

SDICP: Semi-Dense Tracking based on Iterative Closest Points

Laurent Kneip^{1,2}
 laurent.kneip@anu.edu.au
 Zhou Yi^{1,2}
 yi.zhou@anu.edu.au
 Hongdong Li^{1,2,3}
 hongdong.li@anu.edu.au

¹)Research School of Engineering, Australian National University
²)ARC Centre of Excellence for Robotic Vision
³)NICTA Canberra Labs

The paper addresses the problem of *camera tracking*, which denotes the continuous image-based computation of a camera's position and orientation with respect to a reference frame. The method aims at regular cameras, which means that 3D-3D registration methods applicable to RGB-D cameras are not an option. The tracked frame contains only 2D information, thus requiring a solution to the *absolute pose* or *2D-3D registration* problem.

While traditional solutions to camera tracking [3] rely on sparse feature correspondences, the community has recently seen a number of direct photometric registration methods such as Newcombe et al. [8] and Engel et al. [1]. [1] is conceptually similar to [8], however gains computational efficiency by reducing the computation from dense to semi-dense regions that correspond to a thresholded edge-map of the image.

Photometric methods have the more general advantage of compensating for appearance variations caused by perspective view-point changes, whereas classical sparse methods often rely on static feature descriptors only (providing at most rotation and scale invariant properties [5, 6]). However, photometric registration techniques inherently suffer from the disability to overcome large disparities, where *large* sometimes means even just a couple of pixels [7]. Many photometric registration techniques therefore depend on pyramidal subsampling schemes in order to alleviate this problem.

The goal of the present paper is a novel 2D-3D registration paradigm for semi-dense depth maps that relies on the Iterative Closest Point (ICP) technique, and thus a reintroduction of geometric error minimization as a valid alternative for real-time monocular camera tracking in the case of semi-dense features. An example semi-dense depth map is indicated in Figure 1. In comparison to photometric registration techniques, our ICP technique has the conceptual advantage of requiring neither isotropic enlargement of the employed semi-dense regions, nor pyramidal subsampling. The work is in line with Feldmar et al. [2], Tomono [9], and Klein and Murray [4], which already attempt curve or edge registration in 2D using ICP.

Based on a hypothesized relative pose, the basic idea consists of *warping* a reference curve into the tracked image based on a prior 3D model or depth (in our case semi-dense) inside a reference frame. From a mathematical point of view, our idea may be formulated as follows. Let

$$\mathcal{S}^{\mathcal{F}_k} = \{\mathbf{s}_i^{\mathcal{F}_k}\} = \{(d_i^{\mathcal{F}_k})^{-1} \pi(\mathbf{p}_i^{\mathcal{F}_k})\}$$

denote the semi-dense depth map, where $\mathcal{P} = \{\mathbf{p}_i\}$ is the set of pixel locations that defines the semi-dense region in the reference frame \mathcal{F}_k , d_i the inverse depth of a pixel, and $\pi(\mathbf{p}_i) = \mathbf{f}_i$ is a known function that transforms points in the image plane into unit direction vectors located on the unit sphere around the camera center. The warped semi-dense region is easily obtained by

$$\mathcal{O}^{\mathcal{F}_{k+1}} = \{\mathbf{o}_i^{\mathcal{F}_{k+1}}\} = \{\pi^{-1}(\mathbf{R}^T(\mathbf{s}_i^{\mathcal{F}_k} - \mathbf{t}))\},$$

where \mathbf{t} and \mathbf{R} denote the sought position and orientation of the current frame. The final objective results in

$$\{\hat{\mathbf{t}}, \hat{\mathbf{R}}\} = \underset{\mathbf{t}, \mathbf{R}}{\operatorname{argmin}} \sum_{i=1}^N \left\{ \left\| n(\mathbf{o}_i^{\mathcal{F}_{k+1}}) - \mathbf{o}_i^{\mathcal{F}_{k+1}} \right\| \right\},$$

where $n(\mathbf{o}_i^{\mathcal{F}_{k+1}})$ is a function that returns the pixel from $\mathcal{P}^{\mathcal{F}_{k+1}}$ that is closest to $\mathbf{o}_i^{\mathcal{F}_{k+1}}$ under the Euclidean distance metric. We propose iterative minimization of the sum of all distances to the closest points (i.e. ICP) using gradient descent.

The core of our contribution extends ICP-based camera tracking to the case of noisy, outlier-affected semi-dense depth maps. We achieve large convergence basins and robustness with respect to outliers by prior

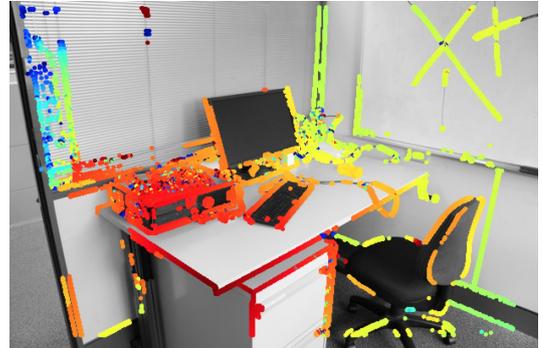


Figure 1: Example of a semi-dense depth map (hot colors=close, cold colors=far). The goal of this work is to track a monocular camera with respect to such a map. As can be observed, the depth map is typically affected by outliers, especially in new regions of the image (left part) and on self-similar background.

evaluation of sparse hypotheses and a robust evaluation of distances in the image plane, respectively. The computation is speeded up by employing a distance transformation and the bin sort technique.

Our results demonstrate successful application to real data, showcasing real-time tracking over large disparities, very large view point changes, and severely corrupted depth maps on a publically available dataset. In addition to an improvement over classical sparse methods, we also show how our approach leads to a larger convergence basin in comparison to a photometric error minimization strategy.

- [1] J. Engel, J. Sturm, and D. Cremers. Semi-dense visual odometry for a monocular camera. In *Proceedings of the International Conference on Computer Vision (ICCV)*, Sydney, Australia, 2013.
- [2] J. Feldmar, N. Ayache, and F. Betting. 3D-2D projective registration of free-form curves and surfaces. In *Proceedings of the International Conference on Computer Vision (ICCV)*, Cambridge, MA, 1995.
- [3] M.A. Fischler and R.C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [4] G. Klein and D. Murray. Full-3D edge tracking with a particle filter. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 1119–1128, Edinburgh, UK, 2006.
- [5] Stephan Leutenegger, Margarita Chli, and Roland Siegwart. BRISK: Binary Robust Invariant Scalable Keypoints. In *Proceedings of the International Conference on Computer Vision (ICCV)*, Barcelona, Spain, 2011.
- [6] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91–110, 2004.
- [7] 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)*, Vancouver, Canada, 1981.
- [8] R. Newcombe, S. Lovegrove, and A. Davison. DTAM: Dense Tracking and Mapping in Real-Time. In *Proceedings of the International Conference on Computer Vision (ICCV)*, Barcelona, Spain, 2011.
- [9] M. Tomono. Robust 3D SLAM based on an edge-point ICP algorithm. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, 2009.