Large-Scale Robotic 3-D Mapping of Urban Structures

Michael Montemerlo and Sebastian Thrun

Stanford AI Lab
Stanford University
{mmde,thrun}@stanford.edu

Abstract. This article present results for building accurate 3-D maps of urban environments with a mobile Segway RMP. The goal of this project is to use robotic systems to rapidly acquire 3-D maps, which seamlessly integrate indoor and outdoor structures. Our approach is based on an information-solution of the SLAM problem, which enables us to seamlessly integrate GPS, IMU, and scan data. 3-D models acquired by the robot are analyzed for navigability using a multi-grid approach, and visualized using a level set technique. Results are presented for a number of environments, some of which combine indoor and outdoor terrain.

1 Introduction

In recent years, there has been a number of projects seeking to map physical environments. Classical work includes mapping from the air [11], the ground [6,23], indoors [5,10], outdoors [22], and even under water [27] and in the subterranean world [3]. The development of techniques for the acquisition of such maps has been driven by a number of desires. They include photo-realistic rendering [1,2], surveillance [26], measurement [3], and robot guidance [27]. Not surprisingly, the best work in this area has emerged from a number of different scientific fields, such as photogrammetry, computer vision, computer graphics [12,20], and robotics [23].

This paper describes a robotic system designed to acquire such maps. Urban terrain possesses a number of characteristic features: It combines large open places with narrowly confined spaces, such as building interiors. GPS is usually inaccurate due to multi-path effects, and places inside building are GPS-denied. From a SLAM (simultaneous localization and mapping) perspective, maps of the size targeted by our research involve 10^7 or more features; gathered over 10^6 poses. Urban terrain is non-flat, hence the robot has to be localized in 6-D. The overall SLAM problem, thus is orders of magnitude more complex than prior work. In fact, the vast majority of SLAM algorithms has only been applied to 2-D models with 3-D poses, to keep the data sets manageable small. Even those that perform 3-D mapping do so via 2-D SLAM [8], with the exception of [19] which offers no provision for closing cycles [7,4,24]. Further, past work has not provided effective means to incorporate occasional GPS measurements.

Fig. 1. The Segbot, a robot based on the Segway RMP platform and developed through the DARPA MARS program.

2 SLAM in Urban Environments

A key problem in building large-scale urban maps pertains to the ability to integrate information from multiple information sources, specifically GPS (global positioning system), IMU (the inertial measurement unit), odometry, and the LIDAR sensor(a laser-light detection and ranging sensor). This mapping problem is a version of the SLAM problem, short for simultaneous localization and mapping. The SLAM problem is characterized by a necessity to estimate the map of an environment while simultaneously localize the sensor relative to the map. Outdoors, GPS provided absolute position labels; indoors, it presently does not.

Our approach builds on prior work on SLAM by Lu and Milios, who proposed Kalman filter-based approach that represents SLAM posteriors through collections of local constraints between nearby poses [13] (see also [7]). However, this algorithm has been reported to suffer from numerical instabilities even for small data sets, and it also does not accommodate GPS measurements. Also related is recent work in [4,18,25], who propose variants of the information filter for solving the SLAM problem. These algorithms are approximate, and they also fail to integrate occasional GPS data when available. However, both families of approaches are related in that they represent SLAM posteriors through local constraints—which is in stark contrast to the classical SLAM solution, the EKF [21], which maintains a full covariance between any two features.

