ETH zürich

ETH Library

Centralized Multi-Robot Monocular SLAM

Conference Paper

Author(s):
Forster, Christian; Kneip, Laurent; Lynen, Simon; Siegwart, Roland

Publication date:
2012

Permanent link:
https://doi.org/10.3929/ethz-a-010034829

Rights / license:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

Centralized Multi-Robot Monocular SLAM

Christian Forster
Robotics and Perception Lab
University of Zürich, Switzerland
Email: forster@ifi.uzh.ch

Laurent Kneip, Simon Lynen and Roland Siegwart
Autonomous Systems Lab
ETH Zürich, Switzerland
Email: {laurent.kneip, simon.lynen, rsiegwart}@ethz.ch

Imagine a swarm of autonomous *Micro Aerial Vehicles* (MAVs) collaboratively searching for victims in a disaster area: a scenario where it is of utmost importance for the robots to effectively localize themselves in the unknown environment. It is the problem of incrementally creating a map and simultaneously localizing itself in it - well known as *Simultaneous Localization and Mapping* (SLAM). The introduction of multiple robots to solve this task increases robustness through redundancy and speeds up the process through parallelism. A fundamental question is of how and at which granularity map-information is shared amongst the robots such that the robots can mutually localize themselves and plan their path for optimal collaboration. One straight forward solution to this problem is to introduce a centralized base-station where all MAVs stream their information to and which thus is able to optimally orchestrate all MAVs. In this paper, we present the first step towards this solution: the *Collaborative Structure from Motion* (**CSfM**) system creates a three dimensional map of the environment in real-time from the input of multiple non-connected cameras and simultaneously localizes them in it.

Using a camera as the only exteroceptive sensor of a MAV is appealing due to its low cost and size. Hence, *Monocular SLAM* algorithms [8], [9], [14] have previously been applied for controlling MAVs [3]. Weiss et al. [4] have further shown how to estimate the absolute scale by fusing the vision data with an IMU. However, very little work on multi-robot SLAM for unconstrained 6-DoF motion has been done by applying monocular vision only: the work in [11] goes into this direction and assumes perfect communication which allows to optimize the full multi-robot pose-graph. In [6] and [5] on the other hand, the robots create local maps and share them upon encounter.

In our scenario, mapping and ego-motion tracking of the MAVs are clearly separated: each MAV tracks its position using a monocular *Visual Odometry* (VO) algorithm and the base station creates a common map for all robots. The minimal structure-from-motion like VO is boosted in terms of robustness and efficiency by the inclusion of incremental relative rotation priors obtained from an IMU [2]. This results in less complicated geometric algorithms that are computing translation only (based on [1]) and further also increases robustness against less favorable feature distributions in the image plane. In our framework, the VO algorithms function as distributed preprocessors by streaming their key-frames and relative position estimates to the **CSfM** system on the base-station. By streaming only the extracted features in the key-frame instead of the whole image, the required bandwidth is kept low. The **CSfM** system on the base-station creates a map for each MAV and constantly refines it by using Bundle Adjustment [10]. The system's **PlaceRecognizer** module is based on the Bag-of-Words model [7], [12] and detects when a MAV observes an environment which previously has been mapped by itself or another MAV. Hence, in the first case, a *loop closure* is initiated and a *map merge* in the latter. Loop-closures are optimized by 7-DoF pose-graph relaxation according to [13]. Otherwise, if the map of two or more MAVs are merged, the relative positions of the MAVs to each other are obtained. The MAVs will then provide information to extend the same map. The key to real-time performance at this stage is the design of data-structures as well as processes which allows multiple threads to concurrently read and modify the same map. Specifically, the **CSfM** system assigns a thread for each MAV to process its received key-frames. The mapping pipeline in the **CSfM** system further estimates the scale-divergence between the VO input and the created map.

The system was implemented in C++ and successfully tested on datasets where it proved the capability for real-time performance. The datasets were recorded with a camera on a hexacopter in the ASL indoor flying room with ground-truth data provided by a Vicon motion capture system. The **CSfM** system was able to reconstruct the relative key-frame positions with an accuracy of 4 cm over a MAV trajectory of 40 meters. Please refer to www.cforster.ch/csfm for videos of the experiments.

