

Context-Aware Correlation Filter Tracker (Supplementary Material)

Matthias Mueller, Neil Smith, and Bernard Ghanem

King Abdullah University of Science and Technology (KAUST), Saudi Arabia
`{matthias.mueller.2, neil.smith, bernard.ghanem}@kaust.edu.sa`

1 Formulation

1.1 Single-Channel Features

Solution in the Primal Domain.

$$\min_{\mathbf{w}} \quad \|\mathbf{A}_0 \mathbf{w} - \mathbf{y}\|_2^2 + \lambda_1 \|\mathbf{w}\|_2^2 + \lambda_2 \sum_{i=1}^k \|\mathbf{A}_i \mathbf{w}\|_2^2 \quad (1)$$

The primal objective function f_p in Eq. (1) can be rewritten by stacking the context image patches below the target image patch forming a new data matrix $\mathbf{B} \in \mathbb{R}^{(k+1)n \times n}$. The new regression target $\bar{\mathbf{y}} \in \mathbb{R}^{(k+1)n}$ concatenates \mathbf{y} with zeros.

$$f_p(\mathbf{w}, \mathbf{B}) = \left\| \begin{bmatrix} \mathbf{A}_0 \\ \sqrt{\lambda_2} \mathbf{A}_1 \\ \vdots \\ \sqrt{\lambda_2} \mathbf{A}_k \end{bmatrix} \mathbf{w} - \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \right\|_2^2 + \lambda_1 \|\mathbf{w}\|_2^2 = \|\mathbf{B} \mathbf{w} - \bar{\mathbf{y}}\|_2^2 + \lambda_1 \|\mathbf{w}\|_2^2 \quad (2)$$

where $\mathbf{w} \in \mathbb{R}^n$, $\mathbf{B} \in \mathbb{R}^{(k+1)n \times n}$, $\bar{\mathbf{y}} \in \mathbb{R}^{(k+1)n}$

Since $f_p(\mathbf{w}, \mathbf{B})$ is convex, it can be minimized by setting the gradient to zero, yielding:

$$\nabla_{\mathbf{w}} f(\mathbf{w}) = 2\mathbf{B}^T (\mathbf{B} \mathbf{w} - \bar{\mathbf{y}}) + 2\lambda_1 \mathbf{w} = 0 \quad (3)$$

Solving for \mathbf{w} :

$$\mathbf{w} = (\mathbf{B}^T \mathbf{B} + \lambda_1 \mathbf{I})^{-1} \mathbf{B}^T \bar{\mathbf{y}} \quad (4)$$

Identity for circulant matrices (\mathbf{F} is the FFT matrix):

$$\left. \begin{aligned} \mathbf{X} &= \mathbf{F} \text{diag}(\hat{\mathbf{x}}) \mathbf{F}^H \\ \mathbf{X}^T &= \mathbf{F} \text{diag}(\hat{\mathbf{x}}^*) \mathbf{F}^H \end{aligned} \right\} \mathbf{X}^T \mathbf{X} = \mathbf{F} \text{diag}(\hat{\mathbf{x}}^* \odot \hat{\mathbf{x}}) \mathbf{F}^H \quad (5)$$

Therefore:

$$\begin{aligned}
\mathbf{B}^T \mathbf{B} &= \mathbf{A}_0^T \mathbf{A}_0 + \dots + \lambda_2 \mathbf{A}_k^T \mathbf{A}_k = \mathbf{A}_0^T \mathbf{A}_0 + \lambda_2 \sum_{i=1}^k \mathbf{A}_i^T \mathbf{A}_i \\
&= \mathbf{F} \text{diag}(\hat{\mathbf{a}}_0^* \odot \hat{\mathbf{a}}_0) \mathbf{F}^H + \lambda_2 \sum_{i=1}^k \mathbf{F} \text{diag}(\hat{\mathbf{a}}_i^* \odot \hat{\mathbf{a}}_i) \mathbf{F}^H \\
&= \mathbf{F} \text{diag} \left(\hat{\mathbf{a}}_0^* \odot \hat{\mathbf{a}}_0 + \lambda_2 \sum_{i=1}^k \hat{\mathbf{a}}_i^* \odot \hat{\mathbf{a}}_i \right) \mathbf{F}^H
\end{aligned} \tag{6}$$

$$\begin{aligned}
\mathbf{B}^T \bar{\mathbf{y}} &= [\mathbf{A}_0^T \ \sqrt{\lambda_2} \mathbf{A}_1^T \ \dots \ \sqrt{\lambda_2} \mathbf{A}_k^T] \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \\
&= \mathbf{A}_0^T \mathbf{y} = \mathbf{F} \text{diag}(\hat{\mathbf{a}}_0^*) \mathbf{F}^H \mathbf{y} = \mathbf{F} \text{diag}(\hat{\mathbf{a}}_0^* \odot \hat{\mathbf{y}})
\end{aligned} \tag{7}$$

Substituting into Eq. 4:

$$\mathbf{w} = \left[\mathbf{F} \text{diag} \left(\hat{\mathbf{a}}_0^* \odot \hat{\mathbf{a}}_0 + \lambda_1 + \lambda_2 \sum_{i=1}^k \hat{\mathbf{a}}_i^* \odot \hat{\mathbf{a}}_i \right) \mathbf{F}^H \right]^{-1} \mathbf{F} \text{diag}(\hat{\mathbf{a}}_0^* \odot \hat{\mathbf{y}}) \tag{8}$$

$$\mathbf{F}^H \mathbf{w} = \left[\text{diag} \left(\hat{\mathbf{a}}_0^* \odot \hat{\mathbf{a}}_0 + \lambda_1 + \lambda_2 \sum_{i=1}^k \hat{\mathbf{a}}_i^* \odot \hat{\mathbf{a}}_i \right) \right]^{-1} \text{diag}(\hat{\mathbf{a}}_0^* \odot \hat{\mathbf{y}}) \tag{9}$$

$$\hat{\mathbf{w}} = \frac{\hat{\mathbf{a}}_0^* \odot \hat{\mathbf{y}}}{\hat{\mathbf{a}}_0^* \odot \hat{\mathbf{a}}_0 + \lambda_1 + \lambda_2 \sum_{i=1}^k \hat{\mathbf{a}}_i^* \odot \hat{\mathbf{a}}_i} \tag{10}$$

Detection formula. The learned filter \mathbf{w} is convolved with image patch \mathbf{z} (search window) in the next frame, where \mathbf{Z} denotes its circulant matrix. The location of the maximum response is the target location within the search window. The primal detection formula in time and frequency domains is given by:

$$\mathbf{r}_p(\mathbf{w}, \mathbf{Z}) = \mathbf{Z} \mathbf{w} \Leftrightarrow \hat{\mathbf{r}}_p = \hat{\mathbf{z}} \odot \hat{\mathbf{w}} \tag{11}$$

Solution in the Dual Domain. Note that the solution in the primal domain in Eq. (4) has the exact same form as the solution of the standard ridge regression problem [2]. Hence, the solution in the dual domain is given by:

$$\boldsymbol{\alpha} = (\mathbf{B} \mathbf{B}^T + \lambda_1 \mathbf{I})^{-1} \bar{\mathbf{y}} \tag{12}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^{(k+1)n}$

$$\begin{aligned}
\mathbf{B}\mathbf{B}^T &= \begin{bmatrix} \mathbf{A}_0 \\ \sqrt{\lambda_2}\mathbf{A}_1 \\ \vdots \\ \sqrt{\lambda_2}\mathbf{A}_k \end{bmatrix} [\mathbf{A}_0^T \ \sqrt{\lambda_2}\mathbf{A}_1^T \ \dots \ \sqrt{\lambda_2}\mathbf{A}_k^T] \\
&= \begin{bmatrix} \mathbf{A}_0\mathbf{A}_0^T & \sqrt{\lambda_2}\mathbf{A}_0\mathbf{A}_1^T & \dots & \sqrt{\lambda_2}\mathbf{A}_0\mathbf{A}_k^T \\ \sqrt{\lambda_2}\mathbf{A}_1\mathbf{A}_0^T & \lambda_2\mathbf{A}_1\mathbf{A}_1^T & \dots & \lambda_2\mathbf{A}_1\mathbf{A}_k^T \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\lambda_2}\mathbf{A}_k\mathbf{A}_0^T & \lambda_2\mathbf{A}_k\mathbf{A}_1^T & \dots & \lambda_2\mathbf{A}_k\mathbf{A}_k^T \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{F} \dots \mathbf{0} \\ \vdots \ddots \vdots \\ \mathbf{0} \dots \mathbf{F} \end{bmatrix} \begin{bmatrix} \text{diag}(\hat{\mathbf{a}}_0 \odot \hat{\mathbf{a}}_0^*) & \dots & \text{diag}(\sqrt{\lambda_2} \hat{\mathbf{a}}_0 \odot \hat{\mathbf{a}}_k^*) \\ \vdots & \ddots & \vdots \\ \text{diag}(\sqrt{\lambda_2} \hat{\mathbf{a}}_k \odot \hat{\mathbf{a}}_0^*) & \dots & \text{diag}(\lambda_2 \hat{\mathbf{a}}_k \odot \hat{\mathbf{a}}_k^*) \end{bmatrix} \begin{bmatrix} \mathbf{F}^H \dots \mathbf{0} \\ \vdots \ddots \vdots \\ \mathbf{0} \dots \mathbf{F}^H \end{bmatrix} \\
&= \bar{\mathbf{F}}\mathbf{D}\bar{\mathbf{F}}^H
\end{aligned} \tag{13}$$

