Cogn Comput (2017) 9:180–193
DOI 10.1007/s12559-017-9449-5

CLASS: Collaborative Low-Rank and Sparse Separation for Moving Object Detection

Aihua Zheng[1] · Minghe Xu[1] · Bin Luo[1] · Zhili Zhou[2] · Chenglong Li[1]

Received: 25 May 2016 / Accepted: 2 January 2017 / Published online: 6 February 2017

Abstract Low-rank models have been successfully applied to background modeling and achieved promising results on moving object detection. However, the assumption that moving objects are modelled as sparse outliers limits the performance of these models when the sizes of moving objects are relatively large. Meanwhile, inspired by the visual system of human brain which can cognitively perceive the physical size of the object with different sizes of retina imaging, we propose a novel approach, called Collaborative Low-Rank And Sparse Separation (CLASS), for moving object detection. Given the data matrix that accumulates sequential frames from the input video, CLASS detects the moving objects as sparse outliers against the low-rank structure background while pursuing global appearance consistency for both foreground and background. The sparse and the global appearance consistent constraints are complementary but simultaneously competing, and thus CLASS can detect the moving objects with different sizes effectively. The smoothness constraints of object motion are also introduced in CLASS for further improving the robustness to noises. Moreover, we utilize the edge-preserving filtering method to substantially speed up CLASS without much losing its accuracy. The extensive experiments on both public and newly created video sequences suggest that CLASS achieves superior performance and comparable efficiency against other state-of-the-art approaches.

Keywords Collaborative model · Cognitive inspired · Low-rank and sparse representation · Global appearance consistency · Fast algorithm

Introduction

Moving object detection, as the fundamental problem in video analysis, plays a crucial role in intelligent transportation [1], law enforcement [2], and many other industries [3]. In the past decades, some representative methods have been proposed for moving object detection, including Gaussian Mixture Model (GMM) [4–6], ViBe [7, 8], optical flow [9, 10], and motion blur detection algorithm [11]. These methods, however, may be stumbling in some challenging scenarios, such as low illumination, intense shadows, camouflage, and dynamic background.

Recently, the low-rank and sparse separation models have been proposed to detect the moving objects as sparse outliers separated from the low-rank structure background [12, 13]. As illustrated in Fig. 1, the matrix of background images can be well approximated by a matrix with the low-rank (<5). Furthermore, the moving objects are usually small and sparse which is credible in most panoramic monitoring. However, the sparse assumption on moving objects usually limits the performance when the sizes of these moving objects are relatively large, which occurs frequently in uncontrolled surveillance systems.

One of the interesting phenomenon of human visual cognitive system of perceiving object is, while holding the physical size of the moving object constant, the distance of the object that the retina perceived varies the occupation

✉ Chenglong Li
lcl1314@foxmail.com

[1] School of Computer Science and Technology, Anhui University, Shushan Qu, China

[2] Jiangsu Engineering Centre of Network Monitoring and School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China

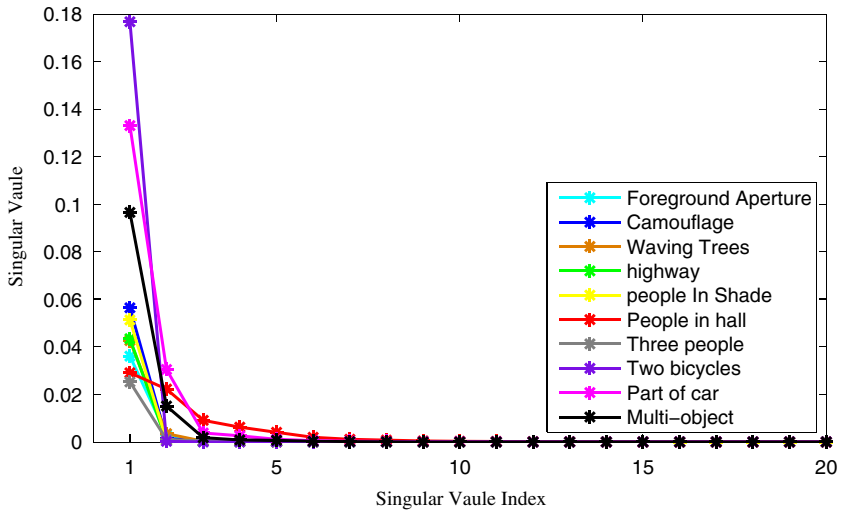

Fig. 1 The demonstration of the low-rank assumption on the background images. The curves indicate the singular values of the background image matrices of the 10 test videos that we evaluated in the experiments. For the detailed information of the videos, please refer to the experiments section

proportions of the human visual field [14, 15]. The neuroscientists at Washington University and the University of Minnesota have discovered that the visual information that the primary visual cortex processed is not the real size of the object but the size of perception [14]. Mel Goodale et al. [15] also discovered that the response of the visual cortex of human brain to bright object is not based on the physical size of the object but according to the perceived size of "afterimage" of object. This may explain why the objects with further distant seem to be so smaller in our visual system, but we still can cognitively feel the real size of the object.

Inspired by the visual cognitive process of human brain, in this paper, we propose a novel Collaborative Low-Rank And Sparse Separation (CLASS) model to robustly detect moving objects with different sizes. Specifically, given the data matrix that accumulates sequential frames from the input video, CLASS detects the moving objects as sparse outliers against the low-rank structure background. To overcome the limitation that the foreground objects should be sparse in conventional methods [14, 15], we incorporate the global appearance consistency of foreground and background into low-rank and sparse separation model and thus pursue a collaborative model for robust moving object detection. In particular, the sparse and the global appearance consistent constraints are complementary but simultaneously competing, and thus CLASS can detect the moving objects with different sizes effectively in challenging scenarios. The proposed collaborative model has the following properties: (i) It can utilize the advantages of low-rank and sparse separation models in background modeling and foreground detection; (ii) It can detect relatively large moving objects by leveraging global appearance consistency. Moreover, the smoothness constraints of object motion are also introduced in CLASS for further improving the robustness to noises.

For optimization, we design an iterative algorithm to efficiently solve the proposed model. Specifically, we iteratively optimize: (i) the background matrix by SOFT-INPUT algorithm [16], and (ii) the foreground mask by solving a Markov Random Field (MRF) model with graph cut algorithm [17, 18]. To further improve the efficiency, we design a fast implementation method without losing much accuracy, called F-CLASS. F-CLASS performs following three steps. First, the input video is down-sampled into low-resolution one. Second, we run CLASS on the low-resolution video to obtain the low-resolution detection results. Finally, regarding the original video as a guidance, we employ the edge-preserving method to recover the full-resolution detection results.

This paper makes the following three contributions:

- It proposes a novel collaborative model, CLASS, for robust moving object detection. CLASS takes advantages of both conventional low-rank and sparse separation models and global appearance consistent constraints and thus can effectively detects the moving objects with different sizes in challenging scenarios.
- It designs an efficient algorithm to solve the associated optimization problem. Moreover, it presents a fast implementation of CLASS based on the edge-preserving filtering, F-CLASS, to substantially speed up CLASS while sacrificing little accuracy.
- It creates several challenging video sequences to comprehensively evaluate the proposed approach against other state-of-the-art methods of moving object detection. These video sequences will be released online for free academic usage.

The rest of this paper is organized as follows. In Section "Related Work", the relevant existing methods are introduced. In Section "CLASS Algorithm", we describe the details of CLASS and F-CLASS and the associated

optimization algorithm. The experimental results on the public and newly created video sequences are shown in Section "Experiments". The final Section "Conclusion" concludes this paper.

Related Work

Over the years, many approaches have been proposed for moving object detection, including background subtraction, frame differencing, temporal differencing, and optical flow [19]. Our method falls into background substraction category, which is considered as one of the most competent approaches for moving object detection.

