the background, we introduce an adaptive distance threshold $R_k(x_i)$ for background matching. Different from ViBe which uses a fixed distance threshold $R_k(x_i) = 20$ for each background sample, we propose an adaptive distance metric through simulating the characteristics of human visual perception (i.e., Weber's law[35]).

Weber's law describes the human response to a physical stimulus in a quantitative fashion. The just noticeable difference (JND) is the minimum amount by which stimulus intensity must be changed in order to produce a noticeable variation in the sensory experience. Ernst Weber, a 19th century experimental psychologist, observed that the size of the JND is linearly proportional to the initial stimulus intensity. This relationship, known as Weber's law, can be expressed as

$$\Delta I_{\mathrm{JND}}/I = c, \tag{3}$$

where $\Delta I_{\mathrm{JND}}$ represents the JND, $I$ represents the initial stimulus intensity, and $c$ is a constant called the Weber ratio.

In visual perception, Weber's law actually describes the ability of the HVS for brightness discrimination, and the Weber ratio can be obtained by a classic experiment[36] which consists of having a subject look at a flat, uniformly illuminated area (with intensity $I$) large enough to occupy the entire field of view, as Fig. 1 shows. An increment of illumination (i.e., $\Delta I$) is added to the field and appears as a circle in the center. When $\Delta I$ achieves $\Delta I_{\mathrm{JND}}$, the subject will give a positive response, indicating a perceivable change. In Weber's law, $\Delta I_{\mathrm{JND}}$ is in direct proportion to $I$. Hence, the $\Delta I_{\mathrm{JND}}$ is small in dark backgrounds and big in bright backgrounds.

In the BS methods, when comparing current intensity with the corresponding background model, the distance threshold can actually be considered as the critical intensity difference in distinguishing foreground objects from the background. Fortunately, Weber's law describes the capacity of the HVS in perceiving noticeable intensity changes, and the JND that the HVS can perceive is in direct proportion to the background illumination. Inspired by Weber's law, we propose our adaptive distance threshold in direct proportion to the background sample intensity; namely, the distance threshold should be low for a dark background sample and high for a bright background sample.

In our method, mapping to Weber's law is as follows: the background sample intensity $B_k(x_i)$ can be regarded as the initial intensity $I$, the difference between the current value and each background sample is the intensity change $\Delta I$, and the distance threshold $R_k(x_i)$ can be regarded as the JND (i.e., $\Delta I_{\mathrm{JND}}$). Consequently, on the basis of Weber's law, we set

$$R_k(x_i)/B_k(x_i) = c. \tag{4}$$

In Eq. (4), $B_k(x_i)$ is the known background sample intensity, and if we want to derive the distance threshold $R_k(x_i)$, we have to first obtain the Weber ratio $c$. However, we cannot directly use the Weber ratio obtained in the classic experiment, because the classic experiment uses a uniformly illuminated area as background, but what we need in our method is a Weber ratio with a complex image as the background. As described in Ref. 37, "for any point or small area in a complex image, the Weber ratio is generally much larger than that obtained in an experimental environment because of the lack of sharply defined boundaries and intensity variations in the background." Moreover, it is also difficult to gain the Weber ratio via redoing the classic experiment using a complex image as the background, because such an experiment will need many subjects and the subjects' evaluation criteria are inconsistent, which will reduce the creditability of the experiment.

Based on the consideration above, we employ a substitute of subjective evaluations in the classic experiment to derive the Weber ratio $c$ for a complex image as the background. Specifically, the substitute is the difference of the peak signal-to-noise ratio (PSNR[38]) presented by the motion picture experts group (MPEG). The MPEG recommends that,[38] for an original reference image ($R$) and two of its reconstructed images ($D_1$ and $D_2$), only when the difference of PSNR (i.e., $\Delta$PSNR) satisfies

$$|\mathrm{PSNR}(D_1, R) - \mathrm{PSNR}(D_2, R)| \geq 0.5\,(\mathrm{dB}), \tag{5}$$

the HVS can perceive that $D_1$ and $D_2$ are different. In Eq. (5), $\mathrm{PSNR}(D, R)$ is used to estimate the level of errors in a distorted image $D$ from its original reference image $R$. For grayscale images with intensities in the range of [0, 255], $\mathrm{PSNR}(D, R)$ is defined as

$$\begin{aligned}\mathrm{PSNR}(D, R) &= 20\ \mathrm{lg}\ \frac{255}{\frac{1}{n}\|D - R\|_1} \\ &= 20\ \mathrm{lg}\ \frac{255}{\frac{1}{n}\sum_{m=1}^{n}|d_m - r_m|}\,(\mathrm{dB}),\end{aligned} \tag{6}$$

where $n$ is the number of pixels in the original image $R$, and $d_m$ and $r_m$ denote the intensities of the $m$'th pixel in $D$ and $R$, respectively.

Since $\Delta$PSNR can objectively reflect the ability of the HVS in discriminating intensity changes, we use $\Delta$PSNR to substitute the subjects' perception in the classic experiment with a complex image as the background. Here, we first construct a complex image. Suppose there is a complex image whose rows and columns are divided into 16 equal parts, respectively. Thus, the complex image is composed of 256 regions of the same size. For each region, the setup is the same as the classic experiment shown in Fig. 1. That is, each region is uniformly illuminated with intensity $I$, and an increment of illumination (i.e., $\Delta I$) is added to the centered circle. Such a region is called a basic region. The complex image consists of 256 basic regions (with $I = 0,1,\ldots,255$), which are randomly permutated, as shown in Fig. 2. In this way, we construct a complex image as the background to simulate the classic experiment in all intensity levels simultaneously, which makes our derivation general and objective.

