Progress and Communication in Sciences

3D Mapping in Tunnel Environments: Maximizing Mission Range

Steven L. Waslander[†], Arun Das[†], and Mark Haley[‡]
[†]University of Waterloo, USA
[‡]Analytical Software Inc., USA

Abstract**—**This paper presents an analysis of the power and computational requirements for 3D tunnel mapping using an autonomous ground vehicle and a single Microsoft Kinect sensor. A comparison of existing mapping algorithms, including RGB-D SLAM and KinectFusion, is performed on a representative hallway to demonstrate the accuracy with which tunnels can be mapped. The power requirements for both the vehicle locomotion and onboard computation are assessed, and representative tunnel mapping ranges are computed. Finally, a system is proposed that minimizes power consumption without sacrificing map quality that could operate continuously for 26 hours and collect over x km of map data, which could be processed online for real-time mapping and refined offline for more accurate results. Extension to UAV/UGV teams is also considered.

Keywords**—**Autonomous ground vehicles, 3D mapping, reconstruction, SLAM.

I. INTRODUCTION

AUTONOMOUS robotic mapping can provide a safe alternative to human exploration of dangerous environments, such as mines, tunnels and war zones with unexploded ordinance. Surveillance, mapping and rescue robots are crucial tools for civilian and military emergencies ranging from the Fukushima nuclear disaster, to disabling improvised explosive devices (IEDs) in Afghanistan and Iraq, exploring coal mines after explosions or in tunnels in war-zones. For example after catastrophes such as Fukushima or for military Intelligence, Surveillance and Reconnaissance (ISR), sometimes the most crucial objective is simply to obtain video and 3D map data to plan the best options for first responders.

Particularly in indoor environments, the multi-channel sensing capabilities of RGB-D sensors such as the Microsoft Kinect provide a low-cost, low-power method to collect accurate and valuable 3D information about the environment from either ground or aerial robots that can be used to construct detailed maps in real-time. These maps can provide both colour and geometric information about narrow passages, and work well in lighted and low-light conditions. In this report, we outline the latest research in the area of 3D mapping of tunnels with RGB-D sensors, and discuss the power requirements for autonomous vehicle missions to generate these maps without the aid of human operators. A summary video of the results and possibilities with UAV/UGV teams is available at http://youtu.be/NsaB7-LSUn8.

The first efforts to employ RGB-D sensors for 3D mapping were presented by Newcombe et al. [1] as Kinect Fusion, using a process that relies on a full voxel grid of the 3D environment to match the current scan to an aggregated map. The resulting maps are spectacularly detailed and can be developed in real time onboard typical CPU/GPU hardware. The main issues with the approach are the reliance on depth information only leading to degeneracy in hallways and tunnels, heavy computational requirement of full scan matching and dense voxel grid maintenance, and the limited volumes that can be mapped even when exploiting the full extent of today's best GPU memory limits. The result is a limited volume, energy hungry method that is not suitable for long range tunnel mapping.

These limitations were immediately observed by multiple groups including ours, and a slew of extensions were produced that alleviated the memory limitation for mapping. The most complete proceeded by tiling the environment and storing compressed surface representations outside an active local voxel grid. Named Kintinuous [2], this work enabled long term mapping with the same map resolution as in Kinect Fusion. The work was expanded in Kintinuous 2.0 [3] which added color information to the scan registration process, improved loop closure and map deformation, leading to successful results over large data sets. Again, computational complexity requires CPU/GPU architectures and is therefore energy hungry.

An alternate approach relies more heavily on the RGB data collected by the sensor, and constructs the map using the dense RGB-D information after loop closure using traditional SLAM methods for features. This method is known as RGB-D SLAM [4], and has been both extensively benchmarked [5] and a computationally light method [6] has been implemented on a small rotorcraft UAV, demonstrating the low computational requirements that are needed for this system to be successful. The drawback of using vision as the primary localization measure instead of depth is that it introduces a requirement for well-lit environments

In summary, both Kintinuous and RGB-D SLAM provide promising approaches for the tunnel environment with strict

Corresponding author: Mark Haley (e-mail: m_haley@faculty.chiba-u.jp).
This paper was submitted on November 04, 2014; revised on December 20, 2014; and accepted on December 23, 2014.

power limitations. When ample light is available, the combined use of vision and depth information exhibited in RGB-D SLAM provides a robust, efficient 3D mapping method. If light conditions are low or the scene presents few variations in colour, however, the active depth sensing of Kintinuous continues to map accurately, as long as there is sufficient geometric richness to properly match new scan images. We have implemented both algorithms, and demonstrate them below in two environments, a tunnel-like hallway for RGB-D SLAM, and a lab environment for a slight variant of Kintinuous.

