

---

# Autonomous MAV-based Indoor Chimney Inspection with 3D Laser Localization and Textured Surface Reconstruction

Jan Quenzel · Matthias Nieuwenhuisen ·  
David Droeschel · Marius Beul · Sebastian  
Houben · Sven Behnke

Received: date / Accepted: date

**Abstract** Inspection of industrial chimneys and smoke pipes induces high costs due to production downtimes and imposes risks to the health of human workers due to high temperatures and toxic gases. We aim at speeding up and automating this process with multicopter micro aerial vehicles. To acquire high quality sensor data, flying close to the walls of the chimney is inevitable, imposing high demands on good localization and fast and reliable control.

In this paper, we present an integrated chimney inspection system based on a small lightweight flying platform, well-suited for maneuvering in narrow space. For navigation and obstacle avoidance, it is equipped with a multimodal sensor setup including a lightweight rotating 3D laser scanner, stereo cameras for visual odometry and high-resolution surface inspection.

We tested our system in a decommissioned industrial chimney at the Zollverein UNESCO world heritage site, and present results from autonomous flights and reconstructions of the chimney surface.

**Keywords** autonomous inspection · SLAM · planning · surface reconstruction

## 1 Introduction

Industrial chimneys must be inspected regularly, which causes costly production downtime—the chimney has to cool down and toxic gases have to dissipate before humans can enter—and poses dangers to human inspection personnel working in large heights. Access to higher parts of the chimney is difficult and requires cranes or scaffolding. Multicopter micro aerial vehicles (MAVs) can carry inspection sensors, such as cameras, to high altitudes, but their manual control inside chimneys close to the inspected surface is not feasible. Especially at higher altitudes, human pilots can hardly assess the exact distance to the surface or the MAV orientation from the ground. Fig. 1 shows examples of challenging situations when operating in

---

Autonomous Intelligent Systems Group  
University of Bonn  
Bonn, Germany  
{quenzel, nieuwenh, droeschel, mbeul, houben, behnke}@ais.uni-bonn.de



Fig. 1: Challenging flight conditions. The perspective of a pilot on the bottom of a chimney makes flight maneuvers in larger heights hard to control without automation (MAV circled red). Left: Manual chimney inspection from the outside in larger distance from the wall is not suitable for 3D surface reconstruction. Center: Autonomous flight close to the walls. Right: Manual operation in larger heights inside the chimney is prohibitive, especially in daylight conditions.

larger heights. Furthermore, turbulences close to the walls require continuous adjustments to the control inputs, making the control even more challenging. Hence, we aim at fully autonomous operation, to make MAV-based inspection of chimneys feasible.

The foremost task of the autonomous system, and key functionality to build upon, is to safely navigate in the chimney in reasonable proximity to its walls. Starting from a coarse geometric model, a detailed 3D model of the chimney for localization and navigation is built by registering and aggregating measurements of a 3D laser scanner which is carried by the MAV. By fusing different sensor modalities, the MAV localizes with respect to this model.

The MAV is equipped with a high-resolution stereo camera to capture the surface of the chimney. To simplify the inspection task for a human expert, we create a highly detailed model of the whole chimney. We employ reconstruction of Structure-from-Motion (SfM) on RGB images to create a detailed 3D model of the chimney surface and visualize it as unwrapped high-resolution orthoimage for the examination by an expert. This expert can specify poses for further inspection and the MAV navigates to these poses autonomously based on the pose estimates and the 3D chimney model.

The use of autonomous MAVs will 1) diminish the risk for human inspectors, 2) reduce the costs of inspections due to shorter production downtimes, and 3) deliver high quality inspection results. We demonstrate the applicability of our system in a decommissioned chimney of a coking plant at the Zollverein UNESCO world heritage site, shown in Fig. 2. Furthermore, we tested and evaluated our system in a narrower chimney mock-up of about 4m in height and 3.5m in diameter. Navigating and maneuvering in this very constrained space is highly challenging since the air flow caused by the MAV is reflected from the chimney walls and the ground and causes turbulences.



Fig. 2: Autonomous chimney inspection. Our lightweight chimney inspection MAV navigates autonomously assisted by real-time 6D laser localization and fast visual odometry. A spiraling scan pattern facilitates good and efficient coverage for camera-based surface reconstruction of the chimney walls. Our MAV is evaluated in a narrow chimney mock-up (left) and a decommissioned industrial chimney (right).

Our main contribution is an integrated lightweight MAV, including camera and 3D laser-based 6D localization and high-resolution 3D surface reconstruction, with a total mass of less than 3.5 kg. The constrained space in chimneys requires reliable navigation close to the walls—a highly demanding task given the lower computational power of the lightweight MAV compared to our prior work (Nieuwenhuisen et al., 2016). To the best of our knowledge, this is the first autonomous MAV-based chimney inspection system.

## 2 Related Work

Employing MAVs for inspection and surveying tasks has been an active field of research in recent years. Still, most MAVs are operated manually or via global navigation satellite system (GNSS) waypoint following in obstacle-free distances to objects (Nex and Remondino, 2014; Chan et al., 2015). High-resolution images are taken at predefined positions for later inspection by an expert or for automated reconstruction of a model employing off-the-shelf software.

One exception is the work of Ortiz et al. (2014) who developed a quadrotor MAV for autonomous vessel inspection. Similar to our approach, they employ a combination of laser localization and visual odometry for navigation, but use a system of mirrors to direct some laser beams to floor and ceiling and employ a 2D localization approach decoupled of the height measurements. In contrast, we perform full 3D simultaneous localization and mapping (SLAM) and 6D localization.

A lightweight MAV system aiming at industrial boiler inspection has been presented by Burri et al. (2012). Their work focuses on agile movements in industrial environments with vision-based state estimation. Our goal is to build complete surface models of chimneys and we are, thus, interested in steady slower, but more accurate movements and a drift-free state estimate.

Intel demonstrated the inspection of an Airbus airplane with an MAV equipped with Intel RealSense sensors and a high-resolution camera (Intel Corp., 2016a).

In contrast to our work, the inspection was performed outdoors in much larger distance from the inspected surface.

In order to estimate depth of object points instantaneously for obstacle avoidance and state estimation, stereo cameras are used on MAVs, e.g., in the works of Schmid et al. (2014) and Park and Kim (2014). Tripathi et al. (2014) use them for reactive collision avoidance. The limited field of view (FoV) of cameras poses a problem when flying in constrained spaces like chimneys where obstacles are necessarily close to the MAV at all times.

To overcome these limitations, some MAVs are equipped with multiple (stereo) cameras. Moore et al. (2014) use a ring of small cameras to achieve an omnidirectional view in the horizontal plane, but rely on optical flow for velocity control, centering, and heading stabilization only.

Grzonka et al. (2012) use a 2D laser scanner to localize the MAV in environments with structures in flight altitude and to avoid obstacles. This limits obstacle avoidance to the measurement plane of the laser scanner.

Other groups combine laser scanners and visual obstacle detection (Tomić et al., 2012; Huh et al., 2013; Jutzi et al., 2014). Still, their perceptual field is limited to the apex angle of the stereo camera (facing forward), and the mostly horizontal 2D measurement plane of the scanner. They do not perceive obstacles above or below this region or behind the vehicle.

Özaslan et al. (2017) inspect tunnels with an MAV. Similar to our work, tunnels are dark and relatively self-similar and thus they also rely on a laser scanner and cameras with on-board illumination for localization. In comparison to our work, the flight is semi-autonomous and the collected data is not exploited by means of, e.g., surface reconstruction.

In chimney inspection, ascending flights are a main direction of movement and the MAV is operating close to surfaces such that omnidirectional obstacle perception is required. We use a continuously rotating laser scanner with a spherical FoV that does not only allow for capturing 3D measurements without moving, but also provides omnidirectional obstacle sensing at comparably high frame rates (2 Hz in our setup).

The proposed MAV extends our own previous work (Beul et al., 2015), an MAV with a dual 3D laser scanner and three wide-angle stereo camera pairs. In contrast to our prior work, the new MAV is significantly smaller and thus more constrained regarding payload. It has less compute power—this puts higher demands on more efficient processing—and only one stereo camera pair instead of three.

An earlier version of the ChimneySpector system has been presented at ICUAS 2017 (Nieuwenhuisen et al., 2017). We extend this with a new localization modality for the transfer to larger chimneys, improved and faster surface reconstruction, and new results from the application in a real industrial chimney.

### 3 System Setup

Our chimney inspection robot is based on the Ascending Technologies Neo hexacopter platform. With a diameter of about only 80 cm, the platform is well-suited for indoor flights. Fig. 3 shows our MAV and the used sensor setup.

The platform is equipped with a front-facing Skybotix VI-Sensor (Nikolic et al., 2014) used as stereo camera system for visual odometry. Each camera has a wide-

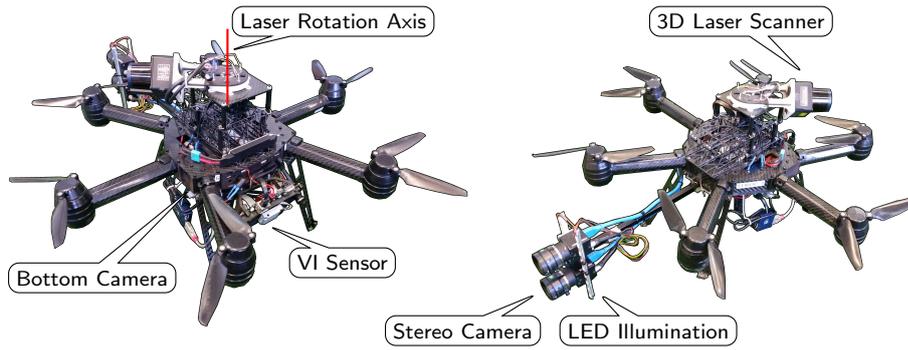


Fig. 3: Sensor setup. Our MAV is a hexarotor equipped with a rotating 3D laser scanner for localization and obstacle avoidance, a bottom camera and a stereo camera system for visual odometry estimation, and a high-resolution stereo camera for surface reconstruction. For better illumination and shorter exposure times, the scanned wall is illuminated by bright LEDs. The laser scanner rotates around the red axis to accumulate scan lines to full 3D scans.