Substituting into Eq. 12:

$$\boldsymbol{\alpha} = [\bar{\mathbf{F}} (\mathbf{D} + \lambda_1 \mathbf{I}) \bar{\mathbf{F}}^H]^{-1} \bar{\mathbf{y}} = \bar{\mathbf{F}} (\mathbf{D} + \lambda_1 \mathbf{I})^{-1} \bar{\mathbf{F}}^H \bar{\mathbf{y}} \tag{14}$$

$$\hat{\boldsymbol{\alpha}} = (\mathbf{D} + \lambda_1 \mathbf{I})^{-1} \hat{\mathbf{y}} = \begin{bmatrix} \text{diag}(\mathbf{d}_{00}) \dots \text{diag}(\mathbf{d}_{0k}) \\ \vdots \quad \ddots \quad \vdots \\ \text{diag}(\mathbf{d}_{k0}) \dots \text{diag}(\mathbf{d}_{kk}) \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{y}} \\ \vdots \\ \mathbf{0} \end{bmatrix} \tag{15}$$

where vectors \mathbf{d}_{jl} with $j, l \in \{1, \dots, k\}$ are given by:

$$\begin{cases} \mathbf{d}_{00} = \hat{\mathbf{a}}_0 \odot \hat{\mathbf{a}}_0^* + \lambda_1 \\ \mathbf{d}_{jj} = \lambda_2 (\hat{\mathbf{a}}_j \odot \hat{\mathbf{a}}_j^*) + \lambda_1, \ j \neq 0 \\ \mathbf{d}_{jl} = \sqrt{\lambda_2} (\hat{\mathbf{a}}_j \odot \hat{\mathbf{a}}_l^*), \ j \neq l \end{cases} \tag{16}$$

Note that the kernel trick can be applied, since all interactions between the image patches occur as bi-products. Hence, the linear correlation can simply be replaced by one of the kernel correlations as derived for conventional kernelized CF trackers [1].

Since all blocks are diagonal, the system can be decomposed into n smaller systems of dimension $\mathbb{R}^{(k+1) \times (k+1)}$. This significantly reduces complexity and allows for parallelization. Instead of solving one large system of dimension $\mathbb{R}^{(k+1)n \times (k+1)n}$ to compute $\boldsymbol{\alpha}$, a separate system is solved for each pixel $p \in \{1, \dots, n\}$ of $\boldsymbol{\alpha}$, as follows:

$$\hat{\boldsymbol{\alpha}}(p) = \begin{bmatrix} \mathbf{d}_{00}(p) \dots \mathbf{d}_{0k}(p) \\ \vdots \quad \ddots \quad \vdots \\ \mathbf{d}_{k0}(p) \dots \mathbf{d}_{kk}(p) \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{y}}(p) \\ \vdots \\ 0 \end{bmatrix} \tag{17}$$

Detection Formula. Note that \mathbf{B} contains the context patches in addition to the target. Consequently, $\boldsymbol{\alpha} \in \mathbb{R}^{(k+1)n}$ is composed of a concatenation of dual variables $\{\boldsymbol{\alpha}_0, \dots, \boldsymbol{\alpha}_k\}$.

$$\mathbf{r}_d(\boldsymbol{\alpha}, \mathbf{B}, \mathbf{Z}) = \mathbf{Z} \mathbf{B}^T \boldsymbol{\alpha} \quad (18)$$

$$\mathbf{r}_d = \mathbf{Z} \begin{bmatrix} \mathbf{A}_0^T & \sqrt{\lambda_2} \mathbf{A}_1^T & \dots & \sqrt{\lambda_2} \mathbf{A}_k^T \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_0 \\ \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_k \end{bmatrix} \quad (19)$$

$$\mathbf{r}_d = \mathbf{F} \begin{bmatrix} \text{diag}(\hat{\mathbf{z}} \odot \hat{\mathbf{a}}_0^*) & \dots & \sqrt{\lambda_2} \text{diag}(\hat{\mathbf{z}} \odot \hat{\mathbf{a}}_k^*) \end{bmatrix} \mathbf{F}^H \begin{bmatrix} \boldsymbol{\alpha}_0 \\ \vdots \\ \boldsymbol{\alpha}_k \end{bmatrix} \quad (20)$$

$$\hat{\mathbf{r}}_d = \begin{bmatrix} \text{diag}(\hat{\mathbf{z}} \odot \hat{\mathbf{a}}_0^*) & \dots & \sqrt{\lambda_2} \text{diag}(\hat{\mathbf{z}} \odot \hat{\mathbf{a}}_k^*) \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\alpha}}_0 \\ \vdots \\ \hat{\boldsymbol{\alpha}}_k \end{bmatrix} \quad (21)$$

$$\hat{\mathbf{r}}_d = \hat{\mathbf{z}} \odot \hat{\mathbf{a}}_0^* \odot \hat{\boldsymbol{\alpha}}_0 + \sqrt{\lambda_2} \sum_{i=1}^k \hat{\mathbf{z}} \odot \hat{\mathbf{a}}_i^* \odot \hat{\boldsymbol{\alpha}}_i \quad (22)$$

1.2 Multi-Channel Features

Solution in the Primal Domain. Now we want to solve the same problem for multi-channel features and effectively learn a joint filter for all feature dimension.

$$\min_{\mathbf{w}_1, \dots, \mathbf{w}_m} \left\| \sum_{i=1}^m \mathbf{A}_{0i} \mathbf{w}_i - \mathbf{y} \right\|_2^2 + \lambda_1 \sum_{i=1}^m \|\mathbf{w}_i\|_2^2 + \lambda_2 \sum_{j=1}^k \left\| \sum_{i=1}^m \mathbf{A}_{ji} \mathbf{w}_i \right\|_2^2 \quad (23)$$

Note that:

$$\left\| \sum_{i=1}^m \mathbf{A}_{0i} \mathbf{w}_i - \mathbf{y} \right\|_2^2 = \left\| \begin{bmatrix} \mathbf{A}_{01} & \dots & \mathbf{A}_{0m} \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_m \end{bmatrix} - \mathbf{y} \right\|_2^2 \quad (24)$$

$$\sum_{i=1}^m \|\mathbf{w}_i\|_2^2 = \left\| \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_m \end{bmatrix} \right\|_2^2 = \|\bar{\mathbf{w}}\|_2^2 \quad (25)$$

$$\sum_{j=1}^k \left\| \sum_{i=1}^m \mathbf{A}_{ji} \mathbf{w}_i \right\|_2^2 = \left\| \begin{bmatrix} \mathbf{A}_{01} & \dots & \mathbf{A}_{0m} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{k1} & \dots & \mathbf{A}_{km} \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_m \end{bmatrix} \right\|_2^2 \quad (26)$$

Therefore, the objective function $f_p(\bar{\mathbf{w}}; \bar{\mathbf{B}})$ in Eq. (23) can be rewritten in a similar fashion as in the case of single-channel features (Eq. (2)) with the difference that $\bar{\mathbf{B}} \in \mathbb{R}^{(k+1)n \times nm}$ contains the base and context image patches as rows and their corresponding features columns. The filters for the different feature dimensions are stacked into $\bar{\mathbf{w}} \in \mathbb{R}^{nm}$.