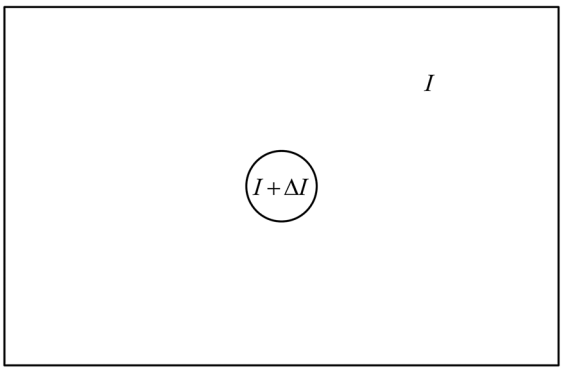


**Fig. 1** Basic experimental setup used to characterize brightness discrimination.

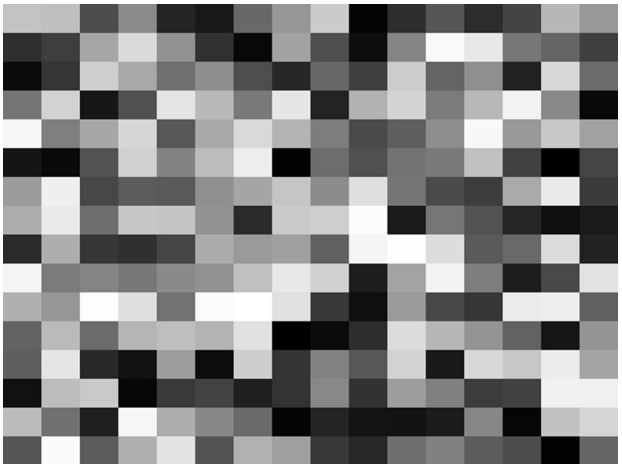

**Fig. 2** Simulated complex image as background.

All the circles in the basic regions of Fig. 2 simultaneously change their intensities with $\Delta I$. When $|\Delta I|$ reaches $\Delta I_{\mathrm{JND}}$ for all the basic regions, the HVS can barely perceive the intensity changes of the complex image (let this image be $D_1$). When $|\Delta I| = \Delta I_{\mathrm{JND}} + \varepsilon$ ($\varepsilon$ is a very small constant, and for a digital image we set $\varepsilon = 1$) for all the basic regions, the HVS can obviously perceive the intensity changes of the complex image (let this image be $D_2$). Suppose the complex image shown in Fig. 2 is the original reference image (i.e., $R$), then $D_1$ and $D_2$ can be regarded as two different distorted images which are reconstructed from the same $R$ and are just perceivably distinguishable by the HVS. Accordingly, on the basis of Eq. (3), the 1-norm of difference between $R$ and $D_1$ is given in Eq. (7), and the 1-norm of difference between $R$ and $D_2$ is provided in Eq. (8),

$$\|D_1 - R\|_1 = \sum_{I=0}^{255}\sum_{w}\Delta I_{\mathrm{JND}} = \sum_{I=0}^{255}\sum_{w} cI = w\sum_{I=0}^{255} cI, \tag{7}$$

$$\begin{aligned}\|D_2 - R\|_1 &= \sum_{I=0}^{255}\sum_{w}(\Delta I_{\mathrm{JND}} + 1) = \sum_{I=0}^{255}\sum_{w}(cI + 1) \\ &= w\sum_{I=0}^{255}(cI + 1),\end{aligned} \tag{8}$$

where $w$ denotes the number of pixels in the circle of each basic region in Fig. 2. In accordance with the recommendation of the MPEG, the difference of PSNR between these two reconstructed images ($D_1$ and $D_2$) meets equality in Eq. (5), i.e., $\Delta\mathrm{PSNR} = 0.5\,(\mathrm{dB})$, that is,

$$20\ \mathrm{lg}\ \frac{255}{\frac{w}{n}\sum_{I=0}^{255} cI} - 20\ \mathrm{lg}\ \frac{255}{\frac{w}{n}\sum_{I=0}^{255}(cI + 1)} = 0.5, \tag{9}$$

where $n$ denotes the number of pixels in the complex image.

Simplifying Eq. (9), we can derive $c = 0.13$. As a result, we conclude that the relationship between the intensity of a background sample and its corresponding distance threshold is: $R_k(x_i) = 0.13B_k(x_i)$.

Nevertheless, according to the description of brightness adaptation of the HVS in Ref. 37, we can infer that, in the extremely dark and extremely bright regions of a

complex image, the linear relationship in Weber's law cannot precisely describe the relation between perceptible intensity changes of the HVS and the background illumination. Therefore, our solution is to cut off the distance threshold for background samples whose intensities are too high or too low. After many experiments, we empirically set [10%, 90%] of the entire intensity range satisfying the linear relationship. Namely, the cut off intensities are $T_1 = \lceil 255 \times 0.1 \rceil = 26$ and $T_2 = \lceil 255 \times 0.9 \rceil = 230$. Consequently, the adaptive distance threshold can be calculated as

$$R_k(x_i) = c \ \min\{\max[B_k(x_i), T_1], T_2\}, \tag{10}$$

which is shown in Fig. 3.

### 3.2 *Background Model Update Mechanism and Impacts of Our Adaptive Distance Threshold Together with Update Mechanism on Detection Results*

It is essential to update the background model $\boldsymbol{B}(x_i)$ to adapt to changes in the background, such as lighting changes and variations of the background. The update of background models is not only for pixels classified as background, but also for their randomly selected eight-connected neighborhood. In detail, when a pixel $x_i$ is classified as background, its current intensity $I(x_i)$ is used to randomly replace one of its background samples $B_k(x_i)$ $(k \in \{1,2,\ldots,N\})$ with a probability $p = 1/\phi$, where $\phi$ is a time subsampling factor similar to the learning rate in GMM (the smaller the $\phi$ we use, the faster the update speed we get). After updating the background model of pixel $x_i$, we randomly select a pixel $x_j$ in the eight-connected spatial neighborhood of pixel $x_i$, i.e., $x_j \in N_8(x_i)$. In light of the spatial consistency of neighboring background pixels, we also use the current intensity $I(x_i)$ of pixel $x_i$ to randomly replace one of pixel $x_j$'s background samples $B_k(x_j)$ $(k \in \{1,2,\ldots,N\})$. In this way, we allow a spatial diffusion of background samples in the process of background model update.

The advantage of this "spatial diffusion" update mechanism is the quick absorption of certain types of ghosts (a set of connected points, detected as in motion but not corresponding to any real moving object[8]). Some ghosts result from removing some parts of the background; therefore, those ghost areas often share similar intensities with their surrounding background. When background samples from surrounding areas try to diffuse inside the ghosts, these samples are likely to match with current intensities at the diffused locations. Thus, the diffused pixels in the ghosts are gradually classified as background. In this way, the ghosts can be progressively eroded until they entirely disappear.

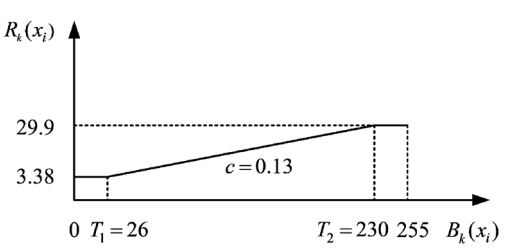


**Fig. 3** The relationship between the intensity of a background sample and its corresponding distance threshold.

However, the "spatial diffusion" update mechanism is disadvantageous for detecting still foreground objects. In environments where the foreground objects are static for several frames, either because the foreground objects share similar intensities with the background, or due to the noise inevitably emerging in the video sequence, some pixels of the foreground objects may be misclassified as background, and then serve as erroneous background seeds propagating foreground intensities in the background models of their neighboring pixels. Since foreground objects are still for several frames, the background models of the neighboring pixels of these misclassified pixels will suffer from more and more incorrect background samples coming from misclassified foreground intensities. In this way, there will be more misclassified foreground pixels, which will lead to the diffusion of misclassification.

