ETH zürich

ETH Library

A UAV System for Inspection of Industrial Facilities

Conference Paper

Author(s):
Nikolic, Janosch; Burri, Michael; Rehder, Joern; Leutenegger, Stefan; Hürzeler, Christoph; Siegwart, Roland

Publication date:
2013

Permanent link:
https://doi.org/10.3929/ethz-a-010061810

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

Originally published in:
https://doi.org/10.1109/AERO.2013.6496959

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

A UAV System for Inspection of Industrial Facilities

Janosch Nikolic
Autonomous Systems Lab
ETH Zurich
janosch.nikolic@mavt.ethz.ch

Michael Burri
Autonomous Systems Lab
ETH Zurich
michael.burri@mavt.ethz.ch

Joern Rehder
Autonomous Systems Lab
ETH Zurich
joern.rehder@mavt.ethz.ch

Stefan Leutenegger
Autonomous Systems Lab
ETH Zurich
stefan.leutenegger@mavt.ethz.ch

Christoph Huerzeler
Autonomous Systems Lab
ETH Zurich
christoph.huerzeler@mavt.ethz.ch

Roland Siegwart
Autonomous Systems Lab
ETH Zurich
rsiegwart@ethz.ch

***Abstract*—This work presents a small-scale Unmanned Aerial System (UAS) capable of performing inspection tasks in enclosed industrial environments. Vehicles with such capabilities have the potential to reduce human involvement in hazardous tasks and can minimize facility outage periods. The results presented generalize to UAS exploration tasks in almost any GPS-denied indoor environment. The contribution of this work is two-fold. First, results from autonomous flights inside an industrial boiler of a coal-fired thermal power plant are presented. A lightweight, vision-aided inertial navigation system provides reliable state estimates under difficult environmental conditions typical of such sites. It relies solely on measurements from an on-board MEMS inertial measurement unit and a pair of cameras arranged in a classical stereo configuration. A model-predictive controller allows for efficient trajectory following and enables flight in close proximity with the boiler surface. As a second contribution, we highlight ongoing developments by displaying state estimation and structure recovery results acquired with an integrated visual-inertial sensor that will be employed on future aerial service robotic platforms. A tight integration in hardware facilitates spatial and temporal calibration of the different sensors and thus enables more accurate and robust ego-motion estimates. Comparison with ground truth obtained from a laser tracker shows that such a sensor can provide motion estimates with drift rates of a only few cm over the period of a typical flight.**

TABLE OF CONTENTS

1 INTRODUCTION 1
2 RELATED WORK 2
3 AERIAL SYSTEM DESIGN AND CONTROL 2
4 INTEGRATED SENSOR DESIGN AND STATE ESTIMATION 3
5 EXPERIMENTS AND RESULTS 4
6 CONCLUSION 6
ACKNOWLEDGMENTS 6
REFERENCES 7
BIOGRAPHY 8

1. INTRODUCTION

Industrial sites often contain areas and facilities that are difficult to access or hazardous to humans. The boiler unit of a coal-fired thermal power plant constitutes one example of that environment type. Fig. 1 shows such a unit, and visual inspection thereof will serve as an application example throughout this work. Whenever it is necessary to inspect such areas, tremendous precautions have to be taken in order to minimize the risks to human workers. Furthermore, scaffolding often needs to be installed in order to grant access to sections that are otherwise hard to reach. These measures result in significant downtime of the plant and are thus associated with vast economic costs.

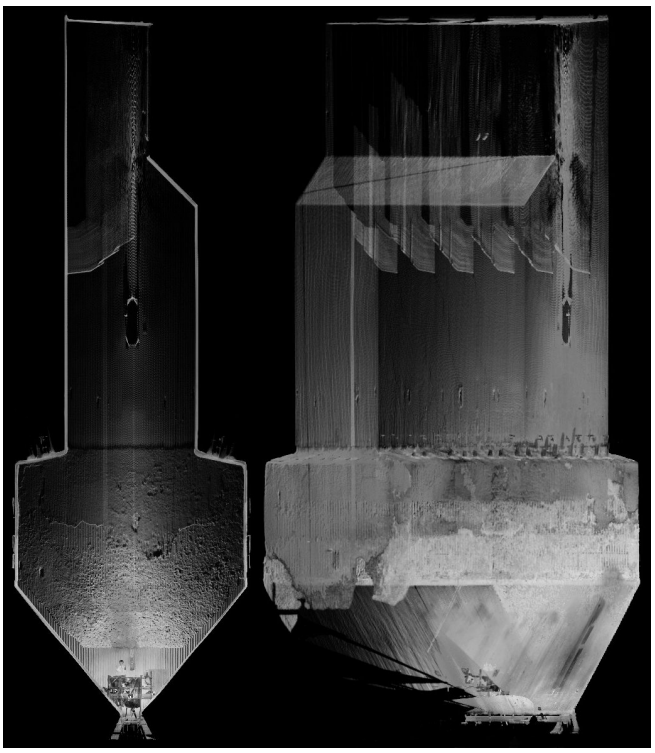


Figure 1. Laser scan of power plant boiler unit II (side view). With a height of almost 50m, this is a medium-sized unit. The flight tests were conducted in the lower part of the boiler.

Autonomous aerial systems have the potential to reduce human involvement in a variety of industrial inspection tasks and can in turn reduce outage periods. They neither require extensive safety measures nor supporting structures to access elevated or confined areas. In contrast to existing and very successfully employed robotic platforms such as e.g. magnetic crawlers [1], flying inspection vehicles can access virtually any area with ease and are not restricted to ferromagnetic surfaces.

