ORIGINAL ARTICLE

# Collaborative model with adaptive selection scheme for visual tracking

Tianshan Liu[1] · Jun Kong[1,2] · Min Jiang[1] · Chenhua Liu[1] · Xiaofeng Gu[1] · Xiaofeng Wang[1]

Received: 4 January 2017 / Accepted: 31 July 2017 / Published online: 23 August 2017

**Abstract** Visual tracking is a challenging task since it involves developing an effective appearance model to deal with numerous factors. In this paper, we propose a robust object tracking algorithm based on a collaborative model with adaptive selection scheme. Specifically, based on the discriminative features selected from the feature selection scheme, we develop a sparse discriminative model (SDM) by introducing a confidence measure strategy. In addition, we present a sparse generative model (SGM) by combining $\ell_1$ regularization with PCA reconstruction. In contrast to existing hybrid generative discriminative tracking algorithms, we propose a novel adaptive selection scheme based on the Euclidean distance as the joint mechanism, which helps to construct a more reasonable likelihood function for our collaborative model. Experimental results on several challenging image sequences demonstrate that the proposed tracking algorithm leads to a more favorable performance compared with the state-of-the-art methods.

**Keywords** Visual tracking · Collaborative model · Adaptive selection scheme · Sparse representation

## 1 Introduction

As one of important issues in computer vision, object tracking plays a vital role in computer vision and has wide applications, such as image compression, video surveillance, activity recognition, human–computer interaction, and so on [1]. While much progress has been made in the past decades [2], designing a robust tracking algorithm is still a challenging problem due to multitudinous factors such as occlusion, illumination change, motion blur, and background clutter.

In order to track objects, many algorithms have been proposed to model object appearance, which can be divided into discriminative approaches [3–10] and generative approaches [11–18]. Discriminative methods regard the object tracking as a classification problem which concentrates on finding a decision boundary to distinguish the target from its background. Grabner et al. utilize an online boosting algorithm to select discriminative features for object tracking [4] and later in [5] a semi-online boosting algorithm is presented to alleviate the drifting problem. Babenko et al. [6] introduce multiple instance learning (MIL) to learn a discriminative model. Zhuang et al. [9] present a discriminative sparse similarity map generated from a multi-task reverse sparse coding approach with Laplacian term for visual tracking. For the purpose of utilizing the discriminative information adequately, we construct a discriminative model based on sparse representation which can select discriminative features adaptively and provide more precise classification results.

Generative tracking methods aim to learn an appearance model and search for the most similar image region to the target. Adam et al. [11] present a patch-division representation (Frag) that incorporates spatial-color features. The incremental visual tracking (IVT) method [12] applies a principal component analysis (PCA) subspace to represent the tracked object and updates the PCA subspace online to account for the

**Electronic supplementary material** The online version of this article (doi:10.1007/s13042-017-0709-1) contains supplementary material, which is available to authorized users.

✉ Jun Kong
kongjun@jiangnan.edu.cn

[1] Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, China

[2] College of Electrical Engineering, Xinjiang University, Urumqi 830047, China

dynamic environment. Mei and Ling [14] use a set of target and trivial templates to sparsely represent the target object for designing a $\ell_1$ tracker. However, the $\ell_1$ tracker needs to solve a series of $\ell_1$-minimization problems with expensive computational complexity. In addition, several methods are proposed to improve the original $\ell_1$ tracker by using accelerated proximal gradient algorithms [15], utilizing orthogonal basis vectors [16] instead of templates. To exploit the strength of both subspace learning and sparse representation, we propose a generative model via combining $\ell_1$ regularization term with PCA reconstruction, which is more robust to handle outliers and able to process high resolution images.

Recently, several hybrid generative discriminative tracking algorithms have been presented to integrate the advantages of both generative and discriminative models [19–22]. Zhong et al. [21] design two independent sparsity-based models and exploit the information from both models to construct a robust collaborative model. Zhao et al. [22] propose a hybrid generative-discriminative appearance model, which utilizes the features of the color, texture and SIFT descriptor. To our knowledge, most hybrid generative discriminative tracking methods adopt straightforward joint mechanism via multiplying the scores or confidence values computed by discriminative model and generative model directly. It can be the bottleneck to restrict the robustness of the trackers since the deviation is accumulated gradually when inaccuracy occurred in either module of the collaborative model. To solve this problem, we present a novel adaptive selection scheme to ensure that the proposed collaborative model does not undergo deterioration in tacking process.

In this paper, we present a robust tracking algorithm based on a collaborative model with adaptive selection scheme. The main work of this paper is as follows. First, we develop a sparse discriminative model (SDM) that is aimed to distinguish the target from the background. After sampling positive and negative templates to constitute the dictionary, we formulate a feature selection scheme to select discriminative features adaptively. Then a confidence measure is given to evaluate the candidates. Second, a sparse generative model (SGM) is presented with subspace representation, which models the target object and noises explicitly. Then we present an effective iterative algorithm to obtain optimal coefficients and introduce a distance function to measure the distinction between a candidate and the target. Third, we propose a novel adaptive selection scheme as the joint mechanism for our collaborative model, which is designed to detect whether the SDM or SGM is degraded and formulate appropriate likelihood function accordingly. Numerous experiments on challenging image sequences compared with state-of-the-art tracking methods demonstrate the effectiveness of the proposed algorithm.

The rest of this paper is organized as follows. Section 2 and Sect. 3 introduce the proposed SDM and SGM respectively in details. Section 4 presents our collaborative model with the proposed adaptive selection scheme. Section 5 shows the proposed tracking method including online update scheme. Extensive experimental results and analysis are discussed in Sect. 6. Section 7 concludes this paper.

## 2 Sparse discriminative model

Motivated by recent advances of sparse representation for vision tasks [17, 23–25], we propose a sparse discriminative model for object tracking.

### 2.1 Templates extraction

The dictionary is formed by $n_m$ positive templates and $n_n$ negative templates. Firstly, we extract $n_m$ image patches from the manually labeled target location in the first frame (e.g., within a rectangular region). After being normalized to the same size, each sampled image patch is put together to constitute the set of positive templates. Similarly, we sample $n_n$ negative templates further away from the marked location (e.g., within an annular region away from the target object) to form the negative template set.

### 2.2 Feature selection

In general, the original feature space is rich yet redundant, so it is necessary to adopt dimension reduction strategy. In this way, the subsequent computational cost can be decreased effectively and the determinative features can be extracted. The discriminative features are selected by following formula:

$$\min_{\mathbf{s}} \left\| \mathbf{A}^{\mathrm{T}}\mathbf{s} - \mathbf{p} \right\|_2^2 + \lambda \|\mathbf{s}\|_1, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{d\times(n_m+n_n)}$ contains $n_m$ positive templates $\mathbf{A}_+$ and $n_n$ negative templates $\mathbf{A}_-$, $d$ denotes the dimension of the original features. Each element of the vector $\mathbf{p} \in \mathbb{R}^{(n_m+n_n)\times 1}$ indicates the label of each template in the dictionary. The nonzero elements in the sparse solution vector $\mathbf{s}$ correspond to discriminative features which we want to select (i.e., the features that facilitate better classification results between the background and the target object). It is noteworthy that the feature selection strategy can choose suitable number of discriminative features adaptively in dynamic tracking environments.

The features are projected to a new subspace by a projection matrix $\mathbf{S}$. Initially, a diagonal matrix $\mathbf{S}'$ is constructed as follows:

$$\mathbf{S}'_{ii} = \begin{cases} 0, & \mathbf{s}_i = 0 \\ 1, & \text{otherwise.} \end{cases} \tag{2}$$

The dimension of the diagonal matrix $\mathbf{S}'$ is $d \times d$. The vector $\mathbf{s}$ is the solution of Eq. (1). After deleting all-zero rows in the diagonal matrix $\mathbf{S}'$, we can obtain the projection matrix $\mathbf{S}$. Both the dictionary $\mathbf{A}$ and the candidates $\mathbf{x}$ are projected to the discriminative feature space as $\mathbf{D} = \mathbf{SA}$ and $\mathbf{y} = \mathbf{Sx}$.

### 2.3 Confidence metric

The proposed SDM is designed based on the assumption that the linear combination of positive templates can represent the target better while the span of negative templates can represent the background better. Given a candidate $\mathbf{x}$, it is represented by the dictionary with the coefficients $\boldsymbol{\alpha}$ computed by

$$\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1, \tag{3}$$

where $\mathbf{y}$ and $\mathbf{D}$ are the projected matrix of the candidates $\mathbf{x}$ and the dictionary $\mathbf{A}$, respectively. $\lambda$ is a weight parameter.

