

Dronument: System for Reliable Deployment of Micro Aerial Vehicles in Dark Areas of Large Historical Monuments

Pavel Petráček¹, Vít Krátký¹, and Martin Saska¹

Abstract—This letter presents a self-contained system for robust deployment of autonomous aerial vehicles in environments without access to global navigation systems and with limited lighting conditions. The proposed system, application-tailored for documentation in dark areas of large historical monuments, uses a unique and reliable aerial platform with a multi-modal lightweight sensory setup to acquire data in human-restricted areas with adverse lighting conditions, especially in areas that are high above the ground. The introduced localization method relies on an easy-to-obtain 3-D point cloud of a historical building, while it copes with a lack of visible light by fusing active laser-based sensors. The approach does not rely on any external localization, or on a preset motion-capture system. This enables fast deployment in the interiors of investigated structures while being computationally undemanding enough to process data online, onboard an MAV equipped with ordinary processing resources.

The reliability of the system is analyzed, is quantitatively evaluated on a set of aerial trajectories performed inside a real-world church, and is deployed onto the aerial platform in the position control feedback loop to demonstrate the reliability of the system in the safety-critical application of historical monuments documentation.

Index Terms—Aerial Systems: Applications, Aerial Systems: Perception and Autonomy, Localization

I. INTRODUCTION

IN recent years, massive advances have been made in the technology of aerial vehicles capable of vertical landing and takeoff, in terms of control, reliability, and autonomy. These multirotor vehicles, which we will refer to as Micro Aerial Vehicles (MAVs), have become extremely popular for their flexibility, diversity, and potential for applicability and amusement. The broad application spectrum of MAV systems ranges from 3-D mapping and deployment in search & rescue scenarios to wildlife & nature conservation.

This letter presents a unique self-localization system for a specialized use of MAV teams - autonomous documentation of historical monuments - derived from the interest expressed by end-users with expertise in restoration and conservation. The

Manuscript received: September 10, 2019; Revised December 11, 2019; Accepted January 9, 2020.

This paper was recommended for publication by Editor Jonathan Roberts upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by project no. DG18P02OVV069 in program NAKI II, by CTU grant no. SGS17/187/OHK3/3T/13, and by the Grant Agency of the Czech Republic under grant no. 17-16900Y.

The authors are with the Faculty of Electrical Engineering, Czech Technical University in Prague, 166 36 Prague 6, Czech Republic {pavel.petracek|vit.kratky|martin.saska}@fel.cvut.cz.

Digital Object Identifier (DOI): see top of this page.

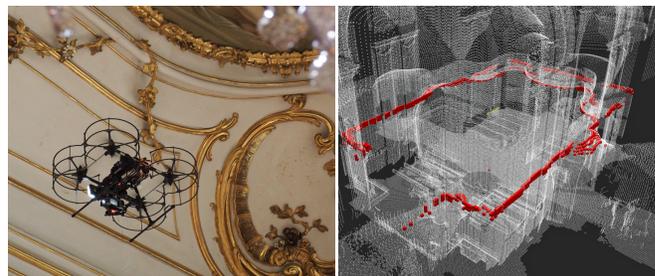


Fig. 1: Demonstration of the presented system during documentation of the parliamentary hall of the UNESCO castle in Kroměříž, Czech Republic, with an example of onboard sensory data registration into the object map

current procedure used during regular studies for restoration works requires a large scaffold to be constructed in order to monitor the condition of a building and its artifacts. An MAV platform can supply the same documentation and inspection techniques as those provided by the experts, but in locations unreachable by people except with the use of a large and expensive scaffolding installation, or in locations which had never been documented before during an initial survey. The MAV platform significantly speeds up the duration and significantly scales down the cost of the restoration works, and offers data acquisition from previously impossible angles and unreachable locations.

The proposed system is designed for deployment in historical monuments, such as ancient or modern, war-damaged, dilapidated or restored cathedrals, chapels, churches, mausoleums, castles, and temples with dimensions varying from small chapels up to large cathedrals. The deployment of robots in these operational environments is a challenging task due to the absence of a global navigation satellite system (GNSS), the adverse lighting conditions, and numerous other challenges summarized later on in Sec. III. An aerial system that handles all the challenges has to provide exceptional robustness, which we propose to achieve by introducing a precise model-based control approach, reliable real-time state estimation, a high level of sensor & actuator redundancy, and feasible mission planning & navigation.

In this letter, we also address in detail the problem of real-time state estimation acting as a state observer for MAV control and mission navigation modules in the tackled *GNSS-denied* environments of large historical buildings. The proposed system relies on a lightweight sensory setup composed of a 2-D laser-scanner and two point-distance rangefind-

ers, and on a map of a historical site pre-generated in the form of a 3-D point cloud provided by a terrestrial laser scanner (TLS). Our laser-inertial approach to indoor localization fuses an onboard IMU and a locally refined global state estimation, while it processes the data onboard an MAV and estimates the global state in real-time.

A. Related Work

Until now, documentation of interiors and exteriors of buildings and facilities has generally been performed manually by generating a 3-D site model using a TLS, together with scan registration post-processing [1] or photogrammetry [2] requiring geo-reference information. The emergence of automation procedures has speeded up the scanning processes. Examples are the *Zebedee* [3] and the *LIPS* [4] hand-held mobile 3-D laser mapping systems, and even ground robot mapping approaches [5].

Using aerial systems introduces the advantage of rapid documentation even in human-unreachable locations. These systems are being employed for documentation purposes [2], [6], but most of them are deployed manually outdoors and require GNSS to obtain geo-referenced data.

Although the lack of GNSS can be bypassed by exploiting a preset external localization system capable of a high accuracy localization, this approach is not scalable for documentation of large structures with limited access time. Other approaches for *GNSS-denied* localization include visual-inertial simultaneous localization and mapping (VI-SLAM), whose recent advances are well described in [7] and [8]. A mono- or stereo-camera SLAM is a thoroughly studied problem for an MAV, due to the lightweight of ubiquitous cameras. One example of a system that attempts to integrate online SLAM is the Open Vision Computer [9], which is an embedded off-the-shelf FPGA module that handles a visual SLAM independent of other onboard subsystems. The state-of-the-art vision-based ORB-SLAM2 [10] was tested in real conditions with lighting conditions similar to the desired environments. However, it was found to be ineffective, and it was disregarded for reasons described in Sec. III. The lighting issues motivated the development of the system presented here, which works under the specified conditions.

