Manifold Regularized Correlation Object Tracking

Hongwei Hu, Bo Ma, *Member, IEEE*, Jianbing Shen, *Senior Member, IEEE*, and Ling Shao, *Senior Member, IEEE*

***Abstract*— In this paper, we propose a manifold regularized correlation tracking method with augmented samples. To make better use of the unlabeled data and the manifold structure of the sample space, a manifold regularization-based correlation filter is introduced, which aims to assign similar labels to neighbor samples. Meanwhile, the regression model is learned by exploiting the block-circulant structure of matrices resulting from the augmented translated samples over multiple base samples cropped from both target and nontarget regions. Thus, the final classifier in our method is trained with positive, negative, and unlabeled base samples, which is a semisupervised learning framework. A block optimization strategy is further introduced to learn a manifold regularization-based correlation filter for efficient online tracking. Experiments on two public tracking data sets demonstrate the superior performance of our tracker compared with the state-of-the-art tracking approaches.**

***Index Terms*— Block circulant, correlation filter, manifold regularization, visual tracking.**

I. Introduction

VISUAL tracking is one of the most important components in a computer vision system. It has been widely used in human–computer interaction, video segmentation, and surveillance [1]–[5], [7]. Target appearance changes caused by variations in illumination, occlusion, deformation, and motion blur have a significant impact on the tracking accuracy. To handle this, numerous tracking approaches [9]–[11] have emerged, and exciting achievements in tracking are obtained on recent tracking data sets. However, as a critical part of tracking, the appearance models of most trackers still lack enough discriminative power to cope with the complicated scenarios during tracking.

Lately, correlation filter-based discriminative visual tracking approaches [6], [8] have achieved great success. Based on existing works, two major observations prompt us to come up with our tracking approach. First, most correlation filter-based tracking methods learn a kernelized ridge regression using only labeled samples, and have no consideration for the intrinsic geometrical manifold structure of the high-dimensional feature space stemming from labeled and unlabeled samples. Based on the manifold assumption, the high-dimensional data are locally smooth in the embedded manifold space, which is locally homeomorphic to the Euclidean space. Under this assumption of feature space in visual tracking, it is believed that a well-learned classifier should assign similar labels for samples close to each other in the manifold space. For regression problems, it implies that the regression function values of these samples are similar. Therefore, in this paper, to further improve the classifier performance, we construct an adjacency graph with an affinity matrix, which preserves the spatial manifold structure of the feature space, and introduce the Laplacian regularized least squares algorithm as the classifier to impose the manifold assumption on the learning model.

Second, almost all correlation filter-based trackers train their classifiers using the translated samples generated from one single positive base sample by taking advantage of the circulant structure theory. But the performance of a classifier learned with these samples only may be suboptimal, since the number of samples is small actually. For example, a base sample with size 32×32 could only generate 1024 translated samples, if these samples are represented by 1024-D intensity features. This may overfit the learned model, and make it susceptible to appearance changes caused by background clutter, similar objects, and fast motion. We observe that the nearly endless base negative samples around the target region are neglected, and these negative samples may be helpful to train a more discriminative classifier, although training with all these available samples is impractical. Negative samples are commonly used in many discriminative tracking methods. They usually crop negative samples around target region according to a certain distribution, such as Gaussian function, and all of these negative samples will be used for classifier training. But the collected negative samples are much redundant samples by these methods. In contrast, we make use of the shift transform of samples and an amount of negative samples for training the classifier. In fact, the performance of a classifier trained with only several base samples based on diagonalization of circulant matrix is similar to the one trained with plenty of generated negative samples [12].

In order to take advantage of these negative samples, we introduce an efficient detector learning method with translated samples from a positive base sample and multiple negative base samples. Since more samples are collected compared with original correlation filter tracking methods, we name these translated samples the augmented samples. Basic kernelized correlation filter (KCF) trackers [6] produce samples from only one base sample, which includes information from both

Manuscript received August 1, 2016; revised January 24, 2017; accepted March 25, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61472036 and Grant 61272359, in part by the National Basic Research Program of China (973 Program) under Grant 2013CB328805, and in part by the Specialized Fund for the Joint Building Program of Beijing Municipal Education Commission. *(Corresponding Author: Bo Ma.)*

H. Hu, B. Ma, and J. Shen are with the Beijing Laboratory of Intelligent Information Technology, School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China (e-mail: bma000@bit.edu.cn; shenjianbing@bit.edu.cn).

L. Shao is with the School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, U.K. (e-mail: ling.shao@uea.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2017.2688448

foreground and background, by applying shift transformation. The shift operation can well model target translation with a uniform background. But with a cluttered background, the samples obtained by the shift operation may not handle the realistic scenario well. For these training samples, the trained regressor can echo with a regression value, indicating how far away the sample is from the target center, but may not respond to negative samples with small values. Obviously, the use of negative samples can alleviate the distraction from the background for trackers and reduce the ambiguity caused by simultaneously shifting foreground and background. Hence, we mine the abundant negative samples keeping certain distance from target, and assign them with very small regression values. The Gram matrix derived from these augmented samples exhibits a block-circulant structure, which enables a linear regression model to be learned efficiently in the Fourier domain [12]. Since Laplacian regularization is utilized in our method, it is crucial to build a block-circulant structural Laplacian matrix. Fortunately, the Laplacian matrix with such a structure can be obtained by a well-designed affinity matrix. Consequently, the final manifold regularized regression model with augmented samples could be solved by an efficient block optimization method with a block diagonalization technique. Though the block-circulant kernel matrix has been investigated in object detection [12], it is new in visual tracking. We believe that effectively mining negative samples is very important for visual tracking, since they may distract the tracker.

In this paper, we present a manifold regularized correlation tracking method with augmented samples (MRCT-AS). Almost all the existing correlation filter-based trackers train their classifiers using one single positive base sample, which neglects the negative samples around target region. Thus, in our tracker, we train the classifier with positive, negative, and unlabeled base samples, and the collection of samples generated by these base samples is named as the augmented samples set. The final result is obtained using only these base samples under diagonalization of block-circulant matrix. Thus, we introduce a block method to solve the final optimization problem. In addition, we introduce the manifold regularization term, which aims to assign similar labels to neighbor samples, into the original ridge regression objective function [6]. Thus, the final classifier in our method is trained under a semi-supervised learning framework. The proposed algorithm is verified on two popular tracking benchmarks: OTB-2013 [13] and TB-100 [14]. The proposed tracker achieves better results compared with the state-of-the-art approaches. Our source code and supplemental materials will be available online.[1]

Compared with the existing visual tracking methods, the proposed approach offers the following main contributions.

1) We present a method to predict the regression values of unlabeled samples based on the manifold assumption of feature space under the correlation filtering framework in a semisupervised manner.
2) In order to collect plenty of training samples, we introduce an augmented sample set generation technique from image patches cropped from both target and nontarget regions, leading to a more discriminative regression model for visual object tracking.
3) A block method for solving the introduced optimization problem is presented, and it results in an efficient learning model.

[1] *http://github.com/shenjianbing/mrctrack*

II. Related Works

A comprehensive survey of tracking algorithms [15] is beyond the scope of this paper, and the following is a brief review of the works that are closely related to ours.

