Learning Collaborative Sparse Correlation Filter for Real-time Multispectral Object Tracking

Yulong Wang, Chenglong Li*, Jin Tang, and Dengdi Sun

School of Computer Science and Technology, Anhui University, Hefei, China
wylemail@qq.com, lcl1314@foxmail.com
tj@ahu.edu.cn, sundengdi@163.com

Abstract. To track objects efficiently and effectively in adverse illumination conditions even in dark environment, this paper presents a novel multispectral approach to deploy the intra- and inter-spectral information in the correlation filter tracking framework. Motivated by brain inspired visual cognitive systems, our approach learns the collaborative sparse correlation filters using color and thermal sources from two aspects. First, it pursues a sparse correlation filter for each spectrum. By inheriting from the advantages of the sparse representation, our filers are robust to noises. Second, it exploits the complementary benefits from two modalities to enhance each other. In particular, we take their interdependence into account for deriving the correlation filters jointly, and formulate it as a $l_{2,1}$-based sparse learning problem. Extensive experiments on large-scale benchmark datasets suggest that our approach performs favorably against the state-of-the-arts in terms of accuracy while achieves in real-time frame rate.

Keywords: Visual tracking · Information fusion · Sparse representation · Correlation filter.

1 Introduction

Visual tracking is an active research area in the computer vision community, since it is an essential and significant task in visual surveillance [20], human-computer interaction [18], and self-driving systems [1]. Despite of many breakthroughs recently, the visual tracking mainly relies on traditional RGB sensors and thus tracking target objects in case of cluttered background and low visibility at night and in bad weather is still regarded as a challenging problem.

Vision is one of the most important ways that the human brain perceives the outside world to acquire information, and the eye is the "window" to receive visual information. It is worth noting that visual cognition of human eyes is a multi-channel system. Therefore, integrating the multi-source information is a nature way to boost vision tasks in challenging scenarios [21]. Moreover, at progressively higher levels of sensory processing, information is carried by fewer

* Corresponding Author

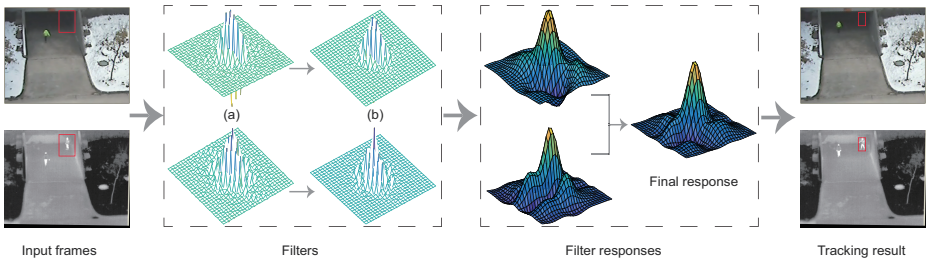

Fig. 1. Pipeline of the proposed approach. (a) and (b) denote the correlation filters optimized by our proposed model without and with collaborative sparse constraint, respectively.

neurons because the system is organized to a near complete a representation with the fewest active neurons. In other terms, the encoding of sensory information gets "sparser" as one moves up into higher levels of sensory processing [7].

Motivated by the above observations, we propose a novel approach that uses collaborative sparse correlation filters to integrate multiple source data, i.e., visible and thermal infrared spectrums, for visual tracking. Our method deploys both the intra- and inter-spectral information in the correlation filter tracking framework. First, instead of using l_2-regularization on the filters [10], we pursue a sparse correlation filter to select most significant parts to enhance the discriminative ability [6, 26]. It will make our filters efficient and robust by inheriting from both advantages of the sparse representation and the correlation filter. The effectiveness of using the sparse is demonstrated in Fig. 1. Second, we exploit the complementary benefits from two modalities to enhance each other. In particular, we observe that different spectrums should have similar filters such that they have consistent localization of the target object, as shown in Fig. 1(b). Therefore, we jointly learn the correlation filters of color and thermal spectrums to collaboratively distinguish the target from the background.

