Target Response Adaptation for Correlation Filter Tracking

Adel Bibi, Matthias Mueller, and Bernard Ghanem
King Abdullah University of Science and Technology (KAUST)

KAUST

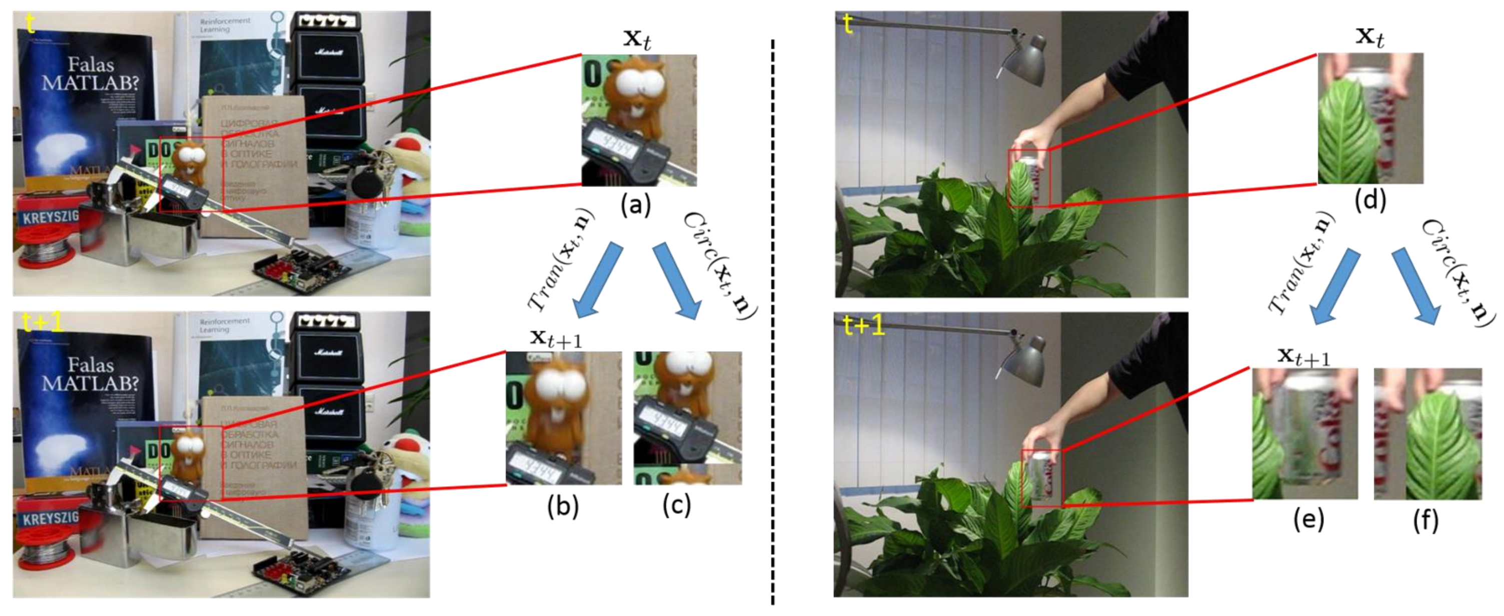
Project page + Code

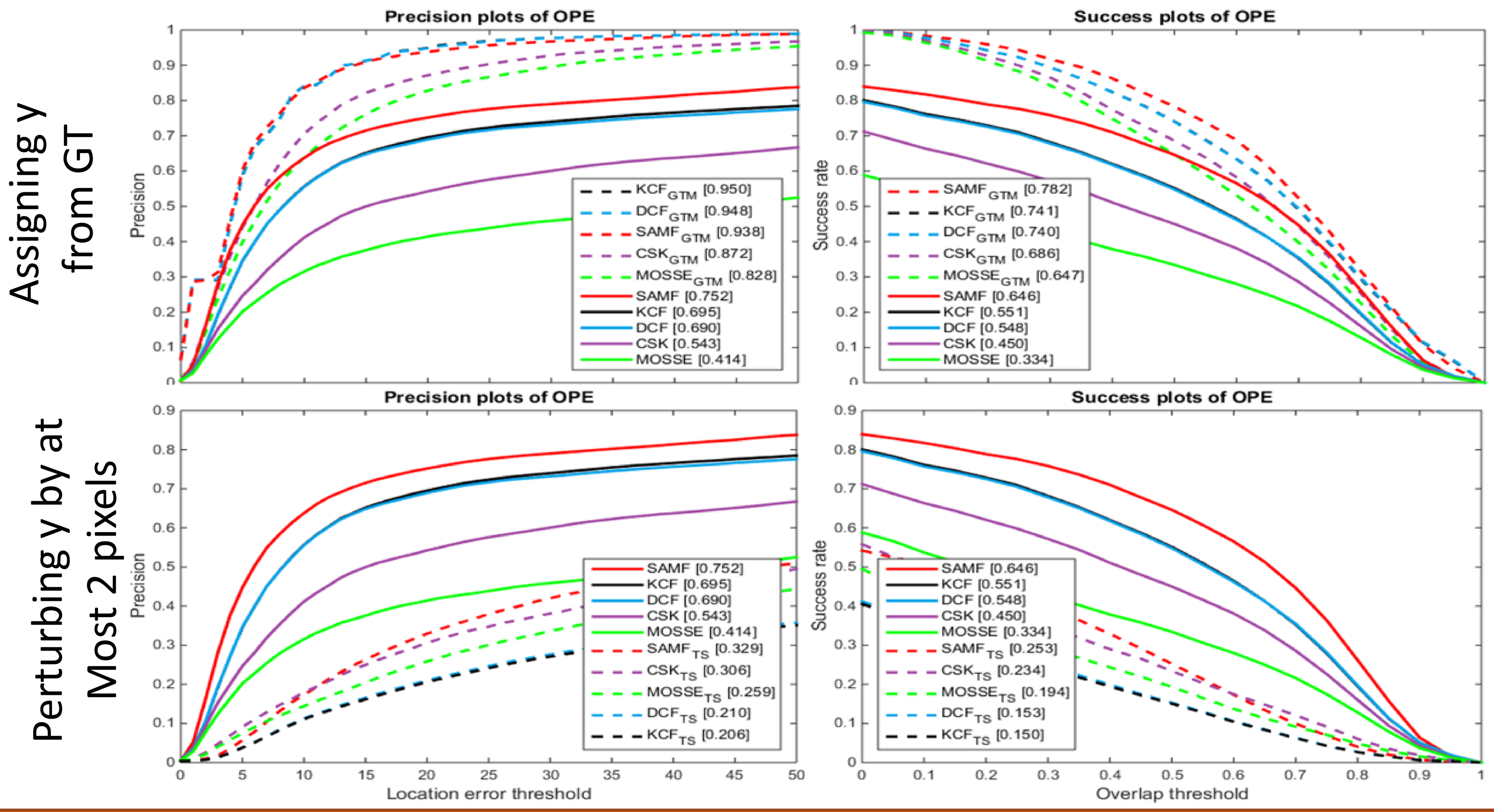
ECCV'16 EUROPEAN CONFERENCE ON COMPUTER VISION

Abstract

- Correlation filter (CF) trackers use circularly shifted patches as a proxy to real translations when training the filter.
- In all CF trackers, correlation scores are regressed to a Gaussian pulse centered around the previous location.
- This limits the trackers' ability to recover from partial occlusion or to track fast moving objects.
- We propose a generic framework, in which we solve for both the optimal filter and target response based on exact translation detection scores from the image.
- The final formulation can use kernels and multiple templates jointly if solved in the dual domain. It improves all baseline CF trackers (SAMF, KCF, DCF, CSK, and MOSSE) by (3 – 13)% on OTB100 [1].

