2016 IEEE International Conference on Multisensor Fusion
and Integration for Intelligent Systems (MFI)
Kongresshaus Baden-Baden, Germany, Sep. 19-21, 2016

An Improved ViBe for Video Moving Object Detection Based on Evidential Reasoning*

Yun Yang, Deqiang Han, *Member, IEEE,* Jiankun Ding, Yi Yang

Abstract— **Visual Background Extractor (ViBe) is a video moving object detection method with simple implementation and fast speed. ViBe uses a detection threshold (neighborhood size) to judge whether a pixel belongs to the background or the foreground. However, in some complicated scenes, the belongingness of the pixels is ambiguous. One cannot well perform the object detection using the ViBe with a single threshold, which is a kind of hard decision without considering the uncertainty incorporated in. In this paper, we use two thresholds to describe the uncertainty in the ViBe-based color video detection, and use the evidence theory to model and handle the uncertainty. Experimental results show that the proposed approach achieves better detection performance compared with the original ViBe method.**

I. Introduction

Video moving object detection is to detect and extract the moving target from the image sequences [1]. Video moving object detection technology is a hot research direction in the field of computer vision. It is an important step for further analysis of video content such as the moving object tracking [2], recognition [3] and behavior understanding [4].

Various moving object detection methods have been proposed, e.g., the optical flow [5, 6, 7], frame difference [8] and background subtraction [9]. The optical flow method calculates the velocity field according to the image sequences to detect the foreground (i.e. the object); however, it has huge computational costs in real-time situations. The frame difference method detects the foreground by the differences between two or more adjacent frames. Although its implementation is simple and the computational cost is small, the foreground tends to contain many blank holes. A series of methods have been proposed based on background subtraction, e.g., the Gaussian Mixture Model (GMM) [10] and codebook [11], where the background pixel model is built according to the recent history frames and the foreground detection is implemented by comparing the current pixels with the corresponding pixel models. Different methods build the background model in different ways. GMM models the pixel values as a mixture of Gaussians to deal with the multi-modal appearance of the dynamic background. It has a high computational cost and the parameters estimation (e.g., the variance) of the model is problematic in a noisy environment [13]. In the codebook method, each pixel is represented by a set of code words, which contains attributes of the corresponding pixel, e.g., the occurrence frequency of the codeword, the first and last access times of the code word [11]. The codebook method needs a high cost of storage.

Olivier Barnich [12, 13] proposed a sample-based background subtraction method for video moving object detection, i.e., the Visual Background Extractor (ViBe). Compared with the other background subtraction methods, it is a faster method for real-time processing. ViBe uses the 1st frame to initialize the background model, which is a sample set of each pixel. In the sequel, it calculates the distance between the current pixel and the corresponding samples in the background model, and makes decision according to the number of pixels with distance less than the given detection threshold. However, in practice (especially in complicated scenes), the belongingness of a pixel is usually ambiguous. Furthermore, the selection of the detection threshold has subjective factors, and both large and small threshold have their own pros and cons. Thus it is not easy to select a single proper detection threshold in ViBe to classify whether the pixels belong to the background or foreground. The detection based on a single threshold here is a kind of hard decision without considering the uncertainty incorporated in.

To address the uncertainty, in this paper we use double thresholds to describe the uncertainty incorporated in the ViBe-based detection. Since the evidence theory is an effective way to deal with uncertain information, we propose an improved ViBe based on evidential reasoning to model and handle the uncertainty in double thresholds based video moving object detection. The uncertainties are modeled by using belief functions in RGB three channels, respectively. Then, the belief functions are combined according to evidence combination rules for decision making. Experimental results show that, compared with the original ViBe and some other related methods, our approach performs better.

