Perception for a River Mapping Robot

Andrew Chambers, Supreeth Achar, Stephen Nuske, Jörn Rehder, Bernd Kitt,
Lyle Chamberlain, Justin Haines, Sebastian Scherer, and Sanjiv Singh

***Abstract*—Rivers in areas with heavy vegetation are hard to map from the air. Here we consider the task of mapping their course and the vegetation along the shores with the specific intent of determining river width and canopy height. A complication in such riverine environments is that GPS may not be available depending on the thickness of the surrounding canopy. We present key components of a multimodal perception system to be used for the active exploration and mapping of a river from a small rotorcraft flying a few meters above the water. We describe three key components that use computer vision and laser scanning to follow the river without the use of a prior map, estimate motion of the rotorcraft, ensure collision-free operation, and create a three dimensional representation of the riverine environment. While the ability to fly simplifies the navigation problem, it also introduces an additional set of constraints in terms of size, weight and power. Hence, our solutions are cognizant of the need to perform multi-kilometer missions with a small payload. We present experimental results from each of the three perception subsystems from representative environments.**

Figure 1: A typical riverine environment that we expect to map. A small rotorcraft is able to fly above the shallow, fast moving water and yet remain below the thick canopy to navigate and map the river. The foliage along the banks is dense enough to block or seriously degrade GPS signals.

I. Introduction

We are developing a minimal sensor suite to be used by a low-flying aircraft to autonomously explore rivers, mapping their width and the surrounding canopy. In some cases, the canopy can be so thick and high around a river that it blocks GPS signals and the problem becomes one of simultaneous localization and mapping in an unstructured fully three-dimensional environment. Exploration from a low-flying vehicle is attractive because it extends the sensing horizon and removes complications of navigating in shallow water and aquatic vegetation. However, a flying solution also adds constraints on the size, weight and power available for perception. This is a significant constraint given that the multi-kilometer missions will force all the sensing/computation to be conducted onboard. It is our estimate that given the size of rotorcraft that could reasonably fly in environments with thick canopy, it will be necessary to keep all the sensing and computation components to less than one kilogram.

These constraints on payload and the inability to rely on GPS have significant implications for our design. First, we will need to depend on perception to produce a high resolution 6DOF pose estimate that is much more stable than can be produced by simply integrating inertial sensors. Second, any active imaging, such as from laser scanning, will be required to be very lightweight and low power and hence will be short range. Third, river-following without a prior map will require a perception system that goes significantly further than could be sensed through laser ranging.

In this paper we describe three key modules for perception. The first is a multi-sensor system for position estimation. While visual odometry is a well-studied problem, we demonstrate its ability to robustly estimate consistent paths spanning more than 10,000 poses in a completely unconstraint, visually-challenging environment while maintaining an accuracy that allows for precise local mapping. The second is a short-range, laser-ranging based system tasked with obstacle detection and the creation of a metric, three-dimensional map. Here we use a continuously rotating, lightweight line laser scanner to produce a three-dimensional scanning pattern that enables collision avoidance and the creation of a three-dimensional representation of the riverine environment. The third is a long range color vision system that uses a forward pointing color camera to automatically find the river even in the face of significant variation in the appearance of the river. We solve this problem with a self-supervised method that continually learns to segment images based only on an estimate of the horizon (from inertial sensing) and some simple heuristics that describe riverine environments. We describe experiments in which each of the three systems has been tested and present the results from these experiments.

A. Chambers, S. Achar, S. Nuske, J. Rehder, L. Chamberlain, J. Haines, S. Scherer, and S. Singh are with The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A., {basti,ssingh}@cmu.edu

B. Kitt is with The Institute of Measurement and Control Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany, bernd.kitt@kit.edu

The work described in this paper is funded by the Office of Naval Research under grant number N00014-10-1-0715.