However, the employment of flying systems poses a number of challenges such as very limited payload capabilities, and

978-1-4577-0557-1/12/ 26.00 ©2012 IEEE.
[1] IEEEAC Paper #xxxx, Version x, Updated dd/mm/yyyy.

hence limited computational resources. One of the main difficulties is reliable state or vehicle ego-motion estimation in absence of GPS signals, high-quality inertial measurement units (IMUs), and the like.

Section 3 describes the platform, its visual/inertial state estimation and the autopilot: a model predictive control scheme enables the vehicle to follow pre-programmed trajectories in close vicinity of the structure efficiently. By arranging these trajectories around an area of interest, a human operator can automatically acquire sets of images for visual inspection. In this respect, the autonomy of our system facilitates efficient inspection of large structures and enables remote inspection in areas, where occlusions or impaired radio reception had rendered it infeasible before.

Section 4 of this work presents a custom-designed visual-inertial sensor-head. The size of our aerial vehicle imposes tight constraints on weight and energy consumption onto the state perception module. At the same time, a high degree of robustness and accuracy is required to enable autonomous operation in poorly and repetitively textured areas. To address these requirements, we are in the course of developing an integrated visual/inertial sensor that allows for precise hardware synchronization of multiple different sensors and on-board estimation of motion and structure. Tight integration of cues from a pair of cameras arranged in a front-looking stereo configuration and a MEMS inertial measurement unit on a hardware level allows for accurate spatial and temporal calibration of the sensor unit and ultimately enables the required level of robustness and accuracy.

Section 5 presents results with respect to the described hardware and algorithms: the system was deployed in a shut-down boiler, in order to assess the on-board state estimation and control capabilities in relation to recorded ground truth motion. Furthermore the integrated sensor performance including a tightly coupled visual-inertial odometry approach is evaluated again with respect to ground truth structure and pose measurements.

Despite the fact that our flight and sensor tests were conducted inside an industrial boiler, we believe that the significance of our results extend to a much broader class of scenarios. With our system, we successfully demonstrate that path following, the predominant prerequisite for all indoor navigation tasks, is possible in almost any GPS-denied, unstructured environment providing a minimum of visible texture.

2. Related Work

Autonomous navigation of aerial vehicles in GPS-denied environments has recently gained attention in the robotics community. Using a laser range sensor, Shen et al. demonstrated indoor navigation of a quadrotor [2]. However, the requirements, imposed onto the system by a localization based on scan matching along with altitude measured by a deflected portion of the scan, are not met in most industrial environments. Approaches using an RGB-D sensor instead of a laser scanner [3] do not share the same constraints on the structure of the environment, but are only applicable to indoor operations. While this would suffice to conduct the experiment presented in this work, it would fail in the case of an outdoor inspection task. Weiss et al. [4] demonstrated visual/inertial enabled outdoor flights of a small aerial vehicle. In contrast to their approach based on a monocular camera, we employ a stereo setup; we believe that the additional information introduced by measurements from synchronized cameras displaced by a known transformation adds necessary robustness.

Most similar projects use either PID or LQR controllers to keep the UAV at a desired position. The main advantage of a PID controller [3] and [5] is the simple design. Furthermore, those controllers do not suffer from offsets in presence of disturbances. However, the control performance achievable with such a basic control schemes is often not satisfying. The second approach is motivated by the theory of optimal control and is used in [2], [6] and [7]. An optimal feedback strategy is calculated that minimizes a quadratic cost function, using a linear model of the quadrotor. The following two requirements are necessary for the solution to be optimal: actuator saturation must not occur and the reference position has to be constant. Both preconditions are typically not fulfilled; in order to overcome these requirements, a model predictive controller is proposed.

There exist several FPGA based stereo vision platforms such as [8][9]. Unlike these systems, the sensor presented in this work also incorporates an inertial measurement unit (IMU), as its application domain is more geared towards motion estimation rather than dense 3D reconstruction. We believe that what sets our integrated system apart from solutions comprised of individual sensors is its precise calibration for inter-sensor spatial transformations and time delays. Commonly, the transformation between a camera and inertial sensors is estimated recursively [10][11], while time delays are determined in a separate process [12]. In contrast, we estimate these quantities in a unified, batch optimal framework, achieving sub-millisecond accuracy in the synchronization of the different sensors [13].

In the computer vision and robotics literature, the Visual Odometry (VO), or visual Simultaneous Localization and Mapping (SLAM) problem is well studied: it is formulated as estimation of both structure, typically in the form of sparse landmarks, and camera poses observing those as two-dimensional projections. Recently, it has been shown that sparse nonlinear optimization over a selection of key-frames is computationally more efficient and provides more accurate results than filtering approaches [14]. Since the optimization problem becomes inherently intractable over time when constantly adding observations, various approaches have been suggested to keep the algorithms limited in complexity and real-time capable: a common procedure, employed e.g. in the monocular real-time SLAM algorithm PTAM [15], is keeping the optimized frames bounded, yielding some form of a sliding-window optimization. PTAM has been adapted [4] to operate on-board MAVs and providing a six-dimensional pose input to an Extended Kalman Filter (EKF) for loosely-coupled IMU-vision integration. In terms of tightly coupling such vision algorithms with inertial sensing, we base our approach on [16], where a nonlinear batch optimization is suggested that jointly optimizes over reprojection error of landmarks as well as deviation of the motion from the prediction based on inertial measurements.

