

GLFP: Global Localization from a Floor Plan

Xipeng Wang

Ryan J. Marcotte

Edwin Olson

Abstract—In this paper, we describe a method for global localization in a previously unvisited environment using only a schematic floor plan as a prior map. The floor plan need not be a precision map – it can be the sort of image found in buildings to guide people or aid evacuation. The core idea is to identify features that are stable across both a drawn floor plan and robot point-of-view LIDAR data, for example wall intersections, which appear as corners from overhead and as vertical lines from the ground. We introduce a factor graph-based global localization method that uses these features as landmarks. The detections of such descriptorless features are noisy and often ambiguous. We therefore propose robust data association based on a pairwise measurement consistency check and max-mixtures error model. We evaluate the resulting system in a real-world indoor environment, demonstrating performance comparable to a baseline system that uses a conventional LIDAR-based prior map.

I. INTRODUCTION

Global localization (i.e. localization in a globally consistent coordinate frame) is critical for planning the paths of autonomous mobile robots, coordinating the actions of multiple agents or creating coherent user interfaces for operators. GPS could supply this positioning information, but GPS does not work well indoors. A robot could also localize using a prior map built with its own sensors, but this requires a previous visit to the site. Such a requirement may be unacceptable in many applications (e.g. search-and-rescue) in which a robot must operate in an environment it is visiting for the first time. How can a robot perform global localization in a new environment where GPS is unavailable?

One solution is to leverage existing global map resources from outside of the robot system. For example, a robot can match visual features in its camera view to aerial or satellite imagery to localize globally in an outdoor environment [1], [2]. Satellite imagery is not available for indoor environments, but we can use an analogous resource: floor plans.

In this paper, we propose and demonstrate GLFP (Global Localization in a Floor Plan), which allows a ground robot to perform global positioning by recognizing landmarks in the floor plan map. In particular, we use a LIDAR sensor on the robot to detect large vertical features (e.g. intersections of walls) that are also easy to detect as corners in the floor plan map.

GLFP can serve as a replacement for GPS in indoor environments, providing global position updates based on a floor plan map. In between such global fixes, the robot relies on local odometry measurements. In our evaluated

The authors are with the Computer Science and Engineering Department, University of Michigan, Ann Arbor, MI 48104, USA. {xipengw, ryanjmar, ebolson}@umich.edu.

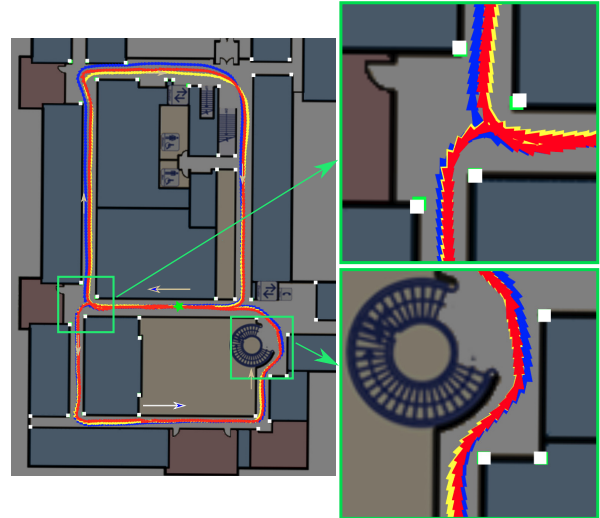


Fig. 1: Comparison of dead-reckoning (blue), laser scan match-based localization [3] (yellow), and our GLFP method (red) over a 180 m looped path in an indoor environment. Unlike the scan matching method, GLFP uses only a floor plan with labeled corners (white squares) as a prior map, then matches vertical edge features (e.g. wall intersections) observed from the robot's point-of-view to those landmarks. GLFP runs online for reasonable indoor speeds: here, the robot moved at about 2 m/s.

system, this odometry is based on wheel encoder and inertial sensors; tracking algorithms such as visual odometry or lidar odometry could also be applied to improve localization between global corrections. Fig. 3 further illustrates the role of GLFP in a robot's localization system.

The primary challenge in our approach is data association. The features are descriptorless and the robot's only prior knowledge is the floor plan. The environment may be dynamic and obstacles may obscure the robot's view of landmarks. Furthermore, the locations of landmarks in the floor plan may be imprecise; this might happen, for example, when a low-resolution map image is scaled. The contributions of this paper target these challenges.