Fortunately, our IViBe employs background matching based on an adaptive distance threshold which can reduce the misclassification inside the foreground objects, can slow down the speed of the misclassification diffusion, and can lower the eaten up speed of still foreground objects. First, IViBe makes full use of the adaptive distance threshold to enhance the discriminating power of similar foregrounds and backgrounds, and then reduces the number of misclassified foreground pixels, which can decrease the possibility of erroneous background seeds occurring. Second, even though misclassification emerges inside the foreground objects for some reason and leads foreground intensities diffusing into the background models of neighboring pixels, the adaptive distance threshold can also cut down the misclassification possibilities of those neighboring pixels inside the foreground objects. Via the aforementioned analysis, we conclude that IViBe has the ability to detect still foreground objects that are present for several frames.

Since we use the adaptive distance threshold as Eq. (10), our threshold for dark areas will be smaller than that of ViBe, hence fewer pixels will be classified as background and then be updated; whereas for bright areas, our threshold will be larger than the fixed threshold used by ViBe, and so more pixels will be classified as background and will then be updated. Accordingly, the updating probability is lower for dark areas and higher for bright areas.

## 4 Experimental Results

In this section, we first list the test sequences and determine the optimal values of parameters in our IViBe method, and then compare our results with those of ViBe in terms of qualitative and quantitative evaluations.

### 4.1 *Experimental Setup*

#### 4.1.1 *Test sequences*

In our experiments, we employ the widely used changedetection.net[26,39] (CDnet) benchmark. We select two sequences to test the capability of these techniques in coping with the camouflaged foreground objects. One sequence is called "lakeSide" from the thermal category, and the other sequence is called "blizzard" from the bad weather category. In the lakeSide sequence, two people are undistinguishable from the lake behind them in thermal imagery after they get out from the lake and have the same temperatures as the lake. This sequence is really a challenging camouflage scenario, for it is even difficult for eyes to discriminate these

people from the background. In the blizzard sequence, a road is covered by heavy snow during bad weather; meanwhile, some cars passing through are white, and some cars with other colors are partially covered by white snow, which makes correct classification difficult.

Besides, to validate the power of IViBe in coping with still foreground objects, we further choose two other typical sequences from the CDnet. One sequence is called "library" from the thermal category, and the other sequence is called "sofa" from the intermittent object motion category. In the library sequence, a man walks in the scene and selects a book, and then sits in front of a desk reading the book for a long time. In the sofa sequence, several men successively sit on a sofa to rest for dozens of frames, and place their belongings (foreground) aside; for example, a box is abandoned on the ground and a bag is left on the sofa.

Moreover, to test the performance of our method in general environments, we also select the baseline category which contains four videos (i.e., highway, office, pedestrians, and PETS2006) with a mixture of mild challenges (including dynamic backgrounds, camera jitter, shadows, and intermittent object motion). For example, the highway sequence endures subtle background motion, the office sequence suffers from small camera jitter, the pedestrians sequence has isolated shadows, and the PETS2006 sequence has abandoned objects and pedestrians that stop for a short while and then move away. These videos are fairly easy but are not trivial to process.[26]

### 4.1.2 *Determination of parameter setting*

There are six parameters in IViBe: number of background samples stored in each pixel's background model (i.e., $N$), ratio of $R_k(x_i)$ to $B_k(x_i)$ (i.e., $c$), cutoff thresholds (i.e., $T_1$ and $T_2$), required number of close background samples when classifying a pixel as background (i.e., $\#_{\min}$), and time subsampling factor (i.e., $\phi$).

In Sec. 3.1, we have determined the parameters of our adaptive distance threshold, namely $c = 0.13$, $T_1 = 26$, and $T_2 = 230$.

In order to evaluate $\#_{\min}$ and $N$ with a variety of values, we introduce the metric called percentage of correct classification[8] (PCC) that is widely used in computer vision to assess the performance of a binary classifier. Let TP be the number of true positives, TN be the number of true negatives, FP be the number of false positives, and FN be the number of false negatives. These raw data (i.e., TP, TN, FP and FN) are summed up over all the frames with ground-truth references in a video. The definition of PCC is given as follows:

$$\mathrm{PCC} = \frac{100(\mathrm{TP} + \mathrm{TN})}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}. \tag{11}$$

Figure 4 illustrates the evolution of the PCC of IViBe on the pedestrians sequence (with 800 ground-truth references) in the baseline category for $\#_{\min}$ ranging from 1 to 20. The other parameters are fixed to $N = 20$, $c = 0.13$, $T_1 = 26$, $T_2 = 230$, and $\phi = 16$. As shown in Fig. 4, when the $\#_{\min}$ increases, the PCC goes down. The best PCCs are obtained for $\#_{\min} = 1$ (PCC = 99.8310), $\#_{\min} = 2$ (PCC = 99.8324) and $\#_{\min} = 3$ (PCC = 99.7923). In our experiments, we find that for stable backgrounds like

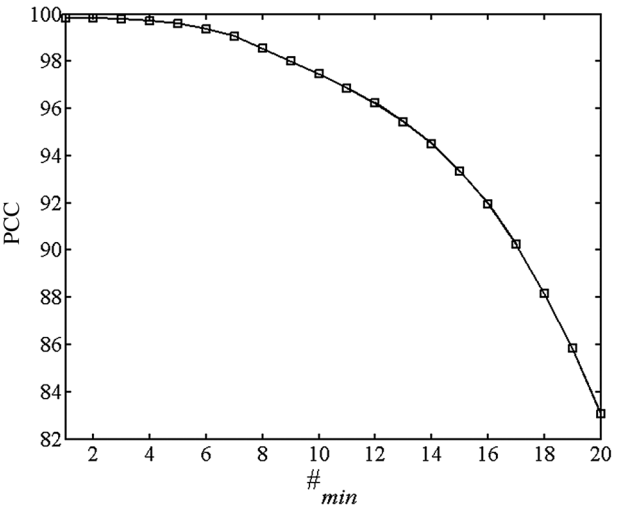


**Fig. 4** PCCs for $\#_{\min}$ ranging from 1 to 20.

those in the baseline category, $\#_{\min} = 1$ can also lead to excellent results. But in more challenging scenarios, $\#_{\min} = 2$ and $\#_{\min} = 3$ are good choices. Since a rise in $\#_{\min}$ is likely to increase the computational cost of IViBe, we set $\#_{\min} = 2$.

Once we set $\#_{\min} = 2$, we study the influence of the parameter $N$ on the performance of IViBe. Figure 5 shows the percentages obtained on the pedestrians sequence for $N$ ranging from 2 to 30. The other parameters are fixed to $\#_{\min} = 2$, $c = 0.13$, $T_1 = 26$, $T_2 = 230$, and $\phi = 16$. We observe that higher values of $N$ provide a better performance. However, the PCCs tend to saturate for $N \geq 20$. Considering that a large $N$ value will induce a large memory cost, we select $N = 20$.