3. Aerial System Design and Control

This section describes the platform that was used during the experiments in boiler unit II. First, the aerial vehicle prototype is described, and modelling, identification and control are presented. The next section then provides an overview over the on-board vision-aided inertial navigation subsystem

that provides real-time motion estimates during flight and allows the controller to stabilize the vehicle.

Modelling and Control

The base of the aerial platform are the AscTec Hummingbird attitude controller and brushless motor controllers. In other words, roll, pitch and yaw velocity are controlled by the base platform, relying on the attitude controllers integrated IMU. A custom airframe protects the propellers during wanted or unwanted contacts with the environment. The platform is lightweight enough in order to survive most crashes, easy to use and generally save for humans to operate.

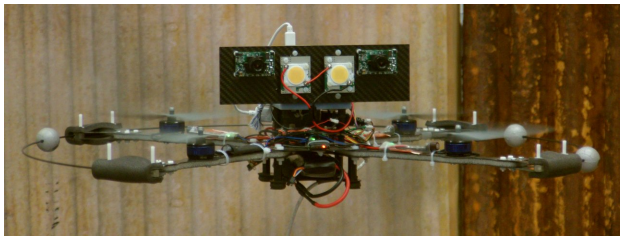


Figure 2. Protoype aerial vehicle used for the experiments in boiler unit II. The sensorhead consists of a pair of cameras arranged in a classical, front-looking stereo configuration. An on-board active illumination module provides sufficient illumination to operate inside the boiler. The IMU is mounted on the back side.

The commands to the attitude controller are the two angles ϕ^* and θ^*, the total thrust T^* and the yaw rate $\dot{\psi}^*$. To find a relation between those commands and the response of the quadrotor a simple point mass model as described in [6] is used. Fig. 3 provides an overview of this approach. The unknown response of the attitude controller to reference commands was identified with a blackbox modelling approach (prediction error method, PEM). Models of first order followed by time delays were found to be accurate enough for the task at hand. The interested reader is referred to [17] for additional details regarding modelling and identification of the vehicle.

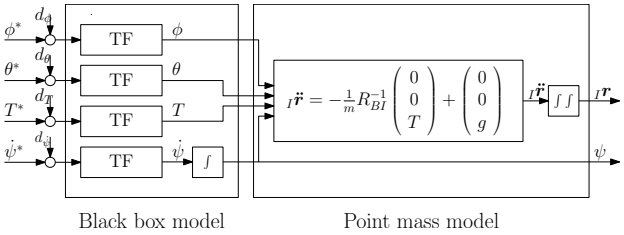


Figure 3. Complete model of the quadrotor using a simple point mass and four transfer functions representing the behaviour of the on-board attitude controller to the commands $(\phi^*, \theta^*, \dot{\psi}^*, T^*)$.

Different approaches for trajectory following were evaluated. A model predictive control strategy is conceptually simple and offers a number of benefits in practice, and was therefore selected. The key idea is to take the current state x, use the model to predict a fixed number N of future states and find an optimal sequence of commands that minimizes the quadratic cost function shown in Eq. 1. Only the first command u_0 is then applied to the vehicle and the problem is solved again in the next time step.

$$J_N^*(x) := \min \frac{1}{2}x_N^T P x_N + \frac{1}{2}\sum_{k=0}^{N-1} x_k^T Q x_k + u_k^T R u_k \quad (1)$$

$$\begin{aligned} \text{s.t.} \quad & x_{k+1} = Ax_k + Bu_k, && \forall k = 0, \ldots, N-1 \\ & u_k \in \mathbb{U}, && \forall k = 0, \ldots, N-1 \\ & x_0 = x \end{aligned}$$

The cost function can be extended to include any desired trajectory, which allows the controller to react to reference changes in advance, reducing the needed control actions, but still following the trajectory in an optimal sense. This leads to less aggressive flight maneuvers, which reduces motion blur in the camera images, an important aspect for the visual navigation subsystem. In addition the cost function can be adapted to punish fast attitude changes to further smoothen the behaviour of the quadrotor. The weighting matrices $Q = Q^T \geq 0$ and $R = R^T > 0$ are therefore chosen to track the position and avoid fast changes in attitude. A terminal penalty matrix P is necessary to guarantee stability.

Vision-Aided Inertial Navigation System

The following subsection provides an overview of the system that was developed for motion estimation of the aerial vehicle. The vehicle state consists of the following variables

$$x = (p_I \quad q_{IB} \quad v_I \quad b_a \quad b_\omega)^T \quad (2)$$

where p_I denotes the position of the sensorhead, q_{IB} the attitude quaternion, v_I the velocity and b_a, b_ω the accelerometer and gyroscope biases, respectively. Fusion of visual and inertial cues is, strictly speaking, implemented in a loosely coupled fashion following the stochastic cloning approach outlined in [18]. Only an overview of the processing pipeline is provided here, and the interested reader is referred to [19] for further details.

In the first stage, a FAST [20] keypoint detector in conjunction with an adaptive thresholding scheme provides well distributed image features that are then tracked in-between camera frames using a hamming matcher based on BRIEF [21] descriptors. A motion hypothesis is provided by the inertial measurements and then refined in a bundle adjustment step over the current and the most recent image pairs. This refined estimate is the used in an update step following a stochastic cloning approach. Gyroscope and accelerometer biases, modelled as random walks, are estimated within the same framework.

