

UC Merced

UC Merced Previously Published Works

Title

Learning Support Correlation Filters for Visual Tracking

Permalink

<https://escholarship.org/uc/item/9b38v04w>

Journal

IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(5)

ISSN

0162-8828

Authors

Zuo, Wangmeng

Wu, Xiaohe

Lin, Liang

et al.

Publication Date

2019-05-01

DOI

10.1109/tpami.2018.2829180

Peer reviewed

Learning Support Correlation Filters for Visual Tracking

Wangmeng Zuo, Xiaohe Wu, Liang Lin, Lei Zhang, and Ming-Hsuan Yang

Abstract—Sampling and budgeting training examples are two essential factors in tracking algorithms based on support vector machines (SVMs) as a tradeoff between accuracy and efficiency. Recently, the circulant matrix formed by dense sampling of translated image patches has been utilized in correlation filters for fast tracking. In this paper, we derive an equivalent formulation of a SVM model with 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) which find the global optimal solution with real-time performance. For a given circulant data matrix with n^2 samples of size $n \times n$, 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.

I. INTRODUCTION

ROBUST visual tracking is a challenging problem due to the large changes of object appearance caused by pose, illumination, deformation, occlusion, distractors, as well as background clutters. Among the state-of-the-art methods, discriminative classifiers with model update and dense sampling have been demonstrated to perform well in visual tracking. On the other hand, correlation filters have been shown to be efficient for locating objects with the use of circulant matrix and fast Fourier transform. Central to the advances in visual tracking are the development of effective appearance models and efficient sampling schemes.

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) [1], [2], [3], [4]. 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, SVM-based trackers usually only use a limited set of samples [3], [4]. As kernel SVM-based tracking methods are susceptible to the *curse of kernelization*, a budget mechanism is introduced for online learning of the structural SVM tracker [3] to restrict the number of support vectors, or an explicit feature mapping function is used to approximate the intersection kernel [4].

While sampling and budgeting may improve tracking efficiency at the expense of accuracy, most SVM-based trackers [2], [3], [4] do not run in real-time.

Correlation filters (CFs) [5], [6], [7], [8] have recently been utilized for efficient visual tracking. The data matrix formed by dense sampling of base sample has circulant structures, which facilitates the use of the discrete Fourier transform (DFT) for efficient and effective visual tracking [5], [6], [7], [8]. However, ridge regression or kernel ridge regression are generally adopted as the predictors in these trackers. Henriques *et al.* [9] apply the circulant property for training of support vector regression efficiently to detect pedestrians. The problem on how to exploit the circulant property to accelerate SVM-based trackers remains unaddressed.

In this paper, we propose a novel SVM-based algorithm via support correlation filters (SCFs) for efficient and accurate visual tracking. Different from the existing SVM-based trackers, the proposed algorithm based on SCFs deals with the sampling and budgeting issues 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 [10] 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 [3], [4] 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 [4], [11], [12].

II. 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). Several surveys and performance evaluation on state-of-the-art tracking algorithms [13], [14], [10], [15] have been reported in the literature, and in this section we discuss the most relevant methods to this work.

A. 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 [16], subspace representations [17], [18], [19], and sparse representations [20], [21], [22] have been developed for object representations. Discriminative appearance methods are usually based on features learned from a large set of examples with effective classifiers. 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 [10].

Discriminative tracking methods are usually based on object detection within local search using classifiers such as boosting methods [23], [24], [25], [26], [27], random forests [28], [29], and SVMs [1], [2], [3]. Among these classifiers, boosting methods [23], [24], [25], [26], [27] and random forests [28], [29] are ensemble learning methods where sampling from large sets of features is indispensable, and that makes it difficult to adapt 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 [26], [27], [30] and multiple instance [23], [24] learning methods. Considering that classification based methods are trained to predict the class label rather than the object location, Hare *et al.* [3] 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 [31], [25], [32].

B. 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 face recognition [33], [34], object detection [35], [9], [36], object alignment [37] and action recognition [38], [39]. A number of correlation filters have been proposed in the literature including the minimum average correlation energy (MACF) [36], optimal trade-off synthetic discriminant filter (OTSDF) [40], unconstrained minimum average correlation energy (UMACE) [34], and minimum output sum of squared error (MOSSE) [5] methods.

Recently, the max-margin CF (MMCF) [41], multi-channel CF [11], [42], [43], [6], and kernelized CF [6], [7], [44] methods have been developed for object detection and tracking. The MMCF [41] scheme combines the localization properties of correlation filters with good generalization performance of SVM. The multi-channel correlation filters [11], [42], [43], [6] 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 [45] for object detection and landmark localization. The kernel tricks are also employed to learn kernelized synthetic discriminant functions (SDF)[44] with correlation filters. We note that the MMCF [45], [41] and kernelized SDF [44] schemes are trained off-line with high computational load, and do not exploit the circulant structure of data matrix formed by translated images of target objects.

In visual tracking, Bolme *et al.* [5] propose the MOSSE method to learn adaptive correlation filters with high efficiency and competitive performance. Subsequently, the kernelized correlation filter (KCF) [6] 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 [11], [42], [43], [6]. Within the tracking methods based on correlation filters, numerous issues such as adaptive scale estimation [11], [12], [8], limited boundaries [46], zero-aliasing [47], tracking failure [48], and partial occlusion [49] have been addressed.

We note existing CF-based tracking methods are developed with ridge regression schemes for locating the target. On the other hand, the SVM-based tracking methods, e.g., Struck [3] and MEEM [4], have been demonstrated to achieve the state-of-the-art performance. One straightforward extension is to integrate SVM-based trackers with the MMCF method[41]. Nevertheless, the MMCF scheme is computationally prohibitive for real-time applications. In this work, we develop novel discriminative tracking algorithms based on SVMs and correlation filters that perform efficiently and effectively.

III. 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, nonlinear and scale-adaptive correlation filters respectively for robust visual tracking.

A. Problem formulation

Given an image \mathbf{x} , the full set of its translated versions forms a circulant matrix \mathbf{X} with several interesting properties [50], 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:

$$\mathbf{X} = F^H \text{Diag}(\hat{\mathbf{x}}) F, \quad (1)$$

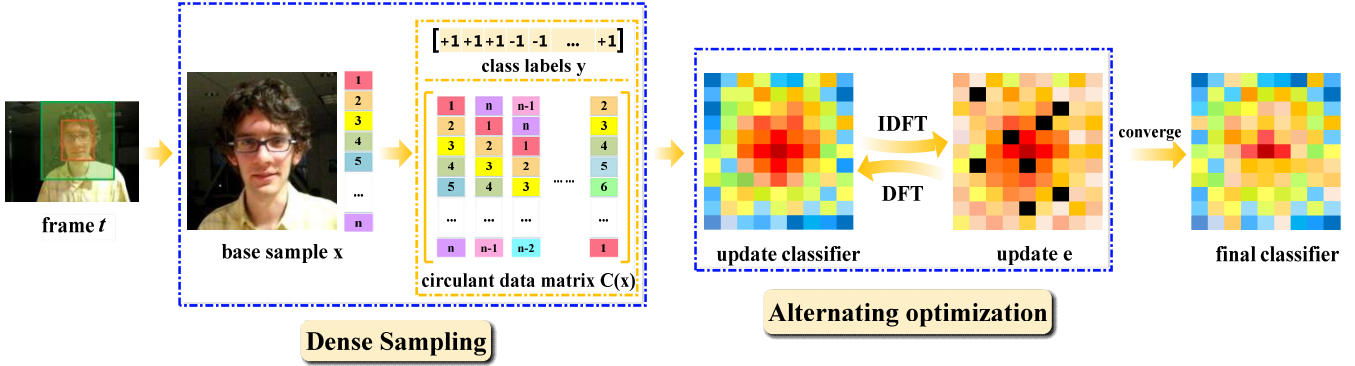


Fig. 1. Illustration of the proposed SCF learning algorithm at the t -th frame. The proposed algorithm iterates between updating \mathbf{e} and updating SVM classifier $\{\mathbf{w}, b\}$ until convergence. In each iteration, only one DFT and one IDFT are required, which make the proposed algorithm computationally 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.

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 $\text{Diag}(\cdot)$ to form a diagonal matrix from a vector, and use $\text{diag}(\cdot)$ to return the diagonal vector of a matrix.

