Supplementary material for the paper: Discriminative Correlation Filter with Channel and Spatial Reliability DCF-CSR

Alan Lukežič[1], Tomáš Vojíř[2], Luka Čehovin Zajc[1], Jiří Matas[2] and Matej Kristan[1]
[1]Faculty of Computer and Information Science, University of Ljubljana, Slovenia
[2]Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic
{alan.lukezic, luka.cehovin, matej.kristan}@fri.uni-lj.si
{vojirtom, matas}@cmp.felk.cvut.cz

Abstract

This is the supplementary material for the paper "Discriminative Correlation Filter with Channel and Spatial Reliability" submitted to the CVPR 2017. Due to spatial constraints, parts not crucial for understanding the DCF-CSR tracker formulation, but helpful for gaining insights, were moved here.

1. Derivation of the augmented Lagrangian minimizer

This section provides the complete derivation of the closed-form solutions for the relations (9,10) in the submitted paper [3]. The augmented Lagrangian from Equation (5) in [3] is

$$\mathcal{L}(\hat{\mathbf{h}}_c, \mathbf{h}, \hat{\mathbf{l}}) = \|\hat{\mathbf{h}}_c^H \mathrm{diag}(\hat{\mathbf{f}}) - \hat{\mathbf{g}}\|^2 + \frac{\lambda}{2}\|\mathbf{h}_m\|^2 + [\hat{\mathbf{l}}^H(\hat{\mathbf{h}}_c - \hat{\mathbf{h}}_m) + \overline{\hat{\mathbf{l}}^H(\hat{\mathbf{h}}_c - \hat{\mathbf{h}}_m)}] + \mu\|\hat{\mathbf{h}}_c - \hat{\mathbf{h}}_m\|^2, \tag{1}$$

with $\mathbf{h}_m = (\mathbf{m} \odot \mathbf{h})$. For the purposes of derivation we will rewrite (1) into a fully vectorized form

$$\mathcal{L}(\hat{\mathbf{h}}_c, \mathbf{h}, \hat{\mathbf{l}}) = \|\hat{\mathbf{h}}_c^H \mathrm{diag}(\hat{\mathbf{f}}) - \hat{\mathbf{g}}\|^2 + \frac{\lambda}{2}\|\mathbf{h}_m\|^2 + \left[\hat{\mathbf{l}}^H(\hat{\mathbf{h}}_c - \sqrt{D}\mathbf{F}\mathbf{M}\mathbf{h}) + \overline{\hat{\mathbf{l}}^H(\hat{\mathbf{h}}_c - \sqrt{D}\mathbf{F}\mathbf{M}\mathbf{h})}\right] + \mu\|\hat{\mathbf{h}}_c - \sqrt{D}\mathbf{F}\mathbf{M}\mathbf{h}\|^2, \tag{2}$$

where $\mathbf{F}$ denotes $D \times D$ orthonormal matrix of Fourier coefficients, such that the Fourier transform is defined as $\hat{\mathbf{x}} = \mathcal{F}(\mathbf{x}) = \sqrt{D}\mathbf{F}\mathbf{x}$ and $\mathbf{M} = \mathrm{diag}(\mathbf{m})$. For clearer representation we denote the four terms in the summation (2) as

$$\mathcal{L}(\hat{\mathbf{h}}_c, \mathbf{h}, \hat{\mathbf{l}}) = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 + \mathcal{L}_4, \tag{3}$$

where

$$\mathcal{L}_1 = \left(\hat{\mathbf{h}}_c^H \mathrm{diag}(\hat{\mathbf{f}}) - \hat{\mathbf{g}}\right)\overline{\left(\hat{\mathbf{h}}_c^H \mathrm{diag}(\hat{\mathbf{f}}) - \hat{\mathbf{g}}\right)}^T, \tag{4}$$

$$\mathcal{L}_2 = \frac{\lambda}{2}\|\mathbf{h}_m\|^2, \tag{5}$$

$$\mathcal{L}_3 = \hat{\mathbf{l}}^H(\hat{\mathbf{h}}_c - \sqrt{D}\mathbf{F}\mathbf{M}\mathbf{h}) + \overline{\hat{\mathbf{l}}^H(\hat{\mathbf{h}}_c - \sqrt{D}\mathbf{F}\mathbf{M}\mathbf{h})}, \tag{6}$$

$$\mathcal{L}_4 = \mu\|\hat{\mathbf{h}}_c - \sqrt{D}\mathbf{F}\mathbf{M}\mathbf{h}\|^2. \tag{7}$$