Having an IMU in the loop which is time-synchronized with the camera shutters allows the estimation of confidence intervals, in the image plane, of where tracked features must re-appear in subsequent camera frames. Hence, once inertial sensor biases are tracked accurately, one can guide feature matching and gain robustness when operating in poorly or repetitively textured scenes. Also, we are able to continuously estimate the vehicle state even when visual tracking fails completely, but only for a short period of time.

4. Integrated Sensor Design and State Estimation

The following section presents the prototype design of a *SLAM in a Box* system, a general-purpose module able to

provide synchronized and calibrated raw-data. Alternatively, it directly provides reliable motion estimates capabilities to different types of robotic platforms, and in a large variety of environments. The hope is that such a module will be helpful for engineers that simply need a sensor that can provide pose estimates in GPS-denied environments, and to researchers in need of high-quality raw data alike.

Sensor Design

The core of the visual/inertial sensor is a Xilinx Spartan 6 FPGA that implements interfaces to up to four Aptina MT9V034 global shutter CMOS image sensors and an Analog Devices ADIS16488 factory-calibrated MEMS IMU. Additionally, a set of low- and medium cost MEMS gyroscopes and accelerometers are mounted for applications where weight or cost are crucial. The FPGA connects through a Gigabit Ethernet interface to a host computer. By routing all sensor streams through the FPGA, it is possible to accurately timestamp all measurements with the same clock source. Fig. 4 shows the sensor hardware.

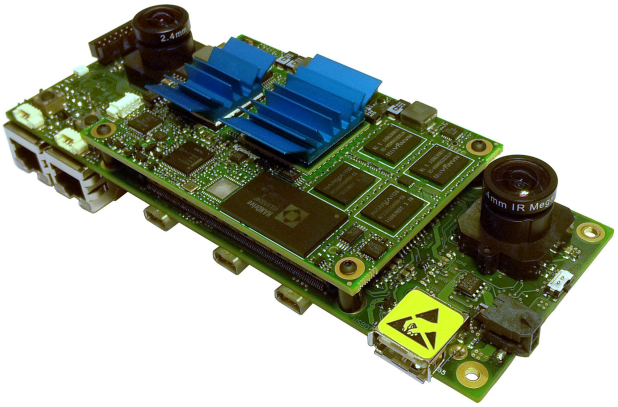


Figure 4. The integrated visual-inertial SLAM sensor. Only the two cameras mounted in a fronto-parallel stereo setup were used in this study. Like the aerial vehicle, the sensor is equipped with a camera exposure synchronous LED flash (not illustrated) to enable operation in the boiler. The FPGA is on the bottom side of the PCB, in the center-top is the commercially available Intel ATOM module.

In order to fuse measurements from a set of sensors contributing towards a motion estimate, the spatial transformation between the sensors has to be known accurately. In addition—and despite our efforts to synchronize the sensors in hardware—there still exists a deterministic time delay between the cameras and inertial sensors, caused by communication delays as well as filter delays in the inertial sensor that have not been accounted for in the time-stamping. Hence, a calibration for these fixed delays is required to achieve optimal estimation results.

In order to calibrate for these inter-sensor transformations and time delays, we employed a unified framework based on continuous-time batch estimation as proposed by Furgale et al. [22] (see [13] for details).

The sensor unit can then either provide synchronized, intrinsically and extrinsically calibrated visual and inertial measurements, a map of features such as e.g. Harris scores, computed in the FPGA, or directly motion estimates computed on the CPU.

Tightly Coupled Visual-Inertial Odometry

We applied a loosely-coupled approach for visual-inertial state estimation [19] running on-board the aerial vehicle: this algorithm serves as the basis for the model-predictive controller described above. As a complementary method, we also experimented with tight integration of inertial sensing and visual odometry. Keypoint observations and associations with map landmarks are created using the BRISK detector [23] and binary descriptor extractor. Inertial error terms are formulated using accelerometer and gyro readings integrated between camera frames using standard IMU kinematics following largely the method described in [16]. Knowledge on the statistical properties of the inertial sensors on the one hand and keypoint detection accuracy on the other hand allow for statistically near-optimal fusion. Inspired by the recent success of key-frame based visual odometry algorithms we applied some generalization of the aforementioned batch fusion to handle a constant size sliding window of key-frames. We select key-frames by field-of-view overlap. Consequently, no key-frames are inserted to the optimization when keeping the pose, which results in almost non-existent drift in this relevant case.

5. Experiments and Results

All experiments presented in this section were conducted in a boiler unit of a thermal power plant in northern Spain. This particular unit was temporarily shut down, and could be safely accessed through a man-hole at the bottom. Fig. 1 shows a side-view of the boiler.

First, results on the flight experiments are discussed, and results obtained with the integrated sensor introduced in Section 4 are presented subsequently.

Flight Experiments

The results presented here aim to demonstrate that the methods presented in Section 3 work in this particular environment. They give an indication of the performance that can be achieved using only on-board sensors, and no external tracking devices or infrastructure.