We summarize our contributions to multispectral object tracking and related applications as follows. First, we propose a novel approach that carries out efficient and effective fusion of multiple spectral data, and performs favorably against the state-of-the-art multispectral trackers on two benchmark datasets, i.e., GTOT [12] and RGBT210 [16]. Second, we employ a sparse- and collaborative sparse-based regularizations on the joint filter to deploy both intra- and inter-modal complementary benefits from of color and thermal spectrums. Third, extensive analysis and evaluation on large-scale benchmark datasets verify the effectiveness and efficiency of the proposed approach.

2 Related Work

The most relevant methods and techniques are discussed. We review the related work to us from two research streams, i.e., multispectral tracking and correlation filter tracking.

2.1 Multispectral Tracking

Given the bounding box of an unknown target in the first frame, the goal of "single target tracking" is to locate this object in subsequent video frames, despite object motion, changes in viewpoint, lighting changes, or other variations. Multispectral object tracking has drawn a lot of attentions in the computer vision community with the popularity of thermal infrared sensors [22, 17, 12, 13, 16, 23, 19]. Yan et al. [23] cognitive fusion of RGB and thermal information provides an effective solution for effective detection and tracking of pedestrians in videos. Wu et al. [22] and Liu and Sun [17] directly employ the sparse representation to calculate the likelihood score using reconstruction residues or coefficients in Bayesian filtering framework. They ignore modality reliabilities in fusion, which may limit the tracking performance when facing malfunction or occasional perturbation of individual sources. Li et al. [12] and Li et al. [16] introduce modality weights to handle this problem, and propose sparse representation based algorithms to fuse color and thermal information. Different from these methods, we make the best use of intra- and inter-spectrum information in the correlation filter tracking framework to perform efficient and effective multispectral tracking.

2.2 Correlation Filter Tracking

Correlation filters have achieved great breakthroughs in visual tracking due to its accuracy and computational efficiency [2, 9, 10, 15, 14, 4, 6, 26, 3]. Bolme et al. [2] first introduce correlation filters into visual tracking, named MOSSE, and achieve hundreds of frames per second, and high tracking accuracy. Recently, many researchers further improve MOSSE from different aspects. For example, Henriques et al. [9, 10] extend MOSSE to non-linear one with kernel trick, and incorporate multiple channel features efficiently by summing all channels in kernel space. To handle scale variations, Danelljan et al. [4] learn correlation filters for translation and scale estimation separately by using a scale pyramid representation. Dong et al. [6] propose a sparse correlation filter for combining the robustness of sparse representation and the efficiency of correlation filter. Zhang et al. [26] integrate multiple parts and multiple features into a unified correlation particle filter framework to perform effective object tracking.

3 Proposed Algorithm

In this section, we first present the technical details of the proposed algorithm and then describe the optimization process of the model.

3.1 Formulation

For a typical correlation filter, many negative samples are used to improve the discriminability of the track-by-detector scheme. In this work, denote $\mathbf{x}_k \in \mathbb{R}^{M \times N \times D}$as the feature vector of k-th spectrum, where M, N, and D indicates

the width, height,and the number of channels, respectively. We consider all the circular shifts of $\mathbf{x}_k$ along the M and N dimensions as training samples. Each shifted sample $\mathbf{x}_{m,n}^k$, $(m,n) \in \{0,1,...,M-1\}\times\{0,1,...,N-1\}$, has a Gaussian function label $y(m,n)$. Let $\mathbf{X}_k = [\mathbf{x}_{0,0},...,\mathbf{x}_{m,n},...\mathbf{x}_{M-1,N-1}]^{\mathrm{T}}$ denote all training samples of the k-th spectrum. The goal is to find the optimal correlation filters $\mathbf{w}_k$ for K different spectrums,

$$\min_{\mathbf{w}_k} \sum_{k=1}^{K} \frac{1}{2}||\mathbf{X}_k\mathbf{w}_k - \mathbf{y}||_2^2 + \lambda_1||\mathbf{w}_k||_2^2, \tag{1}$$

the last term of Eq.(1) is the regularization of $\mathbf{w}_k$. For each spectrum k, only a few possible locations $\mathbf{x}_k^{m,n}$ should be selected to localize where the target object is at next frame. Ideally, only one possible location corresponds to the target object. Based on this observation, we suggest that the regularization should use the l_1 norm instead of the l_2 norm,

$$\min_{\mathbf{w}_k} \sum_{k=1}^{K} \frac{1}{2}||\mathbf{X}_k\mathbf{w}_k - \mathbf{y}||_2^2 + \lambda_1||\mathbf{w}_k||_1. \tag{2}$$

