20th International Conference on the Application of Computer
Science and Mathematics in Architecture and Civil Engineering
K. Gürlebeck and T. Lahmer (eds.)
Weimar, Germany, 20–22 July 2015

TOWARDS STEREO VISION- AND LASER SCANNER-BASED UAS POSE ESTIMATION

J. Kersten * and V. Rodehorst

** Bauhaus-University Weimar*
Bauhausstr. 11, 99423 Weimar
E-mail: jens.kersten@uni-weimar.de

Keywords: Unmanned Aircraft Systems (UAS), Pose Estimation, Odometry.

Abstract. *A central issue for the autonomous navigation of mobile robots is to map unknown environments while simultaneously estimating its position within this map. This chicken-egg-problem is known as simultaneous localization and mapping (SLAM).*
Asctec's quadrotor Pelican is a powerful and flexible research UAS (unmanned aircraft system) which enables the development of new real-time on-board algorithms for SLAM as well as autonomous navigation.
The relative UAS pose estimation for SLAM, usually based on low-cost sensors like inertial measurement units (IMU) and barometers, is known to be affected by high drift rates. In order to significantly reduce these effects, we incorporate additional independent pose estimation techniques using exteroceptive sensors. In this article we present first pose estimation results using a stereo camera setup as well as a laser range finder, individually. Even though these methods fail in few certain configurations we demonstrate their effectiveness and value for the reduction of IMU drift rates and give an outlook for further works towards SLAM.

1 INTRODUCTION

Light-weight unmanned aircraft systems (UAS) are small flying robots which can be used for many different civil-engineering applications like monitoring, assessment and inspection of built-up structures [1]. Besides a camera, standard out of the box UAS are often equipped with a GPS receiver and an inertial measurement unit (IMU) for navigation. Image acquisition is usually done by hand either during manual flights or automatically during pre-planned flights based on GPS waypoints. The latter approach requires prior knowledge about the on-site setting and is restricted to outdoor applications.

Simultaneous localization and mapping (SLAM) [2, 3, 4] enables autonomous exploration of unknown indoor and outdoor environments without or few manual interaction as well as preliminary information. SLAM approaches solve the problem of mapping unknown environments while simultaneously estimating the UAS pose within the map. Pose estimation can be tackled by the fusion of measurements from GPS, IMU and further sensors, for example a compass or a barometer, based on Kalman filtering, extended Kalman filtering [5] or particle filters [6]. Especially in GPS-denied environments pose estimation based on the aforementioned sensors is known to be affected by high drift rates due to dead reckoning. Hence, several systems which additionally incorporate laser range finders and cameras were proposed [7, 8, 9, 10]. These exteroceptive sensors are often used in conjunction since they rely on different measurement principles and therefore offer complementary failure modes.

As a first step in setting up an autonomous indoor UAS, we implemented and tested existing pose estimation approaches using a stereo camera setup as well as a laser range finder.

2 SYSTEM OVERVIEW

We use the off-the-shelf research quadrotor *Pelican* from *Ascending Technologies GmbH* since it is well suited to specific requirements in terms of size, payload, computational power and sensor integration. System implementation is here mainly reduced to method development and algorithm implementation. The stereo camera system consists of two forward looking *BlueFox-MLC200wC* cameras with a resolution of $752 \times 480\ pix$, a field of view of $100°$ and a $90\ Hz$ framerate. The *Hokuyo* laser range finder *UST-20LX* has a scanning range up to $20\ m \pm 4\ cm$, an angular resolution of $0.25°$, a horizontal scan coverage of $270°$ and a scanning frequency of $40\ Hz$. Since method optimization and system implementation is not finished yet, all experiments were conducted manually instead of using data from real test flights.

