ROBUST KEY POINT MATCHING FOR DYNAMIC SCENES

Johannes Furch, Peter Eisert

{johannes.furch, peter.eisert}@hhi.fraunhofer.de – Fraunhofer HHI / Humboldt University, Berlin, Germany

Abstract

Reliable and well distributed correspondences between sparsely sampled photographic images of dynamic scenes are needed in many computer vision applications including image-based rendering and 3D reconstruction. We propose a method based on local affine models that deals with local motion and resolves ambiguities. For most images this results in a considerably larger set of matches and lower error count than classical matching approaches.

Keywords: SIFT, Delaunay, incremental matching, affine.

Fairly reliable matches can be obtained by applying a key point detector and descriptor like SIFT and assigning to each key point the nearest neighbor in a multidimensional descriptor space. To limit false positives due to ambiguities and non-existent correspondences, the distance ratio between the two nearest neighbors is thresholded and the match has to be found in both directions. While discarding many potentially correct matches, a considerable amount of wrong matches remains (figure 1b). This leads to inconsistencies in multi-image key point tracks, suggesting that multiple points in one image represent the same object point. Since global measures to eliminate the wrong matches on a track cannot easily be derived from the descriptor distances, common solution are to drop all of the inconsistent tracks, to only work on image pairs, or to use underlying global models like a fundamental matrix (figure 1c), or a 3D reconstruction to reject outliers. However, those models cannot handle local motion, considerably reduce the matching density, and cannot always be used to resolve ambiguities (figure 1d).

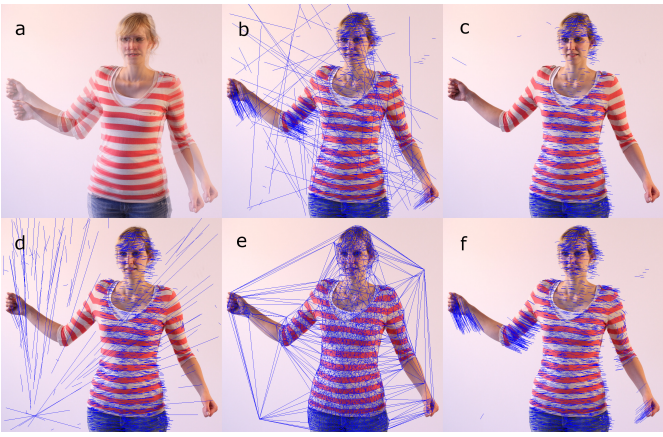

Figure 1: a) input images; b) basic matching (3100 matches); c) F-matrix filtering (2309 matches); d) F-matrix filtering and augmentation (2973 matches); e) Delaunay mesh; f) our method (3500 matches).

Our method is designed to deal with dynamic scenes while keeping or even improving the density of matches (figure 1f). On the basic set of SIFT matches we construct an incremental Delaunay mesh in one of the images and lay it onto the matching points. We then calculate a support value for each point using its transformation error with respect to the affine motion uniquely defined by a nearby triangle (figure 2a) [2].

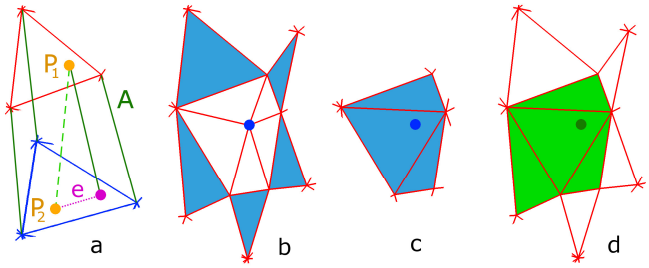

Figure 2: a) local affine model; b-d) neighbourhoods

Transformations derived from the outer triangles and the inner triangulation (figure 2b and c) of the star polygon have proven to give good support even on object edges. The initial set of matches is filtered by incrementally deleting the point with the lowest support until a threshold is reached. For augmentation, we construct a list of match candidates that include all the nearest neighbors up to the distance threshold used to reject points in basic SIFT matching. The ambiguity introduced by multiple match candidates for one key point is addressed by applying the local support measure. For fast localization of supporting triangles, a method already implemented for the incremental Delaunay algorithm [1] is used (figure 2d). Analogously to filtering, augmentation iteratively includes the match candidate with the highest support until a threshold is reached. In each step of filtering and augmentation only the support values of the changed region are updated in a mutable priority queue. The affine matrices calculated for each triangle can also be reused to derive a measure to identify remaining false matches on inconsistent key point tracks.

Acknowledgements:

This research was supported by the European Commission under contract FP7-288238 SCENE.

References

[1] T. Kao. Dynamic maintenance of Delaunay triangulations. *Technical Report, 9*, 32, 1991.

[2] H. Ogawa. Labeled point pattern matching by Delaunay triangulation and maximal cliques. *Pattern Recognition*, *19*(1), 35-40, 1986.