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

$$y_i = \text{sgn}(\mathbf{w}^\top \mathbf{x}_i + b). \quad (2)$$

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

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

where $\mathcal{F}^{-1}(\cdot)$ 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 $\mathbf{X} = [\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_{n^2}]$ with the corresponding class labels $\mathbf{y} = [y_1, y_2, \dots, y_{n^2}]^\top$, we use the squared hinge loss and define the SVM model [51] as follows:

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

where $\boldsymbol{\xi} = [\xi_1, \xi_2, \dots, \xi_i, \dots, \xi_{n^2}]$ is the vector of slack variables.

Based on the circulant property of \mathbf{X} , the SVM model can be equivalently formulated as:

$$\begin{aligned} \min_{\mathbf{w}, b} \|\mathbf{w}\|^2 + C \|\boldsymbol{\xi}\|_2^2 \\ \text{s.t. } \mathbf{y} \circ (\mathcal{F}^{-1}(\hat{\mathbf{x}}^* \circ \hat{\mathbf{w}}) + b\mathbf{1}) \geq \mathbf{1} - \boldsymbol{\xi}. \end{aligned} \quad (4)$$

where \circ denotes the element-wise multiplication, and $\mathbf{1}$ denotes a vector of 1s.

Class labels of the translated images. Let \mathbf{p}^* denote the centre position of the object of interest \mathbf{x}^* , and \mathbf{p}_i as the position of the translated image \mathbf{x}_i . In object detection [31], [32], the overlap function $s(\mathbf{p}^*, \mathbf{p}_i)$ is used to measure the

similarity between \mathbf{x}^* and \mathbf{x}_i . Specifically, the positive samples are defined by all ground truth object windows and the negative samples are defined by those with $s(\mathbf{p}^*, \mathbf{p}_i)$ below a lower overlap threshold. In the proposed discriminative tracking model, we need to set upper and lower thresholds of $s(\mathbf{p}^*, \mathbf{p}_i)$ for assigning binary labels. In Section IV, we determine the optimal upper and lower thresholds for SCF, MSCF and KSCF respectively with experiments.

In this work, we use the following confidence map of object position [8] to define the class label:

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

where γ is a normalization constant, α and β are the scale and shape parameters, respectively. With the confidence map, we define the class labels as follows:

$$y_i = \begin{cases} 1, & \text{if } m(p_i, p^*) \geq \theta_u \\ -1, & \text{if } m(p_i, p^*) \leq \theta_l \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where θ_l and θ_u are the lower and upper thresholds, respectively. With this formulation, we can use the circulant matrix formed by all samples to improve training efficiency, and discard any samples that are not labeled.

Comparisons with existing CF-based trackers. As illustrated in Fig. 2(a), existing CF-based trackers generally follow the ridge regression models. That is, with the continuous confidence map \mathbf{m} , CF-based trackers seek the optimal correlation filter by minimizing the mean squared error (MSE) between the predefined confidence map and actual output,

$$\min_{\mathbf{w}} \lambda \|\mathbf{w}\|^2 + \|\mathbf{X}\mathbf{w} - \mathbf{m}\|_2^2, \quad (6)$$

which has the closed form solution,

$$\hat{\mathbf{w}} = \frac{\hat{\mathbf{x}}^* \circ \hat{\mathbf{m}}}{\hat{\mathbf{x}}^* \circ \hat{\mathbf{x}} + \lambda}. \quad (7)$$

As shown in Fig. 2(b), the proposed model aims to learn a max-margin SVM classifier $\{\mathbf{w}, b\}$ to distinguish the object of interest from the background. Using the label assignment scheme in (5), we can discard uncertain samples in training to alleviate the label ambiguity problem. The importance of

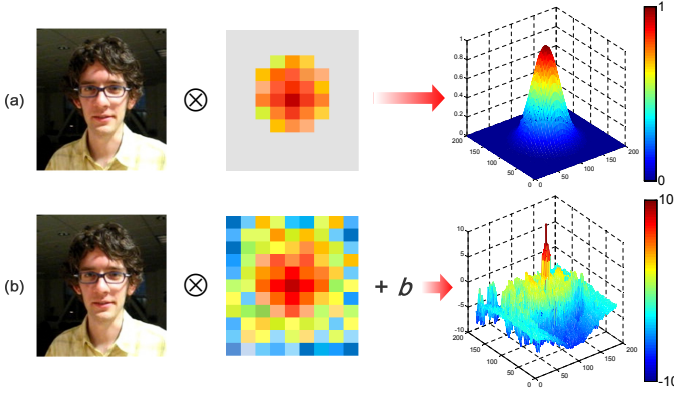


Fig. 2. Differences between the proposed SCF model and existing CF approaches [5], [7], [8]. (a) Existing CF-based models are designed to learn correlation filters that make the actual output being close to the predefined confidence maps. (b) The SCF model aims to learn a support correlation filter together 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.

SVM and label ambiguity issues have been demonstrated in object detection [31]. The proposed model copes with both issues (classification and label ambiguity) for effective visual tracking.

B. Alternating optimization

In this section, we reformulate the model in (3) 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 SVM model in (4) is then reformulated as:

$$\min_{\mathbf{w}, b, \mathbf{e}} \|\mathbf{w}\|^2 + C \|\mathbf{y} \circ (\mathcal{F}^{-1}(\hat{\mathbf{x}}^* \circ \hat{\mathbf{w}}) + b\mathbf{1}) - \mathbf{1} - \mathbf{e}\|^2 \quad (8)$$

s.t. $\mathbf{e} \geq 0$.

With this formulation, the subproblem on \mathbf{e} has a closed form solution when $\{\mathbf{w}, b\}$ is known, and the subproblem on $\{\mathbf{w}, b\}$ has a closed form solution when \mathbf{e} is known. Thus the above model can be efficiently solved using the alternating optimization algorithm by iterating between the following two steps:

Updating \mathbf{e} . Given $\{\mathbf{w}, b\}$, 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:

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

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

$$\mathbf{e} = \max\{\mathbf{e}_0, 0\}. \quad (9)$$

Updating $\{\mathbf{w}, b\}$. Given \mathbf{e} , we let $\mathbf{q} = \mathbf{y} + \mathbf{y} \circ \mathbf{e}$, and the subproblem on $\{\mathbf{w}, b\}$ becomes:

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

The subproblem with $\{\mathbf{w}, b\}$ is a quadratic programming problem. One feasible solution is to let $\mathbf{u} = [\mathbf{w}; b]$ and derive the closed form solution on \mathbf{u} . However, this approach fails to

Algorithm 1 SCF model training

Input: Training image patch \mathbf{x}_t ($n \times n$) class labels \mathbf{y}_t ($n \times n$)
Output: $(\hat{\mathbf{w}}, b)$.

- 1: Initialize $\hat{\mathbf{w}}_0, b_0, k = 1$.
- 2: **while** not converged **do**
- 3: // Lines 4-5 : updating \mathbf{e}_k .
- 4: $\mathbf{d} = \mathbf{y}_t \circ (\mathcal{F}^{-1}(\hat{\mathbf{x}}_t^* \circ \hat{\mathbf{w}}_{k-1}) + b\mathbf{1}) - \mathbf{1}$,
- 5: $\mathbf{e}_k = \max(0, \mathbf{d})$,
- 6: // Lines 7-9 : updating $\mathbf{q}_k, b_k, \mathbf{p}_k$.
- 7: $\mathbf{q}_k = \mathbf{y}_t + \mathbf{y}_t \circ \mathbf{e}_k$,
- 8: $b_k = \text{mean}(\mathbf{q}_k)$,
- 9: $\mathbf{p}_k = \mathbf{q}_k - b_k \mathbf{1}$,
- 10: // Line 11 : updating \mathbf{w}_k .
- 11: $\hat{\mathbf{w}}_k = \frac{\hat{\mathbf{x}}_t^* \circ \hat{\mathbf{p}}_k}{\hat{\mathbf{x}}_t^* \circ \hat{\mathbf{x}}_t + \mathbf{1}/C}$.
- 12: $k \leftarrow k + 1$
- 13: **end while**

exploit the circulant property of \mathbf{X} . Thus, we obtain $\{\mathbf{w}, b\}$ by solving the following system of equations:

$$\hat{\mathbf{w}} = \frac{\hat{\mathbf{x}}^* \circ (\hat{\mathbf{q}} - b\mathbf{1})}{\hat{\mathbf{x}}^* \circ \hat{\mathbf{x}} + \mathbf{1}/C}, \quad (10)$$

$$b = \mathbf{1}^\top (\mathbf{q} - \mathcal{F}^{-1}(\hat{\mathbf{x}}^* \circ \hat{\mathbf{w}})), \quad (11)$$

where $\mathbf{q} = \mathbf{y} + \mathbf{y} \circ \mathbf{e}$. Combining the two equations above and with the property of DFT, we have

$$b = \bar{q}, \quad (12)$$

where \bar{q} is the mean of \mathbf{q} . Given b , we use (10) to obtain the closed form solution to \mathbf{w} .