II. Overview of Video Moving Object Detection

The purpose of video moving object detection is to extract moving objects (foreground) from image sequences. The static or slow moving, parts of the scene are the background. A binary image is usually used to show detection results:

$$result_t(x,y)=\begin{cases}255 & \mu_t(x,y)\in foreground\\ 0 & \mu_t(x,y)\in background\end{cases} \quad (1)$$

*Research supported by the Grant for State Key Program for Basic Research of China (973) (No. 2013CB329405), National Natural Science Foundation (No. 61573275), Science and technology project of SHAANXI Province (No. 2013KJXX-46), and Fundamental Research Funds for the Central Universities (Nos. xjj2014122, xjj2016066).

Yun Yang is with Institute of Integrated Automation, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049 China (e-mail: 773536462@qq.com).

Deqiang Han is with Institute of Integrated Automation, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049 China(corresponding author: +862982667972; fax: +862982668775; e-mail: deqhan@mail.xjtu.edu.cn).

Jiankun Ding is with Institute of Integrated Automation, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049 China (e-mail: d4574b@163.com).

Yi Yang is with School of Aerospace, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049 China (jiafeiyy@mail.xjtu.edu.cn).

where $result_t(x,y)$ denotes the detection result of the pixel $\mu_t(x,y)$ which is located at (x,y) in the t th frame.

In practice, the complexity of the scenes brings difficulties for the moving object detection, such as illumination variation, waving trees and the shadows of the objects. Therefore, the belongingness of a pixel to the background or foreground is usually ambiguous. The difficulties in the moving object detection can be summarized as follows:

A. The Changes of Background

If the background changes, such as illumination change, waving leaves, water ripples, etc., the changing areas would be easily detected as the foreground.

B. The Incomplete Predictability of the Object's Movement

When the moving object stops, the corresponding foreground area may be detected as the background in the following frames, then the object will disappear gradually. When a static object belonging to the background starts moving, the corresponding background area may become a ghost area [12] in the next few frames.

C. Similar Pixel Value between Foreground and Background

When the pixel values of the moving object and its surroundings are similar, the foreground pixels may be classified as the background which will cause blank holes in the detected object.

D. Shadows of the Moving Object

In some lighting scenes, shadows of the moving objects may be detected as the foreground. The shadows should be suppressed or removed timely.

E. Real-time Capability

Videos contain large amount of data. To assure the real-time implementation in practical applications, the moving object detection method should have both low time complexity and low space complexity. Now, the proposed methods are used for some specific scenes. Different methods have their own limitations as introduced in the introduction section I. It is not easy to find a real-time method, which is adaptive to any complicated scenes. ViBe approach has many advantages, e.g., the simple implementation and fast speed. We introduce ViBe in details in the next section.

III. BASICS OF VIBE METHOD

Visual Background Extractor [12] (ViBe), is a sample-based moving object detection method. First, ViBe uses the 1st frame to initialize the background model by building a sample set of each pixel. Then, it starts to detect the foreground from the 2nd frame, and it uses an random sampling strategy to update the background sample set. For color videos, ViBe can be implemented in the RGB color space. The details of each step are listed below.

A. The Initialization of the Background Sample Set

ViBe uses the 1st frame to initialize the background model where each pixel model is built by a set of samples $\mathcal{M}$.

$$\mathcal{M} = \{v_1, v_2, ..., v_N\} \quad (2)$$

where v_i, $i = 1,...,N$ denotes the value of a sample that is randomly chosen from the 8-connected neighborhood of each pixel according to the uniform law.

B. Detection of the Foreground

ViBe starts to classify a new pixel μ from the 2nd frame. ViBe counts the number of samples in the background sample set contained in a sphere $S_R(\mu)$ of radius R centered on μ (see Fig. 1). The pixel is classified as the background once the number is larger than or equal to the cardinality parameter $U_{\min}$. Otherwise, it is classified as the foreground.

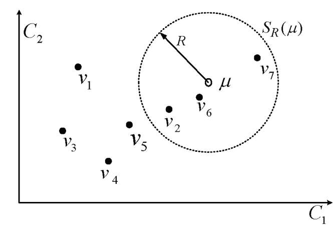


Figure 1. Example of a 2-D Euclidean color space (C_1, C_2).

