Vision-Based Obstacle Avoidance for Micro Air Vehicles Using an Egocylindrical Depth Map

Roland Brockers(✉), Anthony Fragoso, Brandon Rothrock, Connor Lee, and Larry Matthies

Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, USA
{brockers,brandon.rothrock,lhm}@jpl.nasa.gov,
{afragoso,clee}@caltech.edu

Abstract. Obstacle avoidance is an essential capability for micro air vehicles. Prior approaches have mainly been either purely reactive, mapping low-level visual features directly to headings, or deliberative methods that use onboard 3-D sensors to create a 3-D, voxel-based world model, then generate 3-D trajectories and check them for potential collisions with the world model. Onboard 3-D sensor suites have had limited fields of view. We use forward-looking stereo vision and lateral structure from motion to give a very wide horizontal and vertical field of regard. We fuse depth maps from these sources in a novel robot-centered, cylindrical, inverse range map we call an *egocylinder*. Configuration space expansion directly on the egocylinder gives a very compact representation of visible freespace. This supports very efficient motion planning and collision-checking with better performance guarantees than standard reactive methods. We show the feasibility of this approach experimentally in a challenging outdoor environment.

1 Introduction

Micro air vehicles (MAVs) require onboard obstacle detection and avoidance systems with minimal size, weight, power, complexity, and cost, using sensors with a very large field of regard for maneuvering in cluttered spaces. Vision-based approaches have excellent potential to address these needs for many applications. In prior work [13], we used stereo vision for forward-looking depth perception and showed that inverse range maps in image space can be used for MAV motion planning, with configuration space (C-space) obstacle expansion done in image space, dynamically feasible trajectories generated in 3-D Cartesian space, and collision-checking done by projecting candidate trajectories into image space to determine whether they intersect obstacles. This is a very efficient approach to geometric representation and collision checking, and the overall approach is quite effective where the goal is obstacle avoidance rather than mapping.

In this paper, we extend the total field of regard to about 180° by adding side-looking cameras with structure from motion (SfM) for depth perception. We project the depth data from all cameras onto a cylindrical inverse range image we

D. Kulić et al. (eds.), *2016 International Symposium on Experimental Robotics*,
Springer Proceedings in Advanced Robotics 1, DOI 10.1007/978-3-319-50115-4_44

call an *egocylinder*, and perform C-space expansion on the egocylinder. To reduce the computational cost of motion planning for low-speed flight, we currently use a simple method to choose directions toward more distant goals that stay within freespace shown by the egocylinder. This entire architecture is a step toward our ultimate goal of integrating depth data over time with this data structure and formulating more sophisticated motion planning algorithms directly in image space. Experiments in a challenging outdoor environment have demonstrated the promise of this approach.

2 Related Work

Pros and cons of various passive and active sensor options for MAV obstacle avoidance were discussed in [10,13]; recent examples of MAV systems using multiple types of sensors are described in [5,14]. Here we focus on methods that use vision.

Vision-based approaches break down according to how they do vision, scene representation, and planning and control. The main approaches to vision use optical flow, learning, and/or monocular or stereo depth perception. Optical flow methods typically design reactive control algorithms with optical flow input. Control algorithms for provably stable wall-following and corridor-following behavior have been developed this way [11]; however, navigation that requires a discrete choice among alternate directions requires higher-level perception or reasoning. Machine learning methods have also been used to map optical flow and other monocular visual features into reactive obstacle avoidance behaviors [4], but it is difficult to generalize this approach to work in a wide variety of conditions, so most work on MAV obstacle detection uses depth perception. Forward-looking monocular depth perception via structure from motion (SfM) has been used for MAVs [1,3], but requires aircraft motion to measure depth and has poor depth perception near the focus of expansion. Stereo vision overcomes these limitations, works well in many outdoor settings in particular, and small, fast stereo implementations are progressing [9,12,17].

The predominant approach to scene representation has been to use 2-D or 3-D Cartesian probabilistic grid maps, which can be used with motion planning algorithms that vary from reactive to deliberative and from local to global [5,8,17,18]. These methods are particularly useful for mapping, exploration, or obstacle avoidance in areas that require memory of previously examined avenues; however, they use a lot of storage and computation, and scaling to high speed flight requires multiresolution grids.