As illustrated in Fig. 1, when the t -th 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. Given $\{\mathbf{x}^t, \mathbf{y}^t, \mathbf{w}, b\}$, the update 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.

Convergence. 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 A. Based on the optimality condition, we define

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

and adopt the following stopping criterion:

$$\max\{\|\mathbf{r}_1\|_\infty, \max_{e_i > 0}\{\|\mathbf{r}_2(i)\|\}, \max_{e_i = 0}\{\mathbf{r}_3(i)\}\} \leq \epsilon. \quad (13)$$

Comparisons with MMCF [41]. 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 average correlation energy term.

C. Multi-channel SCF

Different local descriptors, e.g., color attributes, HOG, and SIFT [52], [42], [53], 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:

$$\begin{aligned} & \min_{\mathbf{w}, b} \|\mathbf{w}\|^2 + C\|\boldsymbol{\xi}\|_2^2 \\ & \text{s.t. } \mathbf{y} \circ (\mathcal{F}^{-1}(\sum_{l=1}^L (\hat{\mathbf{x}}^l)^* \circ \hat{\mathbf{w}}^l) + b\mathbf{1}) \geq \mathbf{1} - \boldsymbol{\xi} \end{aligned} \quad (14)$$

where L is the number of channels, and \mathbf{x}^l and \mathbf{w}^l denote the l -th channel of the image and correlation filter, respectively. To learn the proposed MSCF model, we adopt the same equations on updating \mathbf{e} and b , and compute \mathbf{w} by solving the following problem:

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

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

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

$$\hat{\mathbf{w}} = (\hat{\mathbf{X}}^H \hat{\mathbf{X}} + \frac{1}{C}\mathbf{I})^{-1} \hat{\mathbf{X}}^H \hat{\mathbf{r}}. \quad (15)$$

where \mathbf{I} is the identity matrix. Note that $\hat{\mathbf{X}}$ is an $n^2 \times Ln^2$ matrix. It is not practical to compute the inverse of $\hat{\mathbf{X}}^H \hat{\mathbf{X}}$ to update $\hat{\mathbf{w}}$. In the multi-channel correlation filters, it is noted that $\hat{\mathbf{X}}$ has the diagonal block structure, and the j -th element of $\hat{\mathbf{r}}$ depends only on $\hat{\mathbf{w}}(j) = [\hat{\mathbf{w}}^1(j); \hat{\mathbf{w}}^2(j); \dots; \hat{\mathbf{w}}^L(j)]$ and $\hat{\mathbf{x}}(j) = [\hat{\mathbf{x}}^1(j); \hat{\mathbf{x}}^2(j); \dots; \hat{\mathbf{x}}^L(j)]$. Thus, the subproblem on $\hat{\mathbf{w}}$ can be further decomposed into n^2 systems of equations:

$$\left(\hat{\mathbf{x}}(j)\hat{\mathbf{x}}(j)^H + \frac{1}{C}\mathbf{I} \right) \hat{\mathbf{w}}(j) = \hat{\mathbf{x}}(j)\hat{\mathbf{r}}(j). \quad (16)$$

In [43], Galoogahi *et al.* solve these n^2 systems of equations by an algorithm with the complexity of $O(n^2 L^3 + Ln^2 \log n)$. We note that the matrix on the left hand of (16) is a rank-one

matrix and a scaled identity matrix. Based on the Sherman-Morrison formula [54], we have

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

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

$$\hat{\mathbf{w}}(j) = \frac{C\hat{\mathbf{x}}(j)\hat{\mathbf{r}}(j)}{1 + C\hat{\mathbf{x}}(j)^H \hat{\mathbf{x}}(j)}. \quad (17)$$

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 several element-wise operations per iteration, and the complexity is $O(n^2 \log n)$.

D. 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:

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

where $\psi(\mathbf{x})$ stands for the nonlinear feature mapping implicitly determined by the kernel function $K(\mathbf{x}, \mathbf{x}')$, and $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_{n^2}]^\top$ 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 [7], 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:

$$\mathbf{K}\boldsymbol{\alpha} = \mathcal{F}^{-1}(\hat{\mathbf{k}}^{\mathbf{xx}} \circ \hat{\boldsymbol{\alpha}}), \quad (18)$$

and we have,

$$\|\mathbf{w}\|^2 = \boldsymbol{\alpha}^\top \mathbf{K}\boldsymbol{\alpha} = \boldsymbol{\alpha}^\top \mathcal{F}^{-1}(\hat{\mathbf{k}}^{\mathbf{xx}} \circ \hat{\boldsymbol{\alpha}}). \quad (19)$$

Based on (18) and (19), the proposed kernelized SCF model is formulated as

$$\begin{aligned} & \min_{\boldsymbol{\alpha}, b} \boldsymbol{\alpha}^\top \mathcal{F}^{-1}(\hat{\mathbf{k}}^{\mathbf{xx}} \circ \hat{\boldsymbol{\alpha}}) \\ & + C(\mathbf{y} \circ (\mathcal{F}^{-1}(\hat{\mathbf{k}}^{\mathbf{xx}} \circ \hat{\boldsymbol{\alpha}}) + b\mathbf{1}) - \mathbf{1} - \mathbf{e})^2 \\ & \text{s.t. } \mathbf{e} \geq 0. \end{aligned} \quad (20)$$

To learn KSCF, we use the alternating optimization method by iteratively solving \mathbf{e} and $\{\mathbf{w}, b\}$. The solution of the subproblem with \mathbf{e} is similar to that in the SCF model, and we update b and \mathbf{w} using the closed form solution of kernel ridge regression. Based on the representation theorem [55], the optimal solution \mathbf{w} in the kernel space can be expressed as the linear combination of the feature maps of the samples: $\mathbf{w}^* = \sum_i \alpha_i^* \varphi(x_i)$. Namely, only the coefficient vector $\boldsymbol{\alpha}$ needs to be learned. In [55], the solution to the kernelized ridge regression in the dual space is given by

$$\boldsymbol{\alpha} = (\mathbf{K} + \lambda\mathbf{I})^{-1} \mathbf{y}.$$

Thus, the closed form solution to our sub-problem on $\boldsymbol{\alpha}$ can be formulated as

$$\boldsymbol{\alpha} = (\mathbf{K} + \lambda\mathbf{I})^{-1} (\mathbf{q} - b\mathbf{1}),$$

TABLE I
RESULTS OF MSCF AND DCF WITH DIFFERENT FEATURE REPRESENTATIONS.

Algorithms Features	MSCF				DCF [6]			
	Raw pixels	CN	HOG	HOG + CN	Raw pixels	CN	HOG	HOG + CN
Mean DP (%)	64.9	66.3	78.4	80.6	44.4	48.0	71.9	76.2
Mean AUC (%)	44.6	44.9	53.7	55.5	31.2	34.8	50.1	53.2
Mean FPS (s)	76	62	64	54	278	210	292	151

TABLE II
RESULTS OF KSCF AND KCF WITH DIFFERENT FEATURE REPRESENTATIONS.

Algorithms Features	KSCF				KCF [6]			
	Raw pixels	CN	HOG	HOG + CN	Raw pixels	CN	HOG	HOG + CN
Mean DP (%)	64.4	68.1	79.3	85.0	55.3	57.3	73.2	75.8
Mean AUC (%)	45.3	46.9	53.2	57.5	40.0	41.8	50.7	53.0
Mean FPS (s)	40	37	44	35	154	120	172	102

where $\mathbf{q} = \mathbf{y} + \mathbf{y} \circ \mathbf{e}$ and $\mathbf{1}$ denotes a vector of 1s. As the kernel matrix \mathbf{K} is circulant and can be diagonalized, the optimal solution of α in the Fourier transform domain can be computed by

$$\hat{\alpha}^* = \frac{\hat{\mathbf{t}}}{\hat{\mathbf{k}}^{xx} + 1/C}, \quad (21)$$

where $\hat{\mathbf{t}} = \hat{\mathbf{y}} + \hat{\mathbf{y}} \circ \hat{\mathbf{e}} - b\hat{\mathbf{1}}$, \mathbf{k}^{xx} is the kernel correlation of \mathbf{x} with itself in the Fourier domain which is known as the kernel auto-correlation.

