International Journal of Computer Vision manuscript No.
(will be inserted by the editor)

Adaptive Correlation Filters with Long-Term and Short-Term Memory for Object Tracking Supplementary Document

Chao Ma · Jia-Bin Huang · Xiaokang Yang · Ming-Hsuan Yang

Received: date / Accepted: date

Abstract In this supplementary document, we present two additional ablation studies on the OTB2013 dataset. First, we show the results of directly minimizing the errors over all the tracked results to update the correlation filters. Second, we analyze the robustness of the proposed method by spatially shifting the ground truth bounding boxes.

By directly minimizing the errors over all the tracked results, we consider *all* the extracted appearances $\{x_j, j = 1, 2, \dots, p\}$ of the target object from the first frame up to the current frame p. The cost function is the weighted average quadratic error over these p frames. We assign each frame j with a weight $\beta_j \geq 0$ and learn correlation filter w by minimizing the following objective function:

$$\min_{w} \sum_{j=1}^{p} \beta_j \left(\sum_{m,n} \left| \left\langle \phi\left(x^j_{m,n}\right), w^j \right\rangle - y^j(m,n) \right|^2 + \lambda \left\langle w^j, w^j \right\rangle \right), \tag{1}$$

where $w^j = \sum_{k,l} a(k,l)\phi(x^j_{k,l})$. We have the solution to (1) in the Fourier domain as:

$$\mathcal{A}^p = \frac{\sum_{j=1}^{p} \beta_j \mathcal{K}^j_x \odot \overline{\mathcal{Y}}}{\sum_{j=1}^{p} \beta_j \mathcal{K}^j_x \odot \left(\mathcal{K}^j_x + \lambda\right)}, \tag{2}$$

where $\mathcal{K}^j_x = \mathscr{F}\left\{k^j_x\right\}$ and $k^j_x(m,n) = k(x^j_{m,n}, x^j)$. We perform a grid search and set the weight $\beta_j = 0.01$ and the update rate $\lambda = 10^{-4}$ for the best accuracy. We restore the parameter $\{\mathcal{K}^j_x\}$, $j = 1, 2, \dots, p-1$, to update the correlation filter in frame j.

Note that such an update scheme is not applicable in practice as it requires a linearly increasing computation and memory storage over the increase of frame number p. The average tracking speed is 2.5 frames per second (fps) vs. 20.8 fps (ours) on the OTB2013 dataset. However, Figure 1 shows that this update scheme does not improve performance. The average distance precision is 83.5% vs. 84.8% (ours), and the average overlap success is 62.0% vs. 62.8% (ours).

C. Ma
The University of Adelaide, South Australia, 5000, Australia. E-mail: c.ma@adelaide.edu.au

J.-B. Huang
Virginia Tech, VA, 24060. E-mail: jbhuang@vt.edu

X. Yang
Shanghai Jiao Tong University, Shanghai, 200240, P. R. China. E-mail: xkyang@sjtu.edu.cn

M.-H. Yang
University of California at Merced, CA, 95344. E-mail: mhyang@ucmerced.edu

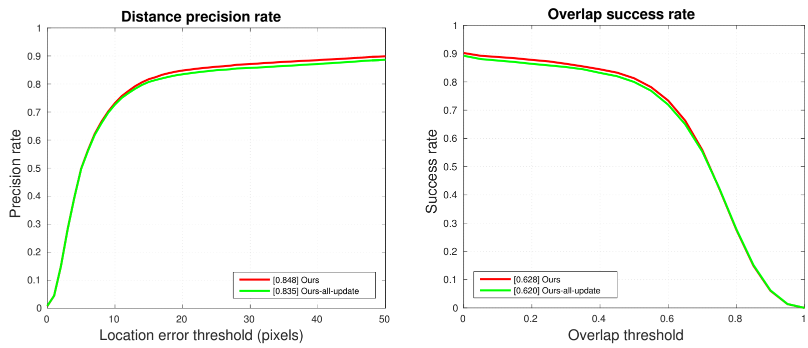

Fig. 1 Performance of different update schemes on the OTB2013 dataset [1] under one pass evaluation (OPE). Considering all the tracked results (ours-all-update) to update the translation filter does not improve tracking performance. The legend of precision plots shows the distance precision scores at 20 pixels. The legend of success plots contains the overlap success scores with the area under the curve (AUC).

We spatially shift the ground truth bounding boxes with eight directions (Figure 2) and rescale the ground truth bounding boxes with scaling factors 0.8, 0.9, 1.1 and 1.2. Figure 3 shows that slightly enlarge the ground truth bounding boxes (with scaling factor 1.1) does not significantly affect the tracking performance.

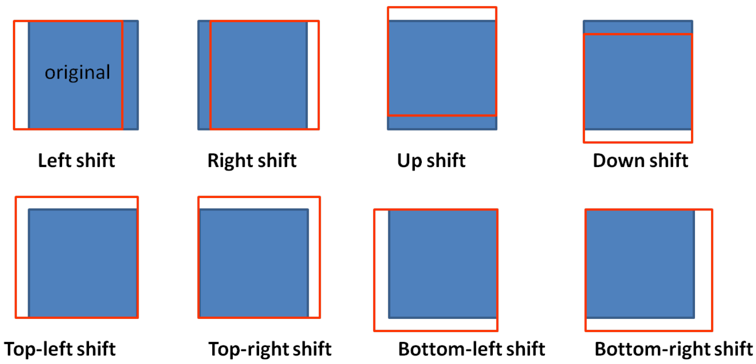

Fig. 2 Spatial shifts. The amount of shift is 10% of width or height of the ground-truth bounding box.

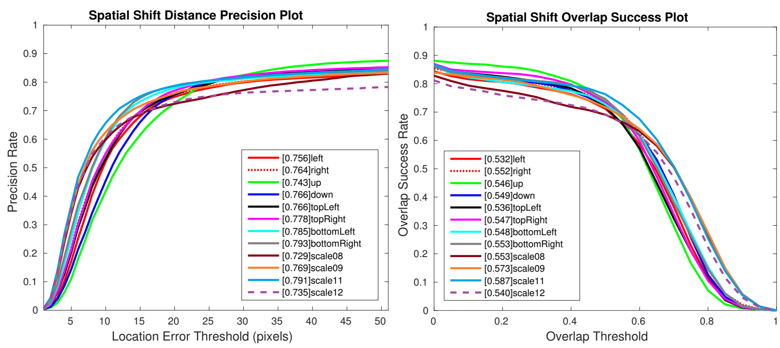

Fig. 3 Tracking performance with spatially shifted ground truth bounding boxes on the OTB2013 dataset [1] under one pass evaluation (OPE).

References

1. Wu, Y., Lim, J., Yang, M.H.: Online object tracking: A benchmark. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2013)