The reconstruction error reflects the similarity between a candidate and the template set. Thus, the confidence value $L_c$ of the candidate $\mathbf{x}$ is formulated as follows:

$$L_c = \exp\left(\left(\varepsilon_n - \varepsilon_p\right)/\sigma\right), \tag{4}$$

where $\varepsilon_n = \left\|\mathbf{y}_c - \mathbf{D}_{\_}\boldsymbol{\alpha}_{\_}\right\|_2^2$ and $\varepsilon_p = \left\|\mathbf{y}_c - \mathbf{D}_+\boldsymbol{\alpha}_+\right\|_2^2$ are the reconstruction error of the candidate $\mathbf{y}_c$ corresponding to the negative template set $\mathbf{D}_{\_}$ and positive template set $\mathbf{D}_+$, respectively. The variable $\sigma$ is a small constant that measures the importance of the SDM module and the SGM module. With this confidence measure, candidates are evaluated in two criterions: both similarity to the foreground and difference from the background.

## 3 Sparse generative model

Despite the sparse discriminative model is successful, it is not sufficient in some challenging situations when we need to characterize the target object itself. Motivated by the success of sparse coding for object tracking [16, 26, 27], we present a sparse generative model which maintains holistic appearance information and handles outliers effectively.

### 3.1 Appearance model with subspace learning and sparse representation

In the proposed SGM, both subspace learning and sparse representation are applied to model the target appearance. It also can be regarded as a combination of $\ell_1$ regularization and PCA reconstruction. Specifically, PCA basis vectors and trivial templates are utilized to model the target appearance and outliers respectively. Given orthogonal PCA subspace $\mathbf{U} \in \mathbb{R}^{d \times k}$, where $d$ and $k$ denote the feature dimension and the number of basis vectors, respectively. The target region $\mathbf{y} \in \mathbb{R}^{d \times 1}$ can be represented by an image subspace with target coefficient $\mathbf{c} \in \mathbb{R}^{k \times 1}$ and an additional error term $\mathbf{s} \in \mathbb{R}^{d \times 1}$:

$$\mathbf{y} = \mathbf{Uc} + \mathbf{s} = [\mathbf{U}\ \mathbf{I}]\begin{bmatrix}\mathbf{c}\\ \mathbf{s}\end{bmatrix}, \tag{5}$$

where $\mathbf{I} \in \mathbb{R}^{d \times d}$ indicates the trivial template set (an identity matrix). Note that the error term $\mathbf{s}$ indicates the noise with arbitrary and sparse properties, so we can formulate the objective function as follows:

$$L(\mathbf{c}, \mathbf{s}) = \frac{1}{2}\|\mathbf{y} - \mathbf{Uc} - \mathbf{s}\|_2^2 + \lambda\|\mathbf{s}\|_1, \tag{6}$$

where $\lambda$ is the regularization parameter. As the illustration in Fig. 1, target coefficients are dense while the trivial coefficients are sparse.

### 3.2 Iterative optimization algorithm

The optimal solution can be obtained as $[\mathbf{c}^*, \mathbf{s}^*] = \arg\min_{\mathbf{c},\mathbf{s}} L(\mathbf{c}, \mathbf{s})$. The objective function in Eq. (6) can be viewed as a combination of the linear regression and a $\ell_1$ regularization term. To the best of our knowledge,

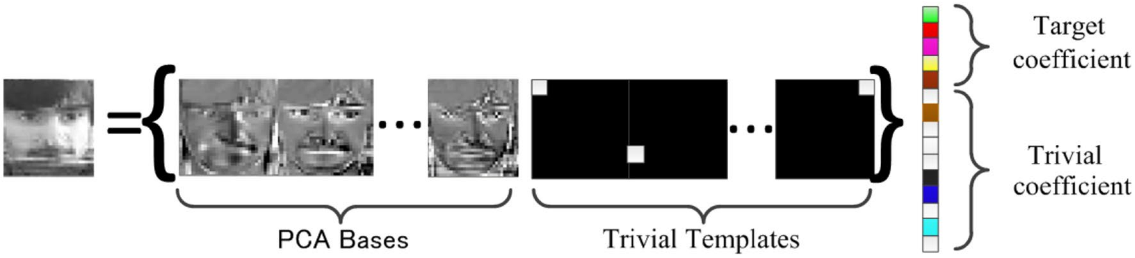

**Fig. 1** Object represented by our generative model

it is convex but not differentiable everywhere. Thus, there is no direct solution for this minimization problem, so an iterative algorithm within linear regression and soft-threshold operation is presented to compute $\mathbf{c}^*$ and $\mathbf{s}^*$.

**Step 1** Fixing the coefficient $\mathbf{s}$, the problem of solving Eq. (6) can be converted to the optimization of $G(\mathbf{c}) = \frac{1}{2}\|(\mathbf{y}-\mathbf{s}) - \mathbf{Uc}\|_2^2$. It is straightforward to obtain the optimal value by the ordinary least squares solution $\mathbf{c}^* = \mathbf{U}^{\mathrm{T}}(\mathbf{y}-\mathbf{s})$.

**Step 2** Fixing the coefficient $\mathbf{c}$, Eq. (6) can be rewritten as $J(\mathbf{s}) = \frac{1}{2}\|\mathbf{s} - (\mathbf{y}-\mathbf{Uc})\|_2^2 + \lambda\|\mathbf{s}\|_1$. The optimal $\mathbf{s}^*$ can be obtained as $\mathbf{s}^* = S_\tau(\mathbf{y}-\mathbf{Uc})$, where $S_\tau(x)$ is a soft-threshold operator [28] expressed as $S_\tau(x) = \operatorname{sgn}(x)\max(|x|-\tau, 0)$ and $\operatorname{sgn}(\cdot)$ is the sign function.

The whole iterative method for solving Eq. (6) is summarized in Algorithm 1.

**Algorithm 1** Algorithm for solving Eq. (6) to obtain $\mathbf{c}^*$ and $\mathbf{s}^*$

```
Input: The candidate sample y, the PCA subspace U, a constant $\tau$.
Initialization: $\mathbf{s}_0 = \mathbf{0}$ and i = 0
Iterate
    $\mathbf{c}_{i+1} = \mathbf{U}^{\mathrm{T}}(\mathbf{y}-\mathbf{s}_i)$
    $\mathbf{s}_{i+1} = S_\tau(\mathbf{y}-\mathbf{U}\mathbf{c}_{i+1})$
    $i \leftarrow i+1$
Until convergence or termination
Output: $\mathbf{c}^*$ and $\mathbf{s}^*$
```

### 3.3 Distance function

The distance is often formulated to vary inversely to the probability of a candidate being the target state. Therefore, a candidate with smaller distance indicates it is more likely to be a target object. Thus, we define the distance between $\mathbf{y}$ and $\mathbf{U}$ as,

$$d(\mathbf{y};\mathbf{U}) = \frac{1}{2}\|\mathbf{y} - \mathbf{Uc}^* - \mathbf{s}^*\|_2^2 + \lambda\|\mathbf{s}^*\|_1, \tag{7}$$

where $\mathbf{c}^*$ and $\mathbf{s}^*$ are calculated by Algorithm 1. As introducing the $\ell_1$ penalty term into the formulation, our distance function is able to reject outliers, which facilitates designing a robust generative model for tracking task.

To maintain consistency with the confidence measure in SDM, we formulate the similarity function based on Eq. (7) to measure the likelihood of the candidates by

$$G_c = \exp\left(-\gamma d\left(\mathbf{y}_c;\mathbf{U}\right)\right), \tag{8}$$

where $\mathbf{y}_c$ denotes the $c$-th candidate, and $\gamma$ is a constant that determines the distribution of the function.

## 4 Collaborative model with adaptive selection scheme

### 4.1 Ordinary collaborative model

Based on the demonstrated success of the collaborative model in [21, 22], we can find that most of the collaborative models are constructed by straightforward multiplicative operation. We denote them as the ordinary collaborative model (OCM). So both the confidence metric computed by the SDM and the similarity function of the SGM are combined to a unified appearance model. Thus, the likelihood function of the OCM within the proposed SDM and SGM is computed by

$$\begin{aligned} H_c &= L_c G_c \\ &= \exp\left(\left(\varepsilon_n - \varepsilon_p\right)/\sigma\right)\exp\left(-\gamma d\left(\mathbf{y}_c;\mathbf{U}\right)\right). \end{aligned} \tag{9}$$

The confidence value $L_c$ assigns higher scores to the candidates regarded as positive samples and penalizes the others [21].

### 4.2 Adaptive selection scheme

Although OCM has been demonstrated to be effective for some challenging factors, it is not always reasonable to construct the likelihood function of the collaborative model by simple multiplicative formula. As we empirically observed, when we use OCM without detecting scheme, slight inaccuracy occurred in either module of the collaborative model will degrade the whole model gradually. To address this problem, we propose a novel adaptive selection scheme (ASS) to supplement the drawbacks of OCM.

Based on the empirical observations that the difference between the target states in consecutive frames is expected to be small, we consider setting a threshold *TH* to determine whether the single model (e.g., SDM or SGM) is degraded or not. And then we can adopt appropriate strategies accordingly. Specifically, we measure the distance between the current tracking result and the previous target state by applying the Euclidean distance. Then we compare the distance with the threshold *TH*. Before introducing the proposed adaptive selection scheme, we firstly define two kinds of distance as follows:

$$d_d = \left|\mathbf{y}_t^{max\_d} - \mathbf{y}_{t-1}^*\right|, \tag{10}$$

$$d_g = \left|\mathbf{y}_t^{max\_g} - \mathbf{y}_{t-1}^*\right|, \tag{11}$$

where $\mathbf{y}_t^{max\_d}$ and $\mathbf{y}_t^{max\_g}$ denote the best candidate with maximum probability for SDM and SGM in the current frame, respectively. And $\mathbf{y}_{t-1}^*$ is the estimated target state in the previous frame.