The authors in [11]–[13] have presented applications, which share a considerable number of common characteristics with indoor documentation of historical structures. A laser-inertial system for inspecting chimneys is presented in [11], while a laser-visual-inertial system for inspecting penstocks and tunnels is presented in [12]. In comparison to our application, environments tackled in [11] and [12] are well-structured and homogeneous for onboard sensors, which makes it possible to tune the system for these specific conditions. On the contrary, our task requires a much higher level of complexity. The authors in [13] focused on inventory applications in warehouses. Their laser-visual-inertial setup is suited for fast flights in complex dynamic environments in order to speed up a periodic inventory audit. However, our targeted scenario requires minimalist MAV dimensions, and slower and more accurate operation with respect to a variable onboard payload. In contrast to our lightweight sensory setup, the systems

in [11]–[13] employ a heavyweight 3-D lidar. Moreover, the systems in [12] and [13] fuse visual information from a set of onboard cameras, which is not suitable for the tackled environments with adverse lighting conditions.

Apart from that, only one work using MAV in the context of documentation of historical buildings has been found [14]. This work evaluates the performance of several SLAM and SFM methods during a 3-D model reconstruction of a single historical site. However, the authors of [14] perform only an offline evaluation of their methods on an outdoor aerial trajectory and do not deploy these methods in *GNSS-denied* environments nor in a position control loop of an MAV.

Documentation systems often extend their applications with a TLS to assist with the digital preservation of the scanned sites [15], which we likewise propose in our system architecture to boost the robustness of the system. In [16], map-based 3-D Monte Carlo localization (MCL) using an RGB-D camera provides a global state estimate. In contrast to this manuscript, our method utilizes a 2-D scanner instead of an RGB-D camera. This provides planar 360° information, making it independent from orientation. Our method goes further by refining the global estimate on a local map by a scan matching technique to yield faster and more accurate results. The authors of manuscripts [17], [18] fuse scan matching output, IMU, and a down-oriented rangefinder. Our proposed method extends the setup with global initialization and fusion of an up-oriented rangefinder. The importance of the up-oriented measurements rises significantly during flights over heterogeneous objects (church benches), where vertical estimate exploits the homogeneous nature of ceilings.

B. Contributions

This letter addresses problems of the deployment of aerial systems in the safety-critical application of historical monument documentation. The stability of the system originates from carrying out tests in real-world historical objects in the course of two years of a research and culture project in close cooperation with the National Heritage Institute of the Czech Republic. The insights into developments for real-world deployment presented here tackle the motivations and constraints of the highly challenging environments guided by end-users from outside the robotic community. The main contributions of this letter are:

- (i) It introduces a unique, highly reliable system for deployment in environments with low feature density and atrocious lighting conditions.
- (ii) It develops a unique hardware and software aerial platform designed in close consultation with restorers and conservationists, using experience from deployment of the system in numerous individual historical objects.
- (iii) It presents and shares the experience of what we believe to be the most comprehensive project in the field of autonomous documentation of historical monuments by an aerial system.
- (iv) It presents a robust light-independent localization system for interiors of historical buildings relying on 2-D lidar as its primary sensor.

The 3-D localization offers precise full 6 degrees-of-freedom estimation, providing fast and robust state estimation integrated into a feedback loop of an MAV position control. Based on a quantitative analysis evaluated on aerial ground-truth data in Sec. VI, our approach reaches persistent RMSE precision below 0.23 m. The drift-free system that is presented yields greater accuracy than map-based localization for autonomous cars [19] and comparable accuracy to a drift-prone method [20] employing a 3-D scanner on a ground vehicle.

II. MOTIVATION

The Dronument (Drone & Monument) project sets out to deploy MAVs for autonomous data acquisition in human-unreachable areas. The self-contained system presented here can be deployed in three modes (manual, semi-autonomous, and fully autonomous), as allowed by the heritage institute and/or the superintendent of the structure. These modes are specified as follows:

- (i) *manual*: a human operator controls all aspects of the flight using an operating transmitter, while the MAV is autonomously localized in the environment to associate gathered data with the 3-D map,
- (ii) *semi-autonomous*: a human operator commands the flight, while onboard systems control the sensory data acquisition and provide control feedback with respect to obstacles in a 3-D neighborhood, and
- (iii) *autonomous*: a human only specifies objects-of-interest (OoI) for documentation, and the system handles each stage of the entire mission - takeoff, stabilization & control, localization, navigation, trajectory optimization, data acquisition, and landing.

In addition to the deployment of a single MAV, the system is prepared for use in cooperative multi-MAV scenarios, as required for some documentation tasks. Typical non-invasive documentation consists of a multiple spectrum survey to obtain specific information valuable for various restoration purposes. For example, the use of different spectra contributes to more precise dating of paintings, as the glow of pigment combinations is unique to a certain period. Examples of single (S) and cooperative (C) tasks are:

- *Direct lighting (S)**: lighting of the scene from an onboard light with the illumination axis collinear with the optical axis of the camera.
- *Reflectance Transformation Imaging (C)*: a photographic technique for capturing the shape of a surface and the color of an object by combining photographs of the objects taken from a semi-static camera on an MAV under various illumination provided by a different MAV [21].
- *Three-point & strong-side lighting (C)**: filming techniques [22] in which 1-3 sources of light are used in different locations relative to the optical axis of the camera. In our previous work [23], a Model Predictive Control (MPC) approach is proposed for controlling a formation of MAVs with respect to the lighting techniques during an aerial deployment of cooperative teams in this task.
- *Radiography & UV screening (C)*: a method for viewing the internal structure of an object (e.g., a statue) by exposing it to X-ray or UV radiation (emission source

onboard the first MAV) captured behind or in front of the object by a detector (the second MAV).

- *3-D reconstruction (S/C)**: a method for aggregating the shape and the appearance of an object by combining laser- and/or vision-based information into a 3-D model.
- *Photogrammetry (S/C)**: a method for extracting measurements of real objects from photographs.

Examples of the tasks marked with (*) can be found within the additional multimedia materials available in [24].