II. Related Work

Previous work in autonomous river mapping has utilized small boats [1] or fixed wing UAVs [2]. While these platforms could be more practical in certain, simplier riverine environments, we aim to develop a platform which can perform well in the most difficult situations such as rapidly flowing water, obstructed waterways, or dense forest canopies.

Visual odometry is a well-studied problem with a broad range of applications in robotics [3], [4], [5]. A variety of approaches have been employed, some of which focus on the implications of epipolar geometry [6] while others address visual odometry as a time-discrete filtering problem of landmarks and camera poses [7], a key-frame bundle adjustment [8], [9], or as a combination of the aforementioned [10]. Scaramuzza et al. simplified the problem by incorporating assumptions about the motion [11]. However, this approach constraints the valid motions significantly, rendering it not well-suited for aerial vehicles.

Choosing the optimal actuation and mounting for a laser scanner is always a trade-off in terms of scan density and detection field of view. [12] provides a thorough analysis of scan patterns for various laser scanner configurations.

III. Vehicle State Estimation

An estimate of the vehicle's pose is required for flight control of the rotorcraft as well as mapping of the environment. In many applications, this estimate heavily depends on GPS measurements. However, in river mapping applications, GPS may not be available or unreliable for large stretches due to occlusion by the canopy. In order to deal with intermittent GPS signals, additional means of localization have to be exploited. Our approach relies on a filter architecture that combines visual odometry with sparse GPS, and IMU measurements to provide a combined state estimate that minimizes drift.

For reasons of robustness, a stereo camera setup was chosen over a monocular camera approach. We use the approach to visual odometry presented in [13]. This approach makes use of the stereo setup by triangulating 3D points based on matches in one stereo pair. The relative motion is then determined by iteratively adjusting the translation and rotation of the camera in order to minimize the reprojection error of these 3D points in the consecutive image pair. An additional Kalman Filter smoothes the estimated motion and bridges gaps, when the temporal matching fails.

In order to fuse inertial measurements, visual odometry and intermittent GPS readings, we employ a graph-based optimization approach comparable to [14], [5]. In this framework, nodes represent the state of the vehicle at different points in time. Sensor readings induce constraints on these states, which are represented as edges connecting state nodes in the graph. Visual odometry and integrated gyroscope measurements provide relative pose measurements, which result in constraints on consecutive nodes in the graph. GPS and IMU information on the other hand impose global constraints on the state, that are represented by edges to a fictitious pose in the origin of the global coordinate frame. Each node is only connected to a small number of other nodes, thus resulting in a sparse system that can efficiently be solved even for a large number of poses.[15] As the effect of constraints added to the last pose decays comparatively fast, the optimization is only performed over a sliding window of last poses rather than the complete graph. This sliding window may further be reduced if exclusivly relative constraints are added as it is the case for visual odometry and gyroscope readings.

Figure 2: Sample image from the river sequence. The scene exhibits large variations in lighting as well as ambiguous patterns in the foliage, resulting in challenging conditions for visual odometry.

IV. Short Range Perception

The rotorcraft must be able to operate in the space between the water's surface and the tree canopy. In this cluttered space, a reliable short range perception system is necessary for obstacle avoidance and mapping. To measure 3D information about the environment, we use an off-axis rotating 2D laser line scanner. As seen in Fig. 3, the Hokuyo UTM-30LX is mounted with the scan plane tilted at 45° with respect to sweep motor axis.

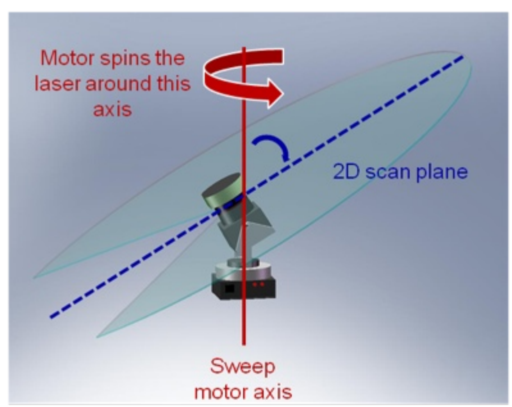


Figure 3: Off-axis spinning scanner

Other laser mounting and actuation configurations such as nodding, spinning on-axis, or roundly swinging [16], could not provide the same scan density or sensing field of view. Our configuration has the advantage of equal detection of horizontal and vertical obstacles and a full 360° field of view.