Specifically, our approach represents the SLAM posterior as an undirected Markov network, where nodes correspond to poses, GPS measurements, and range measurements. The network possesses three types of pairwise node potentials: There are potentials between range measurements and the corresponding pose, at which the measurement was required; there are potentials between subsequent poses, gov-

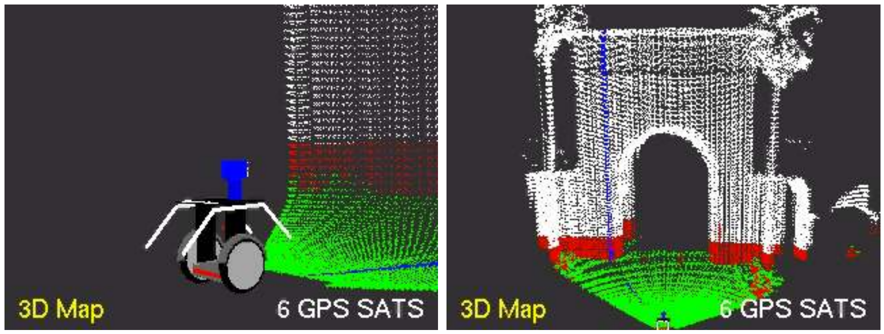


Fig. 2. Data acquisition through a two-directional scanning laser (the blue stripe indicates a vertical scan). The coloring indicates the result of terrain analysis: The ground surface is colored in green, obstacles are red, and structure above the robot's reach are shown in white.

erned by the IMU measurements. And finally, there are absolute location potentials for poses at which GPS data was received. All of these potentials are nonlinear-quadratic; they are composed of a deterministic non-linear projective function (e.g., the robot motion model; the measurement model) with a quadratic penalty function that measures deviations from this non-linear projection. This representation generalizes past work on SLAM, most notably [13], in that the resulting sum of potentials can be thought of as a non-normalized log-likelihood function. However, representing them as potentials avoids numerical instabilities of the covariance representation in [13].

The map is now retrieved by finding the minimum over all state variables in this graph. For this, it is essential to identify measurements that correspond; those are identified using scan matching. With a laser pointed forward on a panning platform, our scan matching algorithm relates scans to a large history window. Further, we use efficient grid-based caching mechanism to identify nearest neighbors in real-time. Once corresponding measurements are found, the resulting Bayes network is collapsed to remove double occurrences of joint landmarks (which is an approximate operation; it would be exact if all potentials were linear-quadratic). Next, all landmarks are integrated out by further shrinking the remaining Bayes network. We arrive at a skeleton network that only contains the path of the robot with potentials added between any two poses at which the same feature was observed. This Bayesian network is then "solved"—meaning we find the maximum likelihood solution in the corresponding probability function—through an efficient conjugate gradient search algorithm.

The advantage of this approach is threefold: It is free of numerical instabilities; it can represent extremely high-dimensional Gaussian SLAM posteriors in 10^8-dimensional space; and the resulting optimization is efficient: Generating an actual map takes in the order of seconds on a low-end workstation. The process performs the mapping in 6-D [19]. The curvature of seemingly urban flat terrain tends to be sufficiently non-flat that SLAM approaches that assume the robot operates on a plane