Among different spectrums, the learned $\mathbf{w}_k$ should select similar circular shifts so that they have similar motion. As a result, the learned $\mathbf{w}_k$ should be similar. In this work, we use the convex $l_{2,1}$ mixed norm to learned their correlation filters jointly to distinguish the target from the background, and the final structure of our model is as follows:

$$\min_{\mathbf{w}_k} \sum_{k=1}^{K} \frac{1}{2}||\mathbf{X}_k\mathbf{w}_k - \mathbf{y}||_2^2 + \lambda_1||\mathbf{w}_k||_1 + \lambda_2||\mathbf{W}||_{2,1}, \tag{3}$$

where $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_K] \in \mathbb{R}^{MN\times K}$, and λ_1, λ_2 are regularization parameters. The definition of the $l_{p,q}$ mixed norm is $||\mathbf{W}||_{p,q} = (\sum_i(\sum_j |[\mathbf{W}]_{ij}|^p)^{\frac{q}{p}})^{\frac{1}{q}}$ and $[\mathbf{W}]_{ij}$ denotes the element at the i-th row and j-th column of $\mathbf{W}$.

3.2 Optimization

In this section, We present algorithmic details on how to efficiently solve the optimization problem (3). We first use the linearized ADM with adaptive penalty to avoid some matrix inversions in optimization. Two auxiliary variables $\mathbf{p}_k \in \mathbb{R}^{MN\times 1}$ and $\mathbf{Q} \in \mathbb{R}^{MN\times K}$ are introduced to make Eq.(3) separable:

$$\min_{\mathbf{w}_k} \sum_{k=1}^{K} \frac{1}{2}||\mathbf{X}_k\mathbf{w}_k - \mathbf{y}||_2^2 + \lambda_1||\mathbf{p}_k||_1 + \lambda_2||\mathbf{Q}||_{2,1}, s.t. \mathbf{w}_k = \mathbf{p}_k, \mathbf{W} = \mathbf{Q} \tag{4}$$

We use the fast first-order Alternating Direction Method of Multipliers (ADMM) to efficiently solve the optimization problem (4). By introducing augmented Lagrange multipliers to incorporate the equality constraints into the objective function, we obtain a Lagrangian function that can be optimized through a sequence

of simple closed form update operations in (5).

$$\begin{aligned}
&\min_{\mathbf{w}_k,\mathbf{p}_k,\mathbf{Q}} \sum_{k=1}^{K} \frac{1}{2}||\mathbf{X}_k\mathbf{w}_k - \mathbf{y}||_2^2 + \lambda_1||\mathbf{p}_k||_1 + \frac{\mu}{2}||\mathbf{w}_k - \mathbf{p}_k||_F^2 + \langle \mathbf{Y}_{1,k}, \mathbf{w}_k - \mathbf{p}_k\rangle \\
&+ \lambda_2||\mathbf{Q}||_{2,1} + \langle \mathbf{Y}_2, \mathbf{W} - \mathbf{Q}\rangle + \frac{\mu}{2}||\mathbf{W} - \mathbf{Q}||_F^2 \\
&= \sum_{k=1}^{K} \frac{1}{2}||\mathbf{X}_k\mathbf{w}_k - \mathbf{y}||_2^2 + \lambda_1||\mathbf{p}_k||_1 + \frac{\mu}{2}||\mathbf{w}_k - \mathbf{p}_k + \frac{\mathbf{Y}_{1,k}}{\mu}||_2^2 - \frac{1}{2\mu}||\mathbf{Y}_{1,k}||_2^2 \\
&+ \lambda_2||\mathbf{Q}||_{2,1} + \frac{\mu}{2}||\mathbf{W} - \mathbf{Q} + \frac{\mathbf{Y}_2}{\mu}||_F^2 - \frac{1}{2\mu}||\mathbf{Y}_2||_F^2
\end{aligned} \tag{5}$$

