Learning Support Correlation Filters for Visual Tracking

Wangmeng Zuo, *Senior Member, IEEE*, Xiaohe Wu, Liang Lin, *Senior Member, IEEE*, Lei Zhang, *Fellow, IEEE*, and Ming-Hsuan Yang, *Senior Member, IEEE*

Abstract—For visual tracking methods based on kernel support vector machines (SVMs), data sampling is usually adopted to reduce the computational cost in training. In addition, budgeting of support vectors is required for computational efficiency. Instead of sampling and budgeting, recently the circulant matrix formed by dense sampling of translated image patches has been utilized in kernel correlation filters for fast tracking. In this paper, we derive an equivalent formulation of a SVM model with the circulant matrix expression and present an efficient alternating optimization method for visual tracking. We incorporate the discrete Fourier transform with the proposed alternating optimization process, and pose the tracking problem as an iterative learning of support correlation filters (SCFs). In the fully-supervision setting, our SCF can find the globally optimal solution with real-time performance. For a given circulant data matrix with n^2 samples of $n \times n$ pixels, the computational complexity of the proposed algorithm is $O(n^2 \log n)$ whereas that of the standard SVM-based approaches is at least $O(n^4)$. In addition, we extend the SCF-based tracking algorithm with multi-channel features, kernel functions, and scale-adaptive approaches to further improve the tracking performance. Experimental results on a large benchmark dataset show that the proposed SCF-based algorithms perform favorably against the state-of-the-art tracking methods in terms of accuracy and speed.

Index Terms—Visual tracking, correlation filters, support vector machine, max-margin learning

1 Introduction

Robust visual tracking is a challenging problem due to large changes of object appearance caused by pose, illumination, deformation, occlusion, distractors, as well as background clutter [38], [41]. Among the state-of-the-art methods, discriminative classifiers with model updating and sampling have been demonstrated to perform well in visual tracking. On the other hand, correlation filters [8], [12], [23], [24] have been shown to be efficient for locating objects using the circulant matrix and fast Fourier transform. Central to the advances in visual tracking are the development of effective appearance models and efficient sampling schemes [41], [42].

Discriminative appearance models have been extensively studied in visual tracking and have achieved the state-of-the-art results. One representative discriminative appearance model is based on support vector machines (SVMs) [2], [4], [21], [42]. To learn classifiers for detecting objects within local regions, SVM-based tracking approaches are developed based on two modules: a *sampler* to generate a set of positive and negative samples, and a *learner* to update the classifier using the training samples. To reduce the computational load, sampling is usually required in SVM-based trackers to select a small set of samples [21], [42]. As kernel SVM-based tracking methods are susceptible to the *curse of kernelization*, budgeting is introduced for online learning of the structural SVM tracker [21] to restrict the number of support vectors, or an explicit feature mapping function is used to approximate the intersection kernel [42]. While sampling and budgeting may improve tracking efficiency at the expense of accuracy, most SVM-based trackers [4], [21], [42] do not run in real-time.

Correlation filters (CFs) [8], [23], [24], [43] have recently been utilized for efficient visual tracking. Here a base sample is defined as the first row of a circulant matrix. Then the data matrix formed by dense sampling (i.e., cyclic shifts) of base sample should have circulant structures, which facilitates the use of the discrete Fourier transform (DFT) for efficient and effective visual tracking [8], [23], [24], [43]. Among the existing CF-based trackers, ridge regression or kernel ridge regression are generally adopted as the predictors. Henriques et al. [22] apply the circulant property for training of support vector regression (SVR) efficiently to detect pedestrians. However, this algorithm is developed to solve an approximation of the SVR model and cannot be used for SVM due to the discrete labels in the classification task. The problem on how to exploit the circulant property to accelerate SVM-based trackers remains unaddressed.

- *W. Zuo and X. Wu are with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China. E-mail: cswmzuo@gmail.com, angela612@126.com.*
- *L. Lin is with the School of Advanced Computing, Sun Yat-Sen University, Guangzhou, Guangdong 510006, China. E-mail: linliang@ieee.org.*
- *L. Zhang is with the Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. E-mail: cslzhang@comp.polyu.edu.hk.*
- *M.-H. Yang is with the School of Engineering, University of California, Merced, CA 95344. E-mail: mhyang@ucmerced.edu.*

Manuscript received 4 Oct. 2015; revised 13 July 2017; accepted 19 Mar. 2018. Date of publication 19 Apr. 2018; date of current version 10 Apr. 2019. (Corresponding author: Xiaohe Wu.)
Recommended for acceptance by I. Reid.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPAMI.2018.2829180

In this paper, we propose a novel SVM-based algorithm via support correlation filters (SCFs) for efficient and accurate visual tracking. Instead of sampling and budgeting, the proposed algorithm based on SCFs deals with the efficiency issue by using the data matrix formed by dense sampling. By exploiting the circulant property, we formulate the proposed SVM-based tracker as a learning problem for support correlation filters and propose an efficient algorithm. By incorporating the discrete Fourier transform in an alternating optimization process, the SVM classifier can be efficiently updated by iterative learning of correlation filters. For an $n \times n$ image, there are n^2 training sample images of the same size in the circulant data matrix and the computational complexity of the proposed algorithm is $O(n^2 \log n)$ whereas that of the standard SVM-based approaches is at least $O(n^4)$. Furthermore, we extend the proposed SCF-based algorithm to multi-channel SCF (MSCF), kernelized SCF (KSCF), and scale-adaptive KSCF (SKSCF) methods to improve the tracking performance.

We evaluate the proposed SCF-based algorithms on a large benchmark dataset with comparison to the state-of-the-art methods [41] and analyze the tracking results. First, with the discriminative strength of SVMs, the proposed KSCF method performs favorably against the existing regression-based correlation filter trackers. Second, by exploiting the circulant structure of training samples, the proposed KSCF algorithm performs well compared with the existing SVM-based trackers [21], [42] in terms of efficiency and accuracy. Third, the proposed KSCF and SKSCF algorithms outperform the state-of-the-art methods including the ensemble and scale-adaptive tracking methods [12], [31], [42].

2 Related Work and Problem Context

Visual tracking has long been an active research topic in computer vision which involves developments of both learning methods (e.g., feature learning and selection, online learning and ensemble models) and application domains (e.g., auto-navigation, visual surveillance and human-computer interactions). Performance evaluation on state-of-the-art tracking algorithms have been reported [38], [41], and we discuss the most relevant methods to this work in this section.

Appearance Models for Visual Tracking. Appearance models play an important role in visual tracking which can be broadly categorized as generative or discriminative. Generative appearance methods based on holistic templates, subspace representations and sparse representations have been developed for object representations [25], [40], [45]. Discriminative appearance methods are usually based on features and classifiers learned from a large set of examples. Visual tracking is posed as a task to distinguish the target objects from the backgrounds. Tracking methods based on discriminative appearance models have been shown to achieve the state-of-the-art results [41].

Discriminative tracking methods are usually based on object detection within local search using classifiers such as boosting methods, random forests, and SVMs [41]. Among these classifiers, boosting methods [3], [28] and random forests [36] are ensemble learning methods, where sampling from large sets of features is indispensable and that makes it difficult to adopt correlation filters in these approaches. In this work, we exploit the discriminative strength of SVMs and efficiency of correlation filters for visual tracking.

Label ambiguity has also been studied for visual tracking, e.g., semi-supervised [19] and multiple instance learning methods [3]. Considering that classification based methods are trained to predict the class label rather than the object location, Hare et al. [21] propose a tracker based on structured SVM. In this work, we alleviate the label ambiguity problem by using the assignment scheme in a way similar to that for object detection and tracking [18].

Correlation Filters for Tracking. A correlation filter uses a designed template to generate strong response to a region that is similar to the target object while suppressing responses to distractors. Correlation filters have been widely applied to numerous problems such as object detection [9] and object alignment [6]. A number of correlation filters have been proposed in the literature including the minimum output sum of squared error (MOSSE) [8] methods. Recently, the max-margin CF (MMCF) [35], multi-channel CF [12], [13], [14], [24], and kernelized CF [23], [24] methods have been developed for object detection and tracking. The MMCF [35] scheme combines the localization properties of correlation filters with good generalization performance of SVM. The multi-channel correlation filters [12], [13], [14], [24] are designed to use more effective features, e.g., histogram of oriented gradients (HOG). In addition, a method that combines MMCF and multi-channel CF is developed [7] for object detection and landmark localization. The kernel tricks are also employed to learn kernelized synthetic discriminant functions (SDF) [32] with correlation filters. We note that the MMCF [7], [35] methods do not exploit the circulant structure of the data matrix in the max-margin loss function.

In visual tracking, Bolme et al. [8] propose the MOSSE method to learn adaptive correlation filters with high efficiency and competitive performance. Subsequently, the kernelized correlation filter (KCF) [24] is developed by exploiting the circulant property of the kernel matrix. Extensions of CF and KCF with multi-channel features are introduced for visual tracking [12], [13], [14], [24]. We note existing CF-based trackers are developed with ridge regression schemes for locating the target. On the other hand, the SVM-based trackers, e.g., Struck [21] and MEEM [42], have been demonstrated to achieve the state-of-the-art performance. One straightforward extension is to integrate SVM-based trackers with the MMCF method [35]. Nevertheless, the MMCF scheme is computationally prohibitive for real-time applications, and the training data matrix for SVM is not circulant. In this work, we develop novel discriminative tracking algorithms based on SVMs and correlation filters that perform both efficiently and effectively.

3 Support Correlation Filtering

We first present the problem formulation and propose an alternating optimization algorithm to learn support correlation filters efficiently. We then develop the MSCF, KSCF and SKSCF methods to learn multi-channel, kernelized and scale-adaptive correlation filters respectively for robust visual tracking.

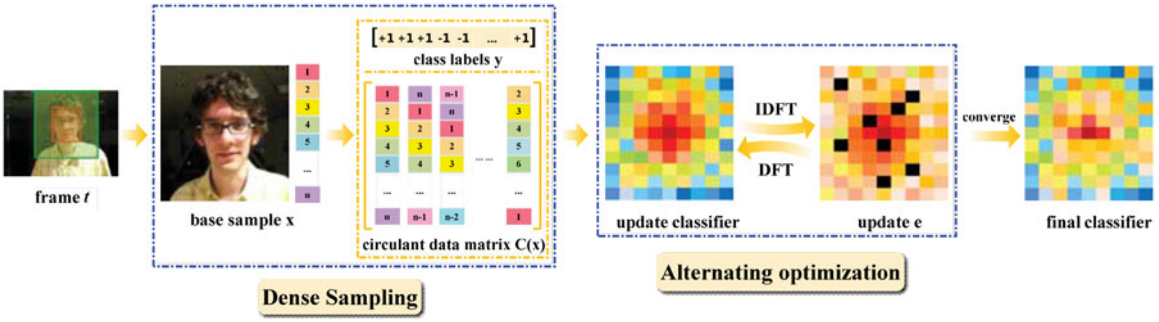


Fig. 1. Illustration of the proposed SCF learning algorithm. The proposed algorithm iterates between updating $\mathbf{e}$ and updating SVM classifier $\{\mathbf{w}, b\}$ until convergence. In each iteration, only one DFT and one IDFT are required, which make the proposed algorithm computationally efficient. The black blocks in $\mathbf{e}$ denote support vectors, and our algorithm can adaptively find and exploit difficult samples (i.e., support vectors) to learn support correlation filters.

3.1 Problem Formulation

Given an image $\mathbf{x}$, the full set of its translated versions forms a circulant matrix $\mathbf{X}$ with several interesting properties [20], where each row represents one possible observation of a target object (see Fig. 1). A circulant matrix consists of all possible cyclic translations of a target image, and tracking is formulated as determining the most likely row. In general, the eigenvectors of a circulant matrix $\mathbf{X}$ are the base vectors F of the discrete Fourier transform:

$$\mathbf{X} = F^H \mathrm{diag}(\hat{\mathbf{x}})F, \tag{1}$$

where F^H is the Hermitian transpose of F and $\hat{\mathbf{x}} = \mathcal{F}(\mathbf{x})$ denotes the Fourier transform of $\mathbf{x}$. In the following, we use $\mathrm{diag}(\cdot)$ to form a diagonal matrix from a vector.

Our goal is to learn a support correlation filter $\mathbf{w}$ and a bias b, to classify any translated image $\mathbf{x}_i$ by

$$y_i = \mathrm{sgn}\big(\mathbf{w}^\top \mathbf{x}_i + b\big). \tag{2}$$

Note that all the translated images $\mathbf{x}_i$ form a circulant matrix $\mathbf{X}$. We can classify all the samples in $\mathbf{X}$ by

$$\mathbf{y} = \mathrm{sgn}\big(\mathcal{F}^{-1}(\hat{\mathbf{x}}^* \circ \hat{\mathbf{w}}) + b\big), \tag{3}$$

where $\circ$ denotes the element-wise multiplication operator. $\mathcal{F}^{-1}(\cdot)$ denotes the inverse discrete Fourier transform (IDFT), and $\hat{\mathbf{x}}^*$ denotes the complex conjugate of $\hat{\mathbf{x}}$. Given the circulant matrix $\mathbf{X}$ generated by an $n \times n$ image $\mathbf{x}$, the computational complexity of classifying every $\mathbf{x}_i$ by (2) is $O(n^4)$, while that of classifying all samples of $\mathbf{X}$ by (3) is $O(n^2 \log n)$.

