Machine Vision and Applications (2018) 29:749–763
https://doi.org/10.1007/s00138-018-0930-2

ORIGINAL PAPER

Visual object tracking via coefficients constrained exclusive group LASSO

Xiao Ma[1] · Qiao Liu[1] · Weihua Ou[2,3] · Quan Zhou[4,5]

Received: 3 August 2016 / Revised: 10 March 2018 / Accepted: 2 April 2018 / Published online: 30 April 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract
Discriminative methods have been widely applied to construct the appearance model for visual tracking. Most existing methods incorporate online updating strategy to adapt to the appearance variations of targets. The focus of online updating for discriminative methods is to select the positive samples emerged in past frames to represent the appearances. However, the appearances of positive samples might be very dissimilar to each other; traditional online updating strategies easily overfit on some appearances and neglect the others. To address this problem, we propose an effective method to learn a discriminative template, which maintains the multiple appearances information of targets in the long-term variations. Our method is based on the obvious observation that the target appearances vary very little in a certain number of successive video frames. Therefore, we can use a few instances to represent the appearances in the scope of the successive video frames. We propose exclusive group sparse to describe the observation and provide a novel algorithm, called *coefficients constrained exclusive group LASSO*, to solve it in a single objective function. The experimental results on CVPR2013 benchmark datasets demonstrate that our approach achieves promising performance.

Keywords Discriminative methods · Samples selection · Template matching · Exclusive group LASSO

1 Introduction

Visual tracking is one of the fundamental tasks in computer vision. Numerous tracking algorithms have been introduced in the past decades [1,2,19,20,26,27,34,38,42,46]; however, visual tracking still suffers many challenging problems, such as varying illumination, pose variations and shape deformation. In general, there are two different paradigms to tackle such variations: generative methods and discriminative methods. The generative methods learn an appearance model to represent the targets and leverage the reconstruction errors to find the optimal results. These methods rely on the representativeness of the appearance model and require the appearance model to cover all appearance variations of targets.

Discriminative methods have been successful used in image segmentation and denoising [7,16,37,39,40,45,57], face and action recognition [8,9,28,30,50–52], handwriting identification [21,23,24] and objects tracking. The discriminative methods learn a discriminative model with supervised information to distinguish the targets from surrounding back-

Xiao Ma and Qiao Liu contributed equally to this work and should be considered co-first authors. Weihua Ou and Quan Zhou are the corresponding authors.

✉ Weihua Ou
ouweihuahust@gmail.com; Weihua.Ou@uts.edu.au

Xiao Ma
turingki@gmail.com

Qiao Liu
liuqiao.hit@gmail.com

Quan Zhou
quan.zhou@njupt.edu.cn

1 School of Computer Science, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, People's Republic of China

2 School of Big Data and Computer Science, Guizhou Normal University, Guiyang, People's Republic of China

3 Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, Australia

4 National Engineering Research Center of Communications and Networking, Nanjing University of Posts and Telecommunications, Nanjing, People's Republic of China

5 Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University, 350121 Fuzhou, People's Republic of China

grounds. To adapt the appearance variations of targets, most existing discriminative methods incorporate online updating strategy to adjust the model. The main concern of online updating is how to tackle the tracked results since they represent the appearances that emerge in past frames. The intuitive approach is to consider these tracked results as positive samples and to retrain the models in updating stages. This approach has been widely used in many discriminative methods [3,17,32,56]. However, in a real application, the positive samples may be very dissimilar to each other, which would lead discriminative models to be easily overfitted on the appearances of larger proportions in the positive sample set. Thus, significant drift would happen when the object is not tracked correctly in some certain frames.

One of remedy approaches is re-weighting some key samples during the training processes [15,19,33,54]. For example, Hare et al. proposed a method named Struck in [19] to automatically re-weight the support vectors under the framework of the structured output support vector machine (SSVM). Struck would randomly remove the old support vectors if they no longer play an important role in the construction of the SSVM classifier. It is not suitable when the targets undergo periodic appearance variations. Li et al. [33] provided a time-weighted sampling strategy to assign larger weight to the recently appended samples and smaller weight for old ones. This strategy would fail when the recently appended samples are misaligned or contaminated with noises. In [54], Yang et al. proposed an active example selection strategy to automatically select the most informative samples using Laplacian regularized least squares and incorporated them into a semi-supervised framework. Gao et al. [15] provided an individual module to re-weight all auxiliary samples (old samples from very early frames) by considering distances between all pairs of samples, which can be used to avoid the tracker from overfitting in the coming frames.

In this paper, we exploit the multiple appearance variations of targets to learn the compact and discriminative templates based on the assumption that all the appearances of a certain object and the templates resize on the same subspace. This assumption has been widely applied in the sparse representation-based visual trackers [5,43]. Specifically, we formulate the template into the linear combination of object appearances and surroundings, as shown in Fig. 1. To enhance the discriminative ability of template, we add a set of surroundings as negative samples in such a linear representation. In fact, appearances of the object vary very little in a certain number of successive video frames in the real application. That means that a sequence consists of many local sequence fragments and the object appearances in each of them are very similar. Therefore, we can select a few appearances from each of the groups as the atoms of the linear representation of the template. We propose

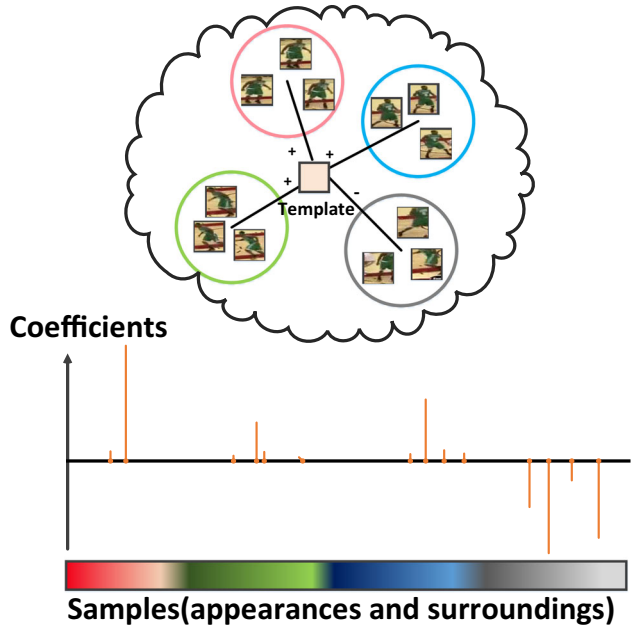


Fig. 1 The motivation of the proposed method. In our approach, the template is represented by a linear combination of sample set, which consists of the appearances of object (marked in color) and surroundings (marked in gray). The appearances can be divided into different groups according to their similarities. We impose two constraints on the coefficients of sparse representation: (1) the coefficients of appearances are positive and are intra-group sparse; (2) the coefficients of surroundings are all negative

coefficients constrained exclusive group LASSO to describe such a selection process and present a novel algorithm to solve the associated model. The experimental results on the CVPR2013 benchmark datasets demonstrate the effectiveness and robustness.

The contributions of this paper are summarized as follows:

- a coefficient constrained exclusive group LASSO model was proposed, which can learn a compact and discriminative template.
- a novel updating algorithm was presented for the coefficients constrained exclusive group LASSO, and the experimental results demonstrate the effectiveness of the proposed method.

The rest of paper is organized as follows. In Sect. 2, we introduce some related works. In Sect. 3, we present the proposed method. Then, we give the experimental results in Sect. 4 and conclude this paper in Sect. 5.

2 Related works

In this section, we briefly review the template matching in Sect. 2.1, present the sparse representation-based visual

tracking in Sect. 2.2 and give the exclusive group LASSO in Sect. 2.3.

2.1 Template matching in tracking

Template matching is a kind of classical method for visual tracking and other computer vision tasks [4,12,41,47]. Lucas and Kanade [4] introduce a popular method to register the deformed view of a template to a reference view by minimizing a linear approximation of intensity image differences with gradient decent approach. Comaniciu and Meer [12] incorporate the mean-shift method to accelerate matching procedure. One of the fundamental problems of template matching is how to design a template updating strategy to update the template with appearance variations of object [41]. To handle this problem, Black and Jepson [6] propose a subspace-based approach to construct and update the template. In [6], the template is composed of a set of "eigen-templates" and the matching procedure is implemented by calculating the reconstruction error between the test samples and the "eigen-templates". Later, Ross et al. [46] extend this idea by integrating the incremental principal component analysis (IPCA) and particle filtering [26]. The works in [46] motivate researchers to find many more effective methods to construct "eigen-templates" and the reconstruction procedure. Among them, the most famous method probably is the sparse representation [27,42,56], which is first introduced in [42]. The sparse representation-based visual tracking can be considered as the multiple templates approach, in which the templates are the atoms of the dictionary.

2.2 Sparse representation-based tracking

Given the target template set $\mathbf{H} = \{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n\} \in \mathbb{R}^{d\times n}$, the target candidate $\mathbf{t} \in \mathbb{R}^{d\times 1}$ can be approximately represented as the linear combination of the templates as follows:

$$\mathbf{t} \approx \mathbf{Ha} = a_1\mathbf{h}_1 + a_2\mathbf{h}_2 + \cdots + a_n\mathbf{h}_n, \tag{1}$$

where $\mathbf{a} = [a_1, a_2, \ldots, a_n]^T \in \mathbb{R}^n$ is the coefficient vector. Based on the sparse assumption, the coefficient vector $\mathbf{a}$ can be obtained via minimizing following objective function:

$$\min_{\mathbf{a}} \frac{1}{2}\|\mathbf{t} - \mathbf{Ha}\|_2^2 \qquad s.t. \quad \|\mathbf{a}\|_0 \le \alpha, \tag{2}$$

where $\|\cdot\|_2$ represent ℓ_2 norm,$\|\cdot\|_0$ represent ℓ_0 norm, and α is the sparsity level, respectively.