A. Conventional Tracking

Traditional tracking-by-detection-based tracking methods consider tracking as a binary classification problem. The most well-known tracking algorithm using tracking-by-detection was proposed by Kalal *et al.* [16]. They decomposed the tracking problem into three steps: tracking, learning, and detection, which was robust to long-term tracking videos. However, the tracking performance of this approach depends on the labeled binary samples, which are used to update the detector. The phenomenon of tracking drift will occur, since the slight inaccuracy would lead to improper sample labels. To handle this situation, Babenko *et al.* [17] presented a tracking method using multiple instance learning method to avoid these problems. Hare *et al.* [18] introduced a tracking framework based on the structured output prediction and the SVMs theory. Their method could avoid the requirement for classification and show good tracking performance. However, these methods consider the target appearance probability as a certain distribution function, which is difficult to deal with appearance changes. Gao *et al.* [19] formulated this probability with Gaussian processes regression to propose a semisupervised tracking framework. Oron *et al.* [20] presented a locally orderless visual object tracker by modeling the target variations with the distance of earth movers. To construct a more robust target appearance model, Chen *et al.* [21] represented object information with complex cells extracted from local descriptors. Despite of these tracking-by-detection methods, some generative tracking methods also perform well for tracking task. For example, Wang and Lu [22] proposed a generative tracking method by using the probability continuous outlier model. Zhong *et al.* [23] proposed a tracking approach by combining both discriminative and generative model based on sparse representations. With the occlusion handling and update strategy, their trackers [43], [44] could deal with appearance variation effectively. These trackers perform well during tracking, but they are far from satisfactory by testing on many challenging videos.

B. Correlation Filter-Based Tracking

Correlation filtering is a valid technique for various tasks, such as recognition [24] and detection [25]. As a pioneer work, Mahalanobis [26] designed an object tracker based on the minimum average correlation energy filter. But this paper did not receive enough attention until an adaptive correlation filter-based tracking method [27]. They presented a tracking

approach using the minimum output sum of squared error filter. Henriques *et al.* [6], [8] exploited the circulant structure to learn a discriminative classifier with a much more efficient diagonalization technique. However, these tracking methods are limited to only determine the target location, and are infeasible to handle difficulties, such as scale changes and rotation. Therefore, lots of efforts have been proposed to address these issues. For example, Li and Zhu [28] solved the fixed template size problem using a scale adaptive scheme, and a feature integration strategy was introduced to further improve tracking performance. Danelljan *et al.* [29] estimated the target scale by learning multiple scale space correlation filters. Zhang *et al.* [30] predicted the target scale and located the target position with an online updating method considering the appearance model of all the previous frames. To handle orientation changes of the target appearance, Du *et al.* [31] trained different orientation-specified models using rotated images from the original target region. Different from addressing the rotation and scale change problems, researchers proposed various methods to improve tracking robustness. Adaptive low-dimensional color attributes [32] and hierarchical convolutional features [33] were applied in a correlation filter-based tracking approach to improve the tracking accuracy and robustness. Danelljan *et al.* [34] introduced spatially regularized discriminative correlation filters to penalize coefficients of the correlation filters according to their spatial positions. Tang and Feng [35] expanded the single kernel of the original correlation filter to the multikernel case. To better deal with the occlusion problem, Liu *et al.* [36] employed a part-based method while Ma *et al.* [37] introduced a redetection strategy on the random fern classifier. Despite these efforts made in correlation filter tracking, few efforts are paid to the spatial manifold structure of both labeled and unlabeled samples.

C. Manifold Regularization-Based Tracking

Manifold regularization falls into the semisupervised learning framework with both labeled and unlabeled samples in [38] and [39]. It is popularized by locally linear embedding [40] and spectral clustering [41]. Manifold regularization has been widely used in visual tracking. To exploit the geometrical structure of the feature space, researchers often construct a Laplacian graph using labeled and unlabeled samples. For example, Bai and Tang [42] estimated the target location by an online Laplacian ranking support vector tracker. Zhuang *et al.* [45] constructed a discriminative sparse similarity map based on Laplacian multitask reverse sparse representation. Ma *et al.* [46] introduced globally linear approximation to nonlinear learning for visual tracking. Unlike correlation filter-based trackers, the samples in these methods are handcrafted or selected from a set of examples with certain heuristic strategies, and failed to utilize the abundant samples during tracking.

III. Manifold Regularized Correlation Tracking With Augmented Samples

A. Manifold Regularized Least Squares

We now focus on the Laplacian regularized least squares algorithm [38], since it provides a closed-form solution to the optimization problem, and considers the geometrical structure of samples. Given l labeled samples $\{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$ and u unlabeled samples $\{\mathbf{x}_i\}_{i=l+1}^{l+u}$, the training aims to find an optimal classification function f^* in reproducing kernel Hilbert space (RKHS) $\mathcal{H}_\kappa$ by minimizing

$$f^* = \arg\min_{f\in\mathcal{H}_\kappa} \frac{1}{l}\sum_{i=1}^{l}(f(\mathbf{x}_i)-y_i)^2 + \lambda\|f\|_\kappa^2 + \frac{\gamma}{n^2}\sum_{i,j=1}^{n}(f(\mathbf{x}_i)-f(\mathbf{x}_j))^2 W_{ij} \tag{1}$$

$$= \arg\min_{f\in\mathcal{H}_\kappa} \frac{1}{l}\sum_{i=1}^{l}(f(\mathbf{x}_i)-y_i)^2 + \lambda\|f\|_\kappa^2 + \frac{\gamma}{n^2}\mathbf{f}^T\mathbf{L}\mathbf{f} \tag{2}$$

where $\|\cdot\|_\kappa$ is the norm induced by a Mercer kernel κ in RKHS $\mathcal{H}_\kappa$, $n = l+u$ the number of all samples, $\mathbf{W}$ an affinity matrix with each element W_{ij} denoting the similarity weight of samples $\mathbf{x}_i$ and $\mathbf{x}_j$, $\mathbf{f} = [f(x_1), f(x_2), \ldots, f(x_n))]^T$, and $\mathbf{L}$ the Laplacian matrix calculated by $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{D}$ is a diagonal matrix with each diagonal element $D_{ii} = \sum_{j=1}^{n} W_{ij}$. λ and γ are two balance factors that control the influences of overfitting and manifold regularization.

According to the representer theorem, the regression value of a sample $\mathbf{v}$ can be represented by

$$f^*(\mathbf{v}) = \sum_{i=1}^{n}\alpha_i\kappa(\mathbf{v},\mathbf{x}_i) \tag{3}$$

where α_i is the ith element of n-dimensional variable $\boldsymbol{\alpha}$, which has the following solution:

$$\boldsymbol{\alpha} = \left(\mathbf{J}\mathbf{K} + \lambda l\mathbf{I} + \frac{\gamma l}{n^2}\mathbf{L}\mathbf{K}\right)^{-1} Y \tag{4}$$

where $\mathbf{J} = \text{diag}(1,\ldots,1,0,\ldots,0) \in \mathbb{R}^{n\times n}$ with the first l diagonal elements set to 1, $\mathbf{I} \in \mathbb{R}^{n\times n}$ is an identity matrix, $Y = [y_1,\ldots,y_l,0,\ldots,0]^T \in \mathbb{R}^n$, and $\mathbf{K}$ is an $n\times n$ Gram matrix with element $K_{ij} = \kappa(\mathbf{x}_i,\mathbf{x}_j)$. Note that the regression problem degenerates to standard regularized least squares when γ is set to 0. The detailed derivation of (4) refers to the Appendix.