Given the training set of a circulant matrix $\mathbf{X} = [\mathbf{x}_1; \mathbf{x}_2; \ldots; \mathbf{x}_{n^2}]$ with the corresponding class labels $\mathbf{y} = [y_1, y_2, \ldots, y_{n^2}]^\top$, the SVM with the squared hinge loss can be defined as:

$$\begin{aligned} &\min_{\mathbf{w},b,\xi} \; \|\mathbf{w}\|^2 + C\sum_i \xi_i^2 \\ &\text{s.t.} \;\; y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \; \forall i, \end{aligned} \tag{4}$$

where $\xi = [\xi_1, \ldots, \xi_i, \ldots, \xi_{n^2}]$ is the vector of slack variables. The squared hinge loss function has been widely adopted in vision applications. Compared to the least-squares SVM (LS-SVM) [39], the squared hinge loss penalizes the estimation error caused by misclassification, and usually helps generalize better and perform robustly [30].

Similar to (3), as $\mathbf{X}$ is circulant, the SVM model can be equivalently formulated as:

$$\begin{aligned} &\min_{\mathbf{w},b,\xi} \; \|\mathbf{w}\|^2 + C\|\xi\|_2^2 \\ &\text{s.t.} \;\; \mathbf{y} \circ (\mathcal{F}^{-1}(\hat{\mathbf{x}}^* \circ \hat{\mathbf{w}}) + b\mathbf{1})) \geq \mathbf{1} - \xi, \end{aligned} \tag{5}$$

where $\mathbf{1}$ denotes a vector of 1s.

Class Labels of Translated Images. Let $\mathbf{p}^*$ denote the center position of the target object $\mathbf{x}^*$, and $\mathbf{p}_i$ as the position of the translated image $\mathbf{x}_i$. In object detection, the overlap ratio of $\mathbf{x}_i$ is used to measure the similarity between $\mathbf{x}^*$ and $\mathbf{x}_i$. In this work, we use the overlap ratio to guide the labeling of the translated image $\mathbf{x}_i$ where samples above a pre-defined upper threshold are considered as positive, and samples below a lower threshold are treated as negative. The optimal upper and lower thresholds for SCF, MSCF and KSCF are empirically determined (see Section 4).

We use the following confidence map of object position [43] to define the class label:

$$m(\mathbf{p}_i, \mathbf{p}^*) = \gamma \exp\Big(-\alpha\|\mathbf{p}_i - \mathbf{p}^*\|^\beta\Big), \tag{6}$$

where γ is a normalization constant, α and β are the scale and shape parameters, respectively. Based on $m(\mathbf{p}_i, \mathbf{p}^*)$, we divide the samples into a labeled subset Ω^l and an unlabeled one Ω^u,

$$i \in \begin{cases} \Omega^u, & \text{if } \theta_l < m(\mathbf{p}_i, \mathbf{p}^*) < \theta_u, \\ \Omega^l, & \text{otherwise}, \end{cases} \tag{7}$$

where θ_l and θ_u are the lower and upper thresholds, respectively. For the labeled samples, we define the class labels as follows:

$$y_i = \begin{cases} 1, & \text{if } m(\mathbf{p}_i, \mathbf{p}^*) \geq \theta_u, \\ -1, & \text{if } m(\mathbf{p}_i, \mathbf{p}^*) \leq \theta_l. \end{cases} \tag{8}$$

The sample $\mathbf{x}_i$ with $i \in \Omega^u$ is treated as an unlabeled sample, and its class label $y_i \in \{-1, 1\}$ is adaptively determined in the learning stage by a semi-supervised learning manner.

Comparisons with Existing CF-Based Trackers. As illustrated in Fig. 2a, existing CF-based trackers generally use the Regularized Least Squares (RLS) model [33]. That is, with the continuous confidence map $\mathbf{m}$, RLS-based CFs seek the optimal correlation filter by minimizing the mean squared error (MSE) between the pre-defined confidence map and actual output,

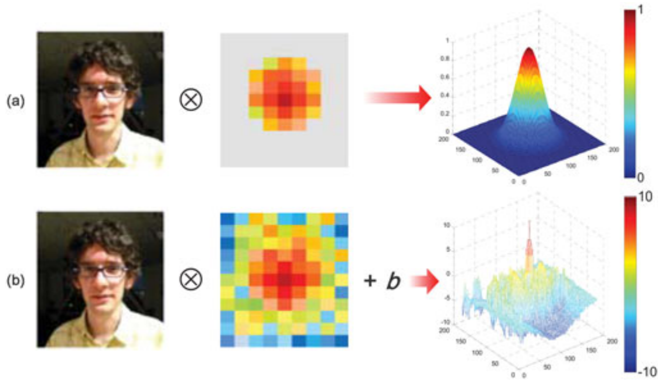


Fig. 2. Differences between the proposed SCF model and existing CF approaches [8], [23], [43]. (a) Existing CF-based models are designed to learn correlation filters that make the actual output being close to the predefined confidence maps. (b) The SCF model aims to learn a support correlation filter together with the bias b for distinguishing a target object from the background based on the max-margin principle. The peak value in the right response map of (b) locates the target object well.

$$\min_{\mathbf{w}} \ \lambda\|\mathbf{w}\|^2 + \|\mathbf{X}^\top\mathbf{w} - \mathbf{m}\|_2^2, \tag{9}$$

which has the closed-form solution,

$$\hat{\mathbf{w}} = \frac{\hat{\mathbf{x}}^* \circ \hat{\mathbf{m}}}{\hat{\mathbf{x}}^* \circ \hat{\mathbf{x}} + \lambda}. \tag{10}$$

The success of CF-based trackers should be attributed more to dense sampling than regularized regression. As most leading non-CF based trackers are based on SVM [21], [42], the incorporation of dense sampling and SVM could further improve tracking performance and outperform RLS-based methods. Unlike RLS-based CF-based methods which impose symmetric penalty to the estimate error, the proposed SCF model adopts the asymmetric squared hinge loss by only penalizing the error caused by misclassification. Thus, the convolutional outputs of SCF can be higher for positive samples, and lower for negative samples, resulting in the non-smooth convolutional map in Fig. 2b.

Using the label assignment schemes in (7) and (8), we can use the unlabeled samples in a semi-supervised learning manner to alleviate the label ambiguity problem. The importance of SVM and label ambiguity issues have been demonstrated in object detection [16]. The proposed model copes with both issues (classification and label ambiguity) in dense sampling setting for effective visual tracking.

3.2 Alternating Optimization

In this section, we modify the model in (5) by taking unlabeled samples into account, and propose an alternating optimization algorithm to learn SCFs efficiently. To exploit the property of the circulant matrix for learning SCFs, we let $\xi = \mathbf{e} + \mathbf{1} - \mathbf{y} \circ (\mathcal{F}^{-1}(\hat{\mathbf{x}}^* \circ \hat{\mathbf{w}}) + b\mathbf{1})$, and the semi-supervised SVM model in (5) is formulated as:

$$\begin{aligned} \min_{\mathbf{w},b,\mathbf{e},y_i(i\in\Omega^u)} \quad & \|\mathbf{w}\|^2 + C\|\mathbf{y}\circ(\mathcal{F}^{-1}(\hat{\mathbf{x}}^*\circ\hat{\mathbf{w}}) + b\mathbf{1}) - \mathbf{1} - \mathbf{e}\|_2^2 \\ \text{s.t.} \quad & \mathbf{e} \geq 0. \end{aligned} \tag{11}$$

With this formulation, the subproblem on each of $\mathbf{e}$, $\{\mathbf{w}, b\}$, and $y_i(i \in \Omega^u)$ has its own closed-form solution when the other variables are given. Therefore, the solution of the above model can be efficiently solved using the alternating optimization algorithm by iterating between the following three steps:

Updating $\mathbf{e}$. Given $\{\mathbf{w}, b\}$ and $y_i(i \in \Omega^u)$, we let $\mathbf{e}_0 = \mathbf{y} \circ (\mathcal{F}^{-1}(\hat{\mathbf{x}}^* \circ \hat{\mathbf{w}}) + b\mathbf{1}) - \mathbf{1}$, and the subproblem on $\mathbf{e}$ becomes:

$$\min_{\mathbf{e}} \ \|\mathbf{e} - \mathbf{e}_0\|^2, \ \text{s.t. } \mathbf{e} \geq 0. \tag{12}$$

The $\mathbf{e}$ subproblem has the closed-form solution:

$$\mathbf{e} = \max\{\mathbf{e}_0, 0\}. \tag{13}$$

Updating $\{\mathbf{w}, b\}$. As $\mathbf{y}$ is the class label vector with $y_i \in \{1, -1\}$, we have $\mathbf{y} \circ \mathbf{y} = \mathbf{1}$ and $\|\mathbf{y} \circ \mathbf{v}\|^2 = \|\mathbf{v}\|^2$. Given $\mathbf{e}$ and $y_i(i \in \Omega^u)$, by letting $\mathbf{q} = \mathbf{y} + \mathbf{y} \circ \mathbf{e}$, the subproblem on $\{\mathbf{w}, b\}$ becomes:

$$\min_{\mathbf{w},b} \ \|\mathbf{w}\|^2 + C\|\mathcal{F}^{-1}(\hat{\mathbf{x}}^* \circ \hat{\mathbf{w}}) + b\mathbf{1} - \mathbf{q}\|_2^2. \tag{14}$$

With $\mathbf{u} = [\mathbf{w}; b]$, we have the closed-form solution on $\mathbf{u}$ since it is a quadratic programming problem. However, this approach fails to exploit the circulant property of $\mathbf{X}$. Instead, we can eliminate b from (14). The mean vector of the circulant matrix $\mathbf{X}$ can be written as $\bar{\mathbf{x}} = \frac{1}{n^2}\sum_{i=1}^{n^2} \mathbf{x}_i = \bar{x}\mathbf{1}$, where $\bar{x} = \frac{1}{n^2}\sum_{j=1}^{n^2} x_{ij}$. By defining $\mathbf{X}_c = \mathbf{X} - \bar{x}\mathbf{1}\mathbf{1}^\top$ and $\mathbf{x}_c = \mathbf{x} - \bar{\mathbf{x}}$, it is clear that $\mathbf{X}_c$ is also a circulant matrix with $\hat{\mathbf{x}}_c = \mathcal{F}(\mathbf{x}_c)$. Then, the closed-form solution to the subproblem on $\{\mathbf{w}, b\}$ can be obtained by the following lemma:

Lemma 1. *The subproblem on* $\{\mathbf{w}, b\}$ *is reformulated as:*

$$\min_{\mathbf{w}} \ \|\mathbf{w}\|^2 + C\|\mathbf{X}_c^\top\mathbf{w} - \mathbf{q}_c\|_2^2. \tag{15}$$

The optimal solutions to $\mathbf{w}$ *and* b *are:*

$$\hat{\mathbf{w}} = \frac{\hat{\mathbf{x}}_c^* \circ \hat{\mathbf{q}}_c}{\hat{\mathbf{x}}_c^* \circ \hat{\mathbf{x}}_c + \frac{1}{C}}, \qquad b = \bar{q}, \tag{16}$$

where $\bar{q}$ *is the mean of* $\mathbf{q}$, *and* $\mathbf{q}_c = \mathbf{q} - \bar{q}\mathbf{1}$.

The proof of Lemma 1 is given in Appendix A.1.

Updating $y_i(i \in \Omega^u)$. Given $\mathbf{e}$ and $\{\mathbf{w}, b\}$, $y_i(i \in \Omega^u)$ can be updated by solving the following subproblem,

$$\min_{y_i\in\{-1,1\}} \left\|y_i(\mathbf{w}^\top\mathbf{x}_i + b) - 1 - e_i\right\|^2, i \in \Omega^u,$$

and the closed-form solution is,

$$y_i = \begin{cases} 1, & \text{if } \mathbf{w}^\top\mathbf{x}_i + b \geq 0 \\ -1, & \text{if } \mathbf{w}^\top\mathbf{x}_i + b < 0. \end{cases}$$

The algorithm above can be easily extended to the fully-supervision setting by excluding the step of updating $y_i(i \in \Omega^u)$. Thus, the proposed model can be used in either fully or semi-supervision settings, and handle both SVM classification and label ambiguity while exploiting the circulant property.

As illustrated in Fig. 1, when the tth frame $\mathbf{x}^t$ with class labels $\mathbf{y}^t$ arrives, the proposed algorithm learns support correlation filters by iterating between updating $\mathbf{e}$ and updating $\{\mathbf{w}, b\}$ until convergence. The complexity of updating $y_i(i \in \Omega^u)$ is $O(n^2)$. Given $\{\mathbf{x}^t, \mathbf{y}^t, \mathbf{w}, b\}$, the updating of $\mathbf{e}$ can be computed element-wise, which has the complexity of $O(n^2)$. Given $\{\mathbf{x}^t, \mathbf{y}^t, \mathbf{e}\}$, the complexity of updating b is $O(n^2)$ and that of updating $\mathbf{w}$ is $O(n^2 \log n)$. Thus, the complexity is $O(n^2 \log n)$ per iteration which makes our algorithm efficient in learning support correlation filters. The main steps of the proposed learning algorithm for support correlation filters are summarized in Algorithm 1.