$$f_p(\bar{\mathbf{w}}; \bar{\mathbf{B}}) = \left\| \begin{bmatrix} \mathbf{A}_{01} & \dots & \mathbf{A}_{0m} \\ \sqrt{\lambda_2} \mathbf{A}_{11} & \dots & \sqrt{\lambda_2} \mathbf{A}_{1m} \\ \vdots & \ddots & \vdots \\ \sqrt{\lambda_2} \mathbf{A}_{k1} & \dots & \sqrt{\lambda_2} \mathbf{A}_{km} \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_m \end{bmatrix} - \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \right\|_2^2 + \lambda_1 \left\| \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_m \end{bmatrix} \right\|_2^2 \quad (27)$$

$$= \|\bar{\mathbf{B}}\bar{\mathbf{w}} - \bar{\mathbf{y}}\|_2^2 + \lambda_1 \|\bar{\mathbf{w}}\|_2^2$$

where $\bar{\mathbf{w}} \in \mathbb{R}^{nm}$, $\bar{\mathbf{B}} \in \mathbb{R}^{(k+1)n \times nm}$, $\bar{\mathbf{y}} \in \mathbb{R}^{(k+1)n}$

Note that the objective function $f_p(\bar{\mathbf{w}}; \bar{\mathbf{B}})$ in Eq. (27) is convex. Hence, this optimization problem can be solved by setting the gradient to zero:

$$\nabla_{\bar{\mathbf{w}}} f(\bar{\mathbf{w}}) = 2\bar{\mathbf{B}}^T(\bar{\mathbf{B}}\bar{\mathbf{w}} - \bar{\mathbf{y}}) + 2\lambda_1 \bar{\mathbf{w}} = 0 \quad (28)$$

Solving for $\bar{\mathbf{w}}$:

$$\bar{\mathbf{w}} = (\bar{\mathbf{B}}^T \bar{\mathbf{B}} + \lambda_1 \mathbf{I})^{-1} \bar{\mathbf{B}}^T \bar{\mathbf{y}} \quad (29)$$

Applying the identity for circulant matrices (Eq. (5)) yields:

$$\bar{\mathbf{B}}^T \bar{\mathbf{y}} = \begin{bmatrix} \mathbf{A}_{01}^T & \sqrt{\lambda_2} \mathbf{A}_{11}^T & \dots & \sqrt{\lambda_2} \mathbf{A}_{k1}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{0m}^T & \sqrt{\lambda_2} \mathbf{A}_{1m}^T & \dots & \sqrt{\lambda_2} \mathbf{A}_{km}^T \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \quad (30)$$

$$= \begin{bmatrix} \mathbf{A}_{01}^T \mathbf{y} \\ \vdots \\ \mathbf{A}_{0m}^T \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{F} \text{diag}(\hat{\mathbf{a}}_{01}^*) & \mathbf{F}^H \mathbf{y} \\ \vdots & \\ \mathbf{F} \text{diag}(\hat{\mathbf{a}}_{0m}^*) & \mathbf{F}^H \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{F} \text{diag}(\hat{\mathbf{a}}_{01}^* \odot \hat{\mathbf{y}}) \\ \vdots \\ \mathbf{F} \text{diag}(\hat{\mathbf{a}}_{0m}^* \odot \hat{\mathbf{y}}) \end{bmatrix}$$

$$\begin{aligned} \bar{\mathbf{B}}^T \bar{\mathbf{B}} &= \begin{bmatrix} \mathbf{A}_{01}^T & \sqrt{\lambda_2} \mathbf{A}_{11}^T & \dots & \sqrt{\lambda_2} \mathbf{A}_{k1}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{0m}^T & \sqrt{\lambda_2} \mathbf{A}_{1m}^T & \dots & \sqrt{\lambda_2} \mathbf{A}_{km}^T \end{bmatrix} \begin{bmatrix} \mathbf{A}_{01} & \dots & \mathbf{A}_{0m} \\ \sqrt{\lambda_2} \mathbf{A}_{11} & \dots & \sqrt{\lambda_2} \mathbf{A}_{1m} \\ \vdots & \ddots & \vdots \\ \sqrt{\lambda_2} \mathbf{A}_{k1} & \dots & \sqrt{\lambda_2} \mathbf{A}_{km} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}_{01}^T \mathbf{A}_{01} + \lambda_2 \sum_{i=1}^k \mathbf{A}_{i1}^T \mathbf{A}_{i1} & \dots & \mathbf{A}_{01}^T \mathbf{A}_{0m} + \lambda_2 \sum_{i=1}^k \mathbf{A}_{i1}^T \mathbf{A}_{im} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{0m}^T \mathbf{A}_{01} + \lambda_2 \sum_{i=1}^k \mathbf{A}_{im}^T \mathbf{A}_{i1} & \dots & \mathbf{A}_{0m}^T \mathbf{A}_{0m} + \lambda_2 \sum_{i=1}^k \mathbf{A}_{im}^T \mathbf{A}_{im} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{F} \dots \mathbf{0} \\ \vdots \ddots \vdots \\ \mathbf{0} \dots \mathbf{F} \end{bmatrix} \begin{bmatrix} \text{diag}(\bar{\mathbf{c}}_{11}) & \dots & \text{diag}(\bar{\mathbf{c}}_{1m}) \\ \vdots & \ddots & \vdots \\ \text{diag}(\bar{\mathbf{c}}_{m1}) & \dots & \text{diag}(\bar{\mathbf{c}}_{mm}) \end{bmatrix} \begin{bmatrix} \mathbf{F}^H \dots \mathbf{0} \\ \vdots \ddots \vdots \\ \mathbf{0} \dots \mathbf{F}^H \end{bmatrix} = \bar{\mathbf{F}} \bar{\mathbf{C}} \bar{\mathbf{F}}^H \end{aligned} \quad (31)$$

where:

$$\begin{aligned}
\bar{\mathbf{c}}_{11} &= \hat{\mathbf{a}}_{01}^* \odot \hat{\mathbf{a}}_{01} + \lambda_2 \sum_{i=1}^k \hat{\mathbf{a}}_{i1}^* \odot \hat{\mathbf{a}}_{i1} \\
\bar{\mathbf{c}}_{1m} &= \hat{\mathbf{a}}_{01}^* \odot \hat{\mathbf{a}}_{0m} + \lambda_2 \sum_{i=1}^k \hat{\mathbf{a}}_{i1}^* \odot \hat{\mathbf{a}}_{im} \\
\bar{\mathbf{c}}_{m1} &= \hat{\mathbf{a}}_{0m}^* \odot \hat{\mathbf{a}}_{01} + \lambda_2 \sum_{i=1}^k \hat{\mathbf{a}}_{im}^* \odot \hat{\mathbf{a}}_{i1} \\
\bar{\mathbf{c}}_{mm} &= \hat{\mathbf{a}}_{0m}^* \odot \hat{\mathbf{a}}_{0m} + \lambda_2 \sum_{i=1}^k \hat{\mathbf{a}}_{im}^* \odot \hat{\mathbf{a}}_{im}
\end{aligned} \tag{32}$$

Substituting into Eq. 29:

$$\bar{\mathbf{w}} = [\bar{\mathbf{F}} (\bar{\mathbf{C}} + \lambda_1 \mathbf{I}) \bar{\mathbf{F}}^H]^{-1} \bar{\mathbf{B}}^T \bar{\mathbf{y}} \tag{33}$$

$$\bar{\mathbf{F}}^H \bar{\mathbf{w}} = (\bar{\mathbf{C}} + \lambda_1 \mathbf{I})^{-1} \bar{\mathbf{F}}^H \bar{\mathbf{B}}^T \bar{\mathbf{y}} \tag{34}$$

$$\begin{aligned}
\hat{\mathbf{w}} &= \begin{bmatrix} \text{diag}(\bar{\mathbf{c}}_{11}) + \lambda_1 \mathbf{I} & \dots & \text{diag}(\bar{\mathbf{c}}_{1m}) \\ \vdots & \ddots & \vdots \\ \text{diag}(\bar{\mathbf{c}}_{m1}) & \dots & \text{diag}(\bar{\mathbf{c}}_{mm}) + \lambda_1 \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \text{diag}(\hat{\mathbf{a}}_{01}^* \odot \hat{\mathbf{y}}) \\ \vdots \\ \text{diag}(\hat{\mathbf{a}}_{0m}^* \odot \hat{\mathbf{y}}) \end{bmatrix} \\
&= \begin{bmatrix} \bar{\mathbf{C}}_{11} & \dots & \bar{\mathbf{C}}_{1m} \\ \vdots & \ddots & \vdots \\ \bar{\mathbf{C}}_{m1} & \dots & \bar{\mathbf{C}}_{mm} \end{bmatrix}^{-1} \begin{bmatrix} \text{diag}(\hat{\mathbf{a}}_{01}^* \odot \hat{\mathbf{y}}) \\ \vdots \\ \text{diag}(\hat{\mathbf{a}}_{0m}^* \odot \hat{\mathbf{y}}) \end{bmatrix}
\end{aligned} \tag{35}$$