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 KCSF model leverages rich features from the nonlinear filters without increasing computational load significantly.

Furthermore, to handle large scale changes, we develop the SKSCF model by maintaining a scaling pool in a way similar to the scale-adaptive CF scheme [12], and the bilinear interpolation is used to resize samples across scales.

IV. PERFORMANCE EVALUATION

We use the benchmark dataset and protocols [10] to evaluate the proposed SCF algorithms. First, we evaluate several variants of the proposed method, i.e., SCF, MSCF, KSCF, and SKSCF, to analyze the effect of feature representations and kernel functions. Next, comprehensive experiments are conducted to compare the proposed methods with other CF-based trackers. Finally, the KSCF and SKSCF algorithms are compared with existing SVM-based and the state-of-the-art methods. The tracking results can be found at <http://faculty.ucmerced.edu/project/scf/> and the source code will be made available to the public.

A. Experimental setup

Datasets and evaluated tracking methods. To assess the performance of the proposed methods, experiments are carried out on a benchmark dataset [10] 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 SKSCF methods. The SCF and MSCF methods are designed in the linear space with raw pixels, and multi-channel features are based on HOG [52] as well as color names (CN) [42], respectively. The KSCF 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 [5], CSK [7], KCF [6], DCF [6], STC [8] and CN [42]), existing SVM based trackers (e.g., Struck [3] and MEEM [4]), and other state-of-the-art methods (e.g., TGPR [56], SCM [16], TLD [27], LIAPG [57], MIL [23], ASLA [20] and CT [58]).

Evaluation protocols. We use the one-pass evaluation (OPE) protocol [10] 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 Xenon 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 (5). In addition, the KSCF method has one extra parameter σ for the Gaussian RBF kernel function, and SKSCF contains a scaling pool parameter s . For online tracking, the model is updated by linear interpolation with the adaption rate ρ [10].

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 2, respectively. The thresholds (θ_l, θ_u) in (5) are set to (0.3, 0.7) for SCF, (0.4, 0.9) for MSCF and (0.5, 0.7) for KSCF, SKSCF. The scale parameter α is set to be $50/(mn)$, which is adaptive to the size $m \times n$ of each target object. The scaling pool s is fixed

TABLE III
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

as $[0.985, 0.990, 0.995, 1, 1.005, 1.010, 1.015]$. The adaption rate ρ is set to 0.075 for raw pixel features, and 0.02 for multi-channel features, respectively. The kernel parameter σ of KSCF is set to 0.2. As for HOG parameters, the orientations and cell size are set to 9 and 4.

B. 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.

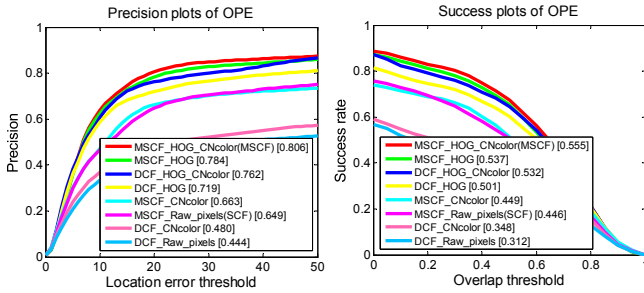


Fig. 3. OPE plots of the MSCF and DCF [6] with different feature representations. the AUC values are shown next to the legends.

We consider three typical feature representations, i.e., raw pixels, HOG features [52], and color names (CN) [42]. The results of the MSCF and KSCF methods are listed in Table I and Table II. The result for each feature representation is optimal by varying the parameters β , ρ , θ_l and θ_u from $[0.5, 1, 1.5, 2]$, $[0.02, 0.04, 0.075]$, $[0.7, 0.8, 0.9]$ and $[0.3, 0.4, 0.5]$. These parameters are then fixed for all the following experiments. For KSCF, the Gaussian RBF kernel with $\sigma=0.2$ is adopted.

The OPE plots of MSCF with linear DCF [6] and KSCF with nonlinear KCF [6] are shown in Fig. 3 and Fig. 4. Compared with raw pixels and color features, the method with HOG representation significantly improves the tracking performance in terms of mean DP and mean AUC. For MSCF, the implementation using color names and HOG features outperforms raw pixels by 1.4% and 13.5% in terms of mean DP. For KSCF, the tracker using color names and HOG features outperforms raw pixels by 3.7% and 14.9% in terms of mean DP.

The MSCF tracker with the combination of color names 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 color names and HOG features. Compared with the DCF [6] and KCF [6] methods,

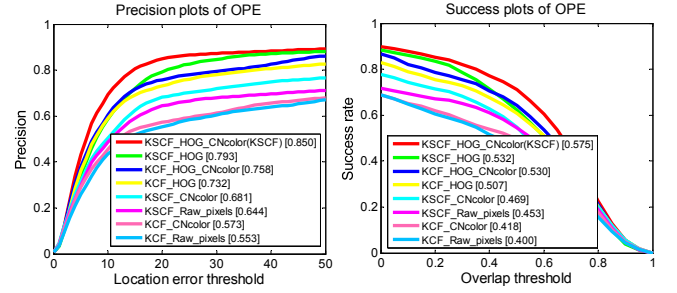


Fig. 4. OPE plots of the KSCF and KCF [6] methods with different feature representations.

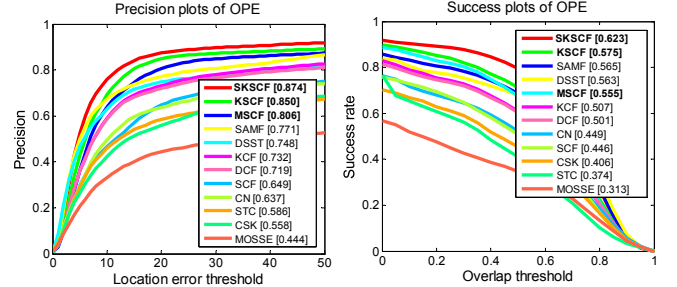


Fig. 5. OPE plots of the SCF methods (i.e., SCF, MSCF, KSCF, and SKSCF) and other CF-based trackers (i.e., MOSSE [5], CSK [7], DCF [6], KCF [6], STC [8], CN [42], DSST [11] and SAMF [12]).

the proposed MSCF and KSCF algorithms achieve higher DP and AUC values for each feature representation. Table I and II show that both KSCF and MSCF methods perform in real-time even using the representation based on HOG and CN features.

We evaluate the effect of kernel functions on KSCF using HOG and CN features, including linear kernel $K_l(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{x}_j$, polynomial kernel $K_p(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^\top \mathbf{x}_j + 1)^d$, and Gaussian RBF kernel $K_g(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\frac{1}{2\sigma^2} \|\mathbf{x}_i - \mathbf{x}_j\|^2)$. For $K_p(\mathbf{x}_i, \mathbf{x}_j)$, the degree d is set as 2. For $K_g(\mathbf{x}_i, \mathbf{x}_j)$, the kernel parameter σ is set as 0.2. Table III 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 IV 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.

C. Comparisons with CF-based trackers

We use the tracking benchmark dataset [10] to evaluate the proposed SCF-based algorithm against existing CF-based methods including MOSSE [5], CSK [7], KCF [6], DCF [6], STC [8], CN [42], DSST [11] and SAMF [12].

TABLE IV
PERFORMANCE OF TRACKING METHODS BASED ON CORRELATION FILTERS: TOP THREE RESULTS ARE SHOWN IN RED, BLUE AND ORANGE.