The paper proceeds as follows. In Section II, we briefly describe the methods employed for 3D SLAM on an autonomous ground vehicle. We then analyze the power requirements and mapping range achievable with different configurations of the system in Section III. Section IV contains representative results of mapping exercises performed in tunnel like and dark environments, and finally, the paper concludes with recommendations for a customized, efficient autonomous platform for 3D mapping in tunnels over long duration missions.

II. 3D SLAM ALGORITHMS

The RGB-D SLAM algorithm is composed of the following sequence of operations, depicted in **Figure 1** (taken from [5]).

1. Each new RGB image is processed to identify unique features and their descriptors using standard techniques such as SIFT, SURF etc.
2. These features are matched to known features in the previous 20 images plus in keyframe images stored to make the map, establishing inter-frame correspondence that can be used in backend optimizations. This matching is significantly more robust than image only matching, as the 3D position of the feature is known thanks to the depth channel.
3. A pairwise 6D transformation is estimated between the new image and the previous one for which matches were made amongst the features. Here, the RANSAC algorithm is employed to aid in the rejection of outliers (points with incorrect correspondence between images).
4. Connections between local image frames and keyframes are aggregated into a factor graph, which is optimized using the research standard g2o optimization library, to define the complete map. This step enables loop closures and large scale map corrections.
5. Finally, the map itself is constructed by aggregating the depth information in the entire sequence of images. The octomap representation is used, which probabilistically aggregates all measurements into a common map based on the trajectory computed in steps 1-4. This map is useful for motion planning and visualization, and can be constructed only after the SLAM optimization stage is complete (once all data has been collected). Interim can of course be made with partial trajectories, but must be discarded and recomputed after each SLAM optimization update.

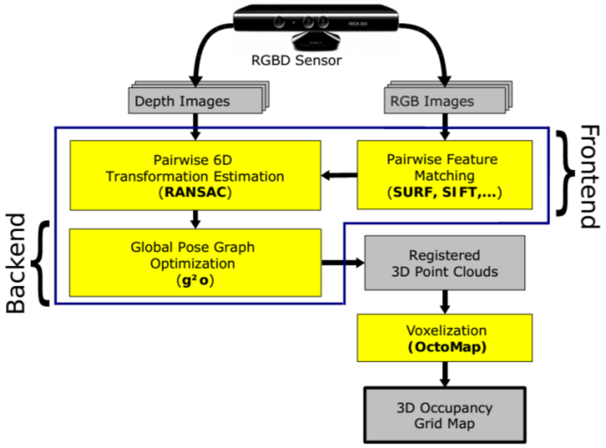

Figure 1 RGB-D SLAM algorithm block diagram [5]

The alternative approach employs no RGB image information, and therefore works equally well with and without light. We employ our own variant of the Kintinuous algorithm, named WaveFusion. An active volume voxel grid is defined around the robot, and the first depth scan is projected into the volume represented as a truncated signed distance function (TSDF). The surface of the environment is represented by a value of zero, and each cell in front of the surface along a particular depth ray from the sensor is assigned its distance to the surface. All cells along the ray behind the surface are given negative values similarly. This representation of the map is used throughout, and the values in each cell are averaged/adjusted every time a new measurement through that cell is received. Given this starting point, the updates proceed as follows.

1. Given an estimate of the inter-frame motion of the robot, the 3D zero surface from the TSDF is reprojected into the current image frame to recreate an expected image from the current map. The true image is registered to the recreated image using the iterative closest point algorithm (ICP), and the process is repeated until convergence.
2. The new depth image can then be added to the map, by projecting each depth measurement into the 3D voxel grid and updating the signed distance values in each cell accordingly. This averaging is what causes the map to have a much smoother appearance than the raw sensor data.
3. If the robot has moved sufficiently away from the center of the 3D voxel grid, the regions of the grid that are outside the field of view are compressed and stored on a hard disk, and new regions are added to the voxel grid in the expected direction of travel.