Algorithm 1. SCF Model Training

```
Input: training image patch $\mathbf{x}_t (n \times n)$, center position $\mathbf{p}^*$
Output: $(\hat{\mathbf{w}}, b)$.
 1: Initialize $\hat{\mathbf{w}}_0$, $b_0$, $\Omega^l, \Omega^u$, $k = 1$,
 2: $\forall i \in \Omega^l, \mathbf{y}_t(i) = \begin{cases} 1, & \text{if } m(\mathbf{p}_i, \mathbf{p}^*) \geq \theta_u \\ -1, & \text{if } m(\mathbf{p}_i, \mathbf{p}^*) \leq \theta_l. \end{cases}$
 3: while not converged do
 4:     // Line 5-6 : update $\mathbf{e}_k$.
 5:     $\mathbf{d} = \mathbf{y}_t \circ \left(\mathcal{F}^{-1}(\hat{\mathbf{x}}^*{}_t \circ \hat{\mathbf{w}}_{k-1}) + b\mathbf{1}\right) - \mathbf{1}$,
 6:     $\mathbf{e}_k = \max(0, \mathbf{d})$,
 7:     // Line 8-10 : update $\mathbf{q}_k$, $b_k$, $\mathbf{p}_k$.
 8:     $\mathbf{q}_k = \mathbf{y}_t + \mathbf{y}_t \circ \mathbf{e}_k$,
 9:     $b_k = mean(\mathbf{q}_k)$,
10:     $\mathbf{p}_k = \mathbf{q}_k - b_k \mathbf{1}$,
11:     // Line 12-13: update $\mathbf{w}_k$.
12:     $\mathbf{x}_c = \mathbf{x}_t - \bar{x}_t \mathbf{1}$
13:     $\hat{\mathbf{w}}_k = \frac{\hat{\mathbf{x}}^*{}_c \circ \hat{\mathbf{p}}_k}{\hat{\mathbf{x}}^*{}_c \circ \hat{\mathbf{x}}_c + \mathbf{1}/C}$,
14:     // Line 15 : update $\mathbf{y}_t$.
15:     $\forall i \in \Omega^u, \mathbf{y}_t(i) = \begin{cases} 1, & \text{if } \mathbf{w}^\top \mathbf{x}_t(i) + b \geq 0 \\ -1, & \text{if } \mathbf{w}^\top \mathbf{x}_t(i) + b < 0 \end{cases}$
16:     $k \leftarrow k + 1$.
17: end while
```

Convergence. In the fully-supervision setting, the objective is a convex quadratic function and the linear constraint $\mathbf{e} \geq 0$ is convex, thereby making it solvable by convex optimization. The proposed algorithm converges to the global optimum with the q-linear convergence rate. For presentation clarity, we give the detailed analysis and proof on its optimality condition, global convergence, and convergence rate in Appendix B. Based on the optimality condition, we define

$$\begin{cases} \mathbf{r}_1 \doteq \mathbf{w} + C\mathcal{F}^{-1}(\hat{\mathbf{x}} \circ \hat{\mathbf{x}}^* \circ \hat{\mathbf{w}} - \hat{\mathbf{r}}), & \\ \mathbf{r}_2(i) \doteq e_i + 1 - (\mathbf{y} \circ (\mathcal{F}^{-1}(\hat{\mathbf{x}}^* \circ \hat{\mathbf{w}}) + b))_i, & \text{if } e_i > 0 \\ \mathbf{r}_3(i) \doteq (\mathbf{y} \circ (\mathcal{F}^{-1}(\hat{\mathbf{x}}^* \circ \hat{\mathbf{w}}) + b))_i - 1, & \text{if } e_i = 0 \end{cases} \tag{17}$$

and adopt the following stopping criterion:

$$\max\{\|\mathbf{r}_1\|_\infty, \max_{e_i > 0}\{\|\mathbf{r}_2(i)\|\}, \max_{e_i = 0}\{\mathbf{r}_3(i)\}\} \leq \epsilon. \tag{18}$$

Comparisons with MMCF. The objective function of maximum margin correlation filters (MMCFs) [7], [35] includes localization and max-margin criteria. The localization criterion is used to exploit the circulant property with the standard CF form, while the max-margin criterion treats each image as one sample and does not consider cyclic translations. In contrast, we directly use the circulant property in the SVM with the max-margin loss and develop a novel alternating minimization algorithm to solve the proposed model.

Specifically, the proposed SCF model and learning algorithm are different from the MMCF approach in three aspects. First, the training samples for MMCF are N images of $n \times n$ pixels, while those for SCF are n^2 translated images of $n \times n$ pixels. We exploit the circulant property of the data matrix $\mathbf{X}$ to develop an efficient learning algorithm. Second, we propose an alternating optimization algorithm to solve the proposed model, which has the complexity of $O(n^2 \log n)$. In contrast, the MMCF method adopts the conventional SMO algorithm with the complexity of $O(N^2 d)$ where d is the dimension of the sample. For visual tracking considered in this work, we have $N = n^2$ and $d = n^2$, and the complexity of MMCF is $O(n^6)$, which is computationally expensive for real-time applications. Third, the proposed model has the squared hinge loss and regularizer terms, while the MMCF method adopts the hinge loss and includes an extra CF-based localization term.

3.3 Multi-Channel SCF

Different local descriptors, e.g., color attributes, HOG, and SIFT, provide rich image features for effective visual tracking. We treat local descriptors as multi-channel images where multiple measurements are associated to each pixel. To exploit multi-dimensional features, we propose the multi-channel SCF as follows:

$$\begin{aligned} &\min_{\mathbf{w}, b, \boldsymbol{\xi}, y_i (i \in \Omega^u)} \|\mathbf{w}\|^2 + C\|\boldsymbol{\xi}\|_2^2 \\ &\text{s.t. } \mathbf{y} \circ \left(\mathcal{F}^{-1}\left(\sum_{l=1}^{L} (\hat{\mathbf{x}}^l)^* \circ \hat{\mathbf{w}}^l\right) + b\mathbf{1}\right) \geq \mathbf{1} - \boldsymbol{\xi}, \end{aligned} \tag{19}$$

where L is the number of channels, and $\mathbf{x}^l$ and $\mathbf{w}^l$ denote the lth channel of the image and correlation filter, respectively. To learn the proposed MSCF model, we adopt the same equations on updating $\mathbf{e}$, b and $y_i (i \in \Omega^u)$, and compute $\mathbf{w}$ by solving the following problem:

$$\min_{\mathbf{w}} \sum_{l=1}^{L} \|\hat{\mathbf{w}}^l\|^2 + C\| \sum_{l=1}^{L} (\hat{\mathbf{x}}^l)^* \circ \hat{\mathbf{w}}^l - \hat{\mathbf{r}}\|_2^2, \tag{20}$$

where $\hat{\mathbf{w}} = [\hat{\mathbf{w}}^1; \hat{\mathbf{w}}^2; \ldots; \hat{\mathbf{w}}^L]$, and $\hat{\mathbf{r}} = \hat{\mathbf{q}} - b\hat{\mathbf{1}}$.

Let $\hat{\mathbf{X}} = [\text{diag}(\hat{\mathbf{x}}^1) \ \text{diag}(\hat{\mathbf{x}}^2) \ \ldots \ \text{diag}(\hat{\mathbf{x}}^L)]$. The closed-form solution for $\hat{\mathbf{w}}$ can be directly obtained by

$$\hat{\mathbf{w}} = \left(\hat{\mathbf{X}}^H \hat{\mathbf{X}} + \frac{1}{C}\mathbf{I}\right)^{-1} \hat{\mathbf{X}}^H \hat{\mathbf{r}}. \tag{21}$$

where $\mathbf{I}$ is the identity matrix. Note that $\hat{\mathbf{X}}$ is an $n^2 \times Ln^2$ matrix, $\hat{\mathbf{w}}$ is an $Ln^2 \times 1$ vector, and $\hat{\mathbf{r}}$ is an $n^2 \times 1$ vector. With $1 \leq l \leq L$ and $1 \leq j \leq n^2$, we use $\hat{\mathbf{w}}^l(j)$ to denote $((l-1)n^2 + j)$th element of $\hat{\mathbf{w}}$, and $\hat{\mathbf{r}}(j)$ denotes the jth element of $\hat{\mathbf{r}}$. In addition, $\hat{\mathbf{w}}(j)$ is introduced to denote the subvector $\hat{\mathbf{w}}(j) = [\hat{\mathbf{w}}^1(j), \hat{\mathbf{w}}^2(j), \ldots, \hat{\mathbf{w}}^L(j)]^\top$ of $\hat{\mathbf{w}}(\hat{\mathbf{r}})$. Similarly, $\hat{\mathbf{x}}^l$ $(1 \leq l \leq L)$ is an $n^2 \times 1$ vector. We use $\hat{\mathbf{x}}^l(j)$ to denote the jth element of $\hat{\mathbf{x}}^l$, and define the sub-vector $\hat{\mathbf{x}}(j) = [\hat{\mathbf{x}}^1(j), \hat{\mathbf{x}}^2(j), \ldots, \hat{\mathbf{x}}^L(j)]^\top$. As $\hat{\mathbf{X}}$ has the diagonal block structure, $\hat{\mathbf{w}}(j)$ only depends on $\hat{\mathbf{x}}(j)$ and $\hat{\mathbf{r}}(j)$. Hence, the subproblem on $\hat{\mathbf{w}}$ can be further decomposed into n^2 systems of equations:

$$\left(\hat{\mathbf{x}}(j)\hat{\mathbf{x}}(j)^H + \frac{1}{C}\mathbf{I}\right)\hat{\mathbf{w}}(j) = \hat{\mathbf{x}}(j)\hat{\mathbf{r}}(j). \tag{22}$$

Detailed derivation of (22) is provided in the supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPAMI.2018.2829180.

In [14], Galoogahi et al. solve these n^2 systems of equations by an algorithm with the complexity of $O(n^2L^3 + Ln^2 \log n)$. We note that the matrix on the left hand of (22) is a rank-one matrix and a scaled identity matrix. Based on the Sherman-Morrison-Woodbury formula [26] we have

$$\left(\hat{\mathbf{x}}(j)\hat{\mathbf{x}}(j)^H + \frac{1}{C}\mathbf{I}\right)^{-1} = C\left(\mathbf{I} - \frac{C\hat{\mathbf{x}}(j)\hat{\mathbf{x}}(j)^H}{1 + C\hat{\mathbf{x}}(j)^H\hat{\mathbf{x}}(j)}\right). \tag{23}$$

The closed-form solution for $\hat{\mathbf{w}}(j)$ is then obtained by

$$\hat{\mathbf{w}}(j) = \frac{C\hat{\mathbf{x}}(j)\hat{\mathbf{r}}(j)}{1 + C\hat{\mathbf{x}}(j)^H\hat{\mathbf{x}}(j)}. \tag{24}$$

The expression in (24) is similar to the solution of single-channel CF, i.e., MOSSE. However, the solution of MOSSE is defined in the batch setting for all elements, but (24) is in the element-wise setting for all channels. From (24), the solution to the lth channel for all elements can be written as:

$$\hat{\mathbf{w}}^l = \frac{C\left(\hat{\mathbf{x}}^l\right)^* \circ \hat{\mathbf{r}}}{1 + C\sum_{l=1}^{L}\left(\hat{\mathbf{x}}^l\right)^* \circ \hat{\mathbf{x}}^l}. \tag{25}$$

When $L = 1$, the solution in (25) degrades to a SCF, which has the same expression but adopts different definition on $\hat{\mathbf{r}}$ with MOSSE. For MOSSE, $\hat{\mathbf{r}}$ denotes the Fourier transform of Gaussian-shaped confidence map. For SCF, $\hat{\mathbf{r}}$ is defined as $\hat{\mathbf{q}} - b\hat{\mathbf{1}}$.

Using the Sherman-Morrison-Woodbury formula, we provide a unified solution to both the single-channel and multi-channel problems. It should be noted that all $\hat{\mathbf{x}}^l$s can be pre-computed with the complexity of $O(n^2 \log n)$. As such, the proposed algorithm only involves one DFT, one IDFT and $O(Ln^2)$ element-wise operations per iteration, and the complexity is $O(n^2 \log n)$.

3.4 Kernelized SCF

Given the kernel function $K(\mathbf{x}, \mathbf{x}') = \langle \psi(\mathbf{x}), \psi(\mathbf{x}') \rangle$, the proposed kernelized SCF model can be extended to learn the nonlinear decision function:

$$f(\mathbf{x}) = \mathbf{w}^\top \psi(\mathbf{x}) + b = \sum_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b, \tag{26}$$

where $\psi(\mathbf{x})$ stands for the nonlinear feature mapping implicitly determined by the kernel function $K(\mathbf{x}, \mathbf{x}')$, and $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_{n^2}]^\top$ is the coefficient vector to be learned.

