Plane-based Odometry using an RGB-D Camera

Carolina Raposo
http://arthronav.isr.uc.pt/~carolina
Miguel Lourenço
http://arthronav.isr.uc.pt/~mlourenco
Michel Antunes
http://www.isr.uc.pt/~michel
João P. Barreto
http://www.deec.uc.pt/~jpbar

Institute of Systems and Robotics,
Faculty of Sciences and Technology,
University of Coimbra,
3030 Coimbra, Portugal

Overview The advent of commodity RGB-D cameras led to intense research towards developing pipelines for estimating self-motion using simultaneously depth and vision cues [1, 2, 3]. State-of-the-art methods [2] are able to provide accurate motion estimations in real-time for the case of short displacement between consecutive frames. However, and since they use a differential registration approach, the performance degrades in the presence of dynamic scenes and/or wide baselines. Situations of wide baseline often arise because of high sensor velocity or increase in image resolution, with the frames taking more time to acquire and/or process.

This article describes a new odometry algorithm for RGB-D cameras that explores the fact of indoor environments being dominated by plane surfaces. The estimation of the relative pose between frames is cast as a plane registration problem that can be elegantly formulated in the dual projective 3D space as the successive estimation of a global rotation and projective scaling. The relative pose between image pairs is then refined by minimizing the photometric error in a set of planar patches.

While state-of-the-art methods use temporal information and motion priors to enforce smooth trajectories, we recover the camera motion by putting together the pairwise estimates without using any global optimization step. Several experiments with real data show that this naive approach is as accurate as the recent DVO algorithm [2] when the baseline is small, and provides substantially better motion estimates for the case of large baselines and/or dynamic foreground. These results suggest that plane features can be an advantageous alternative to point features for the purpose of visual odometry using RGB-D sensors.

Description of the pipeline Given two corresponding sets of 3 planes it is possible to compute the relative pose in a closed-form manner. The problem can be cast as finding the rotation R and translation $\mathbf{t}$ such that

$$\Pi_s^{(i)} \sim \begin{bmatrix} \mathsf{R} & \mathbf{0} \\ -\mathbf{t}^\mathsf{T}\mathsf{R} & 1 \end{bmatrix} \Pi_f^{(i)}, i = 1,2,3 \tag{1}$$

verifies, where $\Pi_f^{(i)}$ and $\Pi_s^{(i)}$ are planes in the first and second reference frames, respectively, in homogeneous representation. Knowing that points and planes are dual entities in 3D, equation (1) can be seen as a projective transformation in $\mathcal{P}^{3*}$ that maps points $\Pi_f^{(i)}$ into points $\Pi_s^{(i)}$. It can be shown that R and $\mathbf{t}$ can be determined separately.

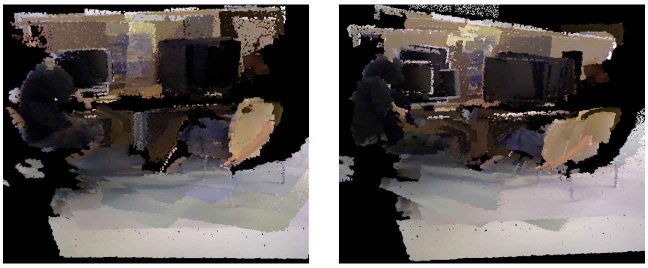
Planes are initially segmented and reconstructed using simultaneously depth and color information. The depth information is used in a RANSAC framework for suggesting plane hypotheses, followed by a plane merging procedure [4]. We reject planes corresponding to small segments, and re-run RANSAC for obtaining the final plane estimates.

Instead of explicitly matching planes between frames, we use an hierarchical scheme where corresponding sets with more planes are selected first. At least 3 non-parallel planes are necessary for computing R and $\mathbf{t}$. Situations where this is not verified are detected and, in case of ambiguity, the strictly necessary point matches are used to obtain an initial estimate.

Given the initial pose estimation, we use a photo-consistency framework to refine the camera motion and plane parameters. We optimize the motion and plane parameters by minimizing the cost function

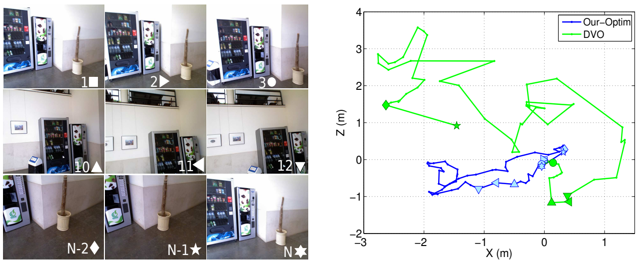
$$\varepsilon = \sum_{\mathbf{q}_f \in \mathcal{N}} \Big[\mathtt{I}_s(\mathbf{w}(\mathbf{q}_f\,;\mathbf{p}+\delta\mathbf{p})) - \mathtt{I}_f(\mathbf{q}_f) \Big]^2, \tag{2}$$

where $\mathbf{w}(\mathbf{q}_f\,;\mathbf{p})$ is a 2D warping function and $\mathbf{p}$ the warping parameter vector that encodes 3 parameters for camera rotation, 3 for translation, and 3 for the plane structure. For robustness against a noisy camera motion initialization, we use a coarse-to-fine registration framework.



(a) Our-Optim (b) DVO [2]

Figure 1: 3D reconstruction for a dynamic dataset. The figure shows the 3D reconstruction computed with 1(a) our method with optimization, and 1(b) the DVO method. The quality of the 3D reconstruction indicates that our algorithm is more robust than DVO for scenes with dynamic motion.



(a) Our-Optim (b) DVO [2]

Figure 2: Loop closing experiment with $N = 59$ images. 2(a) shows some images of the sequence and 2(b) shows the estimated trajectories. Our algorithm enables to keep a reliable trajectory estimation, with a consistent smooth transition between frames. The DVO method diverges after the first couple of frames, providing an erroneous trajectory.

Conclusion We conclude that plane-based registration for relative camera motion estimation is advantageous with respect to point-based registration because: (i) plane-primitives have a more global character, which helps avoiding local minima issues, (ii) scenes are often dominated by large planes, allowing correspondences between wide-baseline frames, (iii) plane primitives are typically in the static background, improving odometry robustness to possible dynamic foreground, and (iv) the fact that the number of plane-features is much smaller than point-features, favors faster correspondence and scalability under increasing image resolution. We show that our algorithm outperforms [2] for large baselines, while keeping similar performance for small baseline sequences.

[1] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard. In *IEEE-ICRA*, 2012.

[2] C. Kerl, J. Sturm, and D. Cremers. Robust odometry estimation for rgb-d cameras. In *IEEE-ICRA*, 2013.

[3] F. Steinbruecker, J. Sturm, and D. Cremers. Real-time visual odometry from dense rgb-d images. In *IEEE-ICCV Workshop on Live Dense Reconstruction with Moving Cameras*, 2011.

[4] Camillo J. Taylor and Anthony Cowley. Parsing indoor scenes using rgb-d imagery. In *Robotics: Science and Systems*, 2012.