Learning Multi-task Correlation Particle Filters for Visual Tracking

Tianzhu Zhang, Changsheng Xu, and Ming-Hsuan Yang

In the supplementary material, we present the details of optimization, more results, and source code.

1 Optimization

In this section, we describe how to solve (1) efficiently by using the accelerated proximal gradient (APG) method, which has been widely applied to solve convex optimization problems with non-smooth terms [1], [2]. Compared to conventional subgradient based methods that converge at the sublinear rate, the APG can obtain globally optimal solution at the quadratic convergence rate and achieves $\mathcal{O}(\frac{1}{t^2})$ residual from the optimal solution after t iterations [2]. At each iteration, the gradient mapping and aggregation steps are involved to solve the optimization problems.

$$\min_{\{\mathbf{z}_{pk}\}} \sum_{p,k} \frac{1}{4}\mathbf{z}_{pk}^\top \mathbf{G}_{pk}\mathbf{z}_{pk} + \frac{1}{4}\lambda \mathbf{z}_{pk}^\top \mathbf{z}_{pk} - \lambda \mathbf{z}_{pk}^\top \mathbf{y} + \gamma \|\mathbf{Z}\|_{2,1}, \quad (1)$$

where γ is a tradeoff parameter between reliable reconstruction and joint sparsity regularization. $\mathbf{Z} = [\mathbf{Z}_1, \cdots, \mathbf{Z}_p, \cdots, \mathbf{Z}_P] \in \mathbb{R}^{MN \times PK}$. Here, for the p-th part, the corresponding filters of the K different features form a matrix $\mathbf{Z}_p = [\mathbf{z}_{p1}, \cdots, \mathbf{z}_{pk}, \cdots, \mathbf{z}_{pK}] \in \mathbb{R}^{MN \times K}$. The definition of the $\ell_{p,q}$ mixed norm is $\|\mathbf{Z}\|_{p,q} = \left(\sum_i \left(\sum_j |[\mathbf{Z}]_{ij}|^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}}$ and $[\mathbf{Z}]_{ij}$ denotes the entry at the i-th row and j-th column of $\mathbf{Z}$. To solve (1),we first denote $f(\mathbf{Z})$ and $g(\mathbf{Z})$ as in [1]:

$$\begin{aligned} f(\mathbf{Z}) &= \sum_{p,k} \frac{1}{4}\mathbf{z}_{pk}^\top \mathbf{G}_{pk}\mathbf{z}_{pk} + \frac{1}{4}\lambda \mathbf{z}_{pk}^\top \mathbf{z}_{pk} - \lambda {\mathbf{z}_{pk}}^\top \mathbf{y}, \\ g(\mathbf{Z}) &= \gamma \|\mathbf{Z}\|_{2,1}, \end{aligned} \quad (2)$$

where $f(\mathbf{Z})$ is the loss function and $g(\mathbf{Z})$ is the regularization term. We note that the objective function in (1) is a composite function of a differential term $f(\mathbf{Z})$ and a non-differential term $g(\mathbf{Z})$. We define

$$\begin{aligned} \Phi\left(\mathbf{Z}, \mathbf{R}_{(t)}\right) &= f\left(\mathbf{R}_{(t)}\right) + \left\langle \nabla f\left(\mathbf{R}_{(t)}\right), \mathbf{Z} - \mathbf{R}_{(t)} \right\rangle \\ &+ \frac{\eta}{2}\left\|\mathbf{Z} - \mathbf{R}_{(t)}\right\|_F^2 + g(\mathbf{Z}), \end{aligned} \quad (3)$$

with the regularization term $g(\mathbf{Z})$ and the approximation of $f(\mathbf{Z})$ by the first order Taylor expansion at point $\mathbf{R}_{(t)}$ with the squared Euclidean distance between $\mathbf{Z}$ and $\mathbf{R}_{(t)}$ as the regularization term. Here, $\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{Tr}\left(\mathbf{A}^\top \mathbf{B}\right)$ denotes the matrix inner product, η is a parameter controlling the step penalty, $\nabla f\left(\mathbf{R}_{(t)}\right)$ denotes the partial derivative of $f(\mathbf{Z})$ with respect to $\mathbf{Z}$ at point $\mathbf{R}$. The derivative computed by:

$$\nabla f(\mathbf{Z}) = \sum_{p=1}^{P}\sum_{k=1}^{K} \frac{1}{2}\mathbf{G}_{pk}\mathbf{Z}\mathbf{I}_{pk}\mathbf{I}_{pk}^\top + \frac{\lambda}{2}\mathbf{Z}\mathbf{I}_{pk}\mathbf{I}_{pk}^\top - \lambda \mathbf{y}_{pk}\mathbf{I}_{pk}^\top, \quad (4)$$

where $\mathbf{I}_{pk} = [\mathbf{0}, \ldots, \mathbf{1}, \ldots, \mathbf{0}]^\top \in \mathbb{R}^{PK \times 1}$ includes PK elements and the pk-th element is 1 and the others are zero.

The solution of (1) at the t-th iteration ($t \geq 1$) can be computed by the following proximal operator:

$$\mathbf{Z}_{(t)} = \arg\min_{\mathbf{Z}} \Phi\left(\mathbf{Z}; \mathbf{R}_{(t)}\right). \quad (5)$$

Here, $\mathbf{R}_{(1)} = \mathbf{Z}_{(0)}$. Given the current estimate $\mathbf{R}_{(t)}$, we can rewrite (5) and obtain $\mathbf{Z}$ at the t-th iteration $\mathbf{Z}_{(t)}$ by solving (6),

$$\mathbf{Z}_{(t)} = \arg\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{Z} - \mathbf{H}\|_F^2 + \frac{\gamma}{\eta}\|\mathbf{Z}\|_{2,1} \quad (6)$$

$$= \arg\min_{\mathbf{Z}^1, \cdots, \mathbf{Z}^m} \sum_{i=1}^{m} \left(\frac{1}{2}\left\|\mathbf{Z}^i - \mathbf{H}^i\right\|_2^2 + \frac{\gamma}{\eta}\left\|\mathbf{Z}^i\right\|_2\right), \quad (7)$$

where $\mathbf{H} = \mathbf{R}_{(t)} - \frac{1}{\eta}\nabla f\left(\mathbf{R}_{(t)}\right)$, and $\mathbf{Z}^i$ as well as $\mathbf{H}^i$ denote the i-th row of the matrix $\mathbf{Z}$, $\mathbf{H}$, respectively. According to [1], for each row of $\mathbf{Z}^i$ in the subproblem (7), an efficient closed-form solution can be computed:

$$\mathbf{Z}_{(t)}^i = \max(0, 1 - \frac{\gamma}{\eta\|\mathbf{H}^i\|_2})\mathbf{H}^i. \quad (8)$$

After obtaining the representation matrix $\mathbf{Z}_{(t)}$, we update the aggregation matrix $\mathbf{R}_{(t)}$ as a linear combination of $\mathbf{Z}_{(t)}$ and $\mathbf{Z}_{(t-1)}$ to store the aggregation of $\mathbf{Z}$ in the previous iterations. As suggested in [1], we update $\mathbf{R}_{(t)}$ as in (9), where α_t is set to $\frac{2}{t+3}$ for $t \geq 1$ and $\alpha_0 = 1$ for $t = 0$ by

$$\mathbf{R}_{(t+1)} = \mathbf{Z}_{(t)} + \alpha_t\left(\frac{1}{\alpha_{t-1}} - 1\right)\left(\mathbf{Z}_{(t)} - \mathbf{Z}_{(t-1)}\right). \quad (9)$$