Denote by $\mathbf{K}$ the kernel matrix with $K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$. As noted in [24], for some kernel functions (e.g., Gaussian RBF and polynomial) which are permutation invariant, the kernel matrix $\mathbf{K}$ is circulant. Let $\mathbf{k}^{\mathbf{xx}}$ be the first row of the circulant matrix $\mathbf{K}$. Therefore, the matrix-vector multiplication $\mathbf{K}\boldsymbol{\alpha}$ can be efficiently computed via DFT:

$$\mathbf{K}\boldsymbol{\alpha} = \mathcal{F}^{-1}(\hat{\mathbf{k}}^{\mathbf{xx}} \circ \hat{\boldsymbol{\alpha}}), \tag{27}$$

and we have,

$$\|\mathbf{w}\|^2 = \boldsymbol{\alpha}^\top \mathbf{K}\boldsymbol{\alpha} = \boldsymbol{\alpha}^\top \mathcal{F}^{-1}(\hat{\mathbf{k}}^{\mathbf{xx}} \circ \hat{\boldsymbol{\alpha}}). \tag{28}$$

Based on (27) and (28), the proposed kernelized SCF model is formulated as

$$\begin{aligned} \min_{\boldsymbol{\alpha}, b, \mathbf{e}, y_i(i\in\Omega^u)} \quad & \boldsymbol{\alpha}^\top \mathcal{F}^{-1}(\hat{\mathbf{k}}^{\mathbf{xx}} \circ \hat{\boldsymbol{\alpha}}) \\ & + C\|\mathbf{y} \circ (\mathcal{F}^{-1}(\hat{\mathbf{k}}^{\mathbf{xx}} \circ \hat{\boldsymbol{\alpha}}) + b\mathbf{1}) - \mathbf{1} - \mathbf{e}\|_2^2 \\ \text{s.t.} \quad & \mathbf{e} \geq 0. \end{aligned} \tag{29}$$

We use the alternating optimization algorithm by iteratively solving $\mathbf{e}$, $\{\boldsymbol{\alpha}, b\}$, and $y_i(i \in \Omega^u)$. The solution of the subproblems with $\mathbf{e}$ and $y_i(i \in \Omega^u)$ are similar to those in the SCF model. By fixing $\mathbf{e}$, the subproblem on $\{\boldsymbol{\alpha}, b\}$ can be reformulated as:

$$\min_{\boldsymbol{\alpha}, b} \ \boldsymbol{\alpha}^\top \mathbf{K}\boldsymbol{\alpha} + C\|\mathbf{K}\boldsymbol{\alpha} + b\mathbf{1} - \mathbf{q}\|_2^2, \tag{30}$$

where $\mathbf{q} = \mathbf{y} + \mathbf{y} \circ \mathbf{e}$. Similar to [37], by considering the circulant property of $\mathbf{K}$ we define the centered kernel matrix $\mathbf{K}_c$ as:

$$\mathbf{K}_c = \mathbf{K} - \bar{k}\mathbf{1}\mathbf{1}^\top, \tag{31}$$

where $\bar{k}$ is the mean of $\mathbf{k}^{\mathbf{xx}}$. Let $\mathbf{k}_c^{\mathbf{xx}}$ denote the first row of the circulant matrix $\mathbf{K}_c$. The closed-form solution is given by the following lemma:

Lemma 2. *The solutions to the* $\{\boldsymbol{\alpha}, b\}$ *in (30) are:*

$$\hat{\boldsymbol{\alpha}}^* = \frac{\hat{\mathbf{q}}_c}{\hat{\mathbf{k}}_c^{\mathbf{xx}} + \frac{1}{C}}, \quad b = \bar{q}, \tag{32}$$

where $\bar{q}$ *is the mean of* $\mathbf{q}$, *and* $\mathbf{q}_c = \mathbf{q} - \bar{q}\mathbf{1}$.

The proof of Lemma 2 is given in Appendix A.2. For image features with L channels, the complexity to compute kernel matrix is $O(Ln^2 \log n)$. After that, the learning process only requires element-wise operations, one DFT and one IDFT per iteration, and the complexity is $O(n^2 \log n)$. Thus, the proposed KSCF model leverages rich features from the nonlinear filters without increasing computational load significantly.

4 Performance Evaluation

We use the benchmark dataset and protocols [41] to evaluate the proposed SCF algorithms. We evaluate several variants of the proposed method, e.g., SCF, MSCF, KSCF, and SKSCF, to analyze the effect of feature representations and kernel functions. Similar to [12], we also implement a scale-adaptive KSCF (SKSCF) method. The tracking results and source code will be available at https://github.com/wuxiaohe/SCF.

4.1 Experimental Setup

Datasets and Evaluated Tracking Methods. To assess the performance of the proposed methods, experiments are carried out on a benchmark dataset [41] of 50 challenging image sequences annotated with 11 attributes. For the first frame of each sequence, the bounding box of the target object is provided for fair comparisons. For comprehensive comparisons, we evaluate the baseline SCF, multi-channel SCF, kernelized SCF and scale-adaptive KSCF methods. The SCF and MSCF methods are designed in the linear space with raw pixels, and multi-channel features based on HOG [11] as well as color names (CN) [13], respectively. The KSCF

TABLE 1
Results of MSCF and DCF [24] (MSCF/DCF) with Different Feature Representations

Features	Raw pixels	CN	HOG	HOG + CN
Mean DP (%)	64.9/44.4	66.3/48.0	78.4/71.9	**80.6**/76.2
Mean AUC (%)	44.6/31.2	44.9/34.8	53.7/50.1	**55.5**/53.2
Mean FPS (s)	76/278	62/210	64/**292**	54/151

and SKSCF algorithms are evaluated by using the Gaussian kernel on multi-channel feature representations. Furthermore, we compare the proposed trackers with the other trackers based on correlation filters (e.g., MOSSE [8], CSK [23], KCF [24], DCF [24], STC [43] and CN [13]), existing SVM based trackers (e.g., Struck [21] and MEEM [42]), and other state-of-the-art methods (e.g., TGPR [15], SCM [45], TLD [28], L1APG [5], MIL [3], ASLA [27] and CT [44]).

Evaluation Protocols. We use the one-pass evaluation (OPE) protocol [41] which reports the precision and success plots based on the position error and bounding box overlap metrics with respect to the ground truth object locations. For precision plots, the distance precision at a threshold of 20 pixels (DP) is reported. For success plots, the area under the curve (AUC) is computed. In addition, the frames per second (FPS) that each method is able to process is discussed.

Parameter Settings. The experiments are carried out on a desktop computer with an Intel Xeon 2 core 3.30 GHz CPU and 32 GB RAM. The proposed SCF-based trackers involve a few model parameters, i.e., trade-off parameter C, scale parameter α and shape parameter β of confidence maps, and lower and upper thresholds (θ_l, θ_u) in (8). The KSCF method has one extra parameter σ for the Gaussian RBF kernel function, and the SKSCF scheme has two other parameters: number of scales S and scale factor a. For online tracking, the model is updated by linear interpolation with the adaption rate ρ [23], [24].

In all experiments, the model parameters are fixed for each SCF-based tracker. For all SCF-based trackers, the trade-off C and shape parameter β are fixed to 10^4 and 1.5, respectively. We empirically find that the optimal setting for the thresholds (θ_l, θ_u) in (8) may be affected by the introduction of multi-channel features, kernel function, and scale estimation. Thus, we set (θ_l, θ_u) to $(0.3, 0.7)$ for SCF, $(0.4, 0.9)$ for MSCF, $(0.5, 0.6)$ for KSCF and $(0.3, 0.6)$ for SKSCF. We choose the number of scales $S = 21$ and scale factor $a = 1.04$ for SKSCF. The adaption rate ρ is set to 0.075 for raw pixel features, and 0.025 for multi-channel features, respectively. The kernel parameter σ is set to 0.2 for KSCF, and 0.5 for SKSCF. The orientations and cell size are set to 9 and 4 for HOG features.

TABLE 2
Results of KSCF and KCF [24] (KSCF/KCF) with Different Feature Representations

Features	Raw pixels	CN	HOG	HOG + CN
Mean DP (%)	64.4/55.3	68.1/57.3	79.3/73.2	**85.0**/75.8
Mean AUC (%)	45.3/40.0	46.9/41.8	53.2/50.7	**57.5**/53.0
Mean FPS (s)	40/154	37/120	44/**172**	35/102

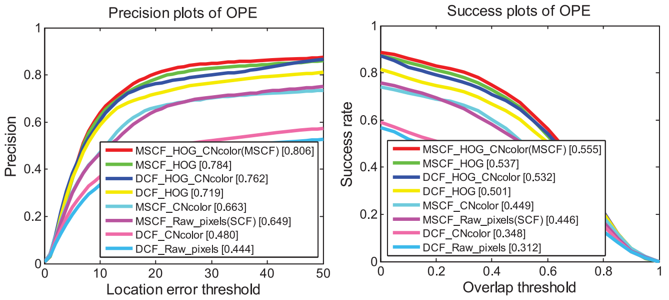


Fig. 3. OPE plots of the MSCF and DCF [24] with different feature representations. The AUC values are shown next to the legends.

4.2 Evaluation on SCF-Based Trackers

In this section, we first evaluate the effect of feature representations and kernel functions, and then compare four variants of the SCF-based trackers, i.e., SCF, MSCF, KSCF, and SKSCF, in terms of both accuracy and efficiency. The results of the corresponding CF-based trackers are also reported for all SCF-based methods.

We consider three typical feature representations, i.e., raw pixels, HOG features [11], and color names (CN) [13]. The results of the MSCF and KSCF methods are listed in Tables 1 and 2. The result for each feature representation is optimal by varying the parameters $\beta \in \{0.5, 1, 1.5, 2\}$, $\rho \in \{0.02, 0.04, 0.075\}$, $\theta_l \in \{0.3, 0.4, 0.5\}$ and $\theta_u \in \{0.6, 0.7, 0.8, 0.9\}$. These parameters are then fixed for all the following experiments. The Gaussian RBF kernel width σ is set to 0.2 for KSCF, and 0.5 for SKSCF.

The OPE plots of MSCF with linear DCF [24] and KSCF with nonlinear KCF [24] are shown in Figs. 3 and 4. Compared with raw pixels and CN features, the method with HOG representation significantly improves the tracking performance in terms of mean DP and mean AUC. For MSCF, the implementation using CN and HOG features outperforms raw pixels by 1.4% and 13.5% in terms of mean DP. For KSCF, the tracker using CN and HOG features outperforms raw pixels by 3.7% and 14.9% in terms of mean DP. The MSCF tracker with the combination of CN and HOG is further improved to 80.6% in terms of DP. Similarly, the performance of the KSCF method is improved to 85.0% in terms of DP with the use of CN and HOG features. Compared with the DCF [24] and KCF [24] methods, the proposed MSCF and KSCF algorithms achieve higher DP and AUC values for each feature representation. Tables 1 and 2 show that both KSCF and MSCF perform in real-time even using the

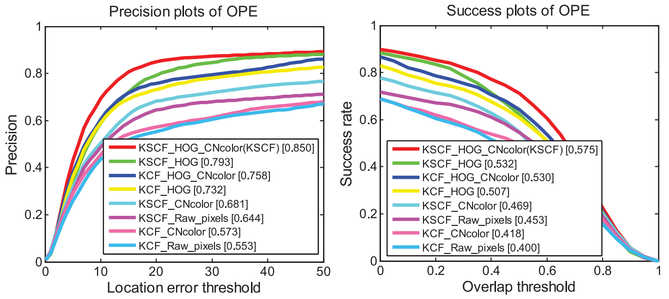


Fig. 4. OPE plots of the KSCF and KCF [24] methods with different feature representations.

TABLE 3
Results of KSCF with Different Kernels

Kernels	Linear	Polynomial	Gaussian
Mean DP (%)	82.0	84.2	**85.0**
Mean AUC (%)	56.2	57.1	**57.5**
Mean FPS (*s*)	**94**	55	35

representation based on HOG and CN features. Note that DCF and KCF [24] are based on the RLS model. The results indicate that SVM with squared hinge loss performs much better than RLS in visual tracking.

Furthermore, we evaluate the effects of kernel functions on KSCF using HOG and CN features, including linear kernel $\mathbf{K}_l(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{x}_j$, polynomial kernel $\mathbf{K}_p(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^\top \mathbf{x}_j + 1)^d$, and Gaussian RBF kernel $\mathbf{K}_g(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\frac{1}{2\sigma^2}\|\mathbf{x}_i - \mathbf{x}_j\|^2)$. For $\mathbf{K}_p(\mathbf{x}_i, \mathbf{x}_j)$, the degree d is set as 2. For $\mathbf{K}_g(\mathbf{x}_i, \mathbf{x}_j)$, the kernel parameter σ is set as 0.2. Table 3 shows the results of KSCF with different kernels. Clearly the KSCF method with a nonlinear kernel outperforms the one with a linear kernel in terms of mean DP and mean AUC, and the one with Gaussian RBF kernel achieves the best performance.

We implement the SKSCF method by extending KSCF with the Gaussian RBF kernel, and compare four variants of the SCF-based trackers, i.e., SCF, MSCF, KSCF, and SKSCF. Table 4 shows the results of four SCF-based trackers, where the SKSCF method performs best, followed by the KSCF approach. On the other hand, the KSCF method is more efficient than the SKSCF approach. In the following experiments, we compare both KSCF and SKSCF methods with the other schemes based on correlation filters, SVMs, and other state-of-the-art tracking approaches.

4.3 Comparisons with CF-Based Trackers

We use the tracking benchmark dataset [41] to evaluate the proposed SCF-based algorithm against existing CF-based methods including MOSSE [8], CSK [23], KCF [24], DCF [24], STC [43], CN [13], DSST [12] and SAMF [31]. For fairness, we also report the results of DCF and KCF based on HOG+CN features.

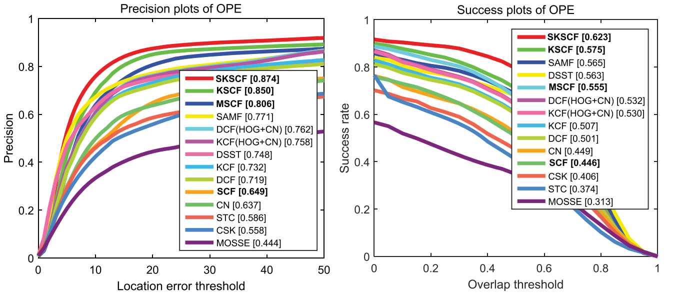


Fig. 5. OPE plots of the SCF methods (i.e., SCF, MSCF, KSCF, and SKSCF) and other CF-based trackers (i.e., MOSSE [8], CSK [23], DCF [24], KCF [24], STC [43], CN [13], DSST [12] and SAMF [31]).