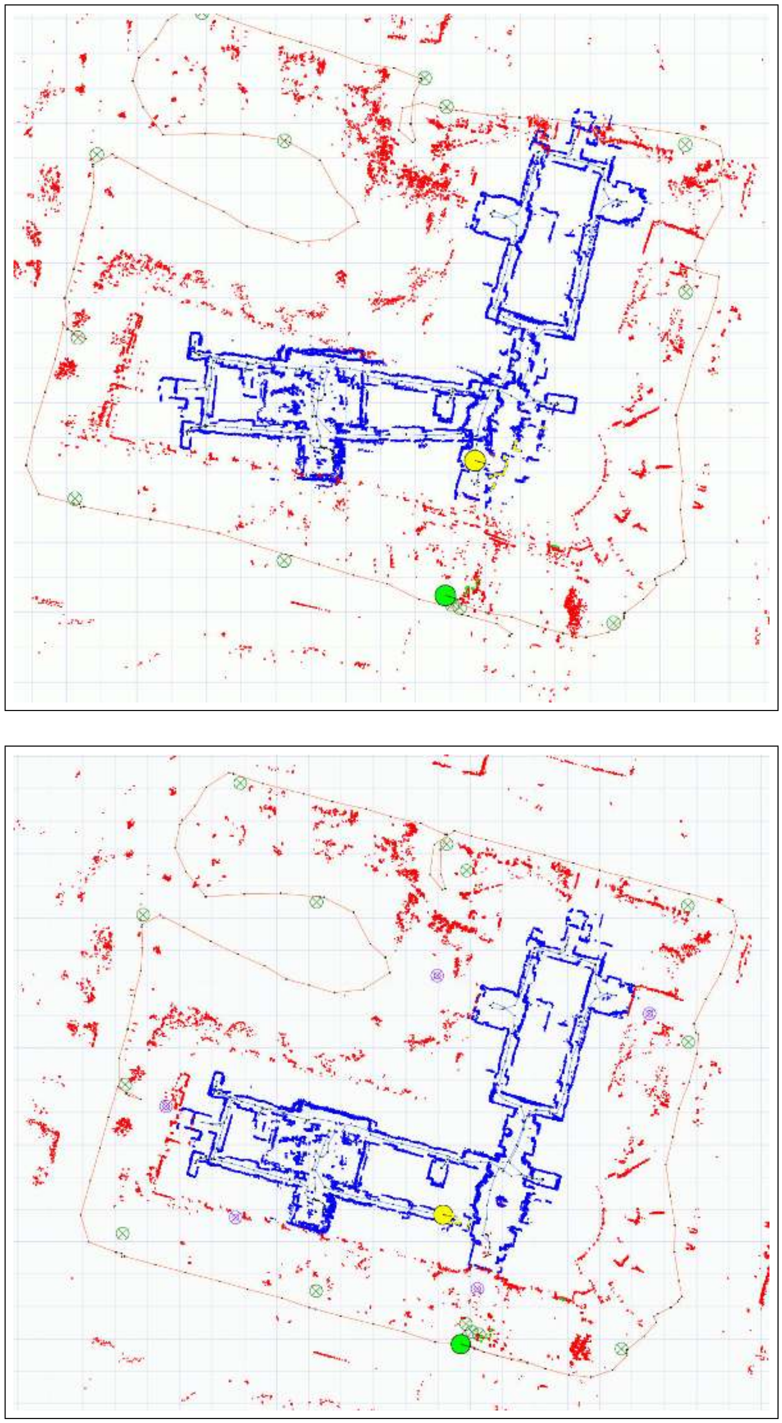
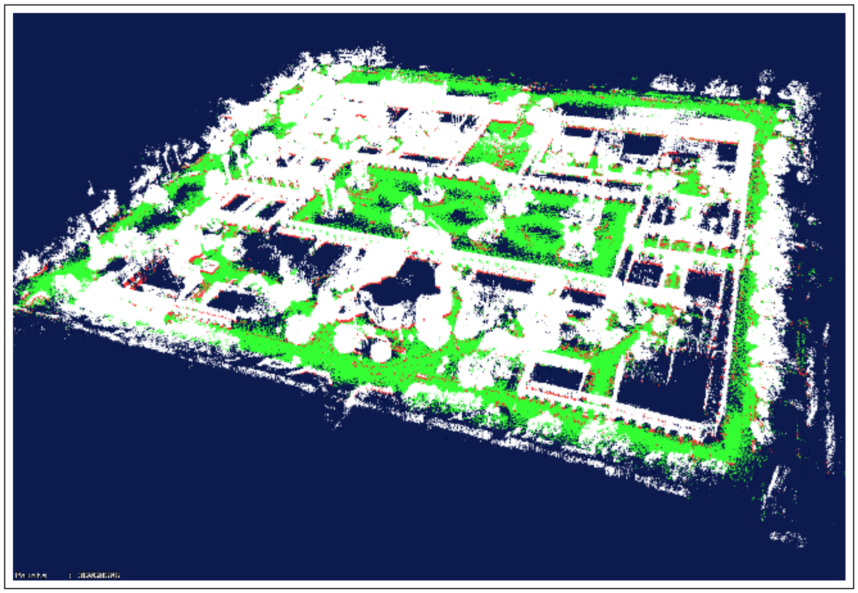


Fig. 3. Indoor mapping. Left: just based on the IMU and SLAM. Right: factoring in GPS data acquired outdoors. This experiment highlights the utility of our hybrid SLAM algorithm that factors in GPS measurements as available.



