

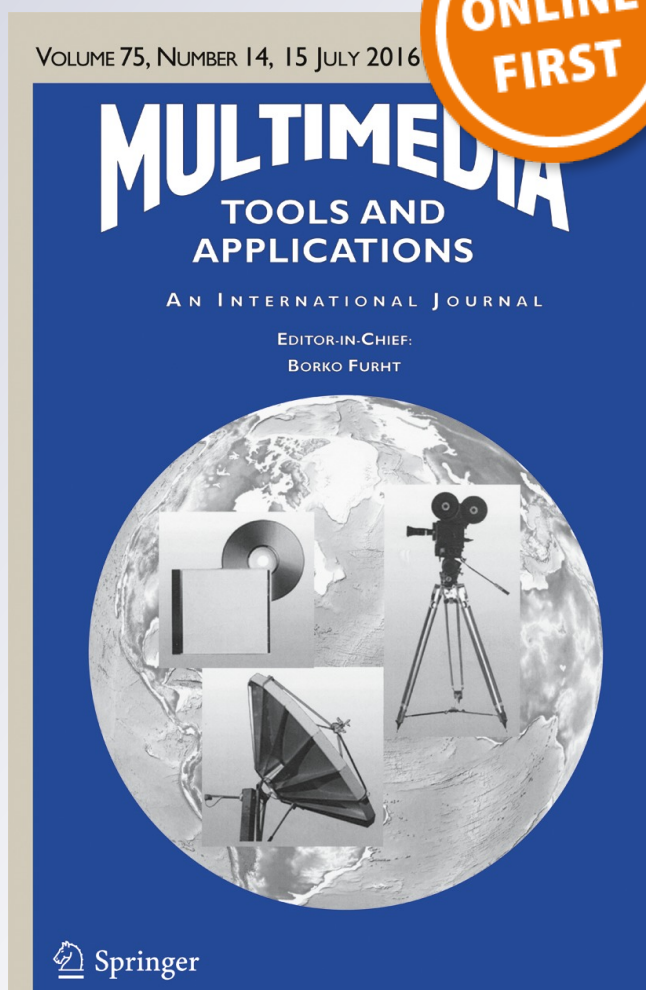
# *Local-to-global background modeling for moving object detection from non-static cameras*

**Aihua Zheng, Lei Zhang, Wei Zhang,  
Chenglong Li, Jin Tang & Bin Luo**

**Multimedia Tools and Applications**  
An International Journal

ISSN 1380-7501

Multimed Tools Appl  
DOI 10.1007/s11042-016-3565-1



**Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# Local-to-global background modeling for moving object detection from non-static cameras

Aihua Zheng<sup>1</sup> · Lei Zhang<sup>1</sup> · Wei Zhang<sup>1</sup> ·  
Chenglong Li<sup>1</sup> · Jin Tang<sup>1,2</sup> · Bin Luo<sup>1,2</sup>

Received: 18 November 2015 / Accepted: 28 April 2016  
© Springer Science+Business Media New York 2016

**Abstract** This paper investigates efficient and robust moving object detection from non-static cameras. To tackle the motion of background caused by moving cameras and to alleviate the interference of noises, we propose a local-to-global background model for moving object detection. Firstly, motion compensation based local location-specific background model is deployed to roughly detect the foreground regions in non-static cameras. More specifically, the local background model is built for each pixel and represented by a set of pixel values drawn from its location and neighborhoods. Each pixel can be classified as foreground or background pixel according to the compensated background model based on the fast optical flow. Secondly, we estimate the global background model by the rough superpixel-based background regions to further separate foregrounds from background accurately. In particular, we use the superpixel to generate the initial background regions based on the detection results generated by local background model to alleviate the noises. Then, a Gaussian Mixture Model (GMM) is estimated for the backgrounds on superpixel level to refine the foreground regions. Extensive experiments on newly created dataset, including 10 challenging video sequences recorded in PTZ cameras and hand-held cameras, suggest that our method outperforms other state-of-the-art methods in accuracy.

**Keywords** Background modeling · Object detection · Non-static cameras · Motion compensation · Random algorithm · Superpixel processing · Gaussian mixture model

---

Part of work in the manuscript has been accepted and published on CCCV 2015.

✉ Aihua Zheng  
ahzheng214@gmail.com

<sup>1</sup> School of Computer Science and Technology, AnHui University, Hefei 230601, China

<sup>2</sup> Key Laboratory of Industry Image Processing and Analysis in Anhui Province, Hefei 230601, China

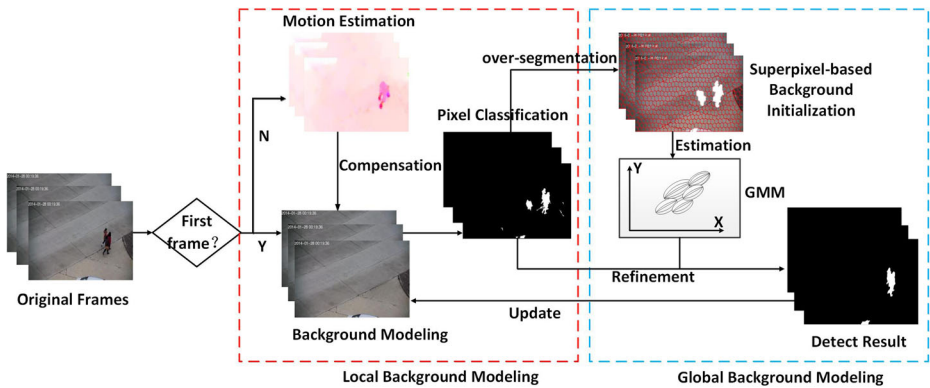
## 1 Introduction

Moving object detection is the fundamental and crucial task in video surveillance [26], and has a broad prospect of applications and research values in the intelligent transportation [24], medical diagnosis [23], security monitoring [8], and many other industries [11]. Generally speaking, moving object detection methods can be divided into two categories, i.e., static background (fixed camera) and non-static background (moving camera). At present, moving object detection in static background has become an increasingly mature technique and many related technologies have been successfully applied to real life in recent decades, such as ViBe [16, 25] and Gaussian Mixture Model (GMM) [12, 27]. However, moving object detection with non-static cameras is still a challenging problem due to the mixed background changing with the object moving.