Background Substraction Background subtraction compares the pixels of input frame with the learnt background model, and the foreground pixels that differ from the background model are considered as moving objects. Thus, the critical task of background subtraction is to build a robust background model. Typical methods include single Gaussian distribution [20], mixture of Gaussian [4], and their variation [5, 21]. Barnich et al. [7] proposed a fast background updating scheme by comparing the current pixel with the sample set that randomly selected from the previous pixel and its neighbor pixels. Some works also introduced the fuzzy concepts into the procedure of the background subtraction process [22]. Pilet et al. [23] proposed a fast background subtraction algorithm for sudden illumination changes by modeling the background illumination as one of the three channels of Gaussian Mixture Model (GMM). Tong et al. [24] and Tu et al. [25] also proposed spatiotemporal saliency models for moving object detection. However, these methods model the background for each pixel independently and lose the consideration of the relations between the consecutive frames, thus they are very sensitive to noises and occlusions.

Low-Rank and Sparse Separation There are many computer vision applications which based on low-rank and sparse theory, such as visual tracking [26], segmentation [27, 28], and object detection [12]. The foreground is detected in the low-rank and sparse-based background modeling by discovering the correlation between the consecutive frames in lower subspace. One pioneering work is Robust Principal Component Analysis (RPCA) [29–31], which decomposes a given matrix/frames into a low-rank background matrix and sparse foreground matrix. Cands et al. [32] proposed to recover the low-rank and sparse components individually by solving a convenient convex program called Principal Component Pursuit (PCP). Zhou et al. [33] proposed Stable Principal Component Pursuit (SPCP) to handle both small entrywise noises and gross sparse errors. Dou et al. [34] proposed an incremental learning-based LRR model using K-SVD for dictionary learning. Different from most of the existing methods relaxing l_0-penalty to l_1-penalty, Zhou et al. [12] proposed DEtecting Contiguous Outliers in the LOw-rank Representation (DECOLOR) to relax the requirement of sparse and random distribution of corruption by preserving l_0-penalty and modeling the spatial contiguity of the sequence. Xin et al. [35] formulated foreground and background separation as a matrix decomposition problem using regularization terms for both the foreground and background matrices. However, most of existing methods ignored the global appearance consistency.

CLASS Algorithm

In this section, we will introduce the proposed model, Collaborative Low-Rank And Sparse Separation(CLASS), in a detailed way, and further present an optimization algorithm to solve it. At last, to improve the efficiency of the model, we develop a fast implementation method for moving object detection.

Problem Formulation

As justified in Fig. 1, background images are generally linearly correlated with each other in video surveillance. Based on this observation, we formulate the problem of foreground detection as a low-rank and sparse separation model. A video sequence $\mathbf{D} = [\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n] \in \mathbb{R}^{m\times n}$ is composed of n frames by of m pixels per frame. $\mathbf{B} \in \mathbb{R}^{m\times n}$ is a background pixel position matrix, which denotes the underlying background images. Our goal is to discover the object mask $\mathbf{S}$ from data matrices $\mathbf{D}$, where $\mathbf{S}_{ij}$ is a binary matrix:

$$\mathbf{S}_{ij} = \begin{cases} 0, & \text{if } ij \text{ is background,} \\ 1, & \text{if } ij \text{ is foreground.} \end{cases} \tag{1}$$

We assume that the underlying background images are linearly correlated and the foregrounds are sparse and contiguous, which has been successfully applied in background modeling [12, 36]. Firstly, the linear correlation of the underlying background images can be formulated by the low-rank constraint. Then, relative to the background region we assume that the foreground is sparse. Meanwhile, in the background region where $\mathbf{S}_{ij} = 0$, we assume that $\mathbf{D}_{ij} = \mathbf{B}_{ij} + \epsilon_{ij}$, where ϵ_{ij} denotes i.i.d. Gaussian noise. Therefore, based on the above assumptions, we have:

$$\begin{aligned} &\min_{\mathbf{B},\mathbf{S}_{ij}\in\{0,1\}} \beta \parallel vec(\mathbf{S}) \parallel_0 \\ &s.t.\ \mathbf{S}_\perp \circ \mathbf{D} = \mathbf{S}_\perp \circ (\mathbf{B} + \epsilon),\ rank(\mathbf{B}) \le r, \end{aligned} \tag{2}$$

where β is a penalized factor, $||\mathbf{X}||_0$ denotes the l_0-norm, which counts the number of nonzero entries, the operator "$\circ$" denotes element-wise multiplication of two matrices. $\mathbf{S}_\perp$ denotes the region of $\mathbf{S}_{ij} = 0$, and r is a constant that constrains the complexity of the background model.

Furthermore, since l_0 norm of the matrix $\mathbf{S}$ is nonconvex, we will introduce the contiguous constraint on $\mathbf{S}$, which is a prior that foreground objects should be contiguous pieces. In this way, $\parallel \mathbf{S} \parallel_0$ and the contigous constraints on $\mathbf{S}$ can be regarded as the unary term and pairwise term of MRF, respectively, which can be solved by graph cuts algorithm [17, 18] (see "S – *subproblem*" for details). This contiguous constraint is formulated as:

$$\sum_{(ij,kl)\in\varepsilon} |\mathbf{S}_{ij} - \mathbf{S}_{kl}| = ||\mathbf{C}\ vec(\mathbf{S})||_1, \tag{3}$$

where ε denotes the edge set connecting spatially neighboring pixels, $\mathbf{C}$ is the node-edge incidence matrix denoting the connecting relationship among pixels, and $vec(\mathbf{S})$ is a vectorize operator matrix $\mathbf{S}$. $||\mathbf{X}||_1 = \sum_{ij} |\mathbf{X}_{ij}|$ denotes the l_1-norm. Based on the above discussion, the formulation can be summarised as:

$$\min_{\mathbf{B},\mathbf{S}_{ij}\in\{0,1\}} \beta \parallel vec(\mathbf{S}) \parallel_0 + \gamma||\mathbf{C}\ vec(\mathbf{S})||_1$$
$$s.t.\ \mathbf{S}_\perp \circ \mathbf{D} = \mathbf{S}_\perp \circ (\mathbf{B} + \epsilon),\ rank(\mathbf{B}) \le r, \tag{4}$$

where γ is a balance parameter to control dependence between adjacent pixels.

However, due to assumptions of local spatial relationships and the sparsity of moving objects, the model Eq. (4) is theoretically not suitable for large object detection. From the statistical point of view, we further assume that foreground and background in the video sequence are Gaussian distributed, which has been widely and successfully used in object detection modeling [4, 6]. Based on this assumption, we introduce a GMM term against the sparse term to enhance detecting ability on large objects. To this end, we integrate the appearance model of foreground and background by global interactions:

$$\sum_{k=0}^{1}\sum_{i=1}^{m}\sum_{j=1}^{n} \delta(k, \mathbf{S}_{ij})\ A_k(i, j) = \sum_{k,i,j} \delta(k, \mathbf{S}_{ij})\ A_k(i, j), \tag{5}$$

where appearance model A consists of two Gaussian Mixture Models over RGB color values, A_0 and A_1 denotes the GMM of background and foreground models. $\delta(k, \mathbf{S}_{ij})$ is the Dirac delta function that denotes the value of k associated with $\mathbf{S}_{ij}$. The value of k is passed to the $A_k(i, j)$ which is a unary potential to evaluate how likely a pixel i is to be foreground or background according to the appearance model of frame j.