Minimization of (Equation 5 in [3]) is an iterative process at which the following minimizations are required:

$$\hat{\mathbf{h}}_c^{\mathrm{opt}} = \arg\min_{\mathbf{h_c}} \mathcal{L}(\hat{\mathbf{h}}_c, \mathbf{h}, \hat{\mathbf{l}}), \tag{8}$$

$$\mathbf{h}^{\mathrm{opt}} = \arg\min_{\mathbf{h}} \mathcal{L}(\hat{\mathbf{h}}_c^{\mathrm{opt}}, \mathbf{h}, \hat{\mathbf{l}}). \tag{9}$$

Minimization w.r.t. to $\hat{\mathbf{h}}_c$ is derived by finding $\hat{\mathbf{h}}_c$ at which the complex gradient of the augmented Lagrangian vanishes, i.e.,

$$\nabla_{\overline{\hat{\mathbf{h}}_c}} \mathcal{L} \equiv 0, \tag{10}$$

$$\nabla_{\overline{\hat{\mathbf{h}}_c}} \mathcal{L}_1 + \nabla_{\overline{\hat{\mathbf{h}}_c}} \mathcal{L}_2 + \nabla_{\overline{\hat{\mathbf{h}}_c}} \mathcal{L}_3 + \nabla_{\overline{\hat{\mathbf{h}}_c}} \mathcal{L}_4 \equiv 0. \tag{11}$$

The partial complex gradients are:

$$\nabla_{\overline{\hat{\mathbf{h}}_c}} \mathcal{L}_1 = \tag{12}$$
$$= \frac{\partial}{\partial \overline{\hat{\mathbf{h}}_c}}\left[\left(\hat{\mathbf{h}}_c^H \mathrm{diag}(\hat{\mathbf{f}}) - \hat{\mathbf{g}}\right)\overline{\left(\hat{\mathbf{h}}_c^H \mathrm{diag}(\hat{\mathbf{f}}) - \hat{\mathbf{g}}\right)}^T\right] =$$
$$= \frac{\partial}{\partial \overline{\hat{\mathbf{h}}_c}}\left[\hat{\mathbf{h}}_c^H \mathrm{diag}(\hat{\mathbf{f}})\mathrm{diag}(\hat{\mathbf{f}})^H \hat{\mathbf{h}}_c - \hat{\mathbf{h}}_c^H \mathrm{diag}(\hat{\mathbf{f}})\hat{\mathbf{g}}^H - \hat{\mathbf{g}}\mathrm{diag}(\hat{\mathbf{f}})^H \hat{\mathbf{h}}_c + \hat{\mathbf{g}}\hat{\mathbf{g}}^H\right] =$$
$$= \mathrm{diag}(\hat{\mathbf{f}})\mathrm{diag}(\hat{\mathbf{f}})^H \hat{\mathbf{h}}_c - \mathrm{diag}(\hat{\mathbf{f}})\hat{\mathbf{g}}^H,$$

$$\nabla_{\overline{\hat{\mathbf{h}}}_c} \mathcal{L}_2 = 0, \tag{13}$$

$$\begin{aligned} &\nabla_{\overline{\hat{\mathbf{h}}}_c} \mathcal{L}_3 = \qquad (14) \\ &= \frac{\partial}{\partial \overline{\hat{\mathbf{h}}}_c} \left[\hat{\mathbf{l}}^H \left(\hat{\mathbf{h}}_c - \sqrt{D}\mathbf{FMh} \right) + \overline{\hat{\mathbf{l}}^H \left(\hat{\mathbf{h}}_c - \sqrt{D}\mathbf{FMh} \right)} \right] = \\ &= \frac{\partial}{\partial \overline{\hat{\mathbf{h}}}_c} \left[\hat{\mathbf{l}}^H \hat{\mathbf{h}}_c - \hat{\mathbf{l}}^H \sqrt{D}\mathbf{FMh} + \hat{\mathbf{l}}^T \overline{\hat{\mathbf{h}}}_c - \overline{\hat{\mathbf{l}}}^H \sqrt{D}\overline{\mathbf{FMh}} \right] = \\ &= \hat{\mathbf{l}}, \end{aligned}$$