Here, $\langle \mathbf{A}, \mathbf{B}\rangle = \mathrm{Tr}(\mathbf{A}^{\mathrm{T}}\mathbf{B})$ denotes the matrix inner product. $\mathbf{Y}_{1,k}$ and $\mathbf{Y}_2$ are Lagrangian multipliers. We then alternatively update one variable by minimizing (5) with fixing other variables. Besides the Lagrangian multipliers, there are three variables, including $\mathbf{W}$, $\mathbf{p}_k$ and $\mathbf{Q}$, to solve. The solutions of the subproblems are as follows:

w-subproblem: Given fixed $\mathbf{p}_k$ and $\mathbf{Q}$, $\mathbf{w}_k$ is updated by solving the optimization problem (6) with the solution (7)

$$\min_{\mathbf{w}_k} \sum_{k=1}^{K} \frac{1}{2}||\mathbf{X}_k\mathbf{w}_k - \mathbf{y}||^2 + \frac{\mu}{2}||\mathbf{w}_k - \mathbf{p}_k + \frac{\mathbf{Y}_{1,k}}{\mu}||_2^2 + \frac{\mu}{2}||\mathbf{W} - \mathbf{Q} + \frac{\mathbf{Y}_2}{\mu}||_F^2, \tag{6}$$

$$\mathbf{w}_k = (\mathbf{X}_k^{\mathrm{T}}\mathbf{X}_k + 2\mu\mathbf{I})^{-1}(\mathbf{X}_k^{\mathrm{T}}\mathbf{y} + \mu(\mathbf{p}_k + \mathbf{Q}_k) - (\mathbf{Y}_{1,k} + \mathbf{Y}_{2,k})). \tag{7}$$

Here, $\mathbf{w}_k$, $\mathbf{P}_k$ and $\mathbf{Q}_k$ denote the k-th column of $\mathbf{W}$, $\mathbf{P}$, $\mathbf{Q}$, respectively. Note that, all circulant matrices are made diagonal by the Discrete Fourier Transform (DFT), regardless of the generating vector. If $\mathbf{X}_k$ is a circulant matrix, it can be expressed with its base sample $\mathbf{x}_k$ as

$$\mathbf{X}_k = F\mathit{diag}(\hat{\mathbf{x}}_k)F^{\mathrm{H}}, \tag{8}$$

where $\hat{\mathbf{x}}_k$ denotes the DFT of the generating vector, $\hat{\mathbf{x}}_k = \mathcal{F}(\mathbf{x}_k)$, and F is a constant matrix that does not depend on $\mathbf{x}_k$. The constant matrix F is known as the DFT matrix. $\mathbf{X}_k^{\mathrm{H}}$ is the Hermitian transpose, i.e., $\mathbf{X}_k^{\mathrm{H}} = (\mathbf{X}_k^*)^{\mathrm{T}}$, and $\mathbf{X}_k^*$ is the complex-conjugate of $\mathbf{X}_k$. For real numbers, $\mathbf{X}_k^{\mathrm{H}} = \mathbf{X}_k^{\mathrm{T}}$. It(Eq.(7)) can be calculated very efficiently in the Fourier domain by considering the circulant structure property of $\mathbf{X}_k$,

$$\mathbf{w}_k = \mathcal{F}^{-1}[\frac{\hat{\mathbf{x}}_k^* \odot \hat{\mathbf{y}} + \mu(\hat{\mathbf{p}}_k + \hat{\mathbf{Q}}_k) - (\hat{\mathbf{Y}}_{1,k} + \hat{\mathbf{Y}}_{2,k})}{\hat{\mathbf{x}}_k^* \odot \hat{\mathbf{x}}_k + 2\mu}]. \tag{9}$$

Here, $\mathcal{F}^{-1}$ denotes the inverse DFT, while $\odot$ as well as the fraction denote the element-wise product and division, respectively. The $\mathbf{x}_k$ is the base sample of circulant matrix $\mathbf{X}_k$.

