3-D Reconstruction in Piecewise Planar Environments

Olaf Kähler, Joachim Denzler

Chair for Computer Vision, Friedrich Schiller University of Jena
07743, Jena, Ernst-Abbe-Platz 2, Germany
{kaehler,denzler}@informatik.uni-jena.de

Our work focuses on the reconstruction of a geometric 3d model from a 2d image sequence. With known camera positions, this reduces to the problem of triangulation, and with a known scene, the camera positions can be calibrated. In our case however, both scene and camera positions are unknown, resulting in the so called *structure-from-motion* problem.

The typical solution to structure-from-motion uses two independent steps [1]. First, a 2d tracking algorithm [2] is applied to identify the motion of salient features in the image sequence, such as corner points or planar surfaces. Second, the recovered motions are used in a geometric reconstruction method [3], to infer the 3d motion of the camera and the scene structure. In our work, these two steps are solved in a combined formulation instead of two independent steps [4]. In the following, we first motivate the benefits of a combined formulation, then introduce the key ideas to implement it, and shortly present some results.

Tracking features in the images is often a demanding task [2]. The features may change their appearance due to illumination or reflections, sensor noise has to be addressed and also ambiguous cases arise e.g. in case of a checkerboard pattern. In a combined formulation, 3d information about the relation of features in the 3d world is implicitly fed back, and the 3d context of the features helps to disambiguate the correspondence problem. On the other hand, the typical least-squares reconstruction algorithms assume a Gaussian distribution of tracking errors [3]. This assumption does not necessarily hold for typical tracking algorithms, and is completely avoided in our combined formulation.

The key idea of our approach is to exploit homographies induced between the images of planar surfaces in the scene. These 2d projective transformations $\mathbf{H}^{(v,p)}$

Fig. 1. Outline of the information flow in our combined approach to tracking and reconstruction.

Dagstuhl Event Proceedings 08422
Computer Vision in Camera Networks for Analyzing Complex Dynamic Natural Scenes
http://drops.dagstuhl.de/opus/volltexte/2009/1867

provide a 2d mapping function between the images I_v and I_0. The homography matrices can also be expressed in terms of the geometric parameters $\mathbf{p}$ relevant to the 3d reconstruction problem. Minimizing the sum of squared errors over all views v, all planes p and all pixels $\mathbf{x}$ within the respective planar regions $\mathcal{R}_p$, the following overall error function ε arises [4]:

$$\varepsilon(\mathbf{p}) = \sum_v \sum_p \sum_{\mathbf{x} \in \mathcal{R}_p} \left(I_v(\pi(\mathbf{H}^{(v,p)}\mathbf{x})) - I_0(\mathbf{x}) \right)^2 \tag{1}$$

The error function ε is then minimized using non-linear optimization approaches like the Levenberg-Iteration.

It is straight forward to apply this estimation approach in an online framework for image sequences. An initial solution for the camera poses is then given by the assumption of small motions between successive views. Although stable convergence is achieved in most cases, a local minimum arises from an ambiguity in the scale-orthographic approximation of the perspective camera model. The second minimum can not be detected algebraically, but only by constructing the alternative solution and comparing the respective residual values of ε.

Overall reconstruction accuracies in the range of millimeters are achieved for a scene $1m$ in front of the camera. An efficient implementation with a slightly simplified cost function and parallel computations allow a performance of 5-10 fps. While the computational effort is higher than for other state-of-the-art methods, the reconstruction accuracy and stability is significantly improved, especially in case of poor image quality.

We were able to show experimentally and theoretically, that the combined approach to tracking and reconstruction allows for an improved tracking robustness and a higher reconstruction accuracy. The improvements are due to the feedback of 3d information to the tracking step and due to avoiding a model for tracking errors. In the sense of statistical estimation, the presented approach gives optimal results in case of Gaussian noise on the image intensities. All computations can be performed in real-time with 5-10 frames per second. Future works aim at transferring the close coupling between tracking and reconstruction to different features like salient image points.

References

1. Pollefeys, M., Gool, L.V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. International Journal of Computer Vision **59** (2004) 207–232
2. Shi, J., Tomasi, C.: Good features to track. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR94). (1994) 593–600
3. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment – a modern synthesis. In: Vision Algorithms: Theory and Practice. Volume 1883 of Lecture Notes in Computer Science., Springer-Verlag (2000) 298–372
4. Kähler, O., Denzler, J.: Robust real-time sfm in a combined formulation of tracking and reconstruction. In: Proceedings of the 13th International Fall Workshop Vision, Modelling, and Visualization (VMV 2008), Konstanz (2008) 283–292