The target and context image patches for each feature dimension $j, l \in \{1, \dots, m\}$ are denoted by \mathbf{a}_{0j} and \mathbf{a}_{ij} respectively. The blocks of $(\bar{\mathbf{C}} + \lambda_1 \mathbf{I})^{-1}$ are defined as:

$$\begin{cases} \bar{\mathbf{C}}_{jj} &= \text{diag} \left(\hat{\mathbf{a}}_{0j}^* \odot \hat{\mathbf{a}}_{0j} + \lambda_2 \sum_{i=1}^k \hat{\mathbf{a}}_{ij}^* \odot \hat{\mathbf{a}}_{ij} \right) + \lambda_1 \mathbf{I} \\ \bar{\mathbf{C}}_{jl} &= \text{diag} \left(\hat{\mathbf{a}}_{0j}^* \odot \hat{\mathbf{a}}_{0l} + \lambda_2 \sum_{i=1}^k \hat{\mathbf{a}}_{ij}^* \odot \hat{\mathbf{a}}_{il} \right), \quad j \neq l \end{cases} \tag{36}$$

Unfortunately, this system cannot be inverted as efficiently as in the single-channel case. However, since all of the blocks are diagonal, the system can be decomposed into n smaller systems of dimension $\mathbb{R}^{m \times m}$. This reduces the complexity significantly and allows for parallelization. Similar to Eq. (17), a separate system is solved for each pixel $p \in \{1, \dots, n\}$ of the filter $\hat{\mathbf{w}}$.

$$\hat{\mathbf{w}}(p) = \begin{bmatrix} \bar{\mathbf{c}}_{11}(p) + \lambda_1 & \dots & \bar{\mathbf{c}}_{1m}(p) \\ \vdots & \ddots & \vdots \\ \bar{\mathbf{c}}_{m1}(p) & \dots & \bar{\mathbf{c}}_{mm}(p) + \lambda_1 \end{bmatrix}^{-1} \begin{bmatrix} [\hat{\mathbf{a}}_{01}^* \odot \hat{\mathbf{y}}](1) \\ \vdots \\ [\hat{\mathbf{a}}_{0m}^* \odot \hat{\mathbf{y}}](1) \end{bmatrix} \tag{37}$$

Detection formula. It is almost the same as in the single-channel case in Eq. (11) with the difference that the image patch \mathbf{z} and the learned filter \mathbf{w} are m -dimensional.

Solution in the Dual Domain. Just like in the case of single-channel features, the multi-channel primal solution (Eq. (29)) also has the exact same form as the solution of the standard ridge regression problem [1] yielding the following solution in the dual domain:

$$\bar{\alpha} = (\bar{\mathbf{B}}\bar{\mathbf{B}}^T + \lambda_1 \mathbf{I})^{-1} \bar{\mathbf{y}} \quad (38)$$

where $\bar{\alpha} \in \mathbb{R}^{kn}$

$$\begin{aligned} \bar{\mathbf{B}}\bar{\mathbf{B}}^T &= \begin{bmatrix} \mathbf{A}_{01} & \dots & \mathbf{A}_{0m} \\ \sqrt{\lambda_2} \mathbf{A}_{11} & \dots & \sqrt{\lambda_2} \mathbf{A}_{1m} \\ \vdots & \ddots & \vdots \\ \sqrt{\lambda_2} \mathbf{A}_{k1} & \dots & \sqrt{\lambda_2} \mathbf{A}_{km} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{01}^T & \sqrt{\lambda_2} \mathbf{A}_{11}^T & \dots & \sqrt{\lambda_2} \mathbf{A}_{k1}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{0m}^T & \sqrt{\lambda_2} \mathbf{A}_{1m}^T & \dots & \sqrt{\lambda_2} \mathbf{A}_{km}^T \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^m \mathbf{A}_{0i} \mathbf{A}_{0i}^T & \sqrt{\lambda_2} \sum_{i=1}^m \mathbf{A}_{0i} \mathbf{A}_{1i}^T & \dots & \sqrt{\lambda_2} \sum_{i=1}^m \mathbf{A}_{0i} \mathbf{A}_{ki}^T \\ \sqrt{\lambda_2} \sum_{i=1}^m \mathbf{A}_{1i} \mathbf{A}_{0i}^T & \lambda_2 \sum_{i=1}^m \mathbf{A}_{1i} \mathbf{A}_{1i}^T & \dots & \lambda_2 \sum_{i=1}^m \mathbf{A}_{1i} \mathbf{A}_{ki}^T \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\lambda_2} \sum_{i=1}^m \mathbf{A}_{ki} \mathbf{A}_{0i}^T & \lambda_2 \sum_{i=1}^m \mathbf{A}_{ki} \mathbf{A}_{1i}^T & \dots & \lambda_2 \sum_{i=1}^m \mathbf{A}_{ki} \mathbf{A}_{ki}^T \end{bmatrix} \\ &= \bar{\mathbf{F}} \begin{bmatrix} \text{diag}(\sum_{i=1}^m \hat{\mathbf{a}}_{0i} \odot \hat{\mathbf{a}}_{0i}^*) & \dots & \text{diag}(\sqrt{\lambda_2} \sum_{i=1}^m \hat{\mathbf{a}}_{0i} \odot \hat{\mathbf{a}}_{ki}^*) \\ \vdots & \ddots & \vdots \\ \text{diag}(\sqrt{\lambda_2} \sum_{i=1}^m \hat{\mathbf{a}}_{ki} \odot \hat{\mathbf{a}}_{0i}^*) & \dots & \text{diag}(\lambda_2 \sum_{i=1}^m \hat{\mathbf{a}}_{ki} \odot \hat{\mathbf{a}}_{ki}^*) \end{bmatrix} \bar{\mathbf{F}}^H \\ &= \bar{\mathbf{F}} \bar{\mathbf{D}} \bar{\mathbf{F}}^H \end{aligned} \quad (39)$$

Substituting into Eq. 38:

$$\bar{\alpha} = [\bar{\mathbf{F}} (\bar{\mathbf{D}} + \lambda_1 \mathbf{I}) \bar{\mathbf{F}}^H]^{-1} \bar{\mathbf{y}} = \bar{\mathbf{F}} (\bar{\mathbf{D}} + \lambda_1 \mathbf{I})^{-1} \bar{\mathbf{F}}^H \bar{\mathbf{y}} \quad (40)$$

$$\hat{\alpha} = (\bar{\mathbf{D}} + \lambda_1 \mathbf{I})^{-1} \hat{\mathbf{y}} = \begin{bmatrix} \text{diag}(\bar{\mathbf{d}}_{00}) & \dots & \text{diag}(\bar{\mathbf{d}}_{0k}) \\ \vdots & \ddots & \vdots \\ \text{diag}(\bar{\mathbf{d}}_{k0}) & \dots & \text{diag}(\bar{\mathbf{d}}_{kk}) \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{y}} \\ \vdots \\ \mathbf{0} \end{bmatrix} \quad (41)$$

where vectors $\bar{\mathbf{d}}_{jl}$ with $j, l \in \{1, \dots, k\}$ are given by:

$$\begin{cases} \bar{\mathbf{d}}_{00} &= \sum_{i=1}^m (\hat{\mathbf{a}}_{0i} \odot \hat{\mathbf{a}}_{0i}^*) + \lambda_1 \\ \bar{\mathbf{d}}_{jj} &= \lambda_2 \sum_{i=1}^m (\hat{\mathbf{a}}_{ji} \odot \hat{\mathbf{a}}_{ji}^*) + \lambda_1, \quad j \neq 0 \\ \bar{\mathbf{d}}_{jl} &= \sqrt{\lambda_2} \sum_{i=1}^m (\hat{\mathbf{a}}_{ji} \odot \hat{\mathbf{a}}_{li}^*), \quad j \neq l \end{cases} \quad (42)$$