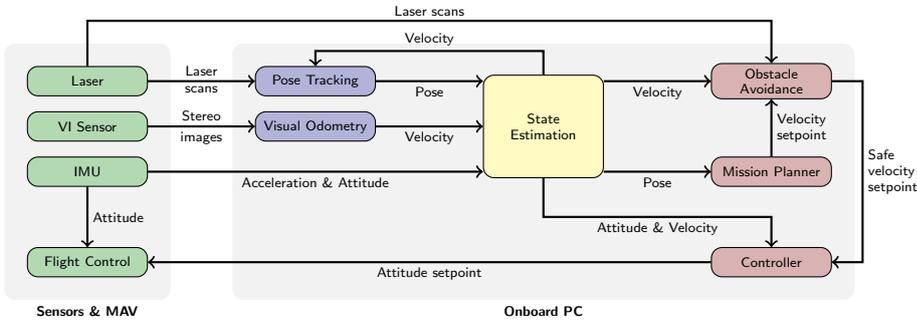


Fig. 4: Schematic overview of system components. Measurements from laser scanner and stereo cameras are processed to generate pose and velocity estimates. We filter these estimates with IMU measurements to get the MAV state with low latency at 100 Hz. A mission planner generates and executes chimney coverage tours or plans paths to all inspection poses for targeted inspection.

angle lens with an apex angle of  $122^\circ$  and a resolution of  $752 \times 480$  pixels. Images are captured with 20 Hz with hardware synchronization. The VI-Sensor is equipped with an IMU, calibrated w.r.t. the two cameras. Another down-facing camera on the bottom of the MAV is used to perceive a marker-based localization system.

On the top of the MAV, we installed a continuously rotating Hokuyo UST-20LX laser scanner for localization and obstacle avoidance. The sensor rotates at a frequency of 1 Hz, yielding a spherical 3D FoV. Due to the  $270^\circ$  apex angle and its mounting pose, it covers the space above the MAV with 2 Hz—chimney inspection starts at the ground and thus unknown obstacles are more likely to be above the MAV—and the space below with 1 Hz.

For surface reconstruction and inspection, we previously used an Intel RealSense SR300 RGB-D camera (Intel Corp., 2016b) mounted at the rear-end of the MAV.

Unfortunately, the estimated depth by the SR300 exhibits a static deformation (Quenzel et al., 2017). Additionally, the camera produces only dark rolling shutter images so that we decided to replace the sensor by a stereo camera rig. Hence, we use a synchronized global shutter USB stereo camera system consisting of two PointGrey Blackfly BFLY-U3-23S6C-C color cameras, which simplifies and improves the surface reconstruction compared to rolling shutter acquisition. The cameras are rotated by  $90^\circ$  to form a vertical stereo pair, running at 10 Hz. The reason for this flipped setup is that the main direction of movement and, thus, the main direction of motion blur is going to be horizontal during the inspection flight. Thus, vertical image gradients will be more likely to be blurred while horizontal gradients that are used for stereo matching by the proposed setup will be retained.

Two stripes of bright LEDs were added to light up the inspected area and to obtain sharp and well-illuminated images, even in higher altitudes where no external light sources exist. We further fixed the exposure time to 8 ms to avoid strong motion blur during rapid motions.

The rig is mounted on a 32 cm long cantilever to bring the sensor closer to the surface while mitigating the effects of turbulences when flying close to walls.

All navigation-relevant sensor data processing, localization, and planning is performed onboard the MAV. Thus, the MAV can operate fully autonomously without depending on an unreliable WiFi connection to a ground control station. To facilitate this, the MAV is equipped with a small and lightweight Intel NUC PC with Intel Core i7-5557U dual core CPU running at 3.1 GHz and 16 GB of RAM. For surface reconstruction, we acquire high-resolution images of size  $1920 \times 1200$  pixels instead of the previous FullHD. Consequently, the two cameras generate approx. 200 MB per second with our settings. To store this data stream, we equipped the onboard PC with a fast Samsung Pro 950 SSD.

The 3D laser scanner and the Skybotix VI-Sensor are connected via Ethernet, and the laser rotator via a serial-to-USB interface. Fig. 4 gives a schematic overview of the flight-relevant system components. The overall weight of the system including all sensors and batteries is about 3.4 kg.

#### 4 Localization and State Estimation

In order to navigate in a chimney, robust localization and state estimation, not relying on GNSS availability are crucial. Our multimodal localization and state estimation pipeline exploits the specific characteristics of all sensors in terms of, e.g., accuracy and speed.

##### Visual Odometry

Our visual odometry estimation is based on LIBVISO2 (Geiger et al., 2011), a fast feature-based visual odometry library for monocular and stereo cameras. The stereo odometry approach is very general and does not require a specific motion model. The only prerequisite is that the input images have to be rectified and the

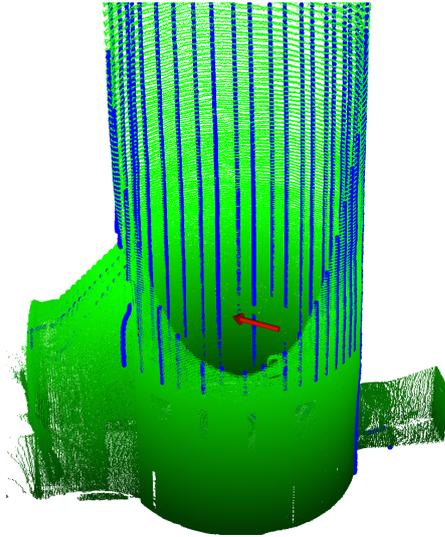


Fig. 5: Laser-based localization. To track the MAV pose (red arrow) in an allocentric frame, we aggregate individual laser scans over 500 ms (blue) and match these to an allocentric map of the chimney (green).

calibration parameters are known. Similar to other feature-based methods, LIBVISO2 extracts and matches features over subsequent stereo frames and estimates the egomotion by minimizing the reprojection error. To be robust to outliers, RANSAC is used for initialization of the minimization step.

The interior of a chimney poses a challenging environment for visual odometry algorithms—the tracked features are always close to the camera due to the restricted space and the repetitive structures are self-similar. Thus, visual odometry is prone to heavy drift and can easily lose track. We use only the position derivative of the visual odometry and integrate it as noisy velocity estimate into our state estimation filter at the stereo camera frequency of 20 Hz.

#### Laser-based Pose Tracking

To localize the MAV in an allocentric chimney frame, we track its pose by registering local multiresolution maps to a global map employing multiresolution surfel registration (MRSR) (Droeschel et al., 2014). This yields a 6D pose estimate in the map frame at 2 Hz (Fig. 5). We build the map of the chimney ad-hoc from the takeoff position before a mission. For larger chimneys, our approach is able to perform SLAM during a first simple exploration flight, e.g., flying straight up and down in the chimney center. The allocentric map is represented by surfels with a uniform size.

Since the laser scanner acquires complete 3D scans with a relatively low rate, we incorporate the filtered egomotion estimate from visual odometry and measurements from the Neo’s IMU to track the pose of the MAV. The egomotion estimate is used as a prior for the motion between two consecutive 3D scans.

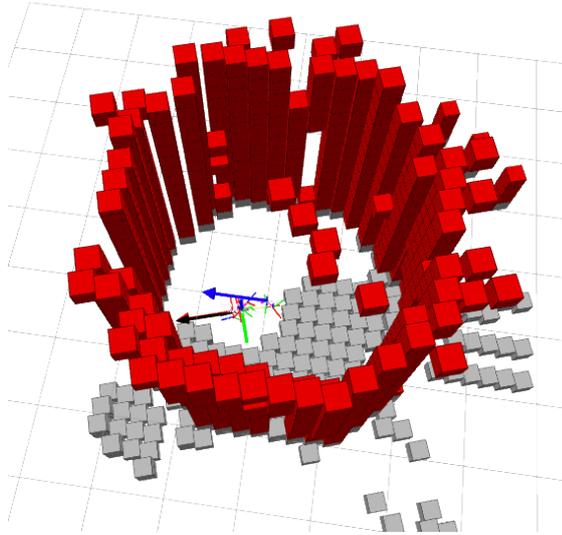


Fig. 6: Laser obstacle map. To avoid collisions, we maintain an obstacle map containing the most recent 3D laser scan and move it with our motion estimate (black arrow). The blue arrow depicts the laser localization pose.

In detail, we track the pose hypothesis by alternating the prediction of the MAV movement given the filter result and alignment of the current local multiresolution map towards the allocentric map of the environment.

The allocentric localization is triggered after acquiring a 3D scan and adding it to the local multiresolution map. We update the allocentric robot pose with the resulting registration transformation. To achieve real-time performance of the localization module, we only track one pose hypothesis. We assume that the initial pose of the MAV is roughly known by starting from a predefined pose in the center of the chimney or, if this is not possible, setting the pose in our control GUI. The approximately known pose is then quickly refined by scan registration. Here, small structures, e.g., a ladder commonly attached to chimney walls, are sufficient to align with the map.

Fig. 5 shows the registration of a 3D scan to the map and an estimated 6D pose. The resulting robot pose estimate from the allocentric localization is used as a measurement update in a lower-level state estimation filter.

For obstacle avoidance, we create a local obstacle map consisting of the most recent individual laser measurements that form a full laser rotation, i.e., 1 s of aggregated laser scans, depicted in Fig. 6. The obstacle map is kept in an egocentric frame by incorporating the MAV egomotion estimate.