This scan pattern detects thin horizontal and vertical obstacles equally well as opposed to a nodding laser, which has difficulty detecting thin horizontal obstacles or an vertically-mounted on-axis spinning laser, which has difficulty detecting thin vertical obstacles. In a natural river environment, both thin vertical and horizontal tree branches should be expected and can be reliably sensed with our configuration.

V. Long Range Perception

The rotorcraft needs to be able to determine the course of the river and follow it. Also, river width measurements are needed to build the river map. The onboard laser sensor has an effective range limit of 15-20 meters which is sufficient for obstacle avoidance but not for guidance or mapping. Our solution to these problems is to use images from an onboard color camera. Images are segmented to find the extent of the river from the current viewpoint of the vehicle. Using knowledge of the vehicle's orientation and height above the river provided by the vehicle's onboard IMU and altimeter, the extent of the river in the image can be projected into a local coordinate frame. This forms a local map of the river for guiding the vehicle along the river's course. By registering and fusing many such local maps over time as the vehicle moves, the output of the segmentation is used to build a global map of the river.

The main challenge in detecting the extent of the river in images taken from a low-flying air vehicle is building a suitable appearance model for the water. Within a single image, the appearance of water can vary greatly due to reflections of the foliage and other structures on the bank, reflections of the sky and dark regions that fall in the shadows. In addition, ripples on the water's surface create variations in texture. This variability in river appearance in an image makes it hard to determine the extent of the river. Also there are large changes in water appearance with variation in weather and illumination conditions. All these factors make it difficult to learn a single classifier to detect river regions that does well in a variety of settings and which can generalize well to previously unseen environments. More details for our self-supervised river classification algorithm can be found in [17].

VI. Results

A. Vehicle State Estimation

Fig. 4 depicts the reconstructed path for different sensor suites, overlaid onto an aerial map of the area. For reference, we acquired highly accurate position information with a differential L1/L2 GPS, which is shown in green. The sequence spans about 1.9 km and roughly 10,000 frames.

Fig. 4a shows the results of the visual odometry for stereo image pairs taken at 15 Hz. For this sequence, the path sensed by visual odometry is smooth and locally accurate. As errors accumulate, it diverges from the reference path quite significantly. The path in Fig. 4b was estimated by fusing visual odometry and inertial measurements recorded at 100 Hz. Integrated gyroscope outputs were used to determine relative rotations, while accelerometers served as inclinometers to constraint the orientation in a global coordinate frame. Incorporating additional sensor information improves the results, but errors still accumulate as the sensors provide mostly relative motion informations. In Fig. 4c, position information from a low-cost L1 GPS receiver was incorporated into the estimation at a rate of 0.2 Hz. For most parts of the sequence, the resulting estimation lines up well with the reference path. Furthermore, the path is locally smooth, thus providing a suitable basis for laser point cloud registration.

B. 3D Point Cloud Mapping

Figure 5 shows examples of the 3D reconstruction built by the laser scanner as the vehicle moves through the environment. Each laser scan is globally registered and placed into a world map by using the filtered state estimate. Since the laser scans occur at a higher frequency than the state estimates, an intermediate state is found by interpolating between neighboring state estimates. The filtered state is locally smooth and accurate enough to build clean 3D reconstructions. The terrain mesh seen in the reconstruction is build from Elevation data and Orthoimagery provided by USGS and the Seamless Data Warehouse.