Note that the linear system is the same as in case of the dual domain solution for single-channel features with the exception that there is now a sum along the feature dimension m . This solution also permits the use of kernels and the linear system can be solved in the same fashion as the single-channel case (Eq. (17)). Since all blocks are diagonal, the system can be decomposed into n smaller systems of dimension $\mathbb{R}^{(k+1) \times (k+1)}$. This significantly reduces complexity and allows for

parallelization. Instead of solving one large system of dimension $\mathbb{R}^{(k+1)n \times (k+1)n}$ to compute $\hat{\bar{\alpha}}$, a separate system is solved for each pixel $p \in \{1, \dots, n\}$ of $\hat{\bar{\alpha}}$, as follows:

$$\hat{\bar{\alpha}}(p) = \begin{bmatrix} \bar{\mathbf{d}}_{00}(p) & \dots & \bar{\mathbf{d}}_{0k}(p) \\ \vdots & \ddots & \vdots \\ \bar{\mathbf{d}}_{k0}(p) & \dots & \bar{\mathbf{d}}_{kk}(p) \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{y}}(p) \\ \vdots \\ 0 \end{bmatrix} \quad (43)$$

Detection Formula. It follows the single-channel feature case with the difference that $\bar{\mathbf{Z}} \in \mathbb{R}^{nm \times n}$ and $\bar{\mathbf{B}} \in \mathbb{R}^{(k+1)n \times nm}$ now have multiple feature dimensions as columns:

$$\mathbf{r}_d(\bar{\alpha}, \bar{\mathbf{B}}, \bar{\mathbf{Z}}) = \bar{\mathbf{Z}} \bar{\mathbf{B}}^T \bar{\alpha} \quad (44)$$

$$\mathbf{r}_d = [\mathbf{Z}_1 \dots \mathbf{Z}_m] \begin{bmatrix} \mathbf{A}_{01}^T & \sqrt{\lambda_2} \mathbf{A}_{11}^T & \dots & \sqrt{\lambda_2} \mathbf{A}_{k1}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{0m}^T & \sqrt{\lambda_2} \mathbf{A}_{1m}^T & \dots & \sqrt{\lambda_2} \mathbf{A}_{km}^T \end{bmatrix} \begin{bmatrix} \bar{\alpha}_0 \\ \bar{\alpha}_1 \\ \vdots \\ \bar{\alpha}_k \end{bmatrix} \quad (45)$$

$$\mathbf{r}_d = [\sum_{i=1}^m \mathbf{Z}_i \mathbf{A}_{0i}^T \dots \sqrt{\lambda_2} \sum_{i=1}^m \mathbf{Z}_i \mathbf{A}_{ki}^T] \mathbf{F}^H \begin{bmatrix} \bar{\alpha}_0 \\ \vdots \\ \bar{\alpha}_k \end{bmatrix} \quad (46)$$

$$\mathbf{r}_d = \mathbf{F} [\sum_{i=1}^m \text{diag}(\hat{\mathbf{z}}_i \odot \hat{\mathbf{a}}_{0i}^*) \dots \sqrt{\lambda_2} \sum_{i=1}^m \text{diag}(\hat{\mathbf{z}}_i \odot \hat{\mathbf{a}}_{ki}^*)] \mathbf{F}^H \begin{bmatrix} \bar{\alpha}_0 \\ \vdots \\ \bar{\alpha}_k \end{bmatrix} \quad (47)$$

$$\hat{\mathbf{r}}_d = [\text{diag}(\sum_{i=1}^m \hat{\mathbf{z}}_i \odot \hat{\mathbf{a}}_{0i}^*) \dots \sqrt{\lambda_2} \text{diag}(\sum_{i=1}^m \hat{\mathbf{z}}_i \odot \hat{\mathbf{a}}_{ki}^*)] \begin{bmatrix} \hat{\alpha}_0 \\ \vdots \\ \hat{\alpha}_k \end{bmatrix} \quad (48)$$

$$\hat{\mathbf{r}}_d = \left(\sum_{i=1}^m \hat{\mathbf{z}}_i \odot \hat{\mathbf{a}}_{0i}^* \right) \odot \hat{\alpha}_0 + \sqrt{\lambda_2} \sum_{j=1}^k \left(\sum_{i=1}^m \hat{\mathbf{z}}_i \odot \hat{\mathbf{a}}_{ji}^* \right) \odot \hat{\alpha}_j \quad (49)$$

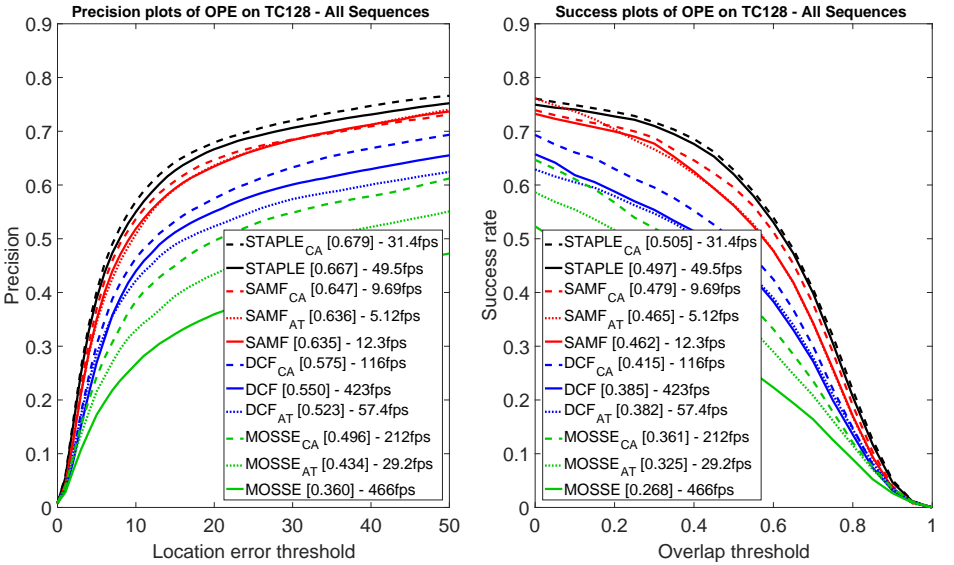
1.3 Energy of data term

$$\begin{aligned} \|\mathbf{A}_0 \mathbf{w} - \mathbf{y}\|_2^2 &= (\mathbf{A}_0 \mathbf{w} - \mathbf{y})^T (\mathbf{A}_0 \mathbf{w} - \mathbf{y}) \\ &= (\mathbf{A}_0 \mathbf{w})^T (\mathbf{A}_0 \mathbf{w}) - (\mathbf{A}_0 \mathbf{w})^T \mathbf{y} - \mathbf{y}^T (\mathbf{A}_0 \mathbf{w}) + (\mathbf{y})^T (\mathbf{y}) \\ &= \|\mathbf{A}_0 \mathbf{w}\|_2^2 - 2(\mathbf{A}_0 \mathbf{w})^T \mathbf{y} + \|\mathbf{y}\|_2^2 \end{aligned} \quad (50)$$

where $\mathbf{A}_0 \mathbf{w}$ is the response

2 Experiments

We also evaluate trackers with the same parameters on two additional datasets to show that our framework improves tracking performance consistently. Figures 1a and 1b show the results on TC-128 [3] and Figures 2a and 2b show the results on UAV-123 [4]. While the results are lower for all trackers due to the higher difficulty of these datasets, the context-aware CF trackers consistently outperform the corresponding baseline CF trackers by a margin. Also note that none of the parameters was adjusted and that the sampling strategy for the context patches is very naive. With further parameter tuning and a smarter sampling strategy the results can be much further improved.