For obstacle avoidance *per se*, less expensive representation and planning algorithms are possible. Often these representations are polar in nature, matching the polar angular resolution of the depth sensors [2,15]. In [16], depth data from two onboard stereo pairs was fused in a cylindrical inverse range map centered on the vehicle. This work introduced C-space obstacle expansion of an image space depth map, though in a limited fashion based on an assumed ground plane. In [13], we generalized the C-space expansion to be based on the actual

depth at each pixel and developed the first combination of an image space depth representation with a dynamics-aware motion planner; feasible trajectories were generated in 3-D and projected into image space to do collision checking by testing for intersections with the C-space expanded depth map. This approach to obstacle representation and collision checking is fast and showed good potential in experiments; however, the field of view was limited and the CL-RRT motion planning algorithm was computationally expensive.

3 Technical Approach

Figure 1 illustrates the sensing, processing, and algorithm architecture of our approach, which is implemented on an AscTec Pelican quadrotor (Fig. 5). To minimize the sensor hardware, we augment forward-looking stereo by adding single cameras aimed 45° to each side, giving a total field of regard of about 180°. Stereo matching is done with local block matching for speed, which works adequately well in our test environments. To obtain depth perception with the side-looking cameras, we examined several options for using two-frame optical flow algorithms, as well as the LSD-SLAM incremental SfM algorithm [6]. LSD-SLAM constrains optical flow search to epipolar lines computed from estimated camera motion, and incrementally improves depth resolution by using each new image to update depth estimates in keyframes. We found this to be much less noisy than unconstrained optical flow algorithms, so we use this approach. Since monocular SfM has an unobservable scale factor, we estimate scale by comparing SfM range measurements with range from stereo in the image overlap areas between the outer SfM cameras and the stereo cameras.

To provide a unified depth representation, we project the stereo and scaled SfM depth maps onto a cylinder centered on the aircraft, which is stored as a

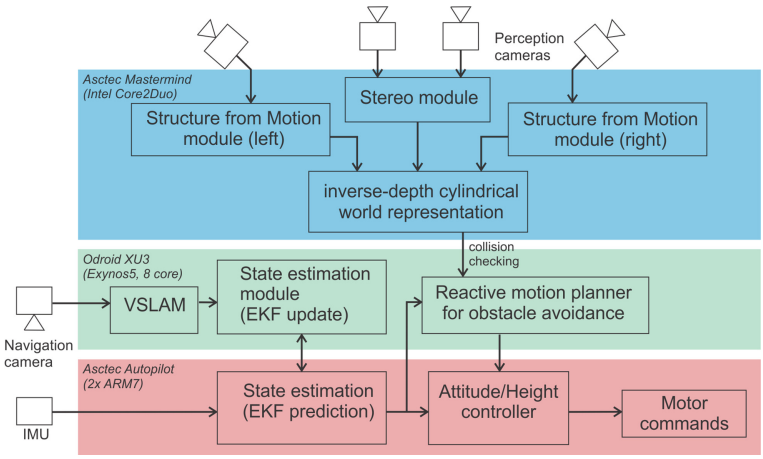

Fig. 1. System architecture.

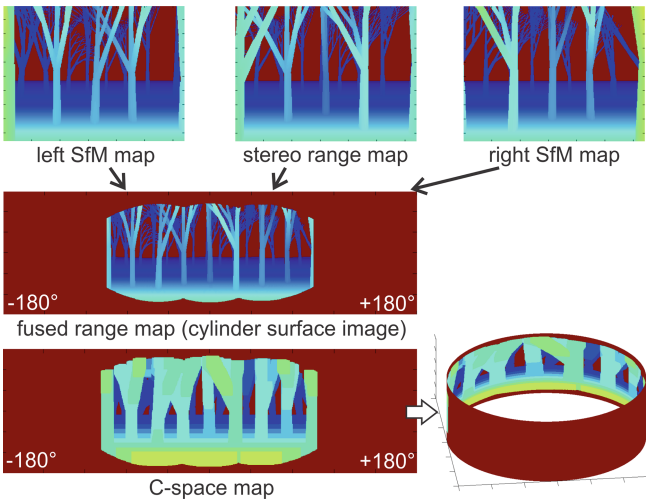

Fig. 2. Schematic illustration of stereo and SfM depth maps fusion into the egocylinder representation, and C-space expansion of the egocylinder. Using inverse range, the expansion widens closer objects more than farther objects.