C. The Background Sample Set Updating

In practice, the scene usually changes over time, e.g., illumination variation, new objects appear or water ripples. The background sample set should be updated continuously to adapt to the dynamic background.

When a pixel is detected as the background, whether this pixel value is used to update the corresponding background sample set depends on a certain probability φ, and the sample to be updated in the sample set of the current pixel model is selected randomly according to a uniform distribution. Moreover, ViBe considers that the neighboring background pixels share a similar temporal distribution. ViBe selects one out of the 8-connected neighborhood background sample sets to be updated. Whether the neighborhood background sample set will be updated or not also depends on the probability φ. This updating mechanism ensures an exponential monotonic decay for the remaining time of the background samples [13].

IV. VIBE BASED ON EVIDENTIAL REASONING METHOD

Although ViBe has many advantages as introduced above, it also has its own limitations. For example, when the scene has a complicated background, e.g., the waving leaves and the shadow of the man, ViBe may bring many false detections as shown in Fig.2.

(a) Original Image (b) The detection result

Figure 2. Example of ViBe detection result.

There exists uncertainties in the detection of a complicated background. ViBe does not perform very well in such cases (e.g., Fig. 2). ViBe uses one detection threshold R to classify a pixel. Whether the detection threshold is appropriate or not will influence the detection of the pixels. However, it is not easy to select a single proper threshold and selecting a threshold lacks objective criteria. To deal with the uncertainties, we propose an improved ViBe based on evidential reasoning (ViBe-ER). Evidence theory [15, 16] is an effective method for uncertainty modeling and reasoning, which is briefly introduced below.

A. Basics of Evidence Theory

Let $\Theta = \{\theta_1, \theta_2, \cdots, \theta_n\}$ is the frame of discernment (FOD) and $\theta_1, \theta_2, \cdots, \theta_n$ are mutually exclusive and exhaustive. 2^Θ is the power set of Θ. Function $m: 2^\Theta \to [0,1]$ is a Basic Belief Assignment (BBA, also called mass function) satisfying:

$$\sum_{A \subseteq \Theta} m(A) = 1, m(\varnothing) = 0 \tag{4}$$

If $A \subseteq \Theta, m(A) > 0$, A is called a focal element.

The Belief (Bel) and Plausibility (Pl) are defined as:

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{5}$$

$$Pl(A) = \sum_{B \cap A \neq \varnothing} m(B) = 1 - Bel(\overline{A}) \tag{6}$$

The interval $[Bel(A), Pl(A)]$ is called the belief interval, which represents the uncertainty degree of A.

Different evidences can be combined according to the evidence combination rules. Suppose that m_1 and m_2 are two mutually independent BBAs in the same FOD. The basic evidence combination rules are introduced as follows:

1) *Dempster's rule [16]*

$$m(A) = \begin{cases} 0 & A = \varnothing \\ \dfrac{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1-K} & A \neq \varnothing \end{cases} \tag{7}$$

where $K = \sum_{A_i \cap B_j = \varnothing} m_1(A_i) m_2(B_j)$ denotes the conflicting coefficient. $1-K$ is used for normalization. Dempster's rule of combination is both commutative and associative.

1) *Yager's rule* [18]

The total conflicting mass assignments is considered as unknown information and is assigned to part of $m(\Theta)$.

$$\begin{cases} m(\varnothing) = 0 \\ m(A) = \sum_{A_i \cap B_j = A \neq \varnothing} m_1(A_i) m_2(B_j) \\ m(\Theta) = m_1(\Theta) m_2(\Theta) + \sum_{A_i \cap B_j = \varnothing} m_1(A_i) m_2(B_j) \end{cases} \tag{8}$$

2) *Smets' rule* [19]

All the evidences are considered reliable. The conflicts come from the incompleteness of the FOD, so the total conflicting mass assignments is assigned to $\varnothing$.

$$m(A) = \begin{cases} \sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j) & A \neq \varnothing \\ \sum_{A_i \cap B_j = \varnothing} m_1(A_i) m_2(B_j) & A = \varnothing \end{cases} \tag{9}$$

3) *Murphy's rule* [20]

Murphy uses equivalent weights to generate average BBA:

$$m_{ave} = \frac{1}{s} \sum_{i=1}^{s} m_i \tag{10}$$

where s is the number of these evidences, then the average BBA m_{ave} is combined with itself $s-1$ times according to Dempster's rule of combination.