III. RANGE AND ENERGY ANALYSIS

The basic hardware configuration employed for these experiments is the Microsoft Kinect sensor, a high-end GPU enabled laptop computer and the Clearpath Husky ground vehicle. A small flashlight was also used in the lab environment to provide illumination from the robot only. Such a system would be a prime candidate for long-range tunnel mapping operations, and could proceed steadily at 1 m/s collecting sufficient data for rich 3D reconstruction that can be performed in real-time on board the vehicle, and conveniently refined and optimized offline for more accurate representations with more energy and computation intensive algorithms.

A. Transport power requirements

The Clearpath Husky is a 50 kg platform capable of operating for up to 8 hours on a single battery charge of its stock lead acid battery (20 VDC, 24 Ah, 480 Wh), assuming a roughly. The average driving load on smooth, flat terrain, therefore, consumes about 60 W of power. It is straightforward to convert such a platform to a LiFe battery technology, for which a much larger battery is available at a similar weight (24 VDC, 100 Ah, 2400 Wh) increasing the vehicle range 5 fold.

B. Computing power requirements

The laptop employed in the experiments is a high-end Lenovo Thinkpad with NVIDIA GPU, which consumes approximately 100-150 W during operation. Improved performance is likely with higher power, as the update rate of the algorithms constrains the speed of travel. It is possible, through a reasonable development effort, to parallelize this computational load over multiple low power CPU/GPU processors such as the NVIDIA TK1, which each consume about 5 W. The benefit would be to drop the computing power requirements below 25 W total.

C. Sensing power requirements

The Microsoft Kinect 1.0 sensor requires 15 W, and operates from USB 3.0, but requires an additional power connection to obtain its full power requirement. Future upgrades to USB 3.0 and to the Kinect sensor may both improve the accuracy, image quality and resolution of the sensor and reduce the overall power consumption, enabling power to be provided by the USB 3.0 interface directly.

D. Lighting power requirements

Optionally, it is possible to use low-power high-lumens LED flood lights to illuminate any dark scenes. A standard 50 W LED flood light provides 4000-5000 lumens, sufficient lighting to illuminate a 120 degree cone out to 30 ft with sufficient light for visual tracking. Note that the resulting shadows may confound visual feature detection, so some development work would be required.

E. Aerial vehicle power requirements

Further, it is possible to extend accessibility in disaster scenarios by deploying aerial companion vehicles when ground vehicle access is blocked or in order to plan safe traversal routes. Again, both RGB and depth data can be collected, and similar sensing densities and update rates are required in order to construct maps of the same quality as onboard the ground vehicle. Flight power requirements for small UAVs can be significantly higher than the Husky vehicle, on the order of 200-300 W for flight of a vehicle with sufficient payload to transport a Kinect sensor. It is expected, however, that the UAV would be used only sporadically for specific mapping issue resolution, and is therefore not considered directly in the range calculations.

The approximate power budgets and range limits for various configurations of the above systems are described in **TABLE I**.

TABLE I POWER BUDGET AND RANGE ANALYSIS FOR DIFFERENT CONFIGURATIONS

System	**Components**	**Power rqmts.**	**Lead-acid Range**	**LiFe Range**
Basic system				
	Husky	60 W		
	Laptop	150 W		
	Kinect	15 W		
	Total	**225 W**	**2h 8m**	**10h 40m**
Lighted system				
	Basic system	225 W		
	LED Light	50 W		
	Total	**275 W**	**1h 45m**	**8h 45m**
Low power system				
	Husky – smooth tires	40 W		
	Low power computing	25 W		
	Kinect	15 W		
	Total	**90 W**	**5h 20m**	**26h 40m**
Low power system with light				
	Low power system	90 W		
	LED Light	50 W		
	Total	**140 W**	**3h 25m**	**17h 8m**

IV. EXPERIMENTAL RESULTS

Three sets of tests were run to demonstrate the effectiveness of the SLAM methods in various environments and lighting conditions. The first tests were representative of the tunnel mapping application in good light, and were performed using RGB-D SLAM. The final map is presented in **Figure 2**, and clearly demonstrates accurate tunnel reconstruction with limited along tunnel warping. Videos of the mapping process and the resulting octomap of the hallway are available at: http://youtu.be/qJOHEZpIHEM, http://youtu.be/T_WVvpbiiR0 It can be observed in the videos that minor Kinect head oscillation was included, to provide sufficient field of view richness to properly construct the tunnel. The biggest issues with this method lie in the extremely large computational load of registering so many RGB-D images with each other, leading to 10 cm/s progress along the tunnel. The current roadblocks in terms of online mapping at greater speed involve simply accelerating the image processing pipeline through parallelization and reducing the update rate for full map reconstruction and visualization, such that no major hurdles are expected in improving this mapping rate to the intended 1 m/s that the vehicle selected is capable of.