C. River Classification Map

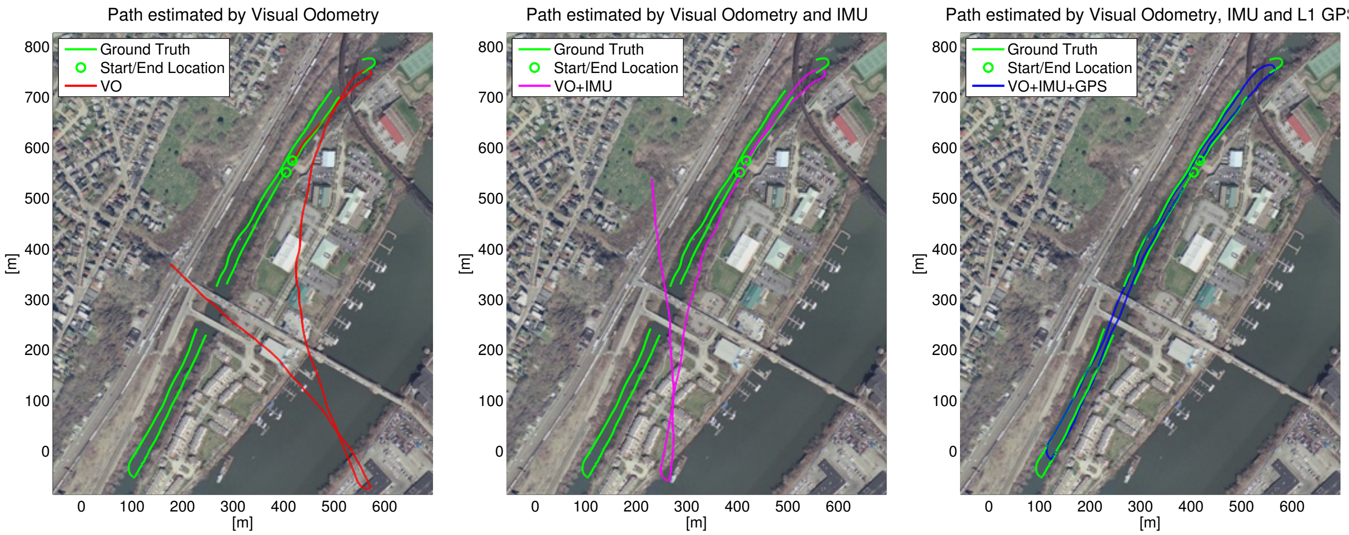
A sample classification results for a single image is shown in Fig. 6. The input image shows the river and an overhead bridge. A lot of the challenges of river imagery are visible in this image. There is a large contrast between distant river and river areas in the shadow. The river appearance itself changes significantly and the bank appearance changes. After classification a confidence of the river class is calculated for each patch. Based on that confidence and the range of the estimate an evidence grid of the shore can be calculated. As multiple classifications are integrated the confidence in the shore estimate will increase.

D. Integrated 3D and River Map

The final desired output of our exploration run is a 3D map describing the structure and a 2D bank giving an estimate of the bank. Figure 7a shows the map at the end of a 2km run moving up and down a channel with our sensor payload. The map shows that even though GPS data is only integrated sparsely (0.2Hz) it is still possible to create a globally registered map of the environment. The individual river classification results are integrated into a map in Fig. 7b. River classification images are calculated at 2Hz and then filtered using an evidence grid into a global representation. The current reliable maximum range for integrating the classification results into the map is about 30 meters and therefore in some regions the bank was not visible to the camera. This range depends on the geometry of the scene and will improve as our flight altitude increases from the current 2.5m to about 6-8m.