There are several kinds of techniques to address such problem. 1) Motion compensation based background modeling. This kind of methods usually compensated the camera motion by computing optical flow of pixels [18, 32] or geometric transformation between adjacent frames [17, 29]. Then, they employed the static-background models to detect moving objects. These methods required a lot of computation time while introduced many noises by motion estimation or background models. Our local background model will follow this direction in an efficient way, and develop a global background model to alleviate aforementioned noises. 2) Motion segmentation. These methods [8, 15, 17, 19] utilize different motions of objects from background's to separate them, but usually are failed when objects and background have similar motions. 3) Trajectory clustering. Other researchers [5, 30, 31, 33] proposed to detect moving objects by clustering the trajectories of all pixels or keypoints into foregrounds or background. These methods processed the videos in a batch way, and were time consuming.

In this paper, we propose a local-to-global background modeling framework to online detect the moving objects from non-static cameras efficiently and robustly. Firstly, in the stage of local background modeling, the motion compensation algorithm is utilized to accommodate the motion of the background. Specifically, the background model of each pixel consists of a set of pixels, which are initialized by its location and neighbors. When new frame arriving, the optical flow algorithm, based on edge-preserving patch matching [4], is employed to their background models from previous frame to current one. Then, every pixel can be classified as the foreground or background pixel by the matching score with their background models. Furthermore, the background models are probabilistically updated in an online fashion to adapt the variation of background. Secondly, due to the noises and false detections, the detection results in the first stage may be unconfident. Therefore, the global background modeling is developed. In particular, we use the superpixel, which can alleviate the interference of noises effectively [1], to generate the initial background regions based on the detection results in first stage. The GMM is then estimated to build the global appearance model of background. We utilize the GMM to refine the detection results produced in the first stage to suppress the effects of background. Figure 1 shows the flowchart of our framework.

The key contributions of this paper are summed up in three aspects. Firstly, a local-to-global background modeling framework is proposed for detecting moving objects with non-static cameras, in which both the local and global information is taken into account for background modeling. Secondly, a robust local background model based on motion compensation is developed and updated by a random algorithm to adapt the local motion and variation of background over time. Thirdly, the global background modeling is developed to refine the



**Fig. 1** Flow chart of our framework

superpixel-based foreground/background regions by GMM. Extensive experiments on newly created dataset suggest that our method outperforms other state-of-the-art methods in accuracy.

This paper is organized as follows. Section 2 reviews some relevant methods moving object detection in non-static cameras, and Section 3 presents the details of our framework. Experimental results and related analysis are shown in Section 4 and Section 5 concludes this paper.

## 2 Related Works

This paper focuses on moving object detection with non-static cameras, and some relevant state-of-the-art methods on moving object detection from non-static cameras have been reviewed in this section.

The first kind of approaches is motion compensation based background modeling [18, 20, 29, 32, 35]. Patel and Parmar [18] used a bilateral filter followed by calculation of motion vector using Lucas - Kanade algorithm to provide accurate and efficient results with the moving background. Zhang et al. [32] used the optic flow as a motion compensation to adaptively update the motion of background. However, the optic flow may introduce lots of noise and false detection. Recently, low-rank matrix approximation based background modeling has been developed [20, 35]. These approaches proposed a moving object detection framework to address several complex scenarios, such as non-rigid motion and dynamic background. They assumed that the transformation between consecutive frames was linear and thus utilized the 2D parametric transforms [29] to model translation, rotation, and planar deformation of the background. Experimental results demonstrated that they can achieve state-of-the-art performance. However, these methods either model the background only on pixel level so they are still sensitive to noise which is ubiquitous [13], or involve many complex operations, which are computationally inefficient.

The second kind of approaches is motion segmentation. Some researchers also proposed to segment the motion to separate the background and the foreground [10, 15, 17, 19]. Schoenemann and Cremers [19] proposed an energy minimization method based on graph-cut to decompose a video into a set of super-resolved moving layers. Kim et al. [8] proposed a fast moving object detection method using optical flow clustering and Delaunay triangulation. Narayana and Hanson [15] develop a segmentation algorithm that clusters pixels that have similar real-world motion irrespective of their depth in the scene according to the optical flow

orientations instead of the complete vectors and exploits the well-known property that under camera translation. Papazoglou and Ferrari [17] proposed a technique for fully automatic video object segmentation in unconstrained settings. However, the aforementioned methods are based on the assumption that the motions between the moving objects and the background are obviously distinguishable. Therefore they fail in the situation that the objects have the similar motion as the background.

The third kind of approaches is trajectory clustering [5, 30, 31, 33]. Zeppelzauer et al. [31] proposed a clustering scheme for spatio-temporal segmentation of sparse motion fields obtained from feature tracking, to segment the meaningful motion components in a scene, such as short and long-term motion of single objects, groups of objects and camera motion. Yu et al. [30] proposed a so called CTraStream clustering algorithm which contains trajectory line segment stream clustering stage and online trajectory cluster updating stage for trajectory data stream. Brox and Malik [5] used long term point trajectories based on dense optical flow to keep the temporal consistent segmentations of moving objects in a video shot. Zhang et al. [33] proposed a video object segmentation system based on a conditional random field model with high-order term which is capable of capturing longer-range spatial and temporal grouping information to take advantages of both point trajectories and region trajectories. However, most of existing methods are time-consuming and rely on the trajectories through time.