disparity map where each pixel records the inverse radial distance to the nearest object in the direction of the pixel (Fig. 2). Currently the egocylinder has the orientation of the body frame, though it could be aligned with the gravity vector. C-space expansion is done on this disparity map similarly to [13], which essentially reduces the range at each pixel and widens objects in the depth map in proportion to the width and height of the aircraft plus a safety margin. Using inverse range conveniently gives a finite representation (zero) to objects beyond the maximum range of the sensors. Quantized inverse range also matches the range uncertainty characteristics of vision-based depth estimation.

The expanded egocylinder allows the aircraft to be treated as a point for collision checking. In this paper, we evaluate the innovations in the perception and representation system with a simple, fast avoidance algorithm that is safe if there are no major perceptual errors. At the obstacle densities and velocities considered here, we employ a reduced dynamical model in which the vehicle can turn with infinite agility but requires a finite distance to come to a stop. Accordingly, we restrict the set of possible vehicle trajectories at any instant to the set of straight lines extending radially from the vehicle. Collision-free trajectories are then extracted from this set by transforming the vehicle velocity into an inverse-range safety horizon that is based on the time required to come to a complete stop. After first checking the goal direction in order to avoid a search if possible, the entire planning horizon is checked against the egocylinder to eliminate flight directions that violate the safety horizon constraint. A simple scan of the remaining pixels then returns the collision-free direction that is closest to the goal direction (Fig. 3).

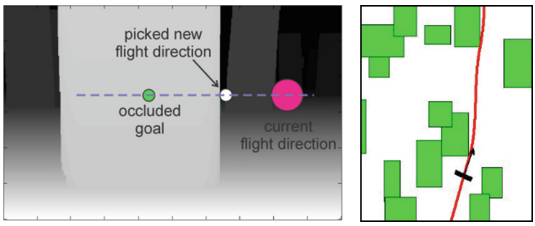

Fig. 3. Motion planning schematic and simulation. Left: selected flight direction to avoid obstacle. Right: simulated flights through cluttered environments without cul-de-sacs were successful up to speeds over 15 m/s (top view).

To maximize safety and visibility of the scene ahead, a low-level controller executes the command by yawing the cameras towards the requested direction while separate PID loops maintain forward velocity and eliminate side slip around the turn. We have also implemented a simple temporal filtering feature that provides robustness against noisy or missing depth data — once a point on the planning horizon is chosen, it is propagated forward with the motion of vehicle for a few cycles and assigned a preference over the egospace pixel scan. In addition to reducing latency by allowing the planning pipeline to be bypassed most of the time, this memory feature tends to smooth out the flight of the vehicle through complicated or noisy environments, in which the target would otherwise change frequently, and prevents dropped frames or other gaps in visual input from ending a flight. This entire planning approach is very fast, safe, and allows us to focus on evaluating perception at the cost of sacrificing algorithmic completeness and strict satisfaction of the full vehicle dynamics. Ongoing work will employ a more sophisticated image space motion planning algorithm that can accommodate these issues.

Figure 1 shows how all parts of the algorithm mapped onto our three-level processor architecture. Images are processed at 384×240 pixel resolution. Stereo runs at 5 fps, LSD-SLAM at 10 fps, and the egocylinder and motion plan are updated at the stereo frame rate of 5 fps. Planning itself takes under a millisecond to verify that the current direction is still safe, and a few milliseconds if it is necessary to search for a new direction. Visual-inertial state estimation is done with a nadir-pointed camera using methods from [19].

Operating outdoors in areas with bright sunlight and deep shadow is difficult, because it creates very large intra-scene (within the same image) and inter-scene (between successive images) illumination variations that greatly exceed the linear dynamic range of available cameras. This has been especially problematic in experiments we have conducted in a grove of trees (see Sect. 4) using a nadir-pointed camera for state estimation. The most effective way to address this is to improve dynamic range at the sensor level. There are multiple potential ways to do this. Some approaches acquire multiple images separated in time and combine these in software; this is impractical on a moving robot. Another approach uses hardware design in the imager that provides a multi-linear exposure mode

that approximates a logarithmic image response. This mode is implemented in the Matrix Vision mvBlueFOX-200w CMOS cameras we use and can extend the total dynamic range from 55 dB to 110 dB. These cameras have three linear segments in their photometric response function, where the slope and transition point of the second and third segments is controlled by a two sets of knee point parameters. Creating a good exposure for given scene conditions requires choosing the total exposure time and setting appropriate knee point parameters.