These contributions are:

- A system that provides global localization in a floor plan map for a ground robot equipped with a LIDAR sensor,
- A factor graph-based localization approach that increases accuracy by modeling uncertainty in the location of labeled landmarks,
- A data association method that increases robustness by using the pairwise measurement consistency checks and max-mixtures error model, and
- Evaluations on real-world and synthetic data demon-

strating robust data association and accuracy comparable to laser scanmatching-based localization with a LIDAR prior map.

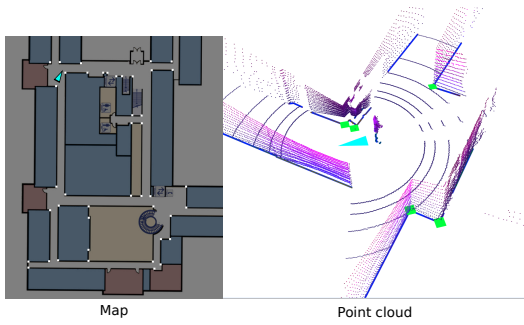


Fig. 2: GLFP performs global localization in a previously unvisited environment. It uses a floor plan (left) as a prior map with labeled corners (white squares) as landmarks. The robot extracts vertical features from a 3D point cloud (right) and matches those features (green squares) to the landmarks in the prior map. It adds the robot's discrete poses and the landmark positions to a factor graph to be optimized.

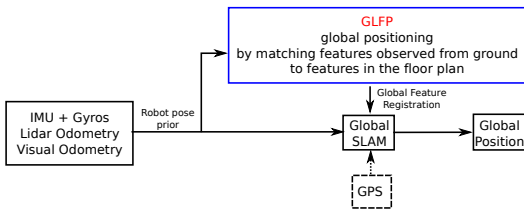


Fig. 3: Example robot localization system architecture showing the role of GLFP as a replacement or augmentation for GPS.

II. RELATED WORK

Our proposed method, GLFP, performs global localization by matching LIDAR data to landmarks in a floor plan map. We split our discussion of related work on global localization into two parts, first focusing on LIDAR-based approaches and then examining methods that use orthographic imagery.

A. Global Localization with LIDAR-based Sensing

Mobile robots typically produce detailed prior maps offline and then perform global localization by matching observed features with landmarks in the map. For example, Hentschel, Wulf, and Wagner [4] applied Monte Carlo Localization to a line feature-based reference map. Levinson, Montemerlo, and Thrun [5] and Levinson and Thrun [6] localize an autonomous vehicle in a pre-built, high-quality road surface map based on the reflectivity of LIDAR. Goeddel, Kershaw, Serafin, *et al.* [3] construct a globally consistent, pose graph-based prior map of the environment and then match 2D laser scans to known locations in the prior map to yield factor potentials in a localization pose graph.

These approaches rely on a prior map generated with the same sensors the robot uses for localization. Typically, this requires the robot to visit and map the environment before operating in it. In contrast, GLFP performs global

localization in a previously unvisited environment, using a floor plan image as a prior map and sensing with LIDAR.

A few existing methods (e.g. [7]–[9]) use aerial images as prior information in graph-based SLAM. These methods seek to use the prior information to make the generated map more globally consistent. In contrast, our paper uses the prior information (i.e. the floor plan map) to perform global localization in real time.

The closest of these methods to our own algorithm is that of Kümmerle, Steder, Dornhege, *et al.* [7]. Their method inserts correspondences found between three-dimensional range data and edges detected in aerial images as constraints into a graph-based formulation of the SLAM problem. The matching algorithms they use do not work well in corridors or other less structured environments because the edge features are aliased from the robot's local perspective. In contrast, our method extracts corner features from the overhead image and then uses several techniques to handle data association challenges.

B. Global Localization in Orthographic Imagery

Other papers (e.g. [2], [10]–[12]) have also considered the problem of localizing without a detailed prior map. Instead, they use an existing map resource such as imagery collected by satellite or UAV. Typically, these methods involve comparing ground-level images to a database of georeferenced overhead images.