### 3 Our Approach

In this section, we present our approach in a detailed way. Firstly, the local background is modeled based on the motion compensation to accommodate the motion of non-static cameras. Then, global superpixel-level background is modeled by GMM to suppress the effects of background and alleviate the noises.

#### 3.1 Local Background Modeling

The main challenge of moving object detection in non-static cameras is the motion of the background. Different from the conventional modeling method for static background, the desired modeling strategy should adapt the background changes. In this paper, the motion of each pixel will be accurately estimated to propagate to their background model to accommodate the motion of the camera.

##### 3.1.1 Motion estimation

Most of existing methods on dense optical flow are computationally inefficient [21]. However, a fast optical flow algorithm based on edge-preserving PatchMatch is recently proposed by Bao et al. [3] with high accuracy and efficiency. Therefore, we employ the edge-preserving PatchMatch optical flow to estimate the optical flow in this work, and briefly review it as follows.

The edge-preserving PatchMatch optical flow is a fast algorithm that employs approximate the nearest neighbor field [16] to handle the large displacement motions and consists of four steps: matching cost computation, correspondence approximation, occlusions and outliers handling, and sub-pixel refinement.

1. Matching Cost Computation. The edge-preserving PatchMatch optical flow follows the traditional local correspondence searching framework [4]. To make the nearest neighbor field preserve the details of the frame, it employs bilateral weights [28] into matching cost calculation, and can be defined as

$$d(a, b) = \frac{1}{W} \sum_{\Delta} w(a, b, \Delta) C(a, b, \Delta) \tag{1}$$

where  $a$  and  $b$  denote two pixels, and  $\Delta$  indicates patches center on  $a$  and  $b$ ,  $W$  is a normalization factor.  $w(\cdot)$  is the bilateral weighting function and  $C(\cdot)$  is the robust cost between  $a$  and  $b$ . More detailed definitions please refer to [3].

2. Correspondence Approximation. To produce high-quality flow fields, this optical flow method utilizes self-similarity propagation and a hierarchical matching scheme to approximate the exact edge-preserving Patch Match [4]. Firstly, self-similarity propagation algorithm is based on the fact that adjacent pixels tend to be similar to each other. Specifically, for each pixel, a set of pixels from its surrounding region is randomly selected and stored into a self-similarity vector in the order of their similarities to the center pixel. Then, its adjacent pixels' vector is merged into its own vector from top-left to bottom-right. This process is reversely repeated. Thanks to the propagation between adjacent pixels, the algorithm can produce reasonably good approximate results in a much faster speed. Secondly, a hierarchical matching scheme is employed to further accelerate the algorithm and similar with Simple Flow method [22].
3. Occlusions and Outliers Handling. FEPM explicitly performs forward-backward consistency check [7] between the two NNFs to detect occlusion regions. Moreover, a weighted median filtering is performed [2] on the flow fields to remove outliers.
4. Sub-pixel Refinement. The edge-preserving PatchMatch optical flow produces sub-pixel accurately with a more efficient technique - paraboloid fitting, which is a 2D extension from the 1D parabola fitting [28].

### 3.1.2 Location-specific background model

Compared with the background models of a static background, the background modeling in dynamic background is difficult to maintain online since the background pixels are also moving. Although estimated optical flow can compensate the background motion, the background model is still sensitive to noises due to incorrect optical flow estimation. Thus, a robust pixel model of background is proposed in this paper to adaptively detect the moving objects in the dynamic background.

For each input video, the first frame is selected to initialize the background model. The background model of each pixel is a set of pixel values, and can be represented as

$$B(p) = \{I(p_1), I(p_2), \dots, I(p_n)\} \tag{2}$$

where  $p_i \in N(p)$ , and  $N(\cdot)$  indicates the neighbors of pixel  $p$ .  $I(\cdot)$  denotes the pixel value. For each pixel,  $n$  samples are selected from itself and its neighboring pixel values to initialize its background model.

### 3.1.3 Pixel classification

Given the background model of previous frame, it can be propagated to the current frame by employing the motion estimation algorithm. Then, every pixel of current frame can be classified as the foreground or background pixel according to the matching scores with their corresponding background model.

For one pixel  $p$ , the matching score with background  $B(p)$  is defined as

$$M(p) = \sum_{i=1}^n \delta(D(I(p), B_i(p)) > R) \tag{3}$$

where  $\delta(\cdot)$  denotes the indicator function, and  $R$  indicates the adaptive threshold of matching cost, which is determined by the variation  $\sigma$  of  $B(p)$ . Herein,  $\sigma$  indicates the complexity of the background, and  $R$  is defined as

$$R = \begin{cases} 20, & \sigma/2 \leq 20 \\ \sigma/2, & 20 < \sigma/2 < 40 \\ 40, & \sigma/2 \geq 40 \end{cases} \tag{4}$$

Then,  $p$  can be classified by

$$U(p) = \begin{cases} 0, & M(p) \geq T_1 \\ 1, & M(p) < T_1 \end{cases} \tag{5}$$

where 0 and 1 indicate the background and foreground, respectively,  $T_1$  denotes the threshold of matching score,  $U$  denotes the detection results.