When lighting conditions are worse, the RGB-D SLAM map quality is reduced significantly, and mapping can fail catastrophically when insufficient features are observed for image registration. Despite this limitation, even with only a small flashlight and a small amount of backlit from an adjacent room, the RGB-D SLAM algorithm was still able to map a lab space environment with reasonable accuracy. Videos of the lab space mapping with only a small flashlight are available at: http://youtu.be/4EQBDyzrTHY, http://youtu.be/BVujJAr18qE The metric consistency of the resulting map is reduced, as can be observed by the wall structures that do not form a clear rectangle. Such errors are less prominent in the hallway scenario with better light. **Figure 3** presents the map construction at an interim point in the lab traversal, with the constructed map in the top half of the figure, the RGB image in the bottom left, the depth image in the bottom middle and the SIFT features in the bottom right.

Finally, the depth only method is demonstrated on the same lab environment, and the resulting map is constructed equivalently regardless of illumination, depicted in **Figure 4**. Although the map quality is locally excellent, the algorithm does not provide for any loop closure correction, and as such, the map tends to drift over time This method leads to maps of similar quality as the low light maps for RGB-D slam in the long run, although the local information is very accurate. Computations on both algorithms are approximately equivalent, and so our recommendation is to use RGB-D SLAM with illumination for the best maps, and either method but with degraded performance for low light conditions.