Algorithm 1 Optimization Procedure to Eq.(4).

Input: The spectra feature matrix $\mathbf{X}_k(k = 1, 2..., K)$ and Gaussian function label $\mathbf{y}$, the parameters λ_1 and λ_2;
Set $\mathbf{w}_k = \mathbf{p}_k = \mathbf{Y}_{1,k} = 0, \mathbf{W} = \mathbf{Q} = \mathbf{Y}_2 = 0, \mu_0 = 0.1, \mu_{max} = 10^{10}, \rho = 1.2, \epsilon = 10^{-15}, maxIter = 10$ and $t = 0$.
Output: The filter $\mathbf{w}_k$.
while not converged **do**
Update $\mathbf{w}_{k,t+1}$ by Eq.(7);
Update $\mathbf{p}_{k,t+1}$ by Eq.(11);
Update $\mathbf{Q}_{t+1}$ by Eq.(13);
Update Lagrange multipliers as follows:
$\mathbf{Y}_{1,k,t+1} = \mathbf{Y}_{1,k,t} + \mu_t(\mathbf{w}_k - \mathbf{p}_k)$;
$\mathbf{Y}_{2,t+1} = \mathbf{Y}_{2,t} + \mu_t(\mathbf{W} - \mathbf{Q})$;
Update μ_{t+1} by $\mu_{t+1} = \min(\mu_{max}, \rho\mu_t)$;
Update t by $t = t + 1$;
Check the convergence condition, i.e. the maximum element changes of $\mathbf{w}_k, \mathbf{p}_k$ and $\mathbf{Q}$ between two consecutive iterations are less than ϵ or the maximum number of iterations reaches maxIter.
end while

p-subproblem: Given fixed $\mathbf{w}_k$ and $\mathbf{Q}$, Eq.(5) can be rewritten as

$$\min_{\mathbf{p}_k} \sum_{k=1}^{K} \frac{\mu}{2} ||\mathbf{w}_k - \mathbf{p}_k + \frac{\mathbf{Y}_{1,k}}{\mu}||_2^2 + \lambda_1 ||\mathbf{p}_k||_1. \tag{10}$$

According to (Lin et al. 2009), an efficient closed-form solution of Eq.(10) can be computed by the soft-thresholding(or shrinkage) method:

$$\mathbf{p}_k = S_{\frac{\lambda}{\mu}}(\mathbf{w}_k + \frac{\mathbf{Y}_{1,k}}{\mu}), \tag{11}$$

where the definition of $S_\lambda(a)$ is $S_\lambda(a) = \text{sign}(a)\max(0, |a| - \lambda)$.

Q-subproblem: Given fixed $\mathbf{W}$, $\mathbf{P}$, Eq.(5) can be rewritten as

$$\begin{aligned} &\min_{\mathbf{Q}} \frac{\mu}{2} ||\mathbf{W} - \mathbf{Q} + \frac{\mathbf{Y}_2}{\mu}||_F^2 + \lambda_2 ||\mathbf{Q}||_{2,1} \\ &= \min_{\mathbf{Q}_1, \ldots, \mathbf{Q}_m} \sum_{i=1}^{m} (\frac{\mu}{2} ||\mathbf{W}_i - \mathbf{Q}_i + \frac{\mathbf{Y}_{2,i}}{\mu}||_2^2 + \lambda_2 ||\mathbf{Q}_i||_2). \end{aligned} \tag{12}$$

$\mathbf{W}_i$, $\mathbf{Q}_i$ and $\mathbf{Y}_{2,i}$ denote the i-th row of the matrix $\mathbf{W}$, $\mathbf{Q}$ and $\mathbf{Y}_2$, respectively. For each row of $\mathbf{Q}_i$ in the subproblem (12), an efficient closed-form solution can be computed:

$$\mathbf{Q}_i = max(0, 1 - \frac{\lambda_2}{\mu ||\mathbf{H}_i||_2}) \mathbf{H}_i, \tag{13}$$

where $\mathbf{H}_i = \mathbf{W}_i + \frac{1}{\mu}\mathbf{Y}_{2,i}$.