#### Visual Detection of a Fiducial Ground Pattern

The chimney is highly symmetric and lacks visual or geometric distinctive features that would allow to continuously and reliably identify an absolute measure of



Fig. 7: The setup and exemplary recording from the down-facing wide-angle camera. LED panels with printed AprilTags are placed on the ground in an arbitrary arrangement (left and center). The robot detects these markers (right) and derives an absolute measure of orientation.

orientation. To circumvent this caveat, we deploy an artificial pattern consisting of several uniformly lit LED panels, each printed with a distinct AprilTag (cf. Fig. 7). The panels are placed on the ground and are recorded with a down-facing wide-angle camera attached to the bottom of the MAV. By sufficiently reducing the exposure time of the camera, the panels are easily segmented within the camera image, both accelerating and robustifying the detection. There is no need to account for lens distortion in a preprocessing step as the used AprilTag detector by Olson (2011) is able to perform well under moderate distortion. We follow a SLAM approach in order to estimate the camera pose and the poses of all tags at the same time, i.e., it is not necessary to choose a particular arrangement of the LED panels. However, in order to define a unique coordinate system, one of the panels is distinguished from the others. It has to be detected first and spans a frame that defines the poses of the other panels and the MAV itself. The state estimation itself is similar to our previous work (Houben et al., 2016a). Briefly, every AprilTag is modeled by its four corner points

$$\begin{aligned} P &= \{p_{i,A_k} : i = 1, \dots, 4\} \\ &= \left\{ \left(-\frac{s}{2}, -\frac{s}{2}, 0\right), \left(\frac{s}{2}, -\frac{s}{2}, 0\right), \left(\frac{s}{2}, \frac{s}{2}, 0\right), \left(-\frac{s}{2}, \frac{s}{2}, 0\right) \right\} \end{aligned}$$

with  $s$  being its size, within its own coordinate frame  $A_k, k = 0, \dots, n$ . Let  $A_0$  be the frame of the distinguished AprilTag. To model the image generation, we transform the corner points into the camera frame  $C$ :

$$\begin{pmatrix} \mathbf{p}_{i,C} \\ 1 \end{pmatrix} = T_{C,A_k} \begin{pmatrix} \mathbf{p}_{i,A_k} \\ 1 \end{pmatrix} \text{ for all } \mathbf{p}_{i,A_k} \in P \text{ and } k.$$

$T_{C,A_k}$  is the  $4 \times 4$  matrix representing the rigid transform between the AprilTag coordinate frame  $A_k$  and the camera frame  $C$ . Subsequently, we project the corner points using the camera matrix  $K_C$  and the bijective lens distortion function  $f : \mathbb{R}^2 \mapsto \Omega \subset \mathbb{R}^2$ :

$$\begin{aligned} \pi(x, y, z) &= \left( \frac{x}{z}, \frac{y}{z} \right), \\ \mathbf{p}_{i,C} &= f(\pi(K_C \mathbf{p}_{i,C})). \end{aligned}$$

This yields two equations for each of the four corner points of each AprilTag. The detection of the distinguished AprilTag with the detected corner points  $\mathbf{u}_{0,i}$ ,  $i = 1, \dots, 4$  allows us to solve

$$\operatorname{argmin}_{T_{C,A_0}} \sum_{i=1}^4 \left( f \left( \pi \left( K_C I_{3 \times 4} T_{C,A_0} \begin{pmatrix} \mathbf{p}^{i,A_0} \\ 1 \end{pmatrix} \right) \right) - \mathbf{u}_{0,i} \right)^2 \quad (1)$$

in order to obtain the relative camera pose. Whenever an AprilTag is detected for the first time, the relative pose of the tag with respect to the camera is computed using Eq. (1). Finally, let  $\mathcal{A}$  be the set of all detected tags in a frame. The camera pose is computed using all present detections:

$$\operatorname{argmin}_{T_{C,A_0}} \sum_{k \in \mathcal{A}} \sum_{i=1}^4 \left( f \left( \pi \left( K_C I_{3 \times 4} T_{C,A_0} T_{A_0,A_k} \begin{pmatrix} \mathbf{p}^{i,A_k} \\ 1 \end{pmatrix} \right) \right) - \mathbf{u}_{k,i} \right)^2.$$

Here  $T_{A_0,A_k}$  denotes the rigid transform between the tags 0 and  $k$ . In regular intervals, all detections of a frame are stored in a new keyframe to enable global bundle adjustment and, thus, refine the estimated poses of the AprilTags and keyframes by minimizing the overall backprojection error.

### State Estimation Filter

We use two filters for state estimation: A low-level filter fuses measurements from accelerometers, gyros, and compass to one 6D attitude and acceleration estimate in the AscTec Trinity flight control unit. The second higher-level filter fuses linear acceleration, velocity, and position information to a state estimate that includes 3D position and velocities. The higher-level filter is based on the Pixhawk Autopilot (Meier et al., 2012) position estimator adapted to use visual odometry velocity estimates and laser pose tracking, running on the onboard computer. IMU measurements are incorporated at 100 Hz and a pose/velocity estimate is published at the same rate.

In our position filter, we estimate the state

$$x = \begin{pmatrix} p_x & p_y & p_z \\ v_x & v_y & v_z \\ a_x & a_y & a_z \end{pmatrix},$$

consisting of position  $\mathbf{p}$ , velocity  $\mathbf{v}$ , and acceleration  $\mathbf{a}$ . The prediction step, based on the current bias-corrected acceleration measurement  $\mathbf{a}_{k,sens}$ , is

$$\begin{aligned} \mathbf{p}_k &= \mathbf{p}_{k-1} + \mathbf{v}_{k-1} \cdot dt + \frac{1}{2} \mathbf{a}_{k,sens} \cdot dt^2, \\ \mathbf{v}_k &= \mathbf{v}_{k-1} + \mathbf{a}_{k,sens} \cdot dt, \\ \mathbf{a}_k &= \mathbf{a}_{k,sens}. \end{aligned}$$

If sensor measurements are available, i.e., the measurement is not timed out, the state is corrected accordingly. For velocity measurements  $\mathbf{v}_{k,sens}$ , coming from, e.g., visual odometry, the state correction is:

$$\begin{aligned} \mathbf{v}_k &= \mathbf{v}_{k-1} + (\mathbf{v}_{k,sens} - \mathbf{v}_{k-1}) \cdot w \cdot dt, \\ \mathbf{a}_k &= \mathbf{a}_{k-1} + (\mathbf{v}_{k,sens} - \mathbf{v}_{k-1}) \cdot w^2 \cdot dt^2. \end{aligned}$$

Here,  $w$  is a weighting factor that indicates the reliability of the inputs.

Position measurements  $\mathbf{p}_{k,sens}$ , coming from, e.g., the laser scanner, are incorporated as

$$\begin{aligned}\mathbf{p}_k &= \mathbf{p}_{k-1} + (\mathbf{p}_{k,sens} - \mathbf{p}_{k-1}) \cdot w \cdot dt, \\ \mathbf{v}_k &= \mathbf{v}_{k-1} + (\mathbf{p}_{k,sens} - \mathbf{p}_{k-1}) \cdot w^2 \cdot dt, \\ \mathbf{a}_k &= \mathbf{a}_{k-1} + (\mathbf{p}_{k,sens} - \mathbf{p}_{k-1}) \cdot w^2 \cdot dt^2.\end{aligned}$$

If no new sensor measurements are received, we gradually degrade the velocity estimate in the correction step until the filter stops.

## 5 Planning and Navigation

Capturing the surface of the chimney with the stereo camera requires a steady flight path with a fixed distance between sensor and walls. Furthermore, the images need sufficient overlap in every direction to build a consistent model for the whole flight. These demands are hard to fulfill in manual operation, especially given the turbulent air movement close to the walls pushing the MAV away and requiring constant control actions. Thus, we operate the MAV fully autonomously except for start and landing. First, we plan an inspection path with 50% image overlap starting 1 m above the ground and moving upwards in a spiraling motion. Fig. 8 shows an example inspection path in the coarse conic chimney model derived from the chimney documentation. In our previous work, we flew smaller vertical circles along this primary spiral to ensure a good image overlap to account for the requirements of our RGB-D sensor setup. This pattern had been found to be advantageous over a simple spiraling motion in preliminary flight tests. With our new improved camera setup, a single spiral is sufficient, speeding-up the capture flight significantly. The inspection mission is planned, given a simple geometric model of the chimney, the sensor characteristics (apex angles, best scanning distance), and the part of the chimney to cover.

After a first complete inspection, the user can specify poses for a targeted second inspection, e.g., to take close-up images of potential defects in the chimney. The MAV processes a set of inspection poses and determines an optimal visiting order to achieve a short inspection flight employing a traveling salesman problem solver (Fig. 9). Selection of poses for targeted inspection is assisted by a graphical tool on the ground station that shows the taken images to an operator and determines corresponding MAV poses for selected images. Thus, the operator can select images where a possible defect is visible and mark these poses for the second inspection.