Classic Correlation Filters. Fig. 5 shows the OPE plots of these trackers. The SCF, MOSSE [8], CSK [23] and STC [43] methods operate on raw pixels in the linear space. We note that the MOSSE method adopts the ridge regression function while the SCF algorithm uses the max-margin model. Although the CSK and STC methods operate on raw pixels, the CSK method is a kernelized CF-based tracker and the STC approach is a scale-adaptive tracking method. Overall, the SCF algorithm performs favorably against these CF-based methods based on regression and nonlinear kernels.

Multi-Channel Correlation Filters. The MSCF, CN [13], and DCF [24] methods are based on correlation filters using multi-channel features. The DCF method is based on HOG features and the CN approach is operated on color attributes, while the MSCF scheme uses the combination of HOG and color representations. Fig. 5 shows that the MSCF method performs well among these three trackers based on correlation filters.

Kernelized Correlation Filters. The KSCF method is compared with the corresponding kernelized KCF [24] and CSK [23] trackers. The CSK and KCF methods are based on raw pixels and HOG features, respectively. As shown in Table 4 and Fig. 6, the KSCF method based on HOG and CN features performs favorably against the KCF and CSK approaches.

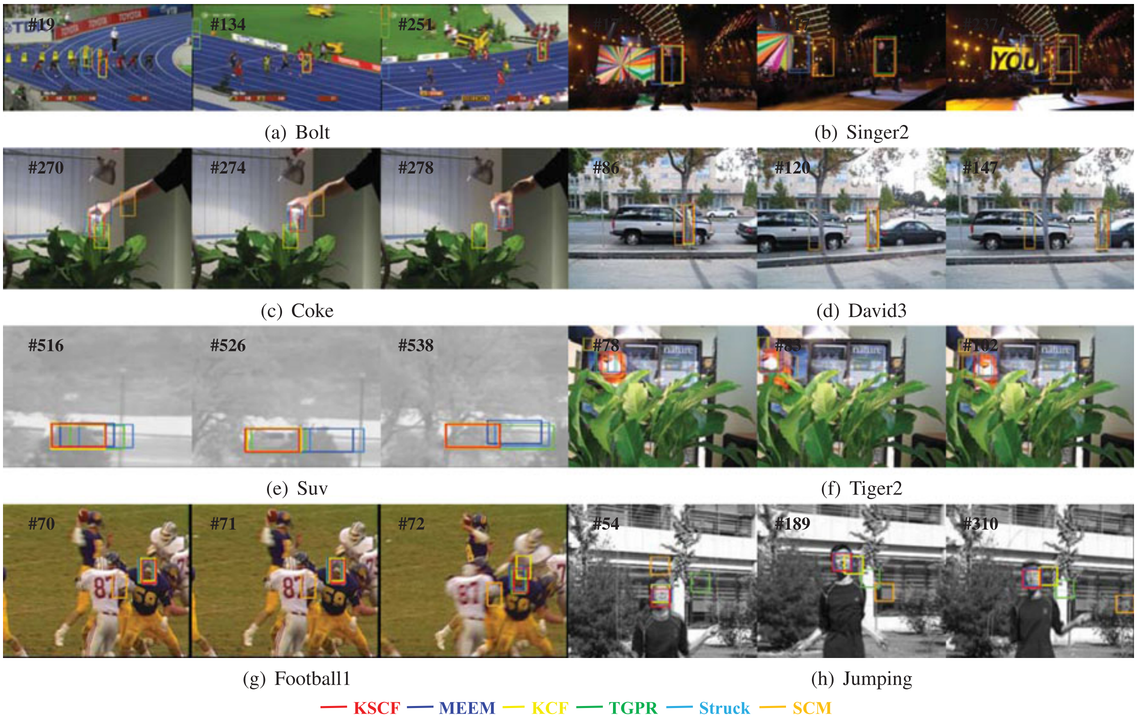
Scale-Adaptive Correlation Filters. The KSCF and SKSCF are evaluated against three scale-adaptive trackers: STC [43], DSST [12] and SAMF [31]. We note that the

TABLE 4
Performance of Tracking Methods Based on Correlation Filters: Top Three Results Are Shown in Red, Blue and Orange

Algorithms	SKSCF	KSCF	MSCF	SCF	KCF (HOG+CN)	DCF (HOG+CN)	SAMF [31]	DSST [12]	KCF [24]	DCF [24]	CN [13]	STC [43]	CSK [23]	MOSSE [8]
Mean DP (%)	**87.4**	**85.0**	**80.6**	62.8	75.8	76.2	77.1	74.8	73.2	71.9	63.7	58.6	55.8	44.4
Mean AUC (%)	**62.3**	**57.5**	55.5	48.9	53.0	53.2	**56.5**	56.3	50.7	50.1	44.9	37.4	40.6	31.3
Mean FPS (*s*)	8	35	54	76	102	151	14	30	172	**292**	79	**557**	151	**421**

TABLE 5
Comparison of KSCF, SKSCF, and the State-of-the-Art Trackers (Top Three Are Shown in Red, Blue and Orange)

Algorithms	SKSCF	KSCF	MEEM [42]	KCF [24]	TGPR [15]	SCM [45]	TLD [28]	ASLA [27]	L1APG [5]	MIL [3]	CT [44]
Mean DP (%)	**87.4**	**85.0**	**83.3**	73.2	71.8	65.2	60.6	54.5	49.4	48.8	41.5
Mean AUC (%)	**62.3**	**57.5**	**57.2**	50.7	51.1	50.1	43.4	44.2	38.6	36.9	30.8
Mean FPS (*s*)	8	**35**	10	**172**	0.5	1	22	8	3	28	**39**



(a) Bolt (b) Singer2 (c) Coke (d) David3 (e) Suv (f) Tiger2 (g) Football1 (h) Jumping

Fig. 6. Screenshots of tracking results on 8 challenging benchmark sequences. For the sake of clarity, we only show the results of six trackers, i.e., KSCF, KCF [24], MEEM [42], TGPR [15], Struck [21] and SCM [45].

DSST [12] and SAMF [31] methods have been shown to perform best and second best trackers in the recent tracking benchmark evaluation [29]. Both KSCF and SKSCF trackers perform significantly better than the STC method. In addition, the KSCF and SKSCF methods also significantly outperform the DSST and SAMF approaches by a large margin. Fig. 7 shows the OPE plots on all the sequences with the attribute of scale variation where the KSCF method performs favorably against the DSST and SAMF trackers. Overall, the KSCF algorithm performs favorably in terms of accuracy and speed.

4.4 Comparisons with SVM-Based Trackers

We evaluate the proposed KSCF and SKSCF with two state-of-the-art SVM-based methods, i.e., Struck [21] and MEEM [42], based on the structured and ensemble learning. Table 6 and Fig. 8 show that both KSCF and SKSCF algorithms perform favorably against the MEEM and Struck methods in all aspects. As shown in Fig. 6, the KSCF algorithm can track target objects more precisely than other methods in the *Singer2*, *Coke*, *Suv* and *Tiger2* sequences. The results show that dense sampling can be efficiently used with SVMs for effective visual tracking. Fig. 6 shows that the KSCF algorithm can track the objects more precisely in all challenging sequences, while the other trackers tend to drift away from the target objects.

4.5 Comparisons with State-of-the-Art Trackers

We evaluate the KSCF algorithm with the other state-of-the-art trackers, including MEEM [42], KCF [24], TGPR [15], SCM [45], TLD [28], L1APG [5], MIL [3], ASLA [27] and CT [44]. Fig. 10 shows the OPE plots, and Table 5 presents the

TABLE 6
Comparison of SVM-Based Trackers

Algorithms	SKSCF	KSCF	MEEM [42]	Struck [21]
Mean DP (%)	**87.4**	85.0	83.3	67.4
Mean AUC (%)	**62.3**	57.5	57.2	48.6
Mean FPS (*s*)	8	**35**	10	10

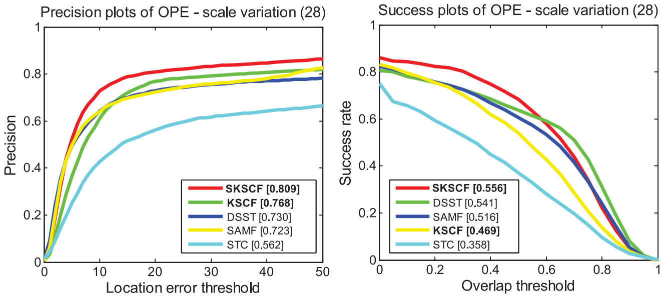


Fig. 7. OPE plots of the KSCF, SKSCF, DSST [12] and SAMF [31] methods on sequences with large scale variation.

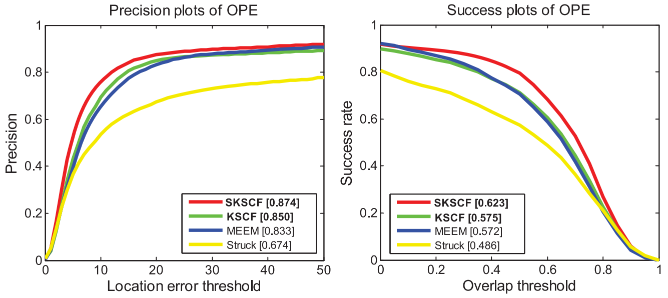


Fig. 8. OPE plots of the KSCF, SKSCF and other SVM-based trackers, including MEEM [42] and Struck [21].

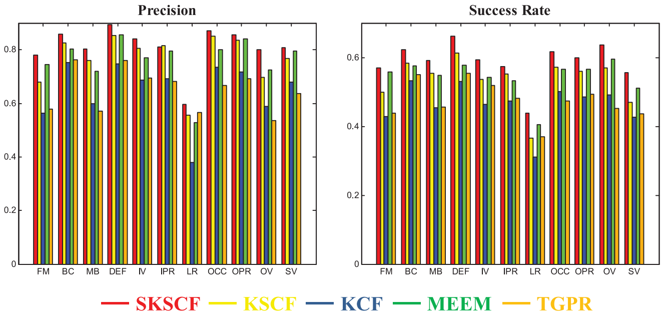


Fig. 9. Precision and success metrics of four top-performing trackers for the 11 attributes.

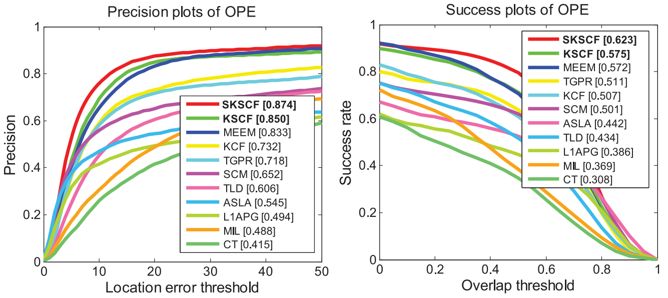


Fig. 10. OPE plots of the KSCF, SKSCF and other state-of-the art trackers, including MEEM [42], TGPR [15], KCF [24], SCM [45], TLD [28], ASLA [27], L1APG [5], MIL [3] and CT [44].

mean DP, AUC and FPS. Overall, the proposed KSCF and SKSCF algorithms perform favorably against the state-of-the-art methods including the TLD, SCM, TGPR and MEEM schemes.

The sequences in the benchmark dataset [41] are annotated with 11 challenging attributes for visual tracking, including illumination variation (IV), scale variation (SV), occlusion (OCC), deformation (DEF), motion blur (MB), fast motion (FM), in-plane rotation (IPR), out-of-plane rotation (OPR), out-of-view (OV), background clutter (BC), and low resolution (LR). Table 7 shows the performance of the SKSCF, KSCF and state-of-the-art methods in terms of precision and success rate with respect to each challenging attributes. Fig. 9 shows the statistics of precision and success rate of the leading trackers (i.e., SKSCF, KSCF, MEEM, KCF and TGPR) with respect to the attributes. We note that MEEM [42] adopts the multiple experts framework to deal with model drift, and performs slightly better than KSCF for attributes FM, LR, OV and SV. Overall, the KSCF algorithm are among the top 3 trackers for any attribute, and the SKSCF algorithm performs best in both metrics for all but one attribute.

5 Conclusions

We propose an effective and efficient approach to learn support correlation filters for real-time visual tracking. By reformulating the SVM model with circulant data matrix as training input, we present a DFT based alternating optimization algorithm to learn support correlation filters efficiently. In addition, we develop the MSCF, KSCF, and SKSCF trackers to exploit multidimensional features, kernelized classifiers, and scale-adaptive schemes. Experiments on a large benchmark dataset show that the proposed KSCF and SKSCF algorithms perform favorably against the state-of-the-art tracking methods in terms of accuracy and speed. Our future work includes developing optimization algorithms to solve SVM with hinge loss, evaluating the effect of classification models on SCF-based tracking, and applying SVMs with circulant training data matrix to other vision tasks such as object detection and localization.