We integrate the global appearance model into Eq. (4) to obtain the formulation of Collaborative Low-Rank And Sparse Separation (CLASS) as:

$$\min_{\mathbf{B},\mathbf{S}_{ij}\in\{0,1\}} \beta \parallel vec(\mathbf{S}) \parallel_0 + \gamma||\mathbf{C}\ vec(\mathbf{S})||_1$$
$$+\mu \sum_{k,i,j} \delta(k, \mathbf{S}_{ij})\ A_k(i, j)$$
$$s.t.\ \mathbf{S}_\perp \circ \mathbf{D} = \mathbf{S}_\perp \circ (\mathbf{B} + \epsilon),\ rank(\mathbf{B}) \le r, \tag{6}$$

Model Optimization

Equation (6) is a NP-hard problem due to the non-convexity of the rank operator on $\mathbf{B}$; to make Eq. (6) tractable, we relax the rank operator with the nuclear norm, where the nuclear norm has proven to be an effective convex surrogate of the rank operator [37]. Therefore, Eq. (6) can be reformulated as:

$$\min_{\mathbf{B},\mathbf{S}_{ij}\in\{0,1\}} \frac{1}{2} ||P_{\mathbf{S}_\perp}(\mathbf{D} - \mathbf{B})||_F^2 + \beta \parallel vec(\mathbf{S}) \parallel_0$$
$$+ \gamma||\mathbf{C}\ vec(\mathbf{S})||_1$$
$$+ \mu \sum_{k,i,j} \delta(k, \mathbf{S}_{ij})\ A_k(i, j) + \alpha \parallel \mathbf{B} \parallel_*, \tag{7}$$

where α is a balance parameter. $||\mathbf{X}||_F = \sqrt{\sum_{ij} \mathbf{X}_{ij}^2}$ is the Frobenius norm, $||\mathbf{X}||_*$ means the nuclear norm, i.e., sum of singular values. $P_{\mathbf{S}_\perp}(\mathbf{X})$ is the complement to $P_{\mathbf{S}}(\mathbf{X})$ which is the orthogonal projection of matrix $\mathbf{X}$ denoted by:

$$P_{\mathbf{S}}(\mathbf{X})(i, j) = \begin{cases} 0, & if\ \mathbf{S}_{ij} = 0, \\ \mathbf{X}_{ij}, & if\ \mathbf{S}_{ij} = 1. \end{cases} \tag{8}$$

Therefore, we adopt an alternating algorithm by separating Eq. (7) over $\mathbf{B}$ and $\mathbf{S}$ in the following two steps.

$\mathbf{B}$ – *subproblem* Given a current estimate of the foreground mask $\hat{\mathbf{S}}$, estimating $\mathbf{B}$ by minimizing Eq. (7) turns

out to be the matrix completion problem. This is to learn a low-rank background matrix from partial observations.

$$\min_{\mathbf{B}} \frac{1}{2} ||P_{\hat{\mathbf{S}}_\perp}(\mathbf{D}-\mathbf{B})||_F^2 + \alpha \parallel \mathbf{B} \parallel_*, \tag{9}$$

The optimal $\mathbf{B}$ in Eq. (12) can be computed by the SOFT-IMPUTE [16] algorithm. which is based on the following Lemma [38]:

Lemma 1 Given a matrix $\mathbf{Z}$, the solution to the optimization problem

$$\min_{\mathbf{X}} \frac{1}{2} ||\mathbf{Z}-\mathbf{X}||_F^2 + \alpha \parallel \mathbf{X} \parallel_*, \tag{10}$$

is given by $\hat{\mathbf{X}} = \Theta_\alpha(\mathbf{Z})$, where Θ_α means the singular value thresholding

$$\Theta_\alpha(\mathbf{Z}) = \mathbf{U}\Sigma_\alpha\mathbf{V}^T, \tag{11}$$

Here, $\Sigma_\alpha = diag[(d_1-\alpha)_+, \cdots, (d_r-\alpha)_+]$, $\mathbf{U}\Sigma_\alpha\mathbf{V}^T$ is the SVD of $\mathbf{Z}$, $\Sigma = diag[d_1 - d_r]$ and $t_+ = max(t, 0)$. Rewriting Eq. (9), we have:

$$\begin{aligned}
&\min_{\mathbf{B}} \frac{1}{2} ||P_{\hat{\mathbf{S}}_\perp}(\mathbf{D}-\mathbf{B})||_F^2 + \alpha \parallel \mathbf{B} \parallel_*, \\
&= \min_{\mathbf{B}} \frac{1}{2} ||[P_{\hat{\mathbf{S}}_\perp}(\mathbf{D}) + P_{\hat{\mathbf{S}}}(\mathbf{B})] - \mathbf{B})||_F^2 + \alpha \parallel \mathbf{B} \parallel_*,
\end{aligned} \tag{12}$$

According to Lemma 1, given an arbitrary initialization $\hat{\mathbf{B}}$, the optimal solution can be obtained by iteratively using Eq. (13):

$$\hat{\mathbf{B}} \longleftarrow \Theta_\alpha(P_{\hat{\mathbf{S}}_\perp}(\mathbf{D}) + P_{\hat{\mathbf{S}}}(\hat{\mathbf{B}})), \tag{13}$$

$\mathbf{S}-$*subproblem* Given a current estimate of the background position matrix $\hat{\mathbf{B}}$, Eq. (7) can be transferred into the following optimization functions:

$$\begin{aligned}
&\min_{\mathbf{S}} \frac{1}{2} ||P_{\mathbf{S}_\perp}(\mathbf{D}-\hat{\mathbf{B}})||_F^2 + \beta \parallel vec(\mathbf{S}) \parallel_0 \\
&\quad + \gamma ||\mathbf{C}\, vec(\mathbf{S})||_1 \\
&\quad + \mu \sum_{k,i,j} \delta(k, \mathbf{S}_{ij})\, A_k(i, j),
\end{aligned} \tag{14}$$

The energy function Eq. (14) can be rewritten in line with the standard form of a first-order Markov Random Fields [39] as:

$$\begin{aligned}
&\frac{1}{2} ||P_{\mathbf{S}_\perp}(\mathbf{D}-\hat{\mathbf{B}})||_F^2 + \beta \parallel vec(\mathbf{S}) \parallel_0 + \gamma ||\mathbf{C}\, vec(\mathbf{S})||_1 \\
&\quad + \mu \sum_{k,i,j} \delta(k, \mathbf{S}_{ij})\, A_k(i, j), \\
&= \frac{1}{2} \sum_{i,j} (\mathbf{D}_{ij} - \hat{\mathbf{B}}_{ij})^2 (1 - \mathbf{S}_{ij}) + \beta \sum_{i,j} \mathbf{S}_{ij} \\
&\quad + \mu \sum_{k,i,j} \delta(k, \mathbf{S}_{ij})\, A_k(i, j) \\
&\quad + \gamma ||\mathbf{C}\, vec(\mathbf{S})||_1, \\
&= \sum_{k,i,j} [(\beta - \frac{1}{2}(\mathbf{D}_{ij} - \hat{\mathbf{B}}_{ij}))^2 \mathbf{S}_{ij} + \mu\, \delta(k, \mathbf{S}_{ij})\, A_k(i, j)] \\
&\quad + \gamma ||\mathbf{C}\, vec(\mathbf{S})||_1 \\
&\quad + \frac{1}{2} \sum_{i,j} (\mathbf{D}_{ij} - \hat{\mathbf{B}}_{ij})^2.
\end{aligned} \tag{15}$$

When $\hat{\mathbf{B}}$ is fixed and $\frac{1}{2}\sum_{i,j}(\mathbf{D}_{ij} - \hat{\mathbf{B}}_{ij})^2$ is constant. Meanwhile, $\mathbf{S}_{ij}$ beside the $(\beta - \frac{1}{2}(\mathbf{D}_{ij} - \hat{\mathbf{B}}_{ij}))^2$ is also constant. Known Markov unary term and pairwise smoothing term, one can easily obtain the optimal foreground matrix though graph cuts method [17, 18] since $\mathbf{S}_{ij} \in \{0, 1\}$ is discrete.

A sub-optimal solution can be obtained by alternating optimizing $\mathbf{B}$ and $\mathbf{S}$ and the algorithm is summarised in Algorithm 1.