III. EXPERIENCE GAINED

Over the last two years, more than 10 objects (mainly in Moravia, Czech Republic) were documented during the ongoing development phase of the presented system. Outputs of these documentation deployments supplied restorers and conservationists with valuable information in state assessment of multiple artifacts within the structures during the initial survey phase. Although some of the documented structures are shown in Figures 1, 2, and 5, the complete list, together with additional multimedia materials, can be found in [24]. During the experimental phase, many lessons from the robotic as well as the restoration point of view have been learned. The acquired experiences for objects of various sizes, shapes, and structures have influenced the system throughout the development and are herein shared.

The indoor surveys are conducted in the close vicinity of heavy-structure buildings. As a consequence, either GNSS is not available at all, or the system is not reliable enough, leading to *GNSS-denied* operations. In order to overcome the absence of GNSS, a local localization system must be used. This is true even when exteriors are being documented, with the intention to document facades in their close proximity in order to capture details of artifacts from various points of view (see exteriors documentation in [24]).

Insufficient *lighting conditions* in a surveyed object is an impediment for two main reasons. First, it degrades the performance of vision-based odometry and SLAM systems, which have been heavily researched over the last three decades. Second, it lowers the quality of the photographs taken in the visible spectrum, as they require decent lighting of the scene. These two issues motivated research in the Dronument project, leading to the development of a novel robust localization system (see Sec. V-B) purposely designed for autonomous flying in specific environments of this kind.

Experience has shown that vision-based localization is limited also by *feature extraction* shaped by two main characteristics. First, it is a common occurrence to fly along protracted segments of white wall lacking any visual features at all. Second, old religious buildings include extensive symmetric and repetitive visual patterns, such as grid flooring and artistic elements. A common example of such an artistic element are repetitive ledge supports shown on the right side of Fig. 2. Together with the lighting conditions, these considerations make most vision-based systems ineffective.

The use of a prior knowledge in the form of a *global map* obtained prior to the deployment of an MAV is extremely beneficial for three main reasons. First, it facilitates the robotic problem, supplies additional robustness to the system and

supports system reliability by serving as a baseline. Second, it yields an opportunity to associate captured onboard data with a 3-D map, which provides well-arranged data output for the end-user. Third, the visualization of the flight plan in a 3-D model is comprehensible for everyone - robotic experts, restorers, filmmakers - and it is necessary for confirmation purposes by an aviation authority, the heritage institute, and/or the administrator of the structure.

The proposed use of an MAV requires us to consider the *diversity of environments* in which it will operate. These environments contain distinct features - cluttered spaces, symmetric blueprints, balconies, stairs, glass windows, vault ceilings, and hanging strings. This forced the system design to include sensor and actuator redundancy, the use of a global map, and mechanical protection of the propellers. The test flights showed that the 2-D lidar that was used is ineffective for detecting thin obstacles, such as chandelier ropes and lighting cables. For this specific reason, a 3-D camera is employed to detect these obstacles in front of an MAV.

Even when MAVs are deployed in historical buildings, the presence of wind gusts ascribed to the stack effect (opened windows, doors) is non-negligible. To maximize the robustness of the system, particular emphasis must be laid on handling the *aerodynamic influence* of the MAV itself, and the wind gusts. First, the control subsystem must be resistant to these aerodynamic disturbances, in order to provide control stability (we rely on low-level stabilization designed in our team for flying in demanding desert conditions [25]). Second, perception modules must maintain their sensory properties when flying in low lighting conditions and in dust clouds, which originate when previously settled dust starts whirling.

An obvious constraint arises from the particular *historical value* of the surveyed objects and their invaluable character, which is the key reason for undertaking the documentation and restoration work. To avoid potential damage to fragile artifacts or to their surroundings at all costs, the MAV has to maximize the reliability and robustness attributes of the hardware and the software systems. The introduction of redundancy, system fault detectors, and safety procedures is a critical requirement for such a safety-challenging environment.

Last but not least, the system has to provide high *payload modularity*, in order to tackle all the documentation tasks in various environments. During documentation works, it may be necessary to document vertical walls and also the ceiling, even with multiple types of payload. For this reason, the hardware platform must be capable of rapidly changing the payload, its position, and the stabilization axes, in order to provide an effective solution to these end-user requirements.

IV. AERIAL PLATFORM

A specialized multi-MAV platform was developed to survey dark areas of historical monuments. This system is distributed to the primary MAV carrying the mission payload and a couple of lighter MAVs carrying additional mission equipment, such as lights. The primary MAV, shown in Fig. 2, is designed to minimize its dimensions, since the task assumes flights in narrow passages close to obstacles. Simultaneously, it is designed to maximize its payload weight capacities, since the payload is defined by the end-users, is interchangeable, and

often cannot be optimized for employment on aerial platforms. In its default configuration, the primary MAV carries an autopilot, an onboard computer, a down-oriented camera, two laser rangefinders, a 2-D laser scanner, a 3-D camera for obstacle detection, and the payload (an onboard light and a 2-axes stabilization hinge with a professional camera with its lens and a first-person view (FPV) system). The total weight is 5 kg, with a payload weight of 1.5 kg. To provide an extra level of safety, the platform is equipped with a mechanical propeller guard system to isolate the propellers from the external environment. The lighter MAVs carry a light with a digitally-controlled pitch angle (a degree-of-freedom required for formation trajectory optimization in [23]), light intensity, and color warmth.



Fig. 2: An application-tailored MAV carrying sensory and mission equipment during documentation of the Klein mausoleum in Sobotín, Czech Republic

V. SYSTEM ARCHITECTURE

The overall system architecture is composed of four main subsystems, which are hereafter described. The high-level pipeline of the system is outlined in Fig. 3.

A. Control Architecture

An MAV disturbance-resistant control pipeline was developed in the previous work of our group [25]–[28]. Beyond others, the MPC-based approach [26] was tested in the harsh environment of the desert in the United Arab Emirates during the MBZIRC 2017 competition, where it outperformed 147 registered teams [25], [28]. The system architecture presented in Fig. 3 is based on experience gained during this competition, which posed similar requirements of reliability, and resistance to wind disturbance and omnipresent dust. However, the task presented here is considerably different due to absence of GNSS and the density of the obstacles, and therefore goes beyond the work presented in [25] and [28].