TABLE 7
Precision (Top) and Success Rate (Bottom) of the Evaluated Trackers (Top Three Are Shown in Red, Blue and Orange)

	Attributes	FM	BC	MB	DEF	IV	IPR	LR	OCC	OPR	OV	SV
Precision	SKSCF	0.779	0.859	0.802	0.893	0.841	0.810	0.596	0.872	0.857	0.800	0.809
	KSCF	0.680	0.825	0.761	0.854	0.805	0.816	0.555	0.852	0.836	0.697	0.768
	MEEM [42]	0.745	0.802	0.721	0.856	0.771	0.796	0.529	0.801	0.840	0.726	0.795
	TGPR [15]	0.579	0.763	0.570	0.760	0.695	0.683	0.567	0.668	0.693	0.535	0.637
	KCF [24]	0.564	0.752	0.599	0.747	0.687	0.692	0.379	0.735	0.718	0.589	0.680
	SCM [45]	0.346	0.578	0.358	0.589	0.613	0.613	0.305	0.646	0.621	0.429	0.672
	TLD [28]	0.557	0.428	0.523	0.495	0.540	0.588	0.349	0.556	0.593	0.576	0.606
	ASLA [27]	0.255	0.496	0.283	0.473	0.529	0.521	0.156	0.479	0.535	0.333	0.552
	L1APG [5]	0.367	0.425	0.379	0.398	0.341	0.524	0.460	0.475	0.490	0.329	0.472
	MIL [3]	0.415	0.456	0.381	0.493	0.359	0.465	0.171	0.448	0.484	0.393	0.471
	CT [44]	0.330	0.339	0.314	0.463	0.365	0.361	0.152	0.429	0.405	0.336	0.448
	Attributes	FM	BC	MB	DEF	IV	IPR	LR	OCC	OPR	OV	SV
Success	SKSCF	0.729	0.795	0.757	0.863	0.743	0.720	0.542	0.788	0.757	0.808	0.682
	KSCF	0.629	0.741	0.689	0.779	0.649	0.690	0.389	0.696	0.697	0.705	0.540
	MEEM [42]	0.706	0.747	0.692	0.711	0.653	0.648	0.470	0.694	0.694	0.742	0.594
	TGPR [15]	0.542	0.713	0.570	0.711	0.632	0.601	0.501	0.592	0.603	0.546	0.505
	KCF [24]	0.516	0.669	0.539	0.668	0.534	0.575	0.358	0.593	0.587	0.589	0.477
	SCM [45]	0.348	0.550	0.358	0.566	0.586	0.574	0.308	0.602	0.576	0.449	0.635
	TLD [28]	0.475	0.388	0.485	0.434	0.461	0.477	0.327	0.455	0.489	0.516	0.494
	ASLA [27]	0.261	0.468	0.284	0.485	0.514	0.496	0.163	0.469	0.509	0.359	0.544
	L1APG [5]	0.359	0.404	0.363	0.398	0.298	0.445	0.458	0.437	0.423	0.341	0.407
	MIL [3]	0.353	0.414	0.261	0.440	0.300	0.339	0.157	0.378	0.369	0.416	0.335
	CT [44]	0.327	0.323	0.262	0.420	0.308	0.290	0.143	0.360	0.325	0.405	0.342

Appendix A
Proof

A.1 Solution to the $\{\mathbf{w}, b\}$ Subproblem

Proof. The subproblem on $\{\mathbf{w}, b\}$ can be rewritten as

$$\min_{\mathbf{w},b} \left\{ f(\mathbf{w}, b) = \|\mathbf{w}\|^2 + C\|\mathbf{X}^\top\mathbf{w} + b\mathbf{1} - \mathbf{q}\|_2^2 \right\}. \tag{33}$$

The optimal solution of b satisfies the following condition

$$\begin{aligned} \frac{\partial f(\mathbf{w}, b)}{\partial b} &= 2C\big(\mathbf{1}^\top\mathbf{X}^\top\mathbf{w} + bn^2 - \mathbf{1}^\top\mathbf{q}\big) = 0 \\ \Rightarrow b &= \frac{1}{n^2}\big(\mathbf{1}^\top\mathbf{q} - \mathbf{1}^\top\mathbf{X}^\top\mathbf{w}\big). \end{aligned} \tag{34}$$

Let $\mathbf{X} = \mathbf{X}_c + \bar{x}\mathbf{1}\mathbf{1}^\top$. Note that $\mathbf{X}_c$ is a centralized matrix and has $\mathbf{X}_c^\top\mathbf{1} = \mathbf{0}$ as well as $\mathbf{X}_c\mathbf{1} = \mathbf{0}$. Thus we have $\mathbf{1}^\top\mathbf{X}^\top\mathbf{w} = \mathbf{1}^\top\mathbf{X}_c^\top\mathbf{w} + \bar{x}\mathbf{1}^\top\mathbf{1}\mathbf{1}^\top\mathbf{w} = n^2\bar{x}\mathbf{1}^\top\mathbf{w}$ and

$$b = \bar{q} - \bar{x}\mathbf{1}^\top\mathbf{w}. \tag{35}$$

By substituting $\mathbf{X}^\top\mathbf{w} = \mathbf{X}_c^\top\mathbf{w} + \bar{x}\mathbf{1}\mathbf{1}^\top\mathbf{w}$ and (35) into (33), we have

$$\min_{\mathbf{w}} \|\mathbf{w}\|^2 + C\|\mathbf{X}_c^\top\mathbf{w} - \mathbf{q}_c\|_2^2. \tag{36}$$

Similar to (34), the optimal solution of $\mathbf{w}$ satisfies the following condition,

$$\begin{aligned} \frac{\partial f(\mathbf{w}, b)}{\partial \mathbf{w}} &= 2\mathbf{w} + 2C\mathbf{X}_c\big(\mathbf{X}_c^\top\mathbf{w} - \mathbf{q}_c\big) = \mathbf{0} \\ \Rightarrow \mathbf{w} &= \left(\mathbf{X}_c\mathbf{X}_c^\top + \frac{1}{C}\mathbf{I}\right)^{-1}\mathbf{X}_c\mathbf{q}_c. \end{aligned} \tag{37}$$

Since $\mathbf{X}_c$ is a circulant matrix, we have

$$\hat{\mathbf{w}} = \frac{\hat{\mathbf{x}}_c^* \circ \hat{\mathbf{q}}_c}{\hat{\mathbf{x}}_c^* \circ \hat{\mathbf{x}}_c + \frac{1}{C}}. \tag{38}$$

From (37), we have

$$\mathbf{1}^\top\mathbf{w} = \mathbf{1}^\top\mathbf{X}_c\mathbf{q}_c - C\mathbf{1}^\top\mathbf{X}_c\mathbf{X}_c^\top\mathbf{w}, \tag{39}$$

As $\mathbf{1}^\top\mathbf{X}_c = \mathbf{0}^\top$, we have $\mathbf{1}^\top\mathbf{w} = 0$. The closed-form solution to b can be rewritten as

$$b = \bar{q}. \tag{40}$$

□

A.2 Solution to the $\{\alpha, b\}$ Subproblem

Proof. Let $g(\boldsymbol{\alpha}, b) = \boldsymbol{\alpha}^\top\mathbf{K}\boldsymbol{\alpha} + C\|\mathbf{K}\boldsymbol{\alpha} + b\mathbf{1} - \mathbf{q}\|_2^2$. The optimal solution to b satisfies the following condition:

$$\begin{aligned} \frac{\partial g(\boldsymbol{\alpha}, b)}{\partial b} &= 2C\big(\mathbf{1}^\top\mathbf{K}\boldsymbol{\alpha} + bn^2 - \mathbf{1}^\top\mathbf{q}\big) = 0 \\ \Rightarrow b &= \bar{q} - \frac{1}{n^2}\mathbf{1}^\top\mathbf{K}\boldsymbol{\alpha}. \end{aligned} \tag{41}$$

Note that $\mathbf{K} = \mathbf{K}_c + \bar{k}\mathbf{1}\mathbf{1}^\top$ and $\mathbf{1}^\top\mathbf{K}_c = \mathbf{0}^\top$. We have

$$b = \bar{q} - \bar{k}\mathbf{1}^\top\alpha = \bar{q} - n^2\bar{k}\bar{\alpha}. \tag{42}$$

By substituting (31) and (42) into (30), we have

$$\min_{\boldsymbol{\alpha}} \boldsymbol{\alpha}^\top\mathbf{K}_c\boldsymbol{\alpha} + \bar{k}\boldsymbol{\alpha}^\top\mathbf{1}\mathbf{1}^\top\boldsymbol{\alpha} + C\|\mathbf{K}_c\boldsymbol{\alpha} - \mathbf{q}_c\|_2^2. \tag{43}$$

Let $\bar{\alpha}$ be the mean of $\boldsymbol{\alpha}$. We have $\boldsymbol{\alpha}_c = \boldsymbol{\alpha} - \bar{\alpha}\mathbf{1}$ and (43) becomes

$$\min_{\boldsymbol{\alpha}_c,\bar{\alpha}} \boldsymbol{\alpha}_c^\top\mathbf{K}_c\boldsymbol{\alpha}_c + n^4\bar{k}\bar{\alpha}^2 + C\|\mathbf{K}_c\boldsymbol{\alpha}_c - \mathbf{q}_c\|_2^2. \tag{44}$$

From (44), the optimal solution to $\bar{\alpha}$ should be $\bar{\alpha} = 0$. Based on kernel ridge regression [33], the optimal solution to $\boldsymbol{\alpha}_c$ (i.e., $\boldsymbol{\alpha}$) can be obtained by

$$\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}_c = \frac{\hat{\mathbf{q}}_c}{\hat{\mathbf{k}}_c^{\mathbf{xx}} + \frac{1}{C}}. \tag{45}$$

By substituting $\bar{\alpha} = 0$ into (42), the solution to b becomes,

$$b = \bar{q}. \tag{46}$$

□

Appendix B
Convergence Analysis

B.1 Optimality Conditions

In the spatial domain, the SCF model can be expressed as:

$$\begin{aligned} (\mathbf{w}, b, \mathbf{e}) &= \arg\min_{\mathbf{w},b,\mathbf{e}} \|\mathbf{w}\|^2 + C\|\mathbf{y}\circ(\mathbf{X}^\top\mathbf{w} + b\mathbf{1}) - \mathbf{1} - \mathbf{e}\|_2^2, \\ &\text{s.t. } \mathbf{e} \geq 0 \end{aligned} \tag{47}$$

Defining the augmented vector $\tilde{\mathbf{x}} = [\mathbf{x}^\top, 1]^\top$ with $\mathbf{x} \in R^n$, we compute the augmented weight vector $\tilde{\mathbf{w}} = [\mathbf{w}^\top, b]^\top$. The above problem can then be reformulated as:

$$\begin{aligned} (\tilde{\mathbf{w}}, \mathbf{e}) &= \arg\min_{\tilde{\mathbf{w}},e} \tilde{\mathbf{w}}^\top\tilde{\mathbf{I}}\tilde{w} + C\|\tilde{\mathbf{X}}^\top\tilde{\mathbf{w}} - \mathbf{y} - \mathbf{y}\circ\mathbf{e}\|_2^2, \\ &\text{s.t. } \mathbf{e} \geq 0 \end{aligned} \tag{48}$$

where $\tilde{\mathbf{X}} = \left[\mathbf{X}^\top, \mathbf{1}\right]^\top$ and $\tilde{\mathbf{I}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{bmatrix}$. We introduce an indicator function $\delta(e_i) = \begin{cases} +\infty \ , & \text{if } e_i < 0 \\ 0 \ , & \text{if } e_i \geq 0 \end{cases}$ and the subdifferential [34] of $\delta(e_i)$ is:

$$\partial\delta(e_i) = \begin{cases} 0, & \text{if } e_i > 0 \\ (-\infty, 0), & \text{if } e_i = 0 \\ \phi(\text{undefined}), & \text{if } e_i < 0. \end{cases} \tag{49}$$

As the loss function (48) is convex, $(\tilde{\mathbf{w}}^*, \mathbf{e}^*)$ is a solution if and only if the subdifferential of the loss at $(\tilde{\mathbf{w}}^*, \mathbf{e}^*)$ contains zero [10]. Thus, the optimality conditions are:

$$\begin{aligned} &\tilde{\mathbf{I}}\tilde{w} + C\tilde{\mathbf{X}}(\tilde{\mathbf{X}}^\top\tilde{\mathbf{w}} - \mathbf{y} - \mathbf{y}\circ\mathbf{e}) = 0 \\ &\begin{cases} e_i + 1 - y_i\tilde{\mathbf{x}}_i^\top\tilde{\mathbf{w}} = 0, & \text{if } e_i > 0, \\ y_i\tilde{\mathbf{x}}_i^\top\tilde{\mathbf{w}} - 1 \leq 0, & \text{if } e_i = 0, \end{cases} \end{aligned} \tag{50}$$

where $\tilde{\mathbf{x}}_i$ is the ith training sample. With $\lambda = \frac{1}{C}$, we have:

$$\begin{aligned} \det\left(\lambda\tilde{\mathbf{I}} + \tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top\right) &= \det\left(\begin{bmatrix} \lambda\mathbf{I} + \mathbf{X}\mathbf{X}^\top & \sum_i \mathbf{x}_i \\ \sum_i \mathbf{x}_i^\top & n^2 \end{bmatrix}\right) \\ &= n^2\det\left(\mathbf{X}\mathbf{X}^\top + \lambda\mathbf{I} - \frac{1}{n^2}\sum_i \mathbf{x}_i \sum_i \mathbf{x}_i^\top\right) \\ &= n^2\det\big(\mathbf{X}_c\mathbf{X}_c^\top + \lambda\mathbf{I}\big), \end{aligned} \tag{51}$$

where $\mathbf{X}_c = [\mathbf{x}_1 - \bar{\mathbf{x}}, \ldots, \mathbf{x}_n - \bar{\mathbf{x}}]$ with $\bar{\mathbf{x}} = \frac{1}{n}\sum_i \mathbf{x}_i$. Thus the matrix $(\lambda\tilde{\mathbf{I}} + \tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top)$ is invertible. For simplicity, let $\mathbf{M} = \tilde{\mathbf{I}} + C\tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top$, from (50) and above equation, we have

$$\tilde{\mathbf{w}} = C\mathbf{M}^{-1}\tilde{\mathbf{X}}(\mathbf{y} + \mathbf{y}\circ\mathbf{e}), \tag{52}$$

$$(C\mathbf{y}\circ\mathbf{M}^{-1}\tilde{\mathbf{X}}(\mathbf{y} + \mathbf{y}\circ\mathbf{e}) - 1)_i \begin{cases} = e_i, & \text{if } e_i > 0, \\ < 0, & \text{if } e_i = 0, \end{cases} \tag{53}$$