To safely navigate in the vicinity of obstacles, e.g., probes, ladders, or open hatches in the chimney, we employ reactive local obstacle avoidance. For this, we extended our previous work on reactive obstacle avoidance (Nieuwenhuisen et al., 2013). Based on a laser map of the vicinity of the MAV (see Fig. 6), we reduce the MAV velocities towards close obstacles and actively push the MAV back from obstacles if the distance to an obstacle falls below a safety distance. The resulting velocity command  $\mathbf{v}_e$  given a target velocity  $\mathbf{v}_t$  and the artificial force  $\mathbf{F}$  induced

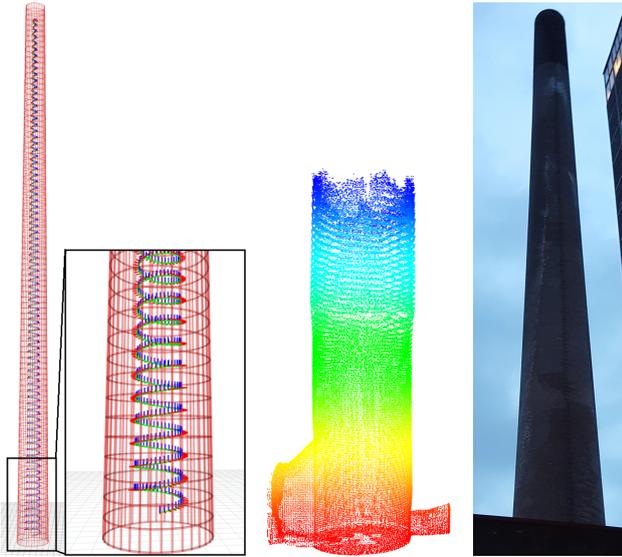


Fig. 8: Coverage tour in the chimney. Based on a few chimney parameters, we plan sensor coverage tours in a coarse geometric chimney model (left, red). The MAV moves from the bottom to the top in a spiraling motion (sensor poses depicted by coordinate axes). In contrast to the initial laser map captured from the starting position (center), the geometric model covers the complete chimney. Right: Photo of the chimney for comparison.

by an obstacle is calculated as

$$\mathbf{v}_o = (\mathbf{v}_t \cdot \hat{\mathbf{F}}) \hat{\mathbf{F}}, \quad (2)$$

$$\mathbf{v}_c = \mathbf{v}_t - s_{\text{slow}} \mathbf{v}_o + s_{\text{push}} \hat{\mathbf{F}}, \quad (3)$$

where  $\mathbf{v}_o$  is the part of the target velocity  $\mathbf{v}_t$  towards the obstacle. Negative parts of  $\mathbf{v}_o$ , i.e., velocities maneuvering the MAV away from the obstacle, are set to zero. The scalar factors  $s_{\text{slow}}$  and  $s_{\text{push}}$  denote the strength of slowing down the MAV and actively pushing it away from obstacles (see Fig. 10). Their value is a linear interpolation between free-space distance and the safety distance, and an interpolation between safety distance and critical distance, respectively. This extends our previous approach on reactive obstacle avoidance to allow for less conservative safety distances, as required by our stereo camera, while still maintaining safe navigation.

Velocity setpoints for the MAV are generated by means of a PID-controller in the integrated mission planning and navigation node when executing coverage or inspection missions. These setpoints are propagated to our obstacle avoidance module. After processing these setpoints, the resulting safe velocity setpoints are fed to the velocity controller to generate attitude and thrust commands. Our velocity controller is based on the linear model predictive controller (MPC) by Kamel et al. (2017), modified to work with velocity-only setpoints. The resulting

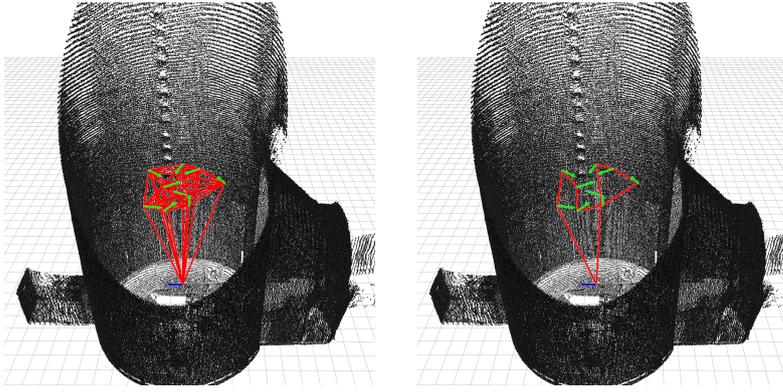


Fig. 9: Targeted inspection. After an operator selected targets to reinspect, the observation poses (green arrows) are sent to the MAV where a mission planner finds an optimal visting order. The left figure depicts all possible paths between view poses and the start/return pose (blue arrow), the right figure shows the best mission path. The MAV navigates to these poses autonomously and hovers there for several seconds to acquire more detailed data of the surface. The chimney map is depicted by black dots.

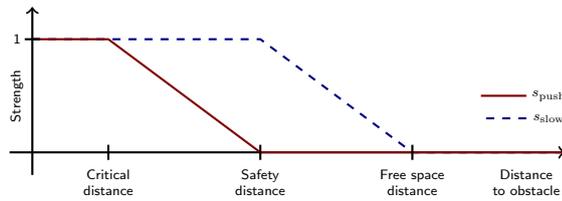


Fig. 10: Scaling factors for reactive obstacle avoidance. If the MAV is too close to an obstacle, it is actively pushed away by a force scaled with  $s_{\text{push}}$ . Entering a safety zone around obstacles is prevented by reducing velocities towards the obstacle with factor  $s_{\text{slow}}$ .

attitude-thrust setpoint is controlled by the low-level AscTec Trinity Autopilot of the MAV.

## 6 Surface Reconstruction

We aim at creating a high-resolution reconstruction of the chimney surface. In contrast to our previous work (Nieuwenhuisen et al., 2017), we use a rear-facing synchronized stereo camera mounted at a cantilever arm with additional illumination. The stereo camera is calibrated offline by using the Kalibr (Furgale et al., 2013) toolkit.

Chimney inspection experts are particularly interested in fine details showing irregularities within the surface, e.g., cracks in the wall. Hence, we use the full resolution of  $1920 \times 1200$  px for the stereo cameras.

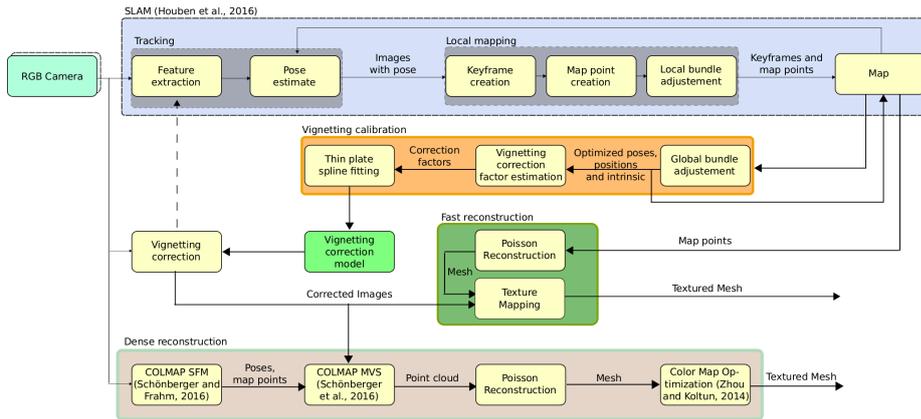


Fig. 11: Schematic overview of our surface reconstruction pipeline. We derive an accurate camera trajectory, estimate the 3D structure of the surface, correct the vignetting and fuse the data to colored meshes. This process is either done onsite with SLAM and the fast reconstruction yielding a first result of the inspection mission and later offsite using the dense reconstruction with SfM and MVS from COLMAP.

Fig. 11 shows an overview of our surface reconstruction pipeline. In contrast to our prior work, we add a new low-latency coarser reconstruction which is done within minutes onsite on a ground station laptop. This fast reconstruction enables the inspection experts to define points of interest for a second flight. The dense reconstruction using the Structure-from-Motion (SfM) and Multi-View Stereo (MVS) pipeline of COLMAP (Schönberger and Frahm, 2016; Schönberger et al., 2016) is done offsite. We further improve the quality of the reconstructed meshes by correcting camera-static color and vignetting effects.

The real-time estimated MAV pose, used for control and navigation, is not accurate enough for surface which demands a higher precision to prevent reconstruction artifacts. This prohibits the direct application of a mapping with known poses approach. Hence, we reestimate the MAV trajectory from the onboard stored stereo camera stream using a modified multi-camera variant (Houben et al., 2016b) of ORB-SLAM (Mur-Artal et al., 2015).

### Fast Reconstruction

The fast reconstruction uses only the keyframes created by the SLAM system with their respective poses and the triangulated map points. The map points are first meshed by the Poisson surface reconstruction included in the Point Cloud Library (Rusu and Cousins, 2011). We map the texture onto the resulting mesh by simply projecting triangles into the keyframes. The keyframe with the largest projected area is chosen as a triangles reference keyframe. We chose the largest projection, since it typically corresponds to the locally closest keyframe with most detail and has the smallest incident angle. The projection area is extracted to texture the

triangle. The resulting textured mesh can then be unwrapped to create the 2D orthoimage.

To cope with the self-similar chimney environment, we added restrictions on loop-closing to perform only if no connections between the candidates local maps are available. So far, the loop-closing was only appearance-based using a Bag-of-Words approach, which did not perform well under varying perspectives that often occur during exploration flights with minimal overlap. Candidate frames were often discarded since the number of matches was too low. To address this issue, we added a pose-dependent loop closing on found candidates based on frustum intersection and local map matching with two-way verification. For each candidate frame with intersecting frustum, we gather the local map points and match these against the current keyframe and vice versa. From these matches, we can establish connections between the local maps, find equivalent map points and estimate a correcting transform between candidate and current keyframe. Afterwards, we run a global bundle adjustment (GBA) using Ceres-Solver (Agarwal et al., 2016) on the whole map. To further improve the accuracy and reliability, we first optimize for the keyframe poses and map point positions with fixed intrinsics. In a second iteration, the intrinsic parameters—including lens distortion—are refined in the optimization. To prevent scale drift, we fix the transformation between the stereo color camera pair. Due to the synchronization of the color cameras, we have only one pose to optimize per frame pair. After convergence, the found ORB-features for each keyframe are again undistorted using the newly found distortion parameters, since only undistorted keypoints are used throughout the SLAM pipeline.

### Color and Vignetting Correction

The intensity of the textured mesh varies strongly, depending on the triangles position in the reference keyframe. This so called vignetting effect decreases the measured intensity with increasing distance to the image center and is a property induced by the camera lens and in our case in composition with the additional illumination. In order to provide a view of the reconstructed surface with uniform illumination and contrast, we adapted our online depth calibration method (Quenzel et al., 2017), previously developed for the SR300, to estimate pixel-wise correction factors for the color image. We show the effect of devignetting in Fig. 12.

