Master's thesis review

Image Matching for Dynamic Scenes

by **Bc. Filip Šrajer**

The thesis of Filip Šrajer describes a 3D reconstruction pipeline that processes images from an unorganized set of cameras and delivers camera poses and reconstructed 3D points. The algorithm is based on the assumption that image correspondences can be explained in piecewise manner by a set of homographies with sufficient low-level correspondence support. The support need not be compact. These are not unrealistic assumptions in static scenes of man-made world that contain many planar surfaces. The benefit of this approach is that the model can explain non-rigid scenes as well, as long as the individual motions are piecewise rigid. That is an interesting problem that has been and still is studied in computer vision.

The core of the approach follows some ideas of the recent work [81]. The support for the homographies is 'grown' from seed correspondences that initialize affinities, by means of fitting progressively more general homographies to the inliers of the initial homography. Filip Šrajer considers a modification that grows supports for multiple such homographies simultaneously. The procedure is heuristic, essentially following a greedy algorithm, but the heuristics seem to be well-chosen, motivated by the call for computational speed. Non-optimal steps are amended by local optimizations that use an optimal (non-heuristic) model.

Grown homographies are then aggregated to groups that are consistent with an epipolar geometry. A pair of homographies over-constrains epipolar geometry. Filip Šrajer verifies that a suitable composition of two homographies compatible with an epipolar geometry is a planar homology. I believe this is known, including the geometric interpretation of the homology. What is interesting here is that one can construct a numerical score that measures the divergence of a homography from a homology. That can be used as a test if two homographies are bound by the same epipolar geometry. The proposed score seems quite heuristic, so it would be interesting to see if one could construct a metric for the distance from a homology, for instance one inspired by [1].

The grouping has several roles: (1) it allows non-planar rigid motions, (2) it serves as a filter that provides additional constraints for inlier detection, and (3) it identifies epipolar geometries corresponding to several rigid motions in the scene. The last role cannot distinguish individual motions, however, which, I believe, is nicely demonstrated by the example discussed in Fig. 7.1 (cf. also Fig. 6.2a, which shows three homography groups): When the epipolar geometries are induced by (nearly) parallel movements of rigid bodies their essential matrix is (nearly) identical.

Given the homography groups it then possible to reconstruct all their inliers in a modification of a structure-from-motion pipeline. This is described in Sec. 4.2 of the thesis. Filip Šrajer has implemented an original pipeline that builds on some existing libraries for linear algebra, image description, matching and optimization. The implementation is summarized in sufficient detail in Chap. 5.

Chap. 6 then presents a well chosen set of experiments that show that the approach performs comparably to the state of the art but exceeds it in processing speed.

The thesis is written in excellent English, contains a very small number of typos and only small inaccuracies in some expositions. The author has a good knowledge of the state of the art methods and their implementations.

I would like to clarify a few points:

1. How does Alg. 2 guarantee that the same homography is not discovered repeatedly? Should not be the set of visited matches $\mathcal{M}_v$ initialized before the main loop begins, not within it?

2. The prescription for cluster merge in (4.57) on p. 28 seems to lack a quantifier.

3. I suppose that the ground truth for image pair F used in Sec. 6.1 was obtained manually in such a way that all correct correspondences were labeled, not only those up to Lowe ratio of 0.8. Otherwise the precision-recall curve in Fig. 6.1 would be meaningless.

In summary, the submitted thesis shows that the student is able to carry out good engineering work in the area of computer vision, including creative modification of existing methods. The thesis is of high standard and I recommend it for presentation and suggest the **A-grade (excellent)**.

References

[1] E. Begelfor and M. Werman. How to put probabilities on homographies. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1666–1670, 2005.

Doc. Dr. Techn. Ing. Radim Šára

Department of Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague
May 3, 2016