Figure 2 Mapping results from RGB-D SLAM in well-lit hallway

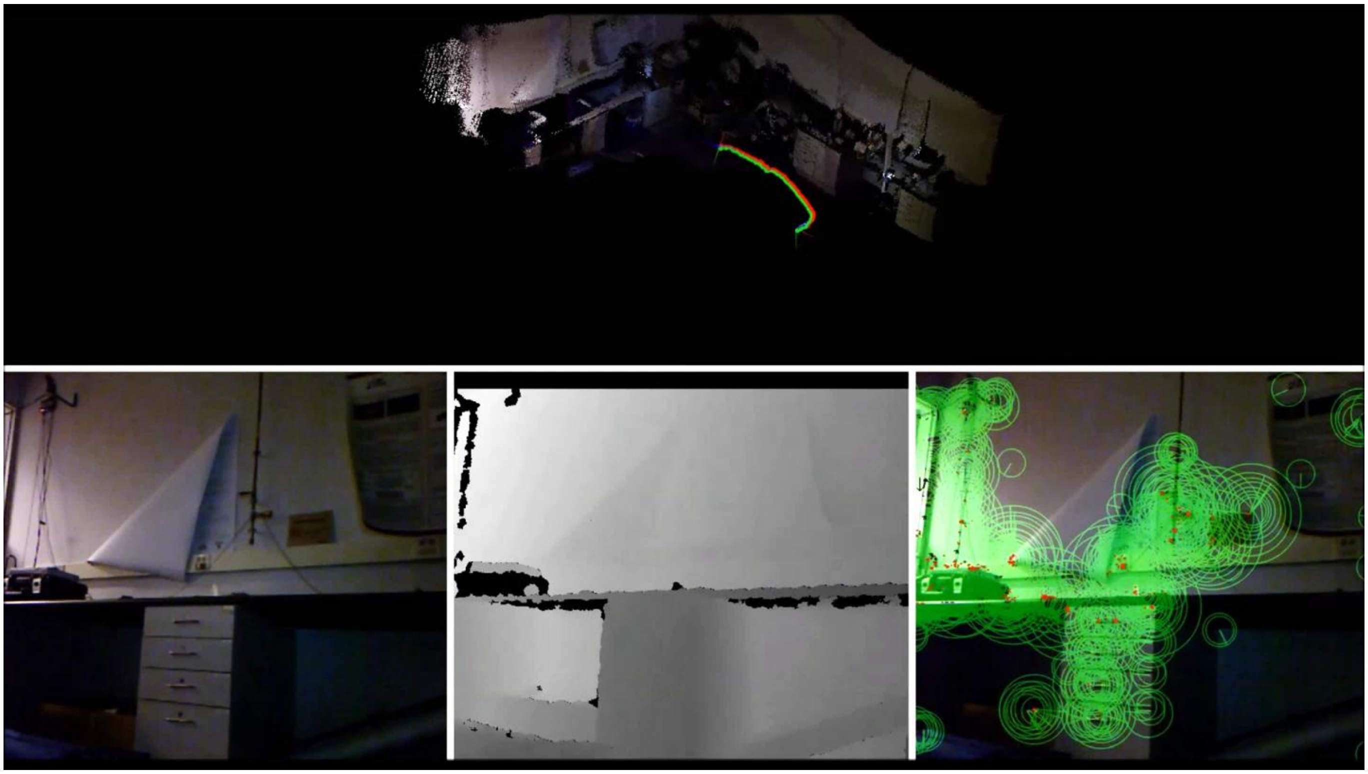

Figure 3 RGB-D SLAM in low-light lab environment

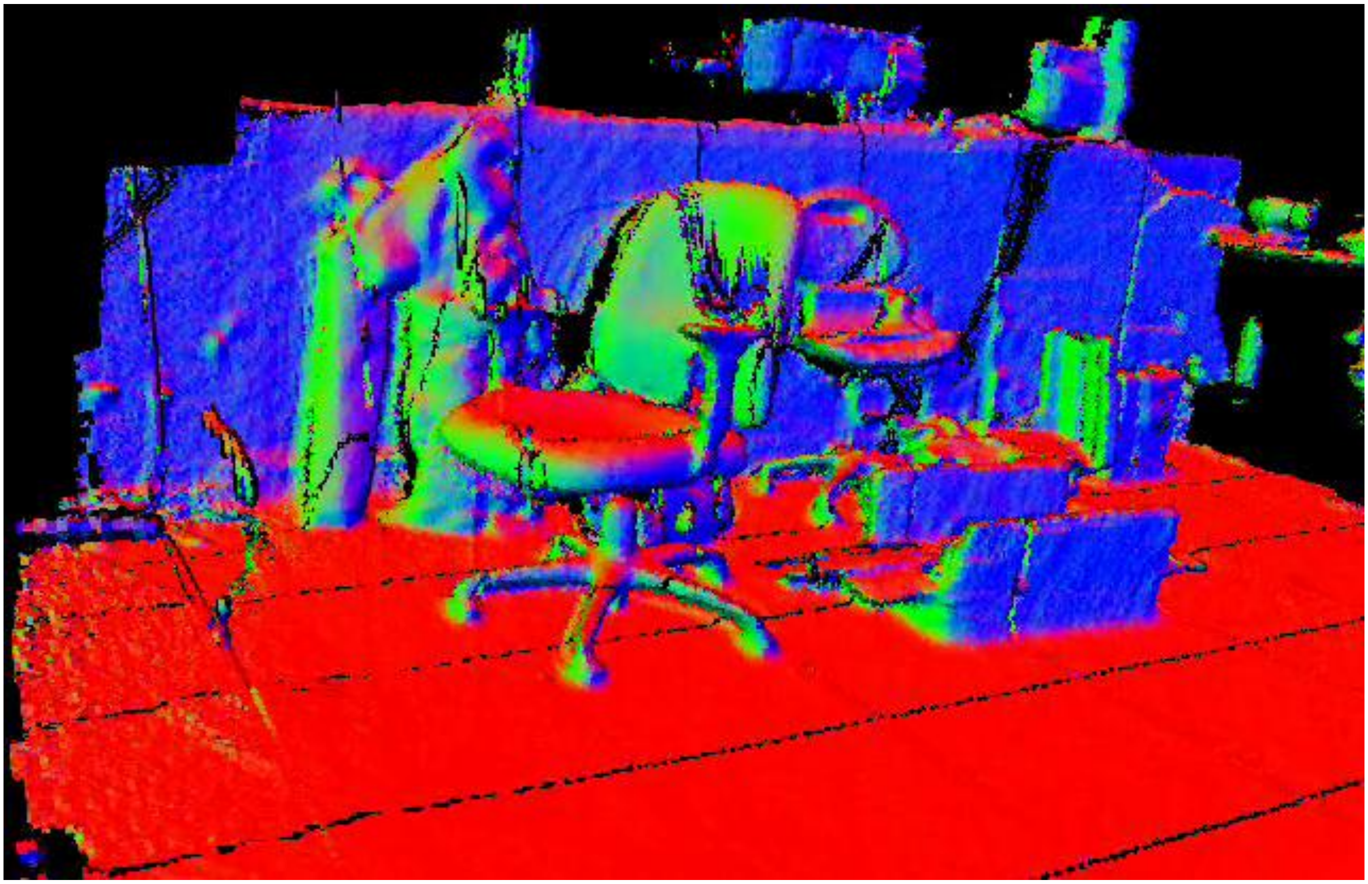

Figure 4 WaveFusion algorithm mapping results in lab environment

V. CONCLUSION

This work demonstrates the feasibility of mapping 3D structures such as tunnels without human intervention or GPS guidance. The power requirements for such a process are driven in nearly equal parts by the locomotion of the platform, the computation of the map, and the potential need for illumination of the environment. Although promising, to truly achieve robust long-duration tunnel mapping will require a concerted effort both in research and development work.

On the research side, today's algorithms are still prone to significant errors in certain specific scenarios, and need to be advanced in terms of their robustness and map quality. This is an exciting research direction, and has many broad implications for robotics, not just for tunnel mapping. By using multiple channels of information effectively, the robot can better track its own motion relative to the environment and therefore improve the overall map quality. The key metric here will be robustness, and can truly be assessed only in representative environments and can be effectively refined only with a representative collection range of real-world datasets.

From the robot development perspective, a fully optimized power efficient platform should be designed to maximize operational time. By taking advantage of the recent revolution in low-power computing driven by the smart phone market as well as the high energy density batteries that have come about thanks to electric automobiles, and by designing a vehicle that can navigate the terrain as efficiently as possible, the overall power consumption can be dramatically reduced, and range can be extended into the 20+ hour range. Such a vehicle need not be particularly expensive, as the parts list for the vehicles described above could easily fall below $5,000.

