Robust correlation filter that combined temporal regularization with spatial weight L1 constraint

Xin He¹, Shukai Duan²,³,⁴,⁵* and Lidan Wang¹,³,⁴,⁵

¹College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China
²College of Artificial Intelligence, Southwest University, Chongqing 400715, China
³Brain-inspired Computing & Intelligent Control of Chongqing Key Lab, Chongqing 400715, China
⁴National & Local Joint Engineering Laboratory of Intelligent Transmission and Control Technology, Chongqing 400715, China
⁵Chongqing Brain Science Collaborative Innovation Center, Chongqing 400715, China

*Corresponding author’s e-mail: duansk@swu.edu.cn

Abstract. The parameter optimization of the Object tracking algorithms has achieved excellent performance based on parameter regularization. Adaptive spatial regularization correlation filter can obtain the optimal weight coefficient for a specific target. Inspired by this, we propose a filter combined temporal regularity with spatial L1 regularization (LTASRCF). The time regularizer is used to remember the data of the previous frame to update the current target position. It performs well in the process of occlusion, deformation and motion blur. Meanwhile, L1 regularization of spatial weights reduces the computational complexity of the coefficients, and improves the robustness of the spatial weight matrix. We have conducted comparative experiments on the OTB-2015 dataset, and the proposed LTASRCF model ranks first among the five popular algorithms, which outperforms by 9.2% and 18.9% than KCF, respectively, and reaches real-time performance.

1. Introduction

Recently, visual tracking algorithms based on the correlation filter has been widely used in some areas, such as video surveillance, automation, scene recognition. Correlation filter uses correlation score between the sample and the filter template to acquire the most satisfied target position, these algorithms have generated a great deal of commercial and military value by continuous optimization.

Discriminative correlation filters (DCFs) learn correlation filter coefficients by converting samples to frequency domain. However, due to the periodic repetitions of the cyclic shift operator, the performance of target tracking is reduced. Thus many scholars begin to pay attention to the optimization of the parameters and contents of the model. These methods can be roughly divided into three aspects: scale-adaptive [1,2,3], feature representation (eg. color names [4], HOG [5] and deep CNN feature [6]), other optimized ideas (statistical method [7], background modeling [8], kernel [9], time regular operator [10], etc). Proposed models improve the tracking accuracy and robustness to a certain extent, however, faced with the challenges of 11 video attributes of OTB-2015 [11] dataset,
their performances are poor. The occlusion, fast motion, motion blur and other attributes during the tracking process require higher robustness to the model and filter parameters. KCF [7] suffered a sharp drop in accuracy coped with most video attributes while increasing speed. SRDCF [12] ignored the effect of scale when introducing spatial regularization, and performed not well in challenges such as rotation. In recent years, many algorithms based on convolutional neural networks have also been proposed, such as DeepSRDCF [13], ECO [14], etc. Most algorithms use neural networks to extract more robust features, although they have achieved certain excellent performance in accuracy, their speed is far from real-time. At the beginning of CVPR2018, a brand new method--Siamese neural networks, has draw a lot of attention. This method firstly to take advantage of the powerful representation capabilities of deep features to achieve the best match. However, there is still no real-time, and the performance is not optimistic. The time complexity of using a neural network-based model is higher and accompanied by a lot of calculation costs. Therefore, balancing the accuracy and speed of object tracking algorithm is still two tricky problems in this field.

Based on the model and performance of the adaptive spatial regular correlation filter (ASRCF [15]), we develop a correlation filter that fuses temporal regularity and spatial L1 regularization. Our contributions are as follows,

1. We introduce a temporal regularization into ASRCF, aiming to relocate current object during the background occlusion or target loss, which greatly improves the reliability of the model.

2. In this paper, the L1 regularization is introduced into the spatial weight regularization, which makes the weight coefficient more robust and reduces the boundary effect.

3. An ADMM algorithm was used to solve the model, this work reduce the computational complexity. We perform a large number of experiments to achieve excellent performance on OTB--2015 dataset.