### 3.1.4 Model update

In this section, we assume that each pixel has been accurately classified by our local-to-global background model (the details are discussed in next section) when new frame arriving. Thus, the background model of each pixel can be updated online by randomly selecting the classified background pixels at the same location or its neighbors. Specifically, for one classified background pixel  $p_b$ , two robust background model updating strategies are adapted to obtain its background model  $B(p_b)$ .

Firstly, one element from  $B(p_b)$  is selected in a uniform probability way to replace  $p_b$ . Secondly, one pixel value is heuristically taken from its neighbors  $N(p_b)$ , and substituted by the element randomly selected in  $B(p_b)$ . Herein, we assume that if one pixel belongs to its background model, its distance to all the values of the background model should be as close as possible. This assumption will be helpful to suppress the effect of the noises. Thus, the selected probability of pixel  $p_b^i$  from  $N(p_b)$  is defined as

$$q_i = \frac{1}{Q} \exp\left(-\frac{1}{n} \sum_{j=1}^n D(I(p_b^i), B_j(p_b))\right), \tag{6}$$

where  $D(\cdot, \cdot)$  denotes the Euclidian distance function, and  $Q$  is a normalization factor.

In addition, to accommodate the change speed of the background, the up-dating probability, called as updating factor and denoted as  $\eta$  in this paper, is introduced to determine whether the above updating is carried out or not.



### 3.2 Global Background Modeling

In order to alleviate the influence of the noises and the false detection caused by the motion estimation and the local background model, we further deploy the global background model to refine the detecting results produced by the local background model.

#### 3.2.1 Superpixel-based background initialization

Superpixel is to generate the irregular and compact connected regions based on the color similarity and can better keep the boundary information of the objects. Therefore, we firstly over-segment the input video frame into a superpixel set [1] as  $\{S_i\}$ . Given the rough detection results in Sect. 3.1.3, we generate the superpixel-level background regions to estimate the global background model. For  $i$ -th superpixel  $S_i$ , we identify it as foreground or background by the ratio  $r(S_i)$ , which is defined as

$$r(S_i) = \frac{1}{|S_i|} \sum_{p \in S_i} U(p) \tag{7}$$

where  $|S_i|$  denotes the total number of pixels in  $S_i$ . From Eq. (26), we can see that  $r(S_i)$  measures the number of foreground pixels respect to the total number of pixels in one superpixel.  $S_i$  is determined to be foreground or background by a threshing operator:

$$F(S_i) = \begin{cases} 1, & r(S_i) \geq T_2 \\ 0, & r(S_i) < T_2 \end{cases} \tag{8}$$

where  $F(\cdot)$  is an indicator function. 1 and 0 indicate the foreground and the background, respectively,  $T_2$  is a predefined threshold. In this way, we generate some initial superpixel regions of background, which will be utilized to estimate the global background model in next section.

#### 3.2.2 Background GMM estimation

In order to effectively resist the false detection (the background superpixel is falsely detected as the foreground), we estimate the GMM of background appearance according to the initial background regions. We first compute the average color values of all background superpixels, and then use these average color values to estimate the GMM for background appearance. The GMM is defined as follows:

$$G(\mathbf{x}) = \sum_{k=1}^K \pi_k N(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{9}$$

where  $N(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  is Gaussian function with mean  $\boldsymbol{\mu}_k$  and variance  $\boldsymbol{\Sigma}_k$ .  $\pi_k$  is weight of  $k$ -th component of GMM, and  $K$  denotes the number of components. Given the computed average colors of background, Eq. (13) can be efficiently learned by EM algorithm [6].

### 3.2.3 Foreground refinement

The foreground refinement is deployed on the basis of estimated background GMM. Each foreground superpixel will be examined by the estimated GMM to remove the false detections. For each foreground superpixel  $S_i$ , we identify it as background region if it has high probability belonging to background appearance model:

$$F(S_i) = 0, \text{ if } G(\bar{S}_i) \geq T_3 \quad (10)$$

where  $\bar{S}_i$  denotes the average color of  $S_i$ , and  $T_3$  is a predefined threshold. Figure 2 demonstrates the effectiveness of our global background model on two typical video sequences. From which we can see that the superpixel initialization (as shown in Fig. 2d) can restore the hollow produced by the local background model and alleviate the interference of the noises but may introduce false foregrounds. The global background model can effectively remove the false foregrounds and generate the satisfying detection results (as shown in Fig. 2e).

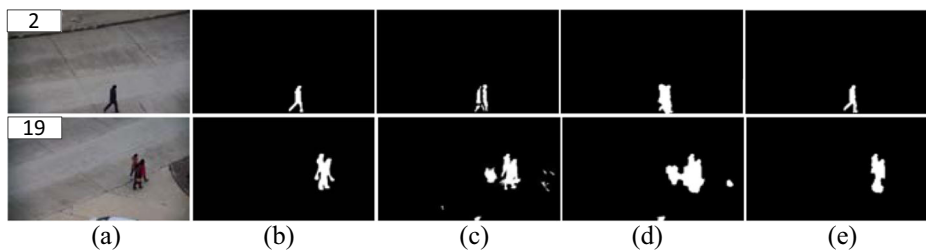
## 4 Experimental Results

In this section, we compare our approach with other state-of-the-art approaches, followed by the discussion of the effectiveness analysis, component analysis and the efficiency analysis of our approach.

### 4.1 Evaluation Settings

The test videos are the real-life videos recorded from the university security monitoring system by PTZ cameras and hand-held cameras with resolution of 320\*180 and frame rate 25fps. The evaluation is performed on 10 challenging videos containing 40,000 frames in total with vary moving objects, including single and multiple pedestrians, animals, cars, motorcycles and bicycles in non-static background of the road, playground or woods. In general, the testing video dataset has taken into account of varying environments with varying number, size, types and speed of moving objects as well as the camera movement and can comprehensively evaluate the performance of the proposed detection algorithm with others.