3 VISUAL ODOMETRY

The essential matrix $\boldsymbol{E}$ describes the relative orientation of two calibrated cameras and can be estimated based on corresponding image points using for example the well-known normalized 8-point algorithm [11]. A singular value decomposition (SVD) of $\boldsymbol{E}$ yields the relative 3D pose (6DOF: relative translation $\boldsymbol{T}$ and rotation $\boldsymbol{R}$) of the cameras. Since $\boldsymbol{E}$ is projectively invariant and independent from the chosen world coordinate system, $\boldsymbol{T}$ can only be estimated up to an unknown scale factor. When a fixed stereo camera setup is used, the scaled $\boldsymbol{T}$ can be computed without additional knowledge, for example ground control points. In a preliminary calibration step the camera intrinsics as well as the relative orientation between

the left and right camera, represented by the rotation matrix $\boldsymbol{R}_{r,l}$ and the translation vector $\boldsymbol{T}_{r,l}$, have to be estimated once. Since the relative pose estimation between consecutive stereo frames is based on the relative orientations $\boldsymbol{E}_l^{1,2}$ and $\boldsymbol{E}_r^{1,2}$, the scale factors μ and λ for the corresponding translations $\boldsymbol{T}_l^{1,2}$ and $\boldsymbol{T}_r^{1,2}$ of the left and right camera are in turn unknown. If temporally tracked points in two subsequent stereo frames are available, the factors can be computed according to the fixed camera pose of the stereo rig [12]: The 3D points $\boldsymbol{X}_{r,l}^1$, triangulated from the stereo setup in the first frame, define a reference scale, whereas the 3D points $\boldsymbol{X}_l^{1,2}$ and $\boldsymbol{X}_r^{1,2}$, triangulated from two subsequent frames of the left and right camera, respectively, can be scaled to $\boldsymbol{X}_{r,l}^1$. The authors of [12] propose to use the nearest five object points for a reliable computation of μ and λ:

$$\mu = 0.2 \cdot \sum_{i=1}^{5} \frac{\left\| \boldsymbol{X}_{l,r}^1 \right\|}{\left\| \boldsymbol{X}_l^{1,2} \right\|}, \lambda = 0.2 \cdot \sum_{i=1}^{5} \frac{\left\| \boldsymbol{X}_{l,r}^1 \right\|}{\left\| \boldsymbol{X}_r^{1,2} \right\|} \tag{1}$$

Whereas this stereo approach (in the following denoted as SVO) is based on 3D correspondences, Perspective-n-Points methods (PnP) use 2D-to-3D correspondences [13]. For the solution of the P3P problem the unknown distances $d_i = \left\| \boldsymbol{X}_i - \boldsymbol{C} \right\|$ between 3D points and the camera position $\boldsymbol{C}$, the inter-point distances $d_{ij} = \left\| \boldsymbol{X}_i - \boldsymbol{X}_j \right\|$ as well as the angle θ between each pair of rays between $\boldsymbol{C}$ and X can be used to form

$$d_{ij}^2 = d_i^2 + d_j^2 - 2 d_i d_j \cos\left(\theta_{ij}\right), \ \forall (i,j) \in \left\{ \{1,2\}, \{1,3\}, \{2,3\} \right\}. \tag{2}$$

This system can be solved by eliminating the unknowns of one equation using the remaining equations, which finally leads to a quartic polynomial.

In this work, the EPnP algorithm [14], which provides a non-linear solution within a linear time, is utilized, since it is known to provide good results compared to other iterative algorithms and is well suited for real-time systems.

Using the aforementioned concepts, visual odometry (VO) consists of two main steps: Feature detection and motion estimation. As stated in [8, 9] an overall estimation frequency of at least $10-15\ Hz$ should be achieved due to the fast flight dynamics. Inspired by [15], we use the Harris-based Shi-Tomasi corner detector [16] along with the Lucas-Canade method [17] for a fast matching of features in all four images of two subsequent stereo frames. The essential matrices for the moving left and right camera are computed using the normalized 8-point algorithm with RANSAC [12] outlier detection. An SVD of $\boldsymbol{E}_l^{1,2}$ and $\boldsymbol{E}_r^{1,2}$ yields four possible stereo camera configurations each, where the wanted parameters are those who lead to triangulated points in front of both cameras. μ and λ can then be computed according to (1).