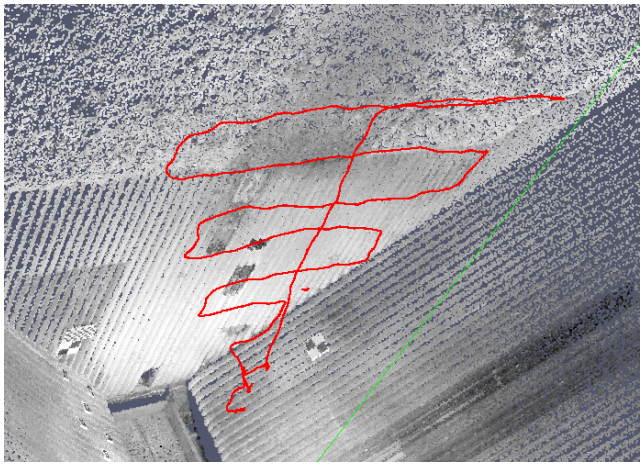


Figure 5. Actual flight path executed by the aerial vehicle, tasked to follow a pre-programmed trajectory with a forward velocity of 0.8m/s.

The task of the aerial vehicle was to follow a pre-defined trajectory at a distance of about one meter to the boiler

surface. For the first run, the desired forward velocity was set to 0.25m/s, and then increased to 0.8m/s for the second run, to test the dynamic capabilities of the vehicle and the visual-inertial motion estimation system to the limit. Fig. 5 shows an overlay of the path the vehicle followed. Fig. 6 shows the vehicle during the flight.

Figure 6. Aerial vehicle during a flight test. At a distance of approximately one meter to the boiler wall, the MAV passes an area where the surface is insulated with an non-ferromagnetic, slightly eroded heat shield. The scene is illuminated by high-power LEDs carried on-board, synchronized with the camera shutter to save power and avoid over-heating.

The MPC loop was closed at a frequency of 50Hz, with a prediction horizon of $N = 30$ steps. Image processing was performed off-board, and the vehicle was provided with power over a tether cable from the ground station.

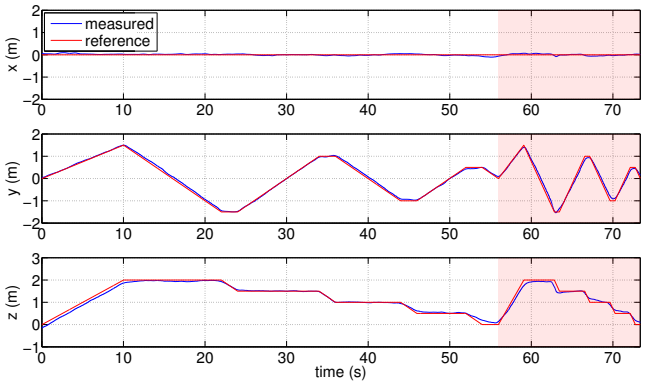


Figure 7. Aerial vehicle following a predefined trajectory with a velocity reference of 0.25 m/s. During the second run (highlighted), the forward velocity was increased to 0.8m/s.

Fig. 7 indicates that the vehicle was able to follow the desired trajectory accurately for both demanded forward velocities. It is important to note that the reference is compared to the pose estimates provided by the on-board sensors, and *not* against an externally measured ground-truth. This is justified by the fact that these estimates are accurate, as demonstrated in the following section, but strictly speaking, not correct.

Compared to alternative control strategies evaluated, the control actions of the MPC controller are smooth, which in turn supports the vision-based motion estimation module. Also, there is virtually no overshoot, an important requirement when flying in close proximity with the environment.

Table 1 lists the RMS tracking errors for the two different forward velocities. Keeping a constant distance to the wall in direction x was not and issue for the controller. However, especially in the vertical (z), external disturbances caused by the tether cable negatively affect the tracking error.

Axis	RMS value for 0.25m/s	RMS value for 0.8m/s
x	3.68 cm	3.88 cm
y	5.25 cm	10.55 cm
z	7.32 cm	11.30 cm

Table 1. Measured position error (RMS) in world coordinates, following the trajectory shown in Fig. 7.

Sensor Evaluation

In order to evaluate the performance of the tightly integrated visual/inertial sensor and motion estimation framework (Section 4), the sensor unit was mounted on an aluminum beam and had to undergo movements that resemble the motion experienced during a typical flight.

For quantitative evaluations, a Leica Total Station TS15 was used to track a reflective prism mounted on the sensor, with sub-millimeter precision. In addition, a visual calibration target (a checkerboard) was attached onto the boiler wall, and visited at the beginning of every dataset. The purpose of this target was two-fold. First, it was used to align the coordinate frame of the Total Station with the somewhat arbitrarily chosen coordinate frame of the motion estimation framework. The alignment was realized by measuring a sufficient amount of corners using the laser tracker and thus determining the rigid-body transformation between the coordinate frame of the tracker and the pattern. Subsequently, about 200 observations of the calibration pattern were used to determine camera poses in the coordinate frame of the tracker by means of a perspective N-point (PnP) algorithm using the implementation of [24]. Associating these poses to the motion estimation by timestamps, the tracker frame was then aligned with the estimation frame in a least square fashion over all 200 pose estimates. The other purpose of visiting the visual target as part of every dataset was to determine the unknown time offset between laser tracker and sensor timestamps. Again, this was implemented as a least square optimization over the position estimates provided by the PnP algorithm and the laser tracker.

Although the Leica Total Station provided measurements at sub-millimeter accuracy, the precision of the quantitative evaluations is limited by the uncertainty in the calibrated chain of transformations from the inertial measurement unit to the reflective prism, as well as jitter in recording of the laser tracker measurements, which amounts to a degradation in accuracy to about a few millimeters.