Algorithm 1 Optimization algorithm to Eq. 7

Input: $\mathbf{D} = [\mathbf{l}_1, \mathbf{l}_2, ..., \mathbf{l}_n] \in \mathbb{R}^{m \times n}$
Set $\mathbf{B} = \mathbf{D}$, $\mathbf{S} = \mathbf{0}$, $\tau = 1e-4$, $maxIter = 20$
Output: $\hat{\mathbf{S}}$, $\hat{\mathbf{B}}$

1: Using SOFT-IMPUTE algorithm to optimize energy function Eq. (9), by computing $\hat{\mathbf{B}}$: $\hat{\mathbf{B}} \longleftarrow \Theta_\alpha(P_{\hat{\mathbf{S}}_\perp}(\mathbf{D}) + P_{\hat{\mathbf{S}}}(\hat{\mathbf{B}}))$
2: **if** $rank(\hat{\mathbf{B}}) \leq K$ **then**
3: tuning parameters α , return to step 1
4: **end if**
5: Using graph cuts algorithm to optimize energy function Eq. (14) by computing $\hat{\mathbf{S}}$: $\hat{\mathbf{S}} = arg\ min_{\mathbf{S}} \sum_{k,i,j}[(\beta - \frac{1}{2}(\mathbf{D}_{ij} - \hat{\mathbf{B}}_{ij}))^2 \mathbf{S}_{ij} + \mu\, \delta(k, \mathbf{S}_{ij})\, A_k(i, j)] + \gamma ||\mathbf{C}\, vec(\mathbf{S})||_1$
6: Check the convergence condition: if the maximum objective change between two consecutive iterations is less than τ or the maximum number of iterations reaches $maxIter$, then terminate the loop.

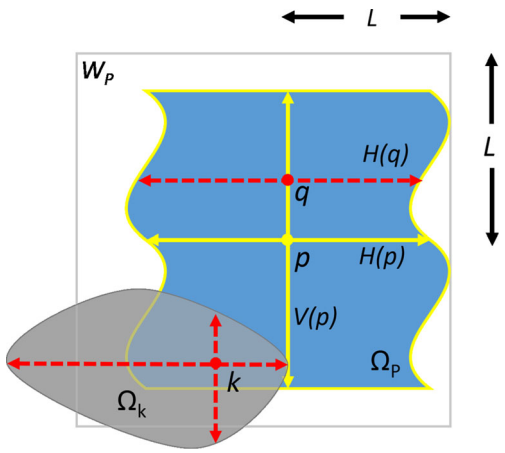

Fig. 2 Illustration of generating a pixelwise shape-adaptive region of CLMF [40]. See text for more details

F-CLASS: Fast Implementation

In order to improve the efficiency, we use an edge-preserving, filtering-based method to speed up our model while sacrificing little accuracy. The fast version of CLASS (F-CLASS) is executed in three steps: firstly, the input video is down-sampled into low resolution by scale of 4 ($\frac{1}{4}W \times \frac{1}{4}H$, where W and H denote the width and hight of video frame, respectively.) of the original resolution though bilinear interpolation. Secondly, we run the CLASS algorithm on the down-sampled frames and obtain the initial detection results. Finally, regarding the original video as a guidance, we employ the edge-preserving up-sampling technique [36] to recover the full-resolution detection results. The edge-preserving up-sampling method is executed in two steps:

1. Shape-Adaptive Region Generation The shape-adaptive region in the frame is generated by Cross-based Local Multipoint Filtering (CLMF) [40] for each pixel.

Table 1 Detailed information of the public test videos

Name of the video	Name of the dataset	Size×no. of frame
Foreground aperture	Wallflower	[160,120]×21
Camouflage	Wallflower	[160,120]×21
Waving trees	Wallflower	[160,120]×39
Highway	2014 Change Detection	[320,240]×27
People In Shade	2014 Change Detection	[380,244]×21

Table 2 Detailed information of the collected test videos

Name of the video	Name of the dataset	Size×no. of frame
People in hall	Shooting	[480,360]×41
Three people	Shooting	[160,120]×21
Two bicycles	Shooting	[240,180]×51
Part of car	Shooting	[240,180]×67
Multi-object	Shooting	[240,180]×51

Specifically, as shown in Fig. 2, for a pixel p centered at a square observation window W_p with the size of $(2L+1)\times(2L+1)$, similarity criterion for a pixel q in the window W_p falls into a color cube as:

$$|I_c(q) - I_c(p)| \leq \tau,\ c \in \{R, G, \mathcal{B}\},\ q \in W_p, \tag{16}$$

where I_c is the intensity of the color channel c of the 3×3 median-smoothed guidance image I and τ controls the size of the color cube to generate the shape-adaptive region Ω_p. CLMF produces a horizontal (left/right) spans (pixels) $H(p)$ and a vertical (up/bottom) spans (pixels) $V(p)$ of the anchor pixel p according to Eq. 16. For $\forall q \in V(p)$, we can construct the the arbitrary-shaped region of p by integrating multiple $H(q)$ sliding along $V(p)$: $\Omega_p = \bigcup_{q\in V(p)} H(q)$.

2. Edge-Preserving Up-sampling We employ the edge-preserving filtering to up-sample the low-resolution frame to the full-resolution without perturbing object edges. For the pixel p on the full-resolution image J, its value can be estimated similarly as the joint bilateral up-sampling [41]:

$$J(p) = \frac{1}{|\Omega_k|}\sum_{k\in\Omega_k} \omega_{p,k} J^l(k), \tag{17}$$

Table 3 F-measure of the proposed method with different parameters

Parameter	Setting	F-measure
β	$0.9\sigma^2$	0.79
	$4.5\sigma^2$	0.93
	$9\sigma^2$	0.92
γ	0.4β	0.91
	1β	0.93
	1.5β	0.91
μ	0.13γ	0.93
	0.27γ	0.93
	0.4γ	0.90

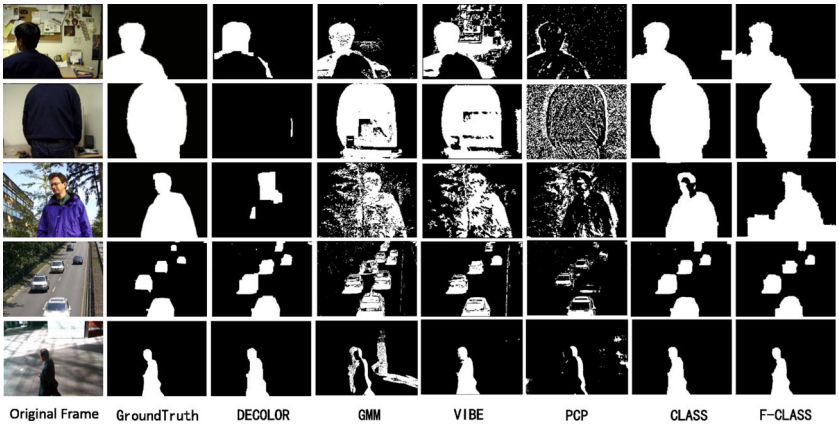

Fig. 3 Example results on five public video sequences. The first row to the fifth row indicate the detected results on video "Foreground aperture", "Camouflage", "Waving trees", "highway" and "people In Shade," respectively

where Ω_k is the shape-adaptive region generated by the guidance image I of pixel k in the low-resolution image J^l, as shown in Fig. 2. $|\Omega_k|$ denotes the number of pixels in Ω_k, and $\omega_{p,k} = \exp(-\frac{||\mathbf{x}_p,\mathbf{x}_k||}{\sigma})$, where $\mathbf{x}_p$ indicates the position of p. In this way, the pixel values are averagely weighted by the spatial distance in the homogeneous regions of the guidance image to form the full-resolution one without perturbing the edges of objects.

Experiments

We evaluate our CLASS and F-CLASS on 10 challenging videos from both public and newly collected video sequences from our on-campus surveillance system.

Evaluation Settings

Datasets We first evaluate our method on five video sequences selected from 2014 Change Detection dataset [42] and Wallflower dataset [43]. The foreground in these sequences are mainly pedestrians and vehicles with various amount, sizes, and velocities. The background is significantly interfered in some videos like swing of the trees, flicker of the screen, shielding between the camera, and the objects. The detailed information of the test videos can be found in Table 1.

We also collected a series of videos on campus real-life scenes with various types, amounts, and sizes of moving objects[1]. Table 2 provides the length and frame size of each video sequence we evaluated in the experiments.

[1] Available at: http://chenglongli.cn/people/lcl/journals.html.