If $d_d \leqslant TH$ and $d_g \leqslant TH$, it indicates that both SDM and SGM are not degraded, so the results computed by the two mentioned models are both reliable. The collaborative model can be treated as OCM. Thus, the likelihood function is formulated by multiplicative operation [i.e., Eq. (9)]. If $d_d > TH$ and $d_g < TH$, it means that SDM is likely to be degraded by incorrectly labeled training examples. It will result in deviation accumulation when we construct the collaborative model as OCM without any handling methods. Thus, we reserve the results from SGM and discard the results computed by SDM (i.e., $H_c = G_c$). Similarly, if $d_d < TH$ and $d_g > TH$, the confidence values from SDM are selected to be the final measure and the results computed by SGM can be negligible (i.e., $H_c = L_c$). If $d_d > TH$ and $d_g > TH$, it indicates that both SDM and SGM are likely to be degraded in varying degrees. Thus, the model with smaller bias is supposed to be selected for representing the collaborative model (e.g., if $d_d < d_g$, $H_c = L_c$, and vice versa). The steps of our adaptive selection scheme are presented in Algorithm 2.

Based on the steps mentioned in Algorithm 2, we can formulate the likelihood function of the proposed collaborative model with adaptive selection scheme (CM-ASS) as follows:

$$H_c = \begin{cases} L_c G_c & d_d \leqslant TH \text{ and } d_g \leqslant TH \\ L_c & d_g > TH \text{ and } d_d < d_g \\ G_c & d_d > TH \text{ and } d_d \geqslant d_g \end{cases}. \tag{12}$$

Compared with the simple multiplicative operation, the proposed adaptive selection scheme provides a more flexible method tending to choose the better candidate. It can avoid introducing the deterioration of the whole collaborative model which caused by the loss in SDM or SGM, thereby enabling the proposed algorithm to alleviate the tracking drift problem. The experimental results presented in Sect. 6 demonstrate the effectiveness of the proposed adaptive selection scheme. And the setting method for threshold value $TH$ is also discussed in Sect. 6.

**Algorithm 2** Adaptive Selection Scheme

```
Input: the predefined threshold $TH$, the pre-computed distance $d_d$ and $d_g$
if $d_g > TH$ and $d_d < d_g$
    $H_c = L_c$
else if $d_d > TH$ and $d_d \ge d_g$
        $H_c = G_c$
    else
        $H_c = L_c G_c$
    end
end
Output: the likelihood function $H_c$
```

## 5 Proposed tracking method

### 5.1 Model update

As the appearance of the target usually changes significantly during the tracking process, it is necessary to adopt online update scheme to alleviate drifting problem. Our model update strategy is categorized into two respects: SDM update and SGM update.

For the SDM module, the negative templates are updated every several frames, and they are all sampled from the image regions away from the current tracking result as mentioned in Sect. 2. The positive templates are unchanged during the tracking process to avoid introducing imprecise samples.It is important to sample the correct and distinct templates to maintain the discriminative properties of SDM.

For the SGM module, we introduce an effective update scheme based on the observation that the non-zero elements in the error term can be used to detect outliers. So we obtain the observation vector of the best candidate state in each frame which is represented by $\mathbf{y}_o$ and infer its corresponding error term $\mathbf{s}_o$. Then the observation vector is reconstructed by using the mean vector $\boldsymbol{\mu}$ to substitute the outliers,

$$y_r^i = \begin{cases} y_o^i & s_o^i = 0 \\ \mu^i & \text{otherwise} \end{cases}, \tag{13}$$

where $y_r^i$, $\mu^i$ and $s_o^i$ denote the $i$th element of the reconstructed vector, the mean vector, and the error vector, respectively. After cumulating enough reconstructed samples, an incremental PCA method [12] is applied to update the SGM (i.e., the PCA subspace $\mathbf{U}$ in Sect. 3).

### 5.2 Tracking with Bayesian inference

In this paper, SDM and SGM are unified into the Bayesian inference framework to develop a robust tracker. Given a series of observed vectors $\mathbf{y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_t\}$, the objective is to estimate the target state variable $\mathbf{x}_t$ by using maximum a posteriori estimation:

$$\mathbf{x}_t^* = \arg\max_{\mathbf{x}_t^i} p\left(\mathbf{x}_t^i | \mathbf{y}_{1:t}\right), \tag{14}$$

where $\mathbf{x}_t^i$ represents the $i$th sample of the state $\mathbf{x}_t$. We estimate the posterior probability $p\left(\mathbf{x}_t | \mathbf{y}_{1:t}\right)$ by the Bayesian theorem recursively:

$$p\left(\mathbf{x}_t | \mathbf{y}_{1:t}\right) \propto p\left(\mathbf{y}_t | \mathbf{x}_t\right) \int p\left(\mathbf{x}_t | \mathbf{x}_{t-1}\right) p\left(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1}\right) d\mathbf{x}_{t-1}, \tag{15}$$

where $p\left(\mathbf{x}_t | \mathbf{x}_{t-1}\right)$ indicates the motion model and $p\left(\mathbf{y}_t | \mathbf{x}_t\right)$ denotes the observation model. The motion model $p\left(\mathbf{x}_t | \mathbf{x}_{t-1}\right)$

describes the state transition between consecutive frames. We use the affine motion model and random walk to formulate the state transition, i.e., $p(\mathbf{x}_t|\mathbf{x}_{t-1}) = N(\mathbf{x}_t;\mathbf{x}_{t-1},\Sigma)$, where $\mathbf{x}_t = \{l_x, l_y, \theta, s, \alpha, \phi\}$ include the x, y translations, rotation angle, scale, aspect ratio and skew of tracked object respectively. $\Sigma$ is a diagonal covariance matrix of these affine parameters. The observation model $p(\mathbf{y}_t|\mathbf{x}_t)$ indicates the likelihood of the observation $\mathbf{y}_t$ at state $\mathbf{x}_t$. Based on the proposed collaborative model with adaptive selection scheme in Sect. 4, the likelihood function is formulated by Eq. (12) as $p(\mathbf{y}_t|\mathbf{x}_t) \propto H_c$.

The details of the proposed tracking method are summarized in Algorithm 3. The flowchart of our tracking algorithm is shown in Fig. 2.

**Algorithm 3** Collaborative model with adaptive selection scheme for visual tracking

```
Input: initial target state $\mathbf{x}_0$
for t=2 to the final frame do
    draw the candidate samples by motion model $p(\mathbf{x}_t \mid \mathbf{x}_{t-1})$, where $\mathbf{x}_{t-1}$ is the previous object state.
    represent each candidate with SDM and SGM.
    measure the similarity between the candidates and dictionary $\mathbf{A}$ or PCA subspace $\mathbf{U}$ using Eq. (4) and Eq. (8).
    calculate the likelihood function according to Eq. (12) and obtain the best object state $\mathbf{x}_t^*$.
    sample new negative templates from the image regions away the current tracking result to update the dictionary $\mathbf{A}$ in SDM every five frames.
    reconstruct the observation vector by Eq. (13), and use an incremental PCA method to update the SGM after cumulating enough reconstructed samples.
end for
Output: Tracking result $\mathbf{x}_t^*$ for each frame
```

## 6 Experiments

The proposed tracker is implemented in MATLAB and runs at 2.5 frames per second on a 3.3 GHz CPU with 8 GB memory without using parallel framework. The location of the target object is manually marked in the first frame for each sequence. To evaluate the performance of our tracker, we implement experiments on fifteen challenge image sequences that contain most challenging factors in object tracking (e.g., severe occlusion, illumination variation, background clutter, motion blur). The proposed tracker is evaluated against eight state-of-the-art algorithms, containing IVT [12], WLCS [29], SCM [21], DSST [9], ASLSA [17], $L_2$-RLS [30], OTSP [16], and $L_0$-RT [31]. Moreover, we combine the proposed SDM and SGM by straightforward multiplicative operation as an OCM tracker to observe the effectiveness of the proposed adaptive selection scheme in Sect. 6.4.