B. Augmented Samples

Given a set of base samples $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n\}$ with $\mathbf{x}_i \in \mathbb{R}^s$, the augmented samples are obtained by translating each base sample $\mathbf{x}_i$ using permutation matrix

$$\mathbf{P} = \begin{bmatrix} \mathbf{0}_{s-1}^T & 1 \\ \mathbf{I}_{s-1} & \mathbf{0}_{s-1} \end{bmatrix} \tag{5}$$

where $\mathbf{0}_{s-1}$ is a column vector with $s-1$ zeros and $\mathbf{I}_{s-1}$ is an identity matrix with size $(s-1)\times(s-1)$. Thus, the augmented sample set is constructed by $\mathcal{X} = \{\mathbf{P}^{t-1}\mathbf{x}_i \mid i = 1,\ldots,n; t = 1,\ldots,s\}$, where $\mathbf{P}^t$ is the tth power of $\mathbf{P}$, and $\mathbf{P}^t\mathbf{x}$ means a cyclic shift of sample $\mathbf{x}$.

Most correlation filter-based tracking methods have no consideration for the negative samples around the target region. In our setting, we crop one base sample from the target region

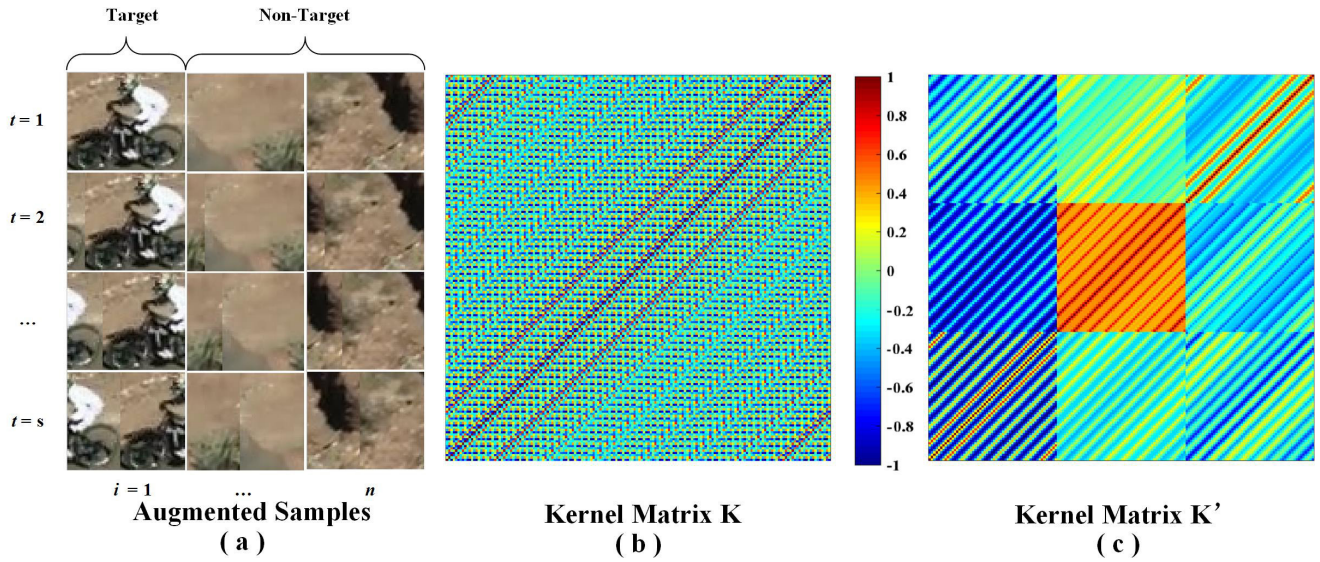


Fig. 1. Structure of the kernel matrix using augmented samples. (a) Three augmented samples that we used to train the classifier. Samples in the first row contain one target image and several background images. (b) Heat map of the Gram matrix $\mathbf{K}$ with block-circulant structure when the sample matrix is organized as (6). (c) Heat map of kernel matrix $\mathbf{K}'$ with circulant blocks when the sample matrix is calculated by (8).

and multiple negative base samples from nontarget regions to generate our augmented samples. Fig. 1(a) shows the samples used in our training procedure. It is clear that all these samples are reconstructed by the base samples in the first row.

C. Block-Circulant Structure of $\mathbf{K}$ and $\mathbf{L}$

With these augmented samples, we show that the resulting kernel matrix and the Laplacian matrix exhibit a very interesting structure. Let

$$\mathbf{X} = [\mathbf{P}^1\mathbf{x}_1, \ldots, \mathbf{P}^1\mathbf{x}_n, \ldots, \mathbf{P}^s\mathbf{x}_1, \ldots, \mathbf{P}^s\mathbf{x}_n]^T \tag{6}$$

denote the sample matrix consisting of all augmented samples. For two samples $\mathbf{x}^{(a)}, \mathbf{x}^{(b)}$ from $\mathbf{X}$, we suppose that they are obtained by translating two base samples $\mathbf{x}_i, \mathbf{x}_j$ with $\mathbf{P}^u, \mathbf{P}^v$, respectively, i.e., $\mathbf{x}^{(a)} = P^u\mathbf{x}_i$ and $\mathbf{x}^{(b)} = P^v\mathbf{x}_j$. We denote one element in the kernel matrix $\mathbf{K}$ of (4) as $K_{(u,v),(i,j)} = \kappa(\mathbf{x}^{(a)}, \mathbf{x}^{(b)})$. Assuming that the Gaussian kernel is chosen as a kernel function, it leads to

$$\begin{aligned} k(\mathbf{x}^{(a)}, \mathbf{x}^{(b)}) &= \exp\left(-\frac{\|\mathbf{x}^{(a)} - \mathbf{x}^{(b)}\|^2}{\sigma^2}\right) \\ &= \exp\left(-\frac{\|\mathbf{x}_i\|^2 + \|\mathbf{x}_j\|^2 - 2\mathbf{x}_i^T P^{v-u}\mathbf{x}_j}{\sigma^2}\right) \end{aligned} \tag{7}$$

where σ is a constant that determines the width of the Gaussian kernel. Since each element in kernel matrix $\mathbf{K}$ depends on $v - u$, $\mathbf{K}$ is a block-circulant matrix [12, Th. 1]. Fig. 1(b) shows the heat map of $\mathbf{K}$. Thus, $\mathbf{K}$ has only $n^2 \times s$ different elements.

Note that if the sample matrix is organized as

$$\mathbf{X} = [\mathbf{P}^1\mathbf{x}_1, \ldots, \mathbf{P}^s\mathbf{x}_1, \ldots, \mathbf{P}^1\mathbf{x}_n, \ldots, \mathbf{P}^s\mathbf{x}_n]^T \tag{8}$$

where the kernel matrix $\mathbf{K}'$ (distinguishing it from $\mathbf{K}$) will form a circulant structure for each block, as shown in Fig. 1(c). Here, n and s are set to 3.

Similar to kernel $\mathbf{K}$, the structure of $\mathbf{L}$ in (2) is a block-circulant matrix as well with an appropriate distance metric function for samples. Fortunately, a simple radial basis function will lead to a desired structure of $\mathbf{L}$. We rewrite the element of similarity matrix $\mathbf{W}$ in (1) as $W_{(u,v),(i,j)}$ that represents the affinity of two samples $\mathbf{x}^{(a)}, \mathbf{x}^{(b)}$

$$W_{(u,v),(i,j)} = \exp\left(-\frac{\|\mathbf{x}_i\|^2 + \|\mathbf{x}_j\|^2 - 2\mathbf{x}_i^T P^{v-u}\mathbf{x}_j}{\varrho^2}\right) \tag{9}$$

where ϱ is a constant. Obviously, $\mathbf{W}$ is a block-circulant matrix, since it exhibits the same structure as kernel matrix $\mathbf{K}$. And because of the diagonal structure of $\mathbf{D}$, the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is still a block-circulant matrix.