Parameters In our model of Eq. 7, the parameter α controls the complexity of the background model which is first roughly estimated by the rank of the background model. The parameter β which controls the sparsity of the foreground masks. The parameter γ controls the spatial smoothness of foreground and background and can be adaptively adjusted by β. The most significant parameter is μ which controls the contribution of global appearance consistency. We determined μ by adjusting its ratio to γ. All the parameters are jointly optimized. We evaluated the parameters with different setting and reported the F-measure in Table 3. For better performance, the parameters are set as: $\{\mu, \gamma, \beta\} = \{0.27\gamma, 1\beta, 4.5\sigma^2\}$ for CLASS where σ^2 is estimated online by the mean variance of $\{\mathbf{D_{ij}} - \hat{\mathbf{B}}_{ij}\}$. Analogously, we empirically set $\{\mu, \gamma, \beta\} = \{0.55\gamma, 0.5\beta, 4.5\sigma^2\}$ for F-CLASS and set the window size L and the similarity threshold τ to be 3 and 10, respectively.

Evaluation Criterion The precision, recall, and F-measure are first comprehensively evaluated, which are defined as follows:

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{\text{TP}+\text{FP}}, \ \text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}, \\ \text{F-measure} &= 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \end{aligned} \tag{18}$$

where TP=true positives, indicating the foreground pixels correctly labeled as foreground. FP=false positives, referring the background pixels incorrectly labeled as foreground. TN=true negatives, corresponding to background pixels correctly labeled as background. FN=false negatives, referring to foreground pixels incorrectly labeled as background [44]. F-measure is a comprehensive measurement to balance the argument between precision and recall.

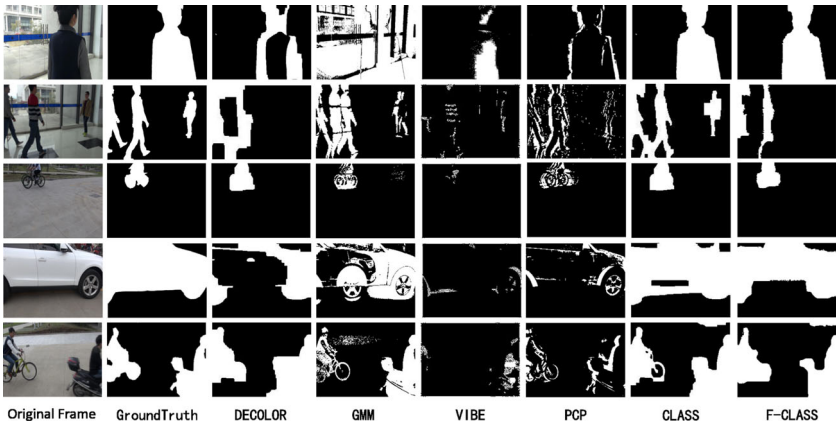

Fig. 4 Example results on five collected video sequences. The first row to the fifth row indicate the detected results on "People in hall", "Three people", "Two bicycle", "Part of car" and "Multi-object," respectively

Furthermore, the mean absolute error (MAE) is evaluated to measure the disagreement between the detected results and the groundtruth:

$$MAE = \frac{1}{N \times \mathcal{F}} \sum_{i=1}^{\mathcal{F}} \sum_{p \in DR, \acute{p} \in GT} XOR(p, \acute{p}) \tag{19}$$

where N denotes and resolution of the frame and F denotes the number of the frames in the video clip. DR and GT indicate the "Detection Result" and the "Ground Truth" respectively. $XOR(*)$ denotes the logic operator "$exclusive\ OR$". $p, \acute{p} \in \{0, 1\}$ denotes the background/foreground pixels.

Comparison Results

We compare our approach with the four state-of-the-art moving object detection algorithms including

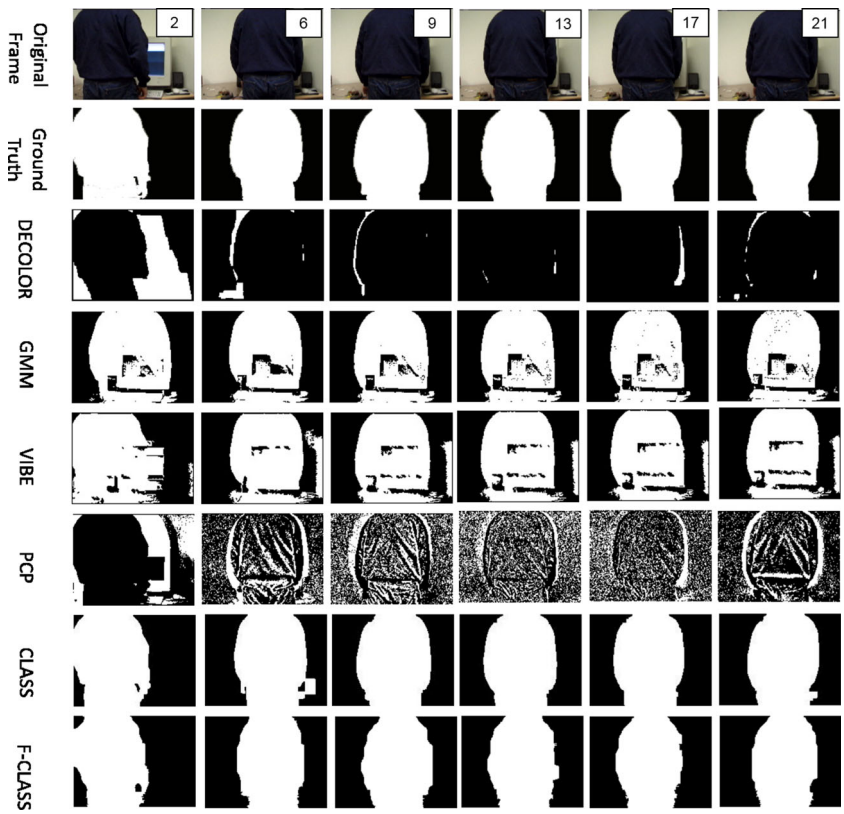

Fig. 5 Detection results on several sequential frames of video "Camouflage"

DECOLOR [12], GMM [5], VIBE [7], and PCP [32]. To keep things fair, we choose the default parameters released by the authors of corresponding methods.

Qualitative Results Figures 3 and 4 demonstrates the detected results on a certain frame of the test video clips listed on Tables 1 and 2. From which we can see, our methods (CLASS and F-CLASS) significantly outperform the state-of-the-art methods on the videos "Foreground aperture", "Camouflage", "Waving trees", "People in hall" and "Part of car" where the moving object dominates the frame (with huge size) and also works satisfactory on the other videos with the normal size moving objects. DECOLOR is designed under the priori assumption that the moving object is sparse (with small size) which limits its application on huge-sized object, while PCP is not robust enough for the influence of contiguous noises and occlusions due to the l_1-penalty. GMM and VIBE work on the original pixel space; therefore, they are quite sensitive to the noises and introduce "ghost."

Figure 5 elaborates the detecting results on several sequential frames of video "Foreground Aperture." DECOLOR estimates the foreground via outlier detection; therefore, it loses the object(s) with abrupt stop while GMM, VIBE, and PCP introduce more noises during detection. Our CLASS and F-CLASS can better capture the whole body of the moving object without introducing extra noises. Note that F-CLASS can further eliminate the false detection in CLASS.

