HVC-Net: Unifying Homography, Visibility, and Confidence Learning for Planar Object Tracking: Supplementary Material

Haoxian Zhang*[1][0000−0001−7078−868X], Yonggen Ling*†[2][0000−0001−8294−6286]

[1] Tencent AI Lab, China
leohxzhang@tencent.com
[2] Tencent Robotics X, China
rolandling@tencent.com

In this supplementary, success plots on the POT dataset [2] between our approach and state-of-the-art methods are shown in Fig. 1. Additional analysis and discussions that could not be fitted into the main paper due of lack of space are also provided. We recommend watching the supplementary video for performance details.

1 Additional analysis of the network

1.1 Confidence estimation with cost volume from different layers

We observe that for a good homography estimate: 1) distributions of constructed cost volumes have minimums with certain patterns and 2) costs in volumes are smaller as scales get larger. The confidence module is developed based on these two observations. If only a subset of cost volumes is used, the predicted confidence may not be reliable.

1.2 Why three pyramid layers are used?

Empirically, more layers (>3) do not lead to noticeable performance improvements. Main video streams are recorded at 30fps. We observe that, in most situations, object movements in the pixel displacement domain are less than 3 times its size in 1 second. That is, object movements are less than 1/10 of its size in consecutive time frames. At the smallest resolution of our pyramid (3 levels), its size is 30 30, and cost volumes are constructed with pixel displacement range [-4, 4]. We can handle object movements less than 4/30 of its size, which are larger than those previously mentioned (i.e. 1/10).

2 Additional experiment discussions

2.1 Performance discussions of different submethod in Fig.8

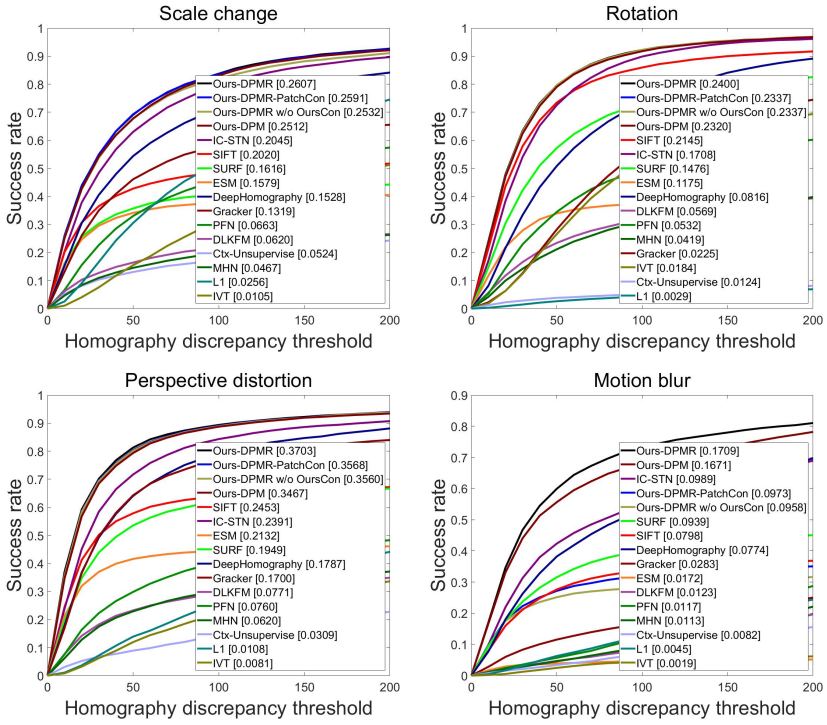
The main difference between submethods in Fig.8 is on the used confidence prediction module. Tracking is relatively easy in the scale-change and rotation conditions, since wrapped objects are with less appearance variations and

* Equal Contribution listed alphabetically. † Corresponding author.

perspective distortions. Ours-DPMR-PatchCon that regresses confidence in the pixel intensity domain is thus effective. However, in the perspective-distortion, motion blur, occlusion, and out-of-view conditions, tracking is hard due to the strong appearance variations, similar object disturbances and perspective distortions. PatchCon can not accurately predict the confidence. The model with this module (Ours-DPMR-PatchCon) sometimes performs worse than the model without it (Ours-DPMR w/o OursCon). Conversely, confidence predicted from our method (Ours-DPMR) is more reliable since we have two cascaded steps (i.e. feature extraction in the intensity domain and cost volume construction in the pixel displacement domain) that explicitly handle mentioned difficulties.

2.2 Why HVC-Net not perform well in 'Cereal' in Table.4

The ground truth of the TMT dataset [34] is labeled based on the results of three methods. The ESM algorithm is one of these three methods. Thus, ESM performs very well on a subset of sequences. Our results on the 'Cereal' sequence is not bad (rank 3). We had checked this sequence visually and found that our qualitative results were very similar to those of ESM or IVT. In addition, our method consistently performs well on all sequences, and thus are more robust than the others.



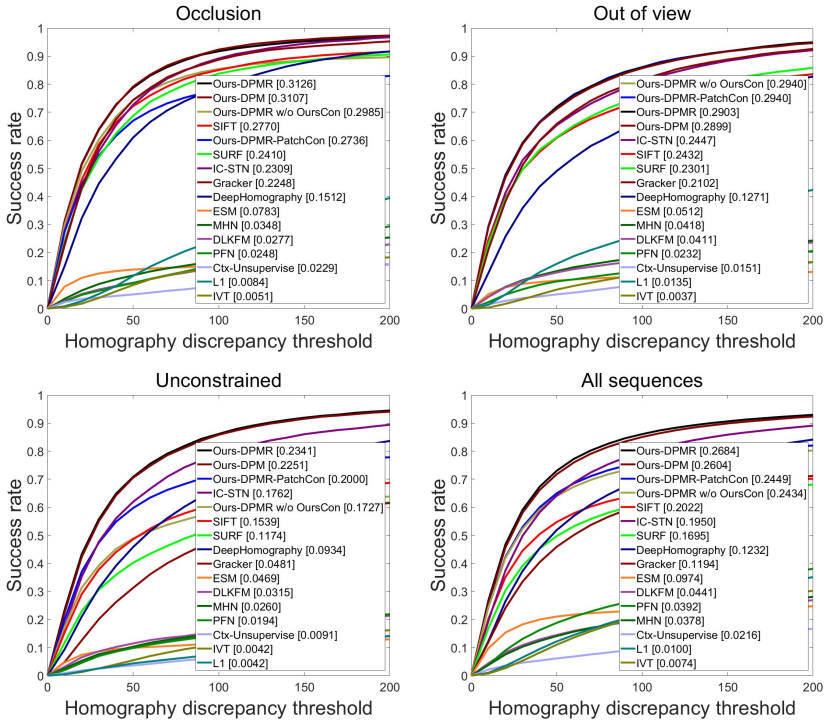


Fig. 1. The comparison of different approaches shown in success plots on the POT dataset [2]. Curves with larger areas are better. The HD at threshold = 10 [1] is illustrated within brackets. Zoom-in is recommended. Video comparisons are in the supplementary material.

References

1. Hare, S., Saffari, A., Torr, P.H.: Efficient online structured output learning for keypoint-based object tracking. In: Proc. of the IEEE Intl. Conf. on Comput. Vis. and Pattern Recognition (2012)
2. Liang, P., Wu, Y., Lu, H., Wang, L., Liao, C., Ling, H.: Planar object tracking in the wild: A benchmark. In: Proc. of the IEEE Intl. Conf. on Robot. and Autom. (2017)