- *Tianzhu Zhang is with National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, P. R. China. E-mail: tzzhang@nlpr.ia.ac.cn.*
- *Changsheng Xu is with National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, P. R. China. E-mail: csxu@nlpr.ia.ac.cn.*
- *Ming-Hsuan Yang is with School of Engineering, University of California, Merced, CA 95344. E-mail: mhyang@ucmerced.edu.*

Algorithm 1: *Proposed APG optimization to solve* (1)

Input : $\mathbf{G}_k$, $\mathbf{y}$, $k = 1, \cdots, K$, λ, γ, and η
Output: $\mathbf{Z}$

Initialize $t \leftarrow 1$, $\mathbf{Z}_{(0)} = \mathbf{0}$, $\mathbf{R}_{(1)} = \mathbf{0}$, $\alpha_0 = 1$
while *not converged* **do**
  Compute $\mathbf{H} = \mathbf{R}_{(t)} - \frac{1}{\eta}\nabla f\left(\mathbf{R}_{(t)}\right)$
  Update $\mathbf{Z}_{(t)}$ via (6)
  $\alpha_t = \frac{2}{t+3}$
  Update $\mathbf{R}_{(t+1)}$ via (9)
  $t \leftarrow t + 1$
end

The main steps of our APG approach for computing the optimization problem (1) is summarized in Algorithm 1. The proposed algorithm stops when the relative change in the solution or objective function falls below a predefined threshold. It is time-consuming to compute the partial derivative (4) because of $\mathbf{G}_{pk} = \mathbf{X}_{pk}\mathbf{X}_{pk}^{\top}$. However, it can be calculated very efficiently in the Fourier domain by considering the circulant structure property of $\mathbf{X}_{pk}$.

In (4), we denote $\mathbf{s}_{pk} = \mathbf{G}_{pk}\mathbf{z}_{pk} = \mathbf{X}_{pk}\mathbf{X}_{pk}^{\top}\mathbf{z}_{pk}$. Assume $\mathbf{x}_{pk}$ is the base sample of $\mathbf{X}_{pk}$, the $\mathbf{s}_{pk}$ can be updated with only the base sample as (10).

$$\hat{\mathbf{s}}_{pk} = \hat{\mathbf{x}}_{pk}^{*} \odot \hat{\mathbf{x}}_{pk} \odot \hat{\mathbf{z}}_{pk}. \tag{10}$$

Here, $\mathbf{x}^*$ is the complex-conjugate of $\mathbf{x}$, $\hat{\mathbf{x}}$ denotes the Discrete Fourier Transform (DFT) of the generating vector $\hat{\mathbf{x}} = \mathcal{F}(\mathbf{x})$, and $\odot$ denotes the element-wise product. Finally, the $\mathbf{s}_{pk}$ can be obtained via $\mathbf{s}_{pk} = \mathcal{F}^{-1}(\hat{\mathbf{s}}_{pk})$. Here, for a 1D signal $\mathbf{x}$, the $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the 1D DFT and its inverse. When $\mathbf{x}$ is 2D, $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the 2D DFT and its inverse. As such, the optimization in the Fourier domain is efficient for 2D image patches with multiple channels. After optimization, we can obtain the multi-task correlation filter $\mathbf{z}_{pk}$ for the p-th part with the k-th feature.

2 Results and Source Code

The video results and the source code of our conference paper [3] are available on the web site: http://nlpr-web.ia.ac.cn/mmc/homepage/tzzhang/lmcpf.html. In the source code, the variances of affine parameters for particle sampling are set to $(0.01, 0.0001, 0.0001, 0.01, 2, 2)$, and the particle number is set to 50. The code can be run under the visual tracking evaluation protocol [4]. To run the code, please download the VGG-Net-19 model. The details are introduced in the readme.txt in the model folder.

To generate the parts, we use the spatial layout as shown in Figure 1 to sample 3 parts based on the targets height-width ratio. If the ratio of height and width for a target object is greater than 1, we use $2/3$ of the height from the top and bottom as well as in the center to obtain the parts. Similarly, we can represent objects if the ratio of height and width is less than 1 (i.e., $2/3$ of the width from left, right and center). We note that this simple representation performs well in practice, and other part-based methods can also be adopted.

