The 9th International Conference on Brain-Inspired Cognitive System, 2018

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

Yulong Wang , Chenglong Li , Jin Tang , Dengdi Sun
wylemail@qq.com , lcl1314@foxmail.com , tj@ahu.edu.cn , sundengdi@163.com
School of Computer Science and Technology, Anhui University, Hefei, China

BICS 2018
Xi'an July,2018

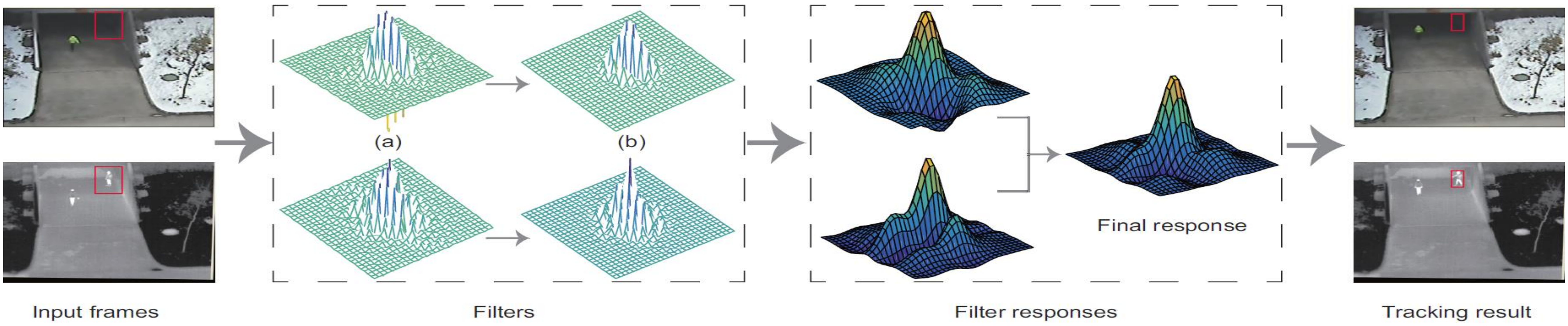


Figure 1: Pipeline of the proposed approach. (a) and (b) denote the correlation filters optimized by our proposed model without and with collaborative sparse constraint, respectively.

Background

- Tracking relies on a single sensor may be failed in challenging scenarios.
- How to perform efficient and effective fusion of different modalities for boosting tracking performance?
- A tracking speed beyond 25fps is considered real-time, some existing methods perform well but cannot be tracked in real time.

Contribution

- Motivated by brain inspired visual cognitive systems, we propose a novel approach that carries out efficient and effective fusion of multiple spectral data.
- we employ a sparse- and collaborative sparse-based regularizations on the joint filter to deploy both intra- and inter-modal complementary benefits from color and thermal spectrums.
- Extensive analysis and evaluation on large-scale benchmark datasets, i.e., GTOT and RGBT210, verify the effectiveness and efficiency of the proposed approach.

Collaborative Sparse Correlation Filter

Given *K* different spectrums, the goal is to find the optimal correlation filters $\mathbf{w}_k$ for *K* different spectrums.

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

Ideally, only one possible location corresponds to the target object. we suggest that the regularization should use the l_1 norm instead of the l_2 norm.

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

Among different spectrums, the learned w_k should select similar circular shifts so that they have similar motion. we use the convex $l_{2,1}$ mixed norm to learned their correlation filters jointly to distinguish the target from the background. Thus, the problem can be finally formulized as follows:

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

Tracking

- **Target position estimation**

For each modality with channel *D*, the response map is :

$$S = \sum_{k=1}^{K} \mathcal{F}^{-1}\left(\sum_{d=1}^{D} \hat{\mathbf{z}}_k^d \odot \hat{\mathbf{w}}_k^d\right)$$

The target location can be estimated by searching for the position of maximum value of the correlation response map *S*.

- **Model update**

Use an incremental strategy to update model as :

$$\mathcal{F}(\mathbf{w_k})^t = (1-\eta)\mathcal{F}(\mathbf{w_k})^{t-1} + \eta\mathcal{F}(\mathbf{w_k})^t$$

Experiments

- **Implementation**

MATLAB + i7-6700K 4.00 GHz CPU with 32 GB RAM.

- **Datasets: GTOT, and RGBT210**
  - C Li et al, Learning collaborative sparse representation for grayscale-thermal tracking, in TIP, 2016
  - C Li et al, Weighted sparse representation regularized graph learning for RGB-T object tracking, in ACM MM, 2017

