

Visual-Inertial SLAM for a Small Helicopter in Large Outdoor Environments

Markus W. Achtelik, Simon Lynen, Stephan Weiss, Laurent Kneip, Margarita Chli, Roland Siegwart

Abstract—In this video, we present our latest results towards fully autonomous flights with a small helicopter. Using a monocular camera as the only exteroceptive sensor, we fuse inertial measurements to achieve a self-calibrating power-on-and-go system, able to perform autonomous flights in previously unknown, large, outdoor spaces. Our framework achieves Simultaneous Localization And Mapping (SLAM) with previously unseen robustness in onboard aerial navigation for small platforms with natural restrictions on weight and computational power. We demonstrate successful operation in flights with altitude between 0.2-70 m, trajectories with 350 m length, as well as dynamic maneuvers with track speed of 2 m/s. All flights shown are performed autonomously using vision in the loop, with only high-level waypoints given as directions.

I. INTRODUCTION

The research towards autonomous Micro Aerial Vehicles (MAVs) has been increasingly active over the past few years, resulting to great progress in the field. We have seen impressive flights with aggressive maneuvers [1], but these rely on external tracking systems (e.g. Vicon) limiting their feasibility to lab-controlled environments. Ego-motion and scene estimation forms the backbone of any task involving autonomy of some degree. Given the strict limitations on weight and power consumption, the choice of sensors, processors and algorithms, impose great technical and scientific challenges. Moreover, while the use of GPS outdoors comes natural, sole reliance on GPS feeds is highly problematic; reception cannot be guaranteed (e.g. in narrow streets with high buildings), and when GPS signal is available, the coordinates provided can be very inaccurate (especially in altitude), compromising the accuracy of position estimates.

The good combination of characteristics offered by cameras (e.g. wealth of information captured, low weight), make them great candidates for MAV sensors in onboard real-time systems. While successful visual SLAM systems [2], [3] have been around for some time, their use onboard an MAV has been very challenging. Approaches like [4] rely on a laser range finder inside the control loop, while a camera is only used for loop-closure detection. However, the additional power and weight requirements of the laser sensor burdens the power consumption onboard the MAV,

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7) under grant agreements n. 231855 (sFly) and n.266470 (myCopter). Markus Achtelik, Simon Lynen, Stephan Weiss and Laurent Kneip are currently PhD students and Margarita Chli is a senior researcher at the ETH Zurich (email: {markus.achtelik, simon.lynen, stephan.weiss, laurent.kneip, margarita.chli}@mavt.ethz.ch). Roland Siegwart is a full professor at ETH Zurich and head of the Autonomous Systems Lab (email: r.siegwart@ieee.org).



Fig. 1: Fusing information from the downward-looking camera and an IMU, the MAV performs SLAM in a disaster area

limiting the capability for aggressive maneuvers and the flight duration and, as a consequence, autonomy.

A minimal sensor setup comprising a camera and an Inertial Measurement Unit (IMU) is very suitable for MAV navigation, but imposes great research challenges. Fusing inertial and visual cues can be very beneficial as they provide complementary information, aiding the robustness of the estimation processes. However, the complexity of the system grows vastly with each extra sensor; one has to take into account synchronization issues, inter-sensor calibration and intelligent management of all the additional data that becomes available. In this work, we use the methodologies developed in [5], [6], [7] to automatically calibrate the onboard sensors in flight, fuse sensor measurements to perform local, keyframe-based SLAM (without global optimization) and finally use this information to control the helicopter.

II. MAV: PLATFORM AND ONBOARD HARDWARE

Our setup consists of a micro helicopter equipped with an IMU, a monocular down-looking camera and an onboard computer. The helicopter is a prototype of the “FireFly” hexacopter from Ascending Technologies (Fig. 1). Compared to the “AscTec Pelican” it has an improved vibration damping for the sensors (camera, Flight Control Unit (FCU)) and can tolerate failure of one rotor. On the high level microcontroller of the FCU, we implemented a position controller based on nonlinear dynamic inversion, reference-model-based waypoint-following [7] and the state prediction of the Extended Kalman Filter (EKF) [5] – all executed at 1 kHz. The platform features a “MasterMind” embedded 1.86 GHz Core2Duo onboard computer from Ascending Technologies, and a MatrixVision “Bluefox” wide VGA camera. Our implementation is exclusively based on ROS middleware and has been made publicly available¹ for the position controller and helicopter interface, the camera interface, our visual odometry and the data fusion EKF module.

¹<http://www.asl.ethz.ch/research/software>

III. ONBOARD VISUAL-INERTIAL SLAM

State Estimation / Data Fusion: Visual measurements are fused with IMU readings (linear acceleration and angular velocity) in an EKF framework. The findings of [8], [9], [10] are applied to not only estimate the pose and velocity of the MAV, but also the sensor biases, the scale of the position estimates and the (inter-sensor) calibration between the IMU and the camera *in real-time*, rendering the system truly power-on-and-go. Since IMU readings are available at 1 kHz at the high-level processor (HLP) of the FCU, we compute the less expensive EKF state prediction on the HLP at the same rate. The covariance propagation and update, as well as the EKF measurement update stage run on the Core2Duo computer due to their complexity. This approach lets us handle the fast dynamics of the helicopter, while allowing the computation of an ideal Kalman gain based on the uncertainties of the state and the measurements (instead of using a fixed Kalman gain). Measurement delays are compensated for by keeping a buffer with states and associated covariances and applying the measurements according to their time-stamp. In [5] we showed that this approach is able to handle measurement delays up to 500 ms, noise with standard deviation up to 20 cm and slow update rates (as low as 1 Hz), while dynamic maneuvers are still possible.