$$\begin{aligned} &\nabla_{\overline{\hat{\mathbf{h}}}_c} \mathcal{L}_4 = \qquad (15) \\ &= \frac{\partial}{\partial \overline{\hat{\mathbf{h}}}_c} \left[\mu \left(\hat{\mathbf{h}}_c - \sqrt{D}\mathbf{FMh} \right) \overline{\left(\hat{\mathbf{h}}_c - \sqrt{D}\mathbf{FMh} \right)}^T \right] = \\ &= \frac{\partial}{\partial \overline{\hat{\mathbf{h}}}_c} \Big[\mu \Big(\hat{\mathbf{h}}_c \hat{\mathbf{h}}_c^H - \hat{\mathbf{h}}_c \sqrt{D} \mathbf{h}^H \mathbf{MF}^H - \\ &\sqrt{D}\mathbf{FMh}\hat{\mathbf{h}}_c^H + D\mathbf{FMhh}^H\mathbf{MF}^H \Big) \Big] = \\ &= \mu \hat{\mathbf{h}}_c - \mu \sqrt{D}\mathbf{FMh}. \end{aligned}$$

Note that $\sqrt{D}\mathbf{FMh} = \hat{\mathbf{h}}_m$ according to our original definition of $\hat{\mathbf{h}}_m$. Plugging (12-15) into (11) yields

$$\text{diag}(\hat{\mathbf{f}})\text{diag}(\hat{\mathbf{f}})^H \hat{\mathbf{h}}_c - \text{diag}(\hat{\mathbf{f}})\hat{\mathbf{g}}^H + \hat{\mathbf{l}}\mu\hat{\mathbf{h}}_c - \mu\hat{\mathbf{h}}_m = 0, \tag{16}$$

$$\hat{\mathbf{h}}_c = \frac{\text{diag}(\hat{\mathbf{f}})\hat{\mathbf{g}}^H + \mu\hat{\mathbf{h}}_m - \hat{\mathbf{l}}}{\text{diag}(\hat{\mathbf{f}})\text{diag}(\hat{\mathbf{f}})^H + \mu},$$

which can be rewritten into

$$\hat{\mathbf{h}}_c = \frac{\hat{\mathbf{f}} \odot \overline{\hat{\mathbf{g}}} + \mu\hat{\mathbf{h}}_m - \hat{\mathbf{l}}}{\hat{\mathbf{f}} \odot \overline{\hat{\mathbf{f}}} + \mu}. \tag{17}$$

Next we derive the closed-form solution of (9). The optimal $\mathbf{h}$ is obtained when the complex gradient w.r.t. $\mathbf{h}$ vanishes, i.e.,

$$\nabla_{\overline{\mathbf{h}}} \mathcal{L} \equiv 0 \tag{18}$$

$$\nabla_{\overline{\mathbf{h}}} \mathcal{L}_1 + \nabla_{\overline{\mathbf{h}}} \mathcal{L}_2 + \nabla_{\overline{\mathbf{h}}} \mathcal{L}_3 + \nabla_{\overline{\mathbf{h}}} \mathcal{L}_4 \equiv 0. \tag{19}$$

The partial gradients are

$$\nabla_{\overline{\mathbf{h}}} \mathcal{L}_1 = 0, \tag{20}$$

$$\begin{aligned} &\nabla_{\overline{\mathbf{h}}} \mathcal{L}_2 = \qquad (21) \\ &= \frac{\partial}{\partial \overline{\mathbf{h}}} \left[\frac{\lambda}{2} \overline{(\mathbf{Mh})}^T (\mathbf{Mh}) \right] = \frac{\partial}{\partial \overline{\mathbf{h}}} \left[\frac{\lambda}{2} \mathbf{h}^H \overline{\mathbf{M}}\mathbf{Mh} \right]. \end{aligned}$$