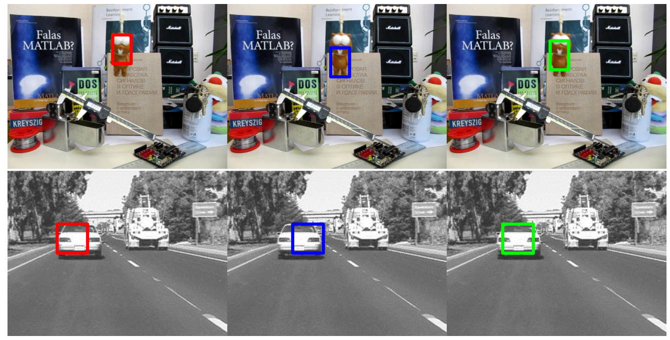

Fig. 1. The sampled 3 parts based on the targets ratio.

For more details, please refer to our other papers [3], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18].

References

[1] X. Chen, W. Pan, J. Kwok, and J. Carbonell, "Accelerated gradient method for multi-task sparse learning problem," in *ICDM*, 2009. 1

[2] Y. Nesterov, "Gradient methods for minimizing composite objective function," in *Center for Operations Research and Econometrics (CORE), Universit catholique de Louvain,76*, 2007. 1

[3] T. Zhang, C. Xu, and M.-H. Yang, "Multi-task correlation particle filter for robust object tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1–9. 2

[4] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *CVPR*, 2013. 2

[5] S. Liu, T. Zhang, X. Cao, and C. Xu, "Structural correlation filter for robust visual tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 2

[6] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via structured multi-task sparse learning," *International Journal of Computer Vision*, vol. 101, no. 2, pp. 367–383, 2013. 2

[7] T. Zhang, S. Liu, N. Ahuja, M.-H. Yang, and B. Ghanem, "Robust Visual Tracking via Consistent Low-Rank Sparse Learning," *International Journal of Computer Vision*, vol. 111, no. 2, pp. 171–190, 2015. 2

[8] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via multi-task sparse learning," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 2

[9] T. Zhang, A. Bibi, and B. Ghanem, "In defense of sparse tracking: Circulant sparse tracker," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 2

[10] T. Zhang, C. Jia, C. Xu, Y. Ma, and N. Ahuja, "Partial occlusion handling for visual tracking via robust part matching," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2014. 2

[11] T. Zhang, S. Liu, C. Xu, S. Yan, B. Ghanem, N. Ahuja, and M.-H. Yang, "Structural sparse tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2015. 2

[12] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Low-rank sparse learning for robust visual tracking," in *Proceedings of European Conference on Computer Vision*, 2012. 2

[13] T. Zhang, B. Ghanem, S. Liu, C. Xu, and N. Ahuja, "Robust Visual Tracking via Exclusive Context Modeling," *IEEE transactions on cybernetics*, vol. 46, no. 1, pp. 51–63, 2016. 2

[14] T. Zhang, C. Xu, and M.-H. Yang, "Robust structural sparse tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–14, 2018. 2

[15] T. Zhang, S. Liu, C. Xu, B. Liu, and M.-H. Yang, "Correlation particle filter for visual tracking," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2676–2687, 2018. 2

[16] J. Gao, T. Zhang, X. Yang, and C. Xu, "Deep relative tracking," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1845–1858, 2017. 2

[17] T. Zhang, C. Xu, and M.-H. Yang, "Learning multi-task correlation particle filters for visual tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–14, 2018. 2

[18] J. Gao, T. Zhang, X. Yang, and C. Xu, "P2t: Part-to-target tracking via deep regression learning," *IEEE Transactions on Image Processing*, vol. PP, no. 99, pp. 1–1, 2018. 2