Visual Localization: As one of the most modern, high-performing systems, we choose to tailor PTAM [3] to the general needs of a computationally limited MAV platform [6]. PTAM is a keyframe-based SLAM system, i.e. the map is defined as a set of keyframes together with their observed features. In order to keep computational complexity constant, we only keep the closest keyframes (by euclidean distance) in the map. During outdoor experiments, we experienced severe issues of self-similarity in the environment such as asphalt in urban areas and grass in rural areas, especially affecting the finest scale image features. Therefore, the finest-scale features are used only for tracking without adding them to the SLAM map. Moreover, an inverted index system on the keyframes structure reduces complexity of feature projection to linear with the number of visible keyframes rather than linear with the number of features in the map.

IV. EXPERIMENTAL RESULTS

The experiments shown in the video, take place in an outdoor disaster-training area and are performed without any prior knowledge of the scene (i.e. no pre-computed map or artificial landmarks), while all computation runs onboard the MAV. For all experiments, the MAV take-off is manual up to ≈ 4 m altitude. After initializing PTAM and a first estimation of the visual scale aided by GPS measurements (≈ 10 s only), the autonomous vision-based position controller is enabled.

The first experiment shows exploration of the area by repetitive flights in a 60×15 m rectangle with a track speed of 2 m/s. Even though SLAM runs by maintaining only the 20 closest keyframes in PTAM, the position error upon completion of the 2nd round was only 1.47 m or 0.4 %. Finally, after 357 m of traveled path, the flight ended due to an empty battery. Further experiments demonstrate robustness

to wind gusts simulated by pulling the MAV with a cord, resilience to challenging light conditions causing saturated brightness in images, and a dynamic trajectory flight in an inclined ellipse of 10×5 m at 2 m/s. Robustness to scale-changes is demonstrated by successful tracking during ascent to an altitude of 70 m and descent. Interestingly, landing is also demonstrated to be performed with autonomous vision-based control, following velocity commands from a joystick. As the MAV approaches the wooden pallet (serving as a landing platform), vision fails at a distance of ≈ 20 cm and landing completes purely based on integration of the IMU readings. We believe that this huge change of scale cannot be handled by a stereo vision setup and thus highlights the strength of our monocular approach, which is independent of scene depth. As the controller of the MAV still needs metric pose estimates, we show how to estimate scale in [8], [5].

V. CONCLUSION

This video demonstrates our visual-inertial methodology for SLAM onboard a MAV, enabling autonomous flights in unknown outdoor environments. The robustness of the system is successfully tested against disturbances, challenging light conditions, large scale-changes and high MAV dynamics enabling *local vision-based stability* of the MAV. Our system provides a basis for further research on high-level algorithms such as relocalization and obstacle avoidance for autonomous exploration, increasing the overall autonomy.

REFERENCES

- [1] D. Mellinger, N. Michael, and V. Kumar, "Trajectory generation and control for precise aggressive maneuvers with quadrotors," in *Proceedings of the International Symposium on Experimental Robotics*, 2010.
- [2] A. J. Davison, N. D. Molton, I. Reid, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 29, no. 6, pp. 1052–1067, 2007.
- [3] G. Klein and D. W. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR)*, 2007.
- [4] S. Shen, N. Michael, and V. Kumar, "Autonomous multi-floor indoor navigation with a computationally constrained MAV," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2011.
- [5] S. Weiss, M. W. Achtelik, M. Chli, and R. Siegwart, "Versatile Distributed Pose Estimation and Sensor Self-Calibration for an Autonomous MAV," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2012.
- [6] S. Weiss, M. W. Achtelik, S. Lynen, M. Chli, and R. Siegwart, "Real-time Onboard Visual-Inertial State Estimation and Self-Calibration of MAVs in Unknown Environments," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2012.
- [7] M. W. Achtelik, M. C. Achtelik, S. Weiss, and R. Siegwart, "Onboard IMU and Monocular Vision Based Control for MAVs in Unknown In- and Outdoor Environments," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2011.
- [8] S. Weiss and R. Siegwart, "Real-time metric state estimation for modular vision-inertial systems," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2011.
- [9] J. Kelly and G. S. Sukhatme, "Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration," *International Journal of Robotics Research (IJRR)*, vol. 30, no. 1, pp. 56–79, 2011.
- [10] F. Mirzaei and S. Roumeliotis, "A Kalman Filter-Based Algorithm for IMU-Camera Calibration: Observability Analysis and Performance Evaluation," *IEEE Transactions on Robotics and Automation*, vol. 24, no. 5, pp. 1143–1156, 2008.