Adaptive Context-Aware Discriminative Correlation Filters for Robust Visual Object Tracking

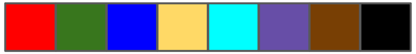
Tianyang Xu, Zhen-Hua Feng, Xiao-Jun Wu, Josef Kittler
University of Surrey
Jiangnan University
tianyang.xu@surrey.ac.uk

Task Definition

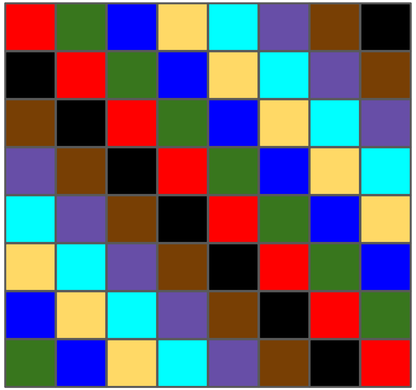
automatically track a target in a video sequence by predicting its location and scale

Background - Discriminative Correlation Filters



x

original vector

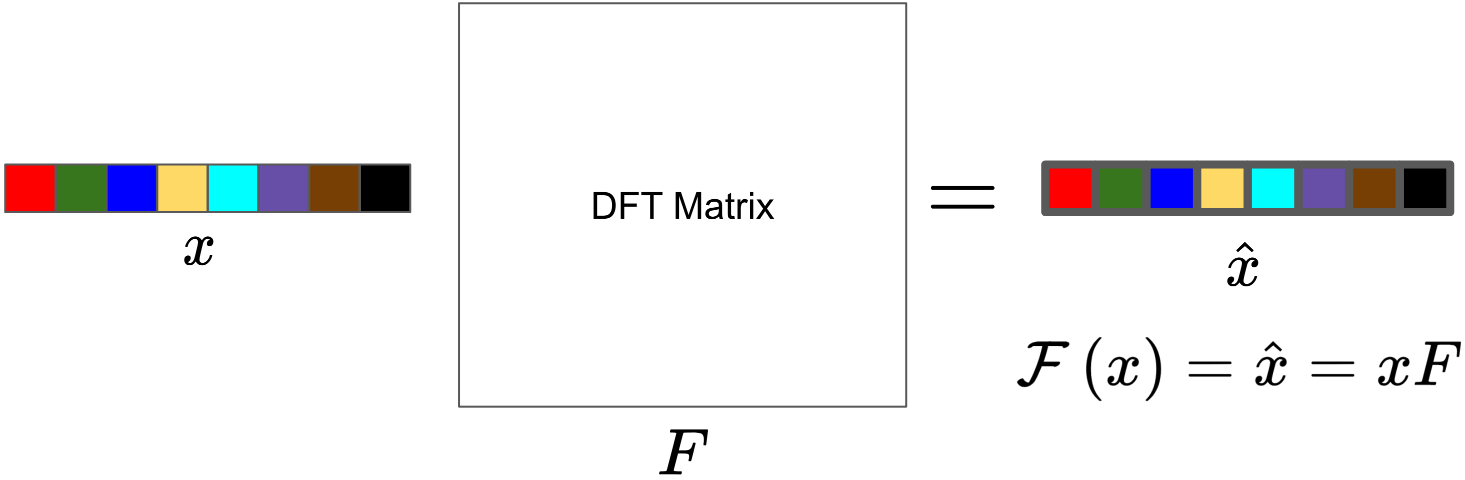


X

circulant matrix

Gray, Robert M. "Toeplitz and circulant matrices: A review." Foundations and Trends® in Communications and Information Theory 2.3 (2006): 155-239.

Background - Discriminative Correlation Filters



$$\mathcal{F}(x) = \hat{x} = xF$$

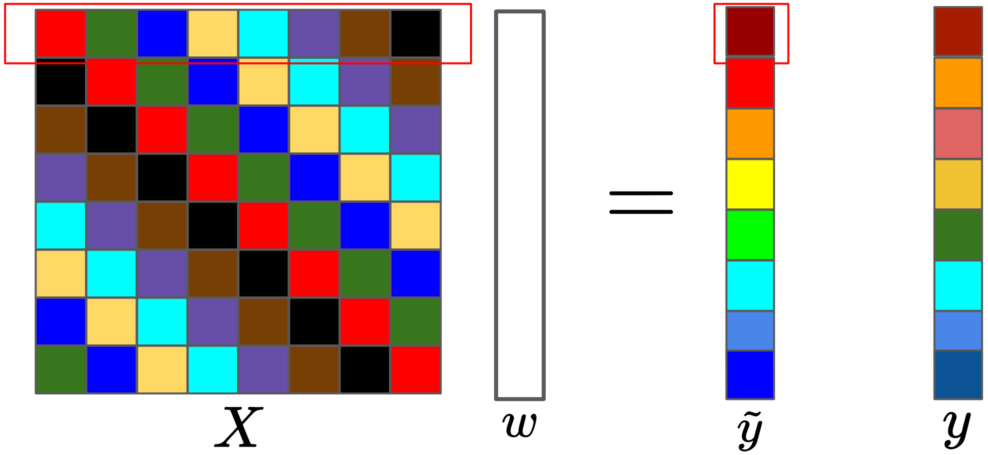
Stockham Jr, Thomas G. "High-speed convolution and correlation." Proceedings of the April 26-28, 1966, Spring joint computer conference. ACM, 1966.

Background - Discriminative Correlation Filters

$$\operatorname{diag}(\hat{x}) = \operatorname{diag}(xF) = F^H X F \qquad X = F \operatorname{diag}(\hat{x}) F^H$$

$$\boxed{X^H X} = F \operatorname{diag}(\hat{x}^* \odot \hat{x}) F^H$$

Background - Discriminative Correlation Filters



$$\min ||\boxed{X}w - y||_2^2 + \lambda||w||_2^2$$

Background - Discriminative Correlation Filters

$$\min \|Xw - y\|_2^2 + \lambda \|w\|_2^2$$

$$w = \left(X^T X + \lambda I\right)^{-1} X^T y \longleftarrow \boxed{\begin{array}{c} X^H X = F \mathrm{diag}\left(\hat{x}^* \odot \hat{x}\right) F^H \\ X = F \mathrm{diag}\left(\hat{x}\right) F^H \end{array}}$$

$$\hat{w} = \frac{\hat{x}^* \odot \hat{y}}{\hat{x}^* \odot \hat{x} + \lambda}$$

the original domain		**the frequency domain**
matrix multiplication	------>	element-wise multiplication
inverse matrix	------>	element-wise division

Henriques, João F., et al. "High-speed tracking with kernelized correlation filters." IEEE transactions on pattern analysis and machine intelligence 37.3 (2015): 583-596.