Quantitative Results Tables 4 and 5 report precision, recall, F-measure, and MAE on public datasets and collected datasets comprehensively. We can see CLASS and F-CLASS outperform the state-of-the-art methods in most

Table 4 The precision, recall, F-measure, and MAE values on five public video sequences, where the italic fonts of results indicate the best performance

		Foreground Aperture	Camouflage	Waving Trees	Highway	People In Shade	Mean	Variance
Precision	DECOLOR	0.5715	0.3399	0.6805	0.6897	0.8549	0.6273	0.0288
	GMM	0.5794	0.9180	0.5489	0.4843	0.3991	0.5859	0.0314
	VIBE	0.5495	0.8859	0.5885	0.6848	*0.8694*	0.7156	0.0195
	PCP	0.1504	0.4036	0.2396	0.6932	0.0552	0.3084	0.0502
	CLASS	0.8987	0.9894	*0.9623*	0.7089	0.7226	0.8564	0.0141
	F-CLASS	*0.9592*	*0.9963*	0.7816	*0.7167*	0.8465	*0.8601*	*0.0111*
Recall	DECOLOR	0.2443	0.0310	0.3558	0.9983	0.9427	0.5144	0.1499
	GMM	0.3357	0.8691	0.6579	0.7465	0.7689	0.6756	0.0334
	VIBE	0.3318	0.9074	0.6356	0.8624	0.9523	0.7379	0.0531
	PCP	0.0611	0.2579	0.1936	0.3466	0.0202	0.1759	0.0147
	CLASS	*0.9866*	*0.9856*	0.8272	*0.9988*	*0.9929*	*0.9582*	0.0043
	F-CLASS	0.9472	0.9173	*0.9656*	0.9181	0.9712	0.9439	*0.0005*
F-measure	DECOLOR	0.3223	0.0474	0.4341	0.8153	0.8913	0.5021	0.0987
	GMM	0.4101	0.8921	0.5796	0.5846	0.5243	0.5981	0.0256
	VIBE	0.3453	0.8962	0.5951	0.7625	*0.9083*	0.7015	0.0445
	PCP	0.0740	0.3109	0.1997	0.4613	0.0291	0.2150	0.0249
	CLASS	0.9404	*0.9875*	*0.8865*	*0.8288*	0.8359	*0.8958*	0.0037
	F-CLASS	*0.9517*	0.9551	0.8592	0.8028	0.9033	0.8944	*0.0033*
MAE	DECOLOR	0.1936	0.5611	0.0750	0.0502	0.0305	0.1821	0.0391
	GMM	0.1655	0.1126	0.1765	0.1181	0.1731	0.1492	0.0008
	VIBE	0.2205	0.1109	0.1416	0.0595	0.0261	0.1117	0.0046
	PCP	0.2878	0.5790	0.3318	0.0901	0.1761	0.2930	0.0276
	CLASS	0.0259	0.0133	0.0231	0.0456	0.0498	*0.0315*	0.0002
	F-CLASS	0.0212	0.0455	0.0372	0.0503	0.0289	0.0366	*0.0001*

Table 5 The precision, recall, F-measure, and MAE values on five collected video sequences, where the italic fonts of results indicate the best performance

		People in hall	Three people	Two bicycles	Part of car	Multi-object	Mean	Variance
Precision	DECOLOR	0.7908	0.6592	0.7898	0.7738	0.7798	0.7587	*0.0025*
	GMM	0.6985	0.5780	0.7347	0.5606	*0.8716*	0.6887	0.0129
	VIBE	0.8384	0.4271	0.0937	0.4383	0.5383	0.4672	0.0570
	PCP	0.1015	0.1631	0.2961	*0.9279*	0.4058	0.3789	0.0865
	CLASS	0.9275	*0.7863*	0.8014	0.7963	0.7731	0.8169	0.0032
	F-CLASS	*0.9850*	0.7399	*0.8543*	0.8005	0.7516	*0.8263*	0.0079
Recall	DECOLOR	0.8409	0.4219	0.9860	0.7676	*0.9890*	0.8011	0.0432
	GMM	0.9112	0.7132	0.4459	0.6627	0.8350	0.7136	0.0256
	VIBE	0.3638	0.1001	0.0455	0.0405	0.1483	0.1396	0.0141
	PCP	0.0344	0.1092	0.3860	0.2722	0.1795	0.1963	0.0152
	CLASS	*0.9946*	*0.9651*	*0.9872*	*0.9449*	0.9814	*0.9746*	*0.0003*
	F-CLASS	0.9754	0.3544	0.9626	0.9193	0.9563	0.8336	0.0587
F-measure	DECOLOR	0.7855	0.4940	0.8764	0.7310	*0.8696*	0.7513	0.0195
	GMM	0.7631	0.6368	0.5523	0.5501	0.8489	0.6702	0.0140
	VIBE	0.5012	0.1621	0.0608	0.0615	0.2262	0.2024	0.0263
	PCP	0.0292	0.1303	0.3144	0.3808	0.2236	0.2157	0.0158
	CLASS	0.9535	*0.8661*	0.8842	*0.8579*	0.8620	*0.8847*	*0.0013*
	F-CLASS	*0.9795*	0.4653	*0.9031*	0.8511	0.8385	0.8075	0.0317
MAE	DECOLOR	0.1213	0.1658	0.0117	0.1183	0.0453	0.0925	0.0031
	GMM	0.2307	0.1644	0.0273	0.2217	0.0578	0.1404	0.0070
	VIBE	0.1709	0.2104	0.0564	0.2604	0.1791	0.1754	0.0045
	PCP	0.3171	0.2977	0.0684	0.2577	0.2008	0.2283	0.0080
	CLASS	0.0098	0.0609	0.0108	0.0574	0.0485	*0.0375*	*0.0005*
	F-CLASS	0.0076	0.1472	*0.0091*	0.0526	0.0607	0.0554	0.0026

of the cases. On the public datasets, CLASS/F-CLASS outperforms the second best method by averagely 20.2, 29.9, 27.7, and 71.8 % in precision, recall, F-measure, and MAE respectively while 8.9, 21.7, 17.8, and 59.5 % on collected datasets. For the videos where the size of the moving object dominates the frame (with huge size), our method significantly beats the state-of-the-art methods in all the precision, recall, and F-measure. For the videos with the normal size moving objects, our CLASS is slightly in the shade of VIBE and DECOLOR in F-measure on "people In Shade" but F-CLASS can further approach to the best performance.

On Table 5, although the precision of CLASS on video "Part of car" looks lower than PCP but with much higher recall thus leads to best F-measure which is the comprehensive criteria between precision and recall. On video "Multi-object," DECOLOR performs slightly better than CLASS, but from Fig. 5, we can see DECOLOR lost the boundary details of the moving objects while our method still achieves the good visual performance. Note that on "Three People", the F-CLASS declines a lot compared to CLASS which is due to the high compression of down-sampling; the performance can be greatly increased by less down-sampling.

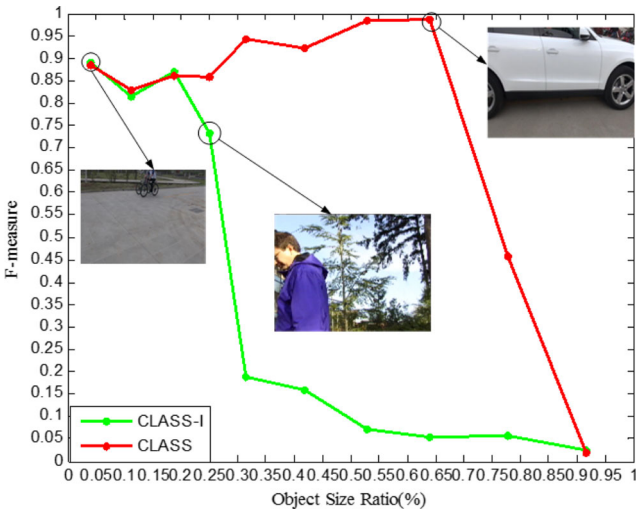

Fig. 6 F-measure against object size ratio

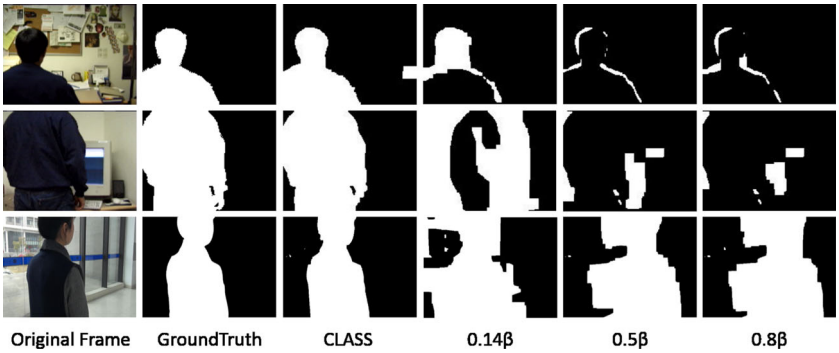

Fig. 7 Detecting results with small coefficient β (weaker constraint of foreground sparsity). The first to the third rows indicate the detected results on "Foreground Aperture", "Camouflage," and "People in hall," respectively

We can also conclude that, due to the smooth-term of CLASS model, CLASS tends to bias the recall, while F-CLASS can improve the precision by exploiting the structural information of the original images through edge-preserving up-sampling. Therefore, F-CLASS appears bias to precision.