(a) TC-128 - Precision Plot

(b) TC-128 - Success Plot

Fig. 1: Average precision and success on TC-128 for all sequences

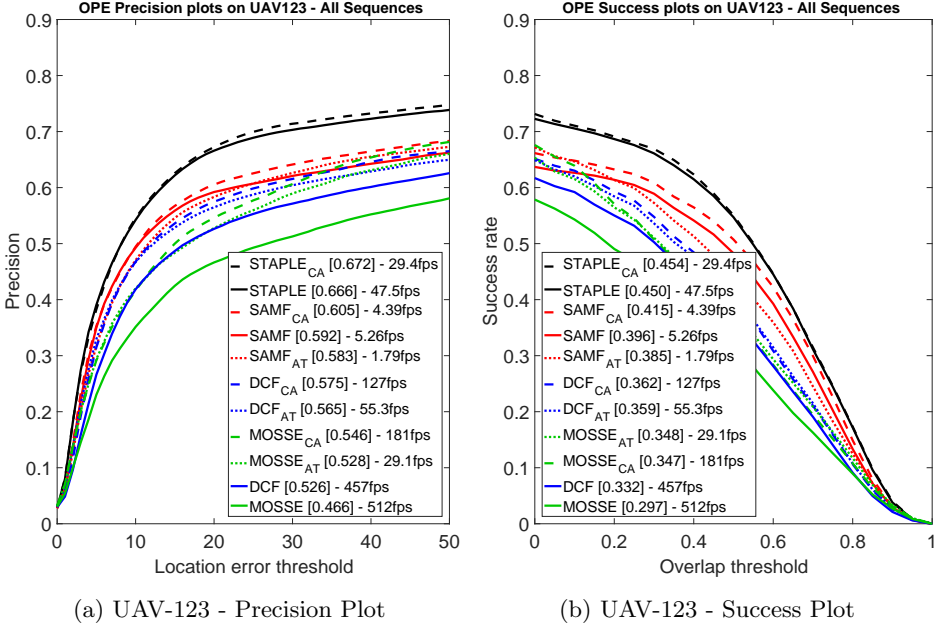


Fig. 2: Average precision and success on UAV-123 for all sequences

The following Figures show the performance for each attribute on OTB-100 [5].

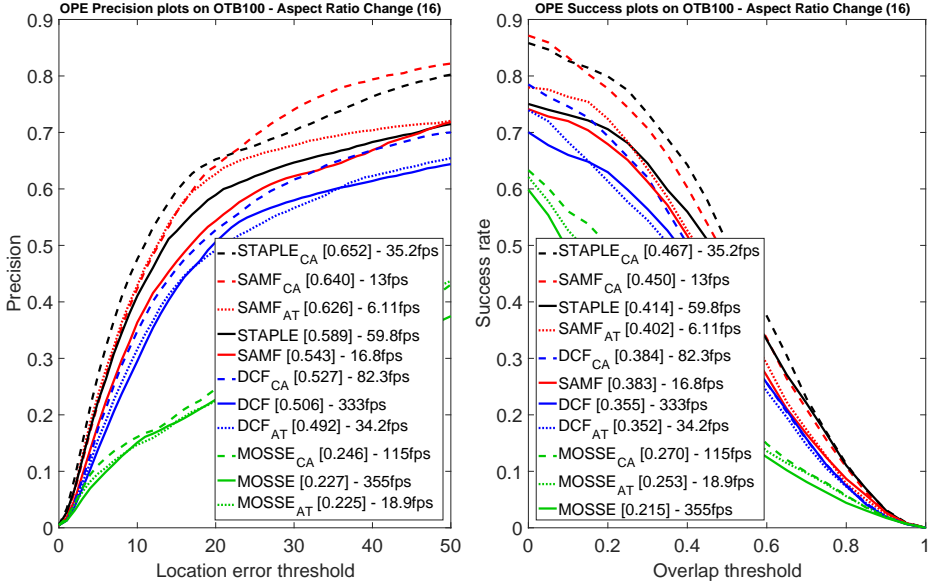


Fig. 3: Average precision and success on OTB-100 for videos with the attribute *Aspect Ratio Change*

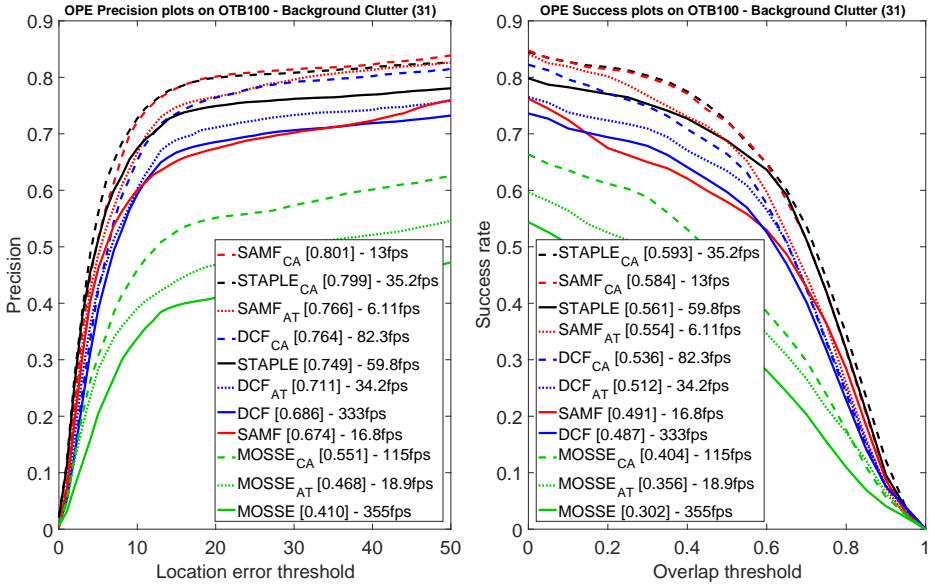


Fig. 4: Average precision and success on OTB-100 for videos with the attribute *Background Clutter*

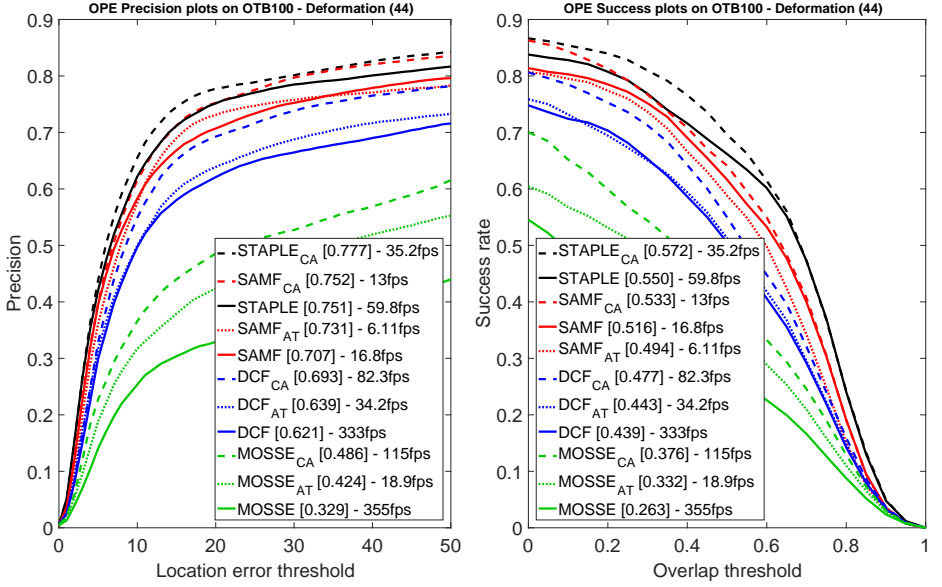


Fig. 5: Average precision and success on OTB-100 for videos with the attribute *Deformation*

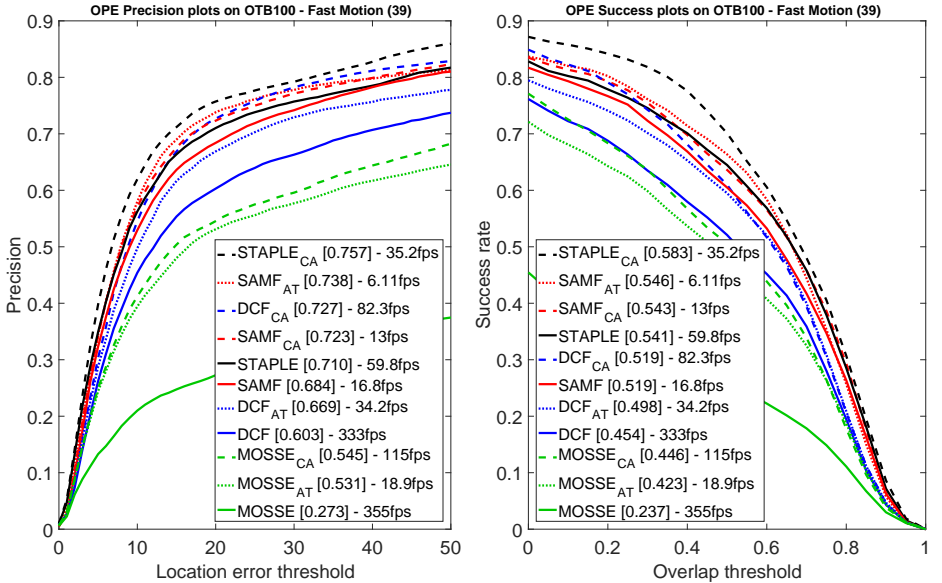


Fig. 6: Average precision and success on OTB-100 for videos with the attribute *Fast Motion*