Usually, we exploit ℓ_1 norm to approximate ℓ_0 norm because the problem (2) is a NP-hard problem. Thus, equation (2) can be rewritten as follows:

$$\widetilde{\mathbf{a}} = arg\min_{\mathbf{a}} \left\{ \frac{1}{2}\|\mathbf{t} - \mathbf{Ha}\|_2^2 + \lambda\|\mathbf{a}\|_1 \right\} \tag{3}$$

where λ is regularization parameter and $\|\cdot\|_1$ denotes ℓ_1 norm.

When the coefficient vector $\widetilde{\mathbf{a}}$ is obtained by Eq. (3), we can evaluate the reconstruction error of each candidate target by the following formulation:

$$E(\mathbf{t}) = \|\mathbf{t} - \mathbf{H}\widetilde{\mathbf{a}}\|_2. \tag{4}$$

Finally, the candidate with smallest reconstruction error is regarded as the tracked result.

A lot of trackers [5,18,22,27,35,36,42,43,49,56] based on sparse representation have been proposed in past several years. Usually, these methods can be classified as two categories: holistic sparse representation [5,18,36,42,43] and local sparse representation [22,27,35,49,56]. Considering the holistic feature, Mei et al. [42] cast the tracking problem as finding a sparse approximation in a template subspace. They adopt the holistic representation of the object as the appearance model and then track the object by solving the ℓ_1 minimization problem (ℓ_1 tracker). To address the computational cost of the ℓ_1 tracker, Bao et al. [5] proposed a new ℓ_1 norm related minimization model based on the accelerated proximal gradient approach (ℓ_1-APG) which can run in real time. Liu et al. [36] also proposed a two-stage sparse optimization algorithm to deal with the computational cost problem. In contrast to the holistic sparse representation, the local sparse representation encodes the each local patch of a target sparsely with an over-complete dictionary, and then aggregate the corresponding sparse codes. For instances, Jia et al. [27] proposed a structural local sparse appearance model which exploits both partial information and spatial information of the target based on a novel alignment-pooling method. Liu et al. [35] also presented a robust tracking algorithm using a local sparse appearance model, which used a static sparse dictionary and a dynamically online updated basis distribution to model the target appearance. Wang et al. [49] proposed an online algorithm based on local sparse representation, where the local image patches of a target are represented by their sparse codes with an over-complete dictionary, and a classifier is learned to discriminate the target from the background.

2.3 Exclusive group LASSO

Least absolute shrinkage and selection operator (LASSO) is a shrinkage and selection method for linear regression. It minimizes the squared errors with a bound on the sum of the absolute values of the coefficients, which represents the sparse constraint in this paper. The exclusive group LASSO was proposed in [29,58] and has been successfully applied in feature selection [58], multi-label image classification [10] and image annotation [11]. Recently, Zhang et al. [55] incorporated the exclusive group method [58] into a sparse

representation-based visual tracking framework, which is different from the regression setting of ours. In [58], the elements of a sparse representation dictionary are divided into three groups: the templates, trivial templates and surrounding contexts. In contrast, our method considers the tracked samples with similar appearances as a group and the original exclusive group LASSO algorithm is not suitable for our purpose (see Sect. 3.1). Specifically, we design a coefficient constrained exclusive group LASSO algorithm, which aims to automatically re-weight the importance of multiple appearances and distinguishes the targets from their surroundings.

3 The proposed approach

In this section, we introduce the proposed method, which includes four parts: the coefficients constrained exclusive group LASSO model (**CCEGroupLASSO**), the algorithm of **CCEGroupLASSO**, the dictionary construction, and the overall tracking algorithm. Specially, we propose the coefficients constrained exclusive group LASSO approach in Sect. 3.1 and present the algorithm in Sect. 3.2. Then, we introduce the details of dictionary construction in Sect. 3.3 and give the overall tracking algorithm which is summarized in Sect. 3.4.

3.1 Coefficients constrained exclusive group LASSO

Given a dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_n] \in \mathbb{R}^{d\times n}$ (d is the dimension of feature vector and n is the size of dictionary), the template $\mathbf{h} \in \mathbb{R}^d$ can be sparsely represented by the dictionary $\mathbf{D}$ as follows:

$$\mathbf{h} \propto \mathbf{D}\boldsymbol{\alpha}, \qquad s.t. \quad \|\boldsymbol{\alpha}\|_0 \le t. \tag{5}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n]^T \in \mathbb{R}^n$, t is the sparsity level. Our goal is to select some representative samples from each group in the dictionary $\mathbf{D}$ to construct the discriminative template $\mathbf{h}$. Therefore, we propose a novel coefficient constrained *exclusive group LASSO*, which is defined below:

$$\begin{aligned} &\min_{\boldsymbol{\alpha}\in\mathbb{R}^n} \{J(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha}) + \lambda\|\boldsymbol{\alpha}\|_{\mathscr{G}}\} \\ &s.t. \quad \delta(\mathbf{d}_i)_i \ge 0, \ \forall i = 1, 2, \ldots, n. \end{aligned} \tag{6}$$

where

$$\delta(\mathbf{d}_i) = \begin{cases} 1 & \mathbf{d}_i \text{ is a positive sample,} \\ -1 & \mathbf{d}_i \text{ is a negative sample.} \end{cases} \tag{7}$$

In Eq. (6), $f(\boldsymbol{\alpha})$ is the loss term, $\|\boldsymbol{\alpha}\|_{\mathscr{G}}$ denotes the exclusive group LASSO penalty term, and $\delta(\mathbf{d}_i)_i \ge 0$ is the coefficient constrained term.

(1) **Loss term** $f(\boldsymbol{\alpha})$: As suggestion in [19], let the training samples be $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)\}$, where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \in \mathbb{R}^{d\times N}$ denotes the training samples and $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T \in \mathbb{R}^N$ denotes desired outputs obtained by calculating the overlap ratios of bounding boxes between the corresponding samples and the tracked result. For each sample, we use the cosine distance to measure the similarities between the template $\mathbf{h}$ and the samples[1]. The loss term $f(\boldsymbol{\alpha})$ includes two parts: 1) the losses on training samples $\mathbf{x}_i \in \mathbf{X}$; 2) the losses on dictionary atoms $\mathbf{d}_i \in \mathbf{D}$:

$$\begin{aligned} f(\boldsymbol{\alpha}) &= \sum_{i=1}^{N}(y_i - \mathbf{x}_i^T\mathbf{h})^2 + \sum_{i=1}^{n}(y_0 - \mathbf{d}_i^T\mathbf{h})^2 \\ &= \sum_{i=1}^{N}(y_i - \mathbf{x}_i^T\mathbf{D}\boldsymbol{\alpha})^2 + \sum_{i=1}^{n}(y_0 - \mathbf{d}_i^T\mathbf{D}\boldsymbol{\alpha})^2, \end{aligned} \tag{8}$$

where $y_0 = 1$ is the desired output for positive samples. The second term in Eq. (8) encourages the template $\mathbf{h}$ to be more discriminative for the target and the background. Meanwhile, the coefficients vector $\boldsymbol{\alpha}$ re-weights the importance of multiple appearances, which relieves the overfitting problem. Using matrix notation, equation (8) can be simplified as:

$$f(\boldsymbol{\alpha}) = \|\mathbf{y}_e - \mathbf{X}_e^T\mathbf{D}\boldsymbol{\alpha}\|_2^2, \tag{9}$$

where $\mathbf{y}_e = [\mathbf{y}; \mathbf{y}_0]$, $\mathbf{y}_0 = [1, 1, \cdots, 1]^T \in \mathbb{R}^n$, $\mathbf{X}_e = [\mathbf{X}, \mathbf{D}] \in \mathbb{R}^{d\times(N+n)}$.

(2) **Exclusive group LASSO penalty term** $\|\boldsymbol{\alpha}\|_{\mathscr{G}}$: In the dictionary $\mathbf{D}$, we divide it into different groups according to the score of the tracked target, as shown in the dictionary part of Fig. 2. More details are introduced in Sect. 3.3. Let $\mathscr{G}$ represents group set which consists of all the groups in the dictionary $\mathbf{D}$. The exclusive group LASSO penalty w.r.t $\boldsymbol{\alpha}$ is defined as:

$$\|\boldsymbol{\alpha}\|_{\mathscr{G}} = \sum_{g\in\mathscr{G}} \|\boldsymbol{\alpha}_g\|_1^2, \tag{10}$$

where g denotes the certain group in $\mathscr{G}$. Such penalty favors $\boldsymbol{\alpha}$ to be zero in each of the groups [29], as shown in exclusive group LASSO part of Fig. 2.

(3) **Coefficient constrained term**: As we know, the exclusive group LASSO does not differentiate the appearance groups and the negative group (see the middle column in Fig. 3), because there is no prior knowledge in the exclusive group LASSO to distinguish between appearance groups and the negative group. When the key samples increase, the exclusive group LASSO would lead to orderless distributions for entries in $\boldsymbol{\alpha}$, which is detrimental to the

[1] In this paper, we assume all the feature vectors are normalized, which means $\frac{<\mathbf{x}_1,\mathbf{x}_2>}{\|\mathbf{x}_1\|_2\|\mathbf{x}_2\|_2} =< \mathbf{x}_1, \mathbf{x}_2 >= \mathbf{x}_1^T\mathbf{x}_2$.