B. Localization

For 6 degrees-of-freedom state estimation, we propose to rely on three laser-based sensors. First, a rigidly-mounted lightweight 2-D scanner produces data in the horizontal plane of the vehicle. Second, two point-distance laser sensors (rangefinders) measure the distance to the ground and ceiling objects. Together with an onboard IMU and a sparse 3-D map, the laser-inertial approach manages to estimate the global position and the attitude in light-independent conditions and without any heavyweight sensory equipment. The whole localization pipeline is summarized in Fig. 4, and will be described in detail in this section.

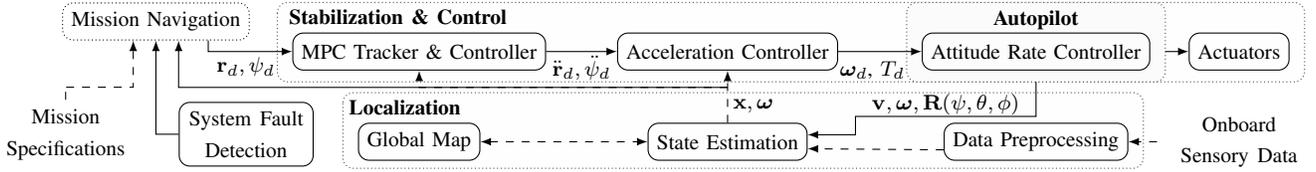


Fig. 3: High-level system pipeline of a single MAV. The stabilization & control pipeline [26] takes reference trajectory \mathbf{r}_d, ψ_d (points sequence of the desired 3-D position and yaw) as a setpoint for the MPC in the MPC tracker, which outputs a command $\ddot{\mathbf{r}}_d, \ddot{\psi}_d$ for the acceleration tracking SO(3) controller. The acceleration controller produces the desired angular rate ω_d and thrust reference T_d for the embedded attitude rate controller. The localization pipeline is described in detail in Sec. V-B.

1) *Global Map*: The localization system proposed in this work is designed to operate with a partially-known map registered from multiple 3-D intensity/color point clouds produced by a terrestrial laser scanner (TLS). The necessary granularity of the map (10 cm in our experiments) depends on the onboard computational resources and on the structural complexity of the documented building. Nonetheless, a modern TLS is capable of producing a scan with millimeter-level granularity, which makes the map subject to data reduction.

A raw map is processed by a set of filters (median filter, outlier rejection, uniform sampling) and is transformed to octree representation in order to employ optimized map operations, such as node traversal, integration of sensor measurements, data access, and tree node queries. During the preprocessing phase, artificial ground data is injected into the map to cope with missing data due to occlusions during scanning. Assuming that the ground is a cavity-free plane, the ground data is augmented by a set of points uniformly sampled from a plane. The parameters of this sampled plane are obtained from fitting it on a set of points withdrawn from the undermost parts of the available map using the RANSAC algorithm. Missing non-ground data is augmented in midair during a mission by the map refinement module.

A smaller structure requires < 5 scans, where a single full-dome scan (360° horizontal and 300° vertical field of view) takes approximately 3 min. During a field operation, there is enough time to produce a map of the object during the preparation of the equipment required for the mission. Example of a map is shown on the right side of Fig. 1 and in Fig. 7.

2) *State Estimation*: An MAV is assumed to have first-order dynamics for a short period of time during hovering and slow flights with negligible tilts (these flight characteristics are required in the confined areas in historical monuments for safety reasons). The linear stochastic discrete state-space model is used as

$$\mathbf{x}_{[k]} = \mathbf{A}_{[k]}\mathbf{x}_{[k-1]} + \mathbf{B}_{[k]}\mathbf{u}_{[k]} + \eta_{[k]}, \quad (1)$$

$$\mathbf{z}_{[k]} = \mathbf{H}_{[k]}\mathbf{x}_{[k]} + v_{[k]}. \quad (2)$$

The state $\mathbf{x}_{[k]}$, system input $\mathbf{u}_{[k]}$, measurements $\mathbf{z}_{[k]}$ and random noises of the system at time k are given as

$$\mathbf{x}_{[k]} = \left(\mathbf{p}_{[k]}^T, \boldsymbol{\Omega}_{[k]}^T \right)^T, \quad \mathbf{u}_{[k]} = \left(\mathbf{v}_{[k]}^T, \boldsymbol{\omega}_{[k]}^T \right)^T, \quad (3)$$

$$\mathbf{z}_{[k]} = \hat{\mathbf{x}}_{[k]}, \quad \eta_{[k]} \sim \mathcal{N}(0, \mathbf{Q}_{[k]}), \quad v_{[k]} \sim \mathcal{N}(0, \mathbf{R}_{[k]}), \quad (4)$$

where the state $\mathbf{x}_{[k]}$ is comprised from global position $\mathbf{p}_{[k]} = (x_{[k]}, y_{[k]}, z_{[k]})^T$ and attitude $\boldsymbol{\Omega}_{[k]} = (\psi_{[k]}, \theta_{[k]}, \phi_{[k]})^T$ consisting of the yaw, pitch, and roll angles; $\mathbf{v}_{[k]}$ is the linear

and $\boldsymbol{\omega}_{[k]}$ is the angular velocity of the IMU frame; $\hat{\mathbf{x}}_{[k]}$ is the measured global state; and $\mathbf{S}_{[k]}$ and $\mathbf{Q}_{[k]}$ are the covariance matrices of the process and the measurement noise at time k . The state-transition model $\mathbf{A}_{[k]}$, the control-input model $\mathbf{B}_{[k]}$, the observation model $\mathbf{H}_{[k]}$, and the covariance matrices $\mathbf{S}_{[k]}$ and $\mathbf{Q}_{[k]}$ are defined as

$$\mathbf{A}_{[k]} = [\mathbf{I}_{6 \times 6}], \quad \mathbf{Q}_{[k]} = \Delta t_{[k]} \begin{bmatrix} \boldsymbol{\Sigma}_{6 \times 6}^{mcl} & \mathbf{0}_{6 \times 6} \\ \mathbf{0}_{6 \times 6} & \boldsymbol{\Sigma}_{6 \times 6}^{icp} \end{bmatrix}, \quad (5)$$

$$\mathbf{B}_{[k]} = \Delta t_{[k]} \begin{bmatrix} \mathbf{R}(\psi_{[k]}, \theta_{[k]}, \phi_{[k]})_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{R}(\psi_{[k]}, \theta_{[k]}, \phi_{[k]})_{3 \times 3} \end{bmatrix},$$