Viswanathan, Pires, and Huber [10] proposed an algorithm that warps images from a ground robot's omnidirectional camera to project them into a top-down view. These projected images are then matched to a grid of satellite locations using hand-crafted interest point-based features. Kim and Walter [2] used two Convolutional Neural Networks (CNN) to transform ground-level and aerial images into a common feature space. To localize a query ground image, they find the closest georeferenced aerial image in that feature space.

These methods based on visual feature matching tend to fail when faced with significant viewpoint or appearance variations (e.g. matching a query image taken at night to a database image taken during the day) and seasonal changes (e.g. matching a query image with snow to one taken during summer). In addition, visual feature matching will not work in a floor plan map as interest point features (e.g. SIFT, SURF, or those learned by a CNN) tend to be the same for different corners because the pixel values around a corner in a floor plan map are artificial.

Recognizing these drawbacks of visual feature matching, we previously proposed FLAG, which directly matches 3D features (edges) in a stereo view to 2D landmarks (corners) in a satellite image [1]. As most environments have at least some 3D features which are stable over time, this approach is robust to visual appearance changes. However, data association without visual feature signatures is still a challenge, particularly in the presence of dynamic or static obstacles near a landmark. FLAG [1] uses a particle filter to explicitly track multiple hypotheses. In this paper, we present a new system that can achieve more robust data

association by applying two data association techniques: the pairwise consistency check and the max-mixtures error model. The max-mixtures model can directly capture multiple hypotheses that arise from uncertain data associations in the graph-based localization formulation. In addition, using graph-based formulation allows us to model the uncertainty in the location of labeled landmarks. In contrast to pure landmark-based localization, our method refines the positions of landmarks as the robot moves. We show that modeling the uncertainty in labeled landmark locations improves global localization accuracy.

III. APPROACH

A key insight with GLFP is to identify features which are co-observable from a low-quality overhead map and from a robot's sensor view. Large vertical edges emerged as a highly-salient, easily-extractable feature that occurs in most indoor environments with enough frequency to be useful but not so much as to confound data association. GLFP extracts these edges from a 3D point cloud and matches them to the known corner features in the floor plan map. It then solves for the robot pose using a factor graph-based localization algorithm.

A. Vertical Edges as Global Map Features

A floor plan is an illustrated, overhead view of the environment. Because it is artificial, we cannot extract interest point-based features (e.g. SIFT or SURF) from it that are also visible from the robot's point-of-view. Instead, we use long vertical edges, which occur regularly indoors at wall intersections or other similar permanent structures, as global features. These features have a characteristic appearance from overhead (i.e. corners) and from the ground (i.e. tall vertical lines). By using these features, a robot can localize in a scene it has not previously observed.

In this paper, we manually label feature locations in the floor plan, though automated extraction would also be feasible. The labeling need not be highly precise: our approach refines the position estimates of the landmarks as the robot navigates.

B. Detecting Features from the Ground

From the ground, our features appear as large vertical lines. Though the same features may be detectable as corners in a 2D point cloud, such an approach would fail to account for the height of the features. An indoor environment may contain many objects (e.g. furniture) that would appear as corners in planar sensor data but are not part of the permanent structure of the environment. Such transient features do not appear in a floor plan and therefore could not be used by GLFP for localization in a previously unvisited environment.

In this paper, we detect vertical edge features using a 3D LIDAR, though one might also use, for example, a stereo vision sensor [1]. We first project the 3D point cloud into 2.5D (i.e. 2D with height information), then detect corners in that data [13]. We discard any corners less than 1 meter in height as these likely do not correspond to permanent structure visible in the floor plan.

C. Factor Graph-Based Localization

The positions of the landmarks labeled in the floor plan may be imprecise, whether because of sloppy annotation or lack of resolution in the floor plan image. We therefore treat both the landmark positions and discrete robot poses as variables to be optimized in a factor graph (see Fig. 4). In this way, the initial landmark position estimates come from the map annotation, and subsequent measurements refine these estimates.

Fig. 4 illustrates the resulting factor graph representation. We treat the label as an observation of the landmark's location and create a factor potential to represent this measurement. We add factor potentials between robot poses based on odometry measurements. In our evaluation, these come from wheel encoders and inertial sensors, though one could also use tracking-based methods such as visual odometry. Finally, the robot measures ranges to landmarks, which are incorporated as factor potentials between landmarks and robot poses. The factor graph is optimized using AprilSAM [14].

