Supplementary Material: Relative Pose from Deep Learned Depth and a Single Affine Correspondence*

Ivan Eichhardt[1,3] and Daniel Barath[1,2]

[1] Machine Perception Research Laboratory, SZTAKI, Budapest, Hungary
[2] VRG, Faculty of Electrical Engineering, Czech Technical University in Prague
[3] Faculty of Informatics, University of Debrecen, Debrecen, Hungary
{eichhardt.ivan,barath.daniel}@sztaki.mta.hu

Validation of 1AC+D in the 1-point RANSAC scheme

The proposed 1AC+D solver estimates the camera pose using a single correspondence and, thus, it can be applied in the 1-point RANSAC scheme. In our proposal, it is also combined with the local optimization of GC-RANSAC [1]. This supplementary material contains a brief evaluation of what inlier/outlier ratios are expected when using deep-learned depth prior [2], combined with affine correspondences [3].

Applicability of local optimization. For this validation, it is interesting to see how many of the obtained correspondences lead to a sufficient number of inliers, when the proposed 1AC+D method is applied in the 1-point RANSAC scheme. We consider an inlier number sufficient if the five-point algorithm, considering only the point coordinates from the correspondences, is applicable. When enough inliers are obtained, local optimization can further refine the inlier set while proposing an improved estimate of the camera pose. Therefore, if at least 6 inlier correspondences are found, the local optimization is applicable.

We took all correspondences of all image pairs from two scenes – *i.e.* 'Alamo' and 'Madrid Metropolis' – of the 1DSfM [4] dataset. To each image pair, we applied the proposed 1AC+D solver on each detected correspondence to determine the relative pose and measured the percentage of its inliers. We took the average over all correspondences of all image pairs and found that in scene 'Madrid Metripolis', $\sim$25.1 % of the correspondences lead to an inlier set of at least 6 elements, while in 'Alamo' $\sim$31.3 % of the correspondences lead to an inlier set of at least 6 elements.In these cases local optimization is applicable to polish the pose parameters on a larger-than-minimal inlier set.

Applicability of 1-point RANSAC. The 1-point RANSAC scheme can handle particularly low number of inliers. Thus, to see the applicability of this scheme, we measured what percentage of correspondences lead to accurate pose estimation, when applying 1AC+D on them.

* Source code available here: https://github.com/eivan/one-ac-pose

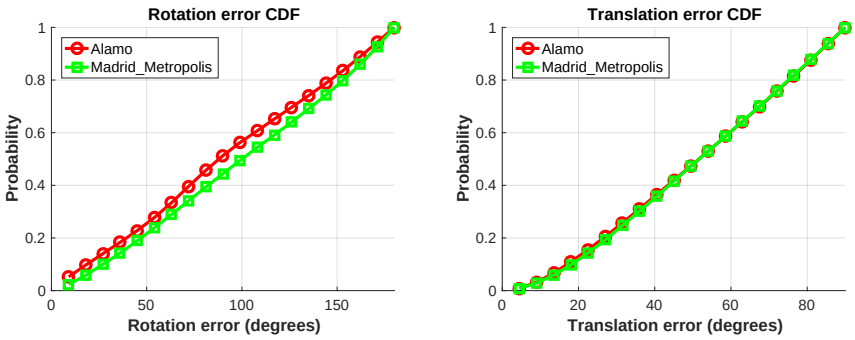

Fig. 1. The CDFs of rotation and translation errors obtained using 1AC+D on all correspondences of all image pairs from two scenes of the 1DSfM [4] dataset. Being more accurate is interpreted as a curve close to the top-left corner.

Fig. 1 shows the cumulative distribution functions (CDFs) of rotation and translation estimation errors for the 'Alamo' and 'Madrid Metropolis' scenes of the 1DSfM [4] dataset. It is visible from the CDFs, that a fairly low percentage of correspondences lead to accurate pose estimates. In fact, the percentage of correspondences that simultaneously yield low rotation and translation errors (20° and 15°, respectively) is even lower, about 2.7 % and 4 % for the two scenes.

Due to the fact that the proposed 1AC+D solver require only a single correspondence, such a low inlier ratio can still be handled and the accurate pose obtained. Given a confidence of 0.99 and an inlier ratio of 4 %, 1-point RANSAC would only need to perform 112 iterations. That is equivalent to only about 14 896 FLOPs when using 1AC+D, not considering the few additional local optimization steps. Consequently, even though the bad quality of the relative depth, the proposed solver is able to recover the sought pose parameters. For more details on the theoretical number of RANSAC iterations and the performance of our proposed method, the Reader is referred to Fig. 2 and Table 1 of our paper.

Conclusion

This brief evaluation suggests that, given the noisy correspondences obtained using the deep-learned depth priors and extracted affine features, 1AC+D produces a fairly low percentage of true inliers. Note that given higher-quality data, *e.g.* when using the metric depth from a depth camera, the above numbers are only expected to improve. Even though, the proposed 1-point RANSAC scheme easily handles such a high outlier ratio, resulting in low number of iterations and accurate results.

References

1. Barath, D., Matas, J.: Graph-cut RANSAC. In: Computer Vision and Pattern Recognition. pp. 6733–6741 (2018)
2. Li, Z., Snavely, N.: MegaDepth: Learning single-view depth prediction from internet photos. In: Computer Vision and Pattern Recognition (June 2018)
3. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/ (2008)
4. Wilson, K., Snavely, N.: Robust Global Translations with 1DSfM. In: Proc. European Conf. on Computer Vision. pp. 61–75 (2014)