D. Block Learning Strategy

Our goal is to calculate the dual variable $\boldsymbol{\alpha}$ in (4), so the regression value of an unlabeled sample $\mathbf{v}$ could be obtained by (3). For efficiency, the block-circulant matrices $\mathbf{K}$ and $\mathbf{L}$ will be decomposed into the block-diagonal forms [see (10) and (12)]. Thus, a more compact form of $\boldsymbol{\alpha}$ will be represented as (18). A block-circulant matrix $\mathbf{K}$ could be decomposed into a block-diagonal form [12]

$$\mathbf{K} = \mathbf{U}^{-1}\bar{\mathbf{K}}\mathbf{U} = \mathbf{U}^{-1}\mathbf{Diag}\left(\bar{\mathbf{K}}(1), \bar{\mathbf{K}}(2), \ldots, \bar{\mathbf{K}}(s)\right)\mathbf{U} \tag{10}$$

where $\bar{\mathbf{K}}(f)$ is the fth block of $\bar{\mathbf{K}}$ with each element $\bar{K}_{ij}(f) = \bar{k}_f(i, j)$, $\mathbf{Diag}(\cdot)$ the block-diagonal operator, and $\mathbf{U} = \mathbf{F}_s \otimes \mathbf{I}_n$, where $\otimes$ is the Kronecker product, $\mathbf{I}_n$ is an $n \times n$ identity matrix, and $\mathbf{F}_s$ is the discrete Fourier transform matrix of size $s \times s$. Since the Fourier transform of vector $\mathbf{x}$ could be represented as the multiplication of $\mathbf{F}_s$ by $\mathbf{x}$, i.e., $\mathcal{F}(\mathbf{x}) = \mathbf{F}_s\mathbf{x}$,

the elements of $\mathbf{K}$ can be obtained by Fourier transform. For each element in the matrix block, we have $\bar{K}_{ij}(f) = \bar{k}_f(i, j)$, where $\bar{k}_f(i, j)$ is the fth element of vector $\bar{\mathbf{k}}(i, j)$. Let $\mathbf{k}(i, j) = [k_1(i, j), k_2(i, j), \ldots, k_s(i, j)]^T$ denote the kernel vector of samples $\mathbf{x}_i$ and $\mathbf{x}_j$, which is actually one block of $\mathbf{K}'$, as shown in Fig. 1(c). Then, vector $\bar{\mathbf{k}}(i, j)$ can be calculated as

$$\bar{\mathbf{k}}(i, j) = \mathcal{F}(\mathbf{k}(i, j)). \tag{11}$$

Similarly, the Laplacian matrix $\mathbf{L}$ can be decomposed as

$$\mathbf{L} = \mathbf{U}^{-1}\bar{\mathbf{L}}\mathbf{U} = \mathbf{U}^{-1}\mathbf{Diag}\left(\bar{\mathbf{L}}(1), \bar{\mathbf{L}}(2), \ldots, \bar{\mathbf{L}}(s)\right)\mathbf{U}. \tag{12}$$

Considering the manifold regularization-based least squares problem in (2), we now introduce the blockwise learning algorithm with augmented samples. For simplicity, we define $\delta = \lambda l$ and $\eta = (\gamma l/n^2)$, and (4) can be rewritten as

$$\boldsymbol{\alpha} = (\mathbf{JK} + \delta\mathbf{I} + \eta\mathbf{LK})^{-1}Y \tag{13}$$

where the block-diagonal matrix

$$\mathbf{J} = \mathbf{Diag}\left(\mathbf{J}_n, \mathbf{J}_n, \ldots, \mathbf{J}_n\right) \in \mathbb{R}^{ns\times ns} \tag{14}$$

$$\mathbf{J}_n = \begin{bmatrix} \mathbf{I}_l & \\ & \mathbf{0}_u \end{bmatrix} \in \mathbb{R}^{n\times n} \tag{15}$$

with identity matrix $\mathbf{I}_l \in \mathbb{R}^{l\times l}$ and zero matrix $\mathbf{0}_u \in \mathbb{R}^{u\times u}$. Thus, it is not difficult to verify that $\mathbf{J} = \mathbf{U}^{-1}\mathbf{J}\mathbf{U}$. And $Y = [\mathbf{y}(1)^T, \mathbf{y}(2)^T, \ldots, \mathbf{y}(s)^T]^T \in \mathbb{R}^{ns\times 1}$ with each element $\mathbf{y}(f) \in \mathbb{R}^{n\times 1}$ representing the regression value vector of all base samples under translating $\mathbf{P}^f$.

The regression values of all the translated samples of negative base samples are set to zero, and those values of target samples refer to [8]. Therefore, the solution in (13) could be written as

$$\begin{aligned}\boldsymbol{\alpha} &= (\mathbf{U}^{-1}\mathbf{J}\mathbf{U}\mathbf{U}^{-1}\bar{\mathbf{K}}\mathbf{U} + \delta\mathbf{U}^{-1}\mathbf{I}\mathbf{U} \\ &\quad + \eta\mathbf{U}^{-1}\bar{\mathbf{L}}\mathbf{U}\mathbf{U}^{-1}\bar{\mathbf{K}}\mathbf{U})^{-1}Y \\ &= \mathbf{U}^{-1}\mathbf{Diag}\left(\Gamma(1), \Gamma(2), \ldots, \Gamma(s)\right)^{-1}\mathbf{U}Y \qquad (16)\\ \Gamma(f) &= \mathbf{J}_n\bar{\mathbf{K}}(f) + \delta\mathbf{I}_n + \eta\bar{\mathbf{L}}(f)\bar{\mathbf{K}}(f). \qquad (17)\end{aligned}$$

According to the lemma of Kronecker products for a linear matrix, we have $\mathbf{U}Y = (\mathbf{F}_s \otimes \mathbf{I}_n)Y = \mathrm{vec}((\mathbf{F}_s\mathbf{Y}_{n\times s}^T)^T)$ [47, Lemma 4.3.1], where $\mathrm{vec}(\mathbf{X})$ means the vectorization of a matrix $\mathbf{X}$, i.e., returning a column vector containing all of the elements in $\mathbf{X}$, and $\mathbf{Y}_{n\times s} = \left[\mathbf{y}(1), \ldots, \mathbf{y}(s)\right] \in \mathbb{R}^{n\times s}$. In other words, $\mathbf{U}Y$ is calculated by executing Fourier transformation for the label vector of every base sample under all possible cyclic shifts. Let $\mathbf{y}_i$ denote the label vector of the ith base sample with all possible transformations, and $\bar{Y} = \mathbf{U}Y = [\bar{Y}(1)^T, \bar{Y}(2)^T, \ldots, \bar{Y}(s)^T]^T$. Then, we have $\bar{Y}_i(f) = \mathcal{F}(\mathbf{y}(f))_i$. Consequently, (16) is written as

$$\boldsymbol{\alpha} = \mathbf{U}^{-1}\begin{bmatrix} \Gamma(1)^{-1} & - & Y(1) \\ \Gamma(2)^{-1} & - & Y(2) \\ & \vdots & \\ \Gamma(s)^{-1} & - & Y(s) \end{bmatrix}. \tag{18}$$

Multiplying (18) by $\mathbf{U}$ on both sides, we get

$$\bar{\boldsymbol{\alpha}} = \mathbf{U}\boldsymbol{\alpha} = \begin{bmatrix} \bar{\boldsymbol{\alpha}}(1) \\ \bar{\boldsymbol{\alpha}}(2) \\ \vdots \\ \bar{\boldsymbol{\alpha}}(s) \end{bmatrix} = \begin{bmatrix} \Gamma(1)^{-1} & - & Y(1) \\ \Gamma(2)^{-1} & - & Y(2) \\ & \vdots & \\ \Gamma(s)^{-1} & - & Y(s) \end{bmatrix}. \tag{19}$$

Therefore, the learning of $\boldsymbol{\alpha}$ in the proposed manifold regularized least squares is converted to solve $\bar{\boldsymbol{\alpha}}$. The solving of the latter is decomposed into s subproblems in a block form, and each block could be calculated independently.