Motivation





Problem Formulation

In our formulation, we solve for both the filter $\mathbf{w}$ and the target response $\mathbf{y}$ in a unified objective. The target response is regressed to $\mathbf{y}_o$. The resultant problem is solved in the dual domain with an efficient solution that allows the use of non-linear kernels and multiple templates jointly. The resulting optimization is:

$$\underset{\mathbf{w},\mathbf{y}}{\text{minimize}} \quad ||\tilde{\mathbf{X}}\mathbf{w}-\mathbf{y}||_2^2+\lambda_1||\mathbf{w}||_2^2+\lambda_2||\mathbf{y}-\mathbf{y}_o||_2^2$$

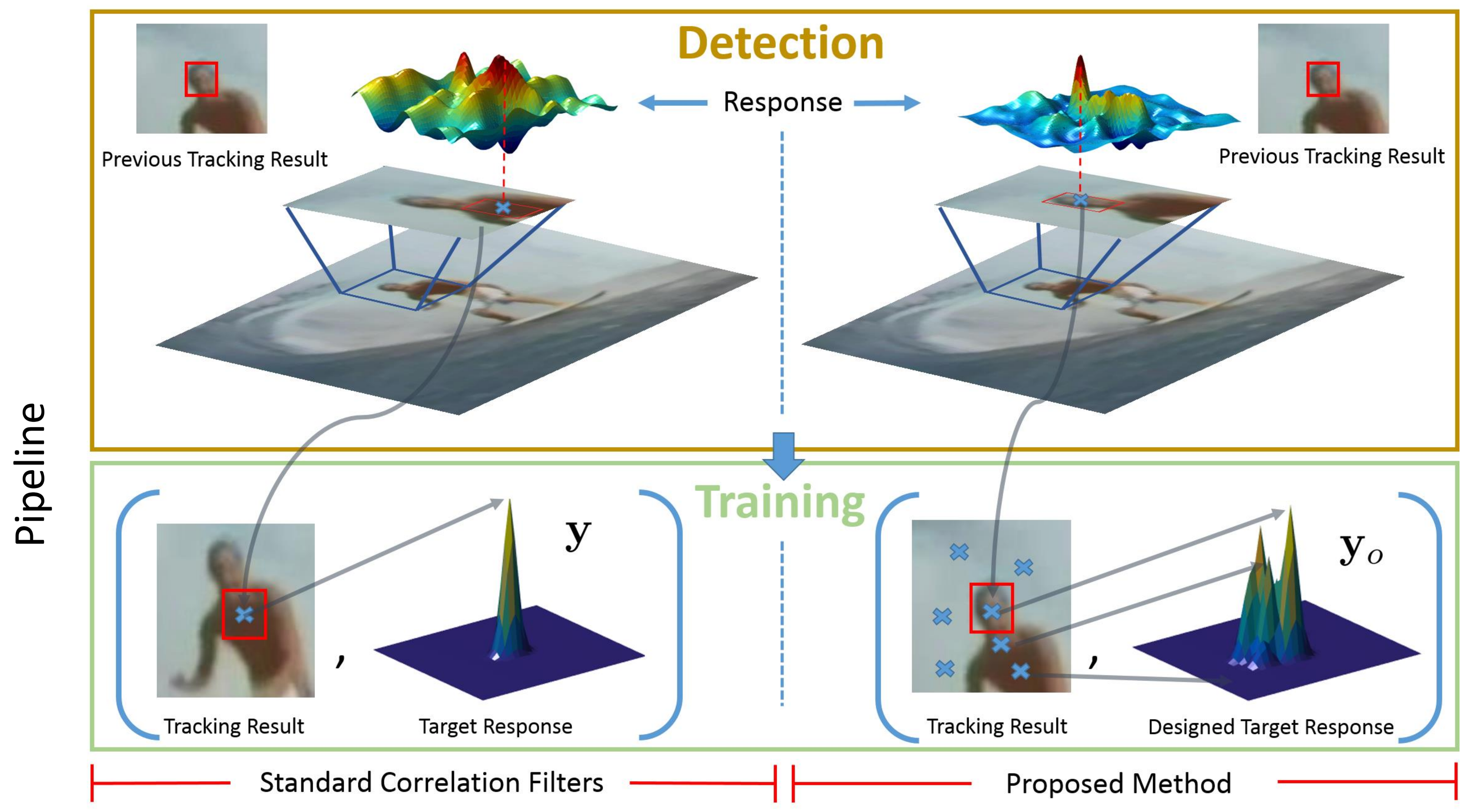
$$\mathbf{D}\tilde{\mathbf{K}}^{-1}\left(\lambda_2\mathbf{D}^T\mathbf{D}+\lambda_1\mathbf{E}^T\mathbf{E}+\tilde{\mathbf{G}}^T\tilde{\mathbf{G}}\right)\tilde{\mathbf{K}}^{-1}\mathbf{D}^T\alpha=\lambda_2\mathbf{D}\tilde{\mathbf{K}}^{-1}\mathbf{D}^T\mathbf{y}_o$$