We compare our approach to the state-of-the-art moving object detection methods including DECOLOR [35], ViBe [16], fastSeg [17], GMM [27]. To make the comparison more comprehensive, the related parameters are empirically fixed as followings:  $n$ , the number of



**Fig. 2** Illustration of detection results generated by our local-to-global background model. **a** Original frames. **b** Ground truths. **c** Detection results by local background model. **d** Superpixe-based initialization. **e** Detection results by global background model

samples to initialize the background model for each pixel, and the updating factor  $\eta$  during the local background modeling, is fixed as 20 and 0.2 respectively. The thresholds  $T_1$ ,  $T_2$  and  $T_3$ , which are for background/foreground classification during the local background modeling, the background initialization and the foreground refinement during the global background modeling are set as 2, 0.2 and  $8 * e-7$  respectively. In summary, we fixed  $\{n, \eta, T_1, T_2, T_3, K\} = \{20, 0.2, 2, 0.2, 8 * e-7, 5\}$  in all evaluations.

### 4.2 Comparison Results

Table 1 presents the average Precision (P), Recall (R) and F-measure on 10 types collected video sequences while the detailed F-measure values on each video sequence are shown in Fig. 3a. The F value is calculated as:

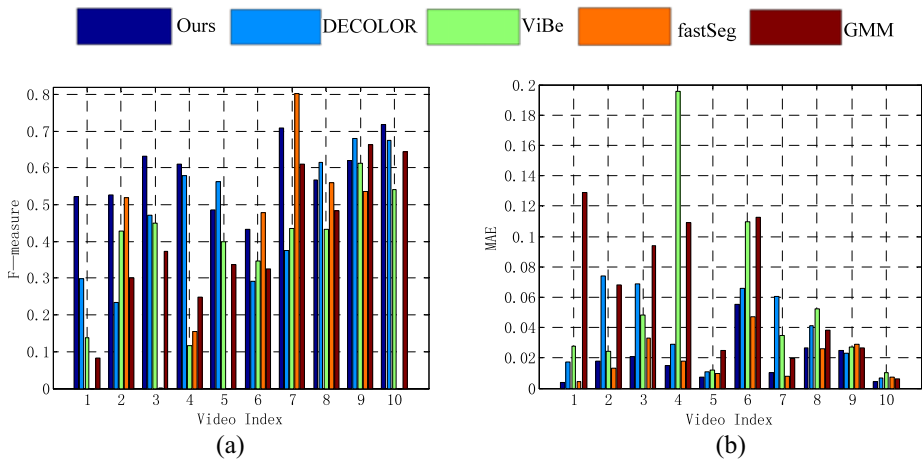
$$F\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

First of all, our method significantly outperforms other state-of-the-art with outstanding Precision in most of videos and with highest average precision. Noted that the false detection, where lots of background may be detected as foreground, may significantly effects recall, as well as the precision. In fact from the following example detection results we can see that the higher recall of other methods produced by the sacrifice of introducing lots of false detection. That's why the recall of our method is relatively lower. The F-measure is acknowledged as the comprehensive indicator to balance the precision and recall, therefore, from the F-measure values in Table 1 and Fig. 3a, we can conclude that our method significantly outperforms the compared methods in most of videos. DECOLOR has outstanding performance in Video 5, 8 and 9. However, the assumption that objects are with small size limits its application on bigger size objects such as cars and crowds although the sparse representation is with satisfying

**Table 1** The P, R and F-measure values on 10 collected video sequences

		1	2	3	4	5	6	7	8	9	10	Aver
P	Our	<b>0.60</b>	0.50	<b>0.69</b>	<b>0.60</b>	<b>0.57</b>	<b>0.64</b>	0.72	0.65	0.67	0.66	<b>0.63</b>
	DECOLOR [35]	0.19	0.15	0.32	0.45	0.45	0.53	0.23	0.46	0.57	0.54	0.39
	ViBe [16]	0.08	0.36	0.37	0.06	0.35	0.28	0.32	0.35	0.56	0.41	0.31
	fastSeg [17]	0	<b>0.68</b>	0	0.59	0	0.92	<b>0.78</b>	<b>0.69</b>	<b>0.73</b>	0	0.44
	GMM [27]	0.05	0.20	0.30	0.16	0.27	0.28	0.54	0.55	0.57	0.67	0.36
R	Our	0.50	0.58	0.64	0.64	0.44	0.34	0.72	0.51	0.60	0.77	0.57
	DECOLOR [35]	<b>0.82</b>	0.69	<b>0.93</b>	<b>0.88</b>	<b>0.76</b>	0.21	0.96	<b>0.92</b>	<b>0.94</b>	<b>0.90</b>	<b>0.80</b>
	ViBe [16]	0.53	0.55	0.68	0.67	0.47	<b>0.46</b>	0.75	0.58	0.70	0.83	0.62
	fastSeg [17]	0	0.46	0	0.10	0	0.32	<b>0.83</b>	0.47	0.43	0	0.26
	GMM [27]	0.61	<b>0.80</b>	0.83	0.79	0.59	0.42	0.78	0.48	0.82	0.68	0.68
F	Our	<b>0.52</b>	<b>0.53</b>	<b>0.63</b>	<b>0.61</b>	0.49	0.43	0.71	0.57	0.62	<b>0.72</b>	<b>0.58</b>
	DECOLOR [35]	0.30	0.23	0.47	0.57	<b>0.56</b>	0.29	0.38	<b>0.61</b>	<b>0.70</b>	0.67	0.48
	ViBe [16]	0.14	0.43	0.45	0.12	0.40	0.35	0.43	0.43	0.61	0.54	0.39
	fastSeg [17]	–	0.52	–	0.15	–	<b>0.48</b>	<b>0.80</b>	0.56	0.54	–	0.31
	GMM [27]	0.08	0.30	0.37	0.25	0.34	0.32	0.61	0.48	0.66	0.64	0.41