- **Ablation Studies**

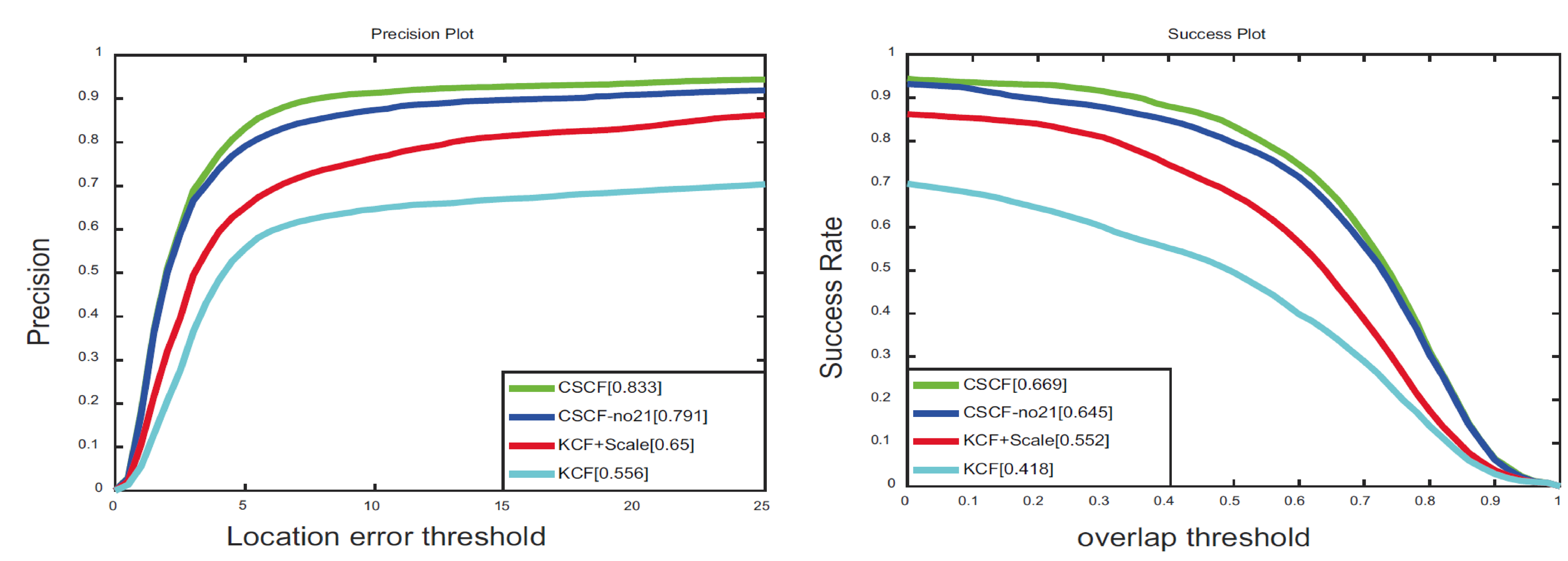


Figure 2: PR and SR plots on GTOT.