Algorithms	SKSCF	KSCF	MSCF	SCF	SAMF [12]	DSST [11]	KCF [6]	DCF [6]	CN [42]	STC [8]	CSK [7]	MOSSE [5]
Mean DP (%)	87.4	85.0	80.6	62.8	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	56.5	56.3	50.7	50.1	44.9	37.4	40.6	31.3
Mean FPS (<i>s</i>)	8	35	54	76	14	30	172	292	79	557	151	421

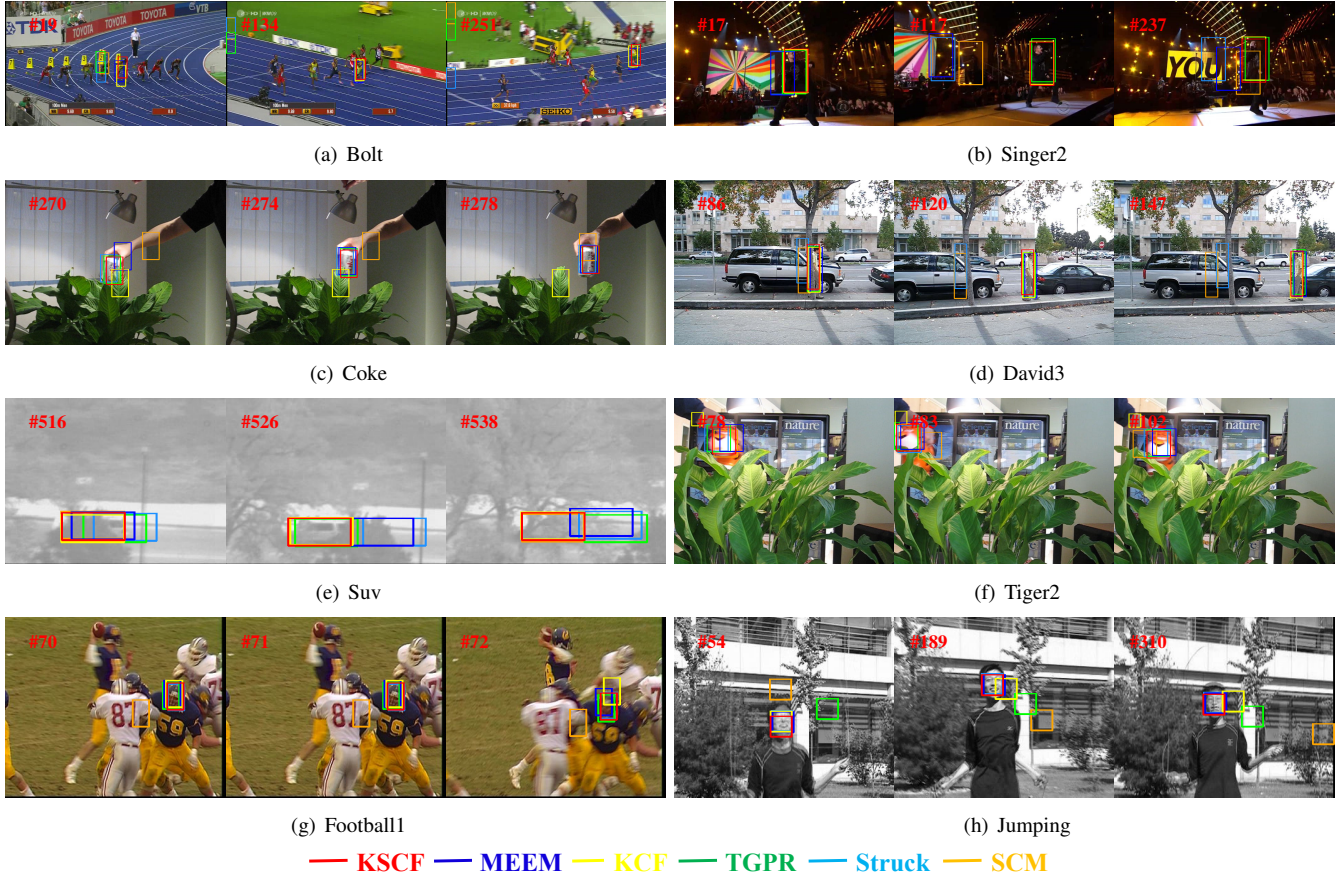


Fig. 6. Screenshots of tracking results on 8 challenging benchmark sequences. For the sake of clarity, we only show the results of five trackers, i.e., KSCF, KCF [6], MEEM [4], TGPR [56], Struck [3] and SCM [16].

TABLE V
COMPARISON OF SVM-BASED TRACKERS.

Algorithms	SKSCF	KSCF	MEEM [4]	Struck [3]
Mean DP (%)	87.4	85.0	83.3	67.4
Mean AUC (%)	62.3	57.5	57.2	48.6
Mean FPS (<i>s</i>)	8	35	10	10

Classic correlation filters. Fig. 5 shows the OPE plots of these trackers. The SCF, MOSSE [5], CSK [7] and STC [8] 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

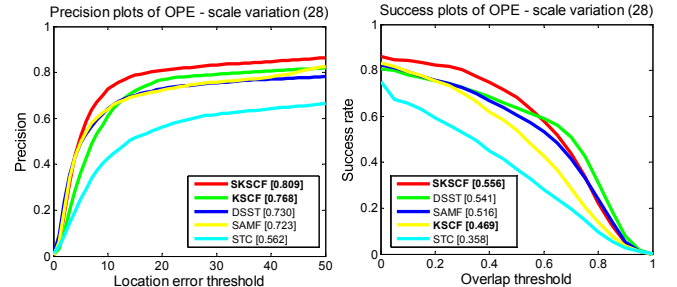


Fig. 7. OPE plots of the KSCF, SKSCF, DSST [11] and SAMF [12] methods on sequences with large scale variation.

regression and nonlinear functions.

Multi-channel correlation filters. The MSCF, CN [42], and DCF [6] methods are based on correlation filters using multi-channel features. The DCF method is based on HOG features

TABLE VI

COMPARISON OF KSCF AND SKSCF METHODS WITH THE STATE-OF-THE-ART TRACKERS. THE TOP THREE RESULTS ARE SHOWN IN RED, BLUE AND ORANGE.

Algorithms	SKSCF	KSCF	MEEM	KCF	TGPR	SCM	TLD	ASLA	L1APG	MIL	CT
			[4]	[6]	[56]	[16]	[27]	[20]	[57]	[23]	[58]
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

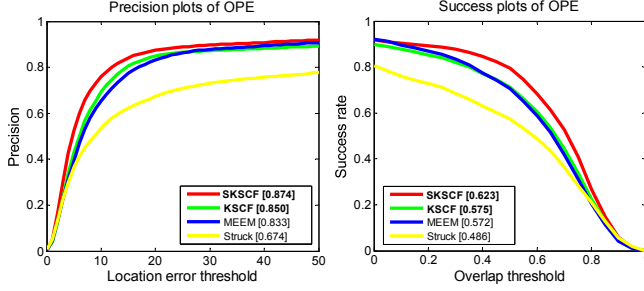


Fig. 8. OPE plots of the KSCF, SKSCF and other SVM-based trackers, including MEEM [4] and Struck [3].

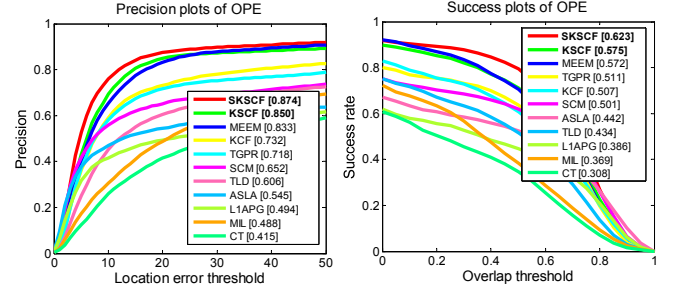


Fig. 10. OPE plots of the KSCF, SKSCF and other state-of-the-art trackers, including MEEM [4], TGPR [56], KCF [6], SCM [16], TLD [27], ASLA [20], L1APG [57], MIL [23] and CT [58].

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 [6] and CSK [7] trackers. The CSK and KCF methods are based on raw pixels and HOG features, respectively. As shown in Table IV and Fig. 6, the KSCF method based on HOG and CN features performs favorably against the KCF and CSK approaches.

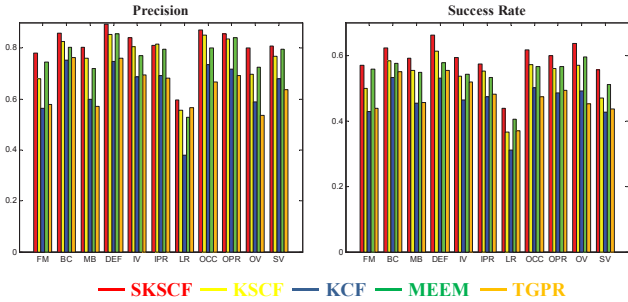


Fig. 9. Precision and success metrics of four top-performing trackers for the 11 attributes.

Scale-adaptive correlation filters. The KSCF and SKSCF are evaluated against three scale-adaptive trackers: STC [8], DSST [11] and SAMF [12]. We note that the DSST [11] and SAMF [12] methods have been shown to perform best and second best trackers in the recent tracking benchmark evaluation [59]. 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.

D. Comparisons with SVM-based trackers

We evaluate the proposed KSCF and SKSCF with two state-of-the-art SVM-based methods, i.e., Struck [3] and MEEM [4], based on the structured and ensemble learning. Table V 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.