### 6.1 Implementation details

Each observation is normalized to $32 \times 32$ pixels, and 16 PCA basis vectors are utilized for PCA representation. To balance the accuracy and speed, 600 particles are adopted and the tracker is updated every five frames. The numbers of positive templates $n_m$ and negative templates $n_n$ are 50 and 200 respectively, since the background usually changes more frequently than the foreground. In order to balance the sparse representation and discriminative power, the weight parameter $\lambda$ of Eq. (1) is fixed to be 0.001, and the variable $\lambda$ in Eq. (3) is set to be 0.01. In addition, the regularization

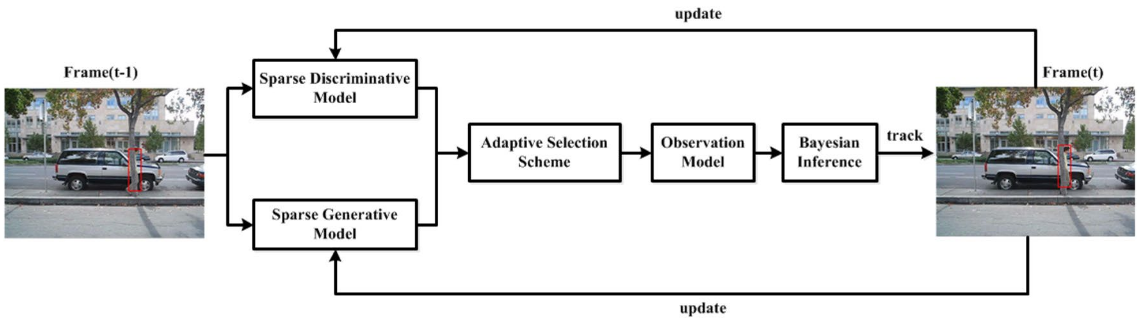

**Fig. 2** Flowchart of the proposed tracking algorithm

constants $\lambda$ in Eqs. (6) and (7) are set to 0.1 in all experiments empirically.

### 6.2 Quantitative evaluation

There are two typical criteria in quantitative evaluation of object tracking: the center location error and the overlap rate. The center location error is defined as the average distance between the predicted center locations and the ground truth. The overlap rate involves the size and pose of the target to evaluate the stability of the tracker. Given the tracking result $R_T$ and the corresponding ground truth $R_G$, the overlap rate is computed by the PASCAL VOC [32] criterion, $score = \frac{area(R_T \cap R_G)}{area(R_T \cup R_G)}$. Note that a smaller center error or a bigger overlap rate means a more favorable result. Tables 1 and 2 report the average overlap rates and average center errors of the evaluated algorithms. Figure 3 further shows the center error curve of each tracking algorithm.

### 6.3 Qualitative evaluation

*Severe Occlusion* We test several sequences (Occlusion1, Occlusion2, Caviar1, Caviar2, Caviar3, DavidOutdoor) with severe partial occlusion, rotation and pose change. Figure 4 demonstrates that the proposed algorithm performs well when the target undergoes severe occlusion. It can be explained by three reasons: (1) the SDM utilizes the discriminative information of the target object and the background to obtain better classification results; (2) the SGM handles outliers effectively and exploits subspace representation to provide sufficient generative information; and (3) the adaptive selection scheme is aimed to avoid degrading the collaborative model by discarding inaccurate model temporarily. The IVT method achieves poor performance under occlusion situation as the weak ability of handling outliers. In the Caviar3 sequence, the targets undergo heavy occlusion and the interference of similar objects. The $L_0$-RT method does not perform well since the lack of enough representative

**Table 1** Average center location error (in pixel)

| **Sequence** | IVT | WLCS | SCM | DSST | ASLSA | $L_2$-RLS | OTSP | $L_0$-RT | OCM | SDM | SGM | CM-ASS(Ours) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Occlusion1 | 12.5 | 4.3 | 3.3 | 17.1 | 9.5 | 4.7 | 4.7 | 5.2 | 5.4 | 6.7 | 5.3 | 4.6 |
| Occlusion2 | 7.8 | 4.7 | 4.7 | 8 | 4.5 | 7.7 | 4 | 3.7 | 3.6 | 7 | 3 | 3 |
| Caviar1 | 87.1 | 1.2 | 1 | 1.9 | 1.6 | 1.5 | 1.7 | 1.6 | 1.6 | 1.9 | 1.5 | 1.3 |
| Caviar2 | 4.3 | 2.7 | 2.3 | 3.6 | 1.7 | 2.3 | 2.2 | 1.9 | 2.2 | 3.9 | 2.2 | 2.2 |
| Caviar3 | 62.7 | 75.2 | 2.8 | 61.5 | 2.2 | 3.2 | 4.4 | 62.6 | 2.8 | 2.8 | 2.6 | 2.6 |
| DavidOutdoor | 52.4 | 240.6 | 67.1 | 73.3 | 86.5 | 8.9 | 8.5 | 7.9 | 52.4 | 6.8 | 52.4 | 6.6 |
| DavidIndoor | 35.9 | 8.5 | 43.5 | 13.2 | 19.9 | 129.3 | 3.2 | 152.9 | 5.7 | 108.9 | 3.8 | 3 |
| Singer1 | 12.8 | 3.6 | 3.3 | 8.8 | 3.8 | 24.9 | 4.8 | 3.6 | 3.5 | 3.5 | 5.4 | 3.4 |
| Car4 | 9.3 | 3.6 | 2.9 | 5.8 | 4.3 | 3.5 | 4 | 3.1 | 3.1 | 5.7 | 3.3 | 2.8 |
| Car11 | 1.7 | 1.8 | 1.8 | 3.5 | 1.8 | 1.7 | 2.2 | 1.9 | 1.6 | 2.8 | 1.5 | 1.5 |
| Deer | 16.5 | 9.8 | 15 | 10.3 | 7 | 9.8 | 11.3 | 12.4 | 9.8 | 15.9 | 9.8 | 9.6 |
| Football | 17.3 | 171.8 | 15.2 | 66.2 | 14.7 | 11 | 33.7 | 7.5 | 23.3 | 155.3 | 23 | 7.5 |
| Jumping | 5 | 95.9 | 3.9 | 6.8 | 5.2 | 5.2 | 5 | 4.5 | 4.7 | 5.4 | 4.8 | 4.4 |
| Owl | 126.7 | 10 | 8.5 | 6 | 5.6 | 7.2 | 47.4 | 186.8 | 111.3 | 233.8 | 119.2 | 5.9 |
| Face | 14.9 | 89.5 | 46.9 | 13.6 | 89.8 | 14.1 | 24.1 | 12.5 | 12.6 | 12.9 | 12.3 | 12 |
| **Average** | 31.1 | 48.2 | 14.8 | 20.0 | 17.2 | 15.7 | 10.7 | 31.2 | 16.2 | 38.2 | 16.7 | 4.5 |

The best three results are shown in red, blue, and green fonts. (Color table online)

**Table 2** Average overlap rate

| **Sequence** | IVT | WLCS | SCM | DSST | ASLSA | $L_2$-RLS | OTSP | $L_0$-RT | OCM | SDM | SGM | CM-ASS(Ours) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Occlusion1 | 0.80 | 0.92 | 0.94 | 0.74 | 0.85 | 0.91 | 0.91 | 0.90 | 0.89 | 0.86 | 0.89 | 0.91 |
| Occlusion2 | 0.73 | 0.82 | 0.82 | 0.73 | 0.83 | 0.74 | 0.84 | 0.83 | 0.82 | 0.72 | 0.86 | 0.86 |
| Caviar1 | 0.28 | 0.90 | 0.90 | 0.87 | 0.90 | 0.89 | 0.89 | 0.87 | 0.89 | 0.85 | 0.88 | 0.89 |
| Caviar2 | 0.62 | 0.77 | 0.80 | 0.76 | 0.83 | 0.69 | 0.71 | 0.76 | 0.81 | 0.77 | 0.81 | 0.81 |
| Caviar3 | 0.14 | 0.14 | 0.85 | 0.14 | 0.85 | 0.82 | 0.81 | 0.15 | 0.85 | 0.84 | 0.86 | 0.86 |
| DavidOutdoor | 0.52 | 0.11 | 0.38 | 0.30 | 0.46 | 0.74 | 0.74 | 0.72 | 0.57 | 0.74 | 0.57 | 0.75 |
| DavidIndoor | 0.44 | 0.69 | 0.33 | 0.57 | 0.45 | 0.23 | 0.76 | 0.27 | 0.71 | 0.36 | 0.75 | 0.77 |
| Singer1 | 0.57 | 0.85 | 0.84 | 0.66 | 0.82 | 0.47 | 0.82 | 0.81 | 0.83 | 0.82 | 0.78 | 0.83 |
| Car4 | 0.68 | 0.88 | 0.90 | 0.77 | 0.88 | 0.91 | 0.91 | 0.91 | 0.92 | 0.88 | 0.91 | 0.92 |
| Car11 | 0.82 | 0.75 | 0.80 | 0.75 | 0.83 | 0.83 | 0.81 | 0.81 | 0.84 | 0.78 | 0.84 | 0.85 |
| Deer | 0.54 | 0.61 | 0.56 | 0.62 | 0.63 | 0.60 | 0.58 | 0.55 | 0.61 | 0.54 | 0.61 | 0.63 |
| Football | 0.58 | 0.27 | 0.61 | 0.48 | 0.59 | 0.66 | 0.62 | 0.64 | 0.50 | 0.27 | 0.51 | 0.70 |
| Jumping | 0.61 | 0.04 | 0.73 | 0.61 | 0.67 | 0.67 | 0.69 | 0.64 | 0.67 | 0.67 | 0.67 | 0.67 |
| Owl | 0.22 | 0.77 | 0.77 | 0.78 | 0.79 | 0.78 | 0.48 | 0.23 | 0.50 | 0.19 | 0.48 | 0.82 |
| Face | 0.71 | 0.47 | 0.56 | 0.75 | 0.30 | 0.72 | 0.68 | 0.75 | 0.75 | 0.76 | 0.75 | 0.76 |
| **Average** | 0.55 | 0.60 | 0.72 | 0.64 | 0.71 | 0.71 | 0.75 | 0.66 | 0.74 | 0.67 | 0.74 | 0.80 |