2. Related Work

2.1 Adaptive Spatial Regularization

MOSSE [16] first introduced CFs (Correlation filters) to object tracking, and used a ridge regression model to fit filter parameters. After transforming to the frequency domain, the objective function was derived to obtain the optimal solution of the filter. However, the periodic repetition of boundary elements reduces the performance of the algorithm. On the basis of MOSSE, the authors of the SRDCF introduced spatial regularity. This spatial regular matrix is a matrix with small values on boundary and large values around of the center of target. The model will penalize pixels that are far away from the target center according to the target space position. Because the background of the target tracking is not modeled over time, it can lead to sub-optimal performance of the tracker. BACF [8] introduced the background cropping operator to model the background. They extracted reliable information from the background as negative samples to distinguish the target foreground and the background. The background information as a real negative sample reduced boundary effect. Therefore, ASRCF learned the advantages of SRDCF and BACF and proposed an adaptive spatial regular correlation filter. The model of ASRCF inherited the original model of MOSSE, remembered the spatial regular expression of SRDCF, and learned the model idea of background, the spatial weight matrix will change for specific targets. This contribution makes the ASRCF model more reliable in the tracking and localization process.

2.2 Temporal regularizer

Temporal regularization use the position information of the previous frame combined with the modeling of the current frame to predict the position and scale information of the target in the current frame. STRCF first introduced temporal operator to correlation filtering and used PA algorithm modeling. This model has two advantages: on the one hand, this model can update the current classifier passively based on the previous frame information; on the one hand, the correct classification of the current pixel was determined aggressively. These two advantages make STRCF
use the remodeling of the target to locate the current target during occlusion and motion blur, and are suitable for large-scale samples. These merits make STRCF locate the current target accurately during occlusion and motion blur.

3. Temporal and Spatial L1 regularization correlation filter

3.1 LTASRCF
Correlation Filter: MOSSE display their model as follows,

$$\text{arg min}_h \| x_k h_k - y \|^2_2$$

where $x_k \in R^{T \times d}$ denotes the sequence sample of $k_{th}$ channel, $h_k \in R^{T \times d}$ denotes filter, $y \in R^{T \times 1}$ is ideal Gaussian-shaped response. * stands for correlation operator.

SRDCF: SRDCF introduces a spatial weight matrix to reduce boundary effects,

$$\text{arg min}_h \frac{1}{2} \sum_{k=1}^{K} x_k h_k - y \|^2_2 + \frac{\lambda_1}{2} \sum_{k=1}^{K} w \cdot h_k \|^2_2$$

w aims to acquire a higher response around the center of the target, $\lambda$ denotes regularization parameter.

ASRCF: ASRCF proposes adaptive spatial regularization correlation filter, merged background crop operator into their job, and extend SRDCF and BACF as:

$$\text{arg min}_h \frac{1}{2} \sum_{k=1}^{K} x_k (P^T h_k) - y \|^2_2 + \frac{\lambda_2}{2} \sum_{k=1}^{K} w \cdot h_k \|^2_2$$

Transforming to frequency domain using Parseval’s theorem:

$$\text{arg min}_h \frac{1}{2} \sum_{k=1}^{K} g^\wedge_k - y \|^2_2 + \frac{\lambda_1}{2} \sum_{k=1}^{K} w \cdot h_k \|^2_2 + \frac{\lambda_2}{2} \|w - w'\|^2_2$$

$$g^\wedge_k = \sqrt{T} F P^T h_k, k = 1, \ldots, K$$

where $P \in R^{T \times T}$ is a diagonal binary matrix to crop the D elements of $x_k$. $\lambda_1$ is regularization parameter. $w'$ is a reference weight $\|w - w'\|^2_2$ stands for spatial regularization operator, $\lambda_2$ denotes adaptive spatial regularization parameter, $g^\wedge$ is a auxiliary variable matrix.

LTASRCF: We develop a temporal regularization and spatial weight L1 constraint based on ASRF and reformulate Eqn. (4) as

$$\text{arg min}_h \frac{1}{2} \sum_{k=1}^{K} g_k - y \|^2_2 + \frac{\lambda_1}{2} \sum_{k=1}^{K} w \cdot h_k \|^2_2 + \frac{\lambda_2}{2} \|w - w'\|^2_2 + \frac{\lambda_3}{2} \|g_k - g_{k-1}\|^2_2 + \frac{\lambda_4}{2} \|w\|_1$$

where $\|g_k - g_{k-1}\|^2_2$ denotes temporal regularization term, $\|w\|_1$ stands for sparse constraint of w.

LTASRCF not only inherits the merits of adaptive spatial regularization algorithm, but also adds two majorization to optimize the filter and weight coefficient.