Pignistic probability transformation [19] is a transformation from a BBA into a probability for probabilistic decision-making:

$$BetP(\theta_i) \triangleq \sum_{\theta_i \in A, A \subseteq \Theta} \frac{m(A)}{|A|} \quad \forall \theta_i \in \Theta \tag{11}$$

where $|A|$ denotes the cardinality of A. The mass of a compound focal element will be equally assigned to each single focal element included.

B. ViBe-ER Method

A color image has different color channels (see Fig. 3). Different color channel reflects different aspect information of the image, respectively. The RGB color space, i.e., red, green and blue, is a widely used color space.

(a) Orignal Image (b) R Channel (c) G Channel (d) B Channel

Figure 3. Example of RGB channel images.

In this paper, we implement the moving object detection by fusing the different information contained in each color channel in RGB.

ViBe uses one detection threshold R to classify a pixel, so the detection result of a pixel is influenced by the detection threshold. If the detection threshold R is smaller, only those pixels which are very close to the background samples will be classified as the background. Such a result is more reliable when the pixels are classified as the background. At the same time, some other pixels belong to the background, but they may be not so close to the background samples. Then, they will be detected as the foreground using this smaller detection threshold. Therefore, the detection of foreground is unreliable to some extent. On the other hand, if the detection threshold R is larger, those pixels far from the background samples are detected as the foreground. Therefore, the result is more reliable when the pixels are detected as the foreground. However, when a new pixel belongs to the foreground but the distance between it and the background samples is lower than the larger detection threshold, it will be mis-detected as the background. Therefore, it is hard to select a single proper detection threshold. In this paper, we use a double thresholds mechanism to address the uncertainties in the detection of pixels. The smaller detection threshold R_b is strict for background detection. If a new pixel is detected as the background, this decision has a higher reliability. Another detection threshold R_f is larger, so the detection of a new pixel as the foreground has a higher reliability. That is, large and small thresholds both have their own pros and cons. Using the double detection thresholds make it possible to take advantages of both the large and small thresholds and to counteract their disadvantages.

According to the double detection thresholds R_b and R_f, the distance range $[0, 255]$ between two pixel values is divided into three intervals: $[0, R_b]$, $[R_b, R_f]$ and $[R_f, 255]$ (In Fig.4, note that the intervals here refer to 1-D distance range in a single color channel and we draw them into a 2-D figure for visual intuition). The distance between a new pixel and a sample pixel in the background sample set in the interval $[0, R_b]$ represents that the new pixel is more likely to be the background. If the distance falls into the interval $[R_f, 255]$, it represents that the new pixel is more likely to be the foreground. The interval $[R_b, R_f]$ represents the uncertainty that either the new pixel belongs to the background or foreground. By counting the numbers of samples in the background sample set whose distance between a new pixel and a background sample pixel falls into the above three intervals respectively, the mass values which support that the new pixel belongs to the background, the foreground and the uncertain interval can be generated in each color channel. Then combine the BBAs according to some evidence combination rules and transform the combined BBA to a probability for making a final decision.

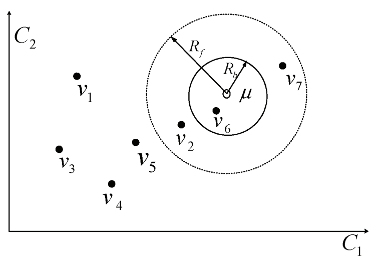


Figure 4. Example of a 2-D Euclidean color space (C_1, C_2).

The details of our proposed method are as follows.