The best three results are shown in red, blue, and green fonts. (Color table online)

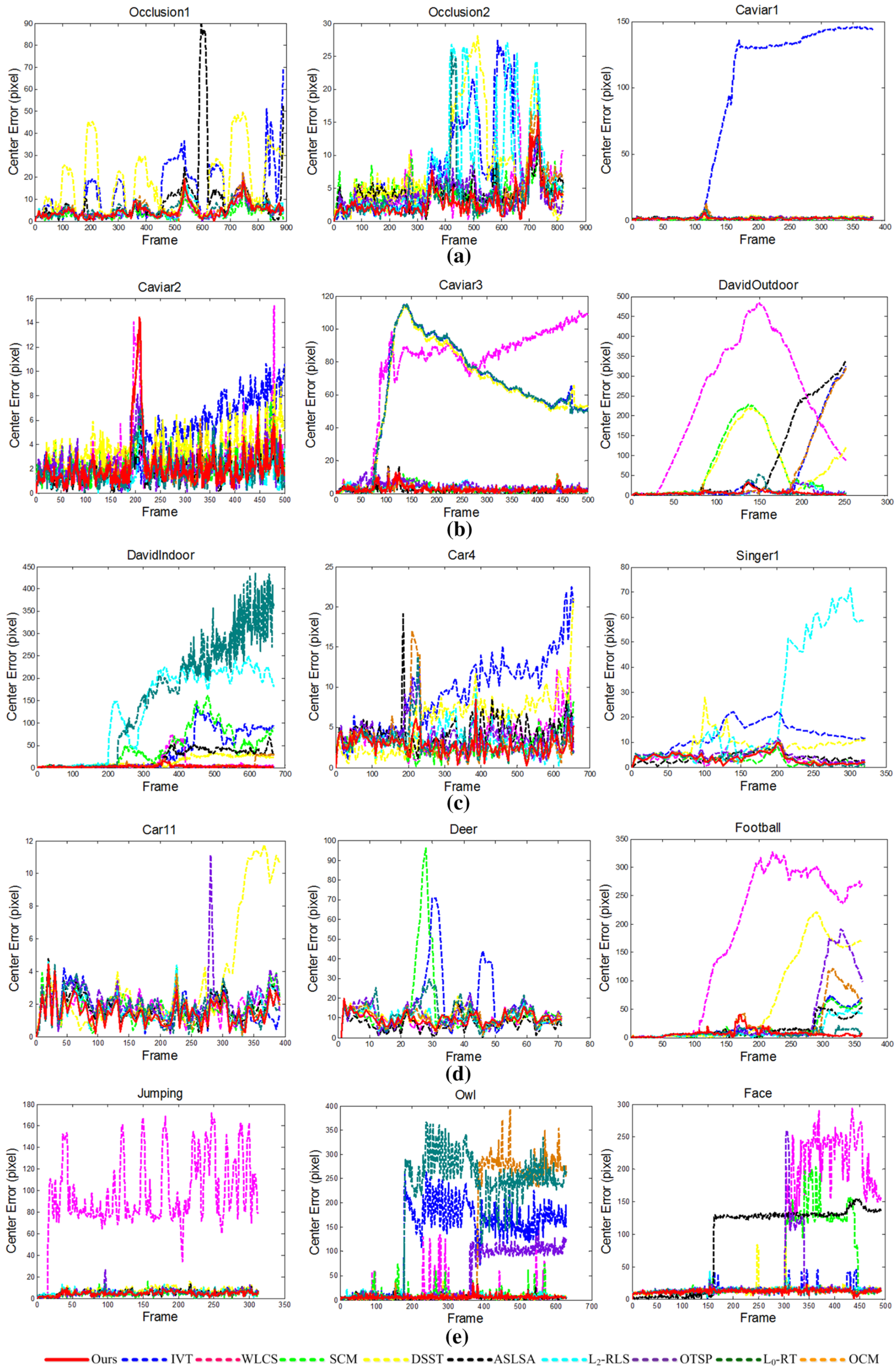

**Fig. 3** Quantitative evaluation of the trackers in terms of center location error (in pixel). **a** Occlusion1, Occlusion2 and Caviar1. **b** Caviar2, Caviar3, and DavidOutdoor. **c** DavidIndoor, Car4 and Singer1. **d** Car11, Deer and Football. **e** Jumping, Owl and Face

power. In the DavidOutdoor sequence, most trackers lose the target while our tracker performs stably during the whole sequences. The IVT, WLCS, SCM and ASLSA methods are susceptible to appearance changes caused by occlusion and pose (e.g., #085, #155), whereas the trackers $L_2$-RLS, OTSP, and $L_0$-RT perform better since they all take outliers into consideration.

*Illumination change* Fig. 5 illustrates the tracking results in the sequences (DavidIndoor, Car4, Singer1) with drastic illumination change and pose variation. Due to the use of discriminative features from SDM and incremental PCA method in SGM, the proposed method performs well in handling the illumination change. In the singer1 sequence, the stage light and scale change drastically. Likewise, in the DavidIndoor sequence, when the person walks from a dark room into a bright area with pose variation, his appearance changes drastically. The $L_2$-RLS tracker is less effective in both of the two sequences (e.g., DavidIndoor #252 and Singer1 #264) for the weak sparse projection coefficient causing the features redundant. The SCM tracker does not perform well in the DavidIndoor sequence (e.g., #380 and #580) as it is susceptible to deformation.

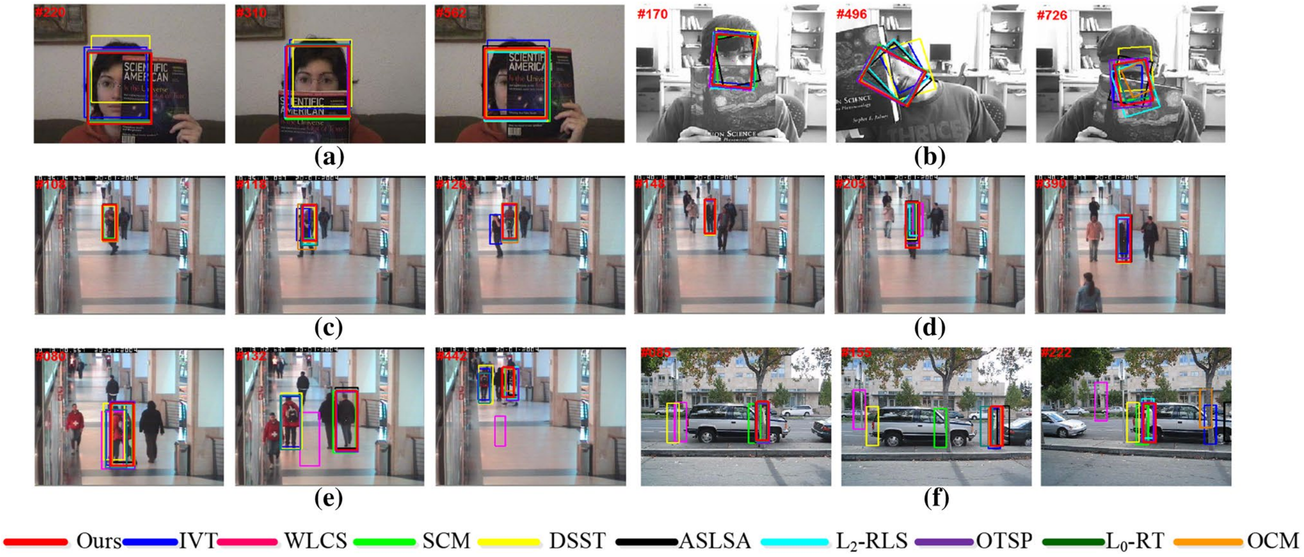

**Fig. 4** Sample tracking results of evaluated algorithms on six image sequences with severe occlusion, in-plane rotation and pose variation. **a** Occlusion1. **b** Occlusion2. **c** Caviar1. **d** Caviar2. **e** Caviar3. **f** DavidOutdoor