$$\tilde{\mathbf{K}}=\left(\lambda_1\mathbf{E}^T\mathbf{E}+\tilde{\mathbf{G}}^T\tilde{\mathbf{G}}\right) \qquad \tilde{\mathbf{G}}=\left[\tilde{\mathbf{X}} \quad -\tilde{\mathbf{I}}\right]\in\mathbb{R}^{kn\times 2n}$$

$$\tilde{\mathbf{I}}^\top=[\mathbf{I}\ \cdots\ \mathbf{I}]\in\mathbb{R}^{n\times kn} \qquad \mathbf{E}=\left[\mathbf{I} \quad \mathbf{0}\right]\in\mathbb{R}^{n\times 2n}$$

$$\mathbf{D}=\left[\mathbf{0} \quad \mathbf{I}\right]\in\mathbb{R}^{n\times 2n}$$

The regressor y_o is generated out of a Gaussian interpolation to few samples generated from the correlation scores of exact translations in the image.



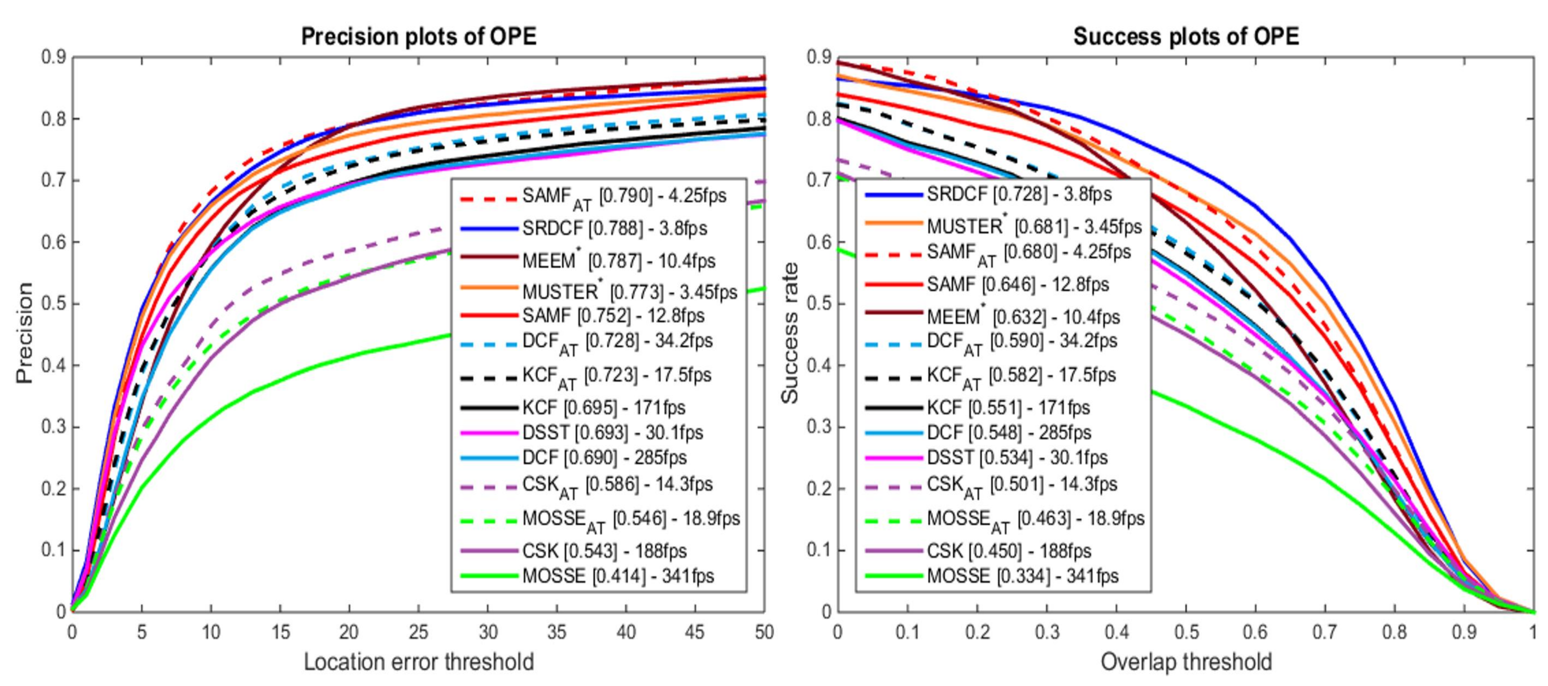
Solution

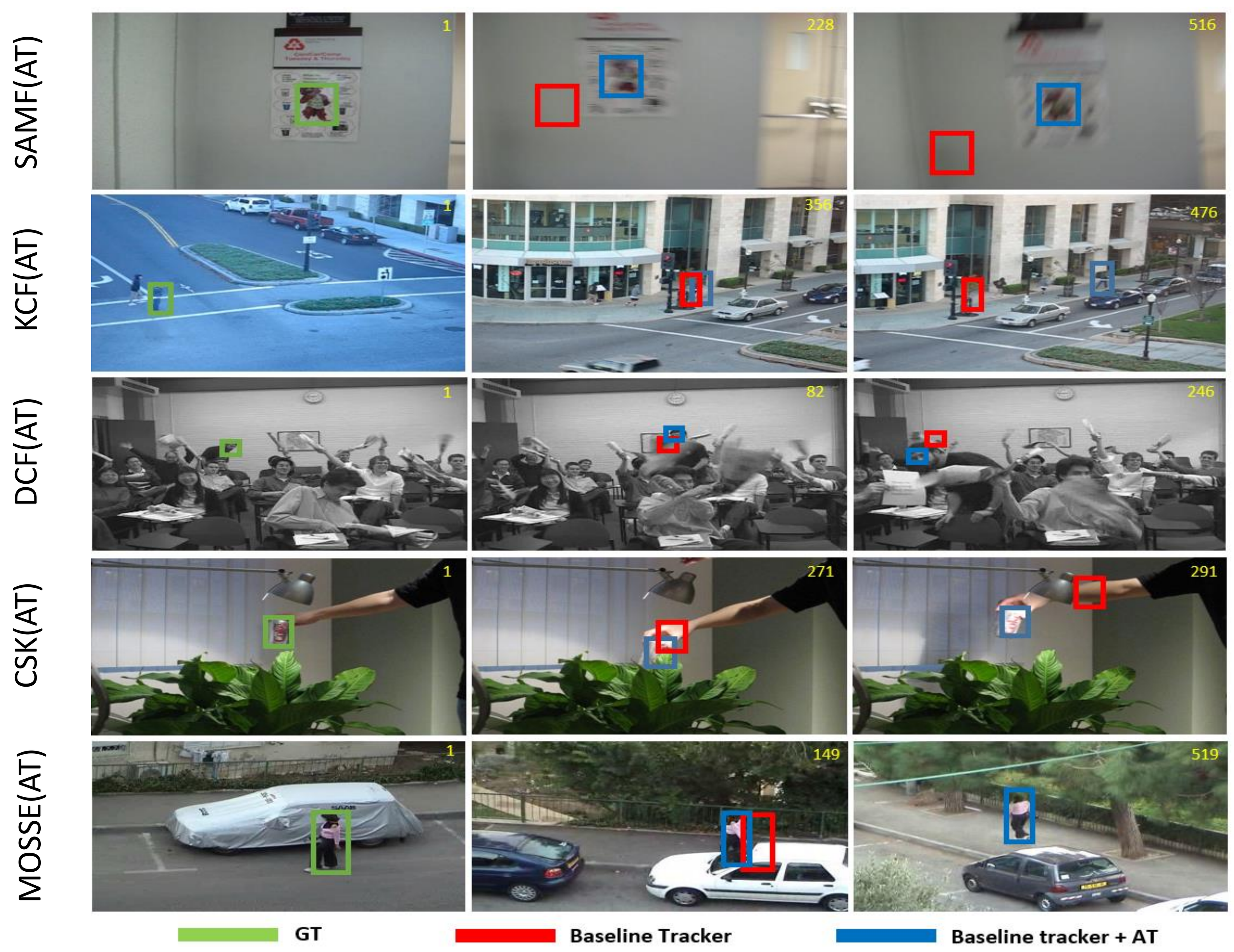
$$\hat{\alpha}^*=\lambda_2 diag^{-1}(\Upsilon)\left(\frac{\frac{1}{k}\left(\sum_i^k\hat{\mathbf{x}}_{1i}^*\right)\odot\left(\sum_i^k\hat{\mathbf{x}}_{1i}\right)\odot\hat{\mathbf{y}}_0^*}{\sum_i^k(\hat{\mathbf{x}}_{1i}^*\odot\hat{\mathbf{x}}_{1i})+\lambda_1-\frac{1}{k}(\sum_i^k\hat{\mathbf{x}}_{1i}^*\odot\sum_i^k\hat{\mathbf{x}}_{1i})}+\frac{\hat{\mathbf{y}}_o^*}{k}\right),$$