Other Discussion Firstly, we shall discuss the influence of object size ratio to CLASS. The object size ratio (OSR) is measured as:

$$OSR = \frac{1}{\mathcal{F}} \sum_{i=1}^{\mathcal{F}} \frac{f_i}{N} \times 100\% \tag{20}$$

where f_i denotes the number of groundtruth foreground pixels of the ith frame, N denotes the resolution of the frame, and $\mathcal{F}$ denotes the number of the frames. Figure 6 shows the performance (F-measure) of CLASS and CLASS-I (without global appearance consistency by setting μ to 0) with different OSR from the 10 test videos. Thinking that the largest object size ratio in the 10 test videos is still less than 55 %, we crop the bounders of the "Foreground Aperture" and "Part of car" to figure out larger object size ratios. From Fig. 6, we can see: (1) CLASS significantly outperforms CLASS-I especially when the size ratio is larger than approximately 20 % which demonstrates the robustness of CLASS while dealing with various size ratios. (2) The performance of CLASS significantly declines when the object size ratio is larger than approximately 65 % which is visually almost full screen while CLASS-I declines at approximately 25 % where the moving object is still a small portion of the screen.

Furthermore, thinking that β controls the sparsity of the foreground masks, one may suggest to set small coefficient β in Eq. 4 for larger objects. However, from the theoretical point of view, setting different parameters for different-sized objects greatly effects the universality of the detection model. In spite of this, we evaluate the detection on large objects by decreasing the coefficient β in Eq. 4 and demonstrating the detection results on Fig. 7. From which we can see, it still fails for detecting large objects with small coefficient β, which is also one of the motivations of our work.

Component Analysis

In order to validate the component contribution of CLASS, we evaluate the components of global appearance consistency and fast implementation and report the results on Table 6, where the precision, recall, and F-measure denote the average values on 10 test video sequences. (1) CLASS: the original Collaborative Low-Rank And Sparse Separation without fast implementation; (2) F-CLASS: the fast implementation on original CLASS; (3) CLASS-I: the original CLASS without global appearance consistency by setting μ to 0; (4) F-CLASS-I: the fast implementation on CLASS-I. From which we can see that: (1) CLASS significantly outperforms CLASS-I and F-CLASS consistently outperforms F-CLASS-I, which justify that the global appearance consistency plays important roles to moving object detection. 2) Although F-CLASS and F-CLASS-I slightly cast into the shade of CLASS and CLASS-I, respectively, the detecting results of the fast implementation are still satisfactory and with near-real-time speed (by 8.52 fps as shown in Table 7). At the same time, in order to understand directly the role of the components of CLASS, we visualize some detection

Table 6 Average precision, recall, and F-measure of our method and its variants on the entire dataset. The italic fonts of results indicate the best performance

Algorithm	Precision	Recall	F-measure
CLASS	0.837	*0.966*	*0.890*
CLASS-I	0.699	0.642	0.612
F-CLASS	*0.843*	0.889	0.851
F-CLASS-I	0.672	0.618	0.591

Table 7 The code type and frames per second (FPS)

	DECOLOR	GMM	VIBE	PCP	CLASS	F-CLASS
Code Type	MATLAB & C++	C++	C++	MATLAB	MATLAB & C++	MATLAB & C++
FPS	2.02	74.26	253.41	23.58	0.43	8.52

results; Fig. 8 demonstrates the detected results on a certain frame of the five test video clips, which visually demonstrates the significance of the global appearance consistency and the fast implementation.

Efficiency Analysis

The experiments are carried out on a desktop with an Intel i7 3.4GHz CPU and 32GB RAM, and implemented on mixing platform of C++ and MATLAB without any code optimization. Runtime of our method against other methods is presented in Table 7, and all frames are with 240 × 180 resolution. From which we can see, F-CLASS can speed up CLASS with almost 20 times (achieving about 8.52 FPS). Though GMM, VIBE, and PCP run much faster than ours, these methods generally perform greatly worse than CLASS and F-CLASS in precision, recall, and F-measure. DECOLOR are comparable with CLASS in efficiency but with much worse accuracy than CLASS and F-CLASS. These demonstrate that F-CLASS keeps a good balance in efficiency and accuracy. Notice that our F-CLASS is near-real-time and can easily obtain real-time performance though code optimization. In addition, we can further reduce the computational burden by increasing the down-sampling scalar with sacrificing slight accuracy or even cloud computing [45, 46].

Furthermore, we present the influences of the down-sampling scalars on the performance and computational efficiency in Table 8. From the results, we can see that the performance decreases slightly while the runtime decreases greatly while turning the down-sampling scalar from 4 to 2. When the down-sampling scalar is between 4 and 5, the results have great changes in both performance and computational efficiency. Therefore, we set the down-sampling scalar to be 4 in this paper to balance accuracy-efficiency trade-off.

Limitations

We also encounter unsatisfying detection results such as on data "Foreground Aperture" and "Camouflage" as shown in Figs. 3 and 5. When the background contains similar color/gray-scale part as the foreground, our method tends to detect it as foreground in some certain frame(s). This may result from the assumption of GMM-based background modeling methods that the foreground is generally with distinguishing appearance to background. One can see that it can be alleviated by decreasing the weight of GMM (the parameter μ), but the ability of our method on detecting large object may consequently decline. This limitation could be refined by the motion changes between the continuous frames.

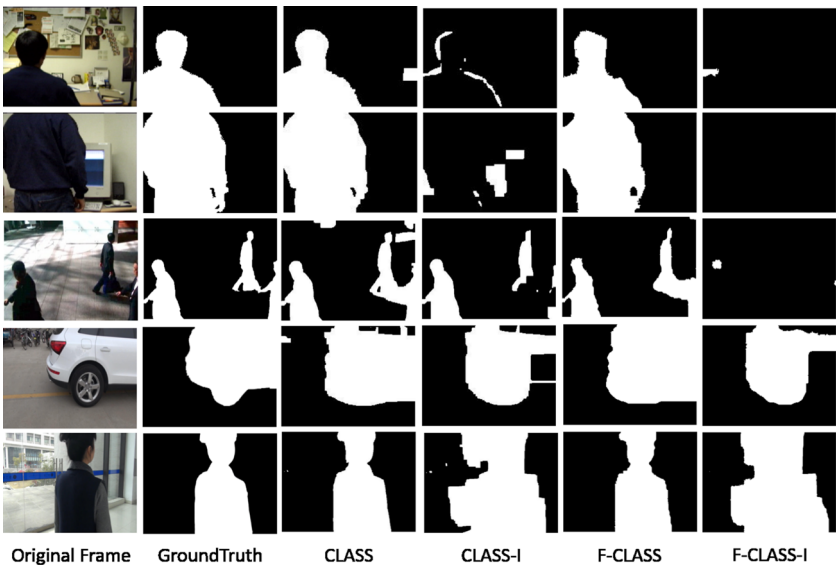

Fig. 8 Example results of our method and its variants on the several datasets. The first to fifth rows indicate the detected results on "Foreground Aperture", "Camouflage", "people In Shade", "Part of car" and "People in hall," respectively

Table 8 The performance and runtime of different down-sampling scalars on several videos

Scalar	F-measure	FPS
2	0.92	5.88
3	0.92	7.85
4	0.91	10.49
5	0.83	18.25

Conclusion

This paper has proposed a collaborative model for robust moving object detection. Our moving object detectors take the foreground as sparse outliers while pursuing the low-rank structure background. In the mean time, our framework also retrained superior global appearance consistency. Through detecting moving objects with various visual sizes, we verified the robustness of our collaborative model. Extensive experiments on the public and collected video sequences suggest that the proposed method outperforms other state-of-the-art detection methods. In future work, we will focus on extending our model to online or streaming fashion for real-life applications and also improve our method to handle the limitations of over-huge objects or similar appearance of foreground and background.

