Noname manuscript No.
(will be inserted by the editor)

Adaptive Channel Selection for Robust Visual Object Tracking with Discriminative Correlation Filters

Supplementary Material

Tianyang Xu · Zhen-Hua Feng · Xiao-Jun Wu · Josef Kittler

Received: date / Accepted: date

1 Optimisation Details

The augmented Lagrangian is as follows:

$$\begin{aligned}\mathcal{L}\left(\mathcal{W},\mathcal{W}',\boldsymbol{\Gamma},\mu\right) = &\left\|\sum_{j=1}^{C}\mathbf{W}^j \circledast \mathbf{X}^j - \mathbf{Y}\right\|_F^2 + \lambda_1\sum_{j=1}^{C}\left\|\mathbf{W}'^j\right\|_F \\ &+ \lambda_2\sum_{j=1}^{C}\left\|\mathbf{W}^j - \mathbf{W}^j_{\text{t-1}}\right\|_F^2 + \frac{\mu}{2}\sum_{j=1}^{C}\left\|\mathbf{W}^j - \mathbf{W}'^j + \frac{\boldsymbol{\Gamma}^j}{\mu}\right\|_F^2,\end{aligned} \tag{1}$$

where $\mathbf{X}^j \in \mathbb{R}^{N\times N}$ is the jth channel of the input tensor (feature) $\mathcal{X} \in \mathbb{R}^{N\times N\times C}$ and $\mathbf{W}^j \in \mathbb{R}^{N\times N}$ is the jth channel of the filter tensor $\mathcal{W} \in \mathbb{R}^{N\times N\times C}$. We introduce the slack variable, $\mathcal{W}' = \mathcal{W}$. $\mathbf{Y} \in \mathbb{R}^{N\times N}$ denotes the desirable Gaussian-shape response map (Henriques et al., 2015). $\boldsymbol{\Gamma}$ is the Lagrangian multiplier of the same size as $\mathcal{W}$, and μ is the corresponding penalty.

T. Xu
Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford GU2 7XH, UK.
E-mail: tianyang.xu@surrey.ac.uk

Z.H. Feng
Department of Computer Science &
Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford GU2 7XH, UK.
E-mail: z.feng@surrey.ac.uk

J. Kitter
Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford GU2 7XH, UK.
E-mail: j.kittler@surrey.ac.uk

X.J. Wu
School of Artificial Intelligence and Computer Science
Jiangnan University, Wuxi 214122, China.
E-mail: wu_xiaojun@jiangnan.edu.cn

To optimise the objective (Eqn. 1), we use the alternating direction method of multipliers (ADMM) (Boyd et al., 2010) to iteratively solve the following sub-problems:

$$\begin{cases} \mathcal{W} = & \arg\min_{\mathcal{W}} \mathcal{L}\left(\mathcal{W}, \mathcal{W}', \boldsymbol{\Gamma}, \mu\right) \\ \mathcal{W}' = & \arg\min_{\mathcal{W}'} \mathcal{L}\left(\mathcal{W}, \mathcal{W}', \boldsymbol{\Gamma}, \mu\right) \\ \boldsymbol{\Gamma} = & \arg\min_{\boldsymbol{\Gamma}} \mathcal{L}\left(\mathcal{W}, \mathcal{W}', \boldsymbol{\Gamma}, \mu\right) \end{cases}. \tag{2}$$

1.1 Updating $\mathcal{W}$

To optimise $\mathcal{W}$, we solve the following sub-problem:

$$\begin{aligned} \mathcal{W} = \arg\min_{\mathcal{W}} & \left\| \sum_{j=1}^{C} \mathbf{W}^j \circledast \mathbf{X}^j - \mathbf{Y} \right\|_F^2 + \lambda_2 \sum_{j=1}^{C} \left\| \mathbf{W}^j - \mathbf{W}^j_{\text{t-1}} \right\|_F^2 \\ & + \frac{\mu}{2} \sum_{j=1}^{C} \left\| \mathbf{W}^j - \mathbf{W}'^j + \frac{\boldsymbol{\Gamma}^j}{\mu} \right\|_F^2, \end{aligned} \tag{3}$$

To efficiently solve Eqn. 3, we transfer the objective to the frequency domain by employing the circulant structure (Gray, 2006):

$$\begin{aligned} \hat{\mathcal{W}} = \arg\min_{\hat{\mathcal{W}}} & \left\| \sum_{j=1}^{C} \hat{\mathbf{W}}^j \odot \hat{\mathbf{X}}^j - \hat{\mathbf{Y}} \right\|_F^2 + \lambda_2 \sum_{j=1}^{C} \left\| \hat{\mathbf{W}}^j - \hat{\mathbf{W}}^j_{\text{t-1}} \right\|_F^2 \\ & + \frac{\mu}{2} \sum_{j=1}^{C} \left\| \hat{\mathbf{W}}^j - \hat{\mathbf{W}}'^j + \frac{\hat{\boldsymbol{\Gamma}}^j}{\mu} \right\|_F^2. \end{aligned} \tag{4}$$