Because the features we use are descriptorless, data association is a primary challenge, especially given that indoor environments may be cluttered with objects that trigger false positive detections. For example, consider the case of a trash can placed near a hallway corner in an office building. The object would not appear in the floor plan, but it may obscure a nearby feature or be mistaken for it. In addition to these types of static obstacles, our system must handle dynamic obstacles, such as a person walking. Filtering out small vertical edges helps with these problems, but we still need a way to improve the robustness of data association.

In contrast to methods which make data association an explicit process prior to inference (e.g. JCBB [15] or IPJC [16]), we move the bulk of data association into the inference process itself using the max-mixtures method [17]. These mixtures are bootstrapped with nearest-neighbor associations.

1) *Pairwise Consistency Check*: To reject erroneous data associations caused by dynamic obstacles, we apply a pairwise consistency check (see Fig. 5). The robot takes multiple measurements of a landmark over the time interval during which the landmark is within view. When combined with the estimate of the robot's pose when the measurement was taken, each observation gives an estimate of the projected position of the landmark. For every pair of these measurements, we compute the distance between the corresponding projected positions of the landmark. If those positions are within a threshold distance, we consider the pair of measurements to be consistent.

We then form an undirected graph in which the measurements are nodes, and we add edges between pairs of nodes corresponding to consistent measurements. Olson, Walter, Leonard, *et al.* [18] presented an approximate solution to find a single cluster of well-connected nodes in a graph. In this paper, we apply a simple variation on this idea. Namely, if there exists a measurement that has enough consistent links to other measurements, we create a factor potential for the measurement and add it into the factor graph. In

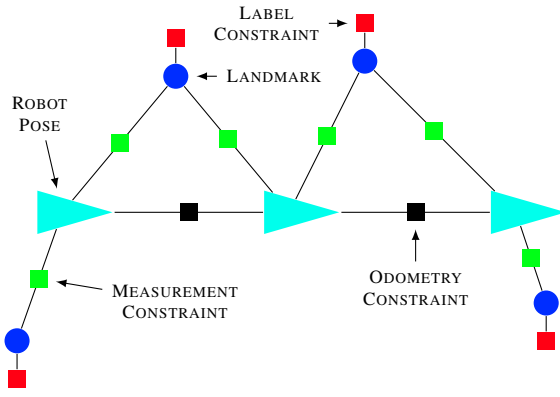


Fig. 4: Illustration of GLFP factor graph. We discretize the robot's trajectory into a series of poses (cyan triangles) that we add to the graph as nodes. We also add the positions of landmarks (blue circles) as graph nodes. Factors in the graph include odometry constraints (black squares), labeled position constraints of landmarks (red squares), and measurement constraints between pose nodes and landmark nodes (green squares). With this formulation, both the robot poses and landmark positions are updated when the graph is optimized.

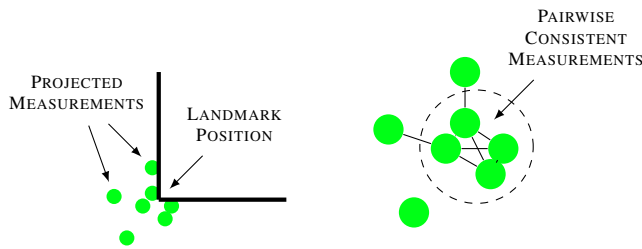


Fig. 5: Illustration of the pairwise consistency check. The green circles (*left*) are the projected positions of landmarks based on distance and bearing measurements. These measurements form an undirected graph (*right*). Edges exist between nodes if the projected landmark positions from the two measurements are within a threshold distance. A clique in the graph represents a set of internally consistent measurements. We reject measurements that are not part of a sufficiently large clique.

this way, a moving obstacle will not yield enough consistent measurements to be added into the graph.

2) *Max-Mixture Error Model*: To reject erroneous data associations caused by static objects near landmarks, we apply a max-mixtures error model [17]. We used two-component mixtures with one being the nominal data association and the second being a "null" hypothesis. With the max-mixtures, the localization system can visit all data association decisions to reject outliers as the robot makes new observations.

Consider again the case of a trash can placed near a hallway corner. As the robot approaches the trash can, associating the object to the nearby feature might reduce error compared to an odometry-only solution. However, as the robot continues to observe the object, the repeated low-likelihood measurements resulting from the association cause error to increase. A max-mixtures error model can revisit those previous data associations and apply a null hypothesis to them.