1) Initialization of the background sample set

In each RGB channel, we use the 1st frame to initialize the background model, where each pixel model is built by a set of samples $\mathcal{M} = \{v_1, v_2, ..., v_N\}$ where v_i, $i = 1, ..., N$ denotes the sample value which is randomly selected from the 8-connected neighborhood of each pixel.

2) The generation of BBAs

The goal of the moving object detection is to classify a pixel as background or foreground, so the FOD is $\Theta = \{B, F\}$. B represents the background and F represents the foreground.

From the 2nd frame, in each RGB channel, calculate the distance between the new pixel and each sample in the sample set. Count the number N_b of samples in the background sample set whose distance is smaller than the threshold R_b, the number N_f of samples in the background sample set whose distance is larger than the threshold R_f. The number of samples in the background sample set whose distance is between R_b and R_f is $N - N_b - N_f$ (Each sample set contains N samples). Then, the BBA in each channel is generated as:

$$\begin{cases} m(B) = N_b / N \\ m(F) = N_f / N \\ m(B \cup F) = \left(N - N_b - N_f\right) / N \end{cases} \tag{12}$$

where $m(B)$ represents the support degree that a new pixel belongs to the background, $m(F)$ represents the support degree that this pixel belongs to foreground and $m(B \cup F)$ is the ignorance degree of this pixel's belongingness.

3) Evidence combination and decision making

In different color channel, we can get different BBAs. By combining them according to some combination rules introduced in the section IV, a combined BBA can be obtained. To make a decision, the combined BBA is transformed to the pignistic probability $BetP$ according to Eq. (11). If $BetP(B) \geq BetP(F)$, the new pixel is detected as the background, otherwise, it is detected as the foreground.

4) The sample set of background updating

We use the ViBe's ground update strategy. We need to update the background sample set for three channels (RGB). The implementation of ViBe-ER is illustrated in Fig. 5.

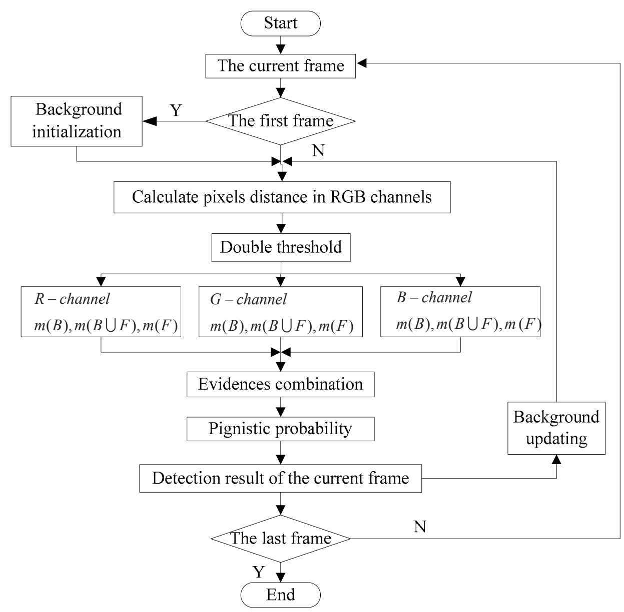
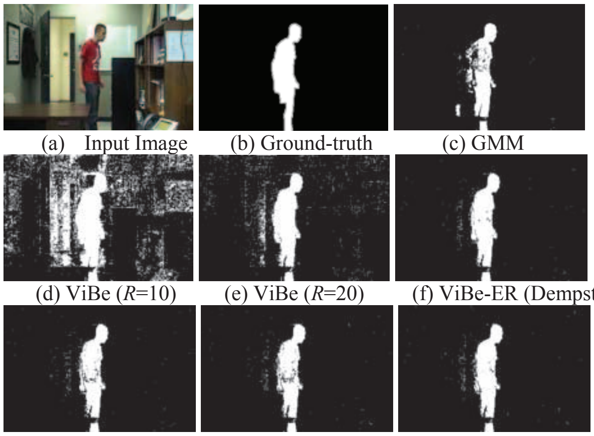


Figure 5. The block diagram of ViBe-ER.