The aforementioned essential matrices are also used for the PnP approach. Further inputs are the triangulated point cloud $\boldsymbol{X}_{l,r}^1$, the corresponding image coordinates in the second frame, the camera intrinsics as well as an optional initial guess of the pose obtained by SVO. Based on homogeneous transforms the current pose $P_{current}$ can then be computed by

$$P_{current} = P_0 \cdot \Delta P_1 \cdot \ldots \cdot \Delta P_{t-1} \cdot \Delta P_t = P_{previous} \cdot \Delta P_t, \text{ with } \Delta P = \begin{bmatrix} R^T & -R^T \cdot T \\ \vec{0} & 1 \end{bmatrix}, \ P_0 = \begin{bmatrix} I & \vec{0} \\ \vec{0} & 1 \end{bmatrix}, \tag{3}$$

where t represents number of frames.

The drawback of (3) is that each pose estimate solely depends on the previous pose as well as the previous and the current view. Therefore, in [9] the UAS motion is estimated across several frames using a bundle adjustment. However, in order to analyze the properties of the above described basic approach no further optimization is done here.

4 LASER ODOMETRY

For the laser scanner-based odometry the hector SLAM approach [18] is utilized, since it enables the processing of measurements acquired with a scan rate of up to 100 Hz. While a navigation filter is used for the 6DOF pose estimation, a 2D SLAM module provides a map as well as the positioning and heading of the UAV related to this map (see figure 1). Please note that for a full 3D pose estimation additional sensors, e.g. a downward looking sonar, is required. Since the used implementation[1] solely uses scan data as well as optional additional IMU data, only 2D pose estimations are possible here.

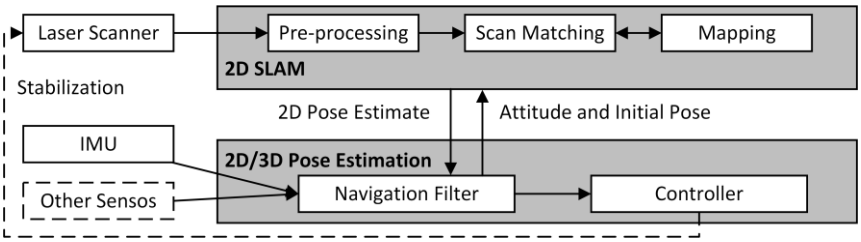

Figure 1: Workflow of hector SLAM [11]. Dashed lines represent optional interfaces.

In order to ensure a real-time pose estimation, SLAM and pose estimation are loosely coupled, i.e., update delays of the SLAM are allowed but don't affect the pose estimation. The motion of a rigid body can in general be described by a nonlinear equation system. Hence, for the pose estimation an extended Kalman filter (EKF) is used. Because IMU-based integrated velocities and positions are known to be affected by high drift rates, the 2D pose estimate from SLAM is used as additional information in EKF filtering.

In the pre-processing step of the SLAM module the scanned point cloud can optionally be filtered or down-sampled. The environment is represented by a 2D occupancy grid map and scan matching aligns laser scans with this map. The scan-based estimation of the current 2D pose is based on a Gauss-Newton approach, where a 3DOF transform of the scanned world coordinates that fits best to the current map is obtained. In order to improve this step, the pose estimate of the EKF is projected to the (x, y)-plane and is used to initialize the scan matching. For more detailed information the reader is kindly referred to [11].

5 EXPERIMENTS

In order to securely test the quality and reliability of the approaches, all experiments were conducted manually, i.e., the UAS equipped with the sensors was carried by hand along pre-defined paths. Since the paths are closed loops, the coordinate differences between the starting and end positions represent a quality measurement. Whereas the 2D/3D SVO and PnP trajectories are compared in the first experiment, the 2D trajectories obtained by hector SLAM

[1] http://wiki.ros.org/hector_slam (June 24, 2015)

with and without IMU fusion are compared. In figure 2 the corresponding results are visualized.