Based on the optimality conditions in (50), we define

$$\begin{cases} \mathbf{r}_1 = \tilde{\mathbf{I}}\tilde{w} + C\tilde{\mathbf{X}}(\tilde{\mathbf{X}}^\top\tilde{\mathbf{w}} - \mathbf{y} - \mathbf{y}\circ\mathbf{e}), \\ \mathbf{r}_2(i) = e_i + 1 - y_i\tilde{\mathbf{x}}_i^\top\tilde{\mathbf{w}} \ \ \forall e_i > 0, \\ \mathbf{r}_3(i) = y_i\tilde{\mathbf{x}}_i^\top\tilde{\mathbf{w}} - 1 \ \ \forall e_i \leq 0, \end{cases} \tag{54}$$

and use the stopping criterion:

$$\max\left\{\|\mathbf{r}_1\|_\infty, \max_{e_i > 0}\|\mathbf{r}_2(i)\|, \max_{e_i = 0}\|\mathbf{r}_3(i)\|\right\} \leq \epsilon, \tag{55}$$

where $\epsilon > 0$ is a predefined threshold.

B.2 Global Convergence

To compute $\mathbf{e}$, we reformulate the subproblem for each entry:

$$\hat{z} = \arg\min_z \frac{1}{2}\|z - z_0\|^2 + \delta(z), \tag{56}$$

where $\delta(z) = \begin{cases} +\infty, & \text{if } z < 0 \\ 0, & \text{if } z \geq 0 \end{cases}$. The solution is given by:

$$\hat{z} = g(z_0) = \begin{cases} z_0, & \text{if } z_0 \geq 0, \\ 0, & \text{if } z_0 < 0, \end{cases} \tag{57}$$

Proposition 1. *For any $a, b \in R$, we have:*

$$\|g(a) - g(b)\|^2 \leq \|a - b\|^2, \tag{58}$$

where the equality holds only if $g(a) - g(b) = a - b$.

Proof.

(1) if $a, b \geq 0$, $\|g(a) - g(b)\|^2 = \|a - b\|^2$, and we also have $g(a) - g(b) = a - b$.

(2) if $a, b < 0$, $\|g(a) - g(b)\|^2 = 0 \leq \|a - b\|^2$, where the equality holds only if $a = b$.

(3) if $ab \leq 0$, e.g., $b \leq 0$, it is easy to see that, $a^2 \leq (|a| + |b|)^2$.

□

For simplicity, let $\mathbf{U} = \tilde{\mathbf{X}}\text{diag}(\mathbf{y})$. We have $\mathbf{U}\mathbf{U}^\top = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top$ and then we get two symmetric positive definite matrices as follows:

$$\mathbf{M} = \tilde{\mathbf{I}} + C\tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top = \tilde{\mathbf{I}} + C\mathbf{U}\mathbf{U}^\top, \tag{59}$$

$$\mathbf{T} = C\mathbf{U}^\top(\tilde{\mathbf{I}} + C\mathbf{U}\mathbf{U}^\top)^{-1}\mathbf{U} = C\mathbf{U}^\top\mathbf{M}^{-1}\mathbf{U}, \tag{60}$$

where $\rho(\mathbf{T}) \leq 1$ and $\rho(\mathbf{T})$ is the spectral radius of matrix $\mathbf{T}$ [26]. With the definitions of $\mathbf{M}$ and $\mathbf{T}$, the updating rules $\tilde{\mathbf{w}}$ and $\mathbf{e}$ can be written as:

$$\mathbf{e}^{k+1} = g(\mathbf{U}^\top\tilde{\mathbf{w}}^k - 1) = g(\mathbf{T}(\mathbf{1} + \mathbf{e}^k) - 1) = g\circ h(\mathbf{e}^k), \tag{61}$$

$$\tilde{\mathbf{w}}^{k+1} = C\mathbf{M}^{-1}\mathbf{U}(\mathbf{1} + \mathbf{e}^{k+1}), \tag{62}$$

Let $h(\mathbf{e}^k) = \mathbf{T}(\mathbf{1} + \mathbf{e}^k) - 1$, we have the following proposition.

Proposition 2. *For any $\mathbf{e} \neq \hat{\mathbf{e}}$, the following inequality holds:*

$$\|h(\mathbf{e}) - h(\hat{\mathbf{e}})\| \leq \|\mathbf{e} - \hat{\mathbf{e}}\|, \tag{63}$$

and the equality holds if and only if $h(\mathbf{e}) - h(\hat{\mathbf{e}}) = \mathbf{e} - \hat{\mathbf{e}}$.

Proof. Note that $\rho(\mathbf{T}) \leq 1$. From the definition of $h(\mathbf{e})$, we have:

$$\|h(\mathbf{e}) - h(\hat{\mathbf{e}})\| = \|\mathbf{T}(\mathbf{e} - \hat{\mathbf{e}})\| \leq \rho(\mathbf{T})\|\mathbf{e} - \hat{\mathbf{e}}\| < \|\mathbf{e} - \hat{\mathbf{e}}\|, \tag{64}$$

Denote the eigen-decomposition of $\mathbf{T}$ by $\mathbf{T} = \mathbf{Q}^\top\Lambda\mathbf{Q}$, where $\mathbf{Q}$ is a full rank orthogonal matrix, and Λ is a diagonal matrix with $0 \leq \lambda_i \leq 1$. The equality $\|h(\mathbf{e}) - h(\hat{\mathbf{e}})\| = \|\mathbf{e} - \hat{\mathbf{e}}\|$ can be written as $\|\mathbf{Q}^\top\Lambda\mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}})\| = \|\mathbf{e} - \hat{\mathbf{e}}\|$. Since $\mathbf{Q}$ is full-rank orthogonal, there is $\|\mathbf{e} - \hat{\mathbf{e}}\| = \|\mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}})\|$. Thus, we have $\|\Lambda\mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}})\| = \|\mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}})\|$.

Let $\Lambda = \text{diag}([\lambda_1; \ldots; \lambda_{n^2+1}])$ and $\mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}}) = [a_1; \ldots; a_{n^2+1}]$. We have $\|\Lambda\mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}})\|^2 = \sum_i \lambda_i^2 a_i^2$ and $\|\mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}})\|^2 = \sum_i a_i^2$. As $0 \leq \lambda_i \leq 1$, we have $\lambda_i^2 a_i^2 \leq a_i^2$. The equality $\sum_i \lambda_i^2 a_i^2 = \sum_i a_i^2$ holds only if we have $\lambda_i^2 a_i^2 = a_i^2$ for any i. That is, for any i, it is required that $\lambda_i = 1$ or $a_i = 0$, and it is equivalent to $\lambda_i a_i = a_i$. Thus, we have $\Lambda\mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}}) = \mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}})$. Multiplying both side $\mathbf{Q}^\top$, we have $\mathbf{T}(\mathbf{e} - \hat{\mathbf{e}}) = h(\mathbf{e}) - h(\hat{\mathbf{e}}) = \mathbf{e} - \hat{\mathbf{e}}$. □

Definition 1 (Fixed point [17]). *Given a linear operator, a point x^* is a fixed point if $x^* = f(x^*)$. We next provide the following property for fixed points of the operator $g\circ h$.*

Lemma 3. *Given any fixed point $\hat{\mathbf{e}}$ of $g\circ h$, for any $\mathbf{e}$, we have:*

$$\|g\circ h(\mathbf{e}) - g\circ h(\hat{\mathbf{e}})\| < \|\mathbf{e} - \hat{\mathbf{e}}\|, \tag{65}$$

unless $\mathbf{e}$ is a fixed point of $g\circ h$.

Proof. From Propositions 1 and 2, it holds:

$$\|g\circ h(\mathbf{e}) - g\circ h(\hat{\mathbf{e}})\| < \|h(\mathbf{e}) - h(\hat{\mathbf{e}})\| < \|\mathbf{e} - \hat{\mathbf{e}}\|, \tag{66}$$

unless $g\circ h(\mathbf{e}) - g\circ h(\hat{\mathbf{e}}) = h(\mathbf{e}) - h(\hat{\mathbf{e}}) = \mathbf{e} - \hat{\mathbf{e}}$. Thus if $g\circ h(\hat{\mathbf{e}}) = \hat{\mathbf{e}}$, we have $g\circ h(\mathbf{e}) = \mathbf{e}$. □

Theorem 1 (Global convergence). *The sequence $\{(\tilde{\mathbf{w}}^k, \mathbf{e}^k)\}$ generated by our algorithm from any starting point $(\tilde{\mathbf{w}}^0, \mathbf{e}^0)$ converges to a solution $(\tilde{\mathbf{w}}^*, \mathbf{e}^*)$ of the optimization problem.*

Proof. First we prove that $\mathbf{e}^k$ converges to a fixed point. Note that $g\circ h$ is non-expansive, thus the sequence $\{\mathbf{e}^k\}$ lies in a compact region and $\mathbf{e}^k$ converges to one limit point $\mathbf{e}^*$ at least. We assume $\mathbf{e}^* = \lim_{j\to\infty}\mathbf{e}^{k_j}$ and let $\hat{\mathbf{e}}$ be any fixed point of $g\circ h$ with $\hat{\mathbf{e}} = g\circ h(\hat{\mathbf{e}})$. Then the following formula is established:

$$\|\mathbf{e}^k - \hat{\mathbf{e}}\| = \|g\circ h(\mathbf{e}^{k-1}) - g\circ h(\hat{\mathbf{e}})\| \leq \|\mathbf{e}^{k-1} - \hat{\mathbf{e}}\|, \tag{67}$$

Based on above, we get the limit as below:

$$\lim_{k\to\infty}\left\|\mathbf{e}^k-\hat{\mathbf{e}}\right\|=\lim_{j\to\infty}\left\|\mathbf{e}^{k_j}-\hat{\mathbf{e}}\right\|=\|\mathbf{e}^*-\hat{\mathbf{e}}\|, \tag{68}$$

which shows that more than one of all limit points of $\{\mathbf{e}^k\}$ have an equal distance to $\hat{\mathbf{e}}$. Because of the continuity of $g\circ h$, we have:

$$g\circ h(\mathbf{e}^*)=\lim_{j\to\infty} g\circ h(\mathbf{e}^{k_j})=\lim_{j\to\infty}\mathbf{e}^{k_j+1}. \tag{69}$$

Thus, $g\circ h(\mathbf{e}^*)$ is also a limit point of sequence $\{\mathbf{e}^k\}$ and it must have an equal distance to $\hat{\mathbf{e}}$:

$$\|\mathbf{e}^*-\hat{\mathbf{e}}\|=\|g\circ h(\mathbf{e}^*)-\hat{\mathbf{e}}\|=\|g\circ h(\mathbf{e}^*)-g\circ h(\hat{\mathbf{e}})\|. \tag{70}$$

According to Lemma 3, we know $g\circ h(\mathbf{e}^*)=\mathbf{e}^*$. Since $\hat{\mathbf{e}}$ is any fixed point of $g\circ h$, with the continuity of $g\circ h(\mathbf{e}^*)$, the convergence: $\lim\limits_{k\to\infty}\mathbf{e}^k=\mathbf{e}^*$ is obtained. We next show that $\mathbf{e}^*$ satisfies the optimization condition in (53). With the definition of $\mathbf{T}$, g and h, we have:

$$\begin{aligned}&g\circ h(\mathbf{e})=g(\mathbf{T}(\mathbf{1}+\mathbf{e})-1)\\&=g(C\mathbf{U}^\top(\tilde{\mathbf{I}}+C\mathbf{U}\mathbf{U}^\top)^{-1}\mathbf{U}(\mathbf{1}+\mathbf{e})-1)\begin{cases}=e_i,\text{if } e_i>0\\<0,\ \text{if } e_i=0,\end{cases}\end{aligned} \tag{71}$$

which can be written as $\mathbf{e}=g\circ h(\mathbf{e})$. Considering $g\circ h(\mathbf{e}^*)=\mathbf{e}^*$, the solution $\mathbf{e}^*$ satisfies the optimization conditions and the proposed algorithm converges to the global optimum. □