Fig. 4. Non-HDR (left) and HDR (right) images in a forest scene. Large areas are saturated or under-exposed in the non-HDR image. The HDR image has a better distribution of intensity values, which leads to better performance of vision algorithms.

We have taken a first step toward exploiting this multi-linear HDR mode in the following camera initialization procedure, which is run once at the start of an experiment (Fig. 4). First, we acquire a series of images while adjusting exposure time via gradient descent to push the average intensity of the image stream towards a target intensity in the middle of the pixel brightness range. Next, we fix the total exposure time while seeking the parameters of each knee point that maximize image entropy. In an iterative process, each knee point is set sequentially to maximize local entropy. This does not simultaneously optimize the setting of both knee points, but it avoids extra parameters and has shown to improve feature tracking performance. Once the exposure parameters are initialized, they are fixed for the duration of the flight, which has been adequate in our test conditions to date. Ideally, exposure should be optimized on every frame; however, our current optimization procedure is too slow for that and large changes of exposure have potential to require changes to feature tracking algorithms to maintain landmark tracking across exposure discontinuities. The latter issue was out of scope for this paper.

4 Results

We have conducted low-speed (< 1 m/s) experimental trials in a grove of trees that provided a relatively high obstacle frequency (Fig. 5). Flights totaled over 500 meters in aggregate length, during which 65 trees were encountered. This area had very difficult illumination conditions due to the combination of brightly

sunlit and deeply shadowed areas in the same image. Figure 6 shows results of the vision pipeline at several points during such a run, as well as a 2-D map produced after the fact from data logs.

Fig. 5. Left: Grove of trees test area; Right: AscTec Pelican with 4 camera head.

The saturated and underexposed areas of the images in Fig. 4 illustrate the dynamic range problem with these illumination conditions. While the C-space expansion effectively fills in many areas that have missing data in depth maps from stereo and SfM, this forest environment was particularly challenging for the visual-inertial state estimation system. Therefore, we focused HDR experiments on the state estimation camera, where use of the HDR mode improved the average percentage of map landmarks that could be matched in each frame from 61% to 79%. Nevertheless, the floor of the forest had many very small, self-similar features, and doing state estimation with a nadir-pointed camera while flying low (< 2 m above the ground) in this environment still made state estimation by far the weakest link in the system.

The detection and evasion portions of the architecture were very reliable in the performance evaluation experiments, which were analyzed quantitatively by noting the frequency and cause of any human intervention required to avoid a collision. These modules were responsible for only a single intervention event during the 521 m recorded, which resulted in successful avoidance of 64 out of 65 trees for an success rate of 98%. The intervention was attributed to a missed detection in which the vehicle had drifted too close to an obstacle and could no longer detect it using stereo matching. LSD-SLAM failed to adequately track features about 25% of the time; with data logging, the LSD-SLAM frame rate dropped to about 8 Hz, which is too slow for this algorithm to be reliable. However, this did not impact overall obstacle avoidance performance, because the control policy of first turning the stereo cameras towards the flight direction, and incorporating a small amount of path hysteresis, provided a high degree of robustness to missed left or right camera LSD-SLAM depth maps, which were seamlessly reacquired beginning on the next frame.