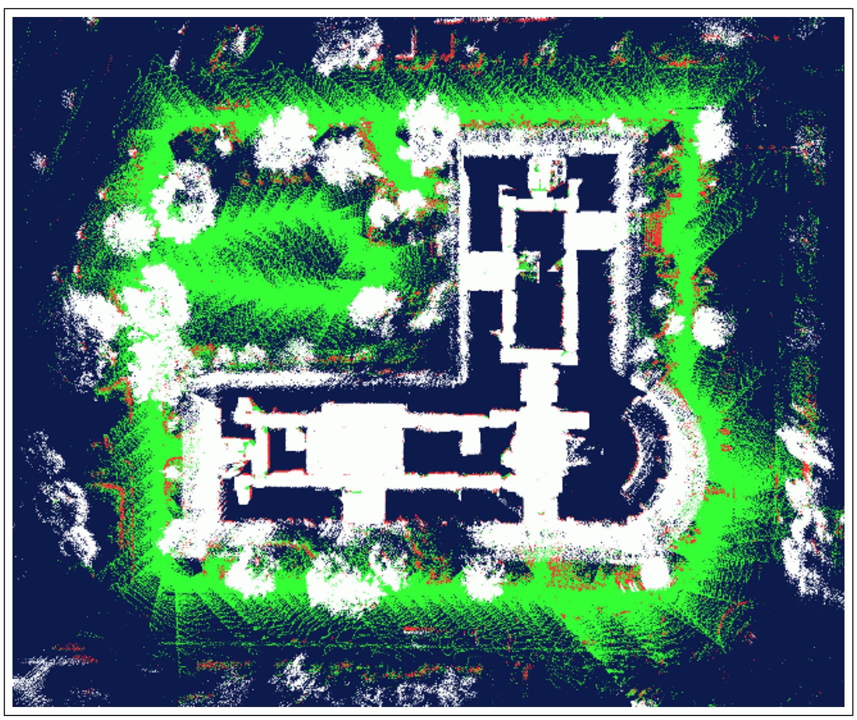


Fig. 4. Top: A map of Stanford University's main campus, whose diameter is approximately 600 meters. Bottom: 3-D map of the Gates Computer Science building and the surrounding terrain.