Note that the symbol $\hat{\cdot}$ stands for the corresponding Fourier representation in the frequency domain (Henriques et al., 2015). We perform vectorisation of the variables involved as follows,

$\mathbf{A} = \left[\text{diag}\left(Vec\left(\hat{\mathbf{X}}^{1\top}\right)\right), \text{diag}\left(Vec\left(\hat{\mathbf{X}}^{2\top}\right)\right), \ldots, \text{diag}\left(Vec\left(\hat{\mathbf{X}}^{C\top}\right)\right)\right] \in \mathbb{C}^{N^2 \times CN^2}$,

$\mathbf{f} = \left[Vec\left(\hat{\mathbf{W}}^{1\top}\right), Vec\left(\hat{\mathbf{W}}^{2\top}\right), \ldots, Vec\left(\hat{\mathbf{W}}^{C\top}\right)\right] \in \mathbb{C}^{CN^2}$,

$\mathbf{r} = \left[Vec\left(\hat{\boldsymbol{\Gamma}}^{1\top}\right), Vec\left(\hat{\boldsymbol{\Gamma}}^{2\top}\right), \ldots, Vec\left(\hat{\boldsymbol{\Gamma}}^{C\top}\right)\right] \in \mathbb{C}^{CN^2}$,

$\mathbf{b} = Vec\left(\hat{\mathbf{Y}}\right) \in \mathbb{C}^{N^2}$

For each $j = 1, 2, \ldots, C$, $Vec\left(\hat{\mathbf{X}}^{j\top}\right) \in \mathbb{C}^{N^2}$, $Vec\left(\hat{\mathbf{W}}^{j\top}\right) \in \mathbb{C}^{N^2}$, and $Vec\left(\hat{\boldsymbol{\Gamma}}^{j\top}\right) \in \mathbb{C}^{N^2}$. $\mathbb{C}$ is the complex domain of the Fourier representations. $Vec(\cdot)$ is the vectorisation operator that flattens the matrix to one dimension. $\text{diag}(\cdot)$ returns a square diagonal matrix with the elements of the vector on its diagonal.

We can rewrite Eqn. (4) as minimising the following function w.r.t. $\mathbf{f}$,

$$\mathcal{H}(\mathbf{f}) = \|\mathbf{A}\mathbf{f} - \mathbf{b}\|^2 + \lambda_2 \|\mathbf{f} - \mathbf{f}_{t-1}\|^2 + \frac{\mu}{2} \|\mathbf{f} - \mathbf{f}' + \frac{\mathbf{r}}{\mu}\|^2. \tag{5}$$

Setting $\partial\mathcal{H}/\partial\mathbf{f} = 0$, we can obtain the closed-form solution,

$$\mathbf{f} = \left(\mathbf{A}^H\mathbf{A} + \lambda_2\mathbf{I} + \frac{\mu}{2}\mathbf{I}\right)^{-1}\left(\mathbf{A}^H\mathbf{b} + \lambda_2\mathbf{f}_{t-1} + \frac{\mu}{2}\mathbf{f}' - \frac{\mathbf{r}}{2}\right). \tag{6}$$

$\cdot^H$ denotes the Hermitian transpose. Note that the term $\mathbf{A}^H\mathbf{A} + \lambda_2\mathbf{I} + \frac{\mu}{2}\mathbf{I}$ is of size $CN^2 \times CN^2$. Hence, it is impossible to calculate its inverse matrix in a real time situation. Revisiting Eqn. (5), we find that it can further be decomposed into the following sub linear system,

$$\left(\mathbf{A}[i]\mathbf{A}[i]^H + \lambda_2\mathbf{I} + \frac{\mu}{2}\mathbf{I}\right)\mathbf{f}[i] = \mathbf{A}[i]\mathbf{b}[i] + \lambda_2\mathbf{f}_{t-1}[i] + \frac{\mu}{2}\mathbf{f}'[i] - \frac{\mathbf{r}[i]}{2}, \tag{7}$$

where $\mathbf{f}[i] \in \mathbb{C}^C$ denotes the elements across all C channels in the i-th spatial unit, $i = 1, 2, \ldots, N^2$. A similar meaning applies to the notation $\mathbf{A}[i] \in \mathbb{C}^C$, $\mathbf{b}[i] \in \mathbb{C}$ and $\mathbf{r}[i] \in \mathbb{C}^C$. Based on the Sherman-Morrison-Woodbury formula, we can obtain the closed-form solution of Eqn. (7) as (Petersen and Pedersen, 2008),

$$\begin{aligned}\mathbf{f}[i] =& \frac{1}{\lambda_2 + \mu}\left(\mathbf{I} - \frac{\mathbf{A}[i]\mathbf{A}[i]^H}{\lambda_2 + \frac{\mu}{2} + \mathbf{A}[i]^H\mathbf{A}[i]}\right) \\ &\times\left(\mathbf{A}[i]\mathbf{b}[i] + \lambda_2\mathbf{f}[i]_{t-1} + \frac{\mu}{2}\mathbf{f}'[i] - \frac{\mathbf{b}[i]}{2}\right)\end{aligned} \tag{8}$$