In addition to accurate position ground truth, full three dimensional structure scans of the boiler were recorded using a static Faro laser scanner, although comparisons to our structure reconstruction will remain qualitative in nature here.

Fig. 8 depicts results of our visual-inertial motion estimation framework for an experiment of approximately 8 minutes length. The camera frame rate had been set to 20 Hz and the IMU returned readings at a rate of 200 Hz. In the figure, the estimated sensor path is marked as a solid blue line, while the dashed black line shows ground truth obtained from laser tracker readings. The red curve shows a section of about 200 frames during which the calibration target was visible in the

video frames and that was used for path alignment and time offset calibration. In this experiment, the time offset was determined to be about 0.2 seconds, which would significantly distort the ground truth comparison when not accounted for. The results show that the estimation framework is in general capable of recovering the motion. However, it also highlights that the estimate diverges from ground truth over the course of the experiment due to accumulation of errors. Fig. 10 depicts this divergence quantitatively. The motion estimation exhibits comparatively low drift, which results in a displacement from ground truth of about 10 cm over the dataset.

Fig. 9 shows a qualitative comparison of the structure as recovered visually and by a laser scanner. For the laser scan only every thousandth measurement is displayed. The visual reconstruction is significantly more sparse and scarcely corrupted by outliers. Nevertheless, the reconstruction of the lateral walls of the boiler aligns accurately with the laser scan.

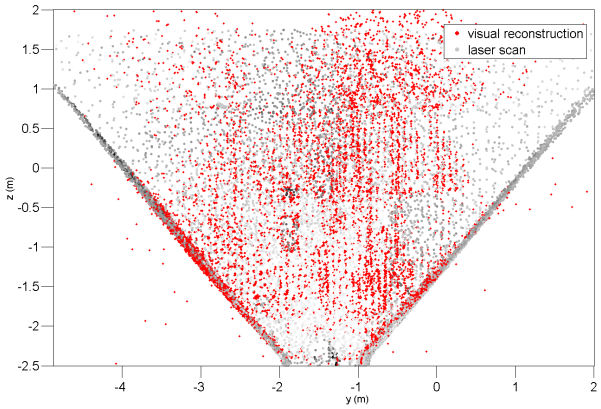


Figure 9. Overlay of a sparse, visual reconstruction of a section of the boiler onto a laser scan. The visually reconstructed point cloud aligns accurately with the laser scan, which becomes particularly apparent for the lateral walls of the boiler.

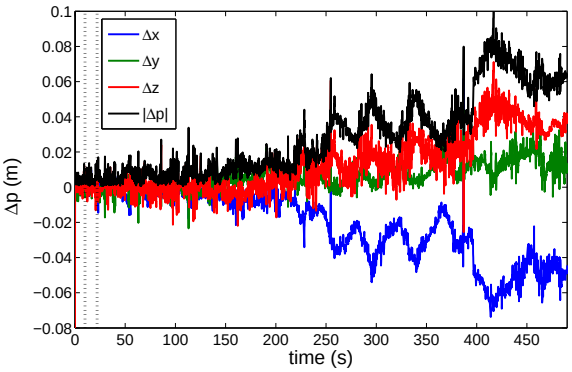
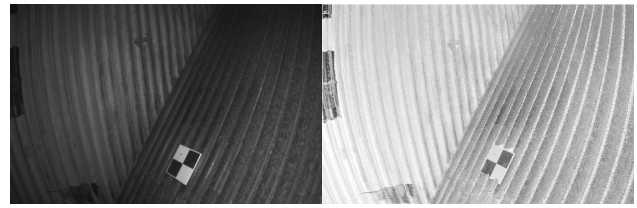


Figure 10. Visualization of the position error of the visual/inertial motion estimation framework. The vertical, dotted lines frame a section of roughly 10 seconds, which was used for aligning the sensor path with the tracker measurements. The results exhibit a low level of drift, amounting to less than 10 cm over a dataset of roughly 9000 frames spanning 500 s.

Fig. 11 illustrates a comparison between a video frame from the dataset and a synthetic view rendered from a laser scan of the structure and the estimated pose of the camera at the moment the video frame was recorded. The rendering is based on camera intrinsics and distortion parameters identical to the ones calibrated for the setup used in conducting the experiments. Although purely qualitatively, this result highlights the orientation component of pose estimation, a quantity that is not accounted for by position ground truth. Camera image and synthetic view exhibit great resemblance, suggesting that not only the camera position but also its orientation has been recovered accurately.



(a) (b)

Figure 11. Comparison of a camera frame from the dataset (11(a)) and a synthetic view (11(b)) using the estimated camera pose and reflectance measurements from a structure scan of the boiler. The rendering applies the camera intrinsics and distortion parameters calibrated for the camera used in the experiments. The figures exhibit great resemblance, suggesting that not only the position of the camera but also its orientation has been estimated to sufficient accuracy.

6. Conclusion

This work presented an unmanned aerial system with application to industrial inspection tasks. To the best of our knowledge, the autonomous flight experiments conducted inside an industrial boiler are unprecedented. As such, the results convey an idea of how autonomous aerial systems could facilitate visual inspection in enclosed industrial environments. By commanding scanning trajectories, we were able to take sets of overlapping, fronto-parallel photographs of the boiler walls at close range, enabling detection of larger structural damages. Quantitative results for our visual-inertial state estimation framework show a small drift in position in the order of around 10 cm over more than 8 minutes. This suggests that following large trajectories is feasible even without external navigation aids.