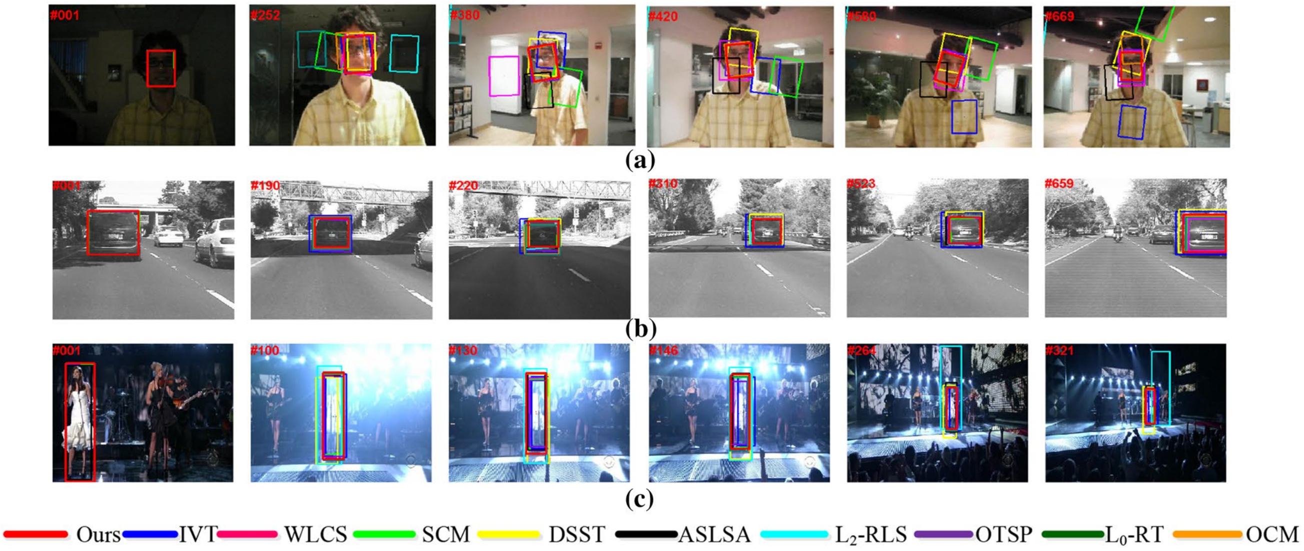

**Fig. 5** Sample tracking results of evaluated algorithms on three image sequences with illumination changes. **a** DavidIndoor. **b** Car4. **c** Singer1

*Background clutter* Fig. 6 shows the tracking results in the sequences (Car11, Deer and Football) with background clutters. Since SGM maintains holistic information and the proposed SDM provides accurate classification results, the proposed algorithm performs better against other methods in these sequences (e.g., Deer #029, Football #320). In Deer sequence, it is difficult to locate the target accurately as the background are similar to the foreground (e.g., #029). In Football sequence, there are many players with similar equipments as the target person in court field. Our tracker locates the target accurately during the sequence whereas most of the other trackers drift away from the target (e.g., #320). The WLCS tracker does not perform well (e.g., #114) since it concentrates on local model and ignores holistic template. The OTSP tracking algorithm loses the target (e.g., #320) since the lack of background information makes its discriminative power poor.

*Motion Blur* Fig. 7 demonstrates the tracking results on the sequences (Jumping, Owl and Face) with motion blur. When the tracked target undergoes motion blur, it is a challenging task to estimate the location of the target precisely. Due to the strong discriminative power provided by SDM and the use of PCA representation in SGM, the proposed method performs more stably than the other approaches (e.g., Owl #568 and Face #302). The WLCS tracker fails to track the target at the beginning of the Jumping sequence (e.g., Jumping #067). In Owl sequence, frequent motion of the camera results in blurred appearance. The OTSP tracker can capture the target in some frames (e.g., #280) but drifts when drastic motion occurs (e.g., #444). For the Face sequence, the ASLSA method does not perform well (e.g., #248) since it only focuses on foreground information.

### 6.4 Analysis and discussion

*Effect of threshold value TH* Since an appropriate threshold value *TH* is important to our adaptive selection scheme, we conduct experiments on four representative image sequences to explore the effects of the threshold value *TH* in a reasonable range. Figure 8 shows the overlap rate curve of the proposed algorithm with different threshold values. Generally, very small value of *TH* leads to a poor performance, because the joint mechanism is sensitive to the difference between two consecutive frames. Thus, in most cases, the tracker is supposed to select a single model (SDM or SGM) as the final measure, which is unable to utilize superiority of the collaborative model. When *TH* is too large, the performance starts to degrade as the constraint is relaxed. So the selection scheme is not able to find the degraded model, and the likelihood function of the collaborative model tends to be OCM [i.e., Eq. (9)] in most cases. It is observed that when *TH* is equal to 0.12, our tracker achieves the best performance over all used image sequences. So we set *TH* = 0.12 for the proposed CM-ASS.

*Effectiveness of adaptive selection scheme* As aforementioned, we propose SDM and SGM respectively, and then combine the proposed SDM and SGM using ASS strategy. To reveal the effectiveness of ASS strategy, we compare the tracker that integrating the proposed SDM and SGM using straightforward multiplicative operation (named as OCM) with the proposed CM-ASS. For fair comparison, these two trackers are evaluated with appropriate parameter settings and the best experimental results for each tracker are presented. Figure 9 illustrates the different tracking centroids in y direction on two sequences DavidOutdoor and Football

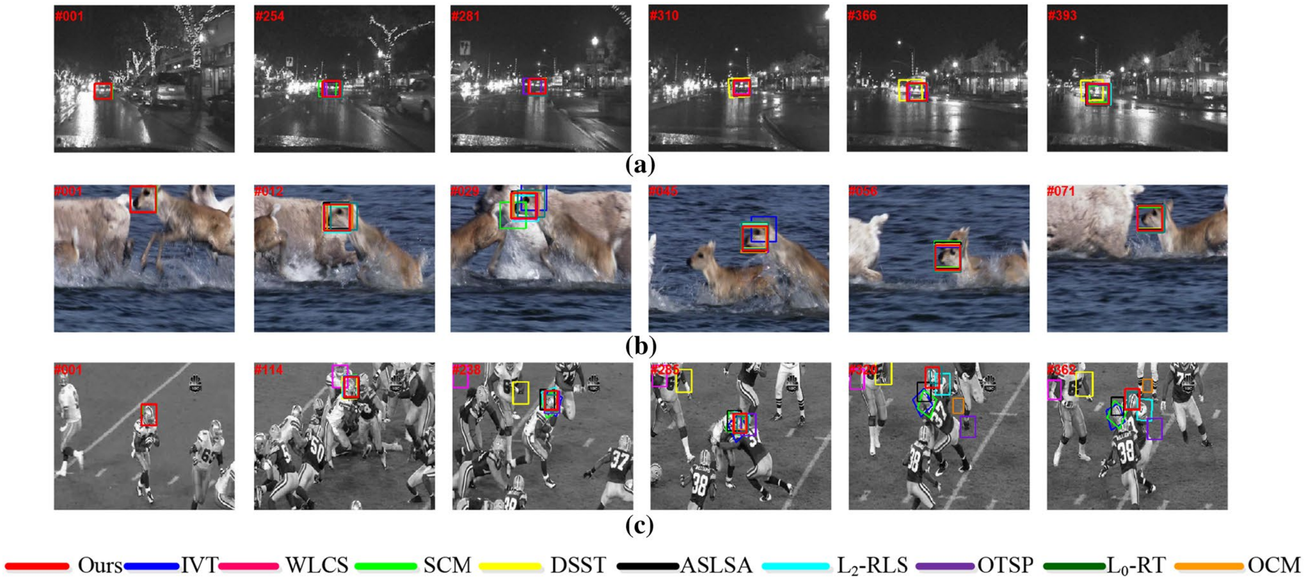

**Fig. 6** Sample tracking results of evaluated algorithms on three image sequences with background cluttered. **a** Car11. **b** Deer. **c** Football

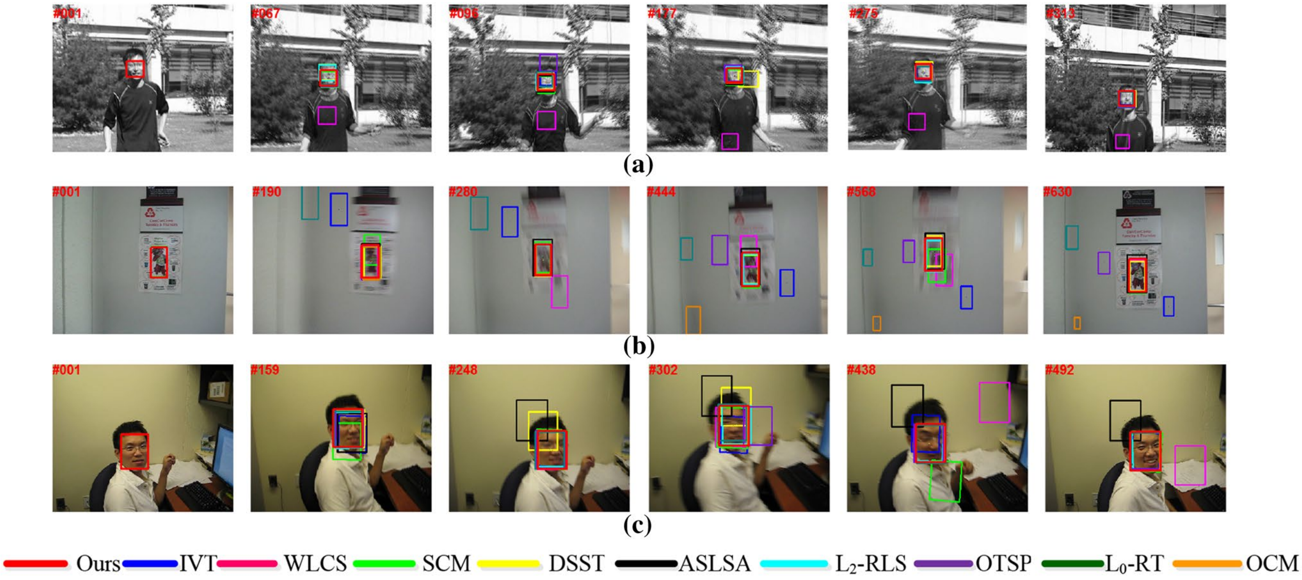

**Fig. 7** Sample tracking results of evaluated algorithms on three image sequences with motion blur. **a** Jumping. **b** Owl. **c** Face