The time subsampling factor $\phi$ is just like the learning rate in the GMM. A large time subsampling factor indicates a small update probability, then the background samples are unable to timely adapt to changes in the real backgrounds, such as gradual illumination changes. That is, when using a large $\phi$, there may be more false positives due to the outdated background model. On the contrary, a small $\phi$ means that the background samples are very likely to be updated according to the current frame, and a still foreground object may be

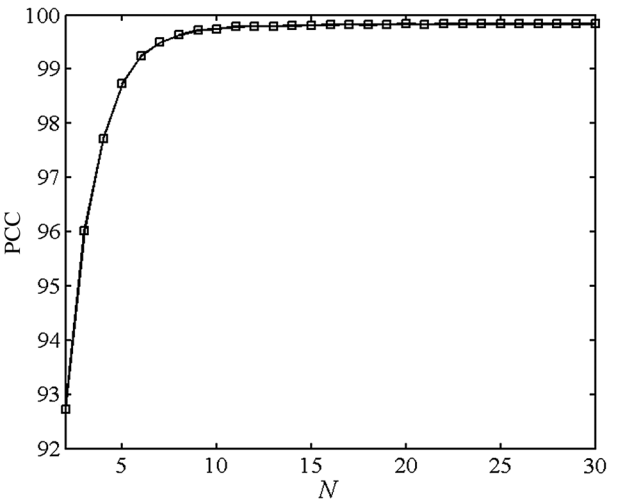


**Fig. 5** PCCs for $N$ ranging from 2 to 30.

much easier to be absorbed into the background to produce more false negatives. Hence, there is a trade-off to adjust $\phi$ in order to balance the false positives and the false negatives. Besides, $\phi$ also affects the speed of our method, because a small $\phi$ will lead to a much higher computational cost for updating. As in ViBe, we also set $\phi = 16$.

Therefore, the parameters of IViBe are set as follows: the number of background samples stored in each pixel's background model is fixed to $N = 20$; the ratio of $R_k(x_i)$ to $B_k(x_i)$ is set to $c = 0.13$; the cutoff thresholds are set to $T_1 = 26$ and $T_2 = 230$; the required number of close background samples when classifying a pixel as background is fixed to $\#_{\min} = 2$; the time subsampling factor is fixed to $\phi = 16$.

As for ViBe, the recommended parameters as suggested in Ref. 8 have been used: $N = 20$, $R_k(x_i) = 20$, $\#_{\min} = 2$, and $\phi = 16$.

#### 4.1.3 *Other settings*

For fair comparison, no postprocessing techniques (such as noise filtering, morphological operations, connected components analysis, etc.) are applied in our test for the purpose of evaluating the unaided strength of each approach.

### 4.2 *Visual Comparisons*

For qualitative evaluation, we visually compare the detection results of our IViBe with those of ViBe in Figs. 6–10 on the test sequences. Although multiple test frames were used for each test sequence, we only show one typical frame for each sequence here due to space limitation.

Figure 6 shows the detection results of the lakeSide sequence. In the input frame shown in Fig. 6(a), after swimming in the lake, the body temperatures of the people are similar to that of the lake; therefore, intensities inside the human bodies (except the heads) are almost the same as the intensities of the lake. In the detection result of ViBe shown in Fig. 6(c), we can find that the child's body is incomplete with many false negatives. This is mainly because ViBe uses a fixed distance threshold $R_k(x_i) = 20$ which is large for dark environments, and unfortunately classifies dark foreground objects as background. However, as shown in Fig. 6(d), our IViBe is able to correctly detect most of the foreground regions due to its utilization of an adaptive distance threshold based upon the perception characteristics of the HVS.

In Fig. 7, the detection results of the blizzard sequence are depicted. To show more clearly, we enlarge two areas that only contain foreground cars to illustrate the improvement of our method. The blizzard sequence is also a very challenging sequence, as shown in Figs. 7(a) and 7(e) [partial enlarged views of the Fig. 7(a)]. Because of snow fall, most of the cars appear white, which will lead to confusion between the passing cars and the road covered by thick snow. As can be seen in Fig. 7(c) and particularly in Fig. 7(g), in the detection result of ViBe, there are holes inside the detected cars, and obviously false detections appear in the areas covered by snow. In contrast to ViBe, our IViBe can discriminate subtle variations using an adaptive distance threshold, and can gain more complete detection results. As shown in Fig. 7(h), our IViBe achieves an evident improvement compared to ViBe.

Figure 8 illustrates the detection results of the library sequence. This is an infrared sequence and contains a lot of noise. In Fig. 8(a), a man is static for a long time while he sits on the chair reading a book. Because of inevitable noise, in the detection result of ViBe, misclassification

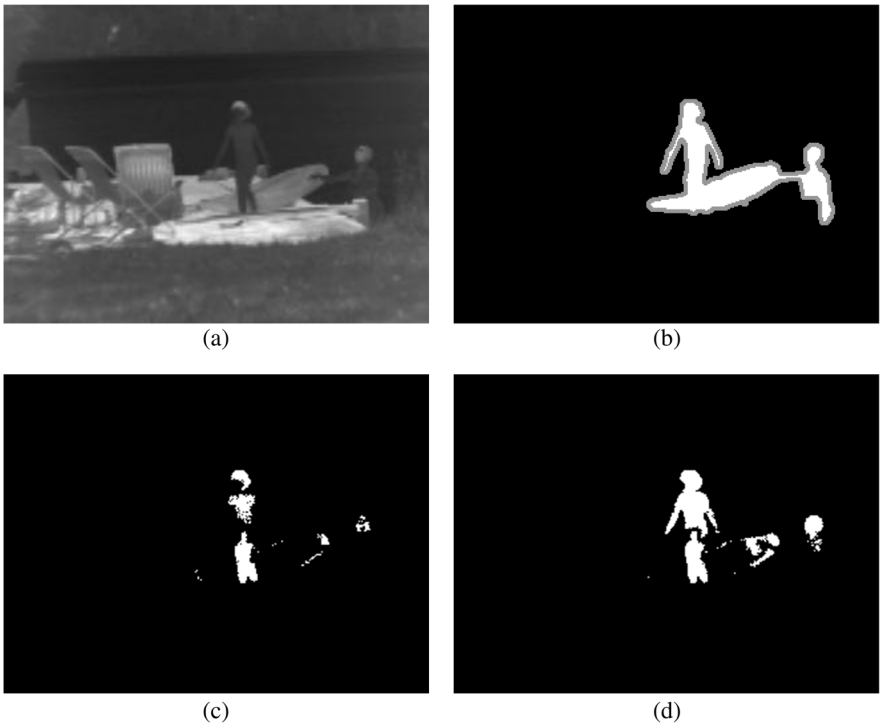


**Fig. 6** Detection results of the lakeSide sequence: (a) frame 2255 of the lakeSide sequence, (b) ground-truth reference, (c) result of ViBe, (d) result of IViBe.