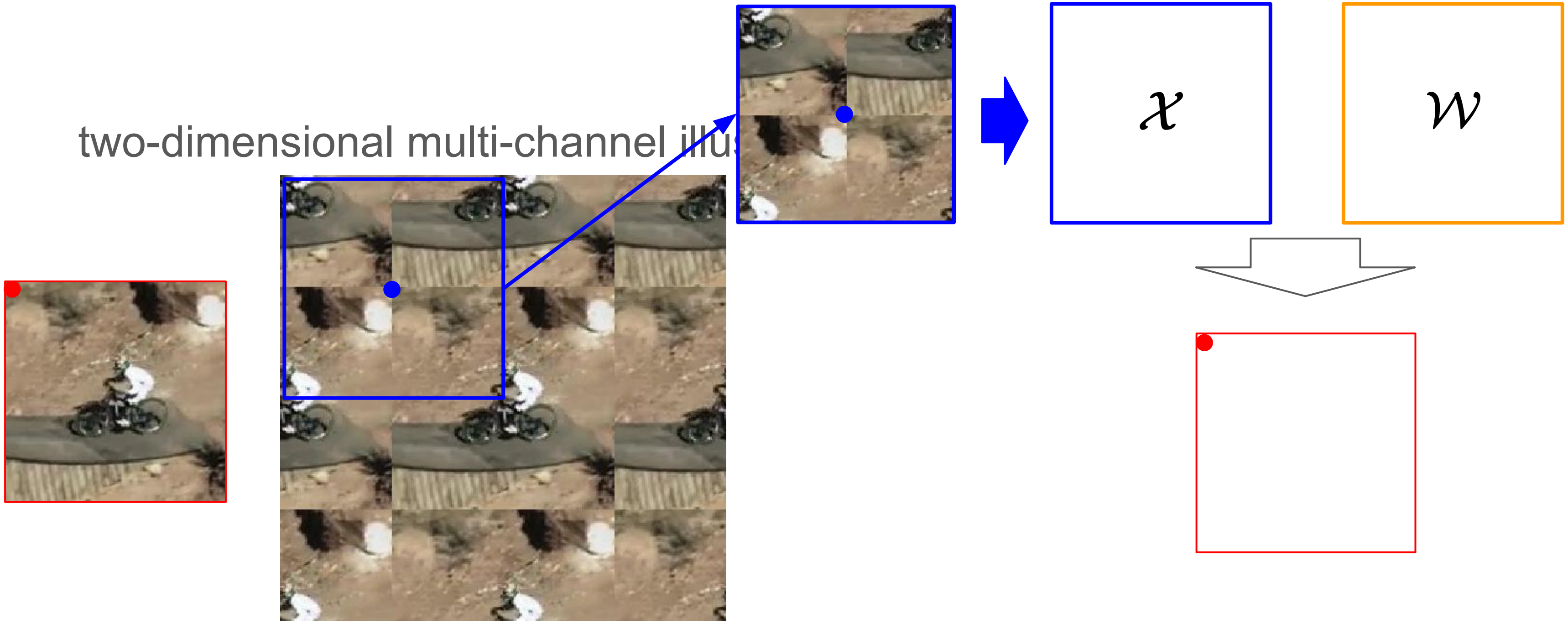
Background - Discriminative Correlation Filters

Advantages:

1. data augmentation
2. superior efficiency

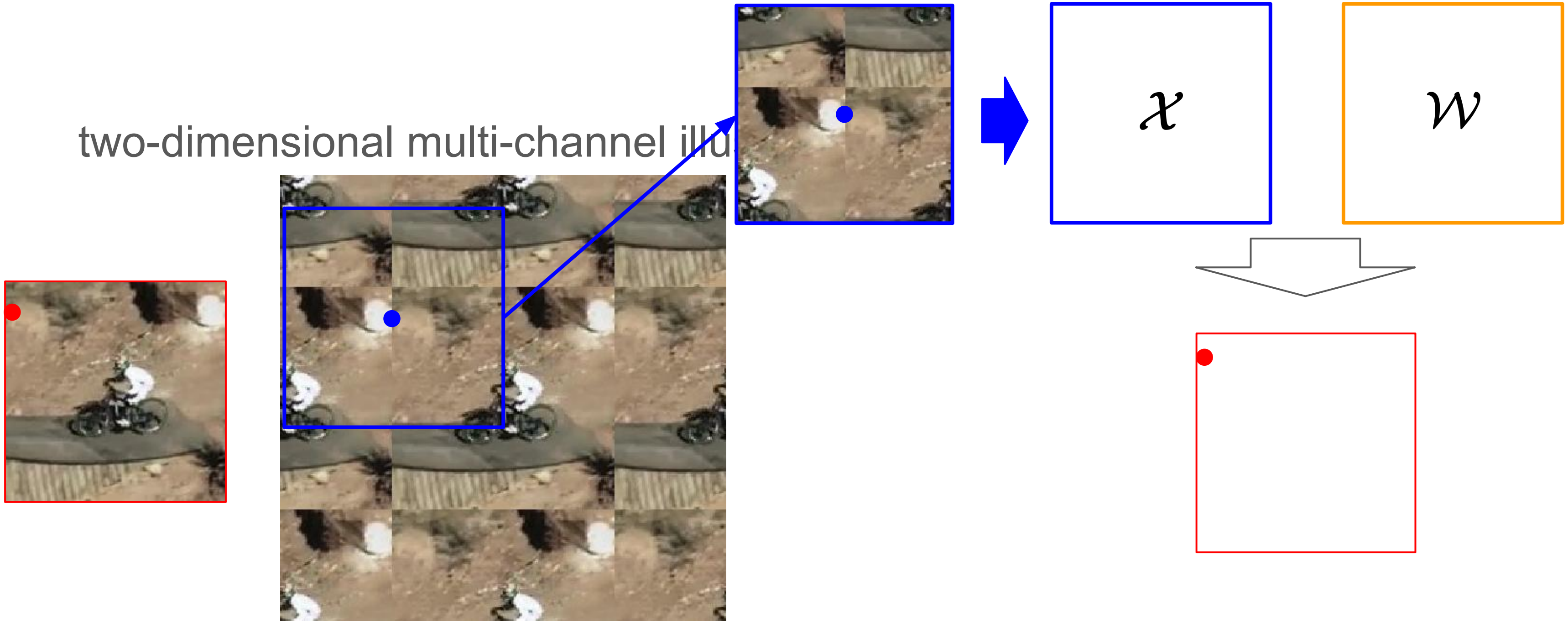
$$1 \rightarrow n$$
$$\mathcal{O}\left(n^{3}\right) \rightarrow \mathcal{O}\left(n \log n\right)$$