where
$$\Upsilon=\left(\frac{\frac{-1}{k}\sum_i^k(\hat{\mathbf{x}}_{1i}^*\odot\hat{\mathbf{x}}_{1i})+\frac{k+\lambda_2}{k}(\sum_i^k\hat{\mathbf{x}}_{1i}^*)\odot(\sum_i^k\hat{\mathbf{x}}_{1i})+\frac{\lambda_1(k+\lambda_2)}{k}}{\sum_i^k(\hat{\mathbf{x}}_{1i}^*\odot\hat{\mathbf{x}}_{1i})+\lambda_1-\frac{1}{k}(\sum_i^k\hat{\mathbf{x}}_{1i}^*)\odot(\sum_i^k\hat{\mathbf{x}}_{1i})}\right)\odot\left(\frac{\frac{1}{k^2}\sum_i^k\hat{\mathbf{x}}_{1i}^*\odot\sum_i^k\hat{\mathbf{x}}_{1i}}{\sum_i^k(\hat{\mathbf{x}}_{1i}^*\odot\hat{\mathbf{x}}_{1i})+\lambda_1-\frac{1}{k}(\sum_i^k\hat{\mathbf{x}}_{1i}^*)\odot(\sum_i^k\hat{\mathbf{x}}_{1i})}+\frac{1}{k}\right)$$

Experiments

We adopt our framework to five different baseline CF trackers: SAMF, KCF, DCF, CSK, and MOSSE. The performance of all baseline trackers improves significantly especially for the Fast Motion (FM), Motion Blur (MB), and Occlusion (OCC) categories. The experiments are done on the OTB100[1] dataset.





References: [1] Wu, Y. et al. Online tracking benchmark, PAMI15.

Acknowledgements. Research reported was supported by competitive funding from King Abdullah University of Science and Technology (KAUST).