E. Comparisons with state-of-the-art trackers

We evaluate the KSCF algorithm with the other state-of-the-art trackers, including MEEM [4], KCF [6], TGPR [56], SCM [16], TLD [27], L1APG [57], MIL [23], ASLA [20] and CT [58]. Fig. 10 shows the OPE plots, and Table VI presents the 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 [10] are annotated with 11 challenging factors 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 clutters (BC), and low resolution (LR). Table VII and Table VIII show the performance of the KSCF

TABLE VII
PRECISION METRICS OF THE TRACKERS FOR 11 ATTRIBUTES. THE TOP THREE RESULTS ARE SHOWN IN RED, BLUE AND ORANGE.

Attributes	FM	BC	MB	DEF	IV	IPR	LR	OCC	OPR	OV	SV
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 [4]	0.745	0.802	0.721	0.856	0.771	0.796	0.529	0.801	0.840	0.726	0.795
TGPR [56]	0.579	0.763	0.570	0.760	0.695	0.683	0.567	0.668	0.693	0.535	0.637
KCF [6]	0.564	0.752	0.599	0.747	0.687	0.692	0.379	0.735	0.718	0.589	0.680
SCM [16]	0.346	0.578	0.358	0.589	0.613	0.613	0.305	0.646	0.621	0.429	0.672
TLD [27]	0.557	0.428	0.523	0.495	0.540	0.588	0.349	0.556	0.593	0.576	0.606
ASLA [20]	0.255	0.496	0.283	0.473	0.529	0.521	0.156	0.479	0.535	0.333	0.552
LIAPG [57]	0.367	0.425	0.379	0.398	0.341	0.524	0.460	0.475	0.490	0.329	0.472
MIL [23]	0.415	0.456	0.381	0.493	0.359	0.465	0.171	0.448	0.484	0.393	0.471
CT [58]	0.330	0.339	0.314	0.463	0.365	0.361	0.152	0.429	0.405	0.336	0.448

TABLE VIII
SUCCESS METRICS OF THE TRACKERS FOR 11 ATTRIBUTES. THE TOP THREE RESULTS ARE SHOWN IN RED, BLUE AND ORANGE.

Attributes	FM	BC	MB	DEF	IV	IPR	LR	OCC	OPR	OV	SV
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 [4]	0.706	0.747	0.692	0.711	0.653	0.648	0.470	0.694	0.694	0.742	0.594
TGPR [56]	0.542	0.713	0.570	0.711	0.632	0.601	0.501	0.592	0.603	0.546	0.505
KCF [6]	0.516	0.669	0.539	0.668	0.534	0.575	0.358	0.593	0.587	0.589	0.477
SCM [16]	0.348	0.550	0.358	0.566	0.586	0.574	0.308	0.602	0.576	0.449	0.635
TLD [27]	0.475	0.388	0.485	0.434	0.461	0.477	0.327	0.455	0.489	0.516	0.494
ASLA [20]	0.261	0.468	0.284	0.485	0.514	0.496	0.163	0.469	0.509	0.359	0.544
LIAPG [57]	0.359	0.404	0.363	0.398	0.298	0.445	0.458	0.437	0.423	0.341	0.407
MIL [23]	0.353	0.414	0.261	0.440	0.300	0.339	0.157	0.378	0.369	0.416	0.335
CT [58]	0.327	0.323	0.262	0.420	0.308	0.290	0.143	0.360	0.325	0.405	0.342

and state-of-the-art methods in terms of DP and AUC with respect to each factor. Fig. 9 shows the precision and success metrics of the leading trackers (i.e., SKSCF, KSCF, MEEM, KCF and TGPR) with respect to the attributes. We note that the MEEM method [4] 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.

V. 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 tracking methods to exploit multidimensional features, nonlinear 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.

ACKNOWLEDGMENTS

This work is supported in part by NSFC grant (61271093), the program of ministry of education for new century excellent talents (NCET-12-0150), NSF CAREER Grant (No. 1149783) and NSF IIS Grant (No. 1152576).

APPENDIX A CONVERGENCE ANALYSIS

In the following, we first analyze the optimality condition of the problem, and then prove the global convergence and convergence rate of the SCF algorithm.

A. Optimality conditions

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

$$(\mathbf{w}, b, \mathbf{e}) = \arg \min_{\mathbf{w}, b, \mathbf{e}} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{2} \|\mathbf{y} \circ (\mathbf{X}^\top \mathbf{w} + b\mathbf{1}) - \mathbf{1} - \mathbf{e}\|^2,$$

$$\text{s.t. } \mathbf{e} \geq 0$$

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:

$$(\tilde{\mathbf{w}}, \mathbf{e}) = \arg \min_{\tilde{\mathbf{w}}, \mathbf{e}} \frac{1}{2} \tilde{\mathbf{w}}^\top \tilde{\mathbf{I}} \tilde{\mathbf{w}} + \frac{C}{2} \|\tilde{\mathbf{X}}^\top \tilde{\mathbf{w}} - \mathbf{y} - \mathbf{y} \circ \mathbf{e}\|^2, \quad (22)$$

$$\text{s.t. } \mathbf{e} \geq 0$$

where $\tilde{\mathbf{X}} = [\mathbf{X}^\top, \mathbf{1}]^\top$ and $\tilde{\mathbf{I}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{bmatrix}$. We introduce

an indicator function $\delta(e_i) = \begin{cases} -\infty, & \text{if } e_i < 0 \\ e_i, & \text{if } e_i \geq 0 \end{cases}$ and the subdifferential [60] of $\delta(e_i)$ is:

$$\partial \delta(e_i) = \begin{cases} 1, & \text{if } e_i > 0 \\ (-\infty, 0), & \text{if } e_i = 0 \\ \phi(\text{undefined}), & \text{if } e_i < 0 \end{cases} \quad (23)$$

As the loss function (22) 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 [61]. Thus the optimality conditions are:

$$\begin{cases} \tilde{\mathbf{w}}\tilde{\mathbf{I}} + C\tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T\tilde{\mathbf{w}} - \mathbf{y} - \mathbf{y} \circ \mathbf{e}) = 0 \\ \begin{cases} e_i + 1 - y_i\tilde{\mathbf{x}}_i^T\tilde{\mathbf{w}} = 0, & \text{if } e_i > 0 \\ y_i\tilde{\mathbf{x}}_i^T\tilde{\mathbf{w}} - 1 < 0 & \text{if } e_i = 0 \end{cases} \end{cases} \quad (24)$$

where $\tilde{\mathbf{X}}_i$ denotes the i -th training sample. With $\lambda = \frac{1}{C}$, we have:

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

where $\mathbf{X}_c = [\mathbf{x}_1 - \bar{\mathbf{x}}, \dots, \mathbf{x}_n - \bar{\mathbf{x}}]$ with $\bar{\mathbf{x}} = \frac{1}{n} \sum_i \mathbf{x}_i$. Thus the matrix $(\lambda\tilde{\mathbf{I}} + \tilde{\mathbf{X}}\tilde{\mathbf{X}}^T)$ is invertible. For simplicity, let $\mathbf{M} = \tilde{\mathbf{I}} + C\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$, from (24) and above equation, we have

$$\tilde{\mathbf{w}} = C\mathbf{M}^{-1}\tilde{\mathbf{X}}(\mathbf{y} + \mathbf{y} \circ \mathbf{e}), \quad (25)$$

$$(C\mathbf{y} \circ \mathbf{M}^{-1}\tilde{\mathbf{X}}(\mathbf{y} + \mathbf{y} \circ \mathbf{e}) - 1)_i \begin{cases} = e_i, & \text{if } e_i > 0 \\ < 0, & \text{if } e_i = 0 \end{cases} \quad (26)$$

Based on the optimality conditions in (24), we define

$$\begin{cases} r_1 = \tilde{\mathbf{w}} + C\tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T\tilde{\mathbf{w}} - \mathbf{y} - \mathbf{y} \circ \mathbf{e}), \\ r_2(i) = e_i + 1 - y_i\tilde{\mathbf{x}}_i^T\tilde{\mathbf{w}} \quad \forall e_i > 0, \\ r_3(i) = y_i\tilde{\mathbf{x}}_i^T\tilde{\mathbf{w}} - 1 \quad \forall e_i \leq 0, \end{cases}$$

and use the stopping criterion:

$$\max\left\{\|r_1\|_\infty, \max_{e_i > 0} \|r_2(i)\|, \max_{e_i = 0} \|r_3(i)\|\right\} \leq \epsilon,$$

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

B. Global convergence

To compute \mathbf{e} , we reformulate the sub-problem for each entry:

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

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

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

Proposition 1. For any $a, b \in \mathbb{R}$, we have:

$$\|g(a) - g(b)\|^2 \leq \|a - b\|^2,$$

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

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

$$\mathbf{M} = \tilde{\mathbf{I}} + C\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T = \tilde{\mathbf{I}} + C\mathbf{U}\mathbf{U}^T,$$

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

where $\rho(\mathbf{T}) < 1$ and $\rho(\mathbf{T})$ is the spectral radius of matrix \mathbf{T} [62].