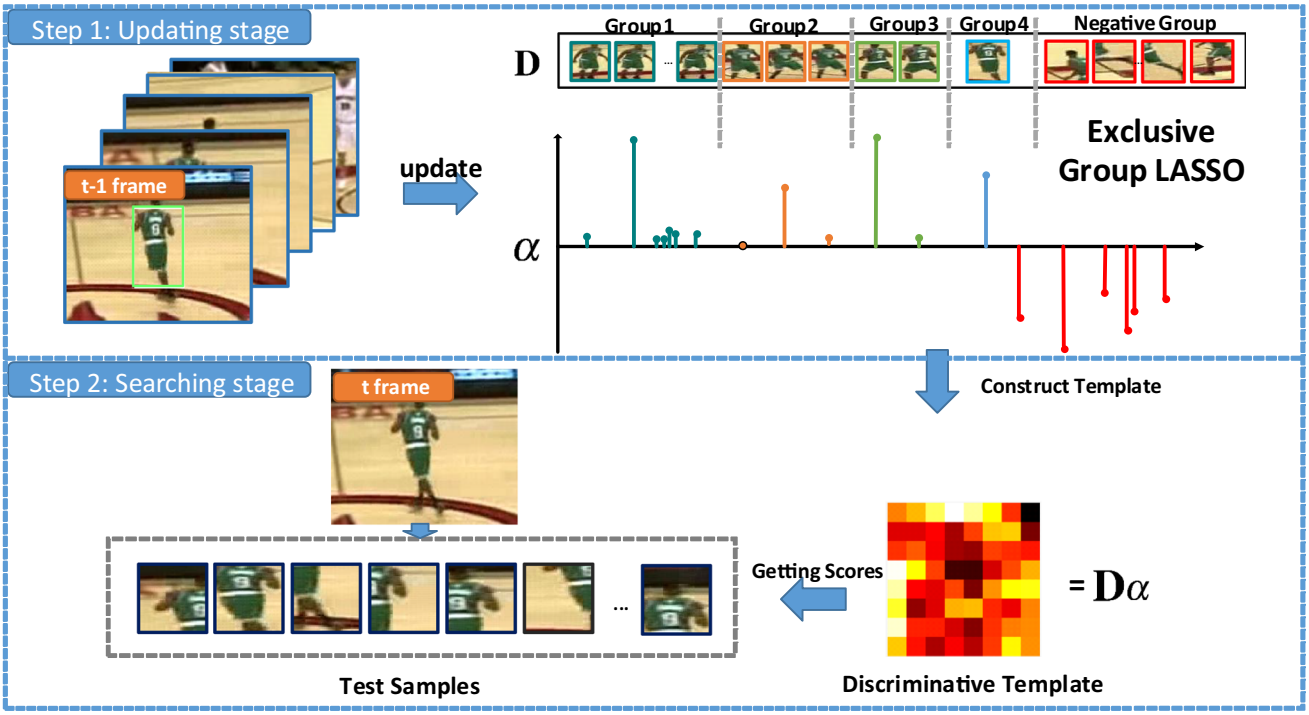


Fig. 2 The framework of our approach. During the tracking process, we maintain a set of key samples, called dictionary $\mathbf{D}$ and its corresponding coefficients vector $\boldsymbol{\alpha}$. The atoms in $\mathbf{D}$ are divided into different groups according to the scores. For the appearances group, their corresponding coefficients are positive and the coefficients of the negative group are negative. Furthermore, the entries of intra-groups are encouraged to be sparse. In the updating stage, we design the coefficients constrained exclusive group LASSO to solve $\boldsymbol{\alpha}$. Utilizing the group LASSO, we can obtain a compact discriminative template set, which are adopted to find the optimal tracked results

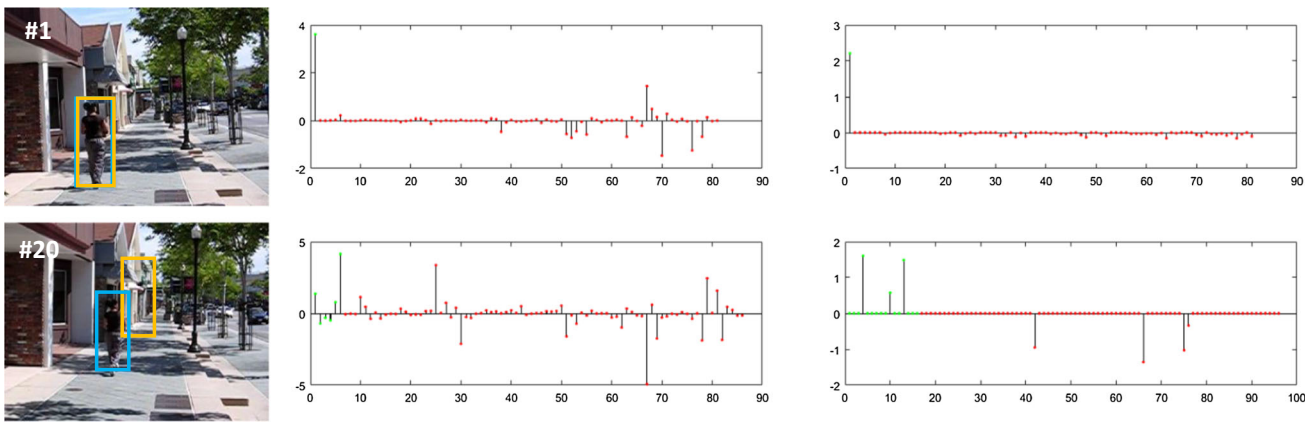


Fig. 3 The comparisons of coefficients vector $\boldsymbol{\alpha}$ between the exclusive group LASSO ([29]) and our method. The left column consists of the 1st and 20th frame of the video sequence *Human9* [53]. The yellow and blue rectangles represent the tracked results of exclusive group LASSO and ours, respectively. The middle and right columns show the entry distributions of $\boldsymbol{\alpha}$ for exclusive group LASSO and ours , where the first row and the second row represent the distribution of the 1st frame and the 20th frame, respectively. Even though the distributions are both sparse, it is clear to see that exclusive group LASSO leads to unordered distributions of $\boldsymbol{\alpha}$, which causes the drift in 20th frame

discriminability. Therefore, we impose the positive–negative groups constraints for exclusive group LASSO. This constraint encourages the coefficients $\boldsymbol{\alpha}$ to satisfy that the entries of appearances groups are positive, and the entries of negative groups are negative.

3.2 Algorithm for CCEGroup LASSO

Incorporating equation (9) and equation (10) into equation (6), we have the objective function:

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^n}\left\{J(\boldsymbol{\alpha})=\|\mathbf{y}_e-\mathbf{X}_e^T\mathbf{D}\boldsymbol{\alpha}\|_2^2+\lambda\sum_{g\in\mathscr{G}}\|\boldsymbol{\alpha}_g\|_1^2\right\} \quad (11)$$
$$s.t.\quad \delta(\mathbf{d}_i)\boldsymbol{\alpha}_i\geq 0,\ \forall i=1,2,\ldots,n.$$

To solve problem (11), we introduce an auxiliary matrix $\mathbf{F}$ for $\sum_{g\in\mathscr{G}}\|\boldsymbol{\alpha}_g\|_1^2$ as in [29]. Therefore, Eq. (11) can be rewritten as follows:

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^n}\left\{J(\boldsymbol{\alpha})=\|\mathbf{y}_e-\mathbf{X}_e^T\mathbf{D}\boldsymbol{\alpha}\|_2^2+\lambda\boldsymbol{\alpha}^T\mathbf{F}\boldsymbol{\alpha}\right\} \quad (12)$$
$$s.t.\quad \delta(\mathbf{d}_i)\boldsymbol{\alpha}_i\geq 0,\ \forall i=1,2,\ldots,n,$$

where $\mathbf{F}\in\mathbb{R}^{n\times n}$ is a diagonal matrix that encodes the exclusive group information. Its i-th diagonal element is given by:

$$\mathbf{F}_{ii}=\sum_{g\in\mathscr{G}}\frac{(\mathbf{I}_g)_i\|\boldsymbol{\alpha}_g\|_1}{|\boldsymbol{\alpha}_i|}, \quad (13)$$

where $\mathbf{I}_g\in\{0,1\}^n$ is the index indicator for group $g\in\mathscr{G}$. For example, suppose $(\mathbf{d}_1,\mathbf{d}_2)$ are elements of group g_0, then $\mathbf{I}_{g_0}=[1,1,0,\ldots,0]^T\in\mathbb{R}^n$. and $\boldsymbol{\alpha}_g=diag(\mathbf{I}_g)\boldsymbol{\alpha}$.

By introducing a label matrix $\mathbf{L}\in\{1,-1\}^{(N+n)\times n}$, equation (12) can be described as:

$$G(\boldsymbol{\alpha}_e)=\|\mathbf{y}_e-(\mathbf{X}_e^T\mathbf{D})\odot\mathbf{L}\boldsymbol{\alpha}_e\|_2^2+\lambda\boldsymbol{\alpha}_e^T\mathbf{F}\boldsymbol{\alpha}_e, \quad (14)$$
$$s.t.\quad \boldsymbol{\alpha}_e\geq 0$$

where $\odot$ is the element-wise product, $\boldsymbol{\alpha}_e=[|\alpha_1|,|\alpha_2|,\ldots,|\alpha_n|]$ is the absolute value of $\boldsymbol{\alpha}$, and $\mathbf{L}_{ij}=\delta(\mathbf{d}_i)$, $i=1,2,\cdots,N+n$, $j=1,2,\ldots,n$.

Let $\mathbf{K}=(\mathbf{X}_e^T\mathbf{D})\odot\mathbf{L}$, equation (14) can be simplified:

$$G(\boldsymbol{\alpha_e})=\|\mathbf{y}_e-\mathbf{K}\boldsymbol{\alpha}_e\|_2^2+\lambda\boldsymbol{\alpha}_e^T\mathbf{F}\boldsymbol{\alpha}_e. \quad (15)$$
$$s.t.\quad \boldsymbol{\alpha}_e\geq 0$$

Problem (15) can be solved by the nonnegative least square method [31,44]. The details are presented in algorithm 1, which is a standard nonnegative least square optimization procedure based on the gradient descent. The main idea is to calculate the negative derivative of the objective, e.g., line 2 of Algorithm 1, and then find a local minimum by iterative optimization.

3.3 Dictionary construction

The tracked results are considered as positive samples and are divided into different groups according to the scores, as shown in updating stage of Fig. 2. This is easy to implement since the similar appearances always emerge in successive frames. Given a group g_i contains a set of tracked results, we append the currently tracked result into group g_i when the score of the tracked result is greater than the given threshold θ. If this score is smaller than the threshold θ , we append the group g_i into the dictionary $\mathbf{D}$. Then, we construct an empty group g_{i+1} and also to do that. To promote the discriminability of $\mathbf{h}$, we add another negative group consisting of the surrounding negative samples emerged recently and drop the negative samples of old frames in the tracking process. Repeating the above steps, we can update the dictionary $\mathbf{D}$ continually.