$$\mathbf{S}_{[k]} = \Delta t_{[k]} \begin{bmatrix} \sigma_p^2 \mathbf{I}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \sigma_\Omega^2 \mathbf{I}_{3 \times 3} \end{bmatrix}, \quad \mathbf{H}_{[k]} = [\mathbf{I}_{6 \times 6} \quad \mathbf{I}_{6 \times 6}]^T,$$

where $\mathbf{I}_{n \times n} \in \mathbb{R}^{n \times n}$ is an identity matrix and $\mathbf{0}_{n \times n} \in \mathbb{R}^{n \times n}$ is an empty matrix, $\boldsymbol{\Sigma}_{6 \times 6}^* \in \mathbb{R}^{6 \times 6}$ is the covariance matrix of the global and local state estimation, $\Delta t_{[k]} = t_{[k]} - t_{[k-1]}$ is the time elapsed since the last KF update, and $\mathbf{R}(\psi_{[k]}, \theta_{[k]}, \phi_{[k]}) \in \mathbb{R}^{3 \times 3}$ is the 3-D attitude. The presence of the rotation matrix in the control-input model copes with the differing global and IMU frames. The input of the system consists of inertial measurements coming at 100 Hz, and observations are produced by two estimation processes running in parallel, incoming at 5 and 20 Hz, which will be described below. The output of the KF correction step is equal to the output of the whole localization process.

3) *Monte Carlo Localization*: The configuration space of a robot inside an a-priori known map of a historical object is immense. This restricts the straight registration of sensory data to the extensive map due to the unknown initial conditions, which MCL provides in the form of a slow global state estimate. Concisely, MCL determines the posterior probability $p(\mathbf{x}|\mathbf{y}, \mathbf{u})$ of an unobservable state \mathbf{x} given sensor observations \mathbf{y} and control inputs \mathbf{u} by computing it on the state space subset in the form of hypotheses, yielding an approximation of the probability density function. The posterior probability can be obtained by employing the Bayes filter, which recursively computes the previous equation in the form of a belief $Bel(\mathbf{x})$ of the posterior probability as

$$Bel(\mathbf{x}) = \eta p(\mathbf{y}|\mathbf{x}) \int p(\mathbf{x}|\hat{\mathbf{x}}, \mathbf{u}) Bel(\hat{\mathbf{x}}) d\hat{\mathbf{x}}, \quad (6)$$

where η is a normalization constant. The derivation of the equation holds under the initial condition $p(\mathbf{x}_0) = p(\mathbf{x}_0|\mathbf{y}_0, \mathbf{u}_0)$ and Markov independence assumptions.

Motion model: An odometry-based model for 2-D mobile robots employing the dead-reckoning principle is expanded to 3-D. In comparison with [29], our application requires slow

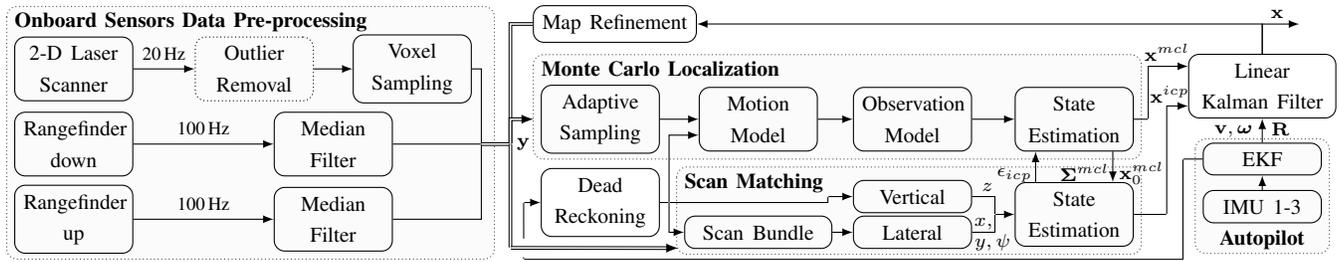


Fig. 4: Workflow diagram of the state estimation process. The laser-inertial pipeline fuses a global map, onboard data from multiple laser-based sensors (a 2-D horizontal scanner and an up- and down-oriented rangefinder), and 3 IMUs to produce state estimation \mathbf{x} . The IMUs are fused by an Extended Kalman Filter outputting linear \mathbf{v} and $\boldsymbol{\omega}$ angular velocity, and attitude $\mathbf{R}(\psi, \theta, \phi)$, where ψ, θ, ϕ are the Euler angles yaw, pitch, and roll. A decoupled local state refinement employing Iterative Closest Point is initialized after a global state estimate \mathbf{x}_0^{mcl} is provided by MCL. The validity of the state estimation processes is observed with respect to the MCL estimate covariance $\boldsymbol{\Sigma}^{mcl}$ and the absolute mean square error ϵ_{icp} of the scan matching.

movement of an airborne vehicle up to 0.5 m s^{-1} , making the variations in roll and pitch negligible and therefore allowing us to reduce the kinematic DoF to 4 (3-D position and heading).

Adaptive sampling: To improve performance, KLD-sampling [30] estimates the sufficient number of hypotheses M by bounding the error introduced by the sample-based representation of the MCL. The estimate is based on drawing from a discrete distribution with p different bins, and for

$$M \approx \frac{p-1}{2\epsilon} \left(1 + \frac{2}{9(p-1)} + \sqrt{\frac{2}{9(p-1)} z_{1-\delta}} \right)^3, \quad (7)$$

guarantees with probability $1 - \delta$ that the Kullback–Leibler distance between the maximum likelihood estimate (MLE) and the true distribution is less than ϵ , with $z_{1-\delta}$ being the upper $1 - \delta$ quantile of the normal $\mathcal{N}(0, 1)$ distribution.

To prevent convergence to an erroneous local minimum, a subset of hypotheses with the lowest weights is replaced in each resampling step with a dynamic-size set of new randomly generated hypotheses over the whole sampling space and a static-size set of new hypotheses matching the position of the latest state estimate with randomly sampled heading. The ratio of newly injected hypotheses is regulated by Augmented-MCL [31], which compares the short-term and long-term likelihood of observations as

$$M_{new} = M \max \left\{ 0, 1 - \frac{w_{fast}}{w_{slow}} \right\}, \quad (8)$$

where $w_{slow} = w_{slow} + \alpha_{slow}(\bar{w} - w_{slow})$ and $w_{fast} = w_{fast} + \alpha_{fast}(\bar{w} - w_{fast})$ for \bar{w} being the weighted average over the whole set of hypotheses, and $0 \leq \alpha_{slow} \ll \alpha_{fast}$ are the decay rates.