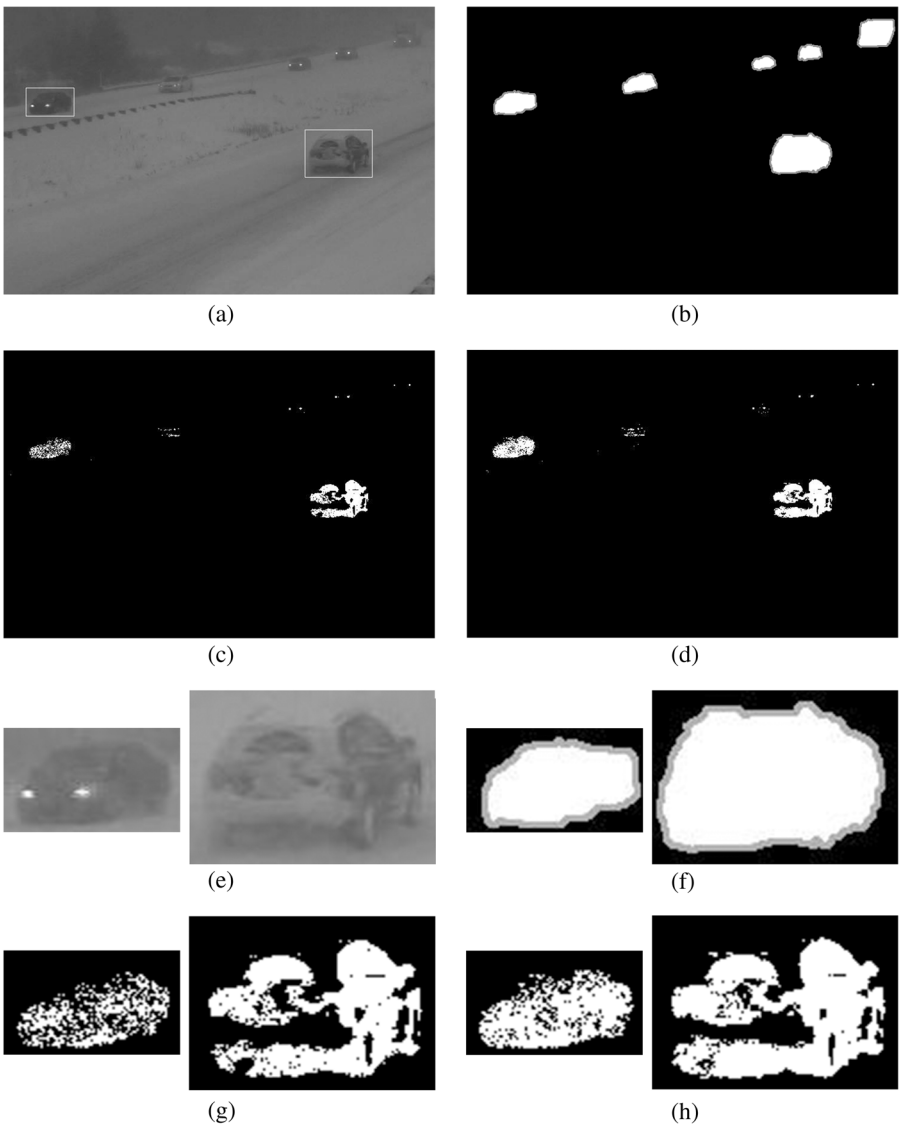


**Fig. 7** Detection results of the blizzard sequence: (a) frame 1266 of the blizzard sequence, (b) ground-truth reference, (c) result of ViBe, (d) result of IViBe, (e)–(h) partial enlarged views of (a)–(d).

emerges in the head, shoulder, and legs of the foreground, later propagates to their neighboring pixels, and finally results in large holes inside the foreground, as shown in Fig. 8(c). However, due to the adaptive distance threshold, our method yields less misclassification in the regions of head, shoulder, and legs of the foreground, and also suppresses the propagation of misclassification. These results prove that our IViBe is more powerful for detecting still foreground objects lasting for some frames in comparison with ViBe.

Figure 9 shows the detection results of the sofa sequence. In Fig. 9(a), we can find an abandoned box (foreground) which is static for a long time on the left corner. Meanwhile, a man sits on the sofa and remains still for quite an extended period. Due to the presence of noise and the adoption of a "spatial diffusion" update mechanism, in the detection result of ViBe, as shown in Fig. 9(c), the box is almost eaten-up and a large number of false negatives appear inside the man. In Fig. 9(d), a notable improvement is shown in our result: the man is more complete and the top surface of the box is well detected. This improvement is mainly the result of the adaptive distance threshold we used.

Figure 10 shows the detection results of the baseline category. For the highway sequence, our method produces more scattered false positives than ViBe in the dark areas of waving trees and their shadows, but detects more complete cars in the top right corner. For the office sequence, a man stands still for some time while reading a book, and Fig. 10(d) shows that IViBe detects more true positives in the legs of the man compared to ViBe. For the pedestrians sequence, both methods yield similar results with evident shadow areas. For the PETS2006 sequence, a man and his bag remain still for a while, and IViBe obviously detects more complete results.

### 4.3 *Quantitative Comparisons*

To objectively assess the detection results, we employ four metrics[26,39] recommended by the CDnet, i.e., recall, precision, $F1$, and percentage of wrong classification (PWC)

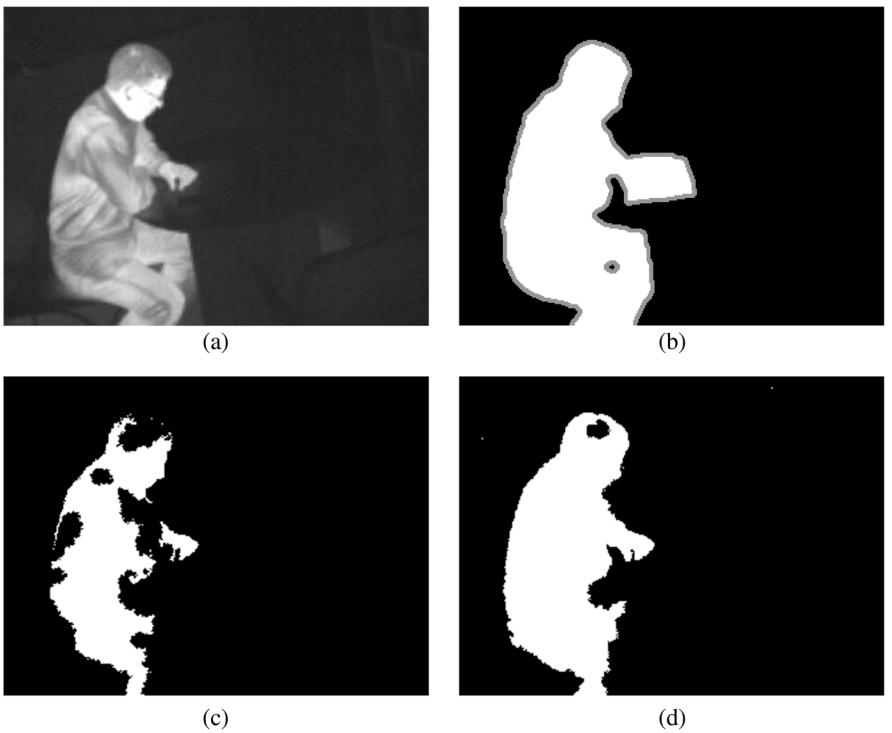

**Fig. 8** Detection results of the library sequence: (a) frame 2768 of the library sequence, (b) ground-truth reference, (c) result of ViBe, (d) result of IViBe.