1) Complexity Analysis: In our implementation, the size of kernel matrix $\mathbf{K}$ is $(sn) \times (sn)$. In order to calculate $\boldsymbol{\alpha}$ in (4), we need to compute the inverse of a $(sn)\times(sn)$ matrix, and its computational complexity is $O\left((sn)^3\right)$. However, in (19), the calculation is divided into s subproblems. Thus, the complexity for computing the inverse matrices of these subproblems is $s \times O(n^3)$. Thus, the computational complexity of (4) is larger than (19).

E. Fast Block Detection

Given a base sample $\mathbf{z}$, according to (3), the regression values of its transformations can be computed as

$$\mathbf{f}(\mathbf{z}) = (\mathbf{K}^z)^T\boldsymbol{\alpha} \tag{20}$$

where $\mathbf{K}^z$ is a kernel matrix for the cyclic shifts of $\mathbf{z}$ and all augmented samples. Assuming that $\mathcal{Z} = \{\mathbf{x}_i\}_{i=l+1}^{l+u}$ denotes all the unlabeled base samples, it is easy to prove that $\mathbf{K}^z \in \mathcal{R}^{(ns)\times(us)}$ is a block-circulant matrix, and can be block-diagonalized as

$$\mathbf{K}^z = \mathbf{U}^{-1}\bar{\mathbf{K}}^z\mathbf{U} \tag{21}$$

$$\bar{\mathbf{K}}^z = \mathbf{Diag}\left(\bar{\mathbf{K}}^z(1), \bar{\mathbf{K}}^z(2), \ldots, \bar{\mathbf{K}}^z(s)\right). \tag{22}$$

Then, (20) can be decomposed as

$$\mathbf{f}(\mathbf{z}) = (\mathbf{U}^{-1}\bar{\mathbf{K}}^z\mathbf{U})^T\boldsymbol{\alpha} = \mathbf{U}^{-1}\begin{bmatrix} (\bar{\mathbf{K}}^z(1))^T & - & \boldsymbol{\alpha}(1) \\ (\bar{\mathbf{K}}^z(2))^T & - & \boldsymbol{\alpha}(2) \\ & \vdots & \\ (\bar{\mathbf{K}}^z(s))^T & - & \boldsymbol{\alpha}(s) \end{bmatrix}. \tag{23}$$

Left multiplying both sides of the above-mentioned equation by $\mathbf{U}$ gives rise to

$$\bar{\mathbf{f}}(\mathbf{z}) = \mathbf{U}\mathbf{f}(\mathbf{z}) = \begin{bmatrix} (\bar{\mathbf{K}}^z(1))^T & - & \boldsymbol{\alpha}(1) \\ (\bar{\mathbf{K}}^z(2))^T & - & \boldsymbol{\alpha}(2) \\ & \vdots & \\ (\bar{\mathbf{K}}^z(s))^T & - & \boldsymbol{\alpha}(s) \end{bmatrix} \tag{24}$$

where $f(\mathbf{z})$ can be calculated rapidly by

$$\mathbf{f}(\mathbf{z}) = \mathbf{U}^{-1}\bar{\mathbf{f}}(\mathbf{z}) \tag{25}$$

$$\mathbf{U}^{-1} = (\mathbf{F}_s \otimes \mathbf{I}_n)^{-1} = \mathbf{F}_s^{-1} \otimes \mathbf{I}_n. \tag{26}$$

According to the property of Kronecker product, $\mathbf{F}_s^{-1}$ is the matrix of inverse discrete Fourier transform.

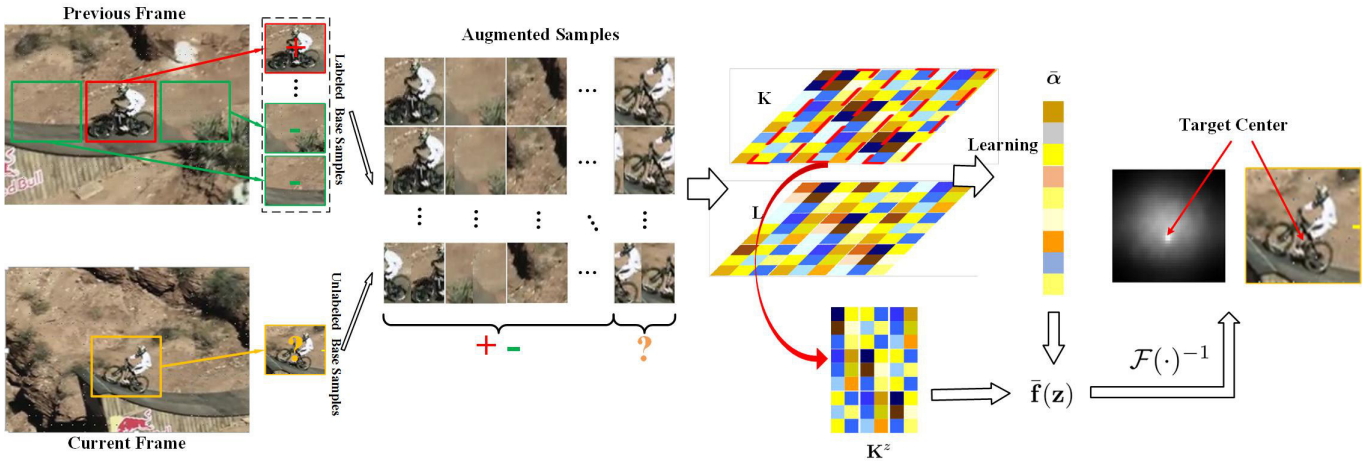


Fig. 2. Flowchart of the proposed manifold regularized correlation tracking with augmented samples. The goal of the method is to estimate the target center in the current frame, and actually, the test kernel matrix $\mathbf{K}^z$ has been calculated in $\mathbf{K}$. The final location center is determined as the position of the max response in the final confidence map.

F. Proposed Tracking Framework

Fig. 2 shows the flow of our tracking method. To collect labeled base samples, we crop one positive sample centered at the target region and several negative samples around target region. The unlabeled base sample is obtained from the current frame centered at the target location of the previous frame. As other correlation-based trackers do, we adopt padding for each base sample, which means the base sample is bigger than a target region. Then, we could calculate the kernel matrix $\mathbf{K}$ and the Laplacian matrix $\mathbf{L}$ using the augmented sample set generated from all base samples. Afterward, the solution of $\bar{\boldsymbol{\alpha}}$ is computed according to (19). To determine the target center in the current frame, we compute a confidence map based on the learned model for the unlabeled base sample under different cyclic shifts. The final target center is estimated as the position with the max response in this map.

1) Model Updating: The target appearance model changes during tracking due to factors, such as illumination changes, occlusion, and deformation. It is significant for the model parameter $\bar{\boldsymbol{\alpha}}$ to adapt to the current target appearance. Our model is updated with a learning rate constant ϵ in every frame once the state of the current target is determined. Consequently, the updating rule can be written as

$$\bar{\boldsymbol{\alpha}}_m^* = (1-\epsilon)\bar{\boldsymbol{\alpha}}_{m-1}^* + \epsilon\bar{\boldsymbol{\alpha}}_m \tag{27}$$

where m is the frame index of the current frame.