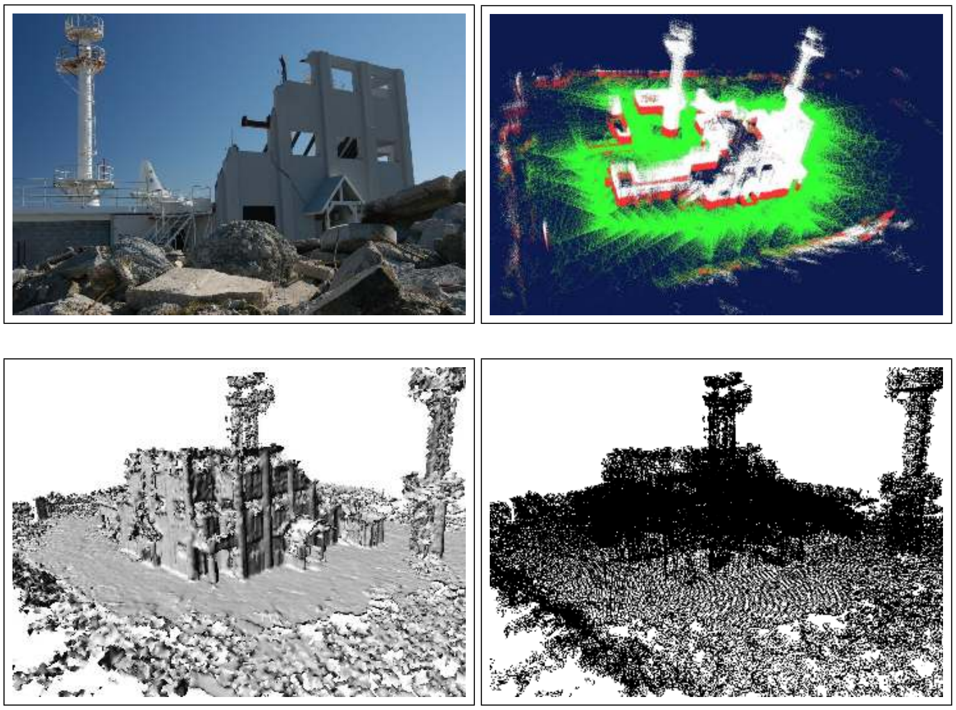


Fig. 5. Visualization of the NASA Ames ARC Disaster Assistance and Rescue Team training site in Moffett Field, CA. This site consist of a partially collapsed building with two large observation platforms. Top: Model. Bottom: Edges.

are simply inapplicable. The 6-D optimization also accommodates the fact that the Segway's pitch is continuously adjusted so as to not lose balance.

3 Navigation

A key component of our approach pertains to the "understanding" of the terrain and surrounding urban structures, to the extent necessary for safe robot navigation. Since our robot sometimes navigates autonomously, our approach also analyzes terrain for possible obstacles, using an algorithm that generalizes related work in [9,14].

Our basic approach analyzes scans for three type of obstacles: (1) terrain that is too steep or too rugged for save traversal such as curbs; (2) obstacles protruding into the workspace of the robot such as overhangs, and (3) locations that lacks the necessary physical support for the robot such as wholes in the ground. The first two obstacle types are "positive obstacles," meaning that they can be detected by analyzing scan measurement points. The third type is commonly called "negative obstacle," to indicate that such obstacles are only detectable by the absence of range measurements.

Our approach identifies positive obstacles by analyzing individual ground scan lines. Each scan line contains a sequence of measurement points. By calculating the derivative of these points in workspace coordinates, our robot can assess the steepness of individual ground patches. In this way, it can avoid obstacles such as steep ramps and upwards staircases. Objects protruding into the robot's workspace are identified by searching up to a limited height for measurement points relative to the ground area. By limiting the search height, our vehicle can navigate indoor environments and through overhangs.

Negative obstacles are detected by lack of supporting ground plane. However, this analysis is somewhat non-trivial, due to the sparseness of measurement points at longer ranges. To perform this analysis, most existing techniques partition the workspace into a grid, similar to the well-known occupancy grid map algorithm [17]. For each grid cell, sensor measurements are integrated using Bayes rule to gradually increase coverage, while at the same time diminishing the effect of sensor noise.

Real-world terrain sensors have limited measurement resolution. For example, our laser range finders can only measure ranges with $0.5°$ accuracy; similar limitations exist for stereo camera systems and sonar sensors. Limited resolution causes two problems with standard evidence grid algorithms: First, the limited resolution may make it impossible to detect small obstacles at a distance. Obstacles like curbs or low-hanging wires can usually only be detected at close range. Second, limited sensor resolution makes it difficult to systematically find navigable terrain at a distance. As a result, a motion planner is either forced to make optimistic assumptions about the nature of terrain at a distance (with the obvious expense of having to replan when negative obstacles are encountered), or must remain confined to nearby regions that have been completely imaged.

