Dense reconstruction of underwater scenes from monocular sequences of images

Yuri Rzhanov and Han Hu

Dense Euclidean reconstruction of 3D scenes still is a challenging task despite significant progress in the last 10 years. Data acquired underwater by a single freely moving (handheld) camera is even more difficult to process, due to lack of reliable salient points, varying and often wide baseline between overlapping views, and inherent for water blurriness and wavelength-dependent light attenuation. Besides, most of the currently collected underwater imagery is acquired without consideration of basic photogrammetric requirements. As neither color nor brightness constancy holds for underwater imagery, the only reliable cue is a texture, which by definition has spatial extent and changes its spatial frequencies depending on the direction of view. This paper proposes a novel algorithm for quasi-dense close-range Euclidean reconstruction motivated by ideas developed for photogrammetric applications. Region-growing approach is initiated with a limited number of reliable matches extracted using SIFT or SURF detector. All matches are ranked according to their trustworthiness (reliability). New matches that satisfy epipolar constraint are searched for in immediate vicinity of previously accepted matches with highest ranking. Points in different views are assumed to match if the normalized cross-correlation score for patches (under some affine transform) around these points is sufficiently high. An optimal affine transformation (which is sometimes called weakly perspective, because it captures slight perspective deformations too) is found iteratively. If iterations do not converge quickly, the matching is considered to fail. Local smoothness constraint requires coefficients of the found affine transform to be close to that of the accepted neighbor. If difference between coefficients exceeds predefined threshold, current patch is likely to include a discontinuity, so the growth in this direction is stopped. After every acceptance of matches they are re-sorted again, so that region growing continues from the current most reliable match. Due to nature of affine matching the conjugate point is found with subpixel accuracy. Quasi-dense matches are triangulated for construction of a point cloud. This cloud is defined in the system of coordinates of the first camera, and the scale presumes that the baseline (distance between camera locations) is arbitrary. Overlap between three consecutive images (typically not exceeding 10 percent of the image area) is sufficient for calculation of a trifocal tensor. The point triples satisfying the trifocal tensor are subjected to bundle adjustment which defines a ratio of two sequential baselines. Bundle adjustment also slightly modifies rotation matrices found from pair-wise matching, but these corrections usually do not exceed 0.01 degrees (for rotation angles).

Sequential point clouds contain point triples determined during calculation of the trifocal tensor. These triples allow for determination of mutual pose and scale in a closed form. Although this information has been calculated previously, it is more convenient to re-calculate it and apply to the second point cloud to concatenate it with the first. The following (third) point cloud again uses known point triples to find rotation, translation and scale necessary for matching with corresponding points from second cloud (already transformed to the system of coordinates of the first cloud), etc..

The final step is a bundle adjustment applied to all matched points (each belonging to two of three sequential views) which corrects for errors accumulated in the cascading process described above. However, without path “closings” the final shape will be imperfect even after the bundle adjustment.

The algorithm has been applied to a set of still images of a WWII seaplane found in Canadian waters.