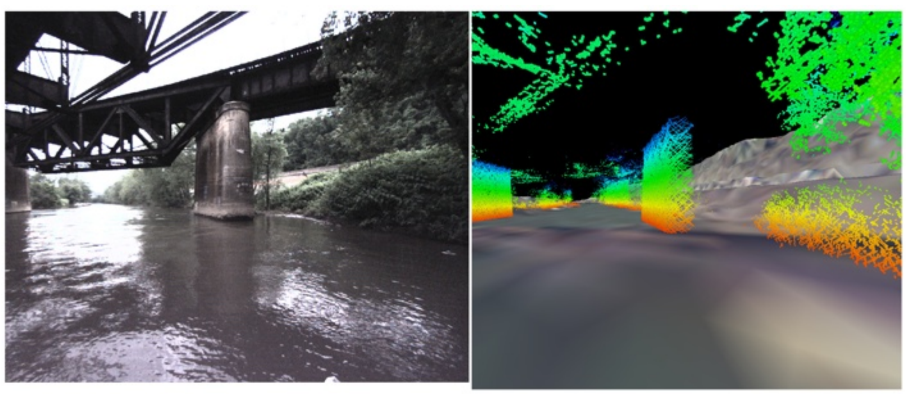
VII. Conclusions and Future Work

We have described the light-weight perception and state estimation systems required to navigate and map a river from

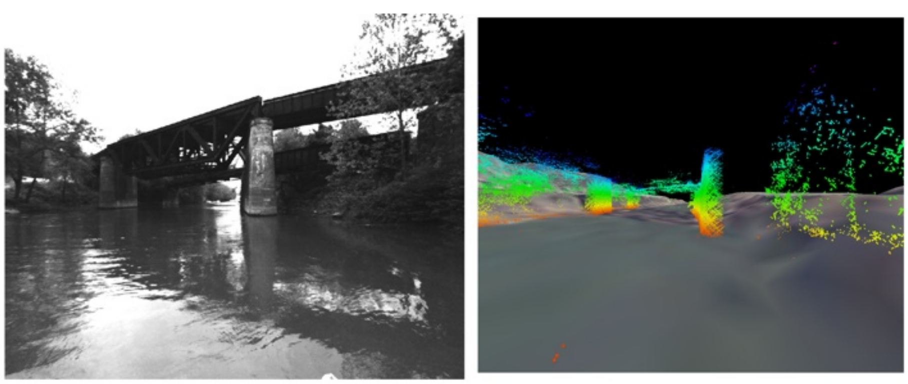


(a) (b) (c)

Figure 4: Estimated path for different sensor suites, overlaid onto an aerial map and a L1/L2 GPS reference path. Fig. 4a depicts the estimated path, if visual odometry is used exclusively. Fig. 4b displays the results, if inertial data is incorporated the estimation. In Fig. 4c intermittent position readings of an L1 GPS were used in addition.

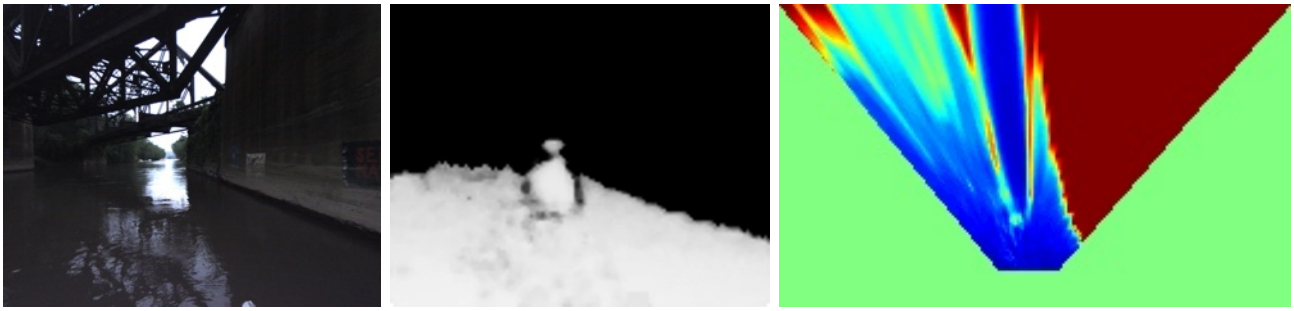


(a)