Algorithm 1 Solve $G(\boldsymbol{\alpha}_e)$ in equation (15)

Inputs: $\mathbf{K}$, $\mathbf{F}$, $\mathbf{y}_e$
Output: $\boldsymbol{\alpha}_e$
1: Set positive set $\mathscr{P}=\{\}$, zero set $\mathscr{Z}=\{1,2,\ldots,n\}$, and $\boldsymbol{\alpha}_e=\mathbf{0}$;
2: Calculate the negative derivative of $G(\boldsymbol{\alpha}_e)$ w.r.t $\boldsymbol{\alpha}_e$: $\nabla G=\mathbf{w}=-(-\mathbf{K}^T(\mathbf{y}_e-\mathbf{K}\boldsymbol{\alpha}_e)+2\lambda\mathbf{F}\boldsymbol{\alpha}_e)$;
3: If $\mathscr{Z}$ is empty or $\mathbf{w}[\mathscr{Z}]\leq 0$, then go to line 11;
4: Find index p, where $\mathbf{w}[p]=\max(\mathbf{w}[\mathscr{Z}])$;
5: Move index p from $\mathscr{Z}$ to $\mathscr{P}$;
6: Let $\mathbf{z}\in\mathbb{R}^n$, init it as $\mathbf{z}=\mathbf{0}$, then let $\mathbf{z}[\mathscr{P}]=(\mathbf{K}[:,\mathscr{P}]^T\mathbf{K}[:,\mathscr{P}]+2\lambda\mathbf{F}[\mathscr{P},\mathscr{P}])^{-1}(\mathbf{K}[:,\mathscr{P}]^T\mathbf{y}_e)$;
7: If $\mathbf{z}[\mathscr{P}]>\mathbf{0}$, then let $\boldsymbol{\alpha}_e=\mathbf{z}$ and go to line 2;
8: Find $q\in\mathscr{P}$ satisfies $\frac{\boldsymbol{\alpha}_e[q]}{\boldsymbol{\alpha}_e[q]-\mathbf{z}[q]}=\min\{\frac{\boldsymbol{\alpha}_e[j]}{\boldsymbol{\alpha}_e[j]-\mathbf{z}[j]},\mathbf{z}[j]\leq 0,j\in\mathscr{P}\}$;
9: Set $\ell=\frac{\boldsymbol{\alpha}_e[q]}{\boldsymbol{\alpha}_e[q]-\mathbf{z}[q]}$ and $\boldsymbol{\alpha}_e=\boldsymbol{\alpha}_e+\ell(\mathbf{z}-\boldsymbol{\alpha}_e)$;
10: Move $j\in\{j|\boldsymbol{\alpha}_e[j]=0,j\in\mathscr{P}\}$ from $\mathscr{P}$ to $\mathscr{Z}$, go to line 6;
11: Output $\boldsymbol{\alpha}_e$.

3.4 Tracking algorithm

The overall tracking algorithm based on the above proposed exclusive group LASSO model is shown in algorithm 2, which includes searching stage and updating stage. In searching stage, we obtain the optimal candidate using the cosine distance between the target template $\mathbf{h}$ and the candidate samples. In updating stage, we update the dictionary $\mathbf{D}$ and the corresponding coefficient vector $\boldsymbol{\alpha}$ to construct the discriminative target template $\mathbf{h}$. The implement details are introduced in Sect. 4.1.

4 Experiments

We evaluate our approach on CVPR2013 benchmark datasets [53] that contains 50 videos sequences. The overall experimental results are illustrated by both precision plots and success plots (see Sect. 4.2). The compare trackers include the 29 visual trackers provided in the benchmark, three baseline trackers (logistic regression, ridge regression and SVM) provided in [48] and three state-of-the-art trackers: KCF [25], DSST [14] and TGPR [15].

4.1 Implementation details

During the tracking process, the cropped images are all resized to 32×32 pixels. In the searching stage, the number

Algorithm 2 Visual Tracking based on Coefficients Constrained Exclusive Group LASSO

Inputs:
N_s: the number of samples during the searching stage;
N: the number of samples during the updating stage;
ϵ: the threshold for updating $\boldsymbol{\alpha}$.
Output:
the estimated optimal candidates in every frames.
Init the dictionary $\mathbf{D}$, the desired outputs $\mathbf{y}$, the corresponding coefficients $\boldsymbol{\alpha}$ and labels;
Init the template $\mathbf{h} = \mathbf{D}\boldsymbol{\alpha}$;
while: the video sequence is not ended **do**
  Searching stage:
  Draw N_s samples $\mathbf{X}_s$ using the particle filtering;
  Calculate $\mathbf{g} = \mathbf{X}_s\mathbf{h}$ to find the optimal candidate samples.
  Updating stage:
  Generate N training samples $\mathbf{X}$ with the desired outputs $\mathbf{y}_o$, whose labels are denoted as $\mathbf{l}_\mathbf{X}$;
  Construct the inner-product matrix $\mathbf{K}$ with $\mathbf{D}$, $\mathbf{X}$;
  Construct label matrix $\mathbf{E}$ using the labels $\mathbf{l}_\mathbf{D}$ and $\mathbf{l}_\mathbf{X}$;
  Construct output vector $\mathbf{y}_e = [\mathbf{y}; \mathbf{y}_0]$;
  Solve $\boldsymbol{\alpha}$ with $\mathbf{K}$, $\mathbf{L}$, $\mathbf{y}_e$ by Algorithm 1.
  Select the key samples from $[\mathbf{X}, \mathbf{D}]$ in which the corresponding coefficients are satisfied $\boldsymbol{\alpha} \geq \epsilon$;
  Update the dictionary $\mathbf{D}$ and $\mathbf{l}_\mathbf{D}$ using the selected key samples.
  Construct new template $\mathbf{h} = \mathbf{D}\boldsymbol{\alpha}$;
end while:

of test samples is set to 750. We use a Gaussian distribution to generate the test sample set with different positions and scales. We use the HOG features [13], in which the cell size is set to 4 and the number of orientation bins is set to 9. Thus, we obtain a 2048 dimension feature vector for each image. In the updating stage, we obtain the training sample set on the polar grid, like Struck [19], where we set the radial to 5 and the angular divisions to 16. Therefore, we get 81 training samples. Note that, we set the tracked results as positive samples while all others as negatives. The maximal number of elements for each group is set to 20, and the maximal number of groups is set to 18. If the number of elements exceeds the maximum, we just discard the old ones. We delete the groups with the smallest number of elements when the number of groups exceeds maximum value. Our tracker is implemented in MATLAB R2015b on a PC with an Intel Core i7 4.0GHz CPU, 32GB RAM and runs about 5 frames per second. Note that, for all of the experiments, we use the fixed parameters.

4.2 Evaluation criteria

We use the precision and success rate for quantitative evaluation:

Precision plot the first evaluation metric is the CLE (center location error), which is defined as the average Euclidean distance between the center locations of the tracked targets and the manually labeled ground truth. The average center location error over all the frames of one sequence is used to evaluate the overall performance. We use the score at the 20 pixels CLE as the representative precision score for each tracker.

Success plot another evaluation metric is the bounding box overlap. We use the typical Pascal VOR (VOC Overlap Ratio) criterion. Given the bounding box B_R of the result and the bounding box B_G of the ground truth, the VOR can be computed as $\text{VOR} = |\frac{B_R \cap B_G}{B_R \cup B_G}|$, where $\cap$ and $\cup$ represent the intersection and union of two regions, respectively, and $|\cdot|$ denotes the number of pixels in the region. To measure the performance on a given sequence, we count the number of successful frames whose VOR is larger than the given threshold t_0. The success plot shows the ratios of successful frames at the thresholds varied from 0 to 1. We use the AUC (area under curve) of each success plot to rank the compared tracking approaches.

To evaluate the robustness of the visual tracking approaches, we use three different evaluations:

- OPE (One pass evaluation): this evaluation runs the compared trackers throughout a test sequence with an initialization ground truth position in the first frame.
- SRE (Spatial robustness evaluation): to evaluate whether a tracking method is sensitive to initialization errors, this evaluation generates the object states by slightly shifting or scaling the ground truth bounding box of a target object. In our experiments, we use eight spatial shifts (four center shifts and four corner shifts), and four scale variations, according to the benchmark [53]. We run the SRE evaluation 12 times, and the final scores of SRE are averaged to rank the compared tracking approaches.
- TRE (Temporal robustness evaluation): in this evaluation, each compared tracking approach is evaluated numerous times from different starting frames across an image sequence. In each test, an algorithm is evaluated from a particular starting frame, with the initialization of the corresponding ground truth object state, until the end of an image sequence. In our experiments, we run TRE 20 times and use the averaged results to generate the TRE scores.