to judge the performance of the BS methods on pixel level. Let TP be the number of true positives, TN be the number of true negatives, FP be the number of false positives, and FN be the number of false negatives. These raw data (i.e., TP, TN, FP and FN) are summed up over all the frames with ground-truth references in a video. For a video $v$ in a category $a$, these metrics are defined as

$$\text{recall}_{v,a} = \frac{\text{TP}_{v,a}}{\text{TP}_{v,a} + \text{FN}_{v.a}}, \tag{12}$$

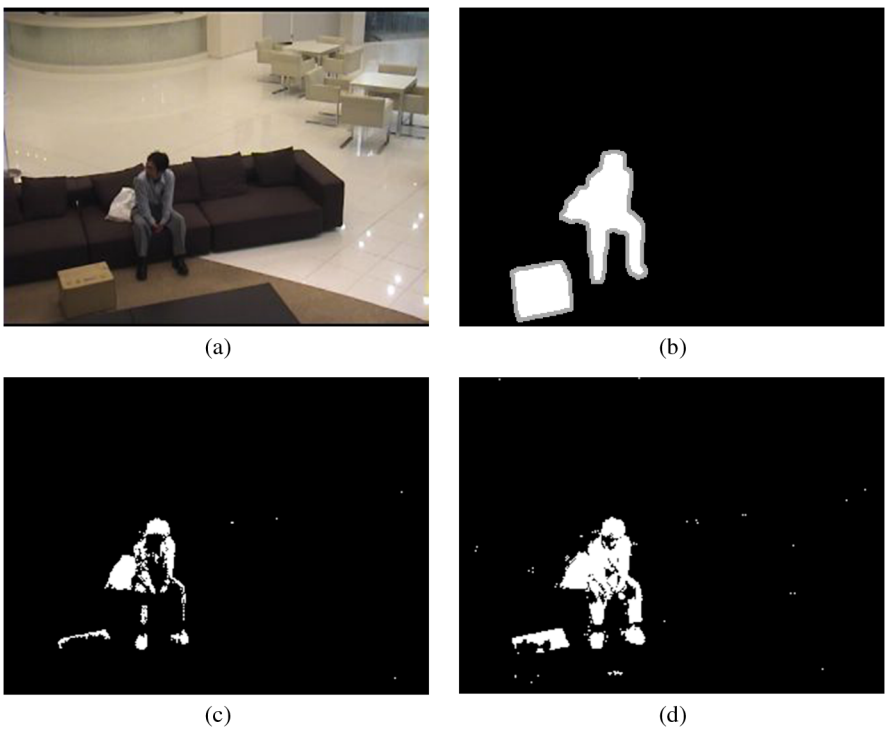

**Fig. 9** Detection results of the sofa sequence: (a) frame 900 of the sofa sequence, (b) ground-truth reference, (c) result of ViBe, (d) result of IViBe.

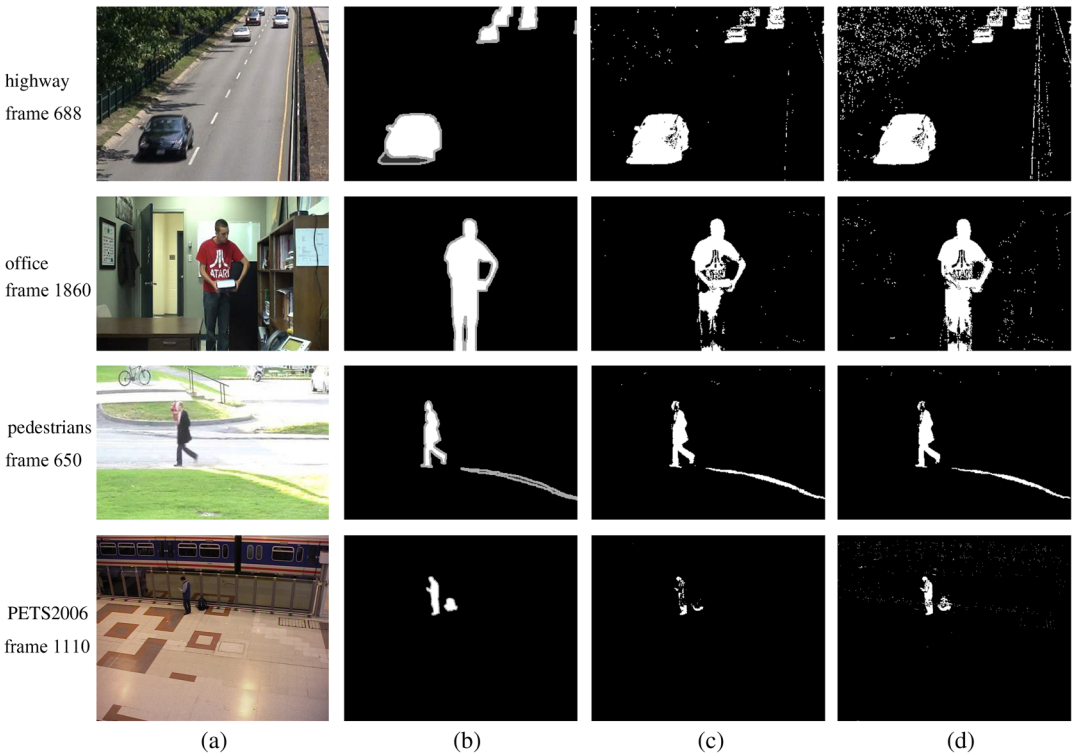


**Fig. 10** Detection results of the baseline category: (a) input frames, (b) ground-truth references, (c) results of ViBe, (d) results of IViBe.

$$\text{precision}_{v,a} = \frac{\text{TP}_{v,a}}{\text{TP}_{v,a} + \text{FP}_{v,a}}, \tag{13}$$

$$F1_{v,a} = 2\frac{\text{recall}_{v,a} \cdot \text{precision}_{v,a}}{\text{recall}_{v,a} + \text{precision}_{v,a}}, \tag{14}$$

$$\text{PWC}_{v,a} = \frac{100(\text{FN}_{v,a} + \text{FP}_{v,a})}{\text{TP}_{v,a} + \text{FN}_{v,a} + \text{FP}_{v,a} + \text{TN}_{v,a}}. \tag{15}$$

Then the average metrics of category $a$ can be calculated as

$$\text{recall}_a = \frac{1}{N_a}\sum_{v=1}^{N_a}\text{recall}_{v,a}, \tag{16}$$

$$\text{precision}_a = \frac{1}{N_a}\sum_{v=1}^{N_a}\text{precision}_{v,a}, \tag{17}$$

$$F1_a = \frac{1}{N_a}\sum_{v=1}^{N_a}F1_{v,a}, \tag{18}$$

$$\text{PWC}_a = \frac{1}{N_a}\sum_{v=1}^{N_a}\text{PWC}_{v,a}, \tag{19}$$

where $N_a$ is the number of videos in the category $a$. These metrics are called category-average metrics.