Background - Discriminative Correlation Filters

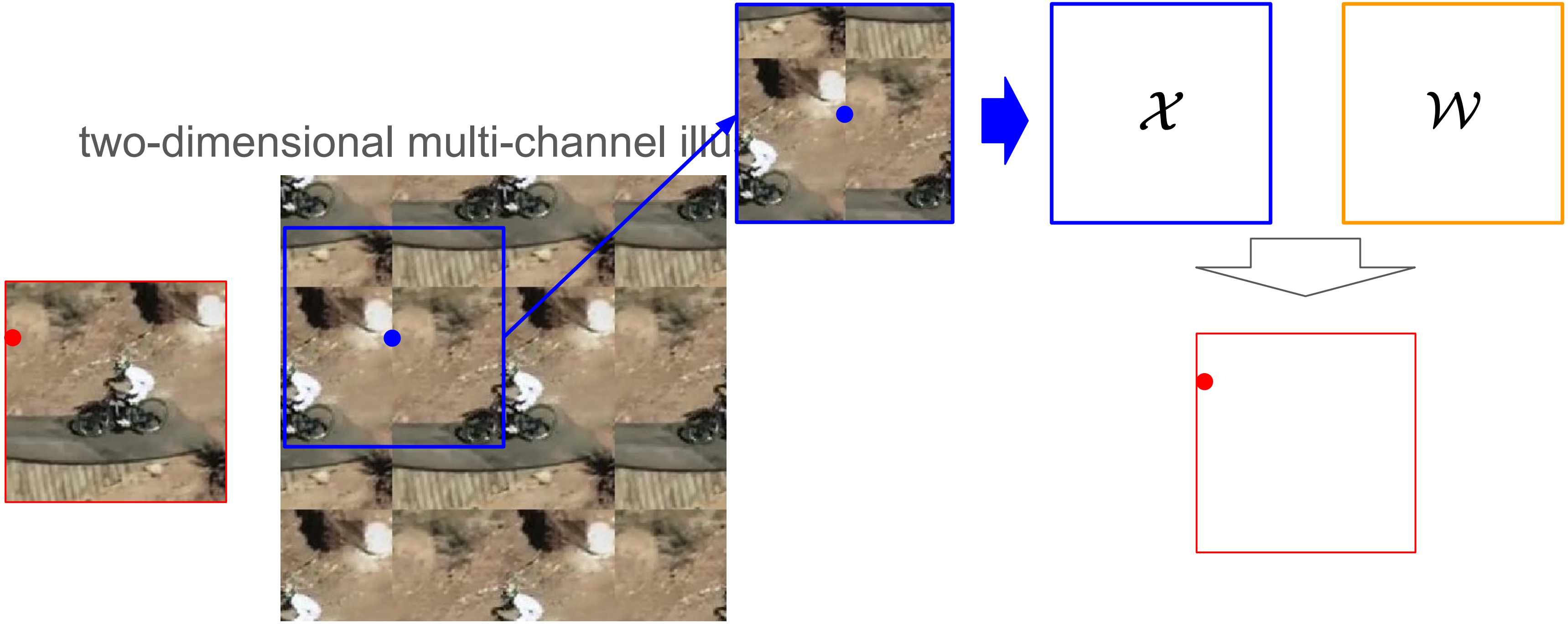


Background - Discriminative Correlation Filters

two-dimensional multi-channel illu

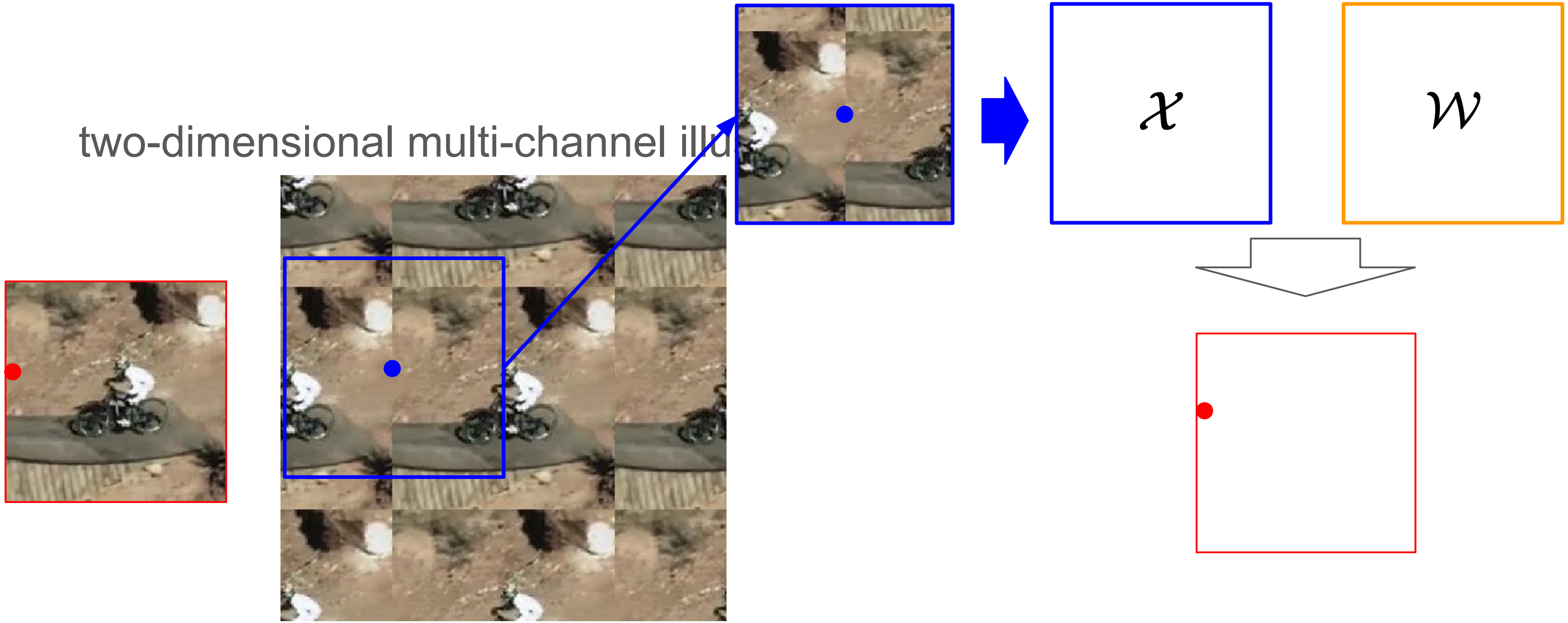


Background - Discriminative Correlation Filters

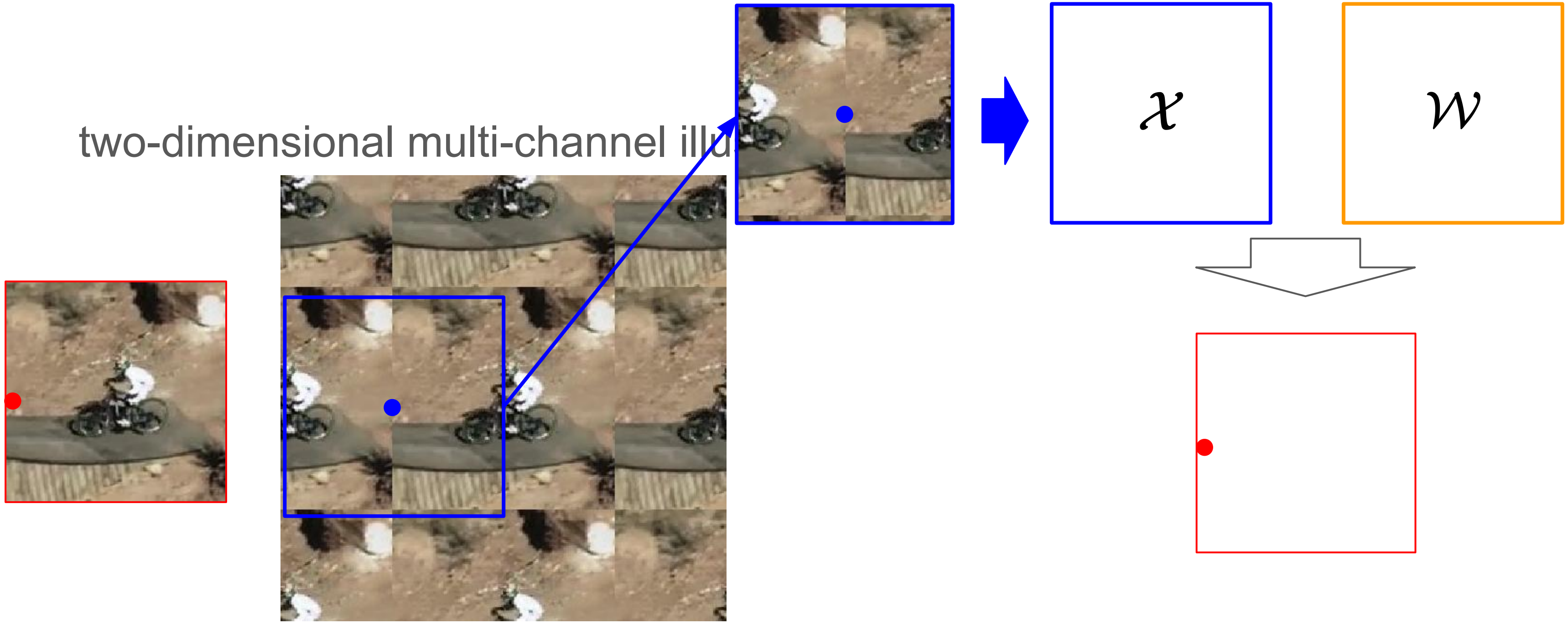


Background - Discriminative Correlation Filters

two-dimensional multi-channel illu

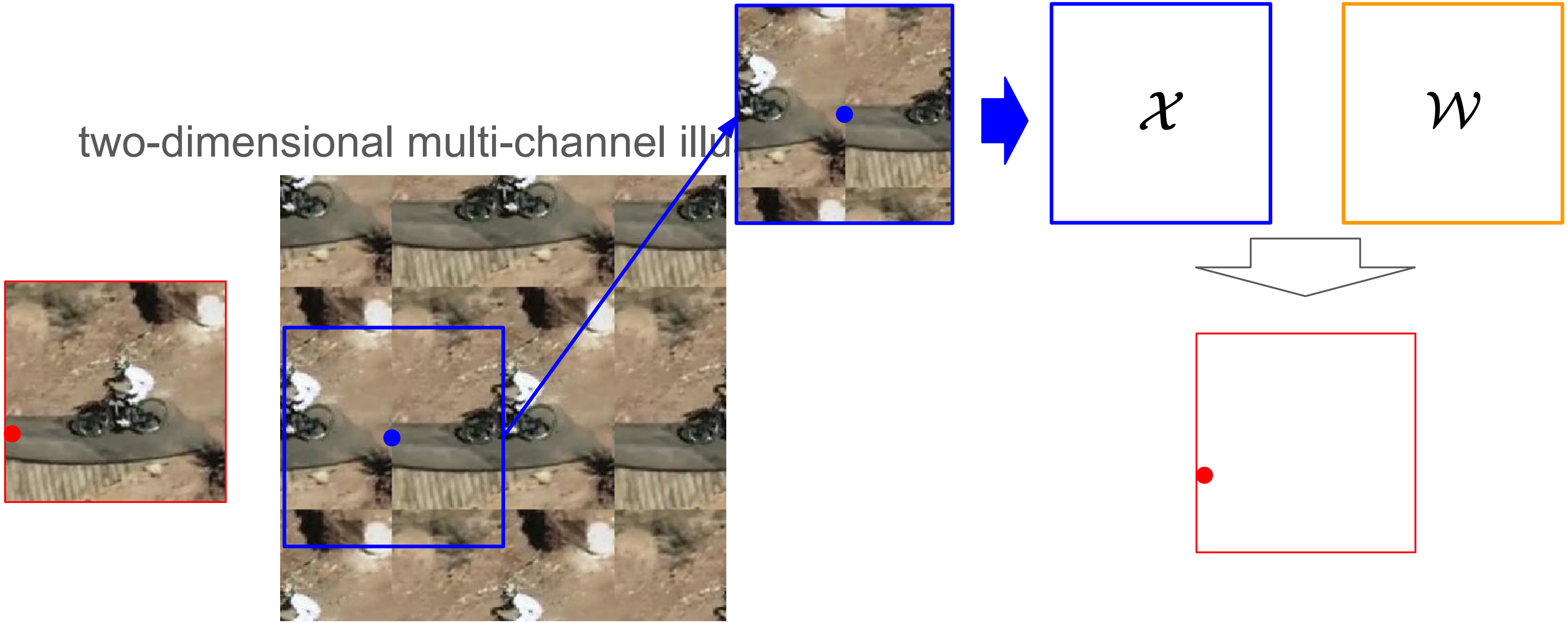


Background - Discriminative Correlation Filters

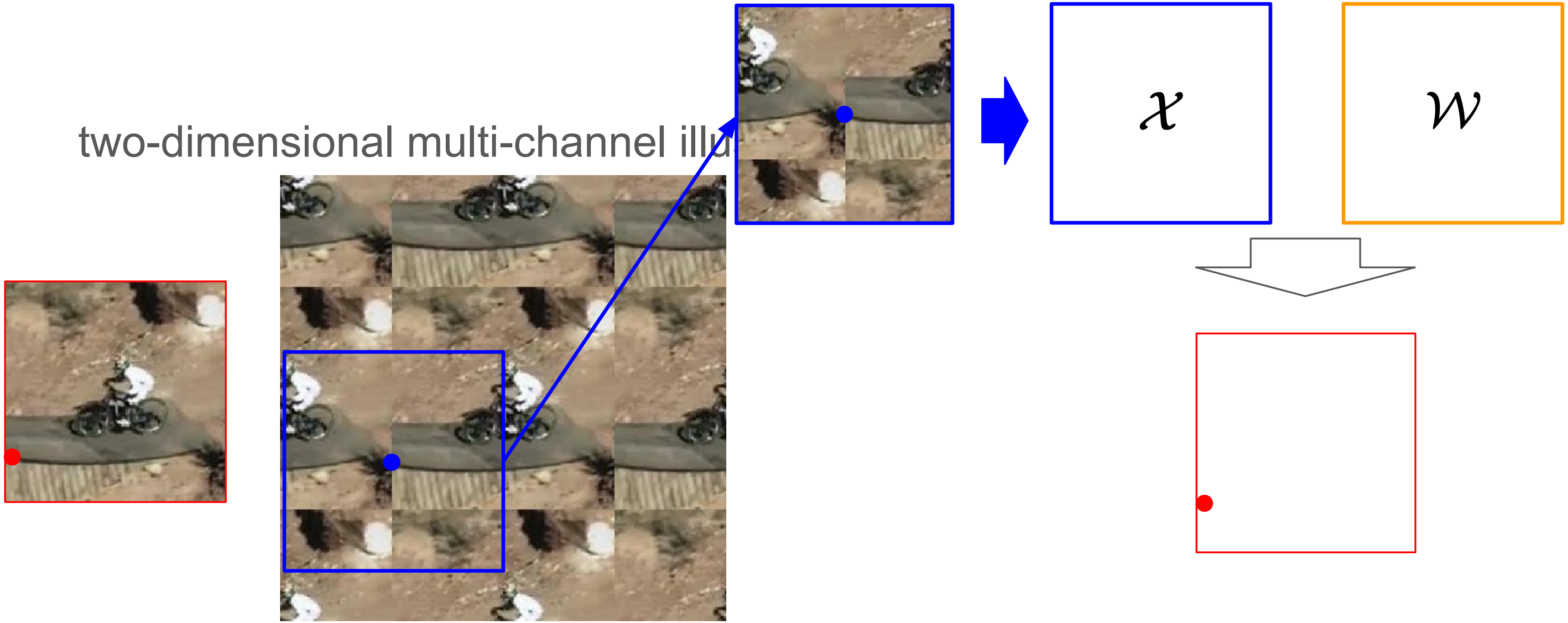


Background - Discriminative Correlation Filters