Compliance with Ethical Standards

Funding This study was funded by the National Nature Science Foundation of China (61502006), the Natural Science Foundation of Anhui Province (1508085QF127), and the Natural Science Foundation of Anhui Higher Education Institutions of China (KJ2014A015, KJ2015A110, KJ2016A114 and KJ2015ZD44).

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Conflict of interests The authors declare that they have no conflict of interest.

References

1. Wen X, Shao L, Xue Y, Fang W. A rapid learning algorithm for vehicle classification. Inf Sci. 2015;295:395–406.
2. Pan Z, Lei J, Zhang Y, Sun X. Fast motion estimation based on content property for low-complexity h.265/hevc encoder. IEEE Trans Broadcast. 2016:1–10.
3. Pang Y, Cao J, Li X. Learning sampling distributions for efficient object detection. IEEE Transactions on Cybernetics. 2016:1–13.
4. Stauffer C, Grimson WEL. Adaptive background mixture models for real-time tracking. 1999 Proceedings of the IEEE International Conference on Computer Vision. Volume 2; 1999. p. 246–252.
5. KaewTraKulPong P, Bowden R. An improved adaptive background mixture model for real-time tracking with shadow detection: Springer; 2002, pp. 135–144.
6. Papazoglou A, Ferrari V. Fast object segmentation in unconstrained video. 2013 Proceedings of the IEEE International Conference on Computer Vision; 2013. p. 1777–1784.
7. Barnich O, Van Droogenbroeck M. Vibe: A universal background subtraction algorithm for video sequences. IEEE Trans Image Process. 2011;20(6):1709–1724.
8. Van Droogenbroeck M, Paquot O. Background subtraction: experiments and improvements for ViBe. 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE; 2012. p. 32-37.
9. Patel MP, Parmar SK. Moving object detection with moving background using optic flow: IEEE; 2014, pp. 1–6.
10. Sun D, Roth S, Black MJ. Secrets of optical flow estimation and their principles. 2010 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2010. p. 2432–2439.
11. Pang Y, Zhu H, Li X, Pan J. Motion blur detection with an indicator function for surveillance machines. IEEE Trans Ind Electron. 2016:5592–5601.
12. Zhou X, Yang C, Yu W. Moving object detection by detecting contiguous outliers in the low-rank representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013;35(3):597–610.
13. Pan J, Li X, Li X, Pang Y. Incrementally detecting moving objects in video with sparsity and connectivity. Cognitive Computation. 2015:1–9.
14. Murray SO, Boyaci H, Kersten D. The representation of perceived angular size in human primary visual cortex. Nat Neurosci. 2006;9(3):429–434.
15. Sperandio I, Chouinard PA, Goodale MA. Retinotopic activity in v1 reflects the perceived and not the retinal size of an afterimage. Nat Neurosci. 2012;15(4):540–5422.
16. Mazumder R, Hastie T, Tibshirani R. Spectral regularization algorithms for learning large incomplete matrices. The Journal of Machine Learning Research. 2010;11:2287–2322.
17. Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001;23(11):1222–1239.
18. Kolmogorov V, Zabin R. What energy functions can be minimized via graph cuts? IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004;26(2):147–159.
19. Kulchandani JS, Dangarwala KJ. Moving object detection: review of recent research trends. 2015 International Conference on Pervasive Computing (ICPC). IEEE; 2015. p. 1-5.
20. Wren CR, Azarbayejani A, Darrell T, Pentland AP. Pfinder: Real-time tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997;19(7):780–785.
21. Santoyo-Morales JE, Hasimoto-Beltran R. Video background subtraction in complex environments. Journal of Applied Research and Technology. 2014;12(3):527–537.
22. Bouwmans T. Background subtraction for visual surveillance: a fuzzy approach. Handbook on Soft Computing for Video Surveillance. 2012:103–134.
23. Pilet J, Strecha C, Fua P. Making background subtraction robust to sudden illumination changes. European conference on computer vision. Springer; 2008. p. 567–580.
24. Yubing T, Cheikh FA, Guraya FFE, Konik H, Trémeau A. A spatiotemporal saliency model for video surveillance. Cognitive Computation. 2011;3(1):241–263.

25. Tu Z, Abel A, Zhang L, Luo B, Hussain A. A new spatio-temporal saliency-based video object segmentation. Cognitive Computation. 2016:1–19.
26. Li C, Cheng H, Hu S, Liu X, Tang J, Lin L. Learning collaborative sparse representation for grayscale-thermal tracking. IEEE Trans Image Process. 2016;25(12):5743.
27. Li C, Lin L, Zuo W, Yan S, Tang J. Sold: sub-optimal low-rank decomposition for efficient video segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 5519–5527.
28. Li C, Lin L, Zuo W, Wang W, Tang J. An approach to streaming video segmentation with sub-optimal low-rank decomposition. IEEE Trans Image Process. 2016;25(5):1947–1960.
29. Torre FDL, Black MJ. A framework for robust subspace learning. Int J Comput Vis. 2003;54(1–3):117–142.
30. Ke Q, Kanade T. Robust l1 norm factorization in the presence of outliers and missing data by alternative convex programming. 2005 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Volume 1. IEEE; 2005. p. 739–746.
31. Jiang B, Ding C, Tang J. Graph-laplacian pca: Closed-form solution and robustness. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2013. p. 3492–3498.
32. Candès EJ, Li X, Ma Y, Wright J. Robust principal component analysis? Journal of the ACM (JACM). 2011;58(3):1–36.
33. Zhou Z, Li X, Wright J, Candes E, Ma Y. Stable principal component pursuit. 2010 IEEE International Symposium on Information Theory. IEEE; 2010. p. 1518–1522.
34. Dou J, Li J, Qin Q, Tu Z. Moving object detection based on incremental learning low rank representation and spatial constraint. Neurocomputing. 2015;168(C):382–400.
35. Xin B, Tian Y, Wang Y, Gao W. Background subtraction via generalized fused lasso foreground modeling. 2015 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015. p. 4676–4684.
36. Li C, Wang X, Zhang L, Tang J, Wu H, Lin L. Weld: Weighted low-rank decomposition for robust grayscale-thermal foreground detection. IEEE Transactions on Circuits and Systems for Video Techniques. 2016;1(1):1–14.
37. Recht B, Fazel M, Parrilo PA. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. 2010;52(3):471–501.
38. Cai JF, Candès EJ, Shen Z. A singular value thresholding algorithm for matrix completion. SIAM J Optim. 2010;20(4):1956–1982.
39. Li SZ. Markov Random field modeling in image analysis. Springer. 2009.
40. Lu J, Shi K, Min D, Lin L, Do MN. Cross-based local multipoint filtering. 2012 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2012. p. 430–437.
41. Kopf J, Cohen MF, Lischinski D, Uyttendaele M. Joint bilateral upsampling. ACM Trans Graph (TOG). 2007;26(3):1–5.
42. Goyette N, Jodoin PM, Porikli F, Konrad J, Ishwar P. Changedetection. net: A new change detection benchmark dataset. 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE; 2012. p. 1–8.
43. Toyama K, Krumm J, Brumitt B, Meyers B. Wallflower: Principles and practice of background maintenance. The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999. Volume 1. IEEE; 1999. p. 255–261.
44. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. International Conference on Machine Learning; 2006. p. 233–240.
45. Fu Z, Sun X, Liu Q, Zhou L, Shu J. Achieving efficient cloud search services: multi-keyword ranked search over encrypted cloud data supporting parallel computing. IEICE Trans Commun. 2015;98(1):190–200.
46. Fu Z, Sun X, Ji S, Xie G. Towards efficient content-aware search over encrypted outsourced data in cloud. IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications. IEEE; 2016. p. 1–9.