Feature-based Mapping with Grounded Landmark and Place Labels

Alexander J. B. Trevor, John G. Rogers III, Carlos Nieto-Granda, Henrik I. Christensen
Robotics & Intelligent Machines, Georgia Institute of Technology

***Abstract*—Service robots can benefit from maps that support their tasks and facilitate communication with humans. For efficient interaction, it is practical to be able to reference structures and objects in the environment, e.g. "fetch the mug from the kitchen table." Towards this end, we present a feature-based SLAM and semantic mapping system which uses a variety of feature types as landmarks, including planar surfaces such as walls, tables, and shelves, as well as objects such as door signs. These landmarks can be optionally labeled by a human for later reference. Support for partitioning maps into labeled regions or places is also presented.**

I. Introduction

Service robots need to have maps that support their tasks. Traditional robot mapping solutions are well-suited to supporting navigation and obstacle avoidance tasks by representing occupancy information. However, it can be difficult to enable higher-level understanding of the world's structure using occupancy-based mapping solutions. One of the most important competencies for a service robot is to be able to accept commands from a human user. Many such commands will include instructions that reference objects, structures, or places, so our mapping system should be designed with this in mind.

Towards this goal, we present GTmapper, a feature based SLAM and semantic mapping system. In contrast to grid based mappers or visual SLAM systems that use features such as SIFT keypoints, our mapper utilizes landmarks that correspond to entities that are meaningful to humans as well as the robot, such as walls, tables, shelves, or objects such as door signs. In conjunction with a human operator, GTmapper allows for labeling of these structures and objects. It also supports partitioning the map into labeled spaces, with the assistance of a human providing the labels. Landmarks such as door signs make labeling particularly easy, as the appropriate label can be read directly from the landmark using optical character recognition (OCR), without the need for human assistance.

We believe that our map representation is well suited to grounding of labels, as each landmark used by our mapper is a distinct entity in the world, and can optionally be labeled by a human. We present a summary of our work on this mapping system, and discuss its suitability for tasks that require spatial dialog.

II. Related Work

The most closely related line of research to this work is the previous work on semantic mapping which focuses on creating maps that capture a higher level of understanding and meaning, such as [5] and [1]. There has also been a great deal of previous work on segmentation of indoor environments into relevant parts for use by service robots, such as [8]. Additionally, another related focus has been on the use of language for referencing spatial information contained in maps, such as understanding of natural language directions with respect to a map, as presented in [4].

For a more detailed treatment of the related work, see [9], [6], and [10].

III. Mapping System

Our feature-based mapping system makes use of the GTSAM (smoothing and mapping) library of Dellaert [2]. GTSAM solves the smoothing and mapping problem using factor graphs that relate landmark poses to robot poses. The factors are nonlinear measurements produced by measuring various features. New factor types can be defined and used by specifying a measurement function along with its derivatives with respect to the robot pose and the landmark pose. The resulting factor graph can be used to optimize the positions of the landmarks as well as the robot poses. Note that we are solving the *full SLAM* problem, recovering not just the current robot pose, but the full trajectory along with the landmark poses.

Our mapping system makes use of the GTSAM optimization and inference tools, and defines a variety of feature types that can be used for SLAM and semantic mapping. We have defined features which correspond to discrete entities in the world, such as planar surfaces or door signs, which can optionally be labeled. A detailed description of earlier work on this system using 2D line measurements as features is presented in [9]. Here we will present our recent work, focusing on semantic labeling of landmarks and places. An example demonstrating the map representation is shown in Figure 1.

The robot used to create these maps is the Jeeves robot, shown in Figure 2. Jeeves is comprised of a Segway RMP-200 mobile base, and is equipped with a variety of sensors. A SICK LMS-291 laser scanner is used for obstacle avoidance, as well as a 3D laser scanner comprised of a Hokuyo UTM-30LX laser scanner on a Directed Perception DP-47-70 pan-tilt unit, and a Prosilica camera used for object recognition. The platform is also equipped with a parallel jaw gripper mounted on a 1-DOF linear actuator, which allows basic manipulation tasks when combined with the Segway's additional degrees of freedom.

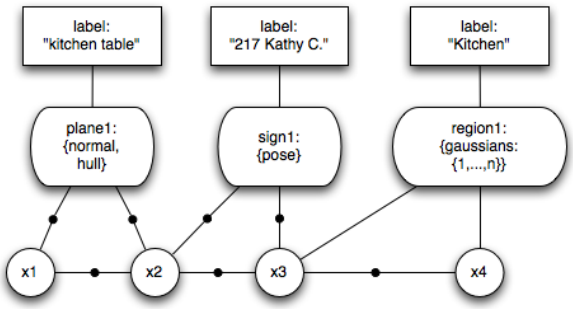

Fig. 1. An example of the type of map representation used by our system, including a robot trajectory (poses x_1 ... x_4), a planar feature with attached label, a door sign feature with a label, and a labeled region. Labels are optional for planar features and sign features. Also note that the regions are not a part of the SLAM factor graph, as they are not used as landmarks for SLAM.