Our approach relies on a multi-grid representation, which combines maps with different resolutions. The map chosen for each measurement depends on the overall range: the further away a measurement point, the coarser the corresponding grid. The advantage of using such a multi-resolution grid is two-fold. First, the coverage in areas further away is increased, without compromising the spatial distribution of the overall map. This leads to improved paths of the robot during autonomous motion. Second, and possibly more importantly, evidence of non-traversibility acquired at short range cannot be overridden by evidence of traversibility acquired at longer range. This effect is the result of using gradients for traversibility analysis: such gradients are necessarily less accurate at a distance, where the density of measurements is reduced. As a result, small obstacles such as curbs are usually not detected at far range. A standard occupancy grid technique would consequently fail to update a cell as non-traversable while the obstacle is still far away. Results in [16] suggest that the time at which small obstacles are detectable at short range may be too short for overriding the evidence acquired at longer ranges; as a result, the robot may run into such obstacles. Our multi-resolution approach overcomes this by maintaining range-specific maps. Small obstacles such as curbs do not show up in the coarse, long-range map, but they do show up in the fine-grained short-range

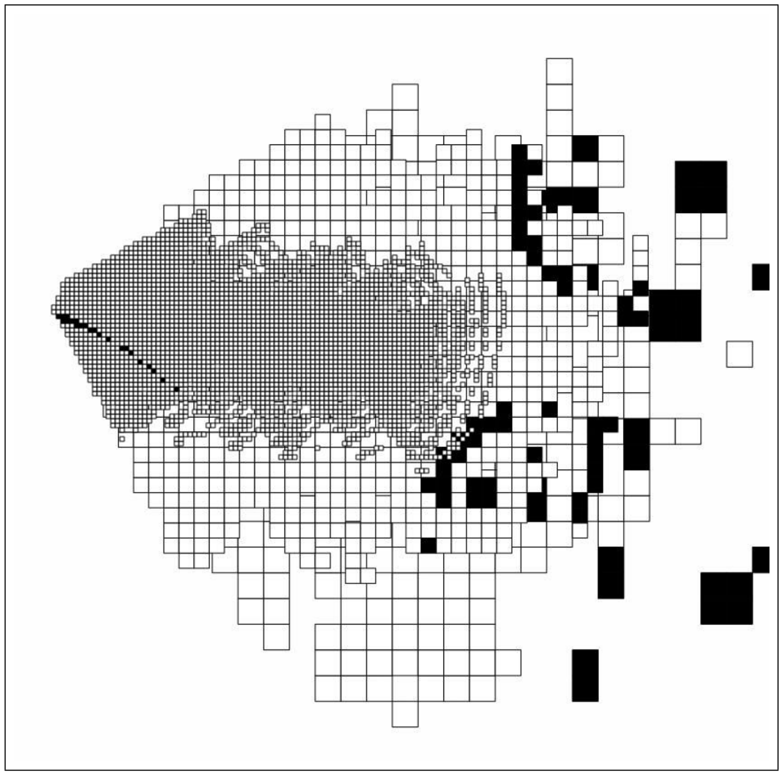


Fig. 6. Multi-resolution pyramid model of the robot's surroundings. The majority of holes in the map are filled in. The terrain map close to the robot is very high resolution, while the area far from the robot is very coarse.

map. When combining maps for assessing the navigability of terrain, preference is given to shorter range maps; however, all maps participate in motion planning.

4 Visualization

To visualize the resulting maps, we use a well-known level set technique. Technically, the data acquired by our robot consists of point clouds. Such point clouds might provide an impression of the surface structure of a building, but they lack an explicit surface description.

Our approach....

5 Results

We conducted a number of experiments, all with the vehicle shown in Figure 1. In particular, we have mapped a number of urban sites, including NASA's Search and

Rescue Facility DART and a large fraction of Stanford's main campus; snapshots of these experiments will be discussed below.

Our experiments either involve the collection of a single large dataset, or a number of datasets. The latter has become necessary since for the environments of the size studied here, the robot possesses insufficient battery capacity to collect all data within a single run. In most experiments, the robot is controlled manually. This is necessary because the urban environments are usually populated with moving objects, such as cars, which would otherwise run danger of colliding with our robot. We have, on several occasions, used our navigation package Carmen [15] to drive the robot autonomously, validating the terrain analysis techniques discussed above.