With the definitions of \mathbf{M} and \mathbf{T} , the updating rules $\tilde{\mathbf{w}}$ and \mathbf{e} can be written as:

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

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

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:

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

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

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

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

Denote the eigen-decomposition of \mathbf{T} by $\mathbf{T} = \mathbf{Q}^T\mathbf{\Lambda}\mathbf{Q}$, where \mathbf{Q} is a full rank orthogonal matrix, and $\mathbf{\Lambda}$ 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 $\|\mathbf{Q}^T\mathbf{\Lambda}\mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}})\| = \|\mathbf{e} - \hat{\mathbf{e}}\|$. Since \mathbf{Q} is full-rank orthogonal, there is $\|\mathbf{e} - \hat{\mathbf{e}}\| = \|\mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}})\|$. Thus, we have $\|\mathbf{\Lambda}\mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}})\| = \|\mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}})\|$. Since $\mathbf{\Lambda}$ is diagonal with $0 \leq \lambda_i \leq 1$, it holds that $\mathbf{\Lambda}\mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}}) = \mathbf{Q}(\mathbf{e} - \hat{\mathbf{e}})$. Multiplying both sides by \mathbf{Q}^T , we have $\mathbf{T}(\mathbf{e} - \hat{\mathbf{e}}) = h(\mathbf{e}) - h(\hat{\mathbf{e}}) = \mathbf{e} - \hat{\mathbf{e}}$. \square

Definition 1. (Fixed point) [63] 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 1. Given any fixed point $\hat{\mathbf{e}}$ of $g \circ h$, for any \mathbf{e} , we have:

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

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

Proof. From Propositions 1 and 2, it holds:

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

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 $g \circ h(\mathbf{e}) = \mathbf{e}$. \square

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 $\mathbf{e}^* = \lim_{j \rightarrow \infty} \mathbf{e}^{k_j}$ 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:

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

Based on above, we get the limit as below:

$$\lim_{k \rightarrow \infty} \|\mathbf{e}^k - \hat{\mathbf{e}}\| = \lim_{j \rightarrow \infty} \|\mathbf{e}^{k_j} - \hat{\mathbf{e}}\| = \|\mathbf{e}^* - \hat{\mathbf{e}}\|,$$

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:

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

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

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

According to Lemma 1, 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: $\lim_{k \rightarrow \infty} \mathbf{e}^k = \mathbf{e}^*$ is obtained. We next show that \mathbf{e}^* satisfies the optimization condition in (26). With the definition of \mathbf{T} , g and h , we have:

$$\begin{aligned} g \circ h(\mathbf{e}) &= g(\mathbf{T}(\mathbf{1} + \mathbf{e}) - 1) \\ &= g(\mathbf{C}\mathbf{U}^\top(\tilde{\mathbf{I}} + \mathbf{C}\mathbf{U}\mathbf{U}^\top)^{-1}\mathbf{U}(\mathbf{1} + \mathbf{e}) - 1) \begin{cases} = \mathbf{e}_i, & \text{if } \mathbf{e}_i > 0 \\ < 0, & \text{if } \mathbf{e}_i = 0 \end{cases}, \end{aligned} \quad (27)$$

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. \square

C. q -linear convergence rate

Theorem 2. (Convergence rate) *The sequence $\{(\tilde{\mathbf{w}}^k, \mathbf{e}^k)\}$ generated by our algorithm satisfies the following 3 conditions:*

- 1) $\|\mathbf{e}^{k+1} - \mathbf{e}^*\| \leq \sqrt{\rho(\mathbf{T}^2)} \|\mathbf{e}^k - \mathbf{e}^*\|,$
- 2) $\|\mathbf{U}^\top(\tilde{\mathbf{w}}^{k+1} - \tilde{\mathbf{w}}^*)\| \leq \sqrt{\rho(\mathbf{T}^2)} \|\mathbf{U}^\top(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^*)\|,$
- 3) $\|\tilde{\mathbf{w}}^{k+1} - \tilde{\mathbf{w}}^*\|_{\mathbf{M}} \leq \sqrt{\rho(\mathbf{T})} \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^*\|_{\mathbf{M}}.$

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

$$\tilde{\mathbf{w}}^{k+1} - \tilde{\mathbf{w}}^* = \mathbf{C}\mathbf{M}^{-1}\mathbf{U}(\mathbf{e}^{k+1} - \mathbf{e}^*), \quad (28)$$

$$\|\mathbf{e}^{k+1} - \mathbf{e}^*\|^2 = \|g \circ h(\mathbf{e}^k) - g \circ h(\mathbf{e}^*)\|^2 \leq \|\mathbf{U}^\top(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^*)\|^2 \quad (29)$$

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

$$\|\mathbf{e}^{k+1} - \mathbf{e}^*\|^2 \leq \|\mathbf{T}(\mathbf{e}^k - \mathbf{e}^*)\|^2,$$

Consequently, we have:

$$\|\mathbf{e}^{k+1} - \mathbf{e}^*\|^2 \leq (\mathbf{e}^k - \mathbf{e}^*)^\top (\mathbf{T}^2) (\mathbf{e}^k - \mathbf{e}^*) \leq \rho(\mathbf{T}^2) \|\mathbf{e}^k - \mathbf{e}^*\|^2.$$

By reformulating above, condition 1 can be satisfied:

$$\|\mathbf{e}^{k+1} - \mathbf{e}^*\| \leq \sqrt{\rho(\mathbf{T}^2)} \|\mathbf{e}^k - \mathbf{e}^*\|. \quad (30)$$

Multiplying $\tilde{\mathbf{X}}^\top$ on both sides of (28), and combining with (29), we obtain:

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

which can be reformulated as:

$$\|\mathbf{U}^\top(\tilde{\mathbf{w}}^{k+1} - \tilde{\mathbf{w}}^*)\| \leq \sqrt{\rho(\mathbf{T}^2)} \|\mathbf{U}^\top(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^*)\|,$$

and satisfies condition 2. From (28), we have:

$$\|\tilde{\mathbf{w}}^{k+1} - \tilde{\mathbf{w}}^*\|_{\mathbf{M}}^2 = (\mathbf{e}^{k+1} - \mathbf{e}^*)^\top \mathbf{T} (\mathbf{e}^{k+1} - \mathbf{e}^*) \leq \rho(\mathbf{T}) \|\mathbf{e}^{k+1} - \mathbf{e}^*\|^2.$$

Combining (29) and the definition of \mathbf{M} , we have:

$$\|\tilde{\mathbf{w}}^{k+1} - \tilde{\mathbf{w}}^*\|_{\mathbf{M}} \leq \sqrt{\rho(\mathbf{T})} \|\mathbf{U}^\top(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^*)\| \leq \sqrt{\rho(\mathbf{T})} \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^*\|_{\mathbf{M}}.$$

Thus, the third condition 3 holds and $\tilde{\mathbf{w}}^k$ converges to $\tilde{\mathbf{w}}^*$ q -linearly [64]. \square

REFERENCES

- [1] S. Avidan, "Support vector tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1064–1072, 2004.
- [2] Y. Bai and M. Tang, "Robust tracking via weakly supervised ranking svm," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1854–1861.
- [3] S. Hare, A. Saffari, and P. H. Torr, "Struck: Structured output tracking with kernels," in *IEEE International Conference on Computer Vision*, 2011, pp. 263–270.
- [4] J. Zhang, S. Ma, and S. Sclaroff, "Meem: Robust tracking via multiple experts using entropy minimization," in *European Conference on Computer Vision*, 2014, pp. 188–203.
- [5] D. S. Bolme, J. R. Beveridge, B. Draper, Y. M. Lui *et al.*, "Visual object tracking using adaptive correlation filters," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2544–2550.
- [6] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.
- [7] —, "Exploiting the circulant structure of tracking-by-detection with kernels," in *European Conference on Computer Vision*, 2012, pp. 702–715.
- [8] K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M.-H. Yang, "Fast visual tracking via dense spatio-temporal context learning," in *European Conference on Computer Vision*, 2014, pp. 127–141.
- [9] J. F. Henriques, J. Carreira, R. Caseiro, and J. Batista, "Beyond hard negative mining: Efficient detector learning via block-circulant decomposition," in *IEEE International Conference on Computer Vision*, 2013, pp. 2760–2767.
- [10] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2411–2418.
- [11] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *British Machine Vision Conference, Nottingham*, 2014.
- [12] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Workshop of European Conference on Computer Vision*, 2014, pp. 254–265.
- [13] Y. Pang and H. Ling, "Finding the best from the second bests-inhibiting subjective bias in evaluation of visual tracking algorithms," in *IEEE International Conference on Computer Vision*, 2013, pp. 2784–2791.