V. EXPERIMENTS AND EVALUATION

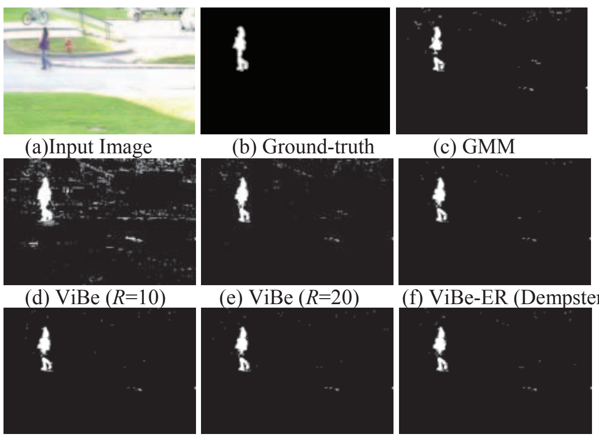
Here, we test three color videos from the dataset 2014 [21]. For better comparisons among different methods, all of the results are obtained directly without any morphological processing or connectivity analysis. In ViBe, we use the detection threshold $R = 10$ and $R = 20$ respectively, the other parameters are set as: $N = 20$, $U_{\min} = 2$ and $\varphi = 1/16$. The double thresholds of ViBe-ER is set as: $R_b = 10$, $R_f = 20$. Other parameters in ViBe-ER are set as: $N = 20$, $\varphi = 1/16$. We use Dempster's rule, Yager's rule, Smets' rule and Murphy's rule for combining BBAs.

A. Experimental Results and Analysis



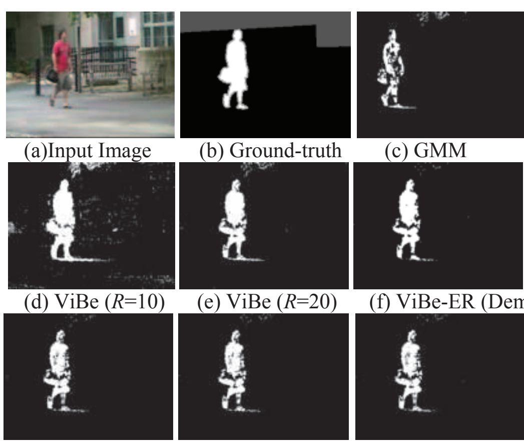
(a) Input Image (b) Ground-truth (c) GMM
(d) ViBe (R=10) (e) ViBe (R=20) (f) ViBe-ER (Dempster)
(g) ViBe-ER (Yager) (h) ViBe-ER (Smets) (I) ViBe-ER (Murphy)

Figure 6. Comparative results of the 661th frame of the Office video.



(a)Input Image (b) Ground-truth (c) GMM
(d) ViBe (R=10) (e) ViBe (R=20) (f) ViBe-ER (Dempster)
(g) ViBe-ER (Yager) (h) ViBe-ER (Smets) (I) ViBe-ER (Murphy)

Figure 7. Comparative results of the 554th frame of the Pedestrians video.



(a)Input Image (b) Ground-truth (c) GMM
(d) ViBe (R=10) (e) ViBe (R=20) (f) ViBe-ER (Dempster)
(g) ViBe-ER (Yager) (h) ViBe-ER (Smets) (I) ViBe-ER (Murphy)

Figure 8. Comparative results of the 1674th frame of the Backdoor video.

TABLE I. FEATURE DESCRIPTIONS OF THE VIDEOS

Videos	Feature Descriptions
Office	An indoor scene with changing light.
Pedestrians	An outdoor scene with uneven light.
Backdoor	An outdoor scene with waving leaves and shadows.

TABLE I lists the feature descriptions of the three scenes.