4) *Local Refinement:* To obtain precise and fast localization, local map registration is performed in a decoupled manner. The decoupling emerges from the sensory setup due to the vast difference between the data volume in the horizontal and in the vertical plane. In contrast to the vertical plane, where only two point-distance measurements are obtained, the horizontal sensor generally provides a greater number of samples (e.g., 16K samples per second for RPLIDAR A3), which needs to be reduced. The vast difference in the data volumes requires decoupling, otherwise the horizontal estimation would heavily outweigh the vertical estimation.

Lateral estimation employs a variant of the Iterative Closest Point (ICP) algorithm. Given a reference set of points \mathbf{P} and a target set of points \mathbf{Q} , the optimization process finds a transformation \mathbf{T} , which minimizes the weighted point-to-point error metric

$$E(\mathbf{T}) = \frac{1}{N} \sum_{i=1}^N (w_i \|\mathbf{T}\mathbf{p}_i - \mathbf{q}_i\|^2) \quad (9)$$

over the set of N correspondence pairs $(\mathbf{p}_i, \mathbf{q}_i, w_i)$, $\mathbf{p}_i \in \mathbf{P}$, $\mathbf{q}_i \in \mathbf{Q}$, $w_i \in \mathbb{R}$, $\forall i \in \langle 1, N \rangle$, $N \in \mathbb{Z}$. An initial solution to Eq. 9 is given by the dead-reckoning principle. Determining of the correspondence pairs involves closest distance pairing and a median filter, duplicate reference matches, and RANSAC-based pairs rejectors. Implementation of the ICP is based on the Point-Cloud library [32].

The reference scan \mathbf{P} is obtained onboard from a 2-D laser scanner, and its data are prepared according to Fig. 4. To provide improved robustness, a short history of the measurements is bundled together using short-time IMU-based dead-reckoning odometry, and is used as the reference scan \mathbf{P} for 3-D scan matching. An example of the scan bundle, registered into a map in the form of a 3-D point cloud, is displayed on the right side of Fig. 1. The target scan $\mathbf{Q} = \mathbf{Q}_{pla} \setminus \mathbf{Q}_{occ}$, $\mathbf{Q}_{pla} \in \mathbf{Q}_{map}$, $\mathbf{Q}_{occ} \in \mathbf{Q}_{pla}$, is derived from \mathbf{Q}_{map} and state estimate from the previous time step $\mathbf{x}_{[k-1]}$. The subset \mathbf{Q}_{pla} represents points of the map located in between two planes parallel to the x - y plane of the 2-D sensor frame at distance $\pm d_{pla}$ on the z axis of the same frame. The subset $\mathbf{Q}_{occ} \in \mathbf{Q}_{pla}$ represents all visually occluded points for which the linear path of a laser beam from a sensor position (rigidly defined by $\mathbf{x}_{[k-1]}$) to $\mathbf{q} \in \mathbf{Q}_{occ}$ is collision-free. A ray-casting algorithm, implemented over an octree representation of the map, is employed to determine the collision status.

During *vertical estimation*, a lateral estimate of the x , y axes, an attitude estimate, and the up- and down-oriented point-distance measurements are used to define a quadratic least squares problem

$$z^* = \arg \min_{z \in \mathbb{R}} (\alpha_{\uparrow} (\hat{\mathbf{p}}(z), y_{\uparrow}^r) \|\mathbf{y}_{\uparrow}^m(\hat{\mathbf{p}}(z)) - y_{\uparrow}^r\|^2 + \alpha_{\downarrow} (\hat{\mathbf{p}}(z), y_{\downarrow}^r) \|\mathbf{y}_{\downarrow}^m(\hat{\mathbf{p}}(z)) - y_{\downarrow}^r\|^2) \quad (10)$$

to find vertical z axis position z^* , where $\hat{\mathbf{p}}(z) = (x, y, z)^T$,

y_{\bullet}^r are real sensor data, and y_{\bullet}^m are map measurements found by map ray-casting. Bear in mind that the attitude and the rigid IMU-sensor transformations are neglected here to maintain simplicity. The validity coefficients α_{\bullet} are defined as

$$\alpha_{\bullet}(\hat{\mathbf{p}}(z), y_{\bullet}^r) = \begin{cases} 0, & \text{if } y_{\bullet}^m(\hat{\mathbf{p}}(z)) \text{ or } y_{\bullet}^r \text{ is invalid,} \\ 1, & \text{otherwise.} \end{cases} \quad (11)$$

Data invalidity emerges directly from invalid sensor measurements or from the absence of a map reference. In addition, the down-oriented sensor detects dynamic obstacles, such as people or map changes, which are observable from an identifiable discrepancy between real and map-based observations. These detections likewise classify the observations as invalid. In the case of $\alpha_{\uparrow} = \alpha_{\downarrow} = 0$, the z axis prediction at time k is given as

$$z_{[k]} = z_{[k-1]} + z_{[k]}^{imu} - z_{[k-1]}^{imu}, \quad (12)$$

where z^{imu} represents the integrated z axis position derived from the IMU-based dead-reckoning odometry.

C. Mission Navigation

To maximize robustness of the system, a visibility-constrained navigation is employed such that an MAV is allowed to maneuver only to obstacle-free areas visible from a front-facing depth camera. This approach supervises lidar-based perception by a redundant check for local obstacles in the camera field-of-view. An MPC-based control for navigation and trajectory optimization for MAV formations in the documentation task is introduced in our previous work [23].