Generally, the recall (known as detection rate) is used in conjunction with the precision (known as positive prediction), and a method is considered good if it reaches high recall values without sacrificing precision.[27] Since recall and precision often contradict each other, the overall indicators ($F1$ and PWC) which integrate false positives and false negatives in one single measure are employed to further compare the results. The first three metrics mentioned above all lie in the range of [0,1], and the higher the metrics are, the better the detection results are. The PWC lies in [0,100]; here, lower is better.

Tables 1–4 show these four metrics for the lakeSide, blizzard, library and sofa sequences using ViBe and our method, respectively. These metrics are calculated utilizing all the ground-truth references available. That is, frames 1000 to 6500 in the lakeSide sequence; frames 900 to 7000 in the blizzard sequence; frames 600 to 4900 in the library sequence; frames 500 to 2750 in the *sofa* sequence.

As illustrated in Table 1, for the lakeSide sequence, the precision value of our method decreases slightly compared to that of ViBe; however, the recall value of our IViBe increases remarkably compared to that of ViBe. With regard to the overall indicators ($F1$ and PWC), our method exhibits

**Table 1** Comparison of metrics for the lakeSide sequence.

| | Recall | Precision | $F1$ | PWC |
|---|---|---|---|---|
| ViBe | 0.2224 | 0.9539 | 0.3607 | 1.5105 |
| Our method | 0.4162 | 0.9157 | 0.5723 | 1.1920 |

**Table 2** Comparison of metrics for the blizzard sequence.

| | Recall | Precision | *F*1 | PWC |
|---|---|---|---|---|
| ViBe | 0.5870 | 0.9870 | 0.7362 | 0.4906 |
| Our method | 0.6486 | 0.9762 | 0.7794 | 0.4281 |

**Table 3** Comparison of metrics for the library sequence.

| | Recall | Precision | *F*1 | PWC |
|---|---|---|---|---|
| ViBe | 0.6321 | 0.9129 | 0.7470 | 8.2532 |
| Our method | 0.7873 | 0.9484 | 0.8604 | 4.9260 |

an impressive improvement over ViBe. As seen in Table 2, for the blizzard sequence, our precision value decreases by 0.01, but our recall value increases by 0.06. For $F1$ and PWC, our method achieves a moderate improvement. As shown in Table 3, for the library sequence, our proposed IViBe produces results with all the metrics better than those of ViBe. Table 4 shows that, for the *sofa* sequence, our precision value decreases by 0.13, while our recall value increases by 0.25. For $F1$ and PWC, our method obtains a remarkable improvement. The experimental results demonstrate that, in the scenarios which contain camouflaged foreground objects, our IViBe can significantly reduce false negatives in the detection results; in the environments where the foreground objects are static for some frames, our IViBe slows the eaten-up speed of those still foreground objects in the detection results.

To calculate the category-average metrics of the baseline category, we also utilize all the ground-truth references available. That is, frames 470 to 1700 in the highway sequence; frames 570 to 2050 in the office sequence; frames 300 to 1099 in the pedestrians sequence; frames 300 to 1200 in the PETS2006 sequence. Table 5 shows the category-average metrics for the baseline category using ViBe and our method. As can be seen in Table 5, our method produces results with a larger recall and a smaller precision; however, the overall indicators ($F1$ and PWC) of both methods are quite similar.

**Table 4** Comparison of metrics for the sofa sequence.

| | Recall | Precision | *F*1 | PWC |
|---|---|---|---|---|
| ViBe | 0.4529 | 0.9078 | 0.6043 | 2.5899 |
| Our method | 0.7062 | 0.7786 | 0.7406 | 2.1599 |

**Table 5** Comparison of category-average metrics for the baseline category.

| | Recall | Precision | *F*1 | PWC |
|---|---|---|---|---|
| ViBe | 0.7888 | 0.9046 | 0.8416 | 1.1012 |
| Our method | 0.8640 | 0.8308 | 0.8425 | 1.0991 |

In general, through quantitative analysis, our IViBe method outperforms ViBe when dealing with camouflaged and still foreground objects, and has a similar performance to ViBe when dealing with normal videos with mild challenges.

## 5 Conclusion

According to the perception characteristics of the HVS concerning the minimum intensity changes under certain background illuminations, we propose an improved ViBe method using an adaptive distance threshold for each background sample in accordance with its intensity. Experimental results demonstrate that our IViBe can effectively improve the ability to deal with camouflaged foreground objects. Since the camouflaged foreground objects are ubiquitous in every real world video sequence, our IViBe has powerful practical value in smart video surveillance systems. Moreover, because of the capacity in dealing with the camouflaged foreground objects, our IViBe not only cuts down the misclassification of foreground pixels as background, but also further suppresses the propagation of misclassification, especially for those pixels inside the still foreground objects. Experimental results also prove that our method outperforms ViBe in scenarios in which foreground objects remain static for several frames.

### *Acknowledgments*

The authors would like to thank the anonymous reviewers for their insightful comments that helped to improve the quality of this paper. This work is supported by the National Natural Science Foundation of China under Grant No. 61374097, the Fundamental Research Funds for the Central Universities (N130423006), the Natural Science Foundation of Hebei Province under Grant No. F2012501001, and the Foundation of Northeastern University at Qinhuangdao (XNK201403).