- **Quantitative Results**

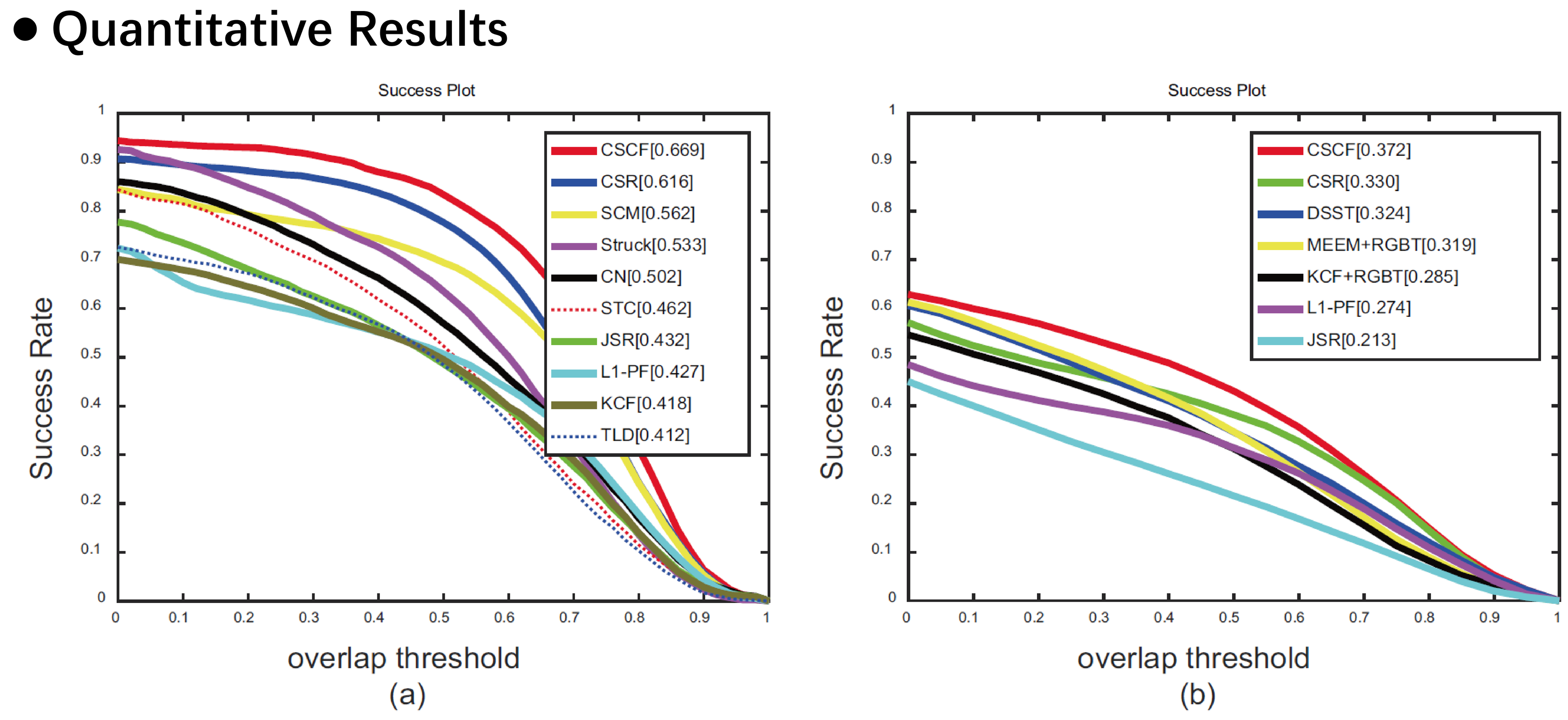
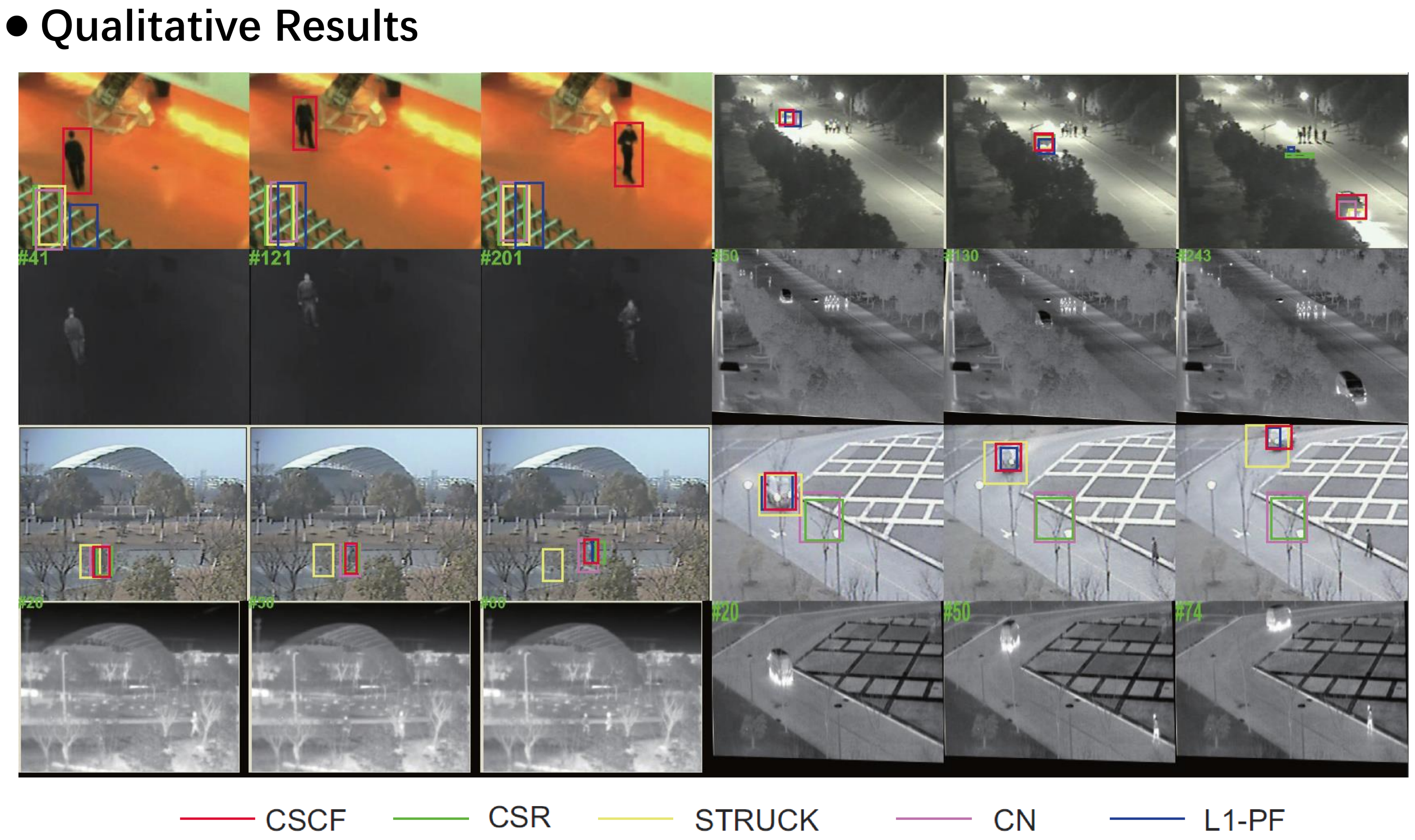


Figure 3: (a) and (b) denote the evaluation results on GTOT and RGB210 dataset, respectively.

- **Qualitative Results**



This work was jointly supported by National Natural Science Foundation of China (61672002, 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.