Transforming the vectorised variables back to Eqn. 4, the closed-form solution of the sub-problem can be obtained as:

$$\hat{\mathbf{w}}[m,n] = \frac{1}{\lambda_2 + \mu}\left(\mathbf{I} - \frac{\hat{\mathbf{x}}[m,n]\,\hat{\mathbf{x}}[m,n]^\top}{\lambda_2 + \frac{\mu}{2} + \hat{\mathbf{x}}[m,n]^\top\,\hat{\mathbf{x}}[m,n]}\right)\mathbf{g} \tag{9}$$

where vector $\hat{\mathbf{w}}[m,n] = \left[\hat{w}^1_{m,n}, \hat{w}^2_{m,n}, \ldots, \hat{w}^C_{m,n}\right] \in \mathbb{C}^C$ denotes the m-th row n-th column units of $\hat{\mathcal{W}}$ through all the C channels, and $\mathbf{g} = \hat{\mathbf{x}}[m,n]\,\hat{y}[m,n] + \frac{\mu}{2}\hat{\mathbf{w}}'[m,n] + \lambda_2\hat{\mathbf{w}}_{\text{t-1}}[m,n] - \frac{\hat{\gamma}[m,n]}{2}$.

1.2 Updating $\mathcal{W}'$

To optimise $\mathcal{W}'$, we minimise the following sub-problem (Yuan and Lin, 2006):

$$\mathcal{W}' = \arg\min_{\mathcal{W}'}\lambda_1\sum_{j=1}^{C}\left\|\mathbf{W}'^j\right\|_F + \frac{\mu}{2}\sum_{j=1}^{C}\left\|\mathbf{W}^j - \mathbf{W}'^j + \frac{\boldsymbol{\Gamma}^j}{\mu}\right\|_F^2. \tag{10}$$

Eqn. 10 can be separated to each channel:

$$\mathbf{W}'^j = \arg\min_{\mathbf{W}'^j}\lambda_1\left\|\mathbf{W}'^j\right\|_F + \frac{\mu}{2}\left\|\mathbf{W}^j - \mathbf{W}'^j + \frac{\boldsymbol{\Gamma}^j}{\mu}\right\|_F^2. \tag{11}$$

Setting the derivative of Eqn. (11) to zero, we obtain $\frac{\lambda_1\mathbf{W}'^j}{\mu\|\mathbf{W}'^j\|_F} + \mathbf{W}'^j = \mathbf{H}^j$, where $\mathbf{H}^j = \mathbf{W}^j + \boldsymbol{\Gamma}^j/\mu$. $\frac{\mathbf{W}'^j}{\|\mathbf{W}'^j\|_F}$ is the unit representation of $\mathbf{W}'^j$. Therefore, $\mathbf{W}'^j$ and $\mathbf{H}^j$ share the same geometric direction, indicating the exchangeability between the unit $\frac{\lambda_1 \mathbf{W}'^j}{\mu\|\mathbf{W}'^j\|_F}$ and $\frac{\lambda_1 \mathbf{H}^j}{\mu\|\mathbf{H}^j\|_F}$. The closed-form optimal solution can directly be derived as Yang et al. (2011):

$$\mathbf{W}'^j = \max\left(0, 1 - \frac{\lambda_1}{\mu\left\|\mathbf{H}^j\right\|_F}\right)\mathbf{H}^j. \tag{12}$$

It is clear that $\mathbf{W}'^j$ tends to shrink to zero by collaboratively integrating the constraints imposed by all the $N \times N$ features among the j-th channel.

1.3 Updating other variables

In each iteration, the penalty μ and the multiplier $\boldsymbol{\Gamma}$ are updated as:

$$\begin{aligned} &\boldsymbol{\Gamma} = \boldsymbol{\Gamma} + \mu\left(\mathcal{W} - \mathcal{W}'\right), \\ &\mu = \min\left(\rho\mu, \mu_{\max}\right), \end{aligned} \tag{13}$$

where ρ controls the strictness of the penalty in each iteration and $\mu_{\max}$ is the maximal penalty value. A parameter K is used to control the maximum number of iterations. As each sub-problem is convex, the convergence of our optimisation is guaranteed (Boyd et al., 2010).

References

Boyd S, Parikh N, Chu E, Peleato B, Eckstein J (2010) Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning 3(1):1–122

Gray RM (2006) Toeplitz and circulant matrices : a review. Foundations and Trends in Communications and Information Theory 2(3):155–239

Henriques JF, Rui C, Martins P, Batista J (2015) High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence 37(3):583–596

Petersen K, Pedersen M (2008) The matrix cookbook. technical university of denmark. Technical Manual

Yang Y, Shen HT, Ma Z, Huang Z, Zhou X (2011) l2, 1-norm regularized discriminative feature selection for unsupervised learning. In: Proceedings-international joint conference on artificial intelligence, pp 1589–1594

Yuan M, Lin Y (2006) Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68(1):49–67