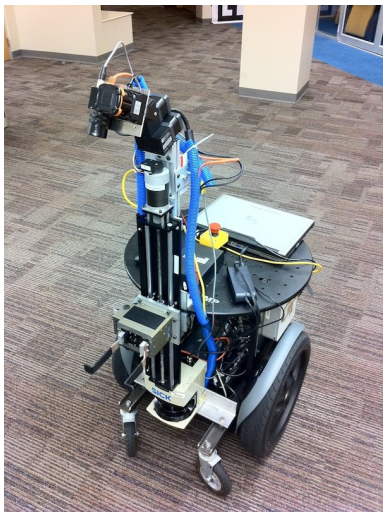

Fig. 2. Jeeves, the robot platform used in this work. The 2D SICK laser scanner is used for obstacle avoidance and place labeling, while the Hokuyo laser scanner combined with the pan-tilt unit is used to generate 3D point cloud data, used for planar surface mapping. The camera used for door sign recognition can also be found on the pan-tilt head.

A. Planes

One type of landmark used by our system is planar surfaces extracted from point cloud data. 3D laser scanners or RGB-D sensors such as the Microsoft Kinect can be used to collect suitable data. Planes are then extracted from the point cloud by an iterative RANdom SAmple Consensus (RANSAC) method, which allows us to find all planes meeting constraints for size and number of inliers. A clustering step is also performed on extracted planes in order to separate multiple coplanar surfaces, such as two tables with the same height, but at different locations. We make use of the Point Cloud Library (PCL) for much of our point cloud processing.

Planes can be represented by the well known equation: $ax + by + cz + d = 0$. Our mapper then represents the planes as: $p = \left[n, hull\right]$ where: $n = \left[a, b, c, d\right]$ and $hull$ is a point cloud of the vertices of the plane's convex hull. As the robot platform moves through the environment and measures planes multiple times, the mapped planes' hulls will be extended with each new portion seen, allowing the use of large planes such as walls where the full extent is typically not observed in any single measurement.

This type of plane can then be used for localization purposes by using the surface normal and perpendicular distance from the robot. In addition to being useful for localization, we believe that these surfaces are also useful for communication with humans. Many service robot tasks may require interaction with objects on horizontal planar surfaces, such as tables or shelves, and navigational tasks may require an understanding of planar surfaces such as walls or doors. In order to support such tasks, our mapping system allows planar surfaces to optionally support a label, such as "kitchen table" or "Henrik's desk," so that they may be easily referenced by a human user. Labels are entered interactively via a command line application. Planes corresponding to walls, the floor, or the ceiling can also be labeled, and multiple planes can share the same label as well. For example, one could label two walls of a hallway as "front hall," which gives the robot an idea of the extent of this structure.

Preliminary results on maps that represent the locations and extent of this type of planar feature are presented in [10]. More recent work includes the use of these features as landmarks for our SLAM, as described above. A map comprised of planar surfaces is shown in Figure 3, and a top-down view is shown in Figure 4 after a large correction was made following a loop-closure. A close-up view of some labeled planar features is shown in Figure 5.

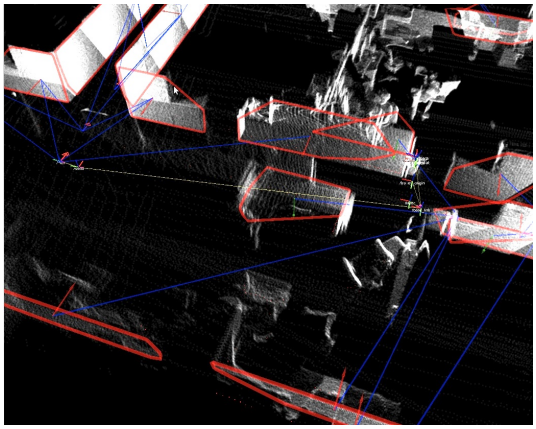

Fig. 3. An example of a map composed of planar surfaces. The mapped area shows several cubicles, with walls, cubicle walls, and desks used as landmarks. The convex hulls of the planar regions are shown in red, and blue lines represent measurements, indicating which poses features were measured from. Surface normals are shown in red for vertical surfaces, and green for horizontal surfaces. The full point clouds are displayed in white, for visualization purposes only; only the extracted planes are used for mapping and localization.