The sparse map is projected into the keyframes and the corresponding color values are extracted. Hence, for each pixel  $\mathbf{u}$  within the image plane  $\Omega$  we might get multiple color triplets or none at all. The median of all measurements for a pixel is used to create a reference value  $I_r$  per channel. We obtain scaling factors  $s_{\mathbf{u}}$  by dividing the reference value by the measurements value, i.e., the recorded color within the image, and calculate a fitting thin plate spline for each color channel. Since not every point has enough scaling factors to obtain independent correction factors, we have to approximate the dense correction function. In fact, the scaling factors are only sparsely distributed. For this reason, the thin plate spline is employed due to excellent fill-in properties, ease of implementation, efficiency and the minimal bending energy property. The two-dimensional thin plate spline is

defined as:

$$f(\mathbf{u}) = \sum_{i=1} c_i \cdot \phi(|\mathbf{u} - \mathbf{d}_i|) + \mathbf{v}^\top \cdot \begin{pmatrix} 1 \\ \mathbf{u} \end{pmatrix}, \quad (4)$$

$$\phi(r) = r^2 \cdot \ln(r). \quad (5)$$

We place the spline control points  $\mathbf{d}_i \in \Omega$  statically in a regular grid of size  $5 \times 7$  on the image plane. The parameters  $\mathbf{c}, \mathbf{v}$  are easily calculated by minimization of the least squared error:

$$\operatorname{argmin}_{\mathbf{c}, \mathbf{v}} \sum_i^M \|f(\mathbf{u}_i) - s_i\|, \quad (6)$$

under the additional constraints:

$$\sum_i^N c_i = 0, \quad (7)$$

$$\sum_i^N c_{i,x} \cdot d_{i,x} = 0, \quad (8)$$

$$\sum_i^N c_{i,y} \cdot d_{i,y} = 0. \quad (9)$$

This minimization is achieved by solving an over-determined linear system of equations using conjugate gradient (CG):

$$\begin{pmatrix} A & X \\ D & 0 \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{s} \\ \mathbf{0} \end{pmatrix}, \quad (10)$$

$$A = \begin{bmatrix} \phi(|\mathbf{u}_1 - \mathbf{d}_1|) & \dots & \phi(|\mathbf{u}_1 - \mathbf{d}_N|) \\ \vdots & \ddots & \vdots \\ \phi(|\mathbf{u}_M - \mathbf{d}_1|) & \dots & \phi(|\mathbf{u}_M - \mathbf{d}_N|) \end{bmatrix},$$

$$X = \begin{pmatrix} 1 & \mathbf{u}_1 \\ \vdots & \vdots \\ 1 & \mathbf{u}_M \end{pmatrix}, \quad D = \begin{pmatrix} 1 & \dots & 1 \\ \mathbf{d}_1 & \dots & \mathbf{d}_N \end{pmatrix}.$$

Robustness against outliers is improved by means of iteratively reweighted least squares (IRLS) (Holland and Welsch, 1977). The usage of CG further speeds up this process since a previous estimate of the solution can be used for the next iteration.

Given the resulting thin plate spline, we can correct the vignetting by applying the inverse attenuation function:

$$\mathbf{a}_\mathbf{u} = f(\mathbf{u}), \forall \mathbf{u} \in \Omega,$$

$$V(\mathbf{u}) = \frac{\mathbf{a}_\mathbf{u}}{\max_{\mathbf{u} \in \Omega} \mathbf{a}_\mathbf{u}},$$

$$I_{a,\mathbf{u}} = I_\mathbf{u} \cdot V(\mathbf{u})^{-1}.$$



Fig. 12: Result of vignetting correction: Left: The brightness decreases towards the corners of the original image. Right: The metal ladder (red circle) becomes more distinct after correction.

The maximum over all scaling factors within a color channel is used to scale the values between zero and one, since vignetting is only observable up to scale (Engel et al., 2016). Vignetting is generally exposure-dependent, but since the cameras have fixed exposure time, we can simply apply the above correction. The thin plate splines are evaluated once for each pixel and used for all images of the same camera.

In contrast to the GBA, the correction is only applied during texture creation and not during SLAM. Application during SLAM would require repeated keypoint detection, descriptor extraction—and possibly matching—for keyframes. This would increase computation time tremendously.

### Dense Reconstruction

After capturing all images, we employ COLMAP (Schönberger and Frahm, 2016) to further improve the accuracy and density of the chimney surface model. The vignetting correction model optimized during the fast reconstruction is then applied to all images for reconstruction. The image corners still contain black regions. Hence, we crop around the image center to  $1775 \times 1100$  px resolution to remove this border. SIFT features are extracted from the images and the reference trajectory is estimated from every fourth image per camera. The estimation includes the intrinsic parameters of both cameras, these are shared between all images of a camera.

For dense reconstruction, we previously used MVE (Fuhrmann et al., 2014). This often resulted in small holes in the mesh where dark regions with low intensity gradients are present. Hence, we decided to employ the MVS module of COLMAP (Schönberger et al., 2016), which automatically fills these. This has the further advantage, that COLMAP makes extensive use of available GPUs to speed up computation, while MVE only relies on the CPU. The final mesh is then exported from the MVS point cloud using Poisson surface reconstruction. The detail of the mesh is high enough to not run Color Map Optimization (Zhou and Koltun, 2014) to refine image poses with respect to the mesh, although this can still increase the accuracy.

In order to use Color Map Optimization, we first sub-sample the mesh multiple times and then optimize the image poses w.r.t. the mesh s.t. the intensity differences of vertices between observing images is reduced. For a better result, we now use all captured and corrected images, instead of only the subset used for sparse and dense reconstruction. The image poses are initialized by linear interpolation from the SfM poses. The final vertex color is obtained as a viewing-dependent weighted average of all vertex observations. A section of the generated high-resolution mesh is shown in Fig. 13. Finally, the reconstructed mesh is unwrapped to yield an easily inspectable 2D visualization, depicted in Fig. 14.

### Mesh Unwrapping and Visualization

As an initial step, we place the chimney mesh into a canonical coordinate system, chosen as being centered at the origin and having the principal axis of the chimney aligned with the z-axis of the world. Therefore, we estimate the center point, the radius and the principal axis that best fit the chimney data using RANSAC. Afterwards, the rotation that aligns the principal axis of the fitted cylinder with the z-axis of the world is computed using the Rodrigues’ formula (Blanco, 2010). Unwrapping the mesh is subsequently performed by computing cylindrical coordinates for each mesh vertex in our canonical coordinate system.

In order for an expert to perform a further inspection of the chimney, we need to render an image large enough for the possible surface defects to be visible on it. Given the size of the mesh and the texture assigned to it, we empirically chose a resolution of  $19\,000 \times 12\,000 \text{ px} = 228 \text{ Mpx}$ . To perform the rendering of such a large image, we rely on the approach of tile rendering. Hence, we divide the image into smaller tiles of size  $800 \times 600 \text{ px}$ , render them individually and assemble them in one image. To prevent sending the whole mesh (approx. 3 GB) to the GPU to be rendered, which will fail in the case of insufficient memory, we rather render the mesh triangle by triangle while removing those that lie outside the frustum of the camera tile.

Not only the visual image of the scene is of interest to the expert, but also depth discontinuities. Hence, we generate a depth map in which each vertex is assigned with the depth deviation to the fitted cylinder. However, the chimney does not have a perfectly cylindrical shape but rather a slightly oval one. In order to cope with this, we make the assumption that bricks and cracks in the wall correspond to high frequency details while the slight oval shape corresponds to lower frequency information. Based on this assumption, we filter the mesh with a Laplacian operator to extract the fine details and effectively push the mesh more

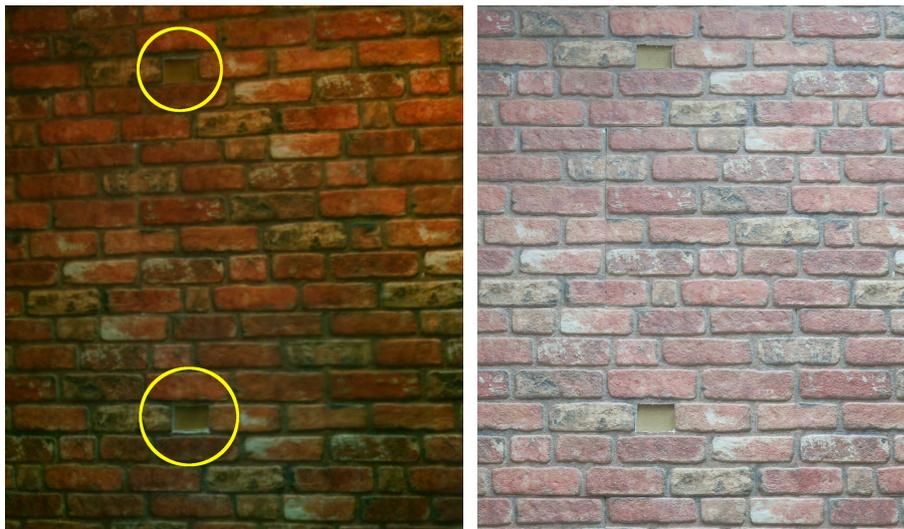


Fig. 13: Comparison of reconstructed surface with a photograph. Left: Close up of the rightmost chimney wall in our reconstruction depicted in Fig. 14. Right: Image of the surface taken with a compact camera. The defects in the wall are clearly visible in the reconstruction (circled yellow).



Fig. 14: 2D visualization of the chimney walls. Reconstructed surface of seven walls from the mock-up chimney shown in Fig. 2. Despite some warping effects caused by the unwrapping of the 3D structure, all details are accurately reconstructed.

in line with the ideal cylinder while still retaining the protruding bricks and cracks on the surface. Results of this mesh unwrapping are depicted in Fig. 14 and Fig. 20.