The CVPR2013 benchmark dataset provides 50 video sequences. To evaluate the trackers robustness, these video sequences are categorized with 11 challenging attributes: fast motion (FM), background clutter (BC), motion blur (MB), deformation (DEF), illumination (IV), in-plane rotation (IPR), low resolution (LR), occlusion (OCC), out-of-view (OV), and out-of-plane rotation (OPR) and scale variations (SV). Each sequence includes several attributes. Among the 50 sequences, there are 39 sequences with OPR attributes, 31 sequences with IPR attributes, 29 sequences with OCC attributes, 28 sequences with SV attributes, 25 sequences with IV attributes, 21 sequences with BC attributes, 19 sequences with DEF attributes, 17 sequences with FM

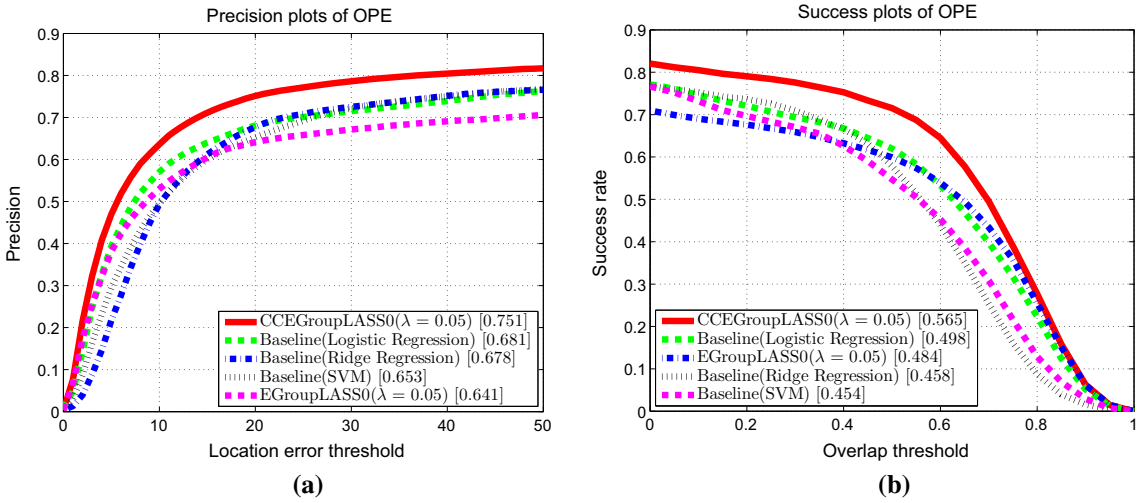


Fig. 4 Distance precision and overlap success plots over 50 sequences on the CVPR2013 benchmark datasets [53] under OPE for our method and the baseline methods

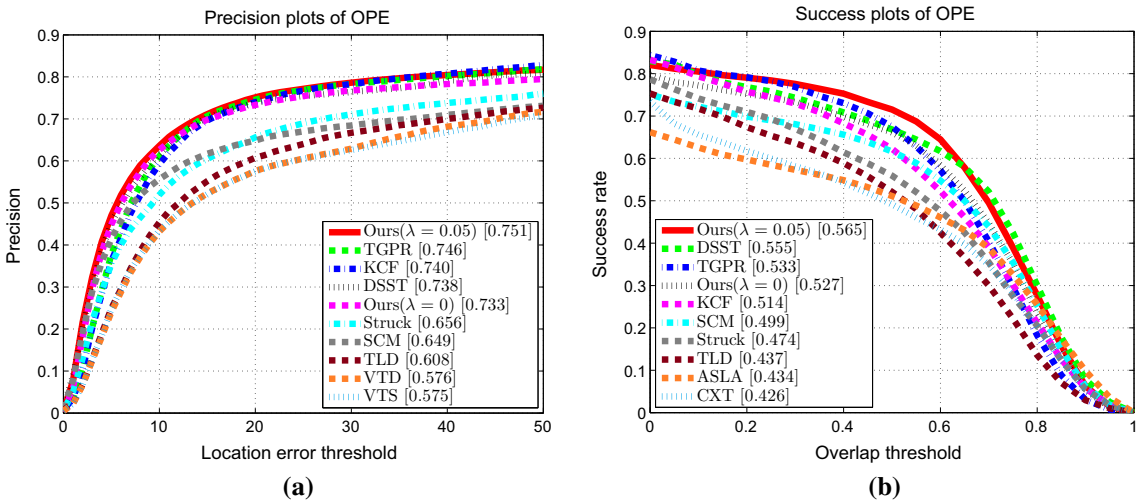


Fig. 5 Distance precision and overlap success plots over standard 50 benchmark sequences [53] using one pass evaluation (OPE). The legend contains the CLE and AUC scores for the top-10 trackers

Table 1 The average comparisons of our approach with the 6 trackers on the 50 sequences of CVPR2013 benchmark datasets in CLE at a threshold of 20 pixels and VOR at a threshold of 0.5

	TGPR [15]	Struck [19]	KCF [25]	DSST [14]	ASLA [27]	SCM [56]	Ours ($\lambda = 0.05$)
CLE	**0.746**	0.656	0.740	0.738	0.532	0.649	*0.751*
VOR	**0.675**	0.559	0.623	0.668	0.511	0.616	*0.715*

The best result of each sequence is highlighted by italics, and the second best is highlighted by bold

attributes, 12 sequences with MB attributes, 6 sequences with OV attributes and 4 with LR attributes.

4.3 Experimental results

We compare our method with three discriminative methods: the logistic regression, ridge regression and SVM. In [48], Wang et al. provide a comprehensive analysis which shows the three methods gain fine experimental results by using a robust feature. Therefore, we use the three methods as the baseline to evaluate our method. To demonstrate the effectiveness of the proposed method, we compare the coefficient constrained exclusive group LASSO(**CCEGroupLASSO**) to originally exclusive group Lasso(**EGroupLASSO**), logis-

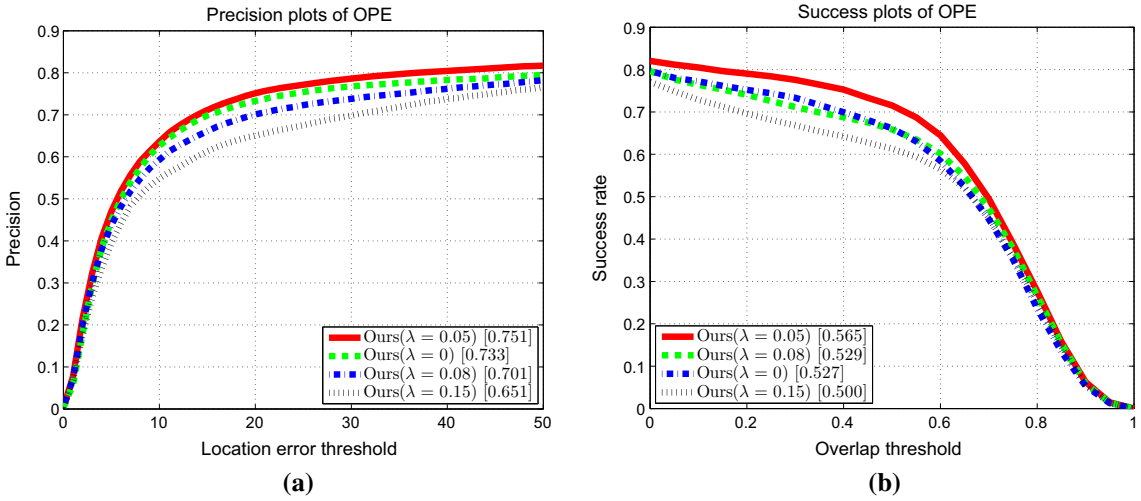


Fig. 6 Distance precision and overlap success plots over standard 50 benchmark sequences [53] using one pass evaluation (OPE). The legend contains the CLE and AUC scores for the proposed tracker with different λ