Our research has led to a number of results. First and foremost, a primary finding is that with our representation, maps with more than 10^8 variables can be computed quickly, even under multiple loop-closure constraints. The time for thinning the network into its skeleton tends to take linear time in the number of robot poses, which is the same order as the time required for data collection. We find that scan matching is easily achieved in real-time, as the robot moves, using a portable laptop computer. This is a long-known result for horizontally mounted laser range finders, but it is reassuring that the same applies to the more difficult scan matching problem involving a vertically panning scanner. More importantly, the relaxation of the pose potentials takes in the order of 30 seconds even for the largest data set used in our research, of an area 600m by 800m in size, and with a dozen cycles. This suggests the appropriateness of our representation an algorithms for large-scale urban mapping.

The second and possibly more important result pertains to the utility of GPS data for indoor maps. We find that indoor maps become more accurate when some of the data is collected outdoors, where GPS measurements are available. Further below, we will discuss an experimental snapshot that documents this result..

Finally, when in autonomous mode, we find that our terrain analysis techniques provide effective navigation at speeds of up to one meter per second. The vehicle navigates collision-free and successfully avoids negative obstacles, but sometimes fails to detect fast moving obstacles fast enough to avoid a collision. The latter is because of the panning motion of the sensor, which requires approximately 2 seconds for a full sweep. We also find that curbs are reliably identified through our multi-resolution approach, where a single-resolution approach using occupancy-grid style update techniques fail.

Experimental snapshots can be found in Figures 3 through 5. Figures 4 and 5 show some of the maps acquired by our system. All maps are substantially larger than previously software could handle, all are constructed with some GPS information. The map shown on the left in Figure 4 corresponds to Stanford's main campus; the one on the right is an indoor-outdoor map of the building that houses the computer science department.

The key result of improved indoor maps through combining indoor and outdoor mapping is illustrated in Figure 3. Here we show a 2-D slice of the 3-D map using SLAM under two different conditions: In the map on the left, the indoor map is constructed independently of the outdoor map, whereas the right map is constructed jointly. As explained, the joint construction lets GPS information affect the building interior through the sequence of potentials liking the outdoor to the indoor. As this figure suggests, the joint indoor-outdoor map is significantly more accurate; in fact, the building possesses a right angle at its center, which well approximated.

Figure 6 shows a snapshot of our multi-resolution grid map for finding negative obstacles. The resolution depends on the distance to the robot. This specific snapshot shows several curbs, some larger ones far away, and one near the robot that a flat approach would have failed to identify.

Acknowledgement

The research has been sponsored by DARPA's MARS Program (contracts N66001-01-C-6018 and NBCH1020014), which is gratefully acknowledged.

References

1. P.K. Allen and I. Stamos. Integration of range and image sensing for photorealistic 3D modeling. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 1435–1440, 2000.
2. R. Bajcsy, G. Kamberova, and L. Nocera. 3D reconstruction of environments for virtual reconstruction. In *Proc. of the 4th IEEE Workshop on Applications of Computer Vision*, 2000.
3. C. Baker, A. Morris, D. Ferguson, S. Thayer, C. Whittaker, Z. Omohundro, C. Reverte, W. Whittaker, D. Hähnel, and S. Thrun. A campaign in autonomous mine mapping. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2004.
4. M. Bosse, P. Newman, M. Soika, W. Feiten, J. Leonard, and S. Teller. An atlas framework for scalable mapping. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2003.
5. S. El-Hakim, P. Boulanger, F. Blais, and J.-A. Beraldin. Sensor based creation of indoor virtual environment models. In *Proc. of the 4th Internetional Conference on Virtual Systems and Multimedia (VSMM)*, Geneva, Switzerland, 1997.
6. A. Elfes. Sonar-based real-world mapping and navigation. *IEEE Journal of Robotics and Automation*, RA-3(3):249–265, June 1987.
7. J.-S. Gutmann and K. Konolige. Incremental mapping of large cyclic environments. In *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA)*, 2000.
8. D. Hähnel, W. Burgard, B. Wegbreit, and S. Thrun. Towards lazy data association in SLAM. In *Proceedings of the 11th International Symposium of Robotics Research (ISRR'03)*, Sienna, Italy, 2003. Springer.
9. K. Hashimoto and S. Yuta. Autonomous detection of untraversability of the path on rough terrain for the remote controlled mobile robots. In *Proceedings of the International Conference on Field and Service Robotics*, Lake Yamanaka, Japan, 2003.
10. L. Iocchi, K. Konolige, and M. Bajracharya. Visually realistic mapping of a planar environment with stereo. In *Proceesings of the 2000 International Symposium on Experimental Robotics*, Waikiki, Hawaii, 2000.