IV. EXPERIMENTAL EVALUATION

We evaluate GLFP in two parts: (1) we test the entire system's performance using real-world data, and (2) we analyze the effect of GLFP's robust data association and factor graph formulation using simulation.

A. Robot Platform

We performed all experiments with the MAGIC2 robot platform developed by the APRIL Robotics Lab. Relevant sensors onboard the robot include wheel encoders and a fiber optic gyroscope (KVH DSP-1715) for robot odometry and a 3D LIDAR (Velodyne VLP-16).

B. Global Localization

We evaluated our method on real-world indoor datasets to demonstrate its overall performance and to compare it to a state-of-the-art laser scan match-based localization method that uses a LIDAR prior map [3], [19].

Our floor plan had a resolution of 0.1 m per pixel. We hand-labeled the initial positions of landmarks at every corner corresponding to an intersection of walls (see white squares in Fig. 1). We assigned an uncertainty of 10 cm to the factor potentials corresponding to these labeled positions. The odometry measurements came from the fusion of wheel encoders and a fiber optic gyroscope; we assigned x, y, θ uncertainty of 10 cm, 5 cm, and 0.1 degrees per meter traveled. We assigned an uncertainty of 5 cm to the ranging measurements between landmarks and robot poses.

We manually drove the robot at a speed of about 2 m/s in a 180 m loop through the halls of an office building (the BBB at the University of Michigan). We followed a figure-8 pattern; the green star in Fig. 1 shows the starting point and the tan arrows indicate the path of travel.

Fig. 1 shows the path of the robot resulting from odometry-based dead-reckoning (blue), a laser scan match-based localization method using a LIDAR prior map [3] (yellow), and our proposed GLFP method (red). GLFP shows clear improvements over dead-reckoning, and its performance is similar to that of laser scan matching approach, which we consider an approximate ground-truth. Recall that the laser scan matching approach uses a LIDAR prior map generated offline, whereas GLFP uses only a floor plan map.

Because GLFP adds the positions of landmarks to the factor graph, it continually updates their positions as the robot travels around and makes observations. The white squares in Fig. 1 are the initial hand-labeled initial positions, and the green squares are the final positions after the whole run.

We evaluate GLFP on four different runs over a 180 m looped path in an indoor environment for 4 different trips. We use a state-of-the-art laser scan match-based method (with a LIDAR prior map) as an approximate ground-truth. Fig. 6 shows the average absolute errors of GLFP and dead-reckoning. GLFP has clear improvements over dead-reckoning.

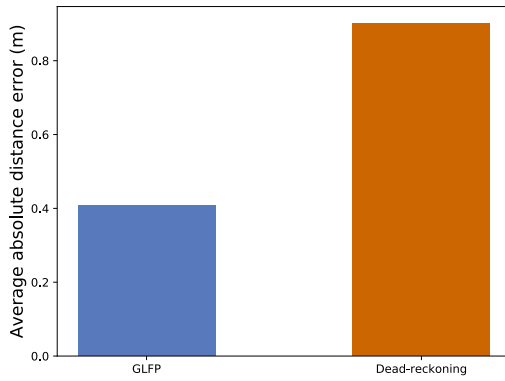


Fig. 6: Average absolute distance errors of GLFP and dead-reckoning over a 180 m looped path in an indoor environment for 4 different runs (Fig. 1 shows one of four runs). We use a state-of-the-art laser scan match-based method (with a LIDAR prior map) as an approximate ground-truth.

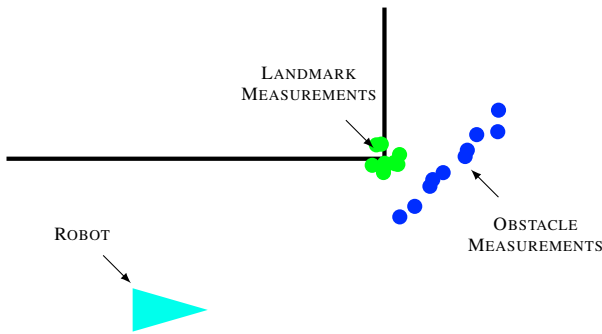


Fig. 7: Illustration of simulated test of the pairwise consistency check algorithm. Green dots show the 10 measurements of the landmark, which are sampled from its true location with variance 10 cm. The blue dots show observations of a dynamic object moving along a straight line, from which measurements are sampled with variance 1 cm. The results are shown in Fig. 8.