## 7 Results

We tested and demonstrated our integrated system in two close-to-application scenarios: Inspection and reconstruction of a narrow chimney mock-up and a decommissioned industrial chimney. Both scenarios were defined in close collaboration with a chimney inspection service contractor. Videos of our experiments can be found on our website.<sup>1</sup>

<sup>1</sup> [http://www.ais.uni-bonn.de/videos/JINT\\_2018\\_ChimneySpector](http://www.ais.uni-bonn.de/videos/JINT_2018_ChimneySpector)

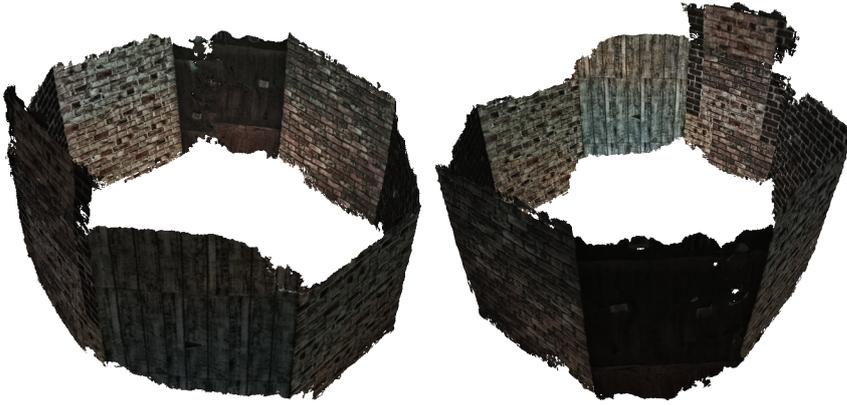


Fig. 15: 3D reconstruction of the chimney mock-up. The highly detailed textured mesh (shown from two perspectives) is the result of our dense surface reconstruction after an initial coverage flight.

### Experiments in Chimney Mock-up

First, we used an octagonal chimney mock-up with an inradius of 1.8 m and a height of 4.4 m, shown in Fig. 2. The mock-up consisted of eight wooden panels with styrofoam structures on the inner sides—resembling the stonework and concrete patterns that can be found in many industrial chimneys. Single structure elements were of size  $1.0 \times 0.5$  m and each of the stone walls—except of the wall containing an entry to the mock-up—was plastered with a single type of elements, resulting in repetitive patterns. In addition, one panel carried a rusty iron surface as found in chimneys with a metal alloy on the inner side. Some bricks in the styrofoam elements were carved out to represent defects. These experiments were performed with an SR300 sensor instead of our stereo camera setup employed in later experiments. The then-required sensor distance of 0.8 m to 1.0 m to the surface yielded a remaining safe navigation space with a diameter of only  $\approx 1.2$  m. The mock-up was designed and built by the chimney inspection service contractor to facilitate the transferability to real inspection applications.

We started with an initial coverage flight to acquire RGB-D data of the chimney surface. For surface coverage, the MAV followed a horizontally and vertically spiraling pattern with the RGB-D sensor directed to the nearest surface in order to enable loop closings in the later surface reconstruction. Initial flight tests showed that the depth data of the used sensor were not sufficiently accurate for surface reconstruction due to the turbulent flight. Thus, we disregarded the depth measurements and processed only the RGB information.

The flight to acquire the data used for the reconstruction shown in Fig. 15 took seven minutes. We covered  $36 \text{ m}^2$  of the chimney surface with a single charge of batteries. All defects were covered by that area. The flight was fully autonomous, with the exception of start and landing.

After the coverage flight, we downsampled the recorded video stream from the RGB-D sensor to speed-up the process of transferring it to a ground control station. Here, an operator can identify and store poses for a more detailed inspection in

the video stream. In the test case, the operator could identify all ten defects in the images. The MAV poses, corresponding to the images showing defects, were saved for a second flight to reinspect those defects. To exemplify the targeted inspection of previously identified defects, the MAV then autonomously planned an inspection mission and followed a path, adopting all stored poses in a useful order and held its position for several seconds to demonstrate reaching the desired position. All poses were successfully reached and the defects were clearly identified in the data captured during the second flight. After the targeted inspection, the acquired data was transferred to the ground control station for further offline processing.

On average, we acquired depth information for 95.5% of the pixels, thus, the sensor was almost always positioned in an optimal distance to the surface.

In a post-processing step, we reconstructed the surface of the chimney walls. We could successfully reconstruct all of the eight segments. All segments covered with artificial stonework were reconstructed without major artifacts. The reconstruction of the rusty metal segment was sufficient for visual inspection. Depth could not be estimated accurately for some parts of the metal wall as the images were too dark to yield enough features—a side effect of the relatively low exposure time required to avoid motion blur. This issue was addressed with a new high-resolution stereo camera setup, additional onboard illumination and an intermittent devignetting step in later experiments. Overall, the representation encompasses all interesting details, in particular, all ten defects carved into the styrofoam wall. Fig. 14 shows the resulting unwrapped 2D visualization of the seven mock-up walls with stonework, Fig. 15 shows the 3D model of the complete chimney.

The computational load during the coverage flight posed a particular challenge due to the laser and visual odometry-based state estimation on the two available physical CPU cores. In combination with recording of the data streams from both the SR300 and stereo cameras desired for optimal post-processing, navigation and MAV control showed an unstable behavior during initial tests. Hence, we refrained from recording all non-crucial data—including the stereo cameras—and estimated the MAV trajectory for surface reconstruction solely from the the monocular vision data from the SR300 sensor. Furthermore, we captured frames at a reduced frequency of 30 Hz instead of the desirable maximum frequency of 60 Hz. Thus, post-processing was more challenging due to the necessary compensation for larger inter-frame motions. Still 200 MB/s were recorded from the SR300 sensor during the coverage flight. With these means towards reducing the system load and by tuning the onboard PC to reach a steady data flow without peaks, we accomplished stable and safe system performance.

The 3D laser scanner served well for low-frequency allocentric localization in a previously recorded map in combination with visual odometry from the VI stereo sensor to compensate for the MAV motion between scans. Furthermore, this combination yielded a high-frequency positioning and state estimation and allowed for stable MAV control when filtered with IMU readings. The stability of this approach allowed us to rely on a pre-chosen initial pose in order to resolve the orientation ambiguity by the symmetric chimney cross section.

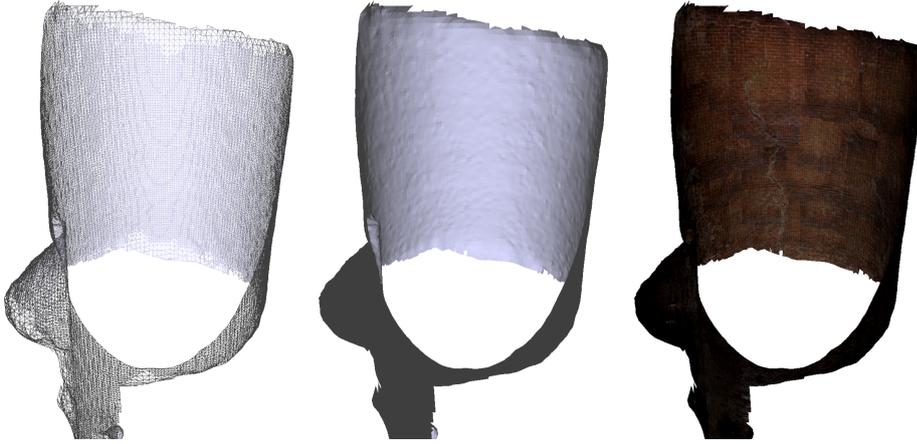


Fig. 16: Fast reconstruction. For visualization purposes, we reconstructed a coarse model of the chimney within minutes. The geometry of the surface is approximated by a very reduced number of vertices compared to the dense reconstruction. We meshed the vertices and applied a texture to the model to provide the operator with an initial visual impression of the inspection mission. From left to right: Wireframe model, untextured surface, textured chimney model.

#### Experiments in a Decommissioned Industrial Chimney

Based on the experiments in the chimney mock-up, we improved our system setup and transferred our system to a scenario closer to the inspection of actual industrial chimneys. We evaluated and demonstrated our system in a chimney of a decommissioned coking plant at the Zollverein Coal Mine Industrial Complex in Essen, Germany. The total height of the chimney is 98 m. The inradius at the bottom is 2.75 m and tapers to 2.2 m on the top. With our improved camera setup, we could increase the distance to the surface to 1.5 m, resulting in a larger covered area. Furthermore, we could omit the smaller vertical spirals for better coverage of the surface. Both improvements yielded a much higher surface coverage speed. The MAV flew at a maximum speed of 1.5 m/s. In total, the area covered with a single charge of batteries could be increased to 140 m<sup>2</sup> due to the faster coverage procedure. The autonomous flight time was 3:27 minutes, plus an additional minute for manual start, landing, and hovering phases. As a result, we covered seven times the surface area per flight time than with our previous setup and reconstruction methods.

After the initial coverage flight, a first 3D model of the chimney was processed, employing our fast surface reconstruction method. This fast reconstruction step took approximately 5:30 minutes (30 s for vignetting correction, 5 min for creating the textured mesh), plus the time to load the data (due to an unoptimized loading step, this equals the flight time). This mesh, shown in Fig. 16, contained 43 463 vertices and 80 980 textured faces. All processing was performed on a notebook equipped with an Intel Core i7-6700 HQ CPU running at up to 3.6 GHz and 32 GB RAM. This fast reconstruction helps to assess if the data collection was successful in a short period of time after the flight.



Fig. 17: 3D reconstruction of the Zollverein chimney. In addition to the spiral coverage of the chimney, the MAV covered a part from the bottom to the begin of the pattern during start, and a higher part of the chimney with a straight ascending flight before landing.

In parallel, an operator has identified three defects for targeted inspection on the ground control station employing the captured image stream. A targeted inspection mission has been planned and all three defects have been reached autonomously in a single flight. The MAV hovered for several seconds at each inspection pose to capture data.