11. G. Konecny. *Geoinformation: Remote Sensing, Photogrammetry and Geographical Information Systems*. Taylor & Francis, 2002.
12. M. Levoy. The digital michelangelo project. In *Proc. of the Second International Conference on 3D Imaging and Modeling*, 1999.
13. F. Lu and E. Milios. Globally consistent range scan alignment for environment mapping. *Autonomous Robots*, 4:333–349, 1997.
14. L. Matthies, T. Litwin, K. Owens, A. Rankin, K. Murphy, D. Coorobs, J. Gilsinn, T. Hong, S. Legowik, M. Nashman, and B. Yoshimi. Performance evaluation of ugv obstacle detection with ccd/flir stereo vision and ladar. In *Proceedings of the Joint Conference on the Science and Technology of Intelligent Systems*, New York, NY, 1998.
15. M. Montemerlo, N. Roy, and S. Thrun. Perspectives on standardization in mobile robot programming: The carnegie mellon navigation (CARMEN) toolkit. In *Proceedings of the Conference on Intelligent Robots and Systems (IROS)*, 2003.
16. M. Montemerlo and S. Thrun. A multi-resolution pyramid for outdoor robot terrain perception. In *Proceedings of the AAAI National Conference on Artificial Intelligence*, San Jose, CA, 2004. AAAI.
17. H. P. Moravec. Sensor fusion in certainty grids for mobile robots. *AI Magazine*, 9(2):61–74, 1988.
18. P. Newman. *On the Structure and Solution of the Simultaneous Localisation and Map Building Problem*. PhD thesis, Australian Centre for Field Robotics, University of Sydney, Sydney, Australia, 2000.
19. A. Nüchter, H. Surmann, K. Lingemann, J. Hertzberg, and S. Thrun. 6D SLAM with application in autonomous mine mapping. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2004.
20. S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Proc. Third International Conference on 3D Digital Imaging and Modeling (3DIM)*, Quebec City, Canada, 2001. IEEEComputer Society.
21. R.C. Smith and P. Cheeseman. On the representation and estimation of spatial uncertainty. *International Journal of Robotics Research*, 5(4):56–68, 1986.
22. S. Teller, M. Antone, Z. Bodnar, M. Bosse, S. Coorg, M. Jethwa, and N. Master. Calibrated, registered images of an extended urban area. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.
23. S. Thrun. Robotic mapping: A survey. In G. Lakemeyer and B. Nebel, editors, *Exploring Artificial Intelligence in the New Millenium*. Morgan Kaufmann, 2002.
24. S. Thrun, W. Burgard, and D. Fox. A real-time algorithm for mobile robot mapping with applications to multi-robot and 3D mapping. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, San Francisco, CA, 2000. IEEE.
25. S. Thrun, Y. Liu, D. Koller, A.Y. Ng, Z. Ghahramani, and H. Durrant-Whyte. Simultaneous localization and mapping with sparse extended information filters. *International Journal of Robotics Research*, 2004. To Appear.
26. C.-C. Wang, C. Thorpe, and S. Thrun. Online simultaneous localization and mapping with detection and tracking of moving objects: Theory and results from a ground vehicle in crowded urban areas. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2003.
27. S. Williams, G. Dissanayake, and H.F. Durrant-Whyte. Towards terrain-aided navigation for underwater robotics. *Advanced Robotics*, 15(5), 2001.