Robust Scene Stitching in Large Scale Mobile Mapping

Filip Schouwenaars[1]
filip.schouwenaars@student.kuleuven.be
Radu Timofte[12]
radu.timofte@esat.kuleuven.be
Luc Van Gool[12]
luc.vangool@esat.kuleuven.be

[1] VISICS, ESAT-PSI / iMinds
KU Leuven
Leuven, Belgium

[2] Computer Vision Lab, D-ITET
ETH Zürich
Zürich, Switzerland

This paper presents a solution for the loop closure problem in an image-based mobile mapping context. A van equipped with stereo cameras collects recordings in an urban environment, simultaneously monitoring GPS information. Using Structure-from-Motion, the position of the van and the surroundings are retrieved. The determination of the translation and orientation of the van's position is recursive: a slight drift can gradually build up to flawed localizations. One can rely on the GPS information to perform adjustments, but its accuracy is not adequate to yield a model with high precision. Yet, visual loop-closing – recognizing that a location is revisited – may help mitigate the issue. The current system does not take into account possible reoccurrences of identical features in distant recordings. This paper adds such loop closure.

Local feature matching in two stages detects when a particular site is revisited, in order to enforce correspondences between images, that may have been taken with large time lapses in between. Our system relies on GPS but does not use odometric information. We extend the original image-to-image matching approach to a pose-to-pose matching approach, combining several images and achieving robust scene matching results. Parameter optimization is followed by extensive experiments. Our pipeline, which facilitates parallel execution, reaches matching rates higher than those reported for typical state-of-the-art algorithms. We also demonstrate robustness to odometric inconsistencies resulting from poor prior model build-up.

Loop closure is crucial for high-accuracy models. The current state-of-the-art in topological mapping are FABMAP [3] and CAT-SLAM [4], but limit themselves to a binary decision, *i.e*. whether or not the location was visited already. In the envisioned application however, it is desirable to have actual image point correspondences to facilitate bundle adjustment. An approach closer to this goal, by directly attempting to match local features among images, is described in [5]. The utilized epipolar constraint to cope with false positives is, however, not error-free. The devised approach method implements more robust error dismissal.

The approach consists of two major steps. First, the issue of detecting revisited sites over time is encountered by clustering GPS information and taking its inaccuracy into account. Next, in every *route* of such a cluster, two poses are selected that are expected to contain common elements, based on a Naive Bayesian matching framework with severely downscaled images. When a so-called *cross-route pose pair* is obtained, re-occurrences of the same physical points are tracked in the associated images. This problem is treated in two steps: *single pose matching* and *cross-route image matching*. The former finds matches and deducts corresponding 3D points using the SURF [1] detector and descriptor among views taken from the same van position. Since camera calibration is available, an epipolar consistency check is straightforward. Using the surviving matches, for every van pose a point cloud results. The latter step of cross-route image matching attempts to match the images from different van poses again using SURF. This practice establishes a link between the two earlier constructed point clouds. PROSAC [2], a prioritized RANSAC algorithm, is applied to robustly calculate the transformation between the two point clouds in a time-optimal way while pruning out false positives. Figure 1 provides an illustration.

This paper has two main contributions. First, our novel loop-closure technique does not depend on single image pairs for correspondences. Instead, a cloud is constructed around two van poses and these clouds are fitted together; an image-to-image approach is extended to a pose-to-pose approach. Second, our method is able to detect matches in challenging wide baseline conditions, were other systems tend to fail.

Since it concerns a system-specific application, a specialized dataset is devised that comprises a substantial amount of images from an urban environment. Separate datasets were used for parameter tuning and subsequent validation. Several time-decreasing techniques that impact on the SURF scheme and the inherent properties of the algorithm allow a balance between timing and performance. An adaptive approach of retrieving 3D information even increases performance while reducing computation significantly. The different steps can be carefully monitored through computation break-down experiments. Fetching the image data from memory constitutes a large part of execution time, followed by SURF detection and description.

Figure 1: Illustration of cross-route image matches after pruning of false positives by means of PROSAC. Note that these are not the only correspondences found; also for other image combinations matches are tracked.

#	#DNB	Descriptor	INC	Matching Rate	Comp. Time
13	100	USURF-64	✗	48/49 (98.0%)	1259 s

Table 1: Concise statement of results for the preferred setup on a 5531 poses dataset. *#DNB* stands for number of descriptors used for Naive Bayes Matching. *Descriptor* provides type and dimension of the descriptor. *INC* shows if the incremental approach is enabled.

Typically, a matching that exceeds 85% is achieved, *i.e*. when a cross-route pose pair is selected, the system is successful in finding correspondences between distant recordings. The number of false positives is zero. Information and results related to the preferred setup are summarized in Table 1.

Mock-up examples to mimic poor prior model build-up show robustness of the technique and prove the applicability in a wide context. Finally, the resulting correspondences between images that associate with distant recordings are embedded in the current systems. Further work is necessary to report on expected accuracy increase following from the promising results.

Our conclusion is that the novel pose-to-pose matching approach for robust scene stitching in a problem-specific context shows promising results, as proven by thorough experimentation. The performance is however not straightforwardly comparable to current state-of-the-art techniques due to the utilization of specialized datasets.

[1] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 2008.

[2] Ondrej Chum and Jiri Matas. Matching with PROSAC - Progressive Sample Consensus. In *CVPR*, 2005.

[3] Mark Cummins and Paul Newman. FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. *The International Journal of Robotics Research*, 2008.

[4] Will Maddern, Michael Milford, and Gordon Wyeth. CAT-SLAM: probabilistic localisation and mapping using a continuous appearance-based trajectory. *Int. J. Rob. Res.*, 2012.

[5] Stephen Se, David Lowe, and Jim Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *The International Journal of Robotics Research*, 2002.