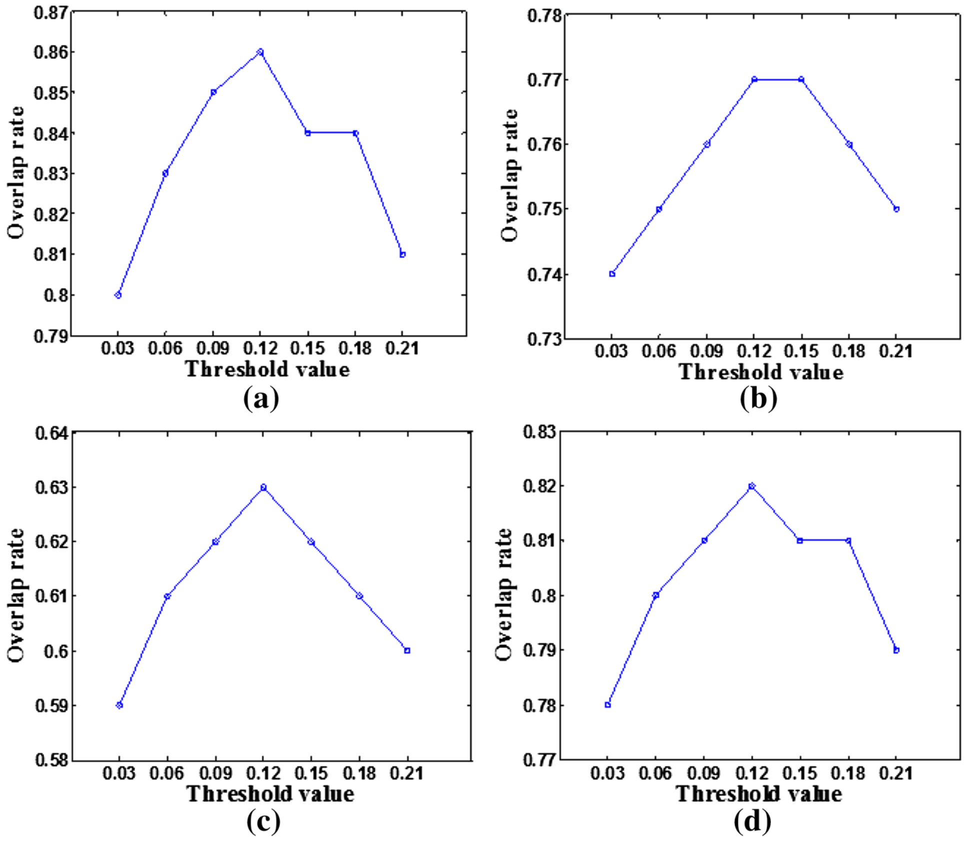

**Fig. 8** Overlap rate curve of the proposed algorithm with different threshold values on four representative image sequences. **a** Caviar3. **b** DavidIndoor. **c** Deer. **d** Owl

using OCM/CM-ASS. If the target undergoes partial occlusion as shown in Fig. 9a, the OCM method will drift away from the target. When the football player undergoes cluttered background situations (e.g., Fig. 9b), the tracker with OCM loses track of the target. While the proposed CM-ASS is able to locate the target successfully and approach to the ground truth curve. This can be attributed to the effectiveness of ASS. The proposed ASS is able to detect whether the SDM or SGM is degraded timely by a distance metric. Then it helps the tracker discard inaccurate model temporarily and construct a more reasonable likelihood function to evaluate the candidates for the current frame. Thus, the proposed ASS is effective to avoid introducing unreliable results.

*CM-ASS vs SCM* SCM is a typical tracking algorithm based on collaborative model. As shown in Tables 1 and 2, CM-ASS achieves better performance than SCM in most cases. Since the integration of holistic information and part-based representation, SCM can perform well when there are no drastic changes in the appearance of the object or less disturbed factors (e.g., Caviar1 and Singer1). Nevertheless, we find out that SCM degrades when the target undergoes large appearance changes, especially in the case of abrupt motion (e.g., #304 of Face in Fig. 7c), illumination variation (e.g., #252 of DavidIndoor in Fig. 5a) and pose change (e.g., #155 of DavidOutdoor in Fig. 4c). This can be explained by the fact combining the discriminative model and the generative model directly without any detecting mechanism is not always reliable. Inaccuracy occurred in one of the collaborative model can deteriorate the whole model gradually with drifts. The proposed CM-ASS tracker is well designed to address this problem and performs more stably against SCM.

*CM-ASS vs SDM/SGM* Since the proposed CM-ASS is based on two modules:SDM and SGM, we further demonstrate how they complement each other and the merits of CM-ASS. The experimental results in term of SDM and SGM are also presented in Tables 1 and 2. When the target object undergoes severe occlusion and drastic illumination change (e.g., DavidOutdoor and Singer1), the SDM tracker achieves better performance than SGM tracker. The reason is that the SDM is designed to differentiate the target from the background and has strong discriminative power. On the other hand, in some cases (e.g., Occlusion2, DavidIndoor and Deer), the SGM tracker is more effective than the SDM tracker. This can be attributed to the fact that SGM handles outliers effectively and maintain enough representative power. Over all, the collaborative model performs better than or equal to the SDM and SGM individually. As shown in Fig. 10, both SDM and SGM tracker lose the target in Owl sequence whereas the proposed CM-ASS performs well. This can be attributed to that the proposed ASS enables the

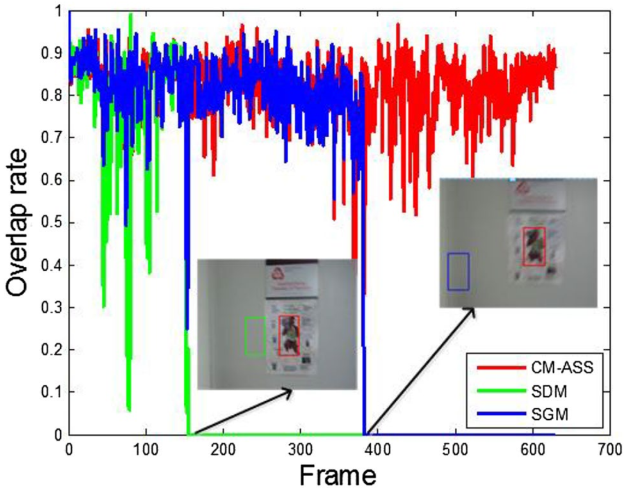

**Fig. 10** Overlap rate of tracking algorithms based on the CM-ASS, SDM and SGM on Owl sequence

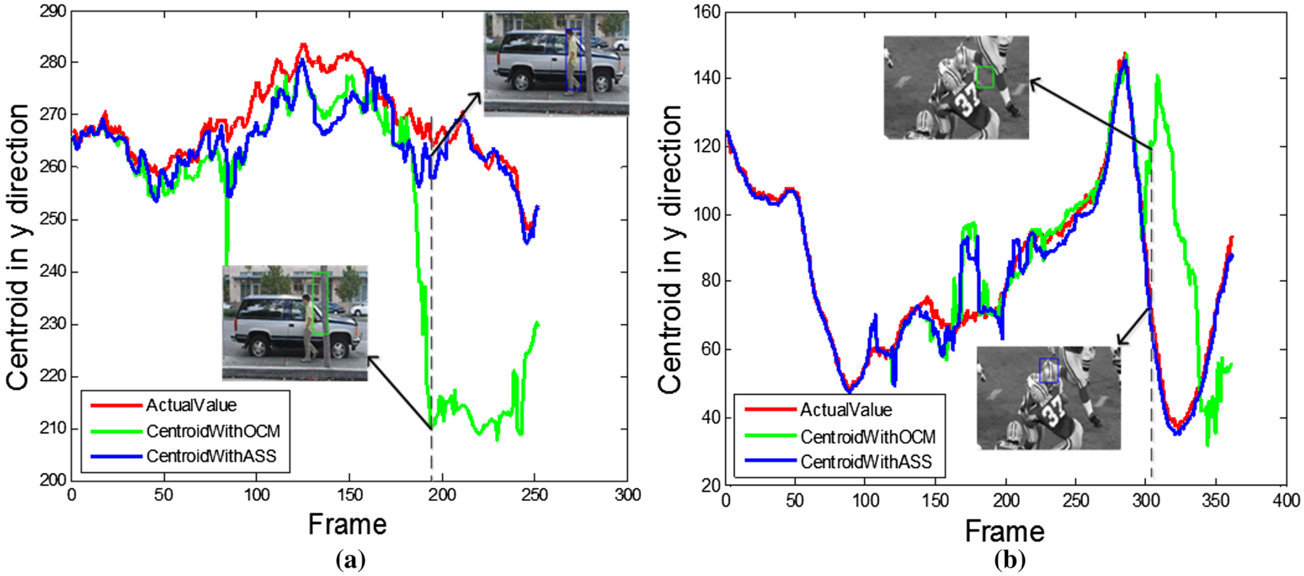

**Fig. 9** Centroids of objects using different joint mechanism, i.e., OCM vs ASS. **a** centroid in *y direction* on DavidOutdoor sequence. **b** centroid in *y direction* on Football sequence