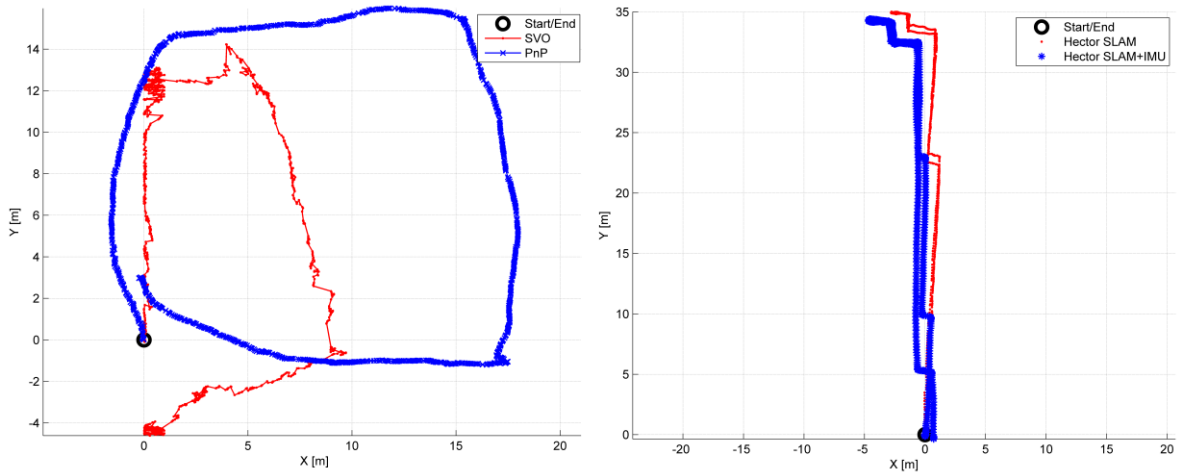

Figure 2: Trajectories of the SVO and PnP method (left) and the hector SLAM approach with and without IMU fusion (right).

For the VO experiment an outdoor path around a house with a size of approximately $19 \times 12\,m$ was chosen. As can be seen in figure 2 (left) SVO yielded very inaccurate results. Even though the rough moving directions are correct, the estimated path differs significantly from the expected result of a smooth rectangular path. Due to an unstable scale factor estimation according to (1), the resulting trajectory tends to be very jagged. Furthermore, the path is too short to circuit the building which implies that the factors tend to be underestimated. Nonetheless, the distances between the start- and endpoint are $4.13\,m$ (2D) and $6.51\,m$ (3D). Based on the PnP approach a much smoother and realistic trajectory could be obtained, whereas a slight underestimation of the relative movements can also be observed here. The start-end differences with $2.97\,m$ (2D) and $5.20\,m$ are slightly but not significantly smaller than using VO. The results confirm the observation that the estimation of the rotation is more stable and precise than the translation [13].

Both approaches solely use point correspondences from two consecutive stereo-frames. Disadvantageous configurations, for example very small or zero translations, therefore lead to errors which accumulate over time. The Lukas-Canade feature matching approach in turn fails for large translations which suggests that another matching method should be used here.

For the hector SLAM experiment a closed indoor trajectory along a straight corridor was chosen, since the laser range finder is sensitive to sunlight. The results solely using the mapping module of hector SLAM produced a smooth trajectory which is close to the true path. With a 2D start-end difference of $0.39\,m$, the estimated path intends to be a very good solution. Nonetheless, compared to the result with additionally incorporated IMU measurements, the red path is slightly curved, which indicates a drift caused by the scan matching. The start- to endpoint difference of $0.72\,m$ is surprisingly larger here.

6 CONCLUSION

In this work results of visual and laser scanner-based pose estimation techniques are presented. The PnP approach as well as the hector SLAM approach using additional IMU data

yielded the best results in our experiments. Nonetheless, the loop closure test and other findings demonstrate that odometry estimates can rapidly exceed drift tolerances defined by the dimensions and richness of detail of the indoor environment to explore.