- [14] A. W. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: an experimental survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1442–1468, 2014.
- [15] H. Yang, L. Shao, F. Zheng, L. Wang, and Z. Song, "Recent advances and trends in visual tracking: A review," *Neurocomputing*, vol. 74, no. 18, pp. 3823–3831, 2011.
- [16] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparsity-based collaborative model," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1838–1845.
- [17] W. Hu, X. Li, W. Luo, X. Zhang, S. Maybank, and Z. Zhang, "Single and multiple object tracking using log-euclidean riemannian subspace and block-division appearance model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 12, pp. 2420–2440, 2012.
- [18] J. Lim, D. A. Ross, R.-S. Lin, and M.-H. Yang, "Incremental learning for visual tracking," in *Advances in neural information processing systems*, 2004, pp. 793–800.
- [19] J. Ho, K.-C. Lee, M.-H. Yang, and D. Kriegman, "Visual tracking using learned linear subspaces," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2004, pp. I–782.
- [20] X. Jia, H. Lu, and M.-H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1822–1829.
- [21] F. Chen, Q. Wang, S. Wang, W. Zhang, and W. Xu, "Object tracking via appearance modeling and sparse representation," *Image and Vision Computing*, vol. 29, no. 11, pp. 787–796, 2011.
- [22] D. Wang, H. Lu, Z. Xiao, and M.-H. Yang, "Inverse sparse tracker with a locally weighted distance metric," *IEEE Transactions on Image Processing*, vol. 24, no. 9, pp. 2646–2657, 2015.
- [23] B. Babenko, M.-H. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 983–990.
- [24] —, "Robust object tracking with online multiple instance learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1619–1632, 2011.
- [25] H. Grabner, M. Grabner, and H. Bischof, "Real-time tracking via on-line boosting," in *British Machine Vision Conference*, vol. 1, 2006.
- [26] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in *European Conference on Computer Vision*, 2008, pp. 234–247.
- [27] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, 2012.
- [28] A. Saffari, C. Leistner, J. Santner, M. Godec, and H. Bischof, "On-line random forests," in *IEEE International Conference on Computer Vision Workshop*, 2009, pp. 1393–1400.
- [29] J. Santner, C. Leistner, A. Saffari, T. Pock, and H. Bischof, "Prost: Parallel robust online simple tracking," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 723–730.
- [30] F. Tang, S. Brennan, Q. Zhao, and H. Tao, "Co-tracking using semi-supervised support vector machines," in *IEEE International Conference on Computer Vision*, 2007, pp. 1–8.
- [31] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.
- [32] J. R. Uijlings, K. E. van de Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *International journal of computer vision*, vol. 104, no. 2, pp. 154–171, 2013.
- [33] B. V. Kumar, M. Savvides, and C. Xie, "Correlation pattern recognition for face recognition," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1963–1976, 2006.
- [34] M. Savvides and B. V. Kumar, "Efficient design of advanced correlation filters for robust distortion-tolerant face recognition," in *Proceedings IEEE Conference on Advanced Video and Signal Based Surveillance*, 2003, pp. 45–52.
- [35] D. S. Bolme, B. Draper, J. R. Beveridge *et al.*, "Average of synthetic exact filters," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 2105–2112.
- [36] A. Mahalanobis, B. V. Kumar, and D. Casasent, "Minimum average correlation energy filters," *Applied Optics*, vol. 26, no. 17, pp. 3633–3640, 1987.
- [37] V. N. Boddeti, T. Kanade, and B. V. Kumar, "Correlation filters for object alignment," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2291–2298.
- [38] I. Gondal, M. Murshed *et al.*, "Action recognition using spatio-temporal distance classifier correlation filter," in *International Conference on Digital Image Computing Techniques and Applications*, 2011, pp. 474–479.
- [39] M. D. Rodriguez, J. Ahmed, and M. Shah, "Action mach a spatio-temporal maximum average correlation height filter for action recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.
- [40] P. Refregier, "Optimal trade-off filters for noise robustness, sharpness of the correlation peak, and hornor efficiency," *Optics Letters*, vol. 16, no. 11, pp. 829–831, 1991.
- [41] A. Rodriguez, V. N. Boddeti, B. V. Kumar, and A. Mahalanobis, "Maximum margin correlation filter: A new approach for localization and classification," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 631–643, 2013.
- [42] M. Danelljan, F. S. Khan, M. Felsberg, and J. van de Weijer, "Adaptive color attributes for real-time visual tracking," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1090–1097.
- [43] H. K. Galoogahi, T. Sim, and S. Lucey, "Multi-channel correlation filters," in *IEEE International Conference on Computer Vision*, 2013, pp. 3072–3079.
- [44] R. Patnaik and D. Casasent, "Fast FFT-based distortion-invariant kernel filters for general object recognition," in *IS&T/SPIE Electronic Imaging*, 2009, pp. 725 202–725 202.
- [45] V. N. Boddeti and B. V. Kumar, "Maximum margin vector correlation filter," *arXiv preprint arXiv:1404.6031*, 2014.
- [46] H. Kiani Galoogahi, T. Sim, and S. Lucey, "Correlation filters with limited boundaries," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [47] J. Fernandez and B. V. Kumar, "Zero-aliasing correlation filters," in *International Symposium on Image and Signal Processing and Analysis*, 2013, pp. 101–106.
- [48] C. Ma, X. Yang, C. Zhang, and M.-H. Yang, "Long-term correlation tracking," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [49] T. Liu, G. Wang, and Q. Yang, "Real-time part-based visual tracking via adaptive correlation filters," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [50] R. M. Gray, *Toeplitz and circulant matrices: A review*. Now publishers inc, 2006.
- [51] C.-P. Lee and C.-b. Lin, "A study on l2-loss (squared hinge-loss) multiclass svm," *Neural computation*, vol. 25, no. 5, pp. 1302–1323, 2013.
- [52] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 886–893.
- [53] H. Zhou, Y. Yuan, and C. Shi, "Object tracking using sift features and mean shift," *Computer Vision and Image Understanding*, vol. 113, no. 3, pp. 345–352, 2009.
- [54] K. B. Petersen, M. S. Pedersen *et al.*, "The matrix cookbook," *Technical University of Denmark*, vol. 7, p. 15, 2008.
- [55] R. Rifkin, G. Yeo, and T. Poggio, "Regularized least-squares classification," *NATO Science Series Sub Series III Computer and Systems Sciences*, vol. 190, pp. 131–154, 2003.
- [56] J. Gao, H. Ling, W. Hu, and J. Xing, "Transfer learning based visual tracking with gaussian processes regression," in *European Conference on Computer Vision*, 2014, pp. 188–203.
- [57] C. Bao, Y. Wu, H. Ling, and H. Ji, "Real time robust l1 tracker using accelerated proximal gradient approach," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1830–1837.
- [58] K. Zhang, L. Zhang, and M.-H. Yang, "Real-time compressive tracking," in *European Conference on Computer Vision*, 2012, pp. 864–877.
- [59] M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, L. Čehovin, G. Nebehay, T. Vojšič, G. Fernandez, A. Lukežič, A. Dimitriev *et al.*, "The visual object tracking vot2014 challenge results," in *Workshop of European Conference on Computer Vision*, 2014, pp. 191–217.
- [60] R. Rockafellar, "On the maximal monotonicity of subdifferential mappings," *Pacific Journal of Mathematics*, vol. 33, no. 1, pp. 209–216, 1970.
- [61] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [62] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.
- [63] K. Goebel and W. Kirk, "A fixed point theorem for asymptotically nonexpansive mappings," *Proceedings of the American Mathematical Society*, vol. 35, no. 1, pp. 171–174, 1972.

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