two-dimensional multi-channel illu

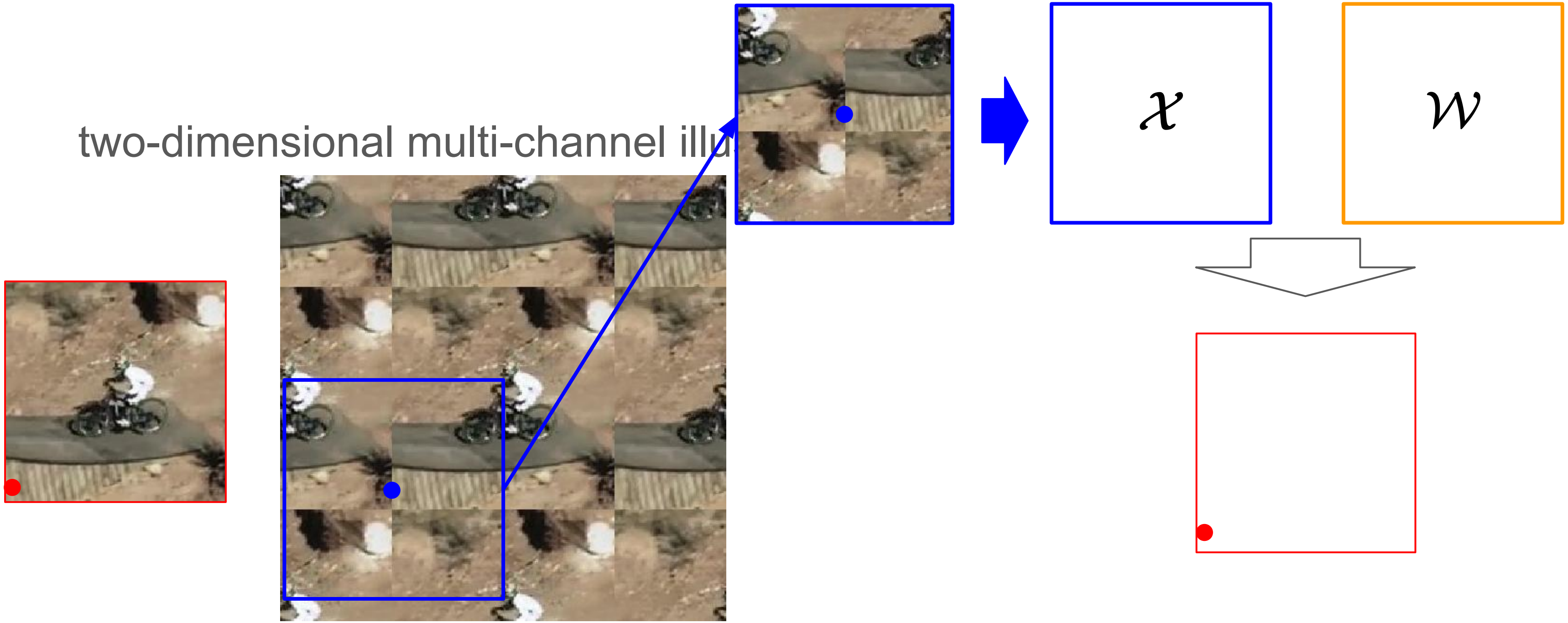


Background - Discriminative Correlation Filters

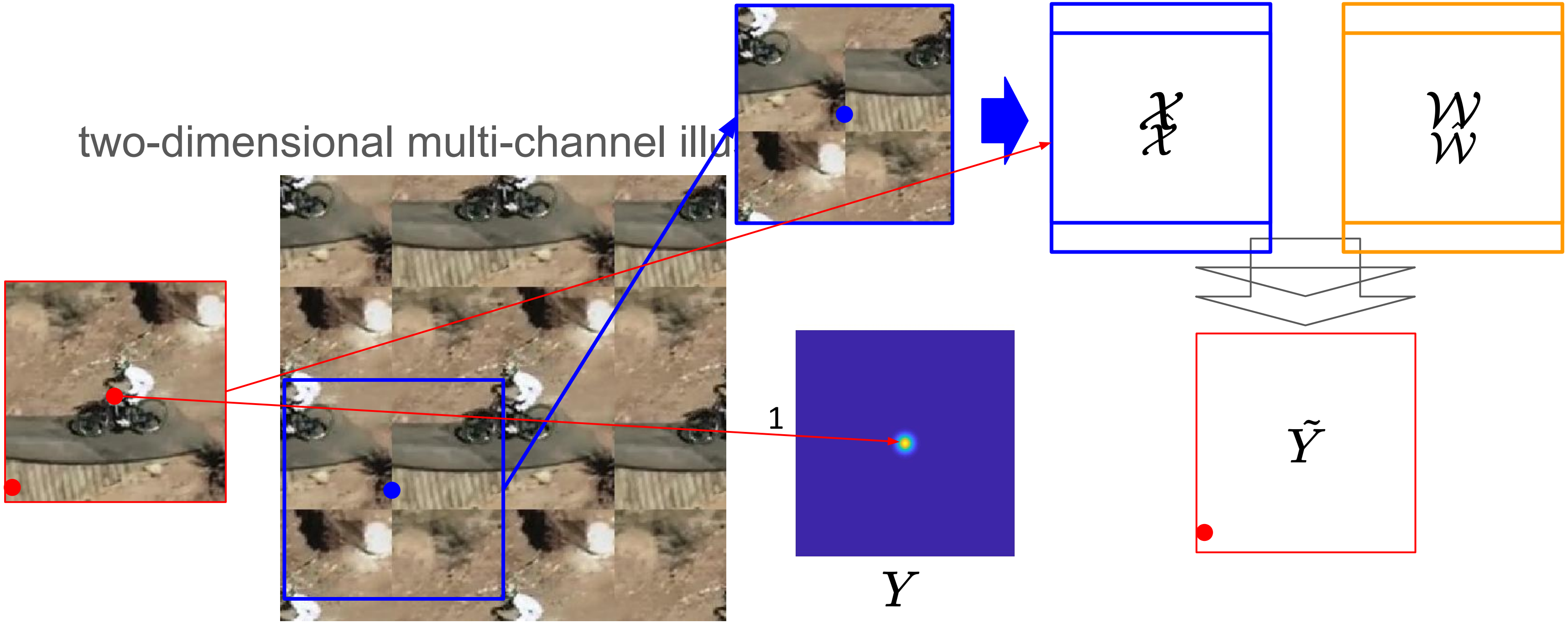


Background - Discriminative Correlation Filters

two-dimensional multi-channel illu



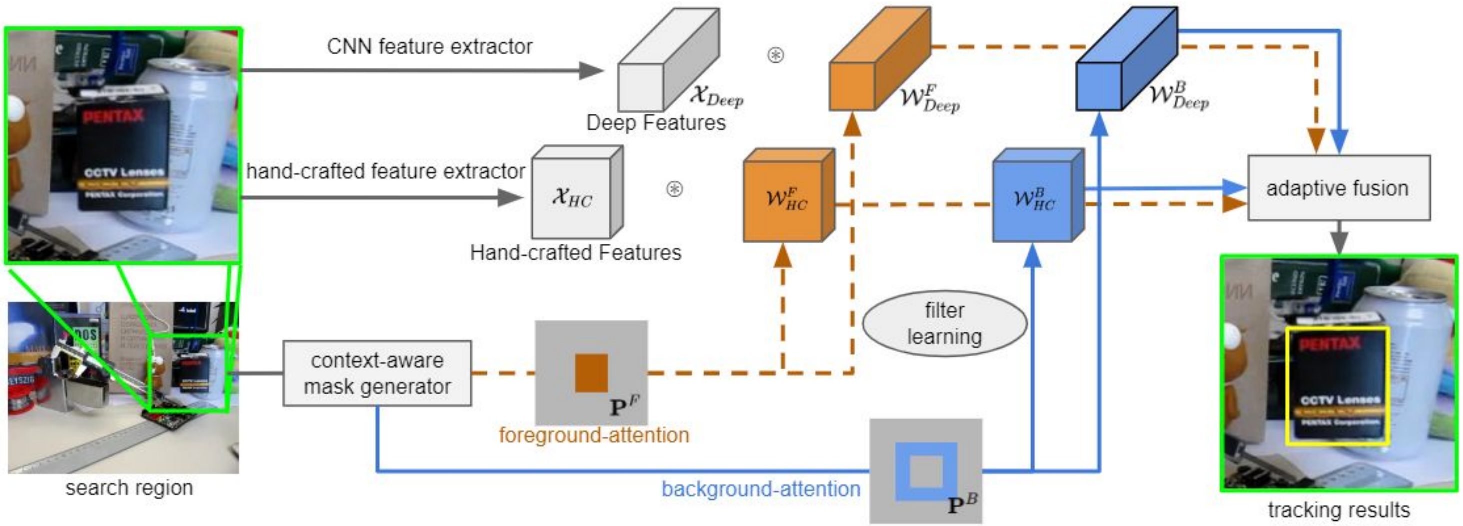
Background - Discriminative Correlation Filters



ACA-DCF - Motivation

Explore the foreground / background context-aware clues within the hand-crafted and deep feature representations to enhance discrimination and robustness.

ACA-DCF - Framework