A highly detailed 3D model was created in an offline processing step offsite. Fig. 17 shows the resulting model, and Fig. 20 the unwrapped orthoimage. The precise dense reconstruction took approximately 3 h to 10 h for the SfM calculations, 48 h for MVS, and another 6 h for fusion and meshing. These calculations were performed on a workstation equipped with eight NVIDIA Titan Xp GPUs—six of them were used for MVS—each with 12 GB RAM and two Intel Xeon e5-2683v4 CPUs running at up to 3 GHz from which up to ten cores were used in parallel for the SfM calculation. The resulting textured mesh contained 80 718 456 colored vertices and 161 643 410 faces.

We compare the estimated trajectories of our laser pose tracking, the trajectory based on AprilTag detections and the filtered trajectory in Fig. 19. Fig. 18 shows the estimated yaw of the MAV during chimney inspection. Both, laser pose tracking and AprilTag detections qualitatively show similar trajectories. The magnification, however, underlines the characteristics of the different modalities. Whereas the laser pose tracking is always available, it shows some rotational misregistra-

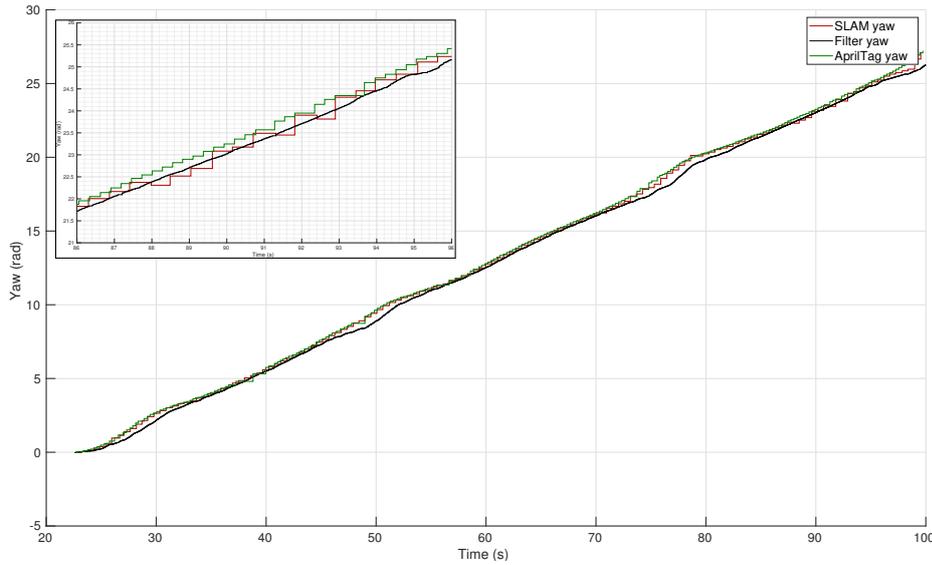


Fig. 18: Yaw of the MAV during chimney inspection. Large: Estimated yaw of the entire inspection flight of laser pose tracking, filter and AprilTag detections. Small: Magnification of the data between 86 s and 96 s.

tions due to the rotational symmetry of the chimney. The yaw estimates based on the fiducial ground pattern are smoother and do not suffer from the symmetries as they measure an absolute angle but might temporarily not be available when the MAV is pitching in larger heights, e.g., after 91 s in the plot. Between 88 s and 89.5 s, the laser yaw estimate had an offset of  $\approx 10^\circ$ , but the AprilTag detections were available. Thus, laser pose tracking and the detection of a fiducial ground pattern can aid the state estimation filter in different situations. Furthermore, while the AprilTag detections give an absolute yaw estimate, the laser pose tracking gives a much better estimate of the position relative to the chimney.

In order to compare the results of the state estimation with the reconstruction, we evaluated the absolute trajectory error (ATE). As a reference we chose the poses found by the SfM reconstruction from COLMAP since we assume this trajectory to be the most accurate. The reference trajectory was rescaled such that the baseline matched the calibrated baseline. The trajectories were then aligned using the method from Horn (1987). The mean ATE without intrinsic refinement during GBA using the fast reconstruction was about 0.311 m. This was reduced with the intrinsic refinement to only 0.074 m. For the laser pose tracking trajectory, we get a mean ATE of 0.20 m with a standard deviation of 0.14 m. The resulting state estimate from our filter had an ATE of 0.13 m (0.06 m).

The color correction improved the dense reconstruction, as can be seen in Fig. 21. The difference between merged observations during fusion of dense depth maps is reduced.

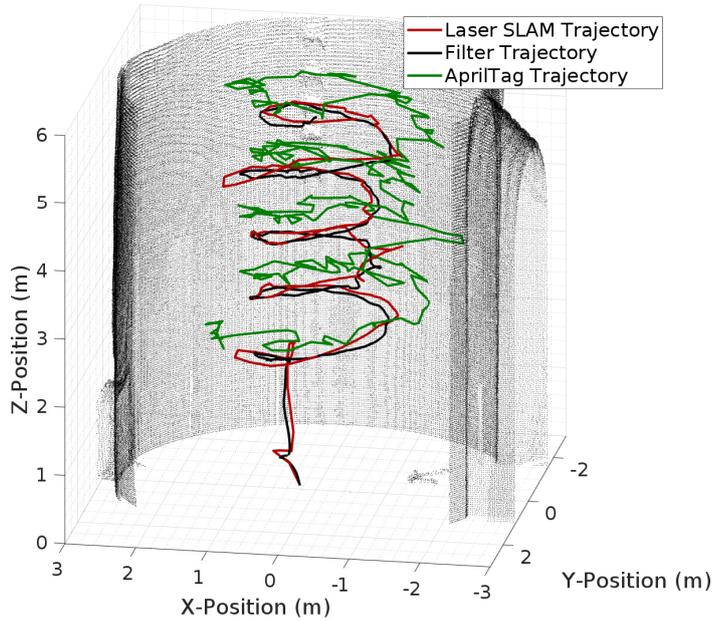


Fig. 19: Estimated trajectories of the chimney inspection. The raw trajectories were estimated by laser pose tracking and the AprilTag detections. While both sensors show significant noise, the filtered trajectory is smooth.

## 8 Conclusions

In this work, we have presented an integrated chimney inspection system based on a lightweight MAV platform. The small form factor severely restricts payload, thus, choosing lightweight sensors and a small onboard PC with limited compute power is essential. To this end, we reduced the onboard sensor setup from our previous work to a smaller subset and still maintain good real-time state estimation and obstacle perception. In addition, we added a new sensor for surface reconstruction. Integrating the whole sensor, state estimation, and control setup on a lightweight MAV posed challenges but could be successfully achieved. Important lessons learnt are that special care has to be taken on required data for online and offline processing and a sensible assessment of the required accuracies and real-time performance for every part of the system has to be performed. Overall, the MAV can cover chimney surfaces autonomously and a high-quality surface reconstruction is possible in a post-processing step. Defects identified by experts can be reinspected at more detail in a second mission.

We have proven the applicability of our system in an experiment in a decommissioned industrial chimney. Here, we were able to cover  $140\text{ m}^2$  of the surface in 3:30 minutes. The total surface of the chimney is approximately  $1520\text{ m}^2$ , thus, we could cover the complete surface with 11 of such flights, resulting in 38:30 min-

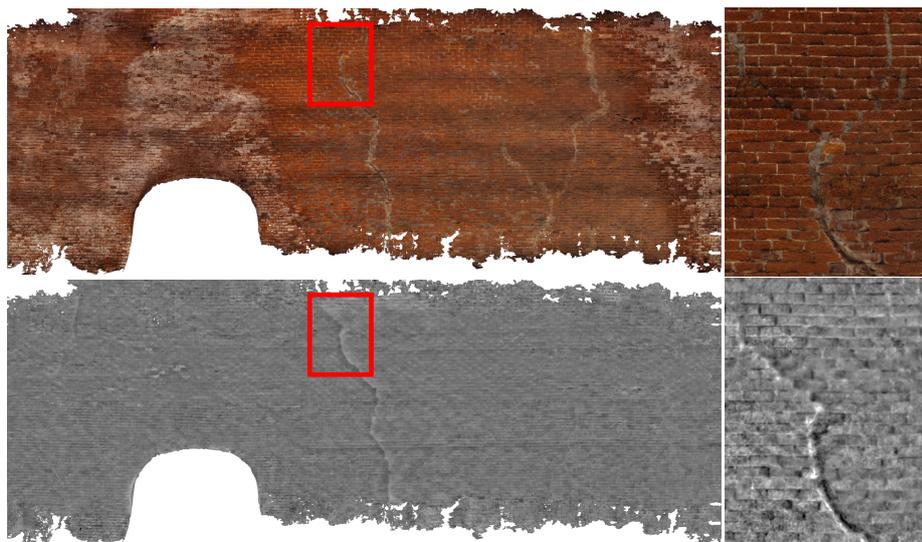


Fig. 20: Unwrapped surface reconstruction. Top: The unwrapped textured mesh of the decommissioned chimney. Bottom: Corresponding depth map. In contrast to the fast reconstruction, the finer geometric structure of the chimney is visible as well. Defects like the large crack in the center of the images can be spotted easily by visual inspection of the depth map.

utes of coverage time. With an estimated overhead of 2 min per flight for ascent, descent, and battery change, an MAV-based inspection system could cover this exemplary industrial chimney in approximately one hour.

### Acknowledgments

The authors wish to thank the Autonomous Systems Lab from ETH Zürich and Ascending Technologies for their technical and organizational support, especially for providing the flying platform. We wish to thank CRN Management GmbH for building the mock-up chimney, for providing expertise for the domain of chimney inspection, and for their support during the evaluation flights. Furthermore, we wish to thank the Zollverein Foundation for the opportunity to develop and evaluate our system at the UNESCO world heritage site *Zeche Zollverein*.