We can see that when the detection thresholds $R=10$ and $R=20$ in ViBe, respectively, a lot of pixels are incorrectly classified as the foreground. In the office video, when light changes indoor, ViBe cannot adapt to the change well, especially in the case of $R=10$, a large area of background is incorrectly detected as the foreground. The detection result of GMM is imcomplete with some blank holes. In the pedestrians' video, some bright areas of the background is also detected as the foreground. In the backdoor video, the waving leaves and the shadow under the object's feet are incorrectly detected as the foreground in the results of ViBe. There are apparent blank holes in the result of GMM. In all the three scenes, the background results of ViBe are not clean, i.e., there exists many false detections.

Using our proposed ViBe-ER method, we can see that the false detections are reduced significantly. ViBe-ER can obtain a relatively clean background with different evidence combination rules. It can adapt to the light changes, and reduce the wrong detection of waving leaves and supress the shadow to some extent. All the results show that the detection performance of a single detection threshold in ViBe is not very well since one threshold cannot deal with the uncertainties in the detection of pixels. ViBe-ER uses double thresholds to model the uncertaities and then use evidence resasoning to supress the uncertainties. We can obtain a better detection performance.

B. Evaluation

The moving object detection is as a binary classification problem. According to [22], a most widely used metric in computer vision to assess the performance of a binary classifier is the percentage of correct classification (PCC). In this paper, we use PCC to assess the performances of different methods. The PCC represents the accuracy of the classification result. The method has a better performance if its PCC is higher. It involves four parameters:

- TP (True Positives): Number of pixels correctly classified as foreground;
- TN (True Negatives): Number of pixels correctly classified as background;
- FP (False Positives): Number of pixels incorrectly classified as foreground;
- FN (False Negatives): Number of pixels incorrectly classified as background.

$$\mathrm{PCC} = \frac{TP+TN}{TP+TN+FP+FN} \tag{14}$$

The evaluation results are listed in TABLE II. We can see that ViBe-ER outperforms ViBe. ViBe-ER can also outperform both GMM and CodeBook to some extent. When R is 10, the PCC of ViBe is the lowest in all the results. Comparing with both $R=10$ and $R=20$ of ViBe, all the PCCs of ViBe-ER by using different evidence combination rules are higher than ViBe. The PCC of ViBe-ER by using the Dempster's rule of combination is the highest. Besides the improvement of the PCC, the main advantage of ViBe-ER is that it does not need to make a big effort to select a perfect detection threshold. By using two thresholds, the uncertainty during the detection process can be addressed. Therefore, ViBe-ER is more proper and it can improve the overall detection accuracy. Note that although ViBe-ER will increase the computational cost, it is only a bit higher than ViBe. In our experiment, ViBe costs about 62.13ms to process one-frame and ViBe-ER costs about 81.39ms.

TABLE II. COMPARATIVE VALUES OF PCC

Methods	Videos		
	Office	Pedestrians	Backdoor
GMM	0.9417	0.9760	0.9576
CodeBook	0.9343	0.9733	0.9558
ViBe (R=10)	0.7283	0.9254	0.8310
ViBe (R=20)	0.9208	0.9710	0.9450
ViBe-ER (Dempster)	***0.9545***	***0.9773***	***0.9582***
ViBe-ER (Yager)	0.9469	0.9760	0.9537
ViBe-ER (Smets)	0.9486	0.9758	0.9537
ViBe-ER (Murphy)	0.9513	0.9760	0.9544

VI. CONCLUSIONS

In this paper, we propose an improved ViBe approach based on evidential reasoning called ViBe-ER. A double thresholds mechanism is used to address the uncertain information during the ViBe-based color object detection. The BBA is used to model the uncertainties, and the evidence combination is used for handling the uncertainty by jointly using the information provided by different color channels. Experimental results show that ViBe-ER can adapt to changes in the background, reduce false detections, and suppress the shadow effectively. At the same time, it will not bring much more computational cost.

Note that our proposed improved ViBe approach can only be used in the object detection problems for color video, since it needs the information provided by the R, G and B color channels. In the future work, we attempt to generalize our idea to grey video.