Since we defined mask $\mathbf{m}$ as real-valued binary mask, the product $\overline{\mathbf{M}}\mathbf{M}$ can be simplified into $\mathbf{M}$ and the result for $\nabla_{\overline{\mathbf{h}}} \mathcal{L}_2$ is

$$\nabla_{\overline{\mathbf{h}}} \mathcal{L}_2 = \frac{\lambda}{2} \mathbf{Mh}. \tag{22}$$

The remaining gradients are as follows:

$$\begin{aligned} &\nabla_{\overline{\mathbf{h}}} \mathcal{L}_3 = \qquad (23) \\ &= \frac{\partial}{\partial \overline{\mathbf{h}}} \left[\hat{\mathbf{l}}^H \left(\hat{\mathbf{h}}_c - \sqrt{D}\mathbf{FMh} \right) + \overline{\hat{\mathbf{l}}^H \left(\hat{\mathbf{h}}_c - \sqrt{D}\mathbf{FMh} \right)} \right] = \\ &= \frac{\partial}{\partial \overline{\mathbf{h}}} \left[\hat{\mathbf{l}}^H \hat{\mathbf{h}}_c - \hat{\mathbf{l}}^H \sqrt{D}\mathbf{FMh} + \hat{\mathbf{l}}^T \overline{\hat{\mathbf{h}}}_c - \overline{\hat{\mathbf{l}}}^H \sqrt{D}\overline{\mathbf{FMh}} \right] = \\ &= -\sqrt{D}\mathbf{MF}^H \hat{\mathbf{l}}, \end{aligned}$$

$$\begin{aligned} &\nabla_{\overline{\mathbf{h}}} \mathcal{L}_4 = \qquad (24) \\ &= \frac{\partial}{\partial \overline{\mathbf{h}}} \left[\mu \overline{\left(\hat{\mathbf{h}}_c - \sqrt{D}\mathbf{FMh} \right)}^T \left(\hat{\mathbf{h}}_c - \sqrt{D}\mathbf{FMh} \right) \right] = \\ &= \frac{\partial}{\partial \overline{\mathbf{h}}} \Big[\mu \Big(\hat{\mathbf{h}}_c^H \hat{\mathbf{h}}_c - \hat{\mathbf{h}}_c^H \sqrt{D}\mathbf{FMh} - \\ &\sqrt{D}\mathbf{h}^H\mathbf{MF}^H\hat{\mathbf{h}}_c + D\mathbf{h}^H\mathbf{Mh} \Big) \Big] = \\ &= -\mu\sqrt{D}\mathbf{MF}^H\hat{\mathbf{h}}_c + \mu D\mathbf{Mh}. \end{aligned}$$

Plugging (20-24) into (19) yields

$$\frac{\lambda}{2}\mathbf{Mh} - \sqrt{D}\mathbf{MF}^H\hat{\mathbf{l}} - \mu\sqrt{D}\mathbf{MF}^H\hat{\mathbf{h}}_c + \mu D\mathbf{Mh} = 0, \tag{25}$$

$$\mathbf{Mh} = \mathbf{M}\frac{\sqrt{D}\mathbf{F}^H(\hat{\mathbf{l}} + \mu\hat{\mathbf{h}}_c)}{\frac{\lambda}{2} + \mu D}.$$

Using the definition of the inverse Fourier transform, i.e., $\mathcal{F}^{-1}(\hat{\mathbf{x}}) = \frac{1}{\sqrt{D}}\mathbf{F}^H\hat{\mathbf{x}}$, (25) can be rewritten into

$$\mathbf{m} \odot \mathbf{h} = \mathbf{m} \odot \frac{\mathcal{F}^{-1}(\hat{\mathbf{l}} + \mu\hat{\mathbf{h}}_c)}{\frac{\lambda}{2D} + \mu}. \tag{26}$$