The proposed adaptive context-aware DCF method using both hand-crafted and deep features. The proposed context-aware mask generator is instrumental in applying foreground-background attention to the learned filters. The corresponding responses are adaptively fused to generate the final result.

ACA-DCF - Context-aware masks

We design our context-aware attention masks with two complementary components, the foreground-attention mask $\mathbf{P}^F$ as:

$$p_{ij}^F = \begin{cases} 1 \times 10^{-3} & (i,j) \in \mathcal{F} \\ 1 & (i,j) \notin \mathcal{F} \end{cases}$$

and the background-attention mask $\mathbf{P}^B$ as:

$$p_{ij}^B = \begin{cases} 1 \times 10^{-3} & (i,j) \in \mathcal{B} \\ 1 & (i,j) \notin \mathcal{B} \end{cases}$$

where p_{ij} is the i-th row j-th column element of $\mathbf{P}$. $\mathcal{F}$ and $\mathcal{B}$ are the target region and surrounding region.

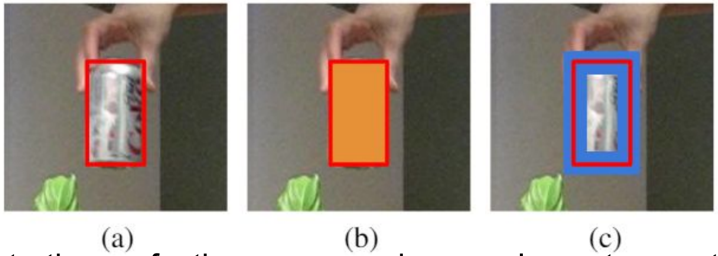


Illustration of the proposed complementary attention mechanism based context-aware mask generator: (a) target bounding box; (b) foreground-attention mask; and (c) background-attention mask.