B. Signs

Door signs commonly found outside of offices are another type of landmark supported by our mapper. These are par-

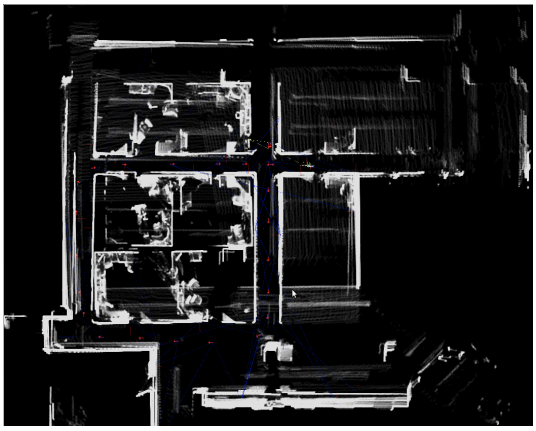

Fig. 4. Another view of a map composed of planar surfaces, after a loop closure has taken place. The full point clouds are displayed for visualization only. Only the extracted planes are used for mapping and localization.

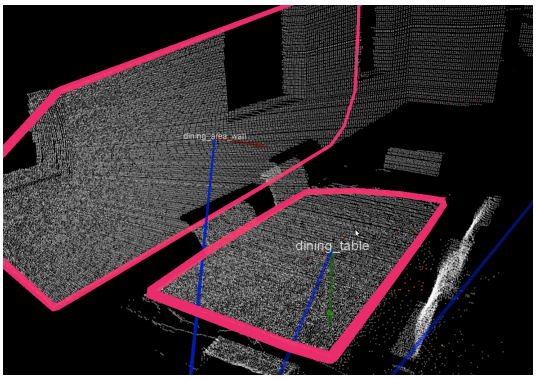

Fig. 5. A visualization of a planar map that include labels. The horizontal table is labeled as "dining_table," and the back wall is labeled as "dining_area_wall."

ticularly interesting from a semantic perspective because the relevant label can be read directly from the object using OCR.

In our previous work [7], signs are recognized in images by first extracting salient regions using the spectral-residual technique of Hou and Zhang in [3]. A Histogram of Oriented Gradients (HOG) feature is computed from each of the salient regions in the image and it is classified by an SVM which was trained on HOG features of signs. If the SVM classifies this HOG feature as a sign, then the text on the image is read by an OCR routine to provide a label; currently our implementation makes a request to the online service *Google Goggles* for this purpose.

Signs and their associated labels are provided as measurements to the mapper where laser scan data is fused to generate a 3D point measurement. Semantic data association between new measurements and mapped signs is performed by matching text strings by analyzing the longest common substring (LCS). The use of the LCS as a similarity measure allows for some minor errors in OCR to occur while still permitting data association.

To validate this technique, we performed a series of large loop-closure runs in an office building. In these test runs, the robot makes observations of door signs as in Figure 6 which enable it to close large loops. An example map can be seen in Figure 7. The semantic text labels on door signs are used for data association and could also be used for grounding human-robot dialogue, e.g. "Robot: Go to room 213".

Fig. 6. An example of an image used to generate a measurement of a door sign landmark. The text read by the OCR program is displayed.

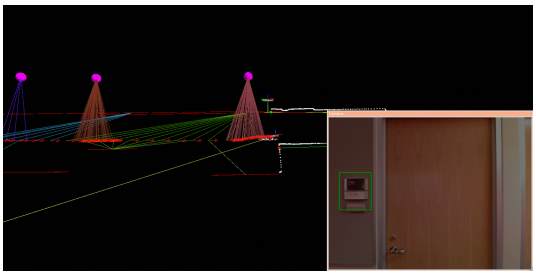

Fig. 7. An example of a map using door sign landmarks. The pink spheres denote the 3D locations of the signs, with lines showing the poses from which each sign was measured. The thick red lines are 2D walls as measured by the robot's laser scanner. The current image is shown as an inset, with the detected sign indicated by the green box.

C. Places

While labeling discrete features is quite useful, it can be helpful for some tasks to have a means of associating a label with a region of a map. Towards this end, we have created a means of interactively labeling places within a map. Our system partitions the map into sets of Gaussian models based on the 2D laser scanner's current view. As the robot moves throughout the map, a human user can enter a label for the current location via a command line application. When a label is provided, a Gaussian is fit to the 2D laser scanner's most recent point cloud, and tagged with the entered label. Each labeled region is represented by one or more such Gaussians in the metric map's coordinate frame, which can produce complex decision boundaries when many locations are tagged with the same label. The result is a collection of regions which represent places in our semantic map. An example of such a partition is shown in Fig. 8. The robot can then use this map to determine its current location by calculating the Mahalanobis distance to each Gaussian region and give the label of the nearest one. The user can also request that the robot move to

a region with a specific label, and it will plan a path and move to the nearest region with the requested label.