D. System Fault Detection

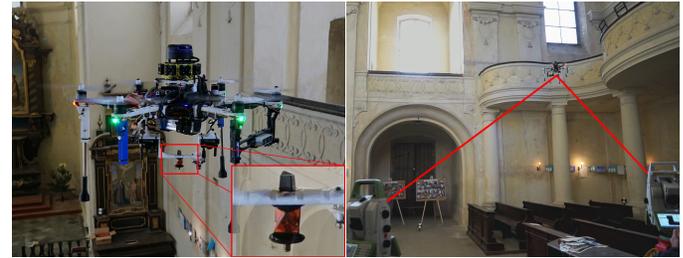
In parallel to the mission controller, a tightly coupled fault detection system supervises all aspects of the mission. That includes supervision of the sensors and battery life status, of the state estimation covariance, or of the divergence from a replanned trajectory. The whole system is implemented as a centralized high-level state machine capable of overriding the mission with an appropriate reaction to fault scenarios. Examples of these safety procedures are enforced controlled landing, trajectory execution termination, or manual take over of the control by a human operator. These safety responses can be likewise triggered by a mission operator, who is required to supervise the mission by an aviation authority.

VI. EXPERIMENTAL EVALUATION

To prove concept of the proposed method, the system was thoroughly verified in simulation (Gazebo 9 coupled with ROS Melodic), before it was deployed in position control feedback loop of an MAV. The main intention of the simulation was to estimate suitability of the developed system for deployment in safety-critical environments of historical buildings, to reduce probability of failures, and to obtain a qualitative analysis of the system behavior. Although the simulation results are omitted here due to lack of space, they can be found in [24].

A. Localization Precision Analysis

This section presents quantitative results of the localization system evaluated inside real church of St. Mary Magdalene in Chlumín, using a prototype MAV with the same sensory setup as is carried by the presented project platform. To obtain ground truth data, two Leica multi-stations were employed to track movement of the MAV equipped with the Leica GRZ101 360° Mini Prism reflector, as shown in 5a, which the stations are able to lock and track throughout 3-D space. Due to the lightweight and miniature dimensions of the particular reflector, the stations were capable to provide only the 3-D position of the reflector relative to a coordinate system of the stations at frequency of 5 Hz. The reference attitude was determined offline by ICP algorithm with parameters set to maximize accuracy. During short occlusions between a station and the target, a predicted trajectory of the target is followed in order to focus back once the occlusions disappear. Hence, the data further used as a ground truth reference contain short time period outages as the stations initialized re-locking procedure.



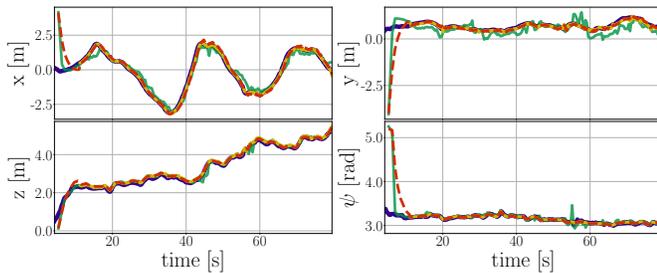
(a) Reflector-mounted platform (b) Automatic tracking demonstration

Fig. 5: An MAV platform equipped with onboard sensors and a reflector tracked by two Leica multi-stations measuring its 3-D position with precision of 1.5 mm at 5 Hz

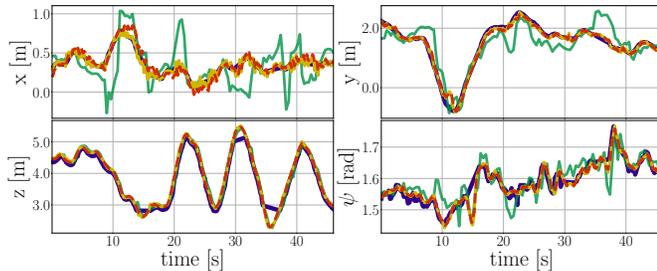
From multiple experimental flights tracked by an outer reference system, three particular trajectories are presented for which the quantitative results are given in Table I. Besides the table, outputs of the distinctive state estimation processes are outlined in Fig. 6 for the first two trials. The analysis of the localization system in real-world conditions exhibits estimation accuracy with translational RMSE less than 0.23 m during each experiment. The experiments also demonstrate minimal time delay, smoothness, and robustness of the state estimation. These attributes are important for reliable deployment as their absence could lead to destabilization of the MAV control. The proposed localization system proves to be a reliable and robust source with sufficient precision of the position estimate.

Trajectory	Trial 1	Trial 2	Trial 3
length [m]	24.055	45.812	21.163
avg linear velocity [m s^{-1}]	0.361	0.485	0.505
max linear velocity [m s^{-1}]	1.586	2.294	1.734
RMSE translation [m]	0.179	0.140	0.230
RMSE absolute orientation [$^{\circ}$]	2.381	2.460	2.747
max translation error [m]	0.385	0.522	0.594
max absolute orientation error [$^{\circ}$]	6.807	6.928	11.302

TABLE I: Quantitative results of the 3-D position and absolute yaw orientation ψ accuracy based on real data taken during real deployment in church of St. Mary Magdalene in Chlumín



(a) Trial 1: verification containing takeoff phase of the flight, where global estimation convergence and scan matching pipeline initialization is visible approx. at time 8 s and altitude of 2.300 m



(b) Trial 2: verification containing ground truth reference interruptions around 15 s, 30 s and 35 s due to visual occlusions between the multi-stations and the tracked target

Fig. 6: State variables x, y, z, ψ for ground truth (5 Hz), global (3 Hz) and local (15 Hz) localization, and fused state estimation during real deployment in church of St. Mary Magdalene in Chlumín

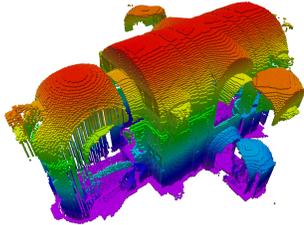


Fig. 7: Single-scan octree map of church of St. Mary Magdalene in Chlumín



Fig. 8: Documentation in church of St. Anne and St. James in Stará Voda

VII. CONCLUSION

This letter presents the first comprehensive study on the use of autonomous MAV systems as an assistive technology for documentation of historical structures. The study shares the experience obtained during developing of the technology in close cooperation with team of restorers and conservationists, and discusses challenges of a robotic deployment. The proposed approach is validated and tuned on a set of identified tasks through extensive experimental flights aimed at collecting of exploitable data from the end-users.