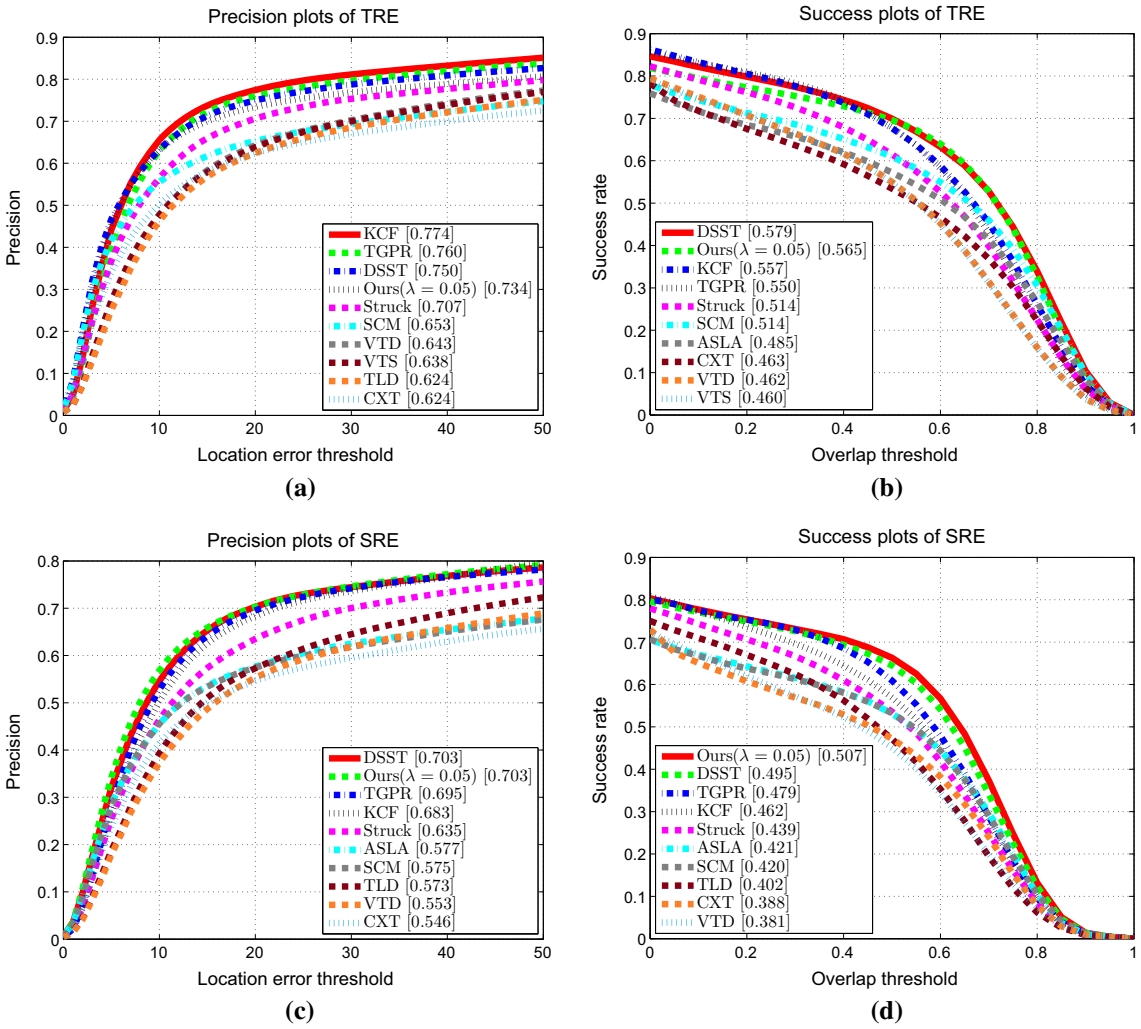
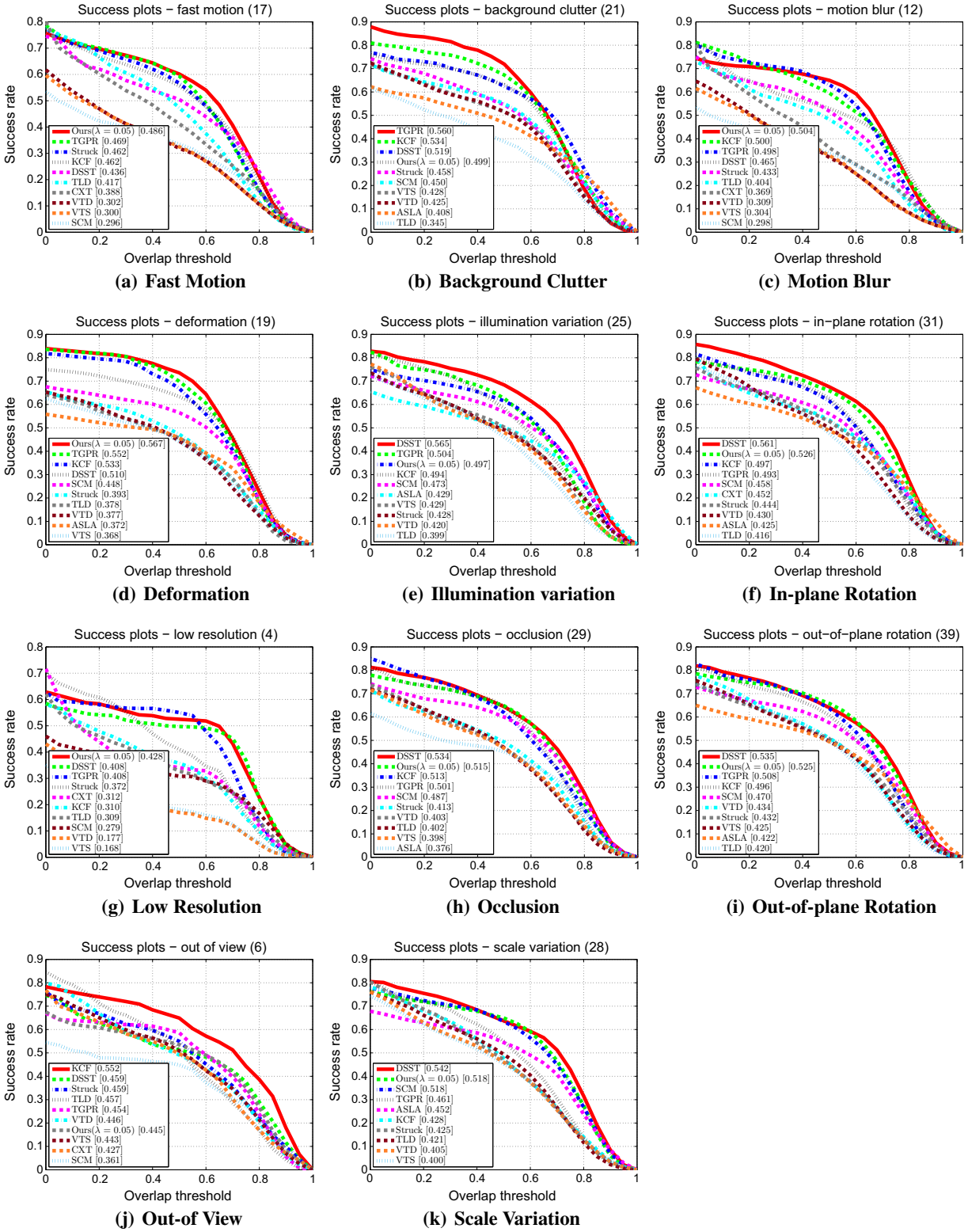


Fig. 7 Distance precision and overlap success plots over 50 sequences on the CVPR2013 benchmark datasets [53] under TRE and SRE



(a) Fast Motion **(b) Background Clutter** **(c) Motion Blur**

(d) Deformation **(e) Illumination variation** **(f) In-plane Rotation**

(g) Low Resolution **(h) Occlusion** **(i) Out-of-plane Rotation**

(j) Out-of View **(k) Scale Variation**

Fig. 8 Overlap success plots over the eleven challenges of fast motion, background clutter, blur, deformation, illumination, in-plane rotation, low resolution, occlusion, out-of-view and out-of-plane rotation and scale variations on the standard CVPR2013 benchmark datasets. The legend contains the AUC score for the top-10 trackers. Our method performs favorably against most other trackers. **a** Fast motion, **b** background clutter, **c** motion blur, **d** deformation, **e** illumination variation, **f** in-plane rotation, **g** low resolution, **h** occlusion, **i** out-of-plane rotation, **j** out-of view, **k** scale variation

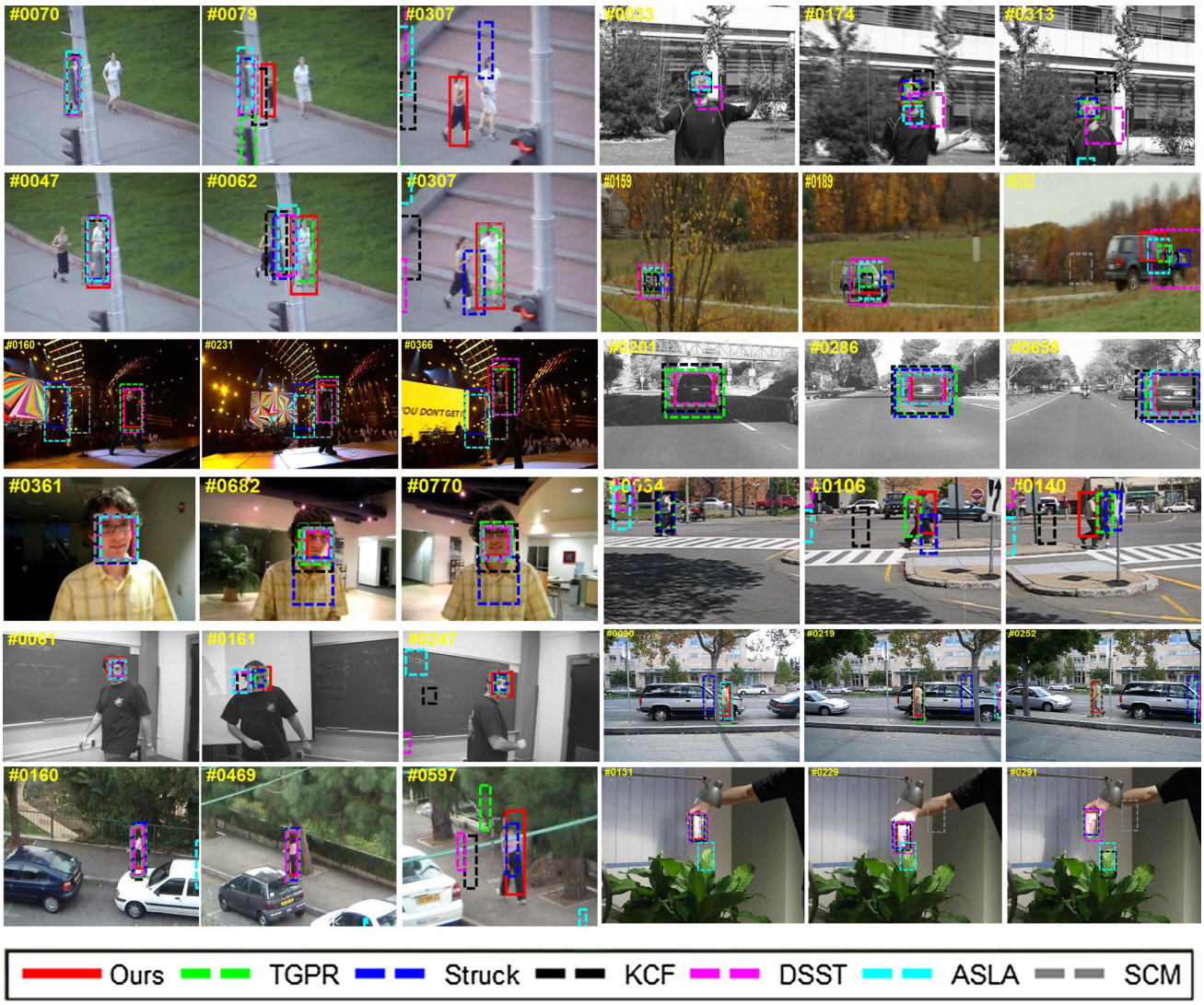


Fig. 9 Tracking results of 6 approaches (Struck [19], TGPR [15], KCF [25], DSST [14], ASLA [27] and SCM [56]) and our approach on 12 challenging sequences (from left to right and top down are Jogging-1, Jogging-2, Singler2, David, Freeman1 , Woman, Jumping, CarScale, Car4, Couple, David3, and Coke)

tic regression, ridge regression and SVM. For **CCEGroupLASSO** and **EGroupLASSO**, we use the same parameter $\lambda = 0.05$. The precision plot and success plot for the 5 compared methods are shown in Fig. 4. It is distinct to see that **CCEGroupLASSO** surpasses **EGroupLASSO** by a large margin and it is also better than the other three baselines. We suggest that is mainly because **CCEGroupLASSO** has a more compact and discriminative template which can distinguish the target from the background more accurate.