ACA-DCF - Formulation

To better illustrate the spatial information, we formulate our ACA-DCF in a 2D and multi-channel manner:

$$\min \left\| \sum_{k=1}^{C} \mathbf{X}^k * \mathbf{W}^k - \mathbf{Y} \right\|_F^2 + \lambda \sum_{k=1}^{C} \left\| \mathbf{P} \odot \mathbf{W}^k \right\|_F^2,$$

where $\mathbf{X}^k \in \mathbb{R}^{N \times N}$ is the k-th channel representation of a 3-rd order feature tensor $\mathcal{X} \in \mathbb{R}^{N \times N \times C}$, $\mathbf{W}^k \in \mathbb{R}^{N \times N}$ is the k-th corresponding filter that is a slice of the filter tensor $\mathcal{W} \in \mathbb{R}^{N \times N \times C}$, and $\mathbf{P}^{N \times N}$ is the designed spatial regularisation mask.

ACA-DCF - Optimization

We use the augmented Lagrange method to optimise the objective function.
We use slack variable $\mathcal{W}' = \mathcal{W}$ (for each k, $\mathbf{W}'^k = \mathbf{W}^k$) and construct the following Lagrange function:

$$\mathcal{L} = \left\| \sum_{k=1}^{C} \mathbf{X}^k * \mathbf{W}^k - \mathbf{Y} \right\|_F^2 + \lambda \sum_{k=1}^{C} \left\| \mathbf{P} \odot \mathbf{W}'^k \right\|_F^2 + \frac{\mu}{2} \sum_{k=1}^{C} \left\| \mathbf{W}^k - \mathbf{W}'^k + \frac{\mathbf{\Gamma}^k}{\mu} \right\|_F^2,$$

$$\begin{cases} \hat{\mathbf{w}}_{i,j} = \left(\mathbf{I} - \dfrac{\hat{\mathbf{x}}_{i,j} \hat{\mathbf{x}}_{i,j}^\top}{\mu/2 + \hat{\mathbf{x}}_{i,j}^\top \hat{\mathbf{x}}_{i,j}} \right) \mathbf{g} \\ \mathbf{W}'^k = (\mathbf{1} - \mathbf{P}) \odot \dfrac{\mu \mathbf{W}^k + \mathbf{\Gamma}^k}{2\lambda + \mu} \\ \Gamma = \Gamma + \mu (\mathcal{W} - \mathcal{W}') \end{cases}$$

$$\mathbf{g} = \left(\hat{\mathbf{x}}_{i,j} \hat{y}_{i,j} + \mu \hat{\mathbf{w}}'_{i,j} - \mu \hat{\gamma}_{i,j} \right) / \mu$$

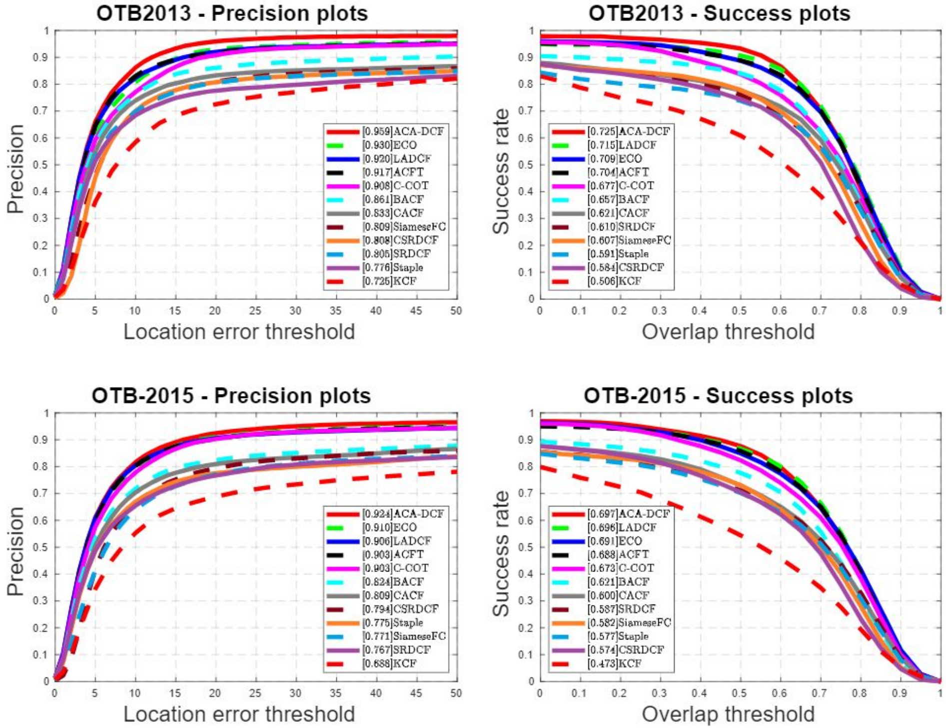
ACA-DCF - Optimization

Solution:

$$\begin{cases} \hat{\mathbf{w}}_{i,j} = \left(\mathbf{I} - \dfrac{\hat{\mathbf{x}}_{i,j}\hat{\mathbf{x}}_{i,j}^{\top}}{\mu/2 + \hat{\mathbf{x}}_{i,j}^{\top}\hat{\mathbf{x}}_{i,j}} \right) \mathbf{g} \\ \mathbf{W}'^{k} = (\mathbf{1} - \mathbf{P}) \odot \dfrac{\mu \mathbf{W}^{k} + \boldsymbol{\Gamma}^{k}}{2\lambda + \mu} \\ \Gamma = \Gamma + \mu \left(\mathcal{W} - \mathcal{W}' \right) \end{cases}$$

$$\mathbf{g} = \left(\hat{\mathbf{x}}_{i,j}\hat{y}_{i,j} + \mu \hat{\mathbf{w}}'_{i,j} - \mu \hat{\gamma}_{i,j} \right) / \mu$$

Experiments



Evaluation on OTB2013 and OTB2015, using the precision plots with DP in the legend and the success plots with AUC in the legend.