### *References*

1. L. Maddalena and A. Petrosino, "The 3DSOBS+ algorithm for moving object detection," *Comput. Vis. Image Und.* **122**, 65–73 (2014).
2. L. Tong et al., "Encoder combined video moving object detection," *Neurocomputing* **139**, 150–162 (2014).
3. O. Oreifej, X. Li, and M. Shah, "Simultaneous video stabilization and moving object detection in turbulence," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(2), 450–462 (2013).
4. J. Guo et al., "Fast background subtraction based on a multilayer codebook model for moving object detection," *IEEE Trans. Circuits Syst. Video Technol.* **23**(10), 1809–1821 (2013).
5. C. Cuevas and N. García, "Improved background modeling for real-time spatio-temporal non-parametric moving object detection strategies," *Image Vis. Comput.* **31**(9), 616–630 (2013).
6. C. Cuevas, R. Mohedano, and N. García, "Kernel bandwidth estimation for moving object detection in non-stabilized cameras," *Opt. Eng.* **51**(4), 040501 (2012).
7. P. Chiranjeevi and S. Sengupta, "Moving object detection in the presence of dynamic backgrounds using intensity and textural features," *J. Electron. Imaging* **20**(4), 043009 (2011).
8. O. Barnich and M. Van Droogenbroeck, "ViBe: a universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.* **20**(6), 1709–1724 (2011).
9. S. Brutzer, B. Höferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *Proc. IEEE Computer Society Conf. on Computer Vision Pattern Recognition*, pp. 1937–1944, IEEE, Piscataway, NJ (2011).
10. Y. Benezeth et al., "Comparative study of background subtraction algorithms," *J. Electron. Imaging* **19**(3), 033003 (2010).

11. A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Comput. Vis. Image Und.* **122**, 4–21 (2014).
12. T. Bouwmans and E. H. Zahzah, "Robust PCA via principal component pursuit: a review for a comparative evaluation in video surveillance," *Comput. Vis. Image Und.* **122**, 22–34 (2014).
13. T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: an overview," *Comput. Sci. Rev.* **11–12**, 31–66 (2014).
14. A. Vacavant et al., "Special section on background models comparison," *Comput. Vis. Image Und.* **122**, 1–3 (2014).
15. T. Bouwmans et al., "Special issue on background modeling for foreground detection in real-world dynamic scenes," *Mach. Vis. Appl.* **25**(5), 1101–1103 (2014).
16. C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8), 747–757 (2000).
17. D.-S. Lee, "Effective Gaussian mixture learning for video background subtraction," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(5), 827–832 (2005).
18. Z. Zivkovic and F. Van Der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognit. Lett.* **27**(7), 773–780 (2006).
19. W. Zhang et al., "Spatiotemporal Gaussian mixture model to detect moving objects in dynamic scenes," *J. Electron. Imaging* **16**(2), 023013 (2007).
20. H. Bhaskar, L. Mihaylova, and A. Achim, "Video foreground detection based on symmetric alpha-stable mixture models," *IEEE Trans. Circuits Syst. Video Technol.* **20**(8), 1133–1138 (2010).
21. T. Elguebaly and N. Bouguila, "Background subtraction using finite mixtures of asymmetric Gaussian distributions and shadow detection," *Mach. Vis. Appl.* **25**(5), 1145–1162 (2014).
22. T. S. F. Haines and T. Xiang, "Background subtraction with Dirichlet process mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(4), 670–683 (2014).
23. A. Elgammal et al., "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proc. IEEE* **90**(7), 1151–1163 (2002).
24. Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(11), 1778–1792 (2005).
25. K. Kim et al., "Real-time foreground-background segmentation using codebook model," *Real-Time Imaging* **11**(3), 172–185 (2005).
26. N. Goyette et al., "Changedetection.net: a new change detection benchmark dataset," in *Proc. IEEE Computer Society Conf. on Computer Vision Pattern Recognition Workshops*, pp. 1–8, IEEE, Piscataway, NJ (2012).
27. L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Trans. Image Process.* **17**(7), 1168–1177 (2008).
28. L. Maddalena and A. Petrosino, "The SOBS algorithm: what are the limits?," in *Proc. IEEE Computer Society. Conf. on Computer. Vision Pattern Recognition Workshops*, pp. 21–26, IEEE, Piscataway, NJ (2012).
29. L. Maddalena and A. Petrosino, "Stopped object detection by learning foreground model in videos," *IEEE Trans. Neural Netw. Learn. Sys.* **24**(5), 723–735 (2013).
30. M. Van Droogenbroeck and O. Barnich, "Background subtraction: experiments and improvements for ViBe," in *Proc. IEEE Computer Society Conf. on Computer Vision Pattern Recognition Workshops*, pp. 32–37, IEEE, Piscataway, NJ (2012).
31. M. Van Droogenbroeck and O. Barnich, "ViBe: a disruptive method for background subtraction," *Background Modeling and Foreground Detection for Video Surveillance* Chapter 7 in T. Bouwmans, F. Porikli, B. Hoferlin, and A. Vacavant, Eds., pp. 7-1–7-23, Chapman and Hall/CRC, London (2014).
32. N. Mould and J. P. Havlicek, "A conservative scene model update policy," in *Proc. IEEE Southwest Symp. on Image Anal. Interpret*, pp. 145–148, IEEE, Piscataway, NJ (2012).
33. C. Liu, P. C. Yuen, and G. Qiu, "Object motion detection using information theoretic spatio-temporal saliency," *Pattern Recognit.* **42**(11), 2897–2906 (2009).
34. V. Mahadevan and N. Vasconcelos, "Spatiotemporal saliency in dynamic scenes," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(1), 171–177 (2010).
35. A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Upper Saddle River, NJ (1989).
36. R. C. Gonzalez and R. E. Woods, Eds., *Digital Image Processing*, 3rd ed., Prentice Hall, Upper Saddle River, NJ (2008).
37. R. C. Gonzalez and P. Wintz, *Digital Image Processing*, Addison-Wesley, Reading, MA (1977).
38. D. Salomon, Ed., *Data Compression: The Complete Reference*, 4th ed., Springer, Berlin, Germany (2007).
39. Y. Wang et al., "CDnet 2014: an expanded change detection benchmark dataset," in *Proc. IEEE Computer Society Conf. on Computer Vision Pattern Recognition Workshops*, pp. 387–394, IEEE, Piscataway, NJ (2014).



**Guang Han** is a lecturer at Northeastern University at Qinhuangdao, China. He received his B.Eng. and M.Eng. degrees from the School of Electronic and Information Engineering, Beihang University, Beijing, China, in 2005 and 2008, respectively. Now he is a PhD candidate at the College of Information Science and Engineering, Northeastern University, Shenyang, China. His current research interests include object detection and object tracking in video sequences.

**Jinkuan Wang** is a professor at Northeastern University at Qinhuangdao, China. He received his B.Eng. and M.Eng. degrees from Northeastern University, China, in 1982 and 1985, respectively, and his PhD degree from the University of Electro-Communications, Japan, in 1993. His current research interests include wireless sensor networks, multiple antenna array communication systems, and adaptive signal processing.

**Xi Cai** is an associate professor at Northeastern University at Qinhuangdao, China. She received her B.Eng. and Ph.D. degrees from the School of Electronic and Information Engineering, Beihang University, Beijing, China, in 2005 and 2011, respectively. Her research interests include image fusion, image registration, object detection, and object tracking.