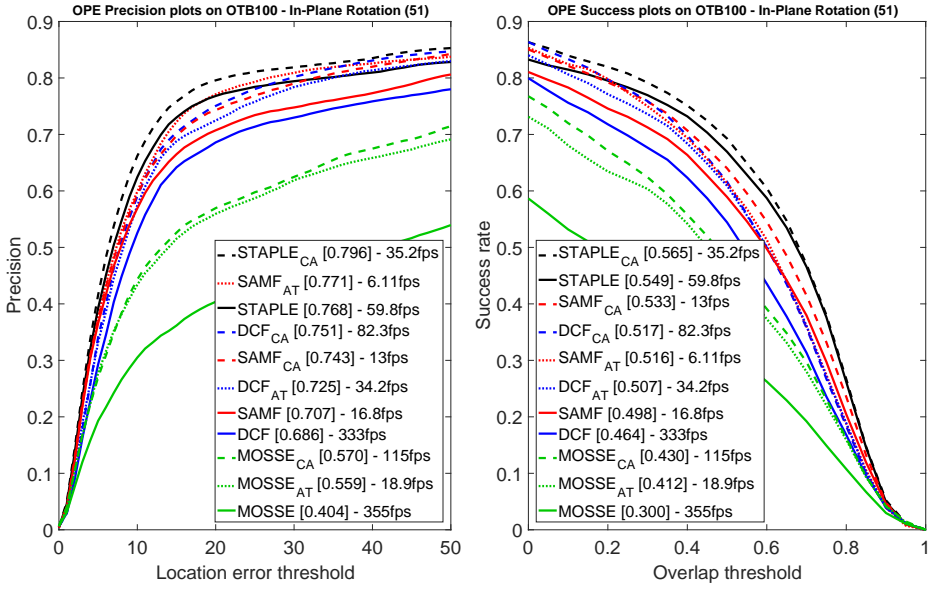


Fig. 7: Average precision and success on OTB-100 for videos with the attribute *In-Plane Rotation*

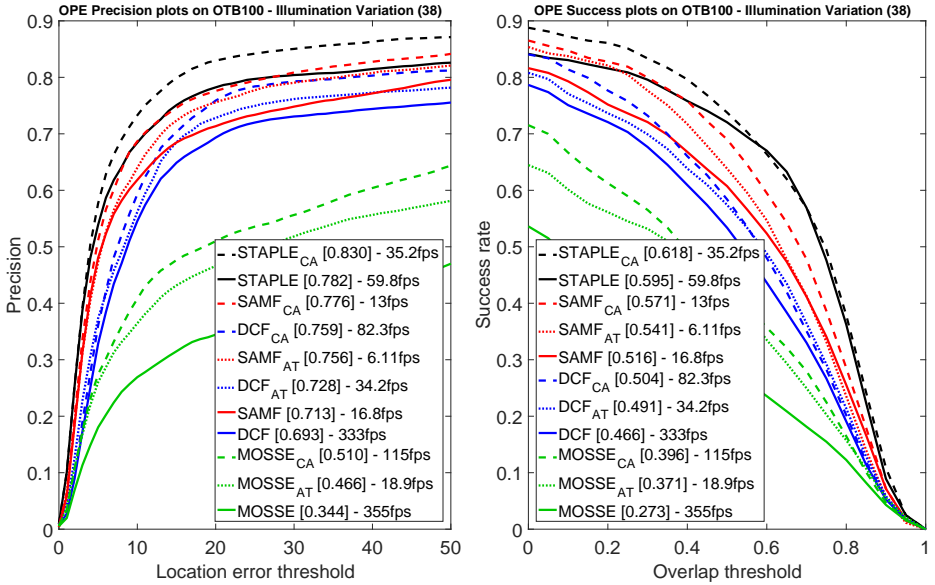


Fig. 8: Average precision and success on OTB-100 for videos with the attribute *Illumination Variation*

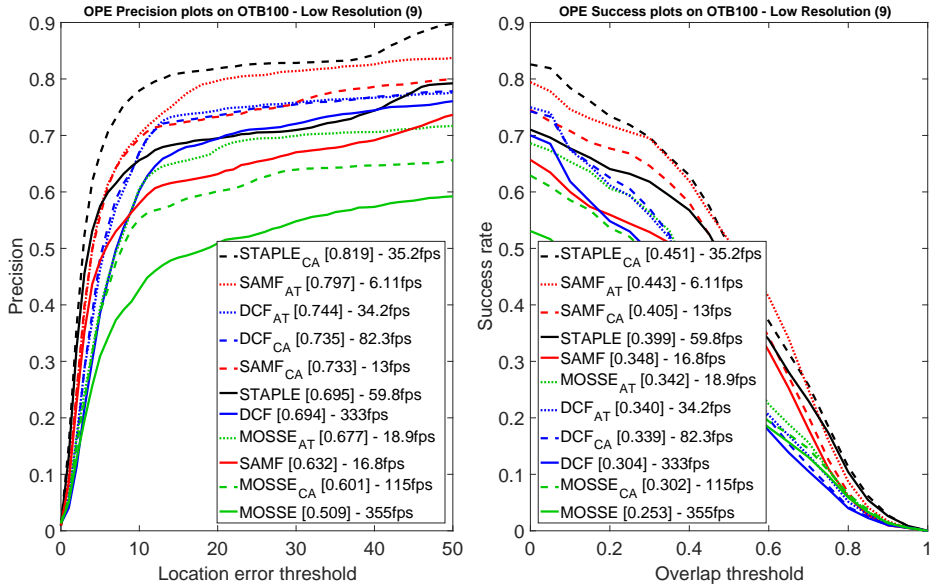


Fig. 9: Average precision and success on OTB-100 for videos with the attribute *Low Resolution*

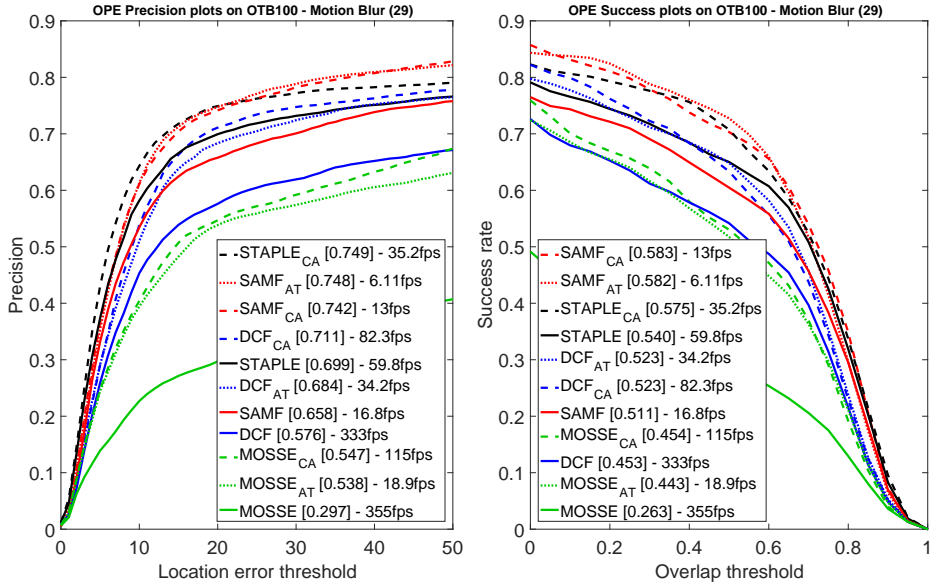


Fig. 10: Average precision and success on OTB-100 for videos with the attribute *Motion Blur*