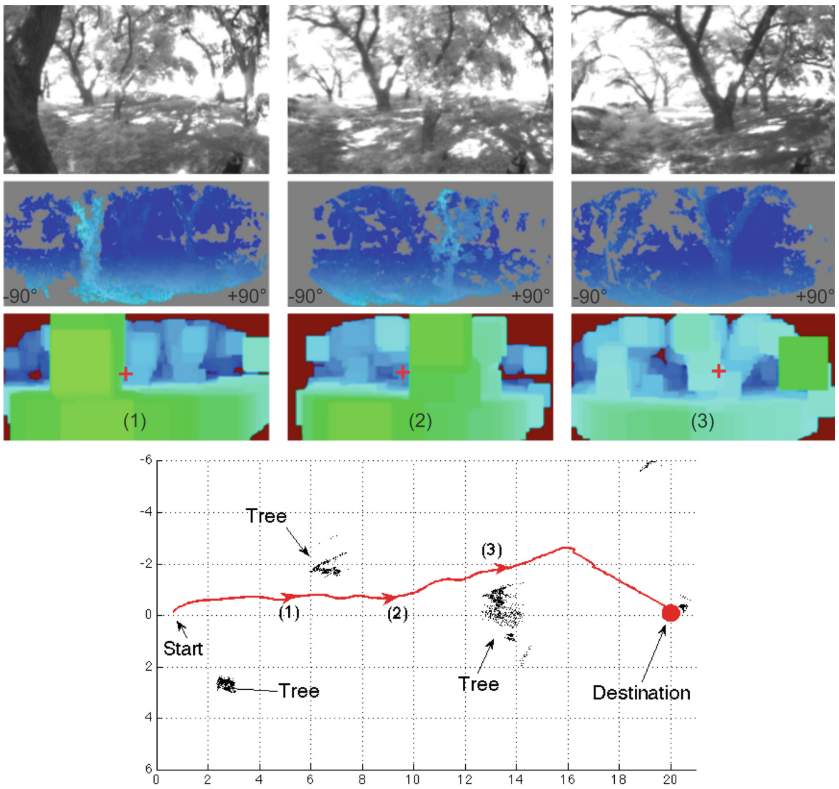

Fig. 6. Results of a 20 m experimental flight through a grove of trees. Top: the results of the perception system for three different locations on the run, showing the left rectified stereo image, the fused egocylinder, and the C-space expanded egocylinder with selected direction of flight (red crosses). This only shows the central 180° of the egocylinder. Bottom: a top down 2-D plot of the trajectory and nearby obstacle pixels from the egocylinder over the whole run. Arrows and numbers on the trajectory show where the three images above were acquired. Vehicle speed was 1 m/s throughout.

5 Main Experimental Insights

Using C-space expansion of image space depth maps for collision checking is a very new approach to obstacle avoidance for MAVs. In our experiments to date, obstacle avoidance has been quite successful; in 521 m of flight in challenging conditions, only one intervention was needed in 65 encounters with obstacles, and no problems with false alarms in freespace were apparent. This is significant, since the approach so far does not include explicit temporal fusion for false alarm suppression or filling in missing data, unlike approaches based on voxel maps. Nevertheless, work is in progress to add temporal fusion to image space representations to address the finite probability that these problems will eventually occur.

By far the biggest performance problem in this system is with visual state estimation. Using a nadir pointed camera while flying low (< 2 m above ground) in a scene with a very high dynamic range of illumination and many small, self-similar features (leaves) on the forest floor seems to be at the heart of the problem. We plan to address this in several ways in ongoing work, including using visual feature tracking in the forward and sideward looking cameras. LSD-SLAM was successful as a source of side-looking depth data, but it requires a high frame rate (>10 Hz) and accurate calibration of camera extrinsics to maintain its usefulness for obstacle detection, both of which were problematic in this implementation. Side-looking stereo cameras might be easier to use, but would lack the potential of exploiting increasing motion baselines to improve depth resolution that exists with recursive approaches to structure from motion. Ultimately, combining both may be a good approach, as is explored in a recent stereo extension of LSD-SLAM [7].

The disparity-space reactive planner extends the advantages of the C-space expansion method and egocylinder to the planning regime — potential trajectories are selected and executed in a highly economical fashion by employing the same framework that allows the egocylinder to represent obstacles compactly and efficiently. Overall, this choice of representation demonstrates decreased planning latency and complexity compared to world-coordinate methods.

There is a close connection between vehicle velocity, uncertainty in the range data, and successful obstacle avoidance — this has not emerged as an issue for the slow speeds of our experiments so far, but for reliable obstacle detection to scale to high speeds, this interplay will require further study. Several approaches may improve the maximum range and range resolution of the system to support higher velocities, including the use of higher resolution imagery and potentially the use of temporal fusion of depth maps for improved range resolution. Scenes involving moving obstacles will require extensions of both the perception and the planning elements of this system.