Since each subproblem of Eq.(4) is convex, we can guarantee that the limit point by our algorithm satisfies the Nash equilibrium conditions. And the main steps of the optimization procedure are summarized in Algorithm 1.

3.3 Target Position Estimation

After solving this optimization problem, we obtain the correlation filter $\mathbf{w}_k$ for each type of modality. Given an image patch in the next frame, the feature vector on the k-th modality is denoted by $\mathbf{z}_k$ and of size $M \times N \times D$. We first transform it to the Fourier domain $\hat{\mathbf{z}}_k = \mathcal{F}(\mathbf{z}_k)$, and then the final correlation response map is computed by

$$S = \sum_{k=1}^{K} \mathcal{F}^{-1}(\hat{\mathbf{z}}_\mathbf{k} \odot \hat{\mathbf{w}}_\mathbf{k}). \tag{14}$$

The target location then can be estimated by searching for the position of maximum value of the correlation response map S of size $M \times N$.

3.4 Model Update

In practice, we adopt an incremental strategy, which only uses new samples $\mathbf{x}_k$ in the current frame to update models as shown in (15), where t is the frame index and η is a learning rate parameter.

$$\mathcal{F}(\mathbf{w}_\mathbf{k})^t = (1-\eta)\mathcal{F}(\mathbf{w}_\mathbf{k})^{t-1} + \eta\mathcal{F}(\mathbf{w}_\mathbf{k})^t. \tag{15}$$

4 Experimental Results

4.1 Experimental Setups

Implementation details: There are two hyperparameters, i.e., λ_1 and λ_2, in Eq.(3). The value of them are estimated by performing a grid search from 0 to 1 with step 0.01. Evaluations show that the best performance is achieved when $\lambda_1 = 0.11$ and $\lambda_2 = 0.14$, and use a kernel width of 0.1 for generating the Gaussian function labels. Their learning rate η in (15) is set to 0.025. To remove the boundary discontinuities, the extracted feature channels are weighted by a cosine window. We implement our tracker in MATLAB on an Intel I7-6700K 4.00 GHz CPU with 32 GB RAM. Furthermore, all the parameter settings are available in the source code to be released for accessible reproducible research.

Datasets: Our algorithm is evaluated on two large datasets: GTOT and RGBT210. GTOT includes 50 aligned RGB-T video pairs with about 12K frames in total, and RGBT210 includes 210 highly-aligned RGB-T video pairs with about 210K frames in total. They are annotated with ground truth bounding boxes and various visual attributes.

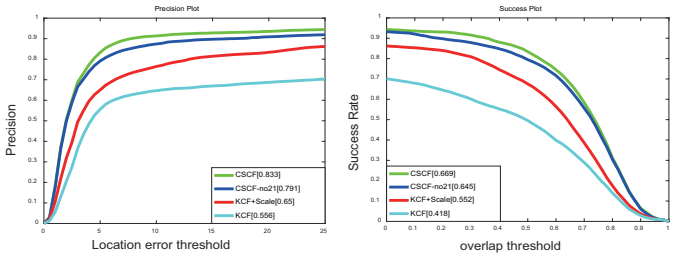

Fig. 2. PR and SR plots on GTOT. The performance of CSCF tracker is improved gradually with the addition of collaborative sparse constraint.

Evaluation protocol: As a measure of tracking results, we use the success rate (SR) to evaluate the RGB-T tracking performance. SR is the ratio of the number of successful frames whose overlap is larger than a threshold. By changing the threshold, the SR plot can be obtained, and we employ the area under curve of SR plot to define the representative SR.