Acknowledgments

The authors would like to thank the following people for their support: Vera De-Vries, Wolfgang Mesters, Elsa Rico, and Eleni Tsigkogianni from Alstom, and Javier Vega Arias from GNF.

The research leading to these results has received funding from the European Communitys Seventh Framework Programme (FP7/2010-2013) under grant agreement no. 248669 (AIRobots).

This project also received funding from the Swiss Commission for Technology and Innovation (CTI), project no. 13394.1 PFFLE-NM (Visual-Inertial 3D Navigation and Mapping Sensor).

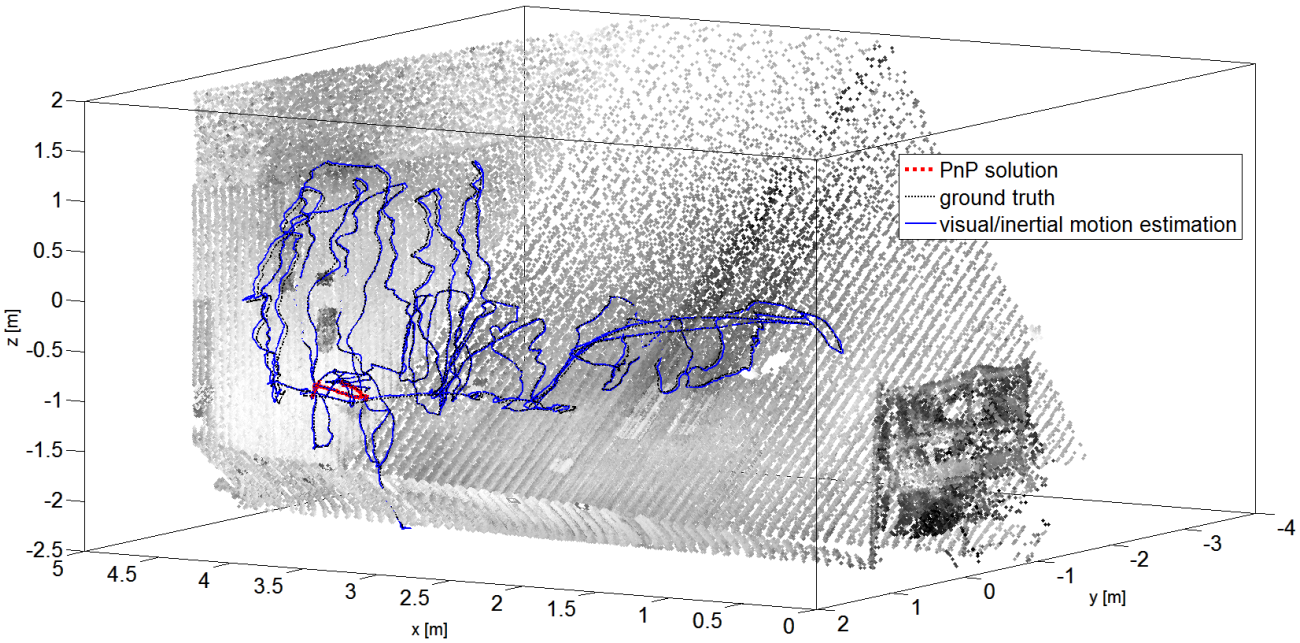


Figure 8. Comparison of ground truth positions and estimated sensor path. The red curve marks the PnP pose estimation used to align the coordinate frames of the motion estimation and the laser tracker. The motion estimation depicted in blue resembles the ground truth shown in black well. Nevertheless, drift becomes apparent over the course of the dataset.

References

[1] E. Stumm, A. Breitenmoser, F. Pomerleau, C. Pradalier, and R. Siegwart, "Tensor voting based navigation for robotic inspection of 3d surfaces using lidar point clouds," *The International Journal of Robotics Research*, 2012.

[2] S. Shen, N. Michael, and V. Kumar, "Autonomous multi-floor indoor navigation with a computationally constrained mav," in *Robotics and automation (ICRA), 2011 IEEE international conference on*. IEEE, 2011, pp. 20–25.

[3] A. S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox, and N. Roy, "Visual odometry and mapping for autonomous flight using an rgb-d camera," in *Proceedings of the International Symposium of Robotics Research (ISRR)*, Flagstaff, AZ, 2011.

[4] S. Weiss and R. Siegwart, "Real-time metric state estimation for modular vision-inertial systems," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, may 2011, pp. 4531 –4537.

[5] F. Fraundorfer, L. Heng, D. Honegger, G. Lee, L. Meier, P. Tanskanen, and M. Pollefeys, "Vision-based autonomous mapping and exploration using a quadrotor mav."

[6] M. Blösch, S. Weiss, D. Scaramuzza, and R. Siegwart, "Vision based mav navigation in unknown and unstructured environments," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, 2010, pp. 21 –28.

[7] A. Bachrach, R. He, and N. Roy, "Autonomous flight in unknown indoor environments," *International Journal of Micro Air Vehicles*, vol. 1, no. 4, pp. 217–228, 2009.

[8] C. Ahlberg, J. Lidholm, F. Ekstrand, G. Spampinato, M. Ekstrom, and L. Asplund, "Gimme-a general image multiview manipulation engine," in *Reconfigurable Computing and FPGAs (ReConFig), 2011 International Conference on*. IEEE, 2011, pp. 129–134.