To provide state estimate in *GNSS-denied* environments, an application-tailored localization system is presented, which was identified as the most important and challenging task in this application. This system provides local 3-D position and attitude without access to GNSS, and with the use of laser-inertial sensory setup copes with bad lighting conditions. That makes it feasible for deployment in indoor areas high above the ground, which are characteristic for historical monuments. The presented analysis of the localization system proves it to be a reliable and robust source of information with sufficient precision, which enabled its deployment into the feedback loop of the position control system.

REFERENCES

- [1] B. Kwoczyńska *et al.*, “The Use of Terrestrial Laser Scanning in Surveying Historic Buildings,” in *BGC Geomatics*, 2016.
- [2] D. Dominić, M. Alicandro, E. Rosciano, and V. Massimi, “Multiscale documentation and monitoring of L’Aquila historical centre using UAV photogrammetry,” *ISPRS*, vol. XLII-5/W1, pp. 365–371, 2017.
- [3] R. Zlot *et al.*, “Efficiently capturing large, complex cultural heritage sites with a handheld mobile 3D laser mapping system,” *JCH*, vol. 15, no. 6, pp. 670 – 678, 2014.
- [4] P. Geneva, K. Eckenhoff, Y. Yang, and G. Huang, “LIPS: LiDAR-Inertial 3D Plane SLAM,” in *IEEE/RSJ IROS*, 2018.
- [5] D. Borrmann *et al.*, “Evaluation of Methods for Robotic Mapping of Cultural Heritage Sites,” *IFAC CESCIT*, 2015.
- [6] D. Cowley *et al.*, “UAVs in Context: Archaeological Airborne Recording in a National Body of Survey and Record,” *Drones*, vol. 2, p. 2, 2017.
- [7] M. Abouzahir *et al.*, “Embedding SLAM algorithms: Has it come of age?” *IEEE RAS*, vol. 100, pp. 14 – 26, 2018.
- [8] G. Huang, “Visual-Inertial Navigation: A Concise Review,” in *IEEE ICRA*, 2019.
- [9] M. Quigley *et al.*, “The Open Vision Computer: An Integrated Sensing and Compute System for Mobile Robots,” in *IEEE ICRA*, 2019.
- [10] R. Mur-Artal and J. D. Tardós, “ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras,” *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.
- [11] M. Nieuwenhuisen *et al.*, “ChimneySpector: Autonomous MAV-based indoor chimney inspection employing 3D laser localization and textured surface reconstruction,” in *IEEE ICUAS*, 2017.
- [12] T. Öztaşlan *et al.*, “Autonomous Navigation and Mapping for Inspection of Penstocks and Tunnels With MAVs,” *IEEE RA-L*, vol. 2, no. 3, pp. 1740–1747, 2017.
- [13] M. Beul *et al.*, “Fast Autonomous Flight in Warehouses for Inventory Applications,” *IEEE RA-L*, vol. 3, no. 4, pp. 3121–3128, 2018.
- [14] A. L. Majdik *et al.*, “Photogrammetric 3D reconstruction of the old slaughterhouse in Budapest,” in *IWCIM*, 2016.
- [15] W. Sheng *et al.*, “Visual Point-Based Analysis of Laser-Scanned Historical Structures,” in *IEEE ICC*, 2015.
- [16] F. J. Perez-Grau *et al.*, “An architecture for robust UAV navigation in GPS-denied areas,” *JFR*, vol. 35, pp. 121–145, 2018.
- [17] G. A. Kumar *et al.*, “A LiDAR and IMU Integrated Indoor Navigation System for UAVs and Its Application in Real-Time Pipeline Classification,” *Sensors*, vol. 17, no. 6, 2017.
- [18] F. Wang *et al.*, “An efficient UAV navigation solution for confined but partially known indoor environments,” in *IEEE ICCA*, 2014.
- [19] L. Wang *et al.*, “Map-Based Localization Method for Autonomous Vehicles Using 3D-LIDAR,” *IFAC World Congress*, 2017.
- [20] H. Ye, Y. Chen, and M. Liu, “Tightly Coupled 3D Lidar Inertial Odometry and Mapping,” *CoRR*, vol. abs/1904.06993, 2019.
- [21] V. Krátký, P. Petráček, V. Spurný, and M. Saska, “Autonomous Reflectance Transformation Imaging by Team of Unmanned Aerial Vehicles,” *IEEE RA-L*, In Press: Accepted for publication on January 15, 2020.
- [22] Y. Zhang *et al.*, “Lighting Design for Globally Illuminated Volume Rendering,” *IEEE TVCG*, vol. 19, no. 12, pp. 2946–2955, 2013.
- [23] M. Saska, V. Krátký, V. Spurný, and T. Báča, “Documentation of Dark Areas of Large Historical Buildings by a Formation of Unmanned Aerial Vehicles using Model Predictive Control,” in *IEEE ETFA*, 2017.
- [24] Multi-Robot Systems group, CTU FEE. (2019) Dronument (Dron & Monument). [Online]. Available: <http://mrs.felk.cvut.cz/dronument>
- [25] G. Loianno *et al.*, “Localization, Grasping, and Transportation of Magnetic Objects by a Team of MAVs in Challenging Desert-Like Environments,” *IEEE RA-L*, vol. 3, no. 3, pp. 1576–1583, 2018.
- [26] T. Baca *et al.*, “Model Predictive Trajectory Tracking and Collision Avoidance for Reliable Outdoor Deployment of Unmanned Aerial Vehicles,” in *IEEE/RSJ IROS*, 2018.
- [27] T. Baca, G. Loianno, and M. Saska, “Embedded Model Predictive Control of Unmanned Micro Aerial Vehicles,” in *IEEE MMAR*, 2016.
- [28] T. Baca, P. Stepan, and M. Saska, “Autonomous Landing On A Moving Car With Unmanned Aerial Vehicle,” in *IEEE EECMR*, 2017.
- [29] A. L. Ballardini *et al.*, “An effective 6DoF motion model for 3D-6DoF Monte Carlo Localization,” in *4th Workshop on Planning, Perception and Navigation for Intelligent Vehicles, IROS*, 2012.
- [30] D. Fox, “KLD-sampling: Adaptive Particle Filters,” in *NIPS*, 2001.
- [31] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. Cambridge, Mass.: MIT Press, 2005.
- [32] R. B. Rusu and S. Cousins, “3D is here: Point Cloud Library (PCL),” in *IEEE ICRA*, Shanghai, China, 2011.