C. Data Association

To evaluate the performance and reliability of our data association system, we performed an ablation study to show the effects of the pairwise consistency test and max-mixture error model, and to measure the effectiveness of refining the floor plan landmark positions.

1) *Pairwise Consistency Check*: The primary purpose of the pairwise consistency check in GLFP is to handle dynamic obstacles in the robot's environment. To test the effect of this component, we constructed a simulated scenario in which a moving object passes in a straight line by an existing landmark (see Fig. 7). We randomly sample 10 measurements from the landmark and 10 measurements from the moving object, both with variance 10 cm. In this experiment, the distance threshold to determine whether measurements are consistent is 10 cm, and the data association threshold is set to be 50 cm. We run 5000 trials for data association with and without pairwise consistency check, incorporating all observations within 50 cm. Fig. 8 shows the result of this test, namely that the pairwise consistency check results in lower error and lower error variance.

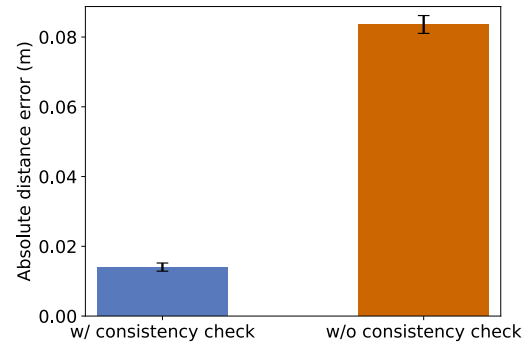


Fig. 8: Localization error in the presence of dynamic obstacles (see Fig. 7). Error bars represent three standard error of the mean across the 5000 trials. Using the pairwise consistency check resulted in lower error.

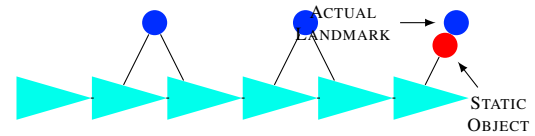


Fig. 9: Illustration of a simulated scenario testing the max-mixture error model. The rightmost landmark is obstructed by a static object, which the robot observes instead of the landmark. For each trial, we randomly sample the location of the static obstacle within a particular radius of the landmark. We plot the localization error of the last pose in Fig. 10.

2) *Max-mixture Error Model*: The primary purpose of GLFP's max-mixture error model is to avoid associating static obstacles with landmarks from the floor plan map. Analogous to the simulation of the previous section, we built a simulated test case in which a robot passes by several landmarks, the last of which is blocked by a static obstacle (see Fig. 9). We randomly sample the location of this static obstacle within a specified radius of the landmark. As in the previous section, we randomly sample measurements with standard deviation 10 cm.

The robot does not observe the final landmark but rather receives measurements from the nearby static obstacle. We vary the maximum distance of the object from the landmark in the range 10-50 cm and conduct 5000 trials at each radius. We measured the localization of the last node in each trial with and without the max-mixture error model. Fig. 10 shows the result of this experiment.

Without the max-mixture error model, error increases proportionally to the distance between the obstacle and the landmark. With the max-mixture model, error increases until the distance reaches approximately 20 cm, after which it levels off. Once measurements from the mis-associated obstacle cause error to rise to a certain point, the max-mixture model "turns off" measurements of that landmark by assigning a highly uncertain noise model to them. This mitigates the effect of the erroneous data association.

3) *Incorporating Landmark Uncertainties*: One source of localization error in our previous approach [1] is not incorporating uncertainties of labeled landmark positions. In

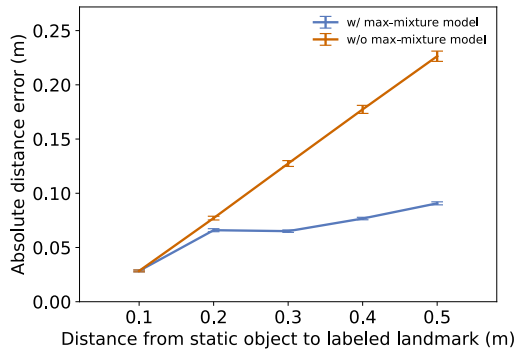


Fig. 10: Localization error in the presence of a static obstacle. Error bars represent three standard error of the mean across the 5000 trials. As the static object gets further from the true landmark localization, the effect of using the max-mixture model becomes more pronounced.