3.2 ADMM solution
By introducing an auxiliary variable $z = w$, the Augmented Lagrangian representation of Eqn. (5) follows the form
\[
L(H, \hat{G}, w, S, f) = \frac{1}{2} \sum_{k=1}^{K} \| x_k \|^2 + \frac{1}{2} \sum_{k=1}^{K} \| w \cdot h_k \|^2 + \frac{\mu_1}{2} \| w - w' \|^2 \\
+ \frac{\lambda_2}{2} \| g_k - g_{k-1} \|^2 + \| \sqrt{T} F P^T h_k + s_k \|_2^2 + \frac{\mu_3}{2} \| z \|_1 
\]

(6)

The ADMM is then adopted by solving alternative subproblem.

**Subproblem h:**

\[
h^*_k = \arg \min_h \left\{ \hat{x}_k \right\} + \frac{\lambda_2}{2} \| w \cdot h_k \|^2 + \frac{\mu_1}{2} \| g_k - \sqrt{T} F P^T h_k + s_k \|_2^2 \}
\]

(7)

**Subproblem g:**

\[
g^*_k = \arg \min_g \frac{1}{2} \sum_{k=1}^{K} \| x_k \cdot g_k - y \|^2 + \frac{\mu_1}{2} \sum_{k=1}^{K} \| g_k - \sqrt{T} F P^T h_k + s_k \|^2 + \frac{\lambda_2}{2} \| g_k - g_{k-1} \|^2
\]

(8)

We process the features on all channels and differentiate the equation by Sherman-Morrison formula,

\[
V_j(\hat{G}) = \arg \min_{V_j(\hat{G})} \frac{1}{2} \| V_j(\hat{X})^T V_j(\hat{G}) - y_j \|^2 + \frac{\mu_1}{2} \sum_{k=1}^{K} \| V_j(\hat{G}) - V_j(\sqrt{T} F P^T H) + V_j(S) \|^2 + \frac{\lambda_2}{2} \| V_j(\hat{G}) - V_j(G_{k-1}) \|^2
\]

(9)

where \( V_j(\hat{G}) \in \mathbb{R}^{K \times 1} \) represents the value of all channels of filter.

\[
V_j(\hat{G}) = \frac{1}{\mu_2 + \lambda_3} \left( I - \frac{V_j(\hat{X})^T V_j(\hat{G})}{V_j(\hat{X})^T V_j(\hat{X}) + \mu_2} + \lambda_3 V_j(\sqrt{T} F P^T F) + \lambda_2 V_j(\hat{G}_{k-1}) - \mu_2 V_j(S) \right)
\]

(10)

**Subproblem w:**

\[
w^*_k = \arg \min_w \frac{\lambda_2}{2} \sum_{k=1}^{K} \| w \cdot h_k \|^2 + \frac{\lambda_3}{2} \| w - w' \|^2 + \| w - z + f \|^2
\]

(11)

\[
= (\lambda_2 \sum_{k=1}^{K} N_k^T N_k + \lambda_3 I + \mu_3)^{-1} (\lambda_2 w' + \mu_3 z - \mu_3 f)
\]

**Subproblem z:**

\[
z^*_k = \frac{\mu_2}{2} \| z \| + \frac{\mu_3}{2} \| w - z + f \|^2
\]

(12)

Update parameter:
4. Experiments

4.1 Equipment and selection
All experiments are conducted under RAM 16GB, Core i7-8700K CPU, NVIDIA GTX 1080Ti, equipped with MATLAB2017b. The parameters of solution are selected as $\lambda_1 = 4$, $\lambda_2 = 0.1$, $\lambda_3 = 6$, $\mu_1 = 5$, the max iterate times of ADMM are 2, penalty factor are $\lambda = 10$ and $\mu_{max} = 10000$.

4.2 Evaluation toolkit
This paper uses the OTB-2015 dataset to evaluate the algorithm, OTB-2015 contains 100 different classes of sequences and 11 different attributes (e.g. occlusion, lighting changes, deformation, fast movement). We use the precision plot and success plot of these 11 attributes to compare with several current algorithms: SRDCF, KCF, DSST [1], SAMF [3].