Experiments

	STAPLE+	EBT [34]	DDC	Staple [17]	MLDF	SSAT	TCNN [35]	C-COT [21]	**ACA-DCF**
EAO	0.286	0.291	0.293	0.295	0.311	0.329	**0.327**	**0.331**	**0.385**
Accuracy	**0.559**	0.465	0.542	0.547	0.492	**0.579**	0.555	0.541	**0.581**
Robustness	0.37	0.25	0.34	0.38	**0.23**	0.29	0.27	**0.24**	**0.21**

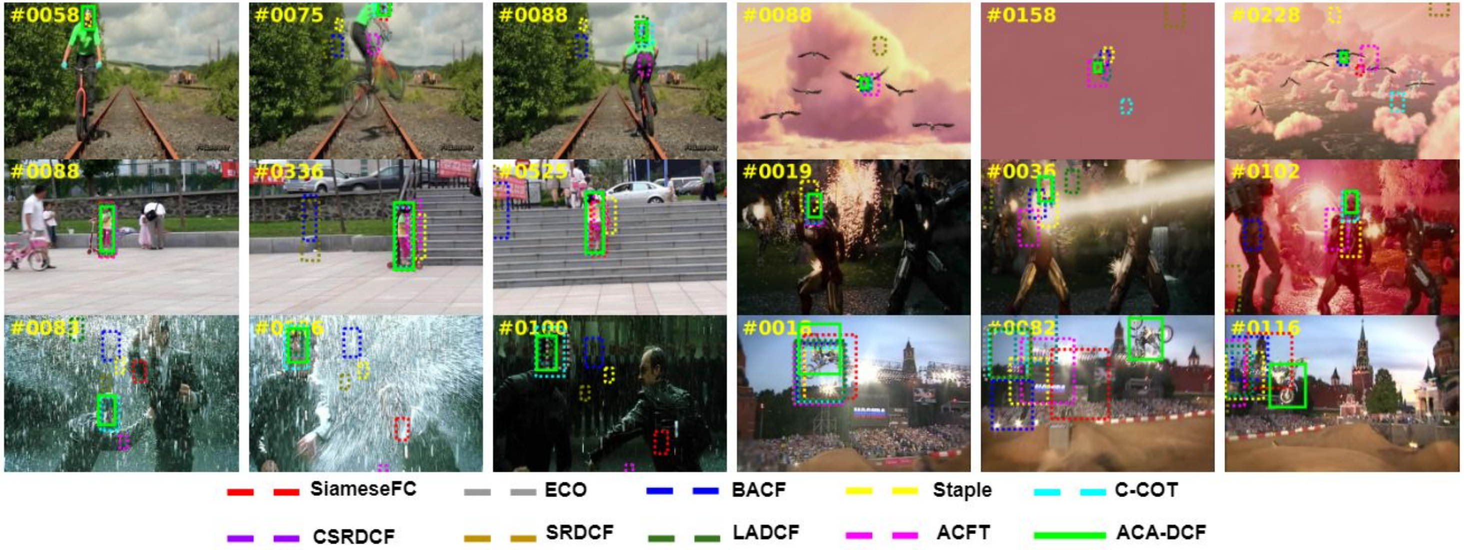
Tracking results on VOT2016.

Experiments

		DEF	IPR	IV	OPR	OV	SV	LR
Mean **DP/AUC** (%/%)	DSST [36] [TPAMI-17]	56.2/40.0	70.7/47.4	73.9/49.6	65.8/43.7	48.1/36.5	64.3/39.5	67.8/30.5
	Staple [17] [CVPR-16]	72.9/54.4	75.3/54.4	76.9/58.8	72.4/52.8	66.1/48.1	72.7/52.5	69.5/39.6
	SRDCF [15] [ICCV-15]	70.7/53.1	69.6/51.7	74.5/58.9	71.3/53.4	58.2/45.8	73.1/55.4	74.2/51.0
	BACF [26] [ICCV-17]	77.8/58.2	79.5/58.4	83.0/64.2	78.7/58.4	76.5/55.2	77.4/57.6	79.5/51.4
	CFNet [24] [CVPR-17]	69.6/50.8	76.8/57.2	70.5/54.9	74.1/54.7	53.6/42.3	72.6/55.0	81.0/58.6
	CSRDCF [27] [CVPR-17]	73.5/52.4	73.6/51.9	73.2/53.9	72.1/51.5	65.1/49.6	75.0/52.8	81.3/45.1
	ACFN [37] [CVPR-17]	77.2/53.5	78.0/54.3	78.8/56.7	77.7/54.3	67.3/49.7	76.0/54.8	81.8/51.5
	STAPLE_CA [38] [CVPR-17]	76.0/56.6	80.6/57.4	81.6/61.3	75.8/55.2	69.7/50.9	75.3/54.1	81.9/44.8
	TRACA [39] [CVPR-18]	76.9/56.1	80.6/57.6	84.1/62.2	82.3/59.3	68.0/53.4	76.1/55.2	86.0/50.2
	CREST [25] [ICCV-17]	77.6/56.9	85,3/61.7	87.6/64.4	84.2/61.5	73.4/56.6	78.6/57.2	86.6/47.3
	SiamFC [4] [ECCV-16]	69.0/50.6	74.2/55.7	73.6/56.8	75.6/55.8	66.9/50.6	73.5/55.2	90.0/**61.8**
	MetaT [40] [ECCV-18]	83.8/62.0	87.7/63.5	86.4/63.4	85.2/62.7	72.3/56.0	80.3/58.2	90.1/47.2
	MCPF [41] [TPAMI-18]	81.6/57.0	**88.8**/62.0	88.1/62.8	86.7/61.9	76.4/55.3	86.2/60.3	**96.3**/58.7
	C-COT [21] [ECCV-16]	**85.9**/61.4	87.7/62.7	88.4/68.2	89.9/65.2	89.5/64.8	**88.1**/65.4	**97.5/62.9**
	ECO [18] [CVPR-17]	**85.9/63.3**	**89.2/65.5**	**91.4/71.3**	**90.7/67.3**	**91.3/66.0**	87.9/**66.6**	88.2/59.1
	LADCF [13] [TIP-19]	**87.2/65.2**	88.3/**65.9**	**89.0/70.5**	**90.4/68.2**	**91.0/67.2**	**88.2/67.0**	87.9/61.7
	ACA-DCF	**89.3/64.8**	**93.1/67.8**	**93.3/72.6**	**91.8/68.0**	**94.0/68.0**	**90.3/68.1**	**99.6/70.7**

The DP and AUC results on OTB2015, parameterised by 7 attributes.

Experiments



Examples of qualitative tracking results on challenging sequences (Left column top to down: Biker, Girl2, and Matrix. Right column top to down: Bird1, Ironman, and MotorRolling). The colour bounding boxes denote the results of SiameseFC, ECO, BACF, Staple, C-COT, CSRDCF, SRDCF, LADCF, ACFT, and ACA-DCF, respectively.

Conclusion

We improve the DCF formulation by designing a novel adaptive context-aware mechanism.

-The information contents of a target and its surroundings are analysed by the proposed complementary foreground-background attention mechanism.

- The two sources of information are fused by a novel adaptive fusion strategy to further improve the robustness with the consideration of the dynamics of target and background appearance variations.

-The experimental results obtained on well-known benchmarking datasets demonstrate the effectiveness and robustness of our method. It was shown to achieve superior performance over the state-of-the-art visual object tracking algorithms.

Thank you