[9] S. Jin, J. Cho, X. Dai Pham, K. Lee, S. Park, M. Kim, and J. Jeon, "Fpga design and implementation of a real-time stereo vision system," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 20, no. 1, pp. 15–26, 2010.

[10] F. Mirzaei and S. Roumeliotis, "A kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation," *Robotics, IEEE Transactions on*, vol. 24, no. 5, pp. 1143–1156, 2008.

[11] J. Kelly and G. Sukhatme, "Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration," *The International Journal of Robotics Research*, vol. 30, no. 1, pp. 56–79, 2011.

[12] J. Kelly and G. S. Sukhatme, "A general framework for temporal calibration of multiple proprioceptive and exteroceptive sensors," in *12th International Symposium on Experimental Robotics, 2010*, Delhi, India, Dec 2010.

[13] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," submitted to *Robotics and Automation (ICRA), 2013 IEEE International Conference on*.

[14] H. Strasdat, J. Montiel, and A. Davison, "Real-time monocular slam: Why filter?" in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, may 2010, pp. 2657 –2664.

[15] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'07)*, Nara, Japan, November 2007.

[16] D. Strelow and S. Singh, "Motion estimation from image and inertial measurements," *The International Journal of Robotics Research*, vol. 23, no. 12, pp. 1157–1195, 2004.

[17] M. Burri, J. Nikolic, C. Huerzeler, G. Caprari, and R. Siegwart, "Aerial service robots for visual inspection of thermal power plant boiler systems," in *Applied Robotics for the Power Industry (CARPI), 2012 2nd International Conference on*, 2012, pp. 70–75.

[18] A. I. Mourikis and S. I. Roumeliotis, "On the treatment of relative-pose measurements for mobile robot localization," in *in Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 15–19.

[19] R. Voigt, J. Nikolic, C. Hurzeler, S. Weiss, L. Kneip, and R. Siegwart, "Robust embedded egomotion estimation," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, sept. 2011, pp. 2694 –2699.

[20] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *In European Conference on Computer Vision*, 2006, pp. 430–443.

[21] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: binary robust independent elementary features," in *Proceedings of the 11th European conference on Computer vision: Part IV*, ser. ECCV'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 778–792. [Online]. Available: http://portal.acm.org/citation.cfm?id=1888089.1888148

[22] P. Furgale, T. Barfoot, and G. Sibley, "Continuous-time batch estimation using temporal basis functions," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 2088–2095.

[23] S. Leutenegger, M. Chli, and R. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2548–2555.

[24] J. Bouguet, "Camera calibration toolbox for matlab," *URL http://www. vision. caltech. edu/bouguetj/calib_doc*, 2010.

BIOGRAPHY

Janosch Nikolic received his MSc degree in signal processing from the University of Edinburgh, UK, in 2009. He is currently pursuing a PhD at the Autonomous Systems Lab of the Swiss Federal Institute of Technology, ETH, in Zurich. His research interests include statistically motivated vision-aided inertial navigation systems and aerial service robots for industrial inspection.

Michael Burri received his MSc degree in robotics and control from the Swiss Federal Institute of Technology, Zurich, in 2011, where he is currently holding the position of a researcher at the Autonomous Systems Lab. His research interests are in control of unmanned aerial systems, embedded systems design and power electronics.

Joern Rehder received his MSc equivalent degree (Diplom) in electrical engineering from the Hamburg University of Technology in 2011. He is currently a PhD student at the Autonomous Systems Lab, ETH Zurich and has been a visiting scholar at the University of California, Berkeley, and the Robotics Institute at Carnegie Mellon University. He is interested in sensor calibration.

Stefan Letutenegger received his BSc and MSc in mechanical engineering with focus on robotics and aerospace engineering from ETH Zurich and he has been a PhD student at ASL since 2009. His research areas range from unmanned solar airplane design and control to sensor fusion including visual-inertial localization and mapping.

Christoph Huerzeler received the M. Eng. degree in mechanical engineering from the Swiss Federal Institute of Technology, Zurich, in 2007, where he is currently pursuing the Ph.D. degree at the Autonomous Systems Lab. During his studies he focused on microrobotics and structure technologies. His main research interest is aerial indoor inspection robots with docking capabilities.

Roland Siegwart is a full professor for autonomous systems at ETH Zurich since July 2006. He has a Diploma in Mechanical Engineering (1983) and Ph.D. in Mechatronics (1989) from ETH Zurich. Previous positions include professor at EPFL (1996-2006) R&D director at MECOS Traxler AG (1990-1990) and senior researcher in robotics at ETH (1990-96). He holds visiting positions at Stanford University (1989/90 and 2005) and at NASA Ames (2005). He was co-initiator and founding Chairman of Space Center EPFL, Vice Dean of the School of Engineering EPFL and currently he is director of studies in ME at ETH. Roland Siegwart is member of the Swiss Academy of Engineering Sciences and the Research Council of the Swiss National Science Foundation, and IEEE Fellow. He served as Vice President for Technical Activities (2004/05), Distinguished Lecturer (2006/07) and AdCom Member (2007- 2009) of the IEEE Robotics and Automation Society. He is coordinator of various large projects and co-founder of several spin-off companies. He currently holds the position of vice persident research and corporate relations of ETH Zurich.