4.3 Comparisons on precision and success plot
![Figure 1](image1.png)

Fig.1 indicates that the precision and success rate of LTASRCF are 78.6% and 73.7%, respectively. LTASRCF takes the first place in precision and success rate, and exceed the KCF tracker by 9.2% and 18.9%, respectively. Experimental results demonstrate that our algorithm outperforms than several other algorithms. As shown in Table 1, LTASRCF is nearly five times higher than SRDCF in FPS.

|         | SRDCF | KCF  | DSST | SAMF | LTASRCF |
|---------|-------|------|------|------|---------|
| Accuracy| 72.6  | 54.9 | 60.0 | 67.2 | **78.6**|
| FPS     | 5.8   | 171.8| 20.4 | 23.2 | **25.6**|

5. Conclusions
In order to overcome the challenges of attributes such as occlusion, deformation and motion blur during tracking process, this paper introduce temporal regularization to update the current target position information. The L1-norm is proposed for the spatial weight matrix, which improves the robustness of the weight matrix. Experimental results on the OTB-2015 show that the accuracy and
success rate of LTASRCF are 0.786 and 0.737 respectively, ranking first among the five algorithms, which are 9.2% and 18.9% higher than the KCF algorithm, respectively. Experiments demonstrate that LTARSCF can effectively improve the robustness for different video attributes and have a high application value.

Acknowledgements
This work was supported by National Key R&D Program of China (Grant Nos. 2018YFB1306600), National Natural Science Foundation of China (Grant Nos. 61571372, 61672436, 61601376), Fundamental Science and Advanced Technology Research Foundation of Chongqing (cstc2017jcyjBX0050, cstc2016jcyjA0547), Fundamental Research Funds for the Central Universities (Grant Nos. XDJK2016A001, XDJK2017A005).

References
[1] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg. (2017) Discriminative scale space tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39:1561-1575.
[2] F. Li, Y. Yao, P. Li, D. Zhang, W. Zuo, and M. H. Yang. (2017) Integrating boundary and center correlation filters for visual tracking with aspect ratio variation. IEEE International Conference on Computer Vision Workshops.
[3] Y. Li and J. Zhu. (2014) A scale adaptive kernel correlation filter tracker with feature integration. European IEEE International Conference on Computer Vision Workshops.
[4] M. Danelljan, F. S. Khan, M. Felsberg, and J. V. D. Weijer. (2014) Adaptive color attributes for real-time visual tracking. IEEE Conference on Computer Vision and Pattern Recognition.
[5] N. Dalal and B. Triggs. (2005) Histograms of oriented gradients for human detection. IEEE Conference on Computer Vision and Pattern Recognition.
[6] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang. (2015) Hierarchical convolutional features for visual tracking. IEEE Conference on Computer Vision.
[7] Alan Lukezic, Tomas Vojir, Luka Cehovin Zajc, Jiri Matas, Matej Kristan. (2017) Discriminative Correlation Filter with Channel and Spatial Reliability. IEEE Conference on Computer Vision and Pattern Recognition.
[8] H. Kiani Galoogahi, A. Fagg, and S. Lucey. (2017) Learning background-aware correlation filters for visual tracking. IEEE Conference on Computer Vision.
[9] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista. (2015) Highspeed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37:583–596.
[10] Feng Li, Cheng Tian, Wangmeng Zuo, Lei Zhang, and Ming-Hsuan Yang. (2018) Learning Spatial-Temporal Regularized Correlation Filters for Visual Tracking. IEEE Conference on Computer Vision and Pattern Recognition.
[11] Y. Wu, J. Lim, and M.-H. Yang. (2015) Object tracking benchmark. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9):1834–1848.
[12] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg. (2015) Learning spatially regularized correlation filters for visual tracking. IEEE Conference on Computer Vision.
[13] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg. (2015) Convolutional features for correlation filter based visual tracking. IEEE Conference on Computer Vision Workshops.
[14] M. Danelljan, G.Bhat, F. S. Khan, and M. Felsberg. (2017) Eco: Efficient convolution operators for tracking. IEEE Conference on Computer Vision and Pattern Recognition.
[15] Kenan Dai, Dong Wang, Huchuan Lu, Chong Sun, Jianhua Li. (2019) Visual Tracking via Adaptive Spatially-Regularized Correlation Filters. IEEE Conference on Computer Vision and Pattern Recognition.
[16] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui. (2010) Visual object tracking using adaptive correlation filters. IEEE Conference on Computer Vision and Pattern Recognition.

[17] Stephen P. Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine earning, 3(1):1–122.