We also compare our method with the 29 classic visual trackers [53] and three state-of-the-art trackers (KCF, DSST and TGPR). To demonstrate the effectiveness of exclusive group LASSO penalty in Eq. (12), we evaluate **CCEGroupLASSO** using two different parameters: $\lambda = 0.05$ and $\lambda = 0$ under the OPE. The precision and success plots under OPE are shown in Fig. 5, and the quantitative comparisons of 6 well-known methods are provided in Table 1. In order to find a optimal λ, we test several different values on the standard 50 benchmark sequences [53] using one pass evaluation (OPE), as shown in Fig. 6. We can see that $\lambda = 0.05$ is the optimal value. We give the SRE and TRE results in Fig. 7. Among the compared visual trackers, KCF, DSST, TGPR and Struck gain better results than other visual trackers. Our approach gains nearly the same performances with DSST in SRE. However, our approach does not gain obvious promotions in TRE. This might be the robustness of our approach relies on the accumulation process of the bags. In TRE, the sequences are divided into several sub-sequences, which is adverse for our approach to collect sufficient groups to construct the template.

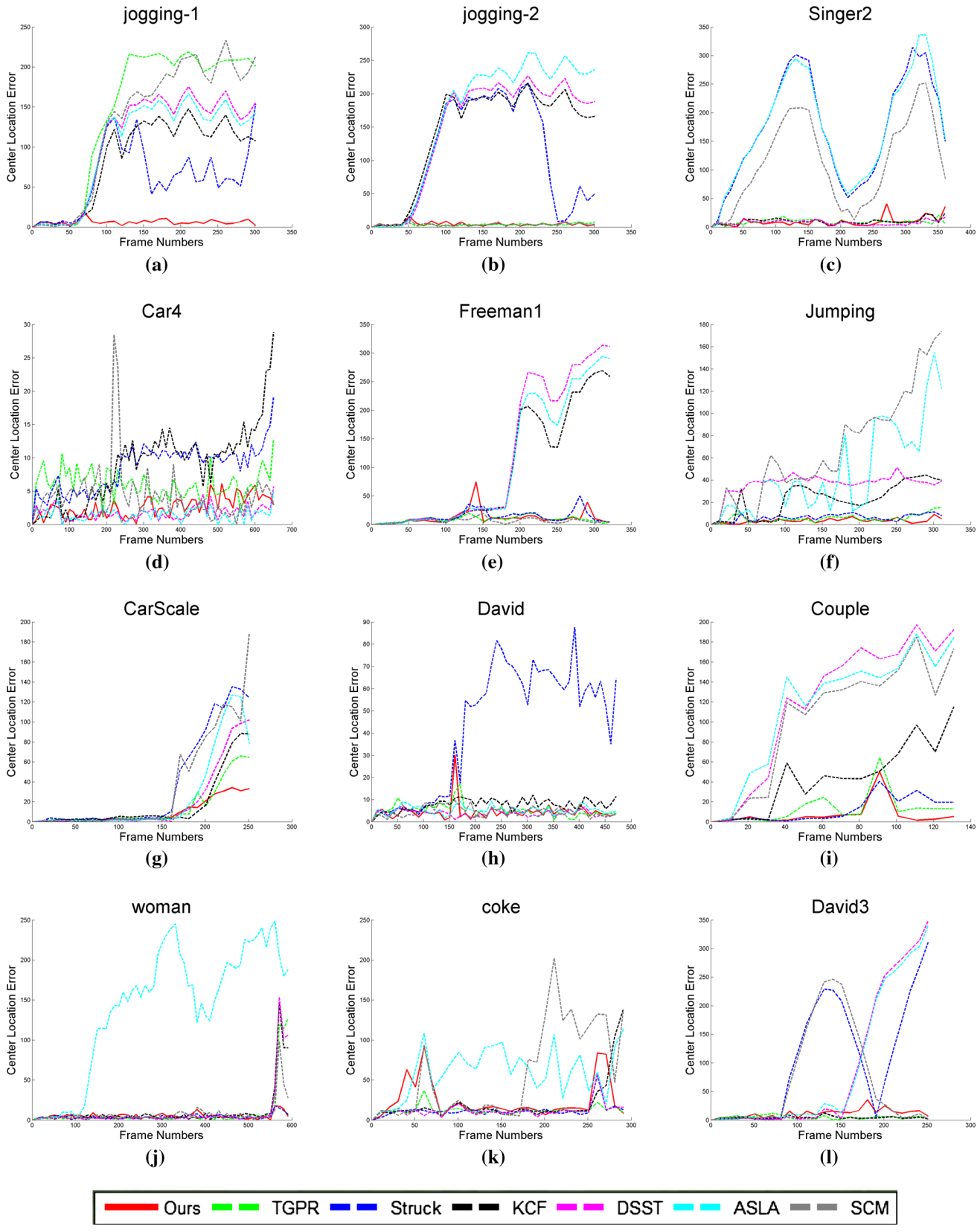


Fig. 10 Frame-by-frame comparison of center location errors on the 12 challenging sequences

In Fig. 8, we give the detailed success plots for each of the 11 challenging attributes. It is obvious to see that our method has achieved the best performance in fast motion, motion blur, deformation and low resolution. However, our tracker cannot perform well in the background clutter, as shown in Fig. 8b, mainly because our method considers the adjacent frame's target as the same, and this causes the tracker more likely to drift to the distractor. The tracking results of 12 sequences are provided in Fig. 9 and the corresponding frame-by-frame comparisons are provided in Fig. 10. From Fig. 9, we can see that our approach tracks the object target accurately, while the most compared trackers failed when the appearances has changed largely, just like the Jogging-1, Jogging-2 and Freeman1 of Fig. 9. As shown in Fig. 10, we also can see that the proposed method achieved the smallest center location error on most challenging sequences. It demonstrated that our approach outperforms the other compared trackers remarkably.

5 Conclusions

In this paper, we provide a novel method to learn a discriminative template by solving a coefficient constrained exclusive group LASSO. In particular, during the tracking process, we preserve the tracked results and surroundings as positive samples and negative samples, respectively. The positive samples are divided into different groups according to the appearances. All negative samples are put into a single negative group. We use those groups to build a dictionary. Then, the template is linearly represented by the dictionary with a coefficients constrained exclusive group LASSO model, which automatically selects a few representative elements from each of the groups. The experimental results show that the proposed method surpasses the original exclusive group LASSO and achieves a promising performance on the CVPR2013 benchmark datasets.

Acknowledgements This work is supported by the National Natural Science Foundation of China (Nos. 61762021, 61402122, 61672183, 61272252, 61401228, 61461008), Science and Technology Planning Project of Guanddong Province (No. 2016B090918047), Shenzhen Research Council (No.JCYJ20160406161948211, JCYJ2016022 6201453085, JSGG20150331152017052), Natural Science Foundation of Guangdong Province (No. 2015A030313544), the 2014 Ph.D. Recruitment Program of Guizhou Normal University, Natural Science Foundation of Guizhou Province (No. 2017[1130]), the Outstanding Innovation Talents of Science and Technology Award Scheme of Education Department in Guizhou Province (Qian jiao KY word[2015]487), the China Scholarship Council (No. 201508525007), Fund of Guizhou Educational Department (KY[2016]027), China Postdoctoral Science Foundation (Grant No. 2015M581841) and Postdoctoral Science Foundation of Jiangsu Province (Grant No. 1501019A).

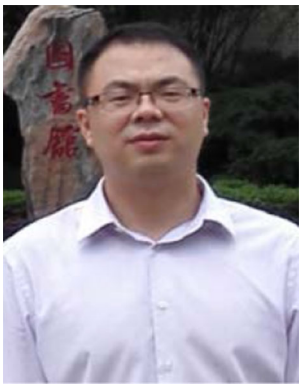
References

1. Adam, A., Rivlin, E., Shimshoni, I.: Robust fragments-based tracking using the integral histogram. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 798–805 (2006)
2. Ahmed, J., Jafri, M.N., Shah, M., Akbar, M.: Real-time edge-enhanced dynamic correlation and predictive open-loop car-following control for robust tracking. Mach. Vis. Appl. **19**(1), 1–25 (2008)
3. Avidan, S.: Support vector tracking. IEEE Trans. Pattern Anal. Mach. Intell. **26**(8), 1064–1072 (2004)
4. Baker, S., Matthews, I.: Lucas-kanade 20 years on: a unifying framework. Int. J. Comput. Vis. **56**(3), 221–255 (2004)
5. Bao, C., Wu, Y., Ling, H., Ji, H.: Real time robust l1 tracker using accelerated proximal gradient approach. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1830–1837 (2012)
6. Black, M.J., Jepson, A.D.: Eigentracking: robust matching and tracking of articulated objects using a view-based representation. Int. J. Comput. Vis. **26**(1), 63–84 (1998)
7. Chen, L., Philip Chen, C.L., Lu, M.: A multiple-kernel fuzzy c-means algorithm for image segmentation. IEEE Trans. Syst. Man Cybern. **41**(5), 1263–1274 (2011)
8. Chen, W., Zhao, Y.: Supervised kernel nonnegative matrix factorization for face recognition. Neurocomputing **205**, 165–181 (2016)
9. Chen, W.-S., Dai, X., Pan, B., Tang, Y.Y.: Semi-supervised discriminant analysis method for face recognition. Int. J. Wavelets Multiresolut. Inf. Process. **13**(06), 1550049 (2015)
10. Chen, X., Yuan, X.-T., Chen, Q., Yan, S., Chua, T.-S.: Multi-label visual classification with label exclusive context. In: IEEE International Conference on Computer Vision (ICCV), pp. 834–841 (2011)
11. Chen, X., Yuan, X., Yan, S., Tang, J., Rui, Y., Chua, T.-S.: Towards multi-semantic image annotation with graph regularized exclusive group lasso. In: International Conference on Multimedia, pp. 263–272 (2011)
12. Comaniciu, D., Ramesh, V., Meer, P.: Real-time tracking of non-rigid objects using mean shift. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 142–149 (2000)
13. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 886–893 (2005)
14. Danelljan, M., Häger, G., Khan, F., Felsberg, M.: Accurate scale estimation for robust visual tracking. In: British Machine Vision Conference (BMVC), pp. 65.1–65.11 (2014)
15. Gao, J., Ling, H., Hu, W., Xing, J.: Transfer learning based visual tracking with Gaussian processes regression. In: European Conference on Computer Vision (ECCV), pp. 188–203 (2014)
16. Ge, Q., Jing, X.-Y., Wu, F., Wei, Z., et al.: Structure-based low-rank model with graph nuclear norm regularization for noise removal. IEEE Trans. Image Process. **26**, 3098–3112 (2016)
17. Gu, B., Sheng, V.S.: A robust regularization path algorithm for ν-support vector classification. IEEE Trans. Neural Netw. Learn. Syst. **28**, 1241–1248 (2017)
18. Han, Z., Jiao, J., Zhang, B., Ye, Q., Liu, J.: Visual object tracking via sample-based adaptive sparse representation (AdaSR). Pattern Recognit. **44**(9), 2170–2183 (2011)
19. Hare, S., Golodetz, S., Saffari, A., et al.: Struck: structured output tracking with kernels. IEEE Trans. Pattern Anal. Mach. Intell. **38**(10), 2096–2109 (2016)
20. He, Z., Li, X., You, X., Tao, D., Tang, Y.Y.: Connected component model for multi-object tracking. IEEE Trans. Image Process. **25**(8), 3698–3711 (2016)
21. He, Z., Chung, A.C.: 3-D B-spline wavelet-based local standard deviation (bwlsd): its application to edge detection and vascular