Acknowledgments. This work was funded by the Army Research Laboratory under the Micro Autonomous Systems & Technology Collaborative Technology Alliance program (MAST-CTA). JPL contributions were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

1. Alvarez, H., Paz, L.M., Sturm, J., Cremers, D.: Collision avoidance for quadratures with a monocular camera. In: Hsieh, M.A., Khatib, O., Kumar, V. (eds.) Experimental Robotics. Springer Tracts in Advanced Robotics, vol. 109, pp. 195–209. Springer, Heidelberg (2016)
2. Bajracharya, M., Howard, A., Matthies, L., Tang, B., Turmon, M.: Autonomous off-road navigation with end-to-end learning for the LAGR program. Field Robot. **26**(1), 3–25 (2009)

3. Daftry, D., Dey, D., Sandhawalia, H., Zeng, S., Bagnell, J.A., Hebert, M.: Semi-dense visual odometry for monocular navigation in cluttered environments. In: IEEE International Conference on Robotics and Automation, Workshop on Recent Advances in Sensing and Actuation for Bioinspired Agile Flight (2015)
4. Dey, D., et al.: Vision and learning for deliberative monocular cluttered flight. In: 10th Conference on Field and Service Robotics (2015)
5. Droeschel, D., Nieuwenhuisen, M., Beul, M., Holz, D., Stuckler, J., Behnke, S.: Multi-layered mapping and navigation for autonomous micro air vehicles. Field Robot. (2015)
6. Engel, J., Schöps, T., Cremers, D.: LSD-SLAM: large-scale direct monocular SLAM. In: European Conference on Computer Vision (ECCV), September 2014
7. Engel, J., Stuckler, J., Cremers, D.: Large-scale direct SLAM with stereo cameras. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, September 2015
8. Fraundorfer, F., Heng, L., Honegger, D., Lee, G.H., Meier, L., Tanskanen, P., Pollefeys, M.: Vision-based autonomous mapping and exploration using a quadrotor MAV. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (2012)
9. Goldberg, S.B., Matthies, L.: Stereo and IMU assisted visual odometry on an OMAP3530 for small robots. In: IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Embedded Computer Vision (2011)
10. Kendoul, F.: Survey of advances in guidance, navigation, and control of unmanned rotorcraft systems. Field Robot. **29**(2), 315–378 (2012)
11. Keshavan, J., Gremillion, G., Alvarez-Escobar, H., Humbert, J.S.: Autonomous vision-based navigation of a quadrature in corridor-like environments. Int. J. Micro Air Veh. **7**(2), 111–123 (2015)
12. Kuhn, M., Moser, S., Isler, O., Gurkaynak, F.K., Burg, A., Felber, N., Kaelin, H., Fichtner, W.: Efficient ASIC implementation of a real-time depth mapping stereo vision system. In: IEEE 46th Midwest Symposium on Circuits and Systems (2003)
13. Matthies, L., Brockers, R., Kuwata, Y., Weiss, S.: Stereo vision-based obstacle avoidance for micro air vehicles using disparity space. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 3242–3249 (2014)
14. Nuske, S., Choudhury, S., Jain, S., Chambers, A., Yoder, L., Scherer, S., Chambelain, L., Cover, H., Singh, S.: Autonomous exploration and motion planning for an unmanned aerial vehicle navigating rivers. Field Robot. **32**(8), 1141–1162 (2015)
15. Oleynikova, H., Honegger, D., Pollefeys, M.: Reactive avoidance using embedded stereo vision for MAV flight. In: IEEE International Conference on Robotics and Automation (2015)
16. Otte, M.W., Richardson, S.G., Mulligan, J., Grudic, G.: Path planning in image space for autonomous robot navigation in unstructured outdoor environments. Field Robot. **26**(2), 212–240 (2009)
17. Schmid, K., Lutz, P., Tomic, T., Mair, E., Hirschmuller, H.: Autonomous vision-based micro air vehicle for indoor and outdoor navigation. Field Robot. **31**(4), 537–570 (2014)
18. Shen, S., Michael, N., Kumar, V.: 3D indoor exploration with a computationally constrained MAV. In: Robotics: Science and Systems (2011)
19. Weiss, S., Achtelik, M., Lynen, S., Achtelik, M., Kneip, L., Chli, M., Siegwart, R.: Monocular vision for long-term micro aerial vehicle state estimation: a compendium. Field Robot. **30**(5), 803–831 (2013)