In order to evaluate our approach, experiments were performed both in simulation and using our robot. Our work on this is presented in [6]. We designed two simulated environments in which the robot can be taught locations and asked to navigate between them. These experiments consisted of a human user providing labels for many regions, including rooms and hallways. An example of the decision boundaries of using this approach for a simulated experiment is shown in Figure 8.

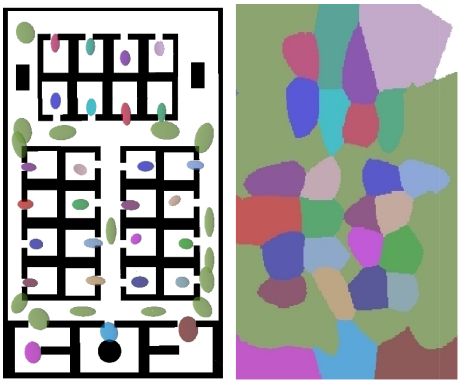

Fig. 8. A simulated environment with many rooms that has been partitioned into several regions is shown on the left, with the corresponding decision boundaries shown on the right.

IV. Conclusion

We have summarized our work on a mapping system capable of using a variety of semantically relevant landmarks as features. The locations and extent of planar surfaces such as tables, shelves, and walls can be represented, labeled, and used for localization and mapping. Door signs can be detected and read by OCR software, and also used for mapping. Finally, these maps can be partitioned into labeled spaces in order to support tasks that require dialogs regarding places in a map, as opposed to specific structures, e.g. "the kitchen" rather than "the kitchen table."

Our efforts thus far have primarily focused on the construction of maps that capture relevant information about spatial structure, as well as applying labels to the appropriate structures or regions. In order to appropriately ground labels in the map, we believe that segmenting the world into meaningful landmarks is a good approach. We have shown our mapper's ability to create maps suitable for localization, and have shown preliminary support for tasks that involve dialog regarding labeled surfaces, objects, and places.

Although we believe that our map representation is well suited for engaging in spatial task related dialogs, our use of the labels so far has been limited to very simple commands such as "go to label." Detailed experiments and user studies on the use of these maps in the context of service robotic tasks are left as future work. As additional future work, we hope to employ a more advanced dialog system allowing a wider range of spatial commands to be understood, such as natural language directions, as in Kollar *et. al* [4].

Acknowledgments

This work was made possible through the Boeing corporation and ARL MAST CTA project 104953. We would also like to thank the reviewers for their helpful comments.

References

[1] P. Beeson, M. MacMahon, J. Modayil, A. Murarka, B. Kuipers, and B. Stankiewicz. Integrating multiple representations of spatial knowledge for mapping, navigation, and communication. In *Symposium on Interaction Challenges for Intelligent Assistants*, AAAI Spring Symposium Series, Stanford, CA, March 2007. AAAI Technical Report SS-07-04.

[2] F. Dellaert and M. Kaess. Square root SAM: Simultaneous localization and mapping via square root information smoothing. *International Journal of Robotics Research*, 25(12):1181–1204, 2006.

[3] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. *CVPR*, 2007.

[4] T. Kollar, S. Tellex, D. Roy, and N. Roy. Toward understanding natural language directions. In *Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 259–266. ACM, 2010.

[5] Ó. Martínez Mozos, R. Triebel, P. Jensfelt, A. Rottmann, and W. Burgard. Supervised semantic labeling of places using information extracted from sensor data. *Robot. Auton. Syst.*, 55(5):391–402, 2007. ISSN 0921-8890.

[6] C. Nieto-Granda, J. G. Rogers, A. J. B. Trevor, and H. I. Christensen. Semantic map partitioning in indoor environments using regional analysis. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 1451–1456. IEEE, 2010.

[7] J. G. Rogers III, A. J. B. Trevor, C. Nieto-Granda, and H.I. Christensen. Simultaneous localization and mapping with learned object recognition and semantic data association. In *Submitted to IEEE Conference on Intelligent Robots and Systems*, 2011.

[8] R.B. Rusu, N. Blodow, Z.C. Marton, and M. Beetz. Close-range Scene Segmentation and Reconstruction of 3D Point Cloud Maps for Mobile Manipulation in Human Environments. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), St. Louis, MO, USA*, 2009.

[9] A. J. B. Trevor, J. G. Rogers III, C. Nieto-Granda, and H.I. Christensen. Applying domain knowledge to SLAM using virtual measurements. *International Conference on Robotics and Automation*, 2010.

[10] A. J. B. Trevor, J. G. Rogers III, C. Nieto-Granda, and H.I. Christensen. Tables, counters, and shelves: Semantic mapping of surfaces in 3d. In *IROS Workshop on Semantic Mapping and Autonomous Knowledge Acquisition*, 2010.