A careful optimization of the single VO steps, especially the feature matching, will be part of the next works and these techniques will be implemented on our quadrotor. Even though the estimated poses can directly be used as additional input for Asctec's on-board autopilot, the application and implementation of other data fusion methods is planned.

REFERENCES

[1] N. Hallermann, G. Morgenthal and V. Rodehorst: Vision-based deformation monitoring of large scale structures using Unmanned Aerial Systems. 37th IABSE Symposium, Engineering for Progress, Nature and People, Madrid, Spain, 2014.

[2] M. Montemerlo: FastSLAM – a factored solution to the simultaneous localization and mapping problem with unknown data association, Ph.D. thesis, Robotics Institute, Carnegie Mellon University, 2003.

[3] H. Durrant-Whyte and T. Bailey: Simultaneous localization and mapping: part I. IEEE Robotics & Automation Magazine, **13-**2, 99-110, 2006.

[4] T. Bailey and H. Durrant-Whyte: Simultaneous localization and mapping: part II. IEEE Robotics & Automation Magazine, **13**-3, 108-117, 2006.

[5] S. S. Haykin. Kalman Filtering and Neural Networks. John Wiley & Sons, Inc., New York, NY, USA, 2001.

[6] B. Ristic, S. Arulampalam and N. J. Gordon. Beyond the Kalman filter: Particle filters for tracking applications. Artech house, 2004.

[7] R. Brockers, M. Hummenberger, S. Weiss and L. Matthies: Towards Autonomous Navigation of Miniature UAV. IEEE Conference on Computer Vision and Pattern Recognition, Workshops (CVPRW), 645-651, 2014.

[8] K. Schmid, T. Tomic, F. Ruess, H. Hirschmuller and M. Suppa: Stereo vision based indoor/outdoor navigation for flying robots. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 3955-3962, 2013.

[9] M. Achtelik, A. Bachrach, R. He, S. Prentice and R. Nicholas: Stereo vision and laser odometry for autonomous helicopters in GPS-denied indoor environments. G. R. Grant, D. W. Gage and C. M. Shoemaker eds. Unmanned Systems Technology XI. SPIE-The International Society for Optical Engineering, Orlando, FL, USA, 2009.

[10] S. Huh, D.H. Shim and J. Kim: Integrated navigation system using camera and gimbaled laser scanner for indoor and outdoor autonomous flight of UAVs. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 3158-3163, 2013.

[11] R. I. Hartley and A. Zisserman: Multiple view geometry in computer vision. Cambridge University Press, 2nd edition, 2004.

[12] V. Rodehorst, M. Heinrichs and O. Hellwich: Evaluation of Relative Pose Estimation Methods for Multi-Camera Setups. Proc. of the XXI Congress of the Int. Society for Photogrammetry and Remote Sensing (ISPRS), Int. Archives of Photogrammetry,

Remote Sensing and Spatial Information Science (IAPRS), **37**-B3b, 135-140, Peking, China, 2008.

[13] L. Quan and Z. Lan: Linear n-point camera pose determination. PAMI, **21**-8, 774-780, 1999.

[14] F. Moreno-Noguer, V. Lepetit and P. Fua: Accurate non-iterative o(n) solution to the pnp problem. In ICCV, Rio de Janeiro, Brazil, 2007.

[15] V. More, H. Kumar, S. Kaingade, P. Gaidhani and N. Gupta: Visual odometry using optic flow for Unmanned Aerial Vehicles. International Conference on Cognitive Computing and Information Processing (CCIP), 1-6, 2015.

[16] J. Shi and C. Tomasi: Good Features to Track. 9th IEEE Conference on Computer Vision and Pattern Recognition, Springer, 1994.

[17] B. D. Lucas and T. Kanade: An iterative image registration technique with an application to stereo vision. In Proceedings of the 7th international joint conference on Artificial intelligence (IJCAI'81), Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, **2**, 674-679, 1981.

[18] S. Kohlbrecher, O. von Stryk, J. Meyer and U. Klingauf: A flexible and scalable SLAM system with full 3D motion estimation. IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), 155-160, 2011.