I. Related Work by the Authors

[1] L. Kneip, D. Scaramuzza, and R. Siegwart, "A Novel Parametrization of the Perspective-Three-Point Problem for a Direct Computation of Absolute Camera Position and Orientation," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[2] L. Kneip, M. Chli, and R. Siegwart, "Robust Real-Time Visual Odometry with a Single Camera and an IMU," *British Machine Vision Conference (BMVC)*, 2011.

[3] M. Blösch, S. Weiss, D. Scaramuzza, and R. Siegwart, "Vision based MAV navigation in unknown and unstructured environments," *IEEE International Conference on Robotics and Automation (ICRA)*, 2010.

[4] S. Weiss, D. Scaramuzza, and R. Siegwart, "Monocular-SLAM-based navigation for autonomous micro helicopters in GPS-denied environments," *Journal of Field Robotics*, vol. 28, no. 6, pp. 854–874, 2011.

[5] A. Martinelli, F. Pont, and R. Siegwart, "Multi-Robot Localization Using Relative Observations," *IEEE International Conference on Robotics and Automation (ICRA)*, 2005.

II. OTHER REFERENCES

[6] M. Cognetti, P. Stegagno, A. Franchi, G. Oriolo, and H. H. Bülthoff. 3D Mutual Localization with Anonymous Bearing Measurements. *IEEE International Conference on Robotics and Automation (ICRA)*, 2012.

[7] Mark Cummins and Paul Newman. Appearance-Only SLAM at Large Scale with FAB-MAP 2.0. *The International Journal of Robotics Research*, pages 1–8, 2011.

[8] Andrew J Davison, Ian D Reid, Nicholas D Molton, and Olivier Stasse. MonoSLAM: Real-Time Single Camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052–67, June 2007.

[9] Georg Klein and David Murray. Parallel Tracking and Mapping for Small AR Workspaces. *IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 1–10, November 2007.

[10] Rainer Kümmerle, Giorgio Grisetti, and Kurt Konolige. g2o: A General Framework for Graph Optimization. *IEEE International Conference on Robotics and Automation (ICRA)*, 2011.

[11] J McDonald, Michael Kaess, C Cadena, José Neira, and John J Leonard. 6-DOF Multi-session Visual SLAM using Anchor Nodes. *European Conference on Mobile Robots*, pages 69–76, 2011.

[12] Josef Sivic and Andrew Zisserman. Video Google: a text retrieval approach to object matching in videos. *Proceedings Ninth IEEE International Conference on Computer Vision*, 2:1470–1477, 2003.

[13] Hauke Strasdat and Andrew J Davison. Scale Drift-Aware Large Scale Monocular SLAM. *Robotics: Science and Systems*, 2010.

[14] Hauke Strasdat, Andrew J Davison, J M M Montiel, and Kurt Konolige. Double Window Optimisation for Constant Time Visual SLAM. *Proceedings of the IEEE International Conference on Computer Vision*, 2011.

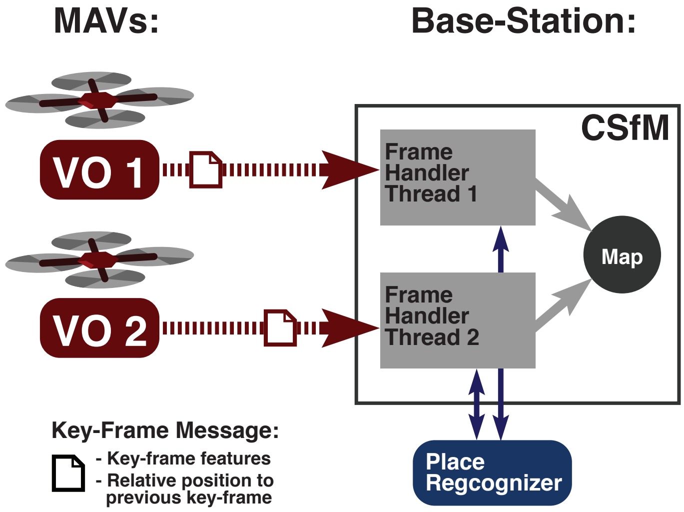

Fig. 1: The **CSfM** system creates a thread for each Visual Odometry (VO) input. Initially, each thread creates its own map. However, if the **Place Recognizer** detects an overlap between the two maps, they are merged. Both threads then update the same map simultaneously.