The bold digits denote the highest value of each measurement P, R, and F-measure



**Fig. 3** F-measure value and MAE value for each video sequence

performance in real applications [9, 34]. Though fastSeg did excellent job on Video 6 and 7, it totally fails on Video 1, 3, 5, 10 since the detection is based on the boundary exaction of objects according to their motion diversity to the background, which results in detection failure when the relative movement between the object(s) and background is small. ViBe and GMM works marvelously in static background but blunder in dynamic background since it’s hard for them to balance the background updating through the video sequence.

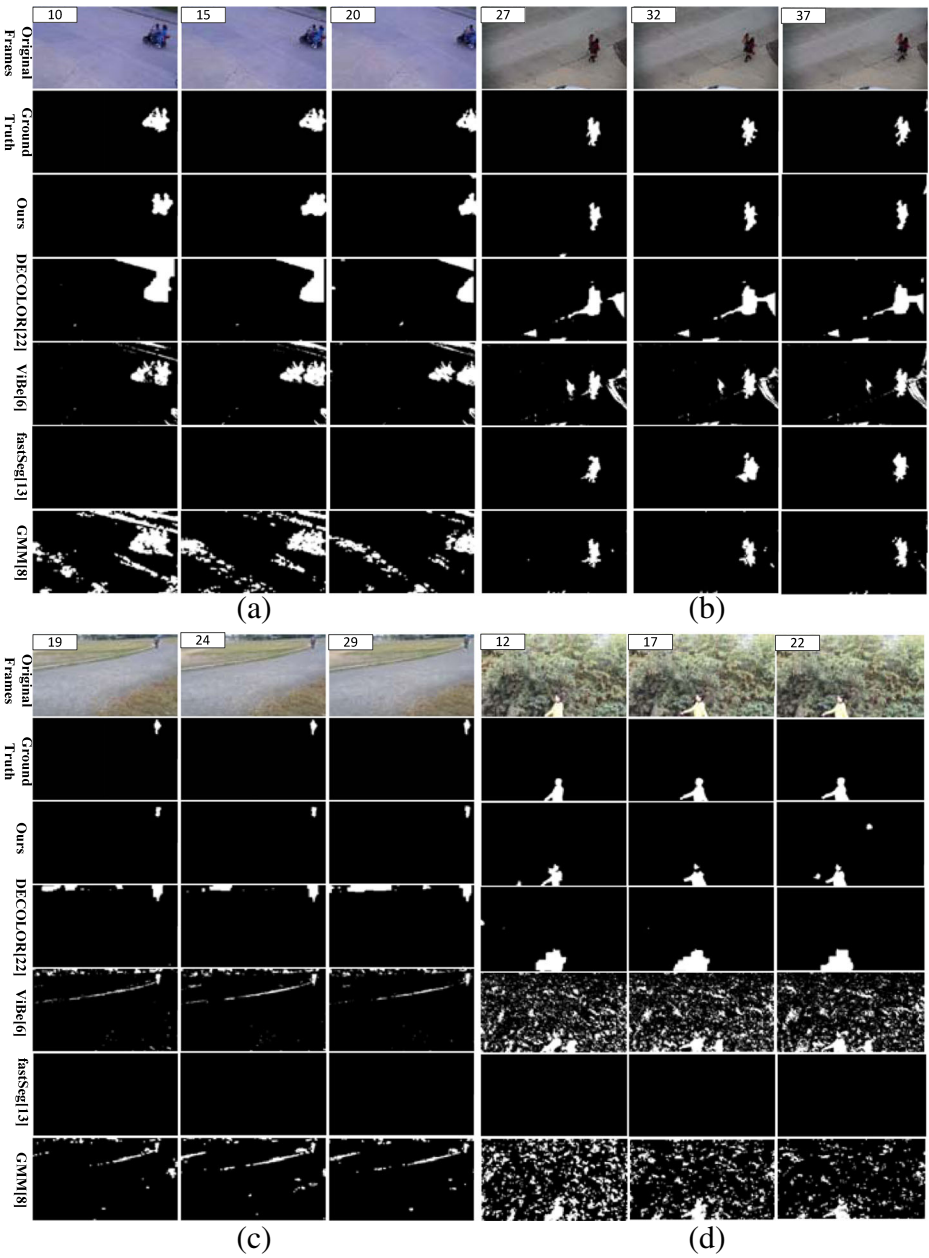
Furthermore, we investigate the Mean Absolute Error (MAE) of our approach on each video sequence as shown in Fig. 3b.

$$MAE = \frac{1}{mn} \sum_{i=1}^n \sum_{p \in F_i, p' \in G_i} XOR(p, p') \quad (12)$$

where the pixels  $p, p' \in \{0, 1\}$  (0, 1 indicate the background and foreground respectively),  $F_i$  and  $G_i$  indicate the detection result and the ground truth of  $i$ -th frame.  $XOR(\bullet)$  is the “exclusive-or” operation,  $n$  is the frame number of a video sequence and  $m$  is the pixel number of one frame.

In general, from the F-measure and MAE value we can see the outperformance of our method compared to the other state-of-the-arts, which demonstrates the robustness of the proposed framework.

