Learning Soft-Consistent Correlation Filters for RGB-T Object Tracking

Yulong Wang [1] , Chenglong Li [1,2] , Jin Tang [1]

wylemail@qq.com , lcl1314@foxmail.com , tj@ahu.edu.cn

[1] School of Computer Science and Technology, Anhui University, Hefei, China

[2] Center for Research on Intelligent Perception and Computing, NLPR, CASIA, Beijing, China

Background

- Tracking relies on a single sensor may be failed in challenging scenarios.

Occlusion Low visibility Thermal crossover

Large scale variation Small object Fast motion

- 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

- In order to take both ***collaboration*** and the ***heterogeneity*** of RGB and thermal spectra into account, a novel soft-consistent correlation filters for RGB-T object tracking is proposed.

- A spectral ***fusion mechanism*** is designed. The spectral weights are obtained according to the response map in the detection phase, and the final response map is obtained by weighted fusion of different spectra.

- Our method performs favorably against a number of state-of-the-art trackers, and the running speed over ***50 frames per second***.

Soft-Consistent Correlation Filter

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

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

The above objective function can equivalently be expressed in its dual form.

$$\min_{\mathbf{z}_k} \sum_{k=1}^{K} \frac{1}{4\lambda_1}\mathbf{z}_k^{\mathrm{T}}\mathbf{G}_k\mathbf{z}_k + \frac{1}{4}\mathbf{z}_k^{\mathrm{T}}\mathbf{z}_k - \mathbf{z}_k^{\mathrm{T}}\mathbf{y}$$

For the ***collaboration***, we observe that the learned $\{z_k\}$ should select similar circular shifts such that they have similar motion. While for the ***heterogeneity***, we intend to allow $\{z_k\}$ have sparse different elements to each other. Taking the above considerations together, we propose a ***soft-consistent constraint*** on $\{z_k\}$ that makes them consistent while allowing the sparse inconsistency exists, and formulated as a l_1-optimization based sparse learning problem. Thus, the problem can be finally formulized as follows:

$$\min_{\mathbf{z}_k} \sum_{k=1}^{K} \frac{1}{4\lambda_1}\mathbf{z}_k^{\mathrm{T}}\mathbf{G}_k\mathbf{z}_k + \frac{1}{4}\mathbf{z}_k^{\mathrm{T}}\mathbf{z}_k - \mathbf{z}_k^{\mathrm{T}}\mathbf{y} + \lambda_2 \sum_{k=2}^{K} ||\mathbf{z}_k - \mathbf{z}_{k-1}||_1$$

Tracking

- **Target position estimation**
  - The k-th correlation response map is computed by:

$$\mathbf{R}_k = \mathcal{F}^{-1}(\hat{\mathbf{s}}_k \odot \hat{\mathbf{x}}_k^* \odot \hat{\mathbf{z}}_k)$$

  - The definition of $APCE$ is:

$$APCE = \frac{|R_{max} - R_{min}|^2}{mean(\sum_{m,n}(R_{m,n} - R_{min})^2)}$$

  - The weights of different spectra are calculated based on $APCE$:

$$\alpha_k = \frac{APCE_k}{\sum_{k=1}^{K} APCE_k}$$

Then the final correlation response map is computed by

$$\mathbf{R} = \sum_{k=1}^{K} \alpha_k \mathbf{R}_k$$

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

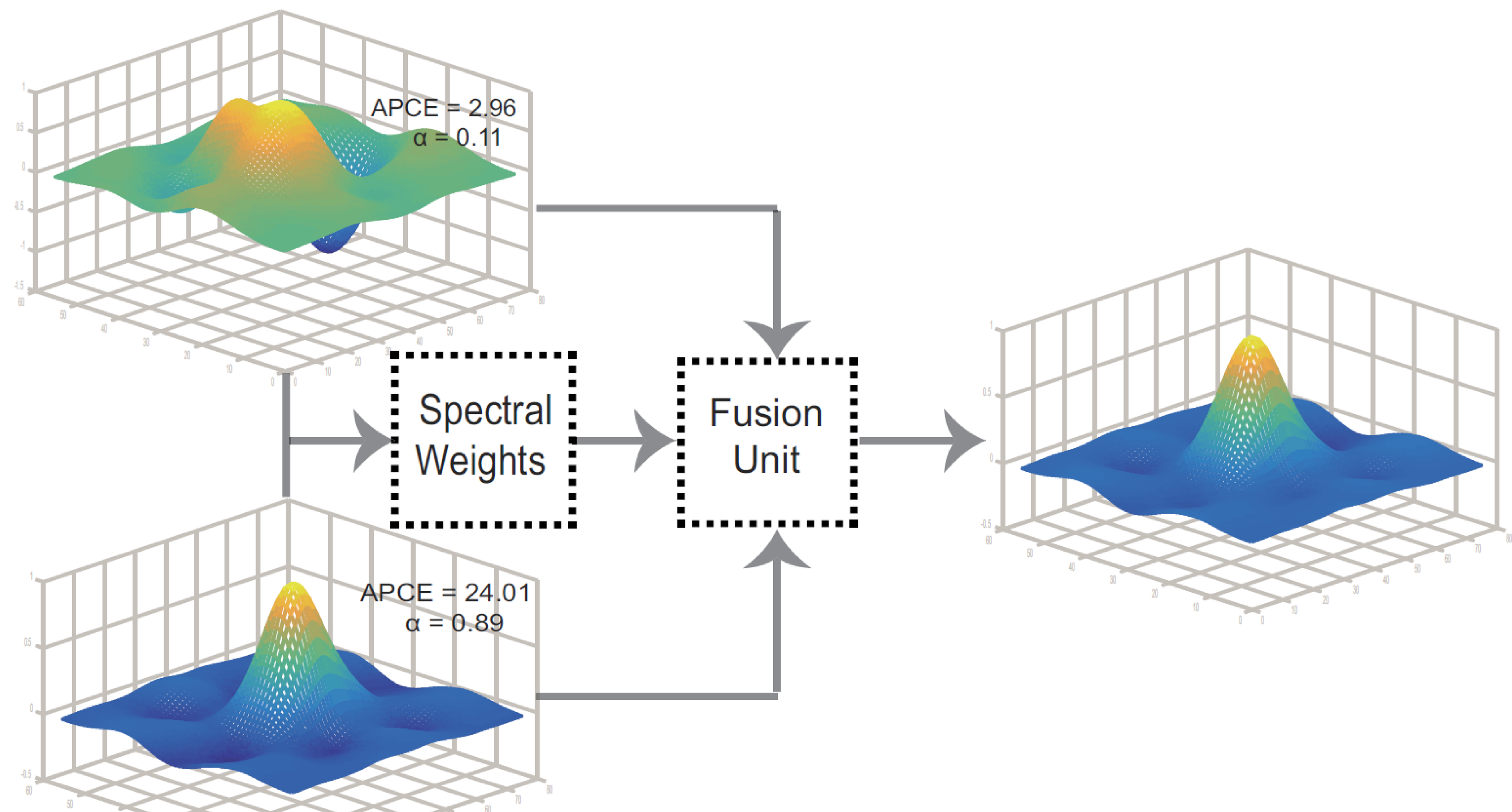

Figure 1: Pipeline of the proposed spectral fusion mechanism.

- **Model update**

Use an incremental strategy to update model as :

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

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

Experiments

- **Implementation**

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

- **Dataset: GTOT**
  - C Li et al, Learning collaborative sparse representation for grayscale-thermal tracking, in TIP, 2016