IV. Experimental Results

A. Experimental Setup

Our tracking method (named MRCT-AS) is verified on two popular public tracking benchmark data sets: OTB2013 [13] and TB=100 [14], and the tracking performance is compared with several state-of-the-art trackers. In our experiments, the correlation filter is trained by combining histogram of oriented gradient [48], intensity, and illumination invariant features (IIF) [49]. The balance factors λ and γ in (2) are set to 10^{-9} and 10^{-7}, respectively. The base sample padding is set to 1.5. And we set the learning rate ϵ in (27) to 0.01 for slow updating. The parameters σ in the Gaussian kernel and ϱ in (9) are set to constant 1. Each base sample is padded to make it bigger than a target region, and the size of padding is 1.5 in our experiments. All the parameters in our tracking algorithm are fixed for different video sequences, which demonstrate the robustness and stability of our method. Note that the selection of these parameters is done by cross validation, which is implemented by manually adjusting one parameter with others fixed on the total OTB-2013 benchmark with 51 video sequences.

B. Discussion and Analysis

The proposed manifold regularization correlation filter considers the similarities of different samples regardless of labeled or unlabeled ones. The estimated regression value of each unlabeled sample is more accurate than the original KCF method. Moreover, we enable the regressor to possess more discriminative power by involving negative base samples into the training procedure. Therefore, the proposed tracking approach is believed to have good handling capacity on difficulties encountered during tracking, such as occlusion, background cluttering, and deformation. Different methods are tested on OTB-2013 (refer to Section IV-C) including ours. The tracking performances of different trackers are evaluated by two criteria. The precision plot calculates the distance precision (DP) percentage with the center location errors of trackers within certain thresholds for successfully tracked frames, and the success plot shows the overlap precision (OP) percentage for overlap rates of tracking results.

To evaluate the influence of manifold regularization and negative base samples, which are the two main components of our method on tracking performance, we compare our results with manifold regularization correlation tracking without negative base samples (MRCT), the correlation tracking without manifold regularization (CT-AS), the original correlation filter-based tracking KCF [8], and KCF with the features we used (KCF + IIF). The comparison results are shown in Fig. 3 (left).

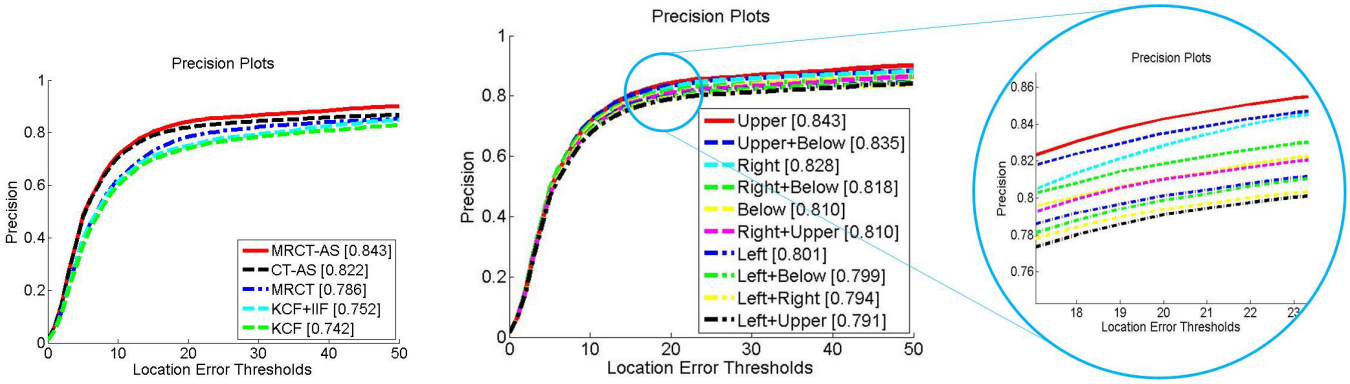


Fig. 3. Left: performance comparison on OTB-2013 with components based on our method. Right: analysis of different selections of negative base samples. For clarity, a partial, enlarged view of this figure is displayed.

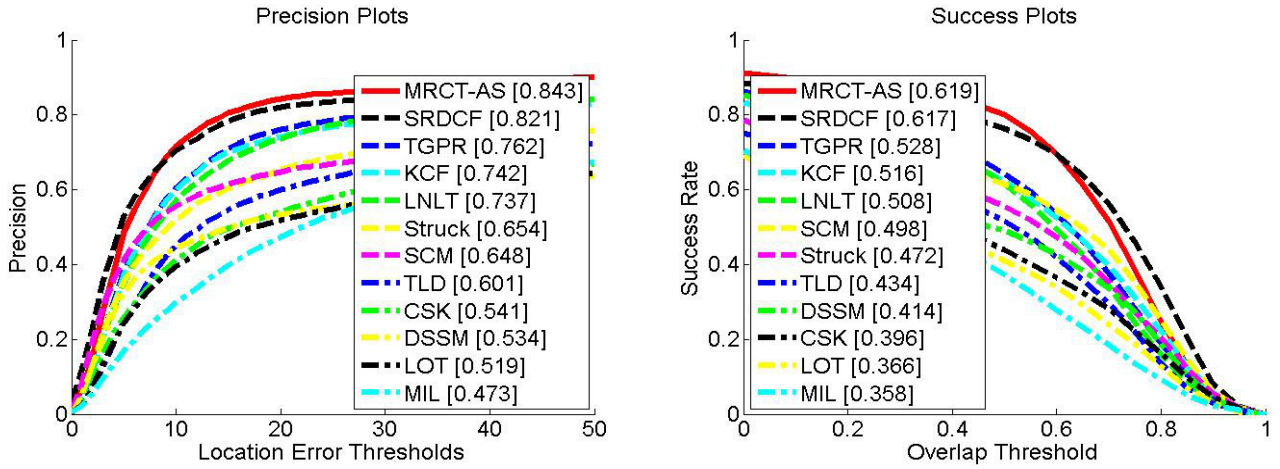


Fig. 4. Results on OTB-2013. Overall performance comparison of DP (left) and OP (right) of different trackers. The overall performance scores of DP at 20 pixels are presented in the legend.

TABLE I
PERFORMANCE SCORES OF ALL THESE TRACKERS ON VTB-2013

	TGPR	KCF	Struck	SCM	CSK	LNLT	SRDCF	TLD	MIL	LOT	DSSM	MRCT-AS
DP	0.762	0.742	0.654	0.648	0.541	0.737	0.821	0.601	0.473	0.519	0.534	**0.843**
OP	0.528	0.516	0.472	0.498	0.396	0.508	0.617	0.434	0.358	0.366	0.414	**0.619**

From Fig. 3, we can see that the performance score of the original KCF is 74.2%, and KCF + IIF is 75.2%. The MRCT improves the score to 78.6%, and CT-AS gets 82.2%. Our method further promotes the score to 84.3%. From Fig. 3, one can see that the two contributions can bring a significant increase on tracking performance for correlation filter-based trackers.

1) Selection of Negative Base Samples: Since different selections of negative base samples have a certain impact on tracking performance, we test our method with different negative base samples, and choose appropriate negative base samples based on the performance. We crop image patches adjacent to the target region (without overlap) in four different directions, i.e., left, right, top, and bottom. As shown in Fig. 2 (top-left), the two green boxes show the left and right negative base samples. For the tracking speed, we consider only two negative base samples at most, which gives us a total of ten combinations of samples. As shown in Fig. 3 (right), the precision plots of all these combinations are listed. From Fig. 3, it is observed that the best performance is obtained when we select a negative base sample located above the target region.