To qualitatively demonstrate the effectiveness of our proposed method against other four methods, we present four sample results, as shown in Fig. 4. DECOLOR segments moving object in image sequence using a framework that detects the outliers to avoid complicated calculation, and uses low rank model to deal with complex background. It's easier to detect relatively dense and continuous region from the group. However, due to the smooth assumption of DECOLOR, more than one closed objects, especially in some occlusions, usually are detected as one single object (Fig. 4a and b). ViBe produces ghost (Fig. 4b and d) and lots of noises (Fig. 4c and d). fastSeg has fine performance on Fig. 4b but fails on the other 3 examples since it cannot successfully initialize the object due to the small relative movement between the objects and background, which is also the limitation of fastSeg. GMM works worst for dynamic background and also introduces lots of noise. In general, from the comparative experiments, we can see that the proposed



**Fig. 4** Comparison results of 4 sample video sequences, where the number on the left corner on each frame indicate the frame index

method outperforms DECOLOR in the details of objects, especially in the case of multiple objects, beats fastSeg when the motion between objects and background is not significantly distinguishable, and is robust to the background interference compared with ViBe and GMM.

**Table 2** Component analysis of our approach

	P	R	F
Ours	<b>0.6316</b>	0.5659	<b>0.5828</b>
Ours-I	0.5960	0.4672	0.5030
Ours-II	0.4770	<b>0.6661</b>	0.5320

The bold digits denote the highest value of each measurement P, R, and F-measure

### 4.3 Component Analysis

In order to validate the effectiveness of the components in our approach, we evaluate the components of superpixel over-segmentation and global background modeling separately and report the results on Table 2, where the P (Precision), R (Recall) and F (F-measure) denote the average values on 10 collected video sequences. Ours-I: without superpixel over-segmentation (the global background modeling is built directly on pixels), Ours-II: without global background modeling. From which we can see that both superpixel refinement and global background modeling play important roles to our model, justifying that both components are vital.

### 4.4 Efficiency Analysis

The experiments are carried out on a desktop with an Intel i7 3.4GHz CPU and 32GB RAM, and implemented on C++ platform without any optimization. In the above experiments, the average runtime of proposed method is 0.17 s per frame while DECOLOR is 20 s per frame. Therefore, the proposed method can substantially meet online detection with 8fps while DECOLOR is a batch method. ViBe costs 0.02 s per frame, but it can only handle weak jitter problem of the camera, and is not suitable for the situation of dynamic background. fastSeg costs 0.4 s exclude the computation of optical flow and over-segmentation, however, it's limited to the situation that objects has distinguish motion to the background. GMM costs 0.002 s which is fastest and well suited for background modeling but stumble for dynamic background.

## 5 Conclusions

In view of the problems of moving object detection in non-static cameras, this paper proposed local and global background modeling based object detection method. The local background model of each pixel is initialized according to the first frame and is propagated to current frame by employing the edge-preserving optical flow algorithm to estimate the motion of each pixel. Each pixel can be finally classified as foreground or background pixel according to the compensated background model which is updated online by the fast random algorithm. Then, the global background model on the superpixel-level is estimated by GMM to alleviate the noise and refine the detecting results during the local background modeling stage. The comparisons with state-of-the-art methods demonstrated the effectiveness of the proposed method. In future works, we will focus on developing more robust moving object detection approaches such as multi-resolution analysis [14].

**Acknowledgments** Our thanks to the support from the National Nature Science Foundation of China (61502006, 61472002), the Natural Science Foundation of Anhui Province (1508085QF127) and the Natural Science Foundation of Anhui Higher Education Institutions of China (KJ2014A015).

## References

- Achanta R, Shaji A, Smith K et al (2012) SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell* 34(11):2274–2282
- Bao L, Song Y, Yang Q, et al (2012) An edge-preserving filtering framework for visibility restoration. 21st IEEE International Conference on Pattern Recognition (ICPR), pp 384–387
- Bao LC, Yang QX, Jin HL (2014) Fast edge-preserving PatchMatch for large displacement optical flow. *IEEE Trans Image Process* 23(12):4996–5006
- Barnes C, Shechtman E, Finkelstein A et al (2009) PatchMatch: a randomized correspondence algorithm for structural image editing. *ACM Trans Graph* 28(3):341–352
- Brox T, Malik J (2010) Object segmentation by long term analysis of point trajectories. In: Proc. European Conference on Computer Vision, vol 6315. pp 282–295
- Dempster AP, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc* 39(1):1–38
- Hosni A, Rhemann C, Bleyer M et al (2013) Fast cost-volume filtering for visual correspondence and beyond. *IEEE Trans Pattern Anal Mach Intell* 35(2):504–511
- Kim J, Wang X, Wang H et al (2013) Fast moving object detection with non-stationary background. *MultimedTools Appl* 67(1):311–335
- Li C, Hu S, Gao S, Tang J (2016) Real-time grayscale-thermal tracking via Laplacian sparse representation. In: Proceedings of International Conference on Multimedia Modeling
- Li C, Lin L, Zuo W, et al (2015) SOLD: sub-optimal low-rank decomposition for efficient video segmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 5519–5527
- Liang Z, Wang M, Zhou X et al (2014) Salient object detection based on regions. *Multimed Tools Appl* 68(3):517–544
- Lin LL, Chen NR (2011) Moving objects detection based on gaussian mixture model and saliency map. *Appl Mech Mater* 2011(63–64):350–354
- Miao Q, Cao Y, Xia G, et al (2015) RBoost: label noise-robust boosting algorithm based on a nonconvex loss function and the numerically stable base learners. *IEEE Trans Neural Netw Learn Syst* 2015:1
- Miao Q, Shi C, Xu P et al (2011) A novel algorithm of image fusion using shearlets. *Opt Commun* 284(6): 1540–1547
- Narayana M, Hanson A et al (2013) Coherent motion segmentation in moving camera videos using optical flow orientations. *IEEE International Conference on Computer Vision (ICCV)*, pp 1577–1584
- Olivier B, Marc V (2011) Vibe: a universal background subtraction algorithm for video sequences. *IEEE Trans Image Process* 20(6):1709–1724
- Papazoglou A, Ferrari V (2013) Fast object segmentation in unconstrained video. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp 1777–1784
- Patel MP, Parmar SK (2014) Moving object detection with moving background using optic flow. *IEEE conference on Recent Advances and Innovations in Engineering (ICRAIE)*, pp 1–6
- Schoenemann T, Cremers D (2008) High resolution motion layer decomposition using dual-space graph cuts. *Proc IEEE Conf Comput Vis Pattern Recognit* 2008:1–7
- Shakeri M, Zhang H (2015) COROLA: a sequential solution to moving object detection using low-rank Approximation. *arXiv preprint arXiv:1505.03566*
- Stauffer C, Grimson WEL (1999) Adaptive background mixture models for realtime tracking. In: *CVPR. Proceedings of CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 2–2246
- Tao M, Bai J, Kohli P, et al (2012). SimpleFlow: a non-iterative, sublinear optical flow algorithm. *Computer graphics forum. Blackwell Publishing Ltd* 31(21):345–353
- Toennies K, Rak M, Engel K (2014) Deformable part models for object detection in medical images. *Biomed Eng Online* 13(suppl1):911–916
- Unzueta L, Nieto M, Barandiaran J et al (2012) Adaptive multi-cue background subtraction for robust vehicle counting and classification. *IEEE Trans Intell Transp Syst* 13(2):527–540

