Dense, Auto-Calibrating Visual Odometry from a Downward-Looking Camera

Jacek Zienkiewicz
j.zienkiewicz12@imperial.ac.uk
Robert Lukierski
r.lukierski12@imperial.ac.uk
Andrew Davison
a.davison@imperial.ac.uk

Department of Computing
Imperial College London
London, UK

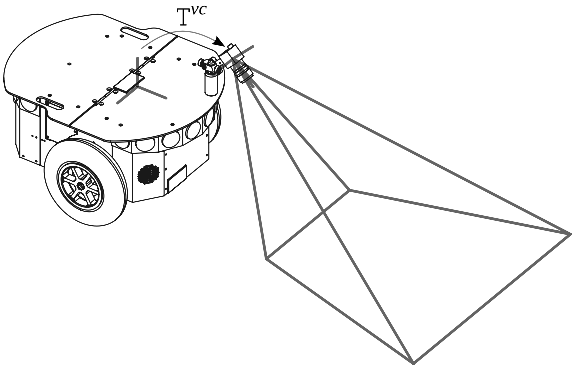

Figure 1: Robot and camera set-up; the camera view frustum intersects the ground plane.

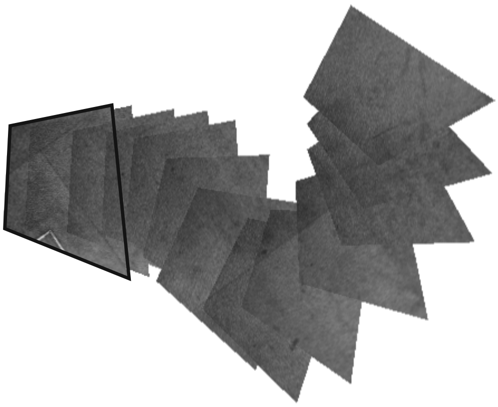

Figure 3: Keyframes registered into a local mosaic.

Computer vision is now widely considered a suitable technology for mobile robot navigation, but few practical mass-market solutions rely solely on cameras for outward-looking sensing. In this paper we present a technique whereby a single camera can be used as a high precision visual odometry sensor in a range of practical settings using simple, computationally efficient techniques. Taking advantage of the local planarity of common floor surfaces, we use real-time dense alignment of a 30Hz video stream as the camera looks down from a fast-moving robot. By using every image pixel rather than sparse features to contribute to a global motion estimate, we find that almost every surface has trackable texture when observed from close enough, and we are able to show robust and accurate real-time performance on a range of normal poorly-textured surfaces such as carpet, vinyl and wood.

The main part of our method uses a whole image alignment approach evolved from the iterative technique introduced by Lucas and Kanade [1]. We define an energy function F which measures the discrepancy between a reference image $\mathcal{I}^r$, and a live image $\mathcal{I}^l$ warped by a plane-induced homography $\mathtt{H}^{lr}$. The homography is parameterised by the incremental planar motion of a robot:

$$F(\mathbf{x}) = \frac{1}{2} \| \mathcal{I}^l(\pi(\mathtt{H}^{lr}(\mathbf{x})\mathbf{p})) - \mathcal{I}^r(\pi(\mathbf{p})) \|_2^2 \, . \tag{1}$$

In order to minimise $F(\mathbf{x})$ we perform an iterative optimisation based on Efficient Second-order Minimisation (ESM) [2, 3]. An example plot of the cost function with respect to the translational degrees of freedom of the robot can be seen in Fig. 2.

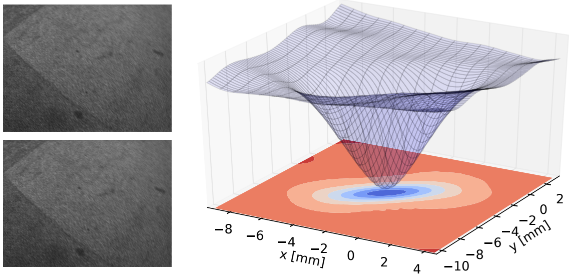

Figure 2: Typical dense tracking cost function.

Our key novelty, and crucial to the practicality of this approach, is rapid and automatic calibration for 6DoF camera extrinsics relative to the robot frame. This procedure has two main steps. First, based purely on the image sequence and the knowledge that the sequence is obtained from planar motion, we can determine the normal of the ground plane relative to the camera, which is equivalent to recovering its elevation and roll angles. An example of a dense local map created as part of the auto-calibration process is shown in Fig. 3. Second, we use relative motion estimates from an external source to calibrate for the remaining parameters, which describe the position of the camera relative to the robot: its yaw angle and metric 3DoF translational position.

Our real-time visual odometry is implemented almost entirely in parallel and achieve a mean frame-rate of about 270–300 FPS on a desktop GPU and about 60–65 FPS on a mid-range laptop GPU.

We experimentally evaluated our visual odometry approach against accurate odometry by making extended runs of the robot with different camera heights on a variety of different surfaces (Fig. 4). Our experiments show unbiased estimation of robot motion and a good performance of the algorithm most of the time. At high linear and angular velocities we see some tendency to underestimation of the velocity resulting mostly from the reduced inter-frame overlap and motion blur. Tracking failure usually only happens though when these effects are combined with disadvantageous other issues such as shadows cast by the robot.

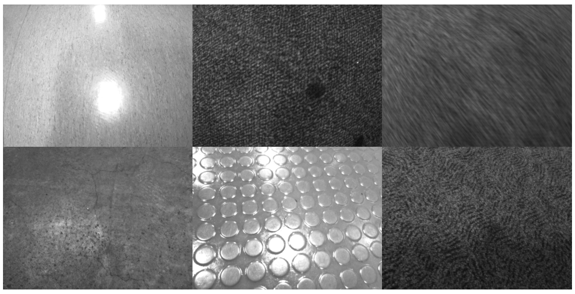

Figure 4: Snapshots of different surfaces used for evaluation.

[1] B. D. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 1981.

[2] E. Malis. Improving vision-based control using efficient second-order minimization techniques. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2004.

[3] C. Mei, S. Benhimane, E. Malis, and P. Rives. Efficient Homography-Based Tracking and 3-D Reconstruction for Single-Viewpoint Sensors. *IEEE Transactions on Robotics (T-RO)*, 24(6):1352–1364, 2008.