ACKNOWLEDGEMENT

This work was primarily supported by the US Army International Technology Center – Pacific (ITC-PAC) with additional funding from the Office of Naval Research Global (ONRG), the Asian Office of Aerospace Research & Development (AOARD), the Deputy Assistant Secretary of the Army for Defense Exports and Cooperation (DASA-DEC) and the U.S. Army Tank Automotive Research, Development and Engineering Center (TARDEC), plus additional assistance from the National Science Foundation (NSF).

REFERENCES

[1] Newcombe, Richard A., Andrew J. Davison, Shahram Izadi, Pushmeet Kohli, Otmar Hilliges, Jamie Shotton, David Molyneaux, Steve Hodges, David Kim, and Andrew Fitzgibbon. "KinectFusion: Real-time dense surface mapping and tracking." In Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on, pp. 127-136. IEEE, 2011.

[2] Whelan, Thomas, Michael Kaess, Maurice Fallon, Hordur Johannsson, John Leonard, and John McDonald. "Kintinuous: Spatially extended kinectfusion.", MIT C-SAIL Technical Report #2012-020 (2012).

[3] Whelan, Thomas, Michael Kaess, John J. Leonard, and John McDonald. "Deformation-based loop closure for large scale dense rgb-d slam." In Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on, pp. 548-555. IEEE, 2013.

[4] Engelhard, Nikolas, Felix Endres, Jürgen Hess, Jürgen Sturm, and Wolfram Burgard. "Real-time 3D visual SLAM with a hand-held RGB-D camera." In Proc. of the RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum, Vasteras, Sweden, vol. 180. 2011.

[5] Sturm, Jürgen, Nikolas Engelhard, Felix Endres, Wolfram Burgard, and Daniel Cremers. "A benchmark for the evaluation of RGB-D SLAM systems." In Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on, pp. 573-580. IEEE, 2012.

[6] Scherer, Sebastian A., and Andreas Zell. "Efficient onbard RGBD-SLAM for autonomous MAVs." In Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on, pp. 1062-1068. IEEE, 2013.