4.2 Model Analysis

KCF [10] is intended as the baseline in this work, and is achieved by using the released code with default parameter settings. KCF+Scale utilize an adaptive multi-scale strategy to handle scale variations. CSCF-no21 is our model without using the mixed $l_{2,1}$ norm. Scale processing also apply to CSCF-no21 and CSCF, and all trackers use grayscale feature. Fig. 2 shows the performance of our model is improved gradually with the addition of l_1 and $l_{2,1}$ norm on the benchmark of GTOT. CSCF-no21 dramatically improves the performance by a SR score of 9.3% compared with KCF+Scale. Our overall model achieves about 4.2%/2.4% improvement with PR and SR metrics in comparison with CSCF-no21, which demonstrates the effectiveness of the proposed mechanism in practical tracking.

4.3 Comparison with State-of-the-Arts

Evaluation on GTOT: We first evaluate our CSCF algorithm with 9 trackers on GTOT, including CSR [12], Struck [8], SCM [27], CN [5], STC [25], KCF [10], L1-PF [22], JSR [17] and TLD [11]. Among all the trackers, our CSCF tracker achieves the best results as shown in Fig. 3(a). Compared with CSR, CSCF achieves about 4.9% improvement with SR on the GTOT dataset. Furthermore, compared with KCF, CSCF achieves much better performance with about 24.7% improvement.

Evaluation on RGBT210: For further demonstrate the effectiveness of the proposed approach, we construct experiments on the public RGBT210 dataset with comparisons to 6 trackers, including DSST [4], MEEM [24]+RGBT, CSR, KCF+RGBT, L1-PF, JSR. Fig. 3(b) shows that our tracker significantly outper-

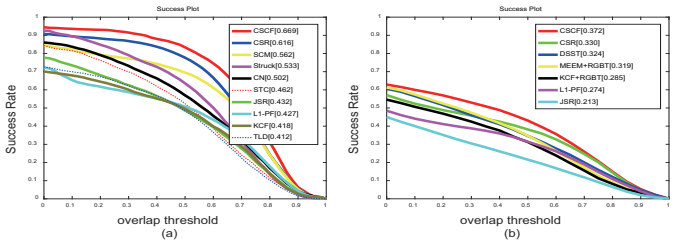

Fig. 3. (a) and (b) denote the evaluation results on GTOT and RGB210 dataset, respectively. The representative score of SR is presented in the legend.

form them. In particular, The proposed CSCF obtains 4.2% performance gains in SR, which is much better than CSR. Most importantly, the proposed tracker performs at about 50 FPS(frames per second), which has enabled real-time object tracking.

5 Conclusion

In this paper, we propose a novel learning collaborative sparse correlation filters for multispectral object tracking. The proposed tracking algorithm can effectively exploit interdependencies among different spectrums to learn their correlation filters jointly. Experimental results compared with several state-of-the-art methods on two visual tracking benchmark datasets demonstrate the effectiveness and robustness of the proposed algorithm.

Acknowledgment

This work was jointly supported by National Natural Science Foundation of China (61702002, 61472002, 61402002), Natural Science Foundation of Anhui Province (1808085QF187), Natural Science Foundation of Anhui Higher Education Institution of China (KJ2017A017, KJ2018A0023) and Co-Innovation Center for Information Supply & Assurance Technology of Anhui University.

References

1. Bengler, K., Dietmayer, K., Farber, B., Maurer, M.: Three decades of driver assistance systems: Review and future perspectives. ITSM **6**(4), 6–22 (2014)
2. Bolme, D.S., Beveridge, J.R., Draper, B.A., Lui, Y.M.: Visual object tracking using adaptive correlation filters. In: Proc. IEEE Conf. CVPR (2010)
3. Danelljan, M., Bhat, G., Khan, F.S., Felsberg, M.: Eco: Efficient convolution operators for tracking. In: Proc. IEEE Conf. CVPR (2017)