C. Comparison Results on OTB-2013

Our method is tested on OTB-2013 [13] with 51 sequences, which contain different tracking difficulties. The testing video clips and ground truth of all clips are available online *(http://visual-tracking.net/)*. In order to evaluate the robustness of these trackers, we adopt the precision plot and success plot as metrics for evaluation. The results of our method are compared with different state-of-the-art tracking approaches, including Struck [18], SCM [23] (which obtain the top two performances out of more than 20 trackers on this data set), MIL [17], LOT [20], TLD [16], two correlation filter-based tracking methods (KCF [8] and CSK [6]), three newly

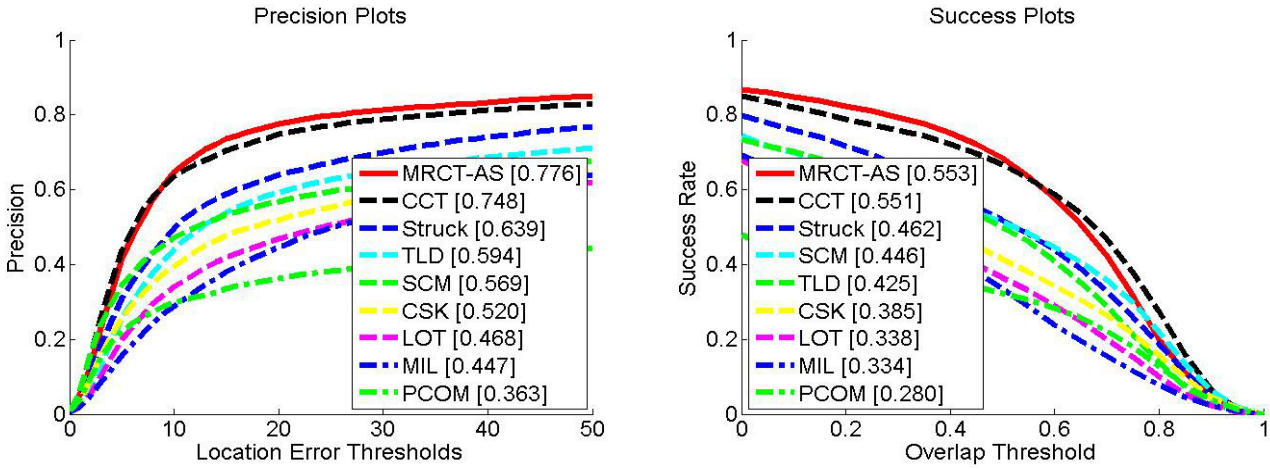


Fig. 5. Results on TB100. Overall performance comparison of DP (left) (the overall performance scores at 20 pixel are presented in the legend) and OP (right) of different trackers.

TABLE II
PERFORMANCE SCORES OF ALL THESE TRACKERS ON TB-100.

	CCT	TLD	Struck	SCM	CSK	PCOM	MIL	LOT	MRCT-AS
DP	0.748	0.594	0.639	0.569	0.520	0.363	0.447	0.468	**0.776**
OP	0.551	0.425	0.462	0.446	0.385	0.280	0.334	0.338	**0.553**

proposed methods (TGPR [19], SRDCF [34], and LNLT [46]), and one manifold regularization-based method (DSSM [45]).

We compare the results of the proposed tracker with other tracking methods, and the precision and success plots are shown in Fig. 4. To sort these trackers, the overall center location error performance scores at 20 pixels and the overlap rate calculated under curves are treated as criteria for ranking as usual. As shown in Fig. 4, we show the performance scores in the legend of precision plots and success plots, respectively. For clarity, we list these performance scores in Table I. It is observed that our tracking approach MRCT-AS outperforms the state-of-the-art trackers in both precision plot score (84.3%) and success plot score (61.9%). Note that SRDCF reaches 61.7% on OP score, which is comparable to ours. In fact, SRDCF introduces a strategy to handle scale change during tracking, while the bounding boxes for each target in our tracker are fixed size. Thus, SRDCF is easy to reach high OP score. Even so, the OP score of our tracker is a little higher than that of SRDCF. And the DP score of ours improves more than 2% than SRDCF, which proves the discriminative power of the introduced classifier.

D. Experimental Results on TB-100

To further verify the tracking performance, the tracking algorithm is tested on visual object tracking benchmark TB-100. This data set is the full object tracking benchmark 2015 [14], which contains 100 video clips. We compare our method with the best performance methods, such as PCOM [22] and CCT [14]. In TB-100, comparisons are measured by DP and mean OP. Fig. 5 shows the precision plot (left) and success plot (right) over all these 100 sequences. For clarity, we list these scores in Table II. CSK is the most relevant work to ours. The performance scores of CSK are 52.0 % and 38.5%, respectively. CCT performs well on this data set (DP: 74.8% and OP: 55.1%). Our tracker (MRCT-AS) gives a mean DP of 77.6% and an OP score of 55.3%, which performs the best.

V. CONCLUSION

We have proposed an MRCT-AS to promote the original correlation filter-based tracking method. By exploiting the manifold spatial structure of both labeled and unlabeled augmented samples, we introduced a semisupervised tracking approach to improve the tracking performance, leading to a more discriminative appearance model. And a blockwise fast learning and detection algorithm has been introduced for online visual tracking. To handle appearance change during tracking, an online model update strategy was applied. Experimental results on OTB-2013 and TB-100 showed that our tracking method performed better than state-of-the-art tracking algorithms.

APPENDIX
DETAILED DERIVATION OF (4)

The manifold regularized least squares algorithm solves the optimization problem in (1)

$$\arg\min_{f\in\mathcal{H}_\kappa} \frac{1}{l}\sum_{i=1}^{l}(f(\mathbf{x}_i)-y_i)^2+\lambda\|f\|_\kappa^2+\frac{\gamma}{n^2}\mathbf{f}^T\mathbf{L}\mathbf{f}$$

where the meaning of each variable refers to Section III-A.

According to the Representer Theorem, the solution of the above-mentioned objective function could be represented

by (3)

$$f^*(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i \kappa(\mathbf{x}, \mathbf{x}_i).$$

Substituting this equation to the above-mentioned objective function, we obtain a convex function over the n-dimensional variable $\boldsymbol{\alpha}$

$$\boldsymbol{\alpha} = \arg\min_{\boldsymbol{\alpha}\in\mathcal{R}^n} \frac{1}{l}(Y - \mathbf{JK}\boldsymbol{\alpha})^T(Y - \mathbf{JK}\boldsymbol{\alpha}) + \lambda\boldsymbol{\alpha}^T\mathbf{K}\boldsymbol{\alpha} + \frac{\gamma}{n^2}\boldsymbol{\alpha}^T\mathbf{KLK}\boldsymbol{\alpha}.$$

Let the derivative of the above-mentioned function (with respective to $\boldsymbol{\alpha}$) be equal to zero

$$\frac{1}{l}(Y - \mathbf{JK}\boldsymbol{\alpha})^T(-\mathbf{JK}) + \left(\lambda\mathbf{K} + \frac{\gamma}{n^2}\mathbf{KLK}\right)\boldsymbol{\alpha} = 0.$$

Finally, we obtain the solution in (4)

$$\boldsymbol{\alpha} = \left(\mathbf{JK} + \lambda l\mathbf{I} + \frac{\gamma l}{n^2}\mathbf{LK}\right)^{-1} Y. \tag{28}$$

REFERENCES

[1] X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, and A. Van Den Hengel, "A survey of appearance models in visual object tracking," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 4, p. 58, Sep. 2013.