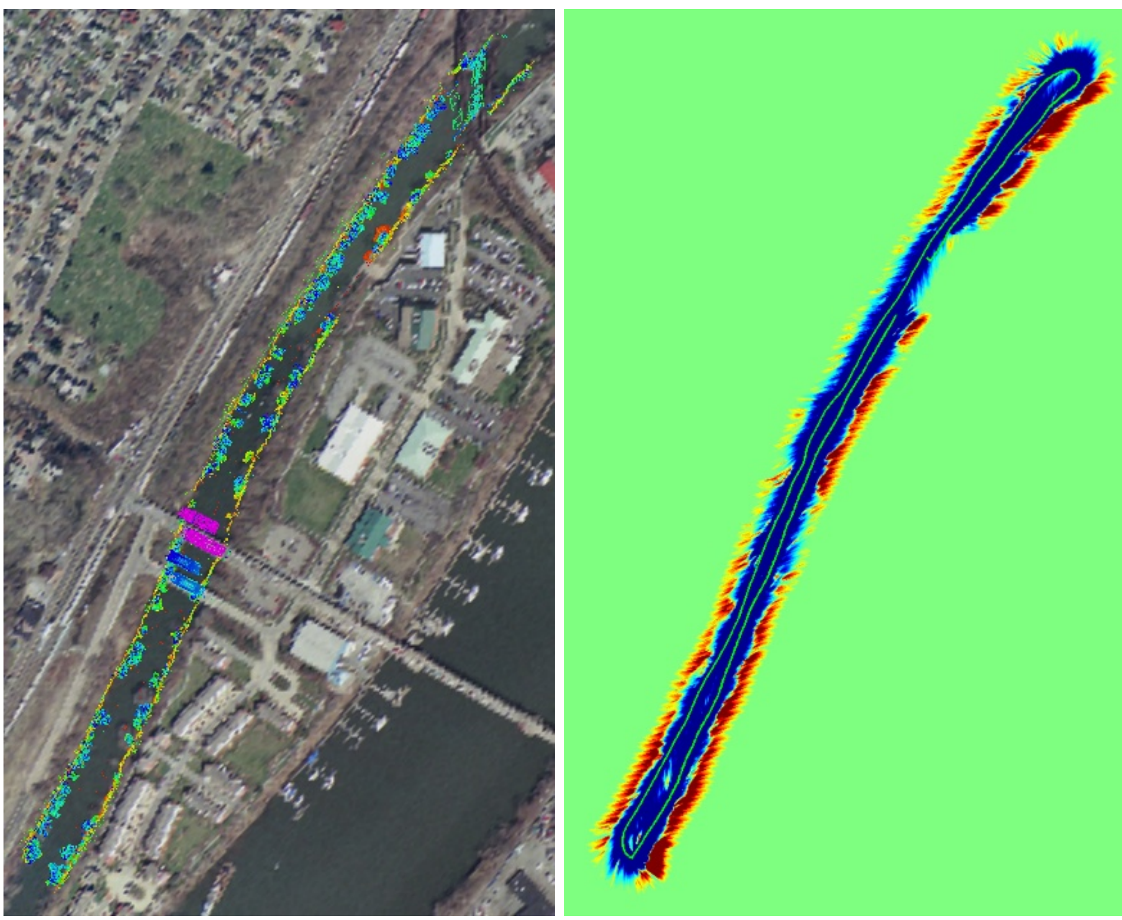
(b)

Figure 5: Example Image and Point Cloud Reconstruction. The left image shows an image of the camera. On the right one can see the reconstructed point cloud based on the integrated filter for pose estimation overlayed with an aerial image and elevation data.