- **Quantitative Evaluation**

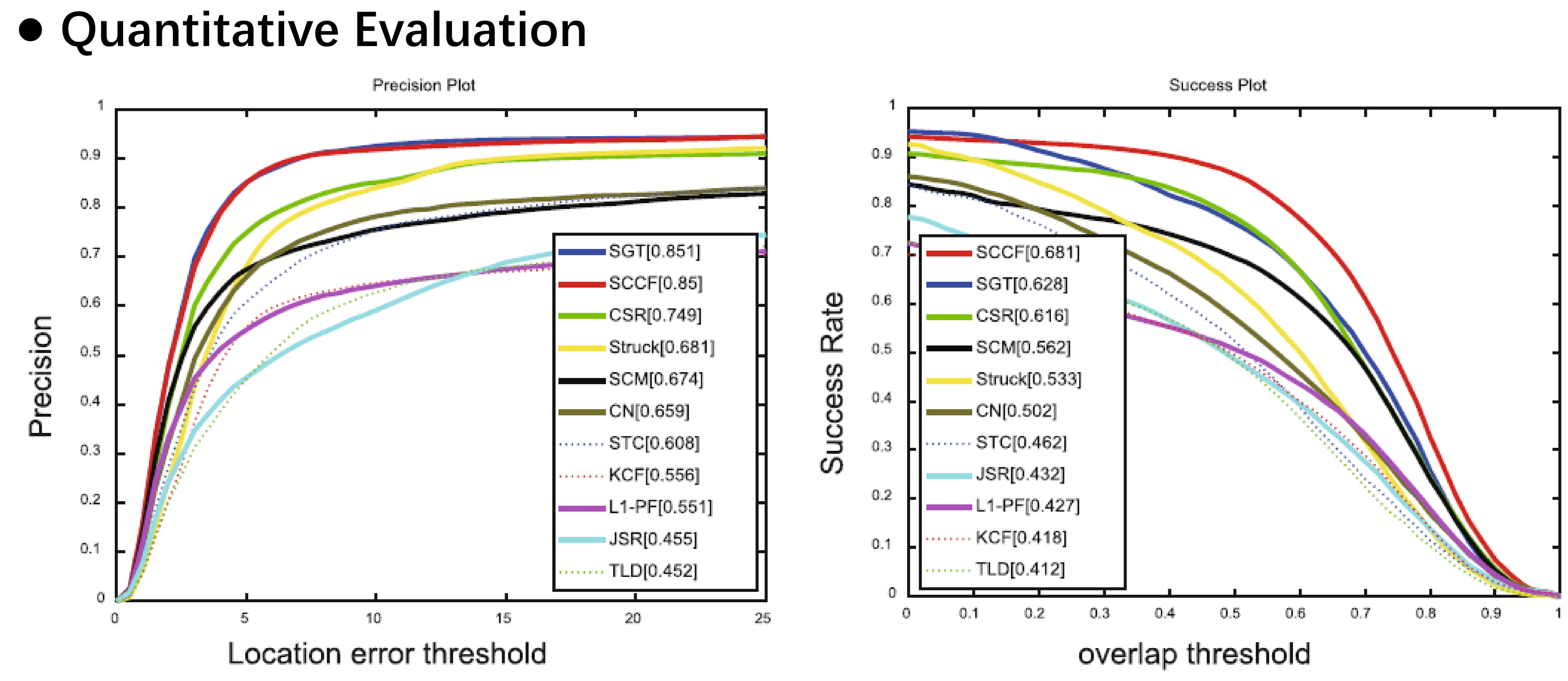

Figure 2: PR and SR plots on GTOT.

- **Qualitative Results**

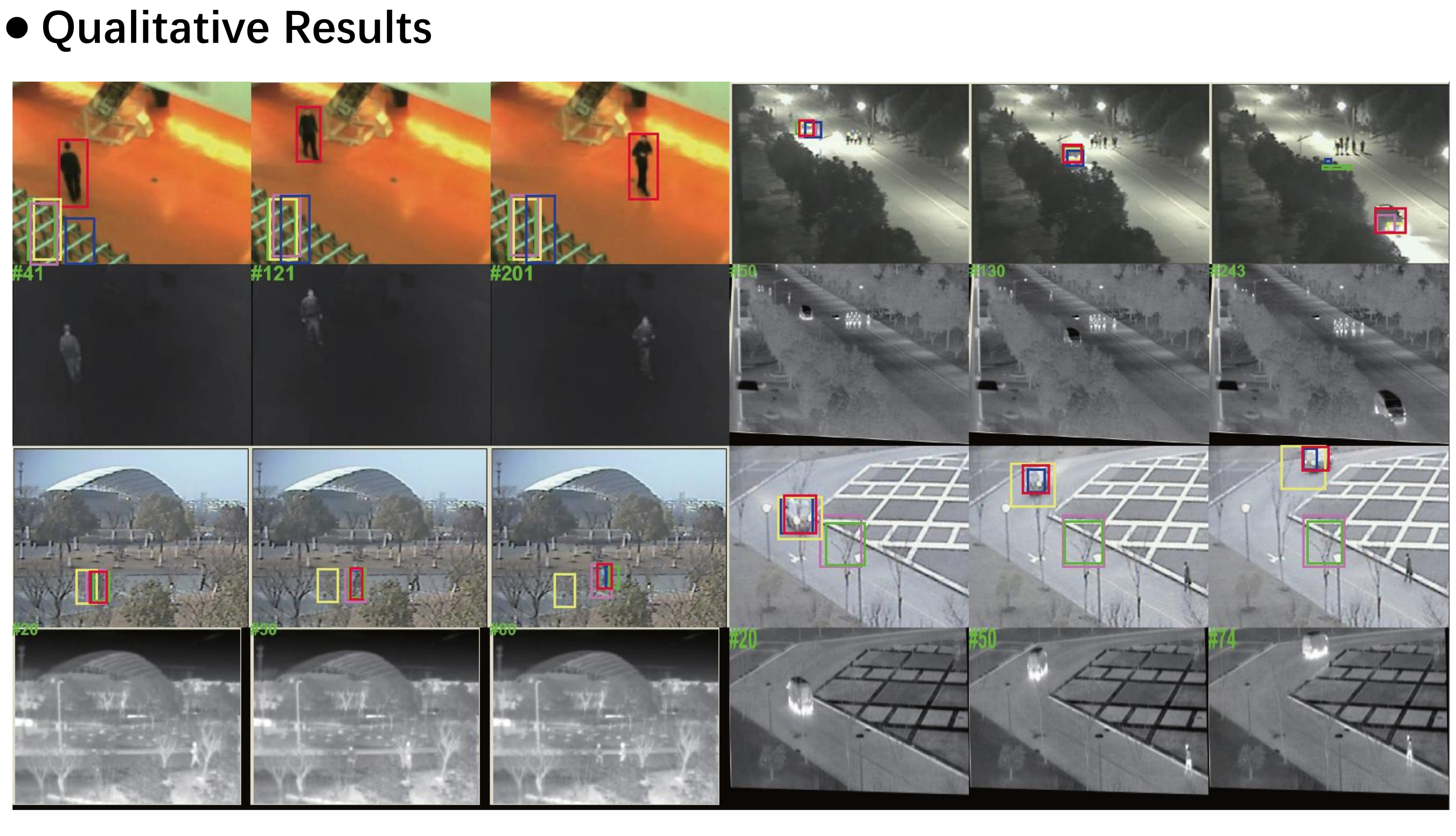

Figure 3: Sample results of our method against other tracking methods.

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