25. Vanogenbroeck M, Paquot O (2012) Background subtraction: experiments and improvements for ViBe. *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2012 I.E. Computer Society Conference, pp 32–37
26. Vieux R, Jenny BP, Domenger JP et al (2012) Segmentation-based multi-class semantic object detection. *Multimed Tools Appl* 60(2):305–326
27. Xu K, Zeng XL, Yan G (2012) Research on moving object detection based on improved gaussian mixture background model. *Sci Mosaic* 2012:12–15
28. Yang Q, Yang R, Davis J, et al (2007) Spatial-depth super resolution for range images. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 1–8
29. Yoon KJ, Kweon IS (2006) Adaptive support-weight approach for correspondence search. *IEEE Trans Pattern Anal Mach Intell* 28(4):650–656
30. Yu Y, Wang Q, Wang X, et al (2012) Trajectory stream mining framework facing to real time query processing. *Chin J Sci Instrum* 33(12)
31. Zeppelzauer M, Zaharieva M, Mitrovic D et al (2010) A novel trajectory clustering approach for motion segmentation. *Lect Notes Comput Sci* 2010(5916):433–443
32. Zhang W, Li CL, Zheng AH, et al (2015). Motion compensation based fast moving object detection in dynamic background. *Computer vision*, Vol. 547. Springer Berlin Heidelberg, pp 247–256
33. Zhang G, Yuan Z, Liu Y et al (2015) Video object segmentation by integrating trajectories from points and regions. *Multimed Tools Appl* 74(21):9665–9696
34. Zhang L, Zhou WD, Li FZ (2013) Kernel sparse representation-based classifier ensemble for face recognition. *Multim Tools Appl* 74(1):123–137
35. Zhou X, Yang C, Yu W (2012) Moving object detection by detecting contiguous outliers in the low-rank representation. *IEEE Trans Pattern Anal Mach Intell* 35(3):597–610



**Aihua Zheng** received her B.Eng. degrees and finished her Master-Doctor combined program in computer science and technology from Anhui University of China in 2006 and 2008 respectively. And received her Ph.D degree in computer science from the University of Greenwich of UK in 2012. After that, she started working with Anhui University till present. Her main research areas are temporal logic, visual based signal processing and pattern recognition.





**Lei Zhang** received his B.E. degree in computer science in 2012 from West Anhui University, Lu'an, China. He is currently pursuing the M.S. degree in computer science and technology at Anhui University. His current research is object segmentation.



**Wei Zhang** received her B.E. degree in computer science and technology in 2012 from Hefei University, Hefei, China. She is currently pursuing the M.S. degree in computer science and technology at Anhui University. Her current research is object detection in the dynamic scenes.



**Chenglong Li** received the B.S. degree in applied mathematics in 2010, and the M.S. degree in computer science in 2013 from Anhui University, Hefei, China. He is currently pursuing the Ph.D. degree in computer science at Anhui University. His current research interests include computer vision, machine learning, and intelligent media technology.



**Jin Tang** received the B.Eng. degree in automation in 1999, and the Ph.D. degree in computer science in 2007 from Anhui University, Hefei, China. Since 2013, he has been a professor at the School of Computer Science and Technology at the Anhui University. His research interests include image processing, pattern recognition, machine learning, and computer vision.



**Bin Luo** received his B.Eng. degree in electronics and M.Eng. degree in computer science from Anhui University of China in 1984 and 1991, respectively. From 1996 to 1997, he worked as a British Council visiting scholar at the University of York under the Sino-British Friendship Scholarship Scheme (SBFSS). In 2002, he was awarded the Ph.D. degree in Computer Science from the University of York, the United Kingdom. His current research interests include graph spectral analysis, large image database retrieval, image and graph matching, statistical pattern recognition, digital watermarking and information security.