(a) Input Image (b) River Likelihood (c) 2D Projection

Figure 6: Example River Classification Result. (a) Input to the algorithm. (b) Shows the classification output (white = river, black = no river). (c) Shows a top down projection of the image based on the pose estimation input and the likelihood from (b). (green= unknown, red = bank, blue = river).



(a) Point Cloud and Overhead Imagery. (b) River Classification Map.

Figure 7: Final 3D Point Cloud and River Maps. These maps show the resulting map of a 2km traverse.

a low-flying rotorcraft. We discussed our state estimation architecture and visual odometry approach. We showed results from our self-supervised river classification algorithm. Finally, we showed how we utilize the pose estimate to build 2D river maps and 3D point cloud maps of the river environment. These maps were shown to match the true environment. Future work will involve feeding the perception and state information back to a motion planner to realize autonomous exploration and mapping.

REFERENCES

[1] J. Leedekerken, M. Fallon, and J. Leonard, "Mapping complex marine environments with autonomous surface craft," in *12th International Symposium on Experimental Robotics*, 2010.

[2] S. Rathinam, P. Almeida, Z. Kim, S. Jackson, A. Tinka, W. Grossman, and R. Sengupta, "Autonomous searching and tracking of a river using an uav," in *American Control Conference, 2007. ACC '07*, 2007, pp. 359 –364.

[3] M. Maimone, Y. Cheng, and L. Matthies, "Two years of visual odometry on the mars exploration rovers," *Journal of Field Robotics*, vol. 24, no. 3, pp. 169–186, 2007.

[4] M. Pollefeys, D. Nistér, J. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. Kim, P. Merrell, *et al.*, "Detailed real-time urban 3d reconstruction from video," *International Journal of Computer Vision*, vol. 78, no. 2, pp. 143–167, 2008.

[5] K. Konolige and M. Agrawal, "Frameslam: From bundle adjustment to real-time visual mapping," *Robotics, IEEE Transactions on*, vol. 24, no. 5, pp. 1066–1077, 2008.

[6] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry for ground vehicle applications," *Journal of Field Robotics*, vol. 23, no. 1, pp. 3–20, 2006.

[7] E. Einhorn, C. Schröter, and H. Gross, "Monocular Scene Reconstruction for Reliable Obstacle Detection and Robot Navigation," *to appear Proc. ECMR*, 2009.

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

[9] C. Engels, H. Stewénius, and D. Nistér, "Bundle adjustment rules," *Photogrammetric Computer Vision*, vol. 2, 2006.

[10] B. Williams and I. Reid, "On combining visual SLAM and visual odometry," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 3494–3500.

[11] D. Scaramuzza, F. Fraundorfer, and R. Siegwart, "Real-time monocular visual odometry for on-road vehicles with 1-point ransac," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, 2009, pp. 4293–4299.

[12] A. Desai and D. Huber, "Objective evaluation of scanning ladar configurations for mobile robots," in *Proceedings of the IEEE/RSJ International Conference on Intelligent RObots and Systems (IROS)*, October 2009.

[13] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *IEEE Intelligent Vehicles Symposium*, Baden-Baden, Germany, June 2011.

[14] R. K
"ummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimization." *Journal of Autonomous Robots*, vol. 30, no. 1, pp. 25–39, 2011.

[15] M. Lourakis, "Sparse non-linear least squares optimization for geometric vision," *Computer Vision–ECCV 2010*, pp. 43–56, 2010.

[16] M. Matsumoto and S. Yuta, "3d laser range sensor module with roundly swinging mechanism for fast and wide view range image," in *Multisensor Fusion and Integration for Intelligent Systems (MFI), 2010 IEEE Conference on*, 2010, pp. 156 –161.

[17] S. Achar, B. Sankaran, S. Nuske, S. Scherer, and S. Singh, "Self supervised segmentation of river scenes," in *Intl. Conf. on Robotics and Automation (ICRA)*, 2011.