**Table 3** Running time comparison results of different tracking algorithms in frames per second (fps)

| Tracker | IVT | WLCS | SCM | DSST | ASLSA | $L_2$-RLS | OTSP | $L_0$-RT | Ours |
|---|---|---|---|---|---|---|---|---|---|
| Average fps | 9.0 | 7.8 | 0.2 | 2.4 | 2.2 | 7.5 | 3.7 | 8.1 | 2.5 |

collaborative model to integrate the superiority of SDM and SGM modules effectively.

### 6.5 Computational complexity

The most time consuming part of the proposed tracking method is to compute the optimal coefficients using the templates and PCA basis vectors for SDM and SGM respectively. For SDM module, the coefficients are computed by LASSO algorithm, so its time complexity is $O\left(d^2 + d\left(n_m + n_n\right)\right)$, where $d$ denotes the dimension of an image observation, $n_m + n_n$is the sum of positive and negative templates. For SGM module, the computational load is mainly from step 3 in Algorithm 1, so the complexity is $O(ndk)$, where $k$ represents the number of PCA basis vectors, $n$ indicates the number of iterations in Algorithm 1. So the computational complexity of our tracking algorithm is $O\left(d^2 + d\left(n_m + n_n\right) + ndk\right)$. The computational efficiency of different trackers is presented in Table 3. For fair comparison, the evaluated tracking algorithms are implemented in MATLAB using source code. As shown in Table 3, the proposed tracker is faster than SCM, DSST, ASLSA tracker. The running time of our algorithm can be further reduced by parallel computation framework.

## 7 Conclusion

In this paper, we propose a robust and effective tracking algorithm based on a collaborative model with adaptive selection scheme. Based on the discriminative features extracted from positive and negative template sets, the sparse discriminative model differentiates the target from the background via the feature selection scheme and confidence measure strategy. The sparse generative model that combines $\ell_1$ regularization with subspace learning is effective to handle outliers and has strong representation power. In addition, a novel adaptive selection scheme based on Euclidean distance is presented as the joint mechanism to construct a more reliable likelihood function, which facilitates better performance compared with the existing hybrid generative discriminative tracking algorithms. The proposed discriminative and generative models are integrated in a Bayesian inference framework by the adaptive selection scheme. Furthermore, the template sets and PCA subspace are updated with different schemes to alleviate drift problem and enhance the proposed algorithm to handle appearance changes during dynamic environments. Quantitative and qualitative evaluations validate that the proposed method can achieve more robust performance compared with several competitive algorithms. In the future, we plan to utilize the local features of image patches for more effective object tracking.

**Acknowledgements** This work was partially supported by the National Natural Science Foundation of China (61362030, 61201429), the Project Funded by China Postdoctoral Science Foundation (2015M581720, 2016M600360), the Project Funded by Jiangsu Postdoctoral Science Foundation (1601216C) and Technology Research Project of the Ministry of Public Security of China (2014JSYJB007).

## References

1. Yilmaz A, Javed O, Shah M (2006) Object tracking: a survey. Acm Comput Surv 38(4):81–93
2. Li X et al. (2013) A survey of appearance models in visual object tracking. Acm Trans Intell Syst Technol 4(4):478–488
3. Avidan S, Ensemble tracking (2007) IEEE Trans Pattern Anal Mach Intell 29(2):261–271
4. Grabner H, Bischof H (2006) On-line boosting and vision. In: IEEE computer society conference on computer vision and pattern recognition, pp 260–267
5. Grabner H, Leistner C, Bischof H (2008) Semi-supervised on-line boosting for robust tracking. In: European conference on computer vision, pp 234–247
6. Babenko B, Yang MH, Belongie S (2011) Robust object tracking with online multiple instance learning. IEEE Trans Pattern Anal Mach Intell 33(8):1619–1632
7. Kalal Z, Mikolajczyk K, Matas J (2012) Tracking-learning-detection. IEEE Trans Pattern Anal Mach Intell 34(7):1409–1422
8. Jiang N, Liu W, Wu Y (2011) Learning adaptive metric for robust visual tracking. IEEE Trans Image Process 20(8):2288–2300
9. Zhuang, Lu H, Xiao Z (2014) Visual tracking via discriminative sparse similarity map. IEEE Trans Image Process 23(4):1872–1881
10. Wu G, Zhao C, Lu W et al (2015) Efficient structured $L_1$ tracker based on laplacian error distribution. Int J Mach Learn Cybern 6(4):581–595
11. Adam A, Rivlin E, Shimshoni I (2006) Robust fragments-based tracking using the integral histogram. In: IEEE computer society conference on computer vision and pattern recognition, pp 798–805
12. Ross J, Lim J, Lin RS, Yang M (2008) Incremental learning for robust visual tracking. Int J Comput Vision 77(1–3):125–141
13. Kong J, Liu C, Jiang M, Wu J, Tian S, Lai H (2016) Generalized $\ell_P$-regularized representation for visual tracking. Neurocomputing 213:155–161
14. Mei X, Ling H (2009) Robust visual tracking using $L_1$ minimization. In: IEEE international conference on computer vision, pp 1436–1443

15. Bao C, Wu Y, Ling H, Ji H (2012) Real time robust $L_1$ tracker using accelerated proximal gradient approach. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1830–1837
16. Wang, Lu H, Yang M-H (2013) Online object tracking with sparse prototypes. IEEE Trans Image Process 22(1):314–325
17. Jia X, Lu H, Yang M (2012) Visual tracking via adaptive structural local sparse appearance model. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1822–1829
18. Liu H, Yuan M, Sun F, Zhang J (2014) Spatial neighborhood-constrained linear coding for visual object tracking. IEEE Trans Ind Inf 10(1):469–480
19. Liu R, Cheng J, Lu H (2009) A robust boosting tracker with minimum error bound in a co-training framework. Proceedings 30(2):1459–1466.
20. Dinh TB, Medioni GG (2011) Co-training framework of generative and discriminative trackers with partial occlusion handling. In: IEEE workshop on the applications of computer vision, pp 642–649.
21. Zhong W, Lu H, Yang M (2014) Robust object tracking via sparse collaborative appearance model. IEEE Trans Image Process 23(5):2356–2368
22. Zhao L, Zhao Q, Chen Y, Lv P (2016) Combined discriminative global and generative local models for visual tracking. J Electron Imaging 25(2):023005
23. Wang Q, Chen F, Xu W, Yang M-H (2012) Online discriminative object tracking with local sparse representation. In: IEEE workshop on the applications of computer vision, pp 425–432.
24. Zhang T, Ghanem B, Liu S, Ahuja N (2013) Robust visual tracking via structured multi-task sparse learning. Int J Comput Vis 101(2):367–383
25. Zhang T, Ghanem B, Liu S, Ahuja N (2012) Low-rank sparse learning for robust visual tracking. In: European conference on computer vision, pp 2042–2049
26. Wang D, Lu H, Yang M (2016) Robust visual tracking via least soft-threshold squares. IEEE Trans Circuits Syst Video Technol 26(9):1709–1721
27. Wang D, Lu H, Bo C (2015) Fast and robust object tracking via probability continuous outlier model. IEEE Trans Image Process 24(12):5166–5176
28. Hale T, Yin W, Zhang Y (2008) Fixed-point continuation for $\ell_1$-minimization: methodology and convergence. Siam J Optim 19(3):1107–1130
29. Wang D, Lu H, Bo C, Visual (2014) Tracking via weighted local cosine similarity. IEEE Trans Cybern 45(9):1838–1850
30. Xiao Z, Lu H, Wang D (2014) L2-RLS-based object tracking. IEEE Trans Circuits Syst Video Technol 24(8):1301–1309
31. Pan J, Lim J, Su Z, Yang M (2014) $L_0$-regularized object representation for visual tracking. In: Proceedings British machine vision conference
32. Everingham M, Van Gool L, C. K. I. Williams, Winn J, Zisserman A (2010) The pascal visual object classes (voc) challenge. Int J Comput Vis 88(2):303–338