[2] X. Liu, D. Tao, M. Song, L. Zhang, J. Bu, and C. Chen, "Learning to track multiple targets," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 5, pp. 1060–1073, May 2015.

[3] L. Zhao, X. Gao, D. Tao, and X. Li, "Learning a tracking and estimation integrated graphical model for human pose tracking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3176–3186, Dec. 2015.

[4] W. Wang, J. Shen, X. Li, and F. Porikli, "Robust video object cosegmentation," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3137–3148, Oct. 2015.

[5] J. Shen, Y. Du, W. Wang, and X. Li, "Lazy random walks for superpixel segmentation," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1451–1462, Apr. 2014.

[6] 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.*, Berlin, Germany, Oct. 2012, pp. 702–715.

[7] W. Wang, J. Shen, and F. Porikli, "Saliency-aware geodesic video object segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3395–3402.

[8] 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.

[9] B. Ma, J. Shen, Y. Liu, H. Hu, L. Shao, and X. Li, "Visual tracking using strong classifier and structural local sparse descriptors," *IEEE Trans. Multimedia*, vol. 17, no. 10, pp. 1818–1828, Oct. 2015.

[10] X. Dong, J. Shen, D. Yu, W. Wang, J. Liu, and H. Huang, "Occlusion-aware real-time object tracking," *IEEE Trans. Multimedia*, vol. 19, no. 4, pp. 763–771, Apr. 2016.

[11] B. Ma, L. Huang, J. Shen, L. Shao, M.-H. Yang, and F. Porikli, "Visual tracking under motion blur," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5867–5876, Oct. 2016.

[12] 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.*, Dec. 2013, pp. 2760–2767.

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

[14] Y. Wu, J. Lim, and M. H. Yang, "Object tracking benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1834–1848, Sep. 2015.

[15] M. Kristan, *et al.*, "The visual object tracking VOT2014 challenge results," in *Proc. Eur. Conf. Comput. Vis. Workshops*, Berlin, Germany, Mar. 2014, pp. 191–217.

[16] 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.

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

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

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

[20] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan, "Locally orderless tracking," *Int. J. Comput. Vis.*, vol. 111, no. 2, pp. 213–228, Jan. 2015.

[21] D. Chen, Z. Yuan, Y. Wu, G. Zhang, and N. Zheng, "Constructing adaptive complex cells for robust visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 1113–1120.

[22] D. Wang and H. Lu, "Visual tracking via probability continuous outlier model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 3478–3485.

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

[24] J. A. Fernandez, V. N. Boddeti, A. Rodriguez, and B. V. K. V. Kumar, "Zero-aliasing correlation filters for object recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1702–1715, Aug. 2015.

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

[26] A. Mahalanobis, "Correlation filters for object tracking, target reacquisition, and smart aim-point selection," *Proc. SPIE*, vol. 3073, pp. 25–32, Mar. 1997.

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

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

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

[30] L. Zhang, Y. Wang, H. Sun, Z. Yao, and S. He, "Robust visual correlation tracking," *Math. Problems Eng.*, vol. 2015, Sep. 2015, Art. no. 238971.

[31] Q. Du, Z. Cai, H. Liu, and Z. L. Yu, "A rotation adaptive correlation filter for robust tracking," in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2015, pp. 1035–1038.

[32] 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.*, Jun. 2014, pp. 1090–1097.

[33] C. Ma, J. B. Huang, X. Yang, and M. H. Yang, "Hierarchical convolutional features for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3074–3082.

[34] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4310–4318.

[35] M. Tang and J. Feng, "Multi-kernel correlation filter for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3038–3046.

[36] T. Liu, G. Wang, and Q. Yang, "Real-time part-based visual tracking via adaptive correlation filters," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4902–4912.

[37] C. Ma, X. Yang, C. Zhang, and M. H. Yang, "Long-term correlation tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5388–5396.

[38] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.

[39] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems*, vol. 16. Cambridge, MA, USA: MIT Press, 2003, pp. 153–160.

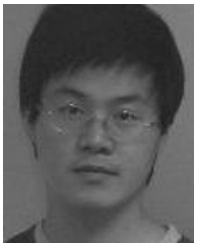
[40] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[41] U. Von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[42] Y. Bai and M. Tang, "Robust tracking via weakly supervised ranking SVM," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1854–1861.
[43] B. Ma, H. Hu, J. Shen, Y. Liu, and L. Shao, "Generalized pooling for robust object tracking," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4199–4208, Sep. 2016.
[44] B. Ma, L. Huang, J. Shen, and L. Shao, "Discriminative tracking using tensor pooling," *IEEE Trans. Cybern.*, vol. 46, no. 11, pp. 2411–2422, Nov. 2016.
[45] B. Zhuang, H. Lu, Z. Xiao, and D. Wang, "Visual tracking via discriminative sparse similarity map," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1872–1881, Apr. 2014.
[46] B. Ma, H. Hu, J. Shen, Y. Zhang, and F. Porikli, "Linearization to nonlinear learning for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4400–4407.
[47] R. A. Hom and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.
[48] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.
[49] S. He, Q. Yang, R. W. H. Lau, J. Wang, and M.-H. Yang, "Visual tracking via locality sensitive histograms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2427–2434.

Hongwei Hu is currently pursuing the Ph.D. degree with the School of Computer Science, Beijing Institute of Technology, Beijing, China.

His current research interests include visual object tracking algorithms.



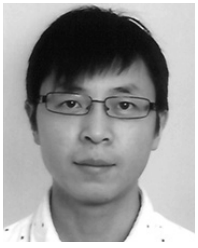
Bo Ma (M'13) received the Ph.D. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 2003.

From 2004 to 2006, he was with the Department of Computer Science, City University of Hong Kong, Hong Kong, where he was involved in research projects in computer vision and pattern recognition. In 2006, he joined the Department of Computer Science, Beijing Institute of Technology, Beijing, China, where he is currently an Associate Professor. He has authored about 40 journal and conference papers, such as the IEEE TRANSACTIONS ON IMAGE PROCESSING and the IEEE International Conference on Computer Vision (ICCV). His current research interests include statistical pattern recognition and computer vision.

Jianbing Shen (M'11–SM'12) received the Ph.D. degree in computer science from Zhejiang University, Hangzhou, China, in 2007.

He was a Visiting Professor with the Department of Information Technology and Electrical Engineering, ETH Zürich, Zürich, Switzerland, and the Department of Computer Science, University of California at Los Angeles, Los Angeles, CA, USA. He has authored about 60 journal and conference papers, such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE Computer Vision and Pattern Recognition, and the IEEE ICCV. He is currently a Professor with the School of Computer Science, Beijing Institute of Technology, Beijing, China. His current research interests include computer vision and machine learning.

Dr. Shen received many flagship honors, including the Fok Ying Tung Education Foundation from the Ministry of Education, the Program for Beijing Excellent Youth Talents from the Beijing Municipal Education Commission, and the Program for New Century Excellent Talents from the Ministry of Education. He serves as an Associate Editor on the editorial boards of *Neurocomputing*.



Ling Shao (M'09–SM'10) is currently a Full Professor with the School of Computing Sciences, University of East Anglia, Norwich, U.K. His current research interests include computer vision, image processing, pattern recognition, and machine learning.

Dr. Shao is a fellow of the British Computer Society and the Institution of Engineering and Technology, and a Life Member of the ACM. He is an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, and other journals.