REFERENCES

[1] Q. Wang and S. B. Zhao, "Comparison and Analysis of Video-based Moving Object Detection Algorithms," *Command Control & Simulation*, vol. 34, no. 6, pp. 36-40, Dec. 2012.

[2] Z Qu, Q Zhang, T Gao, "Moving Object Tracking Based on Codebook and Particle Filter," *Procedia Engineering*, vol. 29, no. 4, pp. 174-178, 2012.

[3] SG. Salve, KC. Jondhale, "Shape matching and object recognition using shape contexts," Pami, vol. 9, no. 4, pp. 471-474, 2015.

[4] S. Vishwakarma, A. Agrawai, "A survey on activity recognition and behavior understanding in video surveillance," Visual Computer, vol. 29, no. 10, pp. 983-1009, 2012.

[5] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1, pp. 185-203, Aug. 1981.

[6] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *The 7th International Joint Conference on Artificial Intelligence*, Vancouver, BC, Canada, Aug. 1981, pp. 674-679.

[7] Daniel D. Doyle, Alan L. Jennings and Jonathan T. Black, "Optical Flow Background Subtraction for Real-Time PTZ Camera Object Tracking," *IEEE International Instrumentation and Measurement Technology Conference*, Minneapolis, Minnesota, USA, May. 2013.

[8] R. Collins, T, Lipton, and T. Kanade, "A system for video surveillance and monitoring: VSAM final report," Technical report CMU-RI-TR-00-12, 2000.

[9] Y. Benezeth, P. M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Review and Evaluation of Commonly-Implemented Background Subtraction Algorithm," *International Conference on Pattern Recognition,* Tampa, Florida, USA, Dec. 2008.

[10] C. Stauffer, W. E. L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Fort Collins, Colorado, USA, 1999.

[11] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Background modeling and subtraction by codebook construction," *International Conference on Image Processing*, Singapore, Oct. 2004, pp. 3061-3064.

[12] O. Barnich and M. V. Droogenbroeck, "ViBe: a powerful random technique to estimate the background in video sequences," *International Conference on Acoustics, Speech and Signal Processing,* Taipei, Apr. 2009, pp. 945-948.

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

[14] X. C. Yang and W. P. Liu, "Moving object detection technology in video sequences," *Computer Applications and Software*, vol. 25, no. 1, pp. 215-217, Jan. 2008.

[15] A. P. Dempster, "Upper and lower probabilities induced by a multiple valued mapping, " *The Annals of Mathematical Statistics*, vol. 38, no. 2, pp. 325-339, 1967.

[16] D. Q. Han, Y. Yang, and C. Z. Han, "Advances in DS evidence theory and related discussions," (in Chinese) *Control and Decision,* vol. 29, no.1, pp. 1-11, Jan, 2014.

[17] C. Z. Han, Y. H. Zhu, and Z. S. Duan, *Multi-source Information Fusion (Second Edition)*, Beijing: Tsinghua university press, 2010, pp. 82-92.

[18] R. R. Yager, "On the dempster-shafer framework and new combination rules," *Information Sciences*, vol. 41, no. 2, pp. 93-137, 1987.

[19] P. Smets, and R. Kennes, "The transferable belief model," *Artificial Intelligence*, vol. 66, no. 2, pp. 191-234, 1994.

[20] C. K. Murphy, "Combining belief functions when evidence conflicts," *Decision Suport Systems*, vol. 29, no. 1, pp. 1-9, Jul. 2000.

[21] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "An Expanded Change Detection Benchmark Dataset," *IEEE Workshop on Change Detection (CDW-2014) , CVPR*, Columbus, Ohio, USA, 2014, pp. 387-394.

[22] S. Elhabian, K. El-Sayed, and S. Ahmed, "Moving object detection in spatial domain using background removal techniques-State-of-art," *Recent Pat. Comput. Sci.*, vol.1, pp. 32-54, Jan. 2008.