segmentation in magnetic resonance angiography. Int. J. Comput. Vis. **87**(3), 235–265 (2010)
22. He, Z., Yi, S., Cheung, Y.-M., You, X., Tang, Y.Y.: Robust object tracking via key patch sparse representation. IEEE Trans. Cybern. **47**, 1–11 (2016)
23. He, Z., You, X., Tang, Y.Y.: Writer identification of Chinese handwriting documents using hidden Markov tree model. Pattern Recognit. **41**(4), 1295–1307 (2008)
24. He, Z., You, X., Zhou, L., Cheung, Y., Jianwei, D.: Writer identification using fractal dimension of wavelet subbands in gabor domain. Integr. Comput. Aided Eng. **17**(17), 157–165 (2010)
25. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 583–596 (2015)
26. Isard, M., Blake, A.: Condensation-conditional density propagation for visual tracking. Int. J. Comput. Vis. **29**(1), 5–28 (1998)
27. Jia, X., Lu, H., Yang, M.-H.: Visual tracking via adaptive structural local sparse appearance model. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1822–1829 (2012)
28. Jing, X., Wu, F.: Multi-spectral low-rank structured dictionary learning for face recognition. Pattern Recognit. **59**(4), 14–25 (2016)
29. Kong, D., Fujimaki, R., Liu, J., Nie, F., Ding, C.: Exclusive feature learning on arbitrary structures via $\ell_{1,2}$-norm. In: Advances in Neural Information Processing Systems, pp. 1655–1663 (2014)
30. Lai, Z., Yong, X., Jin, Z., Zhang, D.: Human gait recognition via sparse discriminant projection learning. IEEE Trans. Circuits Syst. Video Technol. **24**(10), 1651–1662 (2014)
31. Lawson, C.L., Hanson, R.J.: Solving Least Squares Problems, vol. 161. SIAM, Philadelphia (1974)
32. Li, X., Shen, C., Dick, A., van den Hengel, A.: Learning compact binary codes for visual tracking. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2419–2426 (2013)
33. Li, X., Shen, C., Shi, Q., Dick, A., Van den Hengel, A.: Non-sparse linear representations for visual tracking with online reservoir metric learning. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1760–1767 (2012)
34. Li, X., Liu, Q., He, Z., Wang, H., Zhang, C., Chen, W.-S.: A multi-view model for visual tracking via correlation filters. Knowl. Based Syst. **113**, 88–99 (2016)
35. Liu, B., Huang, J., Yang, L., Kulikowsk, C.: Robust tracking using local sparse appearance model and k-selection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1313–1320 (2011)
36. Liu, B., Yang, L., Huang, J., Meer, P., Gong, L., Kulikowski, C.: Robust and fast collaborative tracking with two stage sparse optimization. In: European Conference on Computer Vision, pp. 624–637. Springer (2010)
37. Liu, L., Chen, L.: Weighted joint sparse representation for removing mixed noise in image. IEEE Trans. Cybern. **47**, 600–611 (2016)
38. Liu, Q., Ma, X., Ou, W., Zhou, Q.: Visual object tracking with online sample selection via lasso regularization. Signal Image Video Process. **11**, 881–888 (2017)
39. Liu, R., Tang, Y.: Topological coding and its application in the refinement of sift. IEEE Trans. Cybern. **44**(11), 2155–2166 (2014)
40. Lu, H., Li, B., Zhu, J., et al.: Wound intensity correction and segmentation with convolutional neural networks. Concurr. Comput. Pract. Exp. **29**, 6 (2016)
41. Matthews, I., Ishikawa, T., Baker, S.: The template update problem. IEEE Trans. Pattern Anal. Mach. Intell. **26**(6), 810–815 (2004)
42. Mei, X., Ling, H.: Robust visual tracking using ℓ_1 minimization. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1436–1443 (2009)
43. Mei, X., Ling, H., Wu, Y., Blasch, E., Bai, L.: Minimum error bounded efficient ℓ_1 tracker with occlusion detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1257–1264 (2011)
44. Ou,W., Yuan, D., Liu, Q., Cao, Y.: Object tracking based on online representative sample selection via non-negative least square. Multimed. Tools Appl. (2017). https://doi.org/10.1007/s11042-017-4672-3
45. Qian, J., Fang, B., Yang, W., Luan, X.: Accurate tilt sensing with linear model. IEEE Sens. J. **11**(10), 2301–2309 (2011)
46. Ross, D.A., Lim, J., Lin, R.-S., Yang, M.-H.: Incremental learning for robust visual tracking. Int. J. Comput. Vis. **77**(1–3), 125–141 (2008)
47. Schweitzer, H., Bell, J.W., Wu, F.: Very fast template matching. In: European Conference on Computer Vision (ECCV), pp. 358–372. Springer (2002)
48. Wang, N., Shi, J., Yeung, D.-Y., Jia, J.: Understanding and diagnosing visual tracking systems. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3101–3109 (2015)
49. Wang, Q., Chen, F., Xu, W., Yang, M.-H.: Online discriminative object tracking with local sparse representation. In: IEEE Workshop on Applications of Computer Vision (WACV), pp. 425–432 (2012)
50. Weihua, O., You, X., Tao, D., Zhang, P., Tang, Y., Zhu, Z.: Robust face recognition via occlusion dictionary learning. Pattern Recognit. **47**(4), 1559–1572 (2014)
51. Weihua, O., Shujian, Y., Li, G., Jian, L., Zhang, K., Xie, G.: Multi-view non-negative matrix factorization by patch alignment framework with view consistency. Neurocomputing **204**, 116–124 (2016)
52. Wong, W.K., Lai, Z., Xu, Y., Wen, J., Ho, C.P.: Joint tensor feature analysis for visual object recognition. IEEE Trans. Cybern. **45**(11), 2425–2436 (2015)
53. Wu, Y., Lim, J., Yang, M.-H.: Online object tracking: a benchmark. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2411–2418 (2013)
54. Yang, M., Wu, Y., Pei, M., Ma, B., Jia, Y.: Coupling semi-supervised learning and example selection for online object tracking. In: Asian Conference on Computer Vision (ACCV), pp. 476–491 (2014)
55. Zhang, T., Ghanem, B., Liu, S., Changsheng, X., Ahuja, N.: Robust visual tracking via exclusive context modeling. IEEE Trans. Cybern. **46**(1), 51–63 (2016)
56. Zhong, W., Lu, H., Yang, M.-H.: Robust object tracking via sparsity-based collaborative model. In: IEEE Conference on Computer vision and pattern recognition (CVPR), pp. 1838–1845 (2012)
57. Zhou, Q., Zheng, B., Zhu, W., Latecki, L.J.: Multi-scale context for scene labeling via flexible segmentation graph. Pattern Recognit. **59**(C), 312–324 (2016)
58. Zhou, Y., Jin, R., Hoi, S.: Exclusive lasso for multi-task feature selection. In: International Conference on Artificial Intelligence and Statistics, pp. 988–995 (2010)

Xiao Ma graduated from Communication University of China, Beijing, China, in 2013. He is pursuing the master's degree in computer science with the Research Institute of Biocomputing, School of Computer Science, Harbin Institute of Technology Shenzhen Graduate School, China. His current research interests include object tracking and kernel methods.

Qiao Liu received the master's degree in computer science from the Guizhou Normal University, Guiyang, China, in 2016. He is pursuing the Ph.D. degree with the Research Institute of Biocomputing, School of Computer Science, Harbin Institute of Technology Shenzhen Graduate School, China. His current research interests include object tracking and object recognition.



Weihua Ou received the M.S. degree in Mathematics from the Southeast University, Nanjing, China, in 2006 and the Ph.D. degree in Information and Communication Engineering from Huazhong University of Science and Technology (HUST), China, in 2014. Currently, he is an Associate Professor at the School of Big data and Computer Science in Guizhou Normal University, Guiyang, China. His current research interests include sparse representation, deep learning, and cross-modal retrieval.

Quan Zhou received the B.S. degree in electronics and information engineering in 1998 from China University of Geosciences, Wuhan, China, and M.S. degree and Ph.D. degree in communication and information system in 2006 and 2013, respectively, from Huazhong University of Science and Technology, China. He is now an associated professor of Nanjing University of Posts and Telecommunications. His research interests include computer vision and pattern recognition. He has published more than 30 articles in top journals (e.g., IEEE TIP, EL, and sensors) and conference (ICIP, ICASSP, ACCV, and ICPR) in image processing and computer vision. He has been invited as guest editor of ACM/Springer Mobile Networks and Applications and Multimedia Tools and Applications. He now serves as TPC member of many international conferences and reviewer for a series of SCI journals, including IEEE TIP, IEEE TC, IEEE TSP, IEEE CSVT, PR, and Neurocomputing. He is a member of IEEE.