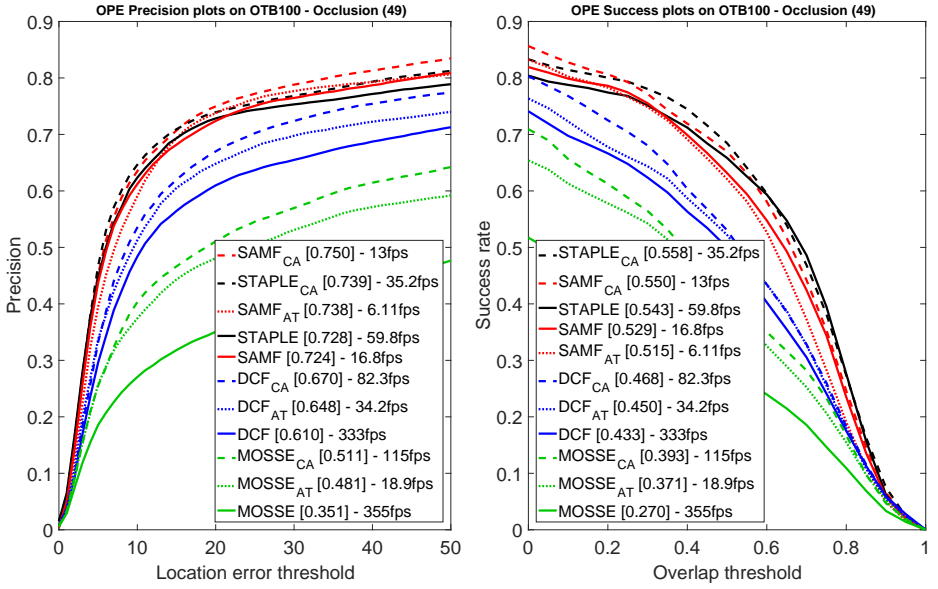


Fig. 11: Average precision and success on OTB-100 for videos with the attribute *Occlusion*

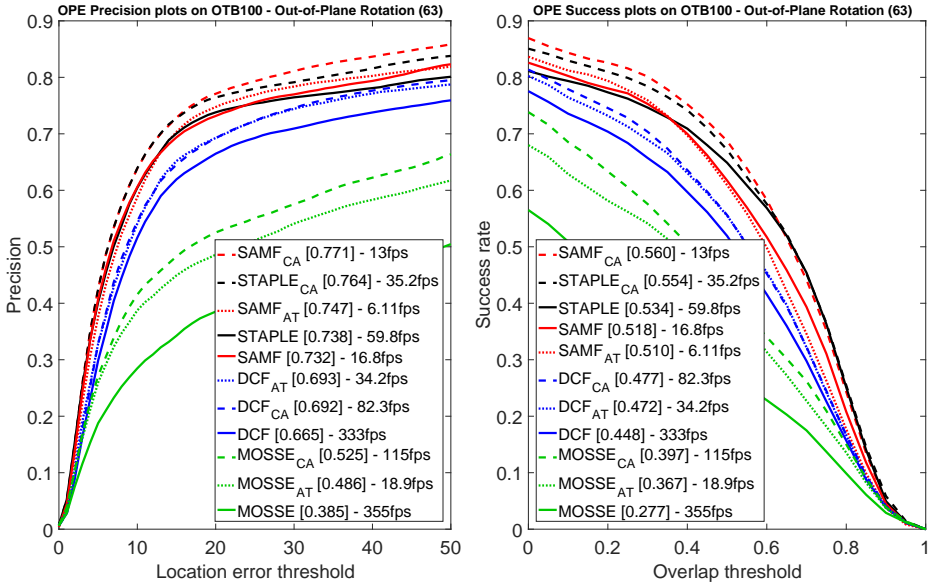


Fig. 12: Average precision and success on OTB-100 for videos with the attribute *Out-of-Plane Rotation*

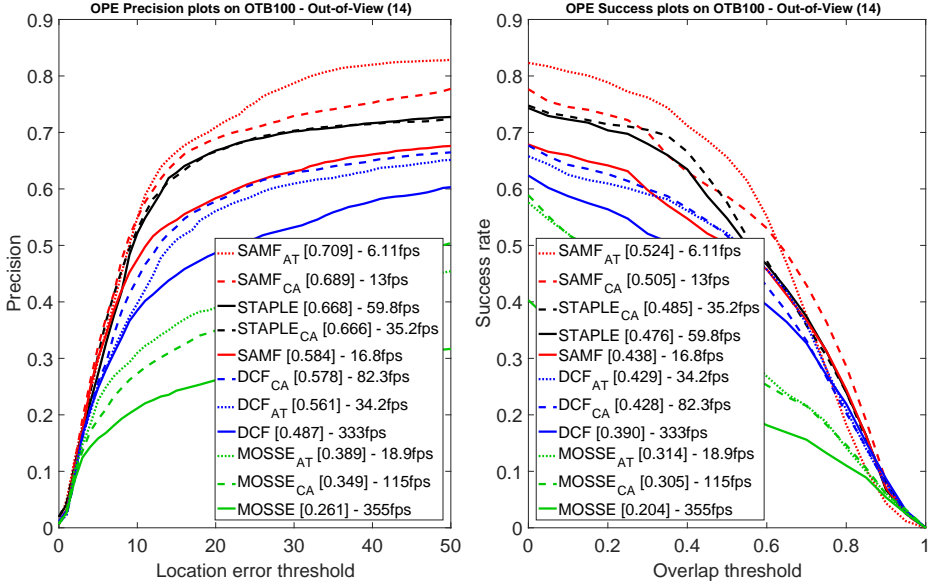


Fig. 13: Average precision and success on OTB-100 for videos with the attribute *Out-of-View*

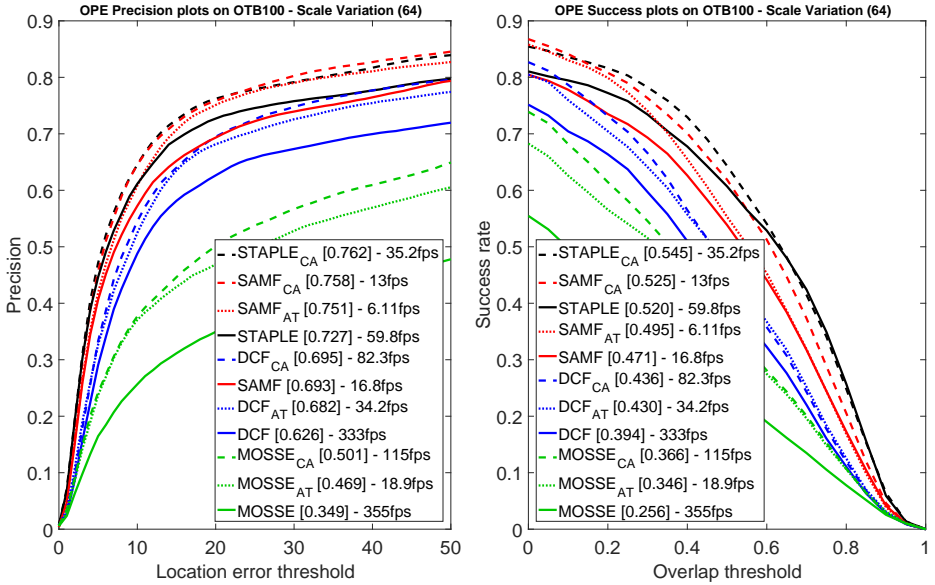


Fig. 14: Average precision and success on OTB-100 for videos with the attribute *Scale Variation*

References

1. J. F. Henriques, R. Caseiro, P. Martins, and J. Batista. Exploiting the circulant structure of tracking-by-detection with kernels. In *European Conference on Computer Vision, ECCV, 2012*.
2. J. F. Henriques, R. Caseiro, P. Martins, and J. Batista. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI, 2015.*, 2015.
3. P. Liang, E. Blasch, and H. Ling. Encoding color information for visual tracking: Algorithms and benchmark. *Image Processing, IEEE ...*, pages 1–14, 2015.
4. M. Mueller, N. Smith, and B. Ghanem. *A Benchmark and Simulator for UAV Tracking*, pages 445–461. Springer International Publishing, Cham, 2016.
5. Y. Wu, J. Lim, and M. H. Yang. Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9):1834–1848, Sept 2015.