This work was partially funded by the European Commission in the FP7 project EuRoC (grant 608849) and by the German Research Foundation (DFG) in the project Mapping on Demand (grants BE 2556/7-2 and BE 2556/8-2).

### References

- Agarwal, S., Mierle, K., and Others (2016). Ceres solver. [online] [ceres-solver.org](http://ceres-solver.org).  
 Beul, M., Krombach, N., Zhong, Y., Droeschel, D., Nieuwenhuisen, M., and Behnke, S. (2015). A high-performance MAV for autonomous navigation

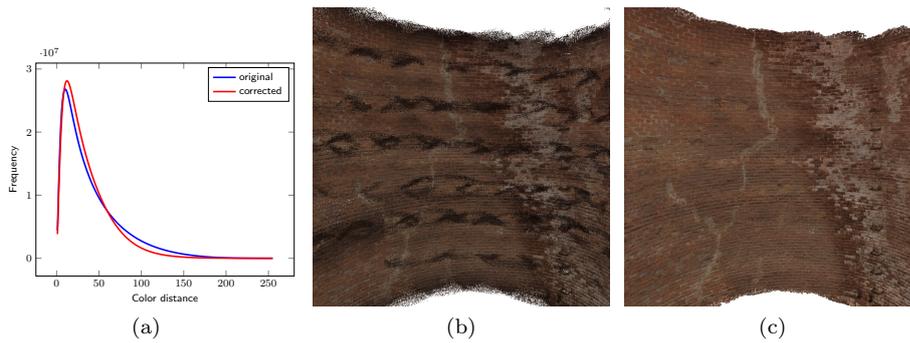


Fig. 21: Vignetting correction. (a) Histogram of the pairwise Manhattan color distances between fused points after dense reconstruction using COLMAP (Schönberger et al., 2016) on the chimney dataset. After color correction, more matching pixel pairs have a smaller color distance. (b) shows the reconstruction result with non-corrected input images and (c) the result with color corrected input images. The darker regions are reduced due to the correction at the image corners.

- in complex 3D environments. In *Int. Conf. on Unmanned Aircraft Systems (ICUAS)*.
- Blanco, J.-L. (2010). A tutorial on SE(3) transformation parameterizations and on-manifold optimization. Technical report, University of Malaga.
- Burri, M., Nikolic, J., Hürzeler, C., Caprari, G., and Siegwart, R. (2012). Aerial service robots for visual inspection of thermal power plant boiler systems. In *Proc. of Int. Conf. on Applied Robotics for the Power Industry (CARPI)*.
- Chan, B., Guan, H., Jo, J., and Blumenstein, M. (2015). Towards UAV-based bridge inspection systems: A review and an application perspective. *Structural Monitoring and Maintenance*, 2(3):283–300.
- Droeschel, D., Stückler, J., and Behnke, S. (2014). Local multi-resolution representation for 6D motion estimation and mapping with a continuously rotating 3D laser scanner. In *Int. Conf. on Robotics and Automation (ICRA)*.
- Engel, J., Usenko, V., and Cremers, D. (2016). A photometrically calibrated benchmark for monocular visual odometry. In *arXiv:1607.02555*.
- Fuhrmann, S., Langguth, F., and Goesele, M. (2014). MVE-a multi-view reconstruction environment. In *EUROGRAPHICS Workshops on Graphics and Cultural Heritage*.
- Furgale, P., Rehder, J., and Siegwart, R. (2013). Unified temporal and spatial calibration for multi-sensor systems. In *Int. Conf. on Intelligent Robots and Systems (IROS)*.
- Geiger, A., Ziegler, J., and Stiller, C. (2011). StereoScan: Dense 3D reconstruction in real-time. In *IEEE Intelligent Vehicles Symposium (IV)*.
- Grzonka, S., Grisetti, G., and Burgard, W. (2012). A fully autonomous indoor quadrotor. *IEEE Trans. on Robotics*, 28(1):90–100.
- Holland, P. and Welsch, R. (1977). Robust regression using iteratively reweighted least-squares. *Communications in Statistics-theory and Methods*, 6(9):813–

827.

- Horn, B. K. P. (1987). Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, 4(4):629–642.
- Houben, S., Droschel, D., and Behnke, S. (2016a). Joint 3D laser and visual fiducial marker based SLAM for a micro aerial vehicle. In *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pages 609–614. IEEE.
- Houben, S., Quenzel, J., and Behnke, S. (2016b). Efficient multi-camera visual-inertial SLAM for micro aerial vehicles. In *Int. Conf. on Intelligent Robots and Systems (IROS)*.
- Huh, S., Shim, D., and Kim, J. (2013). Integrated navigation system using camera and gimbaled laser scanner for indoor and outdoor autonomous flight of UAVs. In *Int. Conf. on Intelligent Robots and Systems (IROS)*.
- Intel Corp. (2016a). Intel and Airbus demo drone inspection of passenger airliners. <https://newsroom.intel.com/chip-shots/intel-airbus-demo-drone-inspection-of-passenger-airliners/>.
- Intel Corp. (2016b). Intel RealSense camera SR300 embedded coded light 3D imaging system with full high definition color camera—product datasheet.
- Jutzi, B., Weinmann, M., and Meidow, J. (2014). Weighted data fusion for UAV-borne 3D mapping with camera and line laser scanner. *International Journal of Image and Data Fusion*, 5(3):226–243.
- Kamel, M., Burri, M., and Siegwart, R. (2017). Linear vs nonlinear MPC for trajectory tracking applied to rotary wing micro aerial vehicles. *IFAC-PapersOnLine*.
- Meier, L., Tanskanen, P., Heng, L., Lee, G., Fraundorfer, F., and Pollefeys, M. (2012). PIXHAWK: A micro aerial vehicle design for autonomous flight using onboard computer vision. *Autonomous Robots*, 33(1-2):21–39.
- Moore, R., Dantu, K., Barrows, G., and Nagpal, R. (2014). Autonomous MAV guidance with a lightweight omnidirectional vision sensor. In *Int. Conf. on Robotics and Automation (ICRA)*.
- Mur-Artal, R., Montiel, J., and Tardos, J. D. (2015). ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Trans. on Robotics*, 31(5):1147–1163.
- Nex, F. and Remondino, F. (2014). UAV for 3D mapping applications: A review. *Applied Geomatics*, 6(1):1–15.
- Nieuwenhuisen, M., Droschel, D., Beul, M., and Behnke, S. (2016). Autonomous navigation for micro aerial vehicles in complex GNSS-denied environments. *Journal of Intelligent and Robotic Systems*, 84(1):199–216.
- Nieuwenhuisen, M., Quenzel, J., Beul, M., Droschel, D., Houben, S., and Behnke, S. (2017). ChimneySpector: Autonomous MAV-based indoor chimney inspection employing 3D laser localization and textured surface reconstruction. In *Int. Conf. on Unmanned Aircraft Systems (ICUAS)*.
- Nieuwenhuisen, M., Schadler, M., and Behnke, S. (2013). Predictive potential field-based collision avoidance for multicopters. In *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. (ISPRS)*, volume XL-1/W2, pages 293–298.
- Nikolic, J., Rehder, J., Burri, M., Gohl, P., Leutenegger, S., Furgale, P. T., and Siegwart, R. (2014). A synchronized visual-inertial sensor system with FPGA pre-processing for accurate real-time SLAM. In *Int. Conf. on Robotics and Automation (ICRA)*.

- Olson, E. (2011). AprilTag: A robust and flexible visual fiducial system. In *Int. Conf. on Robotics and Automation (ICRA)*.
- Ortiz, A., Bonnin-Pascual, F., and Garcia-Fidalgo, E. (2014). Vessel inspection: A micro-aerial vehicle-based approach. *Journal of Intelligent and Robotic Systems*, 76(1):151–167.
- Özaslan, T., Loianno, G., Keller, J., Taylor, C. J., Kumar, V., Wozencraft, J., and Hood, T. (2017). Autonomous navigation and mapping for inspection of penstocks and tunnels with MAVs. *IEEE Robotics and Automation Letters (RAL)*, 2(3):1740–1747.
- Park, J. and Kim, Y. (2014). 3D shape mapping of obstacle using stereo vision sensor on quadrotor UAV. In *AIAA Guidance, Navigation, and Control Conf.*
- Quenzel, J., Rosu, R. A., Houben, S., and Behnke, S. (2017). Online depth calibration for RGB-D cameras using visual SLAM. In *Int. Conf. on Intelligent Robots and Systems (IROS)*.
- Rusu, R. B. and Cousins, S. (2011). 3D is here: Point Cloud Library (PCL). In *Int. Conf. on Robotics and Automation (ICRA)*.
- Schmid, K., Lutz, P., Tomic, T., Mair, E., and Hirschmüller, H. (2014). Autonomous vision-based micro air vehicle for indoor and outdoor navigation. *J. of Field Robotics*, 31(4):537–570.
- Schönberger, J. L. and Frahm, J.-M. (2016). Structure-from-motion revisited. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.
- Schönberger, J. L., Zheng, E., Pollefeys, M., and Frahm, J.-M. (2016). Pixel-wise view selection for unstructured multi-view stereo. In *European Conf. on Computer Vision (ECCV)*.
- Tomić, T., Schmid, K., Lutz, P., Domel, A., Kassecker, M., Mair, E., Grix, I., Ruess, F., Suppa, M., and Burschka, D. (2012). Toward a fully autonomous UAV: Research platform for indoor and outdoor urban search and rescue. *Robotics Automation Magazine, IEEE*, 19(3):46–56.
- Tripathi, A., G Raja, R., and Padhi, R. (2014). Reactive collision avoidance of UAVs with stereovision camera sensors using UKF. In *Int. Conf. on Advances in Control and Optimization of Dynamical Systems (IFAC)*.
- Zhou, Q.-Y. and Koltun, V. (2014). Color map optimization for 3D reconstruction with consumer depth cameras. *ACM Trans. Graph.*, 33(4):155:1–155:10.