The values in $\mathbf{m}$ are either zero or one. Elements in $\mathbf{h}$ that correspond to the zeros in $\mathbf{m}$ can in principle not be recovered from (26) since this would result in division by zero. But our initial definition of the problem was to seek solutions for the filter that satisfies the following relation $\mathbf{h} \equiv \mathbf{h} \odot \mathbf{m}$. This means the values corresponding to zeros in $\mathbf{m}$ should be zero in $\mathbf{h}$. Thus the proximal solution to (26) is

$$\mathbf{h} = \mathbf{m} \odot \frac{\mathcal{F}^{-1}(\hat{\mathbf{l}} + \mu\hat{\mathbf{h}}_c)}{\frac{\lambda}{2D} + \mu}. \tag{27}$$

2. Convergence of the derived optimization

The optimization of augmented Lagrangian (Algorithm 1 in the paper) is central to the proposed tracking algorithm. At each tracking iteration, we constrain the optimization to run for four iterations since the major drop in filter error is achieved within the first few steps. For convenience, we visualize here the convergence profile.

Figure 1 shows the average squared difference between result of the correlation of the the filter constrained by the mask function and the ideal output. This graph was obtained by averaging 60 examples of initializing a filter on a target (one per VOT2015 sequence) and scaling each to an interval between zero and one. It is clear that the error drops by 80% within the first few iterations. The reduction from the remaining iterations is due to reduction of the errors in the sidelobes of the filter.

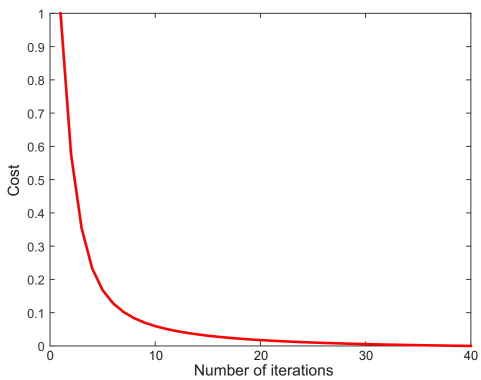

Figure 1: Convergence of our optimization method. The graph is averaged over the initializations in 60 sequences, while for each sequence it is normalized to have maximum cost at 1 and minimum at 0. Due to the different absolute cost values for each optimization, normalization is used to demonstrate only how cost changes during the optimization.

3. Time analysis

The average speed of our tracker measured on the VOT2016 dataset is approximately 13 frames-per-second or 77 milliseconds per-frame. Figure 2 shows the processing time required by each step of the SCR-DCF. A tracking iteration is divided into two steps: (i) target localization and (ii) the visual model update. Target localization takes in average 35 milliseconds at each frame and is composed of two sub-steps: estimation of object translation (23ms) and scale change estimation (12ms). The visual model update step takes on average 42 milliseconds. It consists of three sub-steps: spatial reliability map construction (16ms), filter update (12ms) and scale model update (14ms). Filter optimization, which is part of the filter update step, takes on average 7 milliseconds.

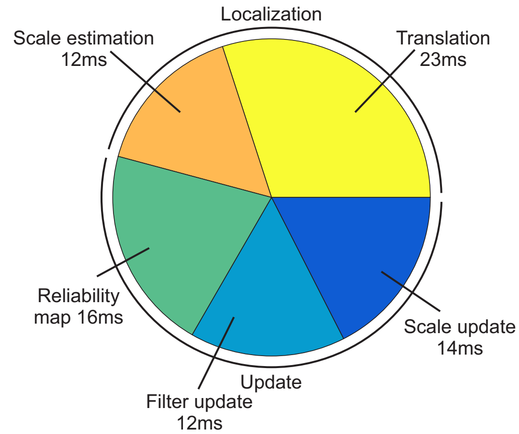

Figure 2: A single iteration processing time decomposed across the main steps of the SCR-DCF.

4. VOT benchmarks

In the paper [3] Figure 7 shows the results of expected average overlap measure for VOT2015 [2] and VOT2016 [1] challenge. For better clarity we showed only top-performing trackers. Full results are presented here. The VOT2015 [2] challenge results with all 61 trackers and the CSR-DCF are shown in Figure 3. Figure 4 shows the results for 70 trackers and CSR-DCF on VOT 2016 [1] challenge.