B.3 Q-Linear Convergence Rate

Theorem 2 (Convergence rate). *The sequence* $\{(\tilde{\mathbf{w}}^k,\mathbf{e}^k)\}$ *generated by our algorithm satisfies the following three conditions:*

$$\begin{aligned}&(1)\ \left\|\mathbf{e}^{k+1}-\mathbf{e}^*\right\|\le\sqrt{\rho(\mathbf{T}^2)}\left\|\mathbf{e}^k-\mathbf{e}^*\right\|,\\&(2)\ \left\|\mathbf{U}^\top(\tilde{\mathbf{w}}^{k+1}-\tilde{\mathbf{w}}^*)\right\|\le\sqrt{\rho(\mathbf{T}^2)}\left\|\mathbf{U}^\top(\tilde{\mathbf{w}}^k-\tilde{\mathbf{w}}^*)\right\|,\\&(3)\ \left\|\tilde{\mathbf{w}}^{k+1}-\tilde{\mathbf{w}}^*\right\|_M\le\sqrt{\rho(\mathbf{T})}\left\|\tilde{\mathbf{w}}^k-\tilde{\mathbf{w}}^*\right\|_M.\end{aligned}$$

Proof. Note that $g\circ h$ is non-expansive, according to Proposition 1, we have:

$$\tilde{\mathbf{w}}^{k+1}-\tilde{\mathbf{w}}^*=C\mathbf{M}^{-1}\mathbf{U}\left(\mathbf{e}^{k+1}-\mathbf{e}^*\right), \tag{72}$$

$$\left\|\mathbf{e}^{k+1}-\mathbf{e}^*\right\|^2=\left\|g\circ h(\mathbf{e}^k)-g\circ h(\mathbf{e}^*)\right\|^2\le\left\|\mathbf{U}^\top(\tilde{\mathbf{w}}^k-\tilde{\mathbf{w}}^*)\right\|^2 \tag{73}$$

Under the definition of $\mathbf{T}$, there is: $\left\|\mathbf{U}^\top(\tilde{\mathbf{w}}^k-\tilde{\mathbf{w}}^*)\right\|^2=\left\|\mathbf{T}(\mathbf{e}^k-\mathbf{e}^*)\right\|^2$, and thus

$$\left\|\mathbf{e}^{k+1}-\mathbf{e}^*\right\|^2\le\left\|\mathbf{T}(\mathbf{e}^k-\mathbf{e}^*)\right\|^2, \tag{74}$$

Consequently, we have:

$$\left\|\mathbf{e}^{k+1}-\mathbf{e}^*\right\|^2\le\left(\mathbf{e}^k-\mathbf{e}^*\right)^\top(\mathbf{T}^2)\left(\mathbf{e}^k-\mathbf{e}^*\right)\le\rho(\mathbf{T}^2)\left\|\mathbf{e}^k-\mathbf{e}^*\right\|^2. \tag{75}$$

By reformulating above, condition 1 can be satisfied:

$$\left\|\mathbf{e}^{k+1}-\mathbf{e}^*\right\|\le\sqrt{\rho(\mathbf{T}^2)}\left\|\mathbf{e}^k-\mathbf{e}^*\right\|. \tag{76}$$

Multiplying $\tilde{\mathbf{X}}^\top$ on both sides of (72), and combining with (73), we obtain:

$$\begin{aligned}\left\|\mathbf{U}^\top\left(\tilde{\mathbf{w}}^{k+1}-\tilde{\mathbf{w}}^*\right)\right\|^2=\left\|\mathbf{T}(\mathbf{e}^{k+1}-\mathbf{e}^*)\right\|^2&\le\rho(\mathbf{T}^2)\left\|\mathbf{e}^{k+1}-\mathbf{e}^*\right\|^2\\&\le\rho(\mathbf{T}^2)\left\|\mathbf{U}^\top\left(\tilde{\mathbf{w}}^k-\tilde{\mathbf{w}}^*\right)\right\|^2,\end{aligned} \tag{77}$$

which can be reformulated as:

$$\left\|\mathbf{U}^\top(\tilde{\mathbf{w}}^{k+1}-\tilde{\mathbf{w}}^*)\right\|\le\sqrt{\rho(\mathbf{T}^2)}\left\|\mathbf{U}^\top(\tilde{\mathbf{w}}^k-\tilde{\mathbf{w}}^*)\right\|, \tag{78}$$

and satisfies condition 2. From (72), we have:

$$\left\|\tilde{\mathbf{w}}^{k+1}-\tilde{\mathbf{w}}^*\right\|_M^2=\left(\mathbf{e}^{k+1}-\mathbf{e}^*\right)^\top\mathbf{T}\left(\mathbf{e}^{k+1}-\mathbf{e}^*\right)\le\rho(\mathbf{T})\left\|\mathbf{e}^{k+1}-\mathbf{e}^*\right\|^2. \tag{79}$$

Combining (73) and the definition of $\mathbf{M}$, we have:

$$\left\|\tilde{\mathbf{w}}^{k+1}-\tilde{\mathbf{w}}^*\right\|_M\le\sqrt{\rho(\mathbf{T})}\left\|\mathbf{U}^\top(\tilde{\mathbf{w}}^k-\tilde{\mathbf{w}}^*)\right\|\le\sqrt{\rho(\mathbf{T})}\left\|\tilde{\mathbf{w}}^k-\tilde{\mathbf{w}}^*\right\|_M. \tag{80}$$

Thus, condition 3 holds and $\tilde{\mathbf{w}}^k$ converges to $\tilde{\mathbf{w}}^*$ q-linearly [1]. □

Acknowledgments

This work is supported in part by the National Defense Science and Technology Innovation Special Zone Project of China Grant (17-163-11-ZT-003-024-01), NSFC Grant (61671182), NSF CAREER Grant (No.1149783), NSF IIS Grant (No. 1152576), and HK RGC GRF Grant (PolyU 152240/15E).

References

[1] M. Allain, J. Idier, and Y. Goussard, "On global and local convergence of half-quadratic algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1130–1142, May 2006.

[2] S. Avidan, "Support vector tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 1064–1072, Aug. 2004.

[3] B. Babenko, M.-H. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 983–990.

[4] Y. Bai and M. Tang, "Robust tracking via weakly supervised ranking svm," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1854–1861.

[5] C. Bao, Y. Wu, H. Ling, and H. Ji, "Real time robust l1 tracker using accelerated proximal gradient approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1830–1837.

[6] V. N. Boddeti, T. Kanade, and B. V. Kumar, "Correlation filters for object alignment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2291–2298.

[7] V. N. Boddeti and B. V. Kumar, "Maximum margin vector correlation filter," CoRR abs/1404.6031, 2014.

[8] D. S. Bolme, J. R. Beveridge, B. Draper, Y. M. Lui, et al., "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2544–2550.

[9] D. S. Bolme, B. Draper, J. R. Beveridge, et al., "Average of synthetic exact filters," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 2105–2112.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 1, pp. 886–893.

[12] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proc. Brit. Mach. Vis. Conf.*, 2014, pp. 1–11.

[13] M. Danelljan, F. S. Khan, M. Felsberg, and J. van de Weijer, "Adaptive color attributes for real-time visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1090–1097.

[14] H. K. Galoogahi, T. Sim, and S. Lucey, "Multi-channel correlation filters," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 3072–3079.

[15] J. Gao, H. Ling, W. Hu, and J. Xing, "Transfer learning based visual tracking with gaussian processes regression," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 188–203.

[16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 580–587.

[17] K. Goebel and W. Kirk, "A fixed point theorem for asymptotically nonexpansive mappings," *Proc. Amer. Math. Soc.*, vol. 35, no. 1, pp. 171–174, 1972.

[18] H. Grabner, M. Grabner, and H. Bischof, "Real-time tracking via on-line boosting," in *Proc. Brit. Mach. Vis. Conf.*, 2006, vol. 1, pp. 47–56.

[19] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 234–247.

[20] R. M. Gray, *Toeplitz and Circulant Matrices: A Review*. Breda, Netherlands: Now publishers Inc., 2006.

[21] S. Hare, A. Saffari, and P. H. Torr, "Struck: Structured output tracking with kernels," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 263–270.

[22] J. F. Henriques, J. Carreira, R. Caseiro, and J. Batista, "Beyond hard negative mining: Efficient detector learning via block-circulant decomposition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2760–2767.

[23] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 702–715.

[24] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.

[25] J. Ho, K.-C. Lee, M.-H. Yang, and D. Kriegman, "Visual tracking using learned linear subspaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, pp. I-782–I-789, 2004.

[26] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[27] X. Jia, H. Lu, and M.-H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1822–1829.

[28] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1409–1422, Jul. 2012.

[29] M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, L. Čehovin, G. Nebehay, T. Vojíř, G. Fernandez, A. Lukežič, A. Dimitriev, et al., "The visual object tracking vot2014 challenge results," in *Proc. Workshop Eur. Conf. Comput. Vis.*, 2014, pp. 191–217.

[30] C.-P. Lee and C.-B. Lin, "A study on l2-loss (squared hinge-loss) multiclass SVM," *Neural Comput.*, vol. 25, no. 5, pp. 1302–1323, 2013.

[31] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Proc. Workshop Eur. Conf. Comput. Vis.*, 2014, pp. 254–265.

[32] R. Patnaik and D. Casasent, "Fast FFT-based distortion-invariant kernel filters for general object recognition," in *Proc. IS&T/SPIE Electron. Imag.*, 2009, pp. 725202–725202.

[33] R. Rifkin, G. Yeo, and T. Poggio, "Regularized least-squares classification," *NATO Sci. Series Sub Series III Comput. Syst. Sci.*, vol. 190, pp. 131–154, 2003.

[34] R. Rockafellar, "On the maximal monotonicity of subdifferential mappings," *Pacific J. Math.*, vol. 33, no. 1, pp. 209–216, 1970.

[35] A. Rodriguez, V. N. Boddeti, B. V. Kumar, and A. Mahalanobis, "Maximum margin correlation filter: A new approach for localization and classification," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 631–643, Feb. 2013.

[36] A. Saffari, C. Leistner, J. Santner, M. Godec, and H. Bischof, "On-line random forests," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop*, 2009, pp. 1393–1400.

[37] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.

[38] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1442–1468, Jul. 2014.

[39] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, 1999.

[40] D. Wang, H. Lu, Z. Xiao, and M.-H. Yang, "Inverse sparse tracker with a locally weighted distance metric," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2646–2657, Sep. 2015.

[41] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2411–2418.

[42] J. Zhang, S. Ma, and S. Sclaroff, "Meem: Robust tracking via multiple experts using entropy minimization," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 188–203.

[43] K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M.-H. Yang, "Fast visual tracking via dense spatio-temporal context learning," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 127–141.

[44] K. Zhang, L. Zhang, and M.-H. Yang, "Real-time compressive tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 864–877.

[45] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparsity-based collaborative model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1838–1845.

Wangmeng Zuo (M'09, SM'15) received the PhD degree in computer application technology from the Harbin Institute of Technology, Harbin, China, in 2007. He is currently a professor with the School of Computer Science and Technology, Harbin Institute of Technology. His current research interests include image enhancement and restoration, image and face editing, object detection, visual tracking, and image classification. He has published more than 70 papers in top-tier academic journals and conferences. He has served as a Tutorial Organizer in ECCV 2016, an associate editor of the *IET Biometrics* and *Journal of Electronic Imaging*. He is a senior member of the IEEE.

Xiaohe Wu received the BE degree from the Harbin Institute of Technology (HIT), Harbin, China, in 2013. She is currently working toward the PhD degree in the School of Computer Science and Technology, HIT. In 2014, she was a research assistant with the Department of Computing, the Hong Kong Polytechnic University, Hong Kong. In 2017, she was as a joint-training PhD in University of California at Merced. Her current research interests include object visual tracking, support vector machines and related problems.

Liang Lin (M'09, SM'15) is the executive R&D director of SenseTime Group Limited and a full professor of Sun Yat-sen University. He currently leads the SenseTime R&D teams to develop cutting-edges and deliverable solutions on computer vision, data analysis and mining, and intelligent robotic systems. He has authored and coauthored on more than 100 papers in top-tier academic journals and conferences. He has been serving as an associate editor of *IEEE Trans. Human-Machine Systems*, *The Visual Computer and Neurocomputing*. He served as Area/Session chairs for numerous conferences such as ICME, ACCV, ICMR. He was supported by several promotive funds for his works such as the Excellent Young Scientists Funds from National Natural Science Foundation of China. He is a Fellow of IET. He is a senior member of the IEEE.

Lei Zhang (M'04, SM'14, F'18) received the BSc degree from the Shenyang Institute of Aeronautical Engineering, Shenyang, P.R. China, in 1995, and the MSc and PhD degrees in control theory and engineering from Northwestern Polytechnical University, Xi'an, P.R. China, respectively in 1998 and 2001, respectively. Since July 2017, he has been a chair professor with the Department of Computing, The Hong Kong Polytechnic University. His research interests include Computer Vision, Pattern Recognition, Image and Video Analysis, and Biometrics, etc. He has published more than 200 papers in those areas. He is an associate editor of *IEEE Trans. on Image Processing, SIAM Journal of Imaging Sciences and Image and Vision Computing*, etc. He is a "Clarivate Analytics Highly Cited Researcher" from 2015 to 2017. More information can be found in his homepage http://www4.comp.polyu.edu.hk/∼cslzhang/. He is a fellow of the IEEE.

Ming-Hsuan Yang received the PhD degree in computer science from the University of Illinois at Urbana-Champaign, in 2000. He is a professor in electrical engineering and computer science with the University of California, Merced. Yang has served as an associate editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *International Journal of Computer Vision*, Computer Vision and Image Understanding, Image and Vision Computing and Journal of Artificial Intelligence Research. He received the NSF CAREER award in 2012, Senate Award for Distinguished Early Career Research at UC Merced in 2011, and Google Faculty Award in 2009. He is a senior member of the IEEE and the ACM.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.