Towards On-Line Intensity-Based Surface Recovery from Monocular Images

Oliver Ruepp[1]
ruepp@in.tum.de
Darius Burschka[1]
burschka@cs.tum.edu
Robert Bauernschmitt[2]
bauernschmitt@dhm.mhn.de

[1] Institut für Informatik VI
Technische Universität München
Boltzmannstraße 3
85748 Garching, Germany

[2] Deutsches Herzzentrum München
Lazarettstr. 36
80636 München, Germany

In this paper, we present a method that allows simultaneous surface reconstruction and camera localization from monocular images in static scenes. The novel aspect of the method is its independence from any explicit feature detection schemes. Instead, it uses method similar to intensity-based bundle adjustment. Thus, it is better suited for 3D reconstruction of weakly textured surfaces.

A number of methods with similar functionality have already been described [4, 6]. All of these methods, however, rely on some kind of feature detection schemes, such as such as SIFT [2] features, FAST corner detection [8], and so on.

The basic concept of our algorithm can be summarized as follows: In traditional bundle adjustment, coordinates of 3D points that are associated with feature points are recovered from a set of 2D feature position measurements. This approach will obviously work only if a feature detecting scheme can be used at all. In our case, we do not assume that robust feature extraction is possible, and thus we do not work with 2D positions, but with image intensities.

Originally, our method was inspired by a stereo disparity tracking method developed by Ramey [7]. The generalization that we are suggesting leads to an optimization problem that corresponds to intensity-based bundle-adjustment that is restricted to two frames. Thus, our solution shares some characteristics with typical bundle-adjustment algorithms [1, 9].

Our method establishes a depth map of the region of interest within a template image that has been chosen by the user. That depth map is then a function $S_{\mathbf{d}}(u,v)$ mapping a k-dimensional parameter vector $\mathbf{d}$ together with image coordinates $(u,v) \in \mathbb{R}^2$ to a depth value $\lambda \in \mathbb{R}$ at the specified coordinate. Given intrinsic camera parameters, this depth map can actually be interpreted as a 3D surface. Let

- $\mathbf{d_n}$ denote the k-dimensional vector of parameters of the model describing the depth map.
- $S_{\mathbf{d}}(u,v)$ denote a function of type $\mathbb{R}^k \times \mathbb{R}^2 \to \mathbb{R}$ that maps model parameters together with image pixel coordinates to 1D pixel depth values.
- $\mathbf{p}_n = (\mathbf{t}_n, \mathbf{q}_n)$ denote the extrinsic camera parameters corresponding to image n, consisting of translation vector $\mathbf{t}_n \in \mathbb{R}^3$ and rotation quaternion $\mathbf{q}_n \in \mathbb{R}^4$.
- $T(\mathbf{t},\mathbf{q},\mathbf{p}) : \mathbb{R}^3 \times \mathbb{R}^4 \times \mathbb{R}^3 \to \mathbb{R}^3$ is a transformation mapping 3D spatial coordinates $\mathbf{p}$ to 3D coordinates in the camera frame described by a translation vector $\mathbf{t}$ and a rotation quaternion $\mathbf{q}$.
- $\pi(\mathbf{p})$ be the projection of a 3D point $\mathbf{p}$ to 2D image coordinates, according to the internal camera calibration parameters of the camera used.
- $I_n(x,y)$ be the image function of image n, containing all pixel values. I_0 is hence the reference image function.
- $(u_1,v_1),\ldots,(u_m,v_m)$ denote the pixel coordinates of the m reference pixels, chosen from the ROI in the reference image.

The problem of determining surface parameters and camera position can then be stated as minimization problem for the following objective function $o(\mathbf{d},\mathbf{p}_n)$:

$$o(\mathbf{d},\mathbf{p}_n) = \sum_{i=1}^{m} (c(I_n(\pi(T(\mathbf{p}_t, r_{u_i,v_i}(S_{\mathbf{d}}(u_i,v_i))) - I_0(u_i,v_i)))^2 \qquad (1)$$

In other words, we are seeking a function that warps image coordinates from the reference image I_0 to the current image I_n such that intensity differences are minimized. The function c is a cost function that should be chosen to be robust against outliers, such as the Pseudo-Huber [1, p. 619] cost function.

It is clear that, to actually recover the model parameters from the scene, we need some method to minimize the cost function described above. Since we are dealing with a constrained problem, an adequate method for optimization is Sequential Quadratic Programming (SQP). For a more detailed description of the method, the reader is referred to [5].

The approach as described so far worked well in situations where camera movement is sufficiently smooth and no large pixel displacements occur between subsequent frames. However, problems occured when that was not the case. This was to be expected, since the algorithm operates on intensity values and will have trouble aligning with the correct values again if they are too far away.

We addressed this problem using Lucas-Kanade optical flow [3] to determine the camera position in a seperate optimization step before the intensity-based optimization is taking place:

$$o'(\mathbf{p}_n) = \sum_{i=1}^{m} c'((u_i',v_i') - \pi(T(\mathbf{p}_n, r_{u_i,v_i}(S_{\mathbf{d}}(u_i,v_i))))) \qquad (2)$$

Thus, we can summarize our strategy as using optic flow for bridging large gaps in camera movement, while intensity-based optimization refines the model and camera parameters and essentially prevents drifting away from the original point intensity values. This could easily occur if only optical flow based optimization would be performed.

[1] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[2] David G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, pages 1150–1157, 1999.

[3] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision (darpa). In *Proceedings of the 1981 DARPA Image Understanding Workshop*, pages 121–130, April 1981.

[4] Richard Newcombe and Andrew Davison. Live dense reconstruction with a single moving camera. In *2010 IEEE Conference on Computer Vision and Pattern Recognition CVPR'10*, 2010.

[5] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, August 2000. ISBN 0387987932.

[6] Q. Pan, G. Reitmayr, and T. Drummond. ProFORMA: Probabilistic Feature-based On-line Rapid Model Acquisition. In *Proc. 20th British Machine Vision Conference (BMVC)*, London, September 2009.

[7] Nicholas A. Ramey, Jason J. Corso, William W. Lau, Darius Burschka, and Gregory D. Hager. Real Time 3D Surface Tracking and Its Applications. In *Proceedings of Workshop on Real-time 3D Sensors and Their Use (at CVPR 2004)*, 2004.

[8] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In *European Conference on Computer Vision*, volume 1, pages 430–443, May 2006. doi: 10.1007/11744023_34.

[9] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *ICCV '99: Proceedings of the International Workshop on Vision Algorithms*, pages 298–372, London, UK, 2000. Springer-Verlag. ISBN 3-540-67973-1.