References

[1] M. Kristan, A. Leonardis, J. Matas, M. Felsberg, R. Pflugfelder, L. Čehovin, T. Vojir, G. Häger, A. Lukežič, and G. et al. Fernandez. The visual object tracking vot2016 challenge results. In *Proc. European Conf. Computer Vision*, 2016. 3, 4

[2] M. Kristan, J. Matas, A. Leonardis, M. Felsberg, L. Čehovin, G. Fernandez, T. Vojir, G. Häger, G. Nebehay, and R. et al. Pflugfelder. The visual object tracking vot2015 challenge results. In *Int. Conf. Computer Vision*, 2015. 3, 4

[3] A. Lukežič, T. Vojíř, L. Čehovin Zajc, J. Matas, and M. Kristan. Discriminative correlation filter with channel and spatial reliability. In *Comp. Vis. Patt. Recognition*, 2017. 1, 3

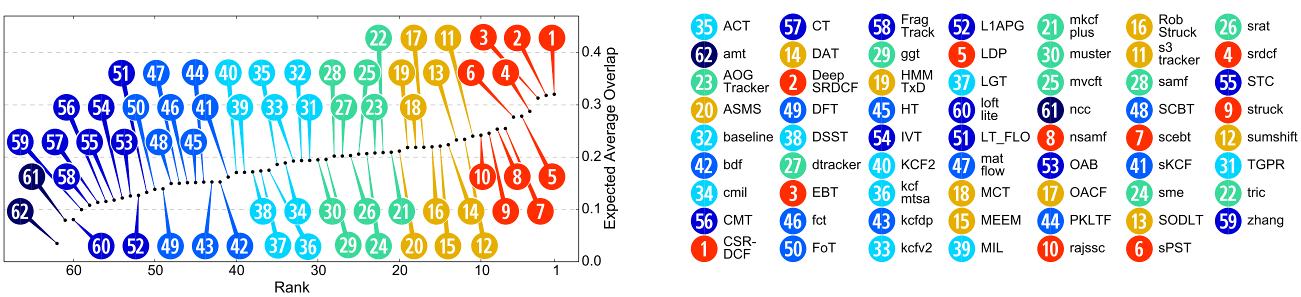

Figure 3: Expected average overlap plot for full VOT2015 [2] (left) benchmark with the proposed CSR-DCF tracker. Legend is shown on the right side.

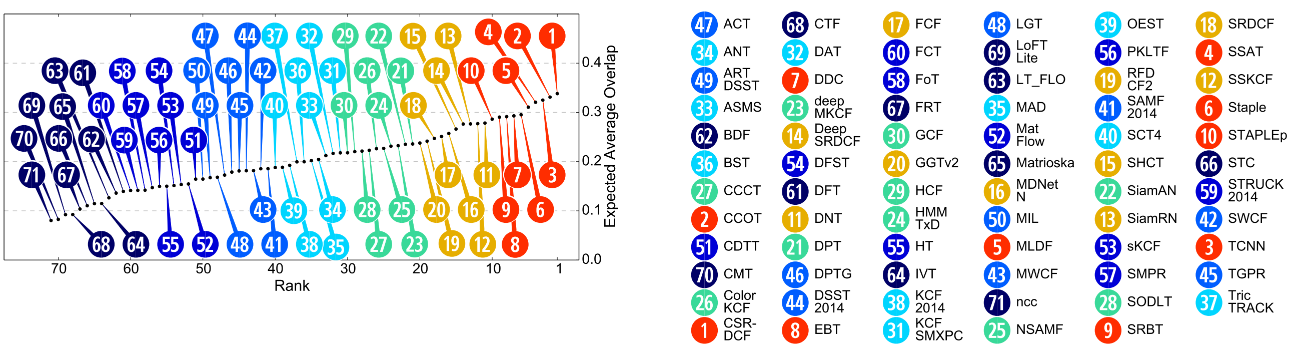

Figure 4: Expected average overlap plot for full VOT2016 [1] (left) benchmark with the proposed CSR-DCF tracker. Legend is shown on the right side.