4. Danelljan, M., Hager, G., Khan, F., Felsberg, M.: Accurate scale estimation for robust visual tracking. In: Proc. BMVC (2014)
5. Danelljan, M., Khan, F.S., Felsberg, M., Weijer, J.V.D.: Adaptive color attributes for real-time visual tracking. In: Proc. IEEE Conf. CVPR. pp. 1090–1097 (2014)
6. Dong, Y., Yang, M., Pei, M.: Visual tracking with sparse correlation filters. In: Proc. IEEE ICIP. pp. 439–443 (2016)
7. Donoho, D.L.: Compressed sensing. IEEE IT **52**(4), 1289–1306 (2006)
8. Hare, S., Saffari, A., Torr, P.H.S.: Struck: Structured output tracking with kernels. In: Proc. IEEE ICCV. pp. 263–270 (2011)
9. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: Exploiting the circulant structure of tracking-by-detection with kernels. In: Proc. ECCV (2012)
10. Henriques, J.F., Rui, C., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. IEEE TPAMI **37**(3), 583–596 (2015)
11. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. IEEE TPAMI **34**(7), 1409–1422 (2012)
12. Li, C., Cheng, H., Hu, S., Liu, X., Tang, J., Lin, L.: Learning collaborative sparse representation for grayscale-thermal tracking. IEEE TIP **25**(12), 5743–5756 (2016)
13. Li, C., Hu, S., Gao, S., Tang, J.: Real-time grayscale-thermal tracking via laplacian sparse representation. In: Proc. MMM. pp. 54–65. Springer (2016)
14. Li, C., Liang, X., Lu, Y., Zhao, N., Tang, J.: Rgb-t object tracking: Benchmark and baseline. arXiv:1805.08982 (2018)
15. Li, C., Lin, L., Zuo, W., Tang, J., Yang, M.: Visual tracking via dynamic graph learning. arXiv:1710.01444 (2018)
16. Li, C., Zhao, N., Lu, Y., Zhu, C., Tang, J.: Weighted sparse representation regularized graph learning for rgb-t object tracking. In: Proc. ACM MM. pp. 1856–1864 (2017)
17. Liu, H., Sun, F.: Fusion tracking in color and infrared images using joint sparse representation. Information Sciences **55**(3), 590–599 (2012)
18. Liu, L., Xing, J., Ai, H., Xiang, R.: Hand posture recognition using finger geometric feature. In: Proc. ICPR. pp. 565–568 (2013)
19. Ren, J., Orwell, J., Jones, G.A., Xu, M.: Tracking the soccer ball using multiple fixed cameras. CVIU **113**(5), 633–642 (2009)
20. Ren, J., Xu, M., Orwell, J., Jones, G.A.: Multi-camera video surveillance for real-time analysis and reconstruction of soccer games. MVA **21**(6), 855–863 (2010)
21. Wang, Z., Ren, J., Zhang, D., Sun, M., Jiang, J.: A deep-learning based feature hybrid framework for spatiotemporal saliency detection inside videos. Neurocomputing (2018)
22. Wu, Y., Blasch, E., Chen, G., Bai, L., Ling, H.: Multiple source data fusion via sparse representation for robust visual tracking. In: Proc. ICIF. pp. 1–8 (2011)
23. Yan, Y., Ren, J., Zhao, H., Sun, G., Wang, Z., Zheng, J., Marshall, S., Soraghan, J.: Cognitive fusion of thermal and visible imagery for effective detection and tracking of pedestrians in videos. Cognitive Computation (9), 1–11 (2017)
24. Zhang, J., Ma, S., Sclaroff, S.: Meem: Robust tracking via multiple experts using entropy minimization. In: Proc. ECCV. pp. 188–203 (2014)
25. Zhang, K., Zhang, L., Liu, Q., Zhang, D., Yang, M.H.: Fast visual tracking via dense spatio-temporal context learning. In: Proc. ECCV. pp. 127–141 (2014)
26. Zhang, T., Xu, C., Yang, M.H.: Multi-task correlation particle filter for robust object tracking. In: Proc. IEEE Conf. CVPR. pp. 4819–4827 (2017)
27. Zhong, W., Lu, H., Yang, M.H.: Robust object tracking via sparse collaborative appearance model. IEEE TIP **23**(5), 2356 (2014)