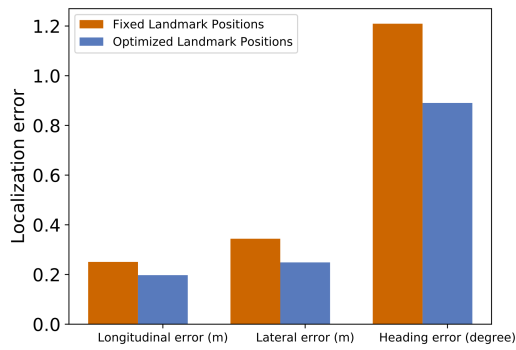


Fig. 11: Effect of modeling uncertainty in labeled landmark positions on localization performance. Including the landmark positions reduces the longitudinal, lateral, and heading errors.

this paper, we model those uncertainties by formulating the localization problem as a factor graph-based SLAM problem that optimizes both robot positions and landmark positions. Fig. 11 shows the effect of this optimization on the average localization error measured on the real-world dataset of Sec. IV-B. We compare the performance of GLFP to that of an otherwise identical method that considers the locations of the landmarks to be static. Incorporating the uncertainty in landmark positions leads to smaller longitudinal, lateral, and heading errors.

V. CONCLUSION

In this paper, we presented GLFP, a method that enables a ground robot to localize itself globally by identifying landmarks visible in a floor plan map. GLFP is a factor graph-based approach that uses a pairwise measurement consistency check and max-mixtures error model. This results in more robust and accurate localization than our previous work [1]. We demonstrated that our system has comparable performance to laser scan matching-based localization with a LIDAR prior map, showing GLFP's feasibility for real-time global localization without GPS in a previously unvisited environment.

Acknowledgments

This material is based upon work partially supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1256260.

REFERENCES

- [1] X. Wang, S. Vozar, and E. Olson, "FLAG: Feature-based Localization between Air and Ground," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2017.
- [2] D.-K. Kim and M. R. Walter, "Satellite image-based localization via learned embeddings," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2017.
- [3] R. Goedel, C. Kershaw, J. Serafin, and E. Olson, "FLAT2D: Fast localization from approximate transformation into 2D," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016.
- [4] M. Hentschel, O. Wulf, and B. Wagner, "A GPS and laser-based localization for urban and non-urban outdoor environments," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2008.
- [5] J. Levinson, M. Montemerlo, and S. Thrun, "Map-based precision vehicle localization in urban environments," in *Proceedings of the Robotics: Science & Systems Conference*, 2007.
- [6] J. Levinson and S. Thrun, "Robust vehicle localization in urban environments using probabilistic maps," in *Proceedings of the IEEE International Conference on Robotics and Automation*, IEEE, 2010.
- [7] R. Kümmerle, B. Steder, C. Dornhege, A. Kleiner, G. Grisetti, and W. Burgard, "Large scale graph-based SLAM using aerial images as prior information," 2011.
- [8] M. P. Parsley and S. J. Julier, "Towards the exploitation of prior information in SLAM," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2010.
- [9] —, "Exploiting prior information in GraphSLAM," in *Proceedings of the IEEE International Conference on Robotics and Automation*, IEEE, 2011.
- [10] A. Viswanathan, B. R. Pires, and D. Huber, "Vision based robot localization by ground to satellite matching in GPS-denied situations," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014.
- [11] T.-Y. Lin, Y. Cui, S. Belongie, and J. Hays, "Learning deep representations for ground-to-aerial geolocalization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [12] M. Bansal, H. S. Sawhney, H. Cheng, and K. Daniilidis, "Geolocalization of street views with aerial image databases," in *Proceedings of the 19th ACM International Conference on Multimedia*, ACM, 2011.
- [13] E. B. Olson, "Robust and efficient robotic mapping," 2008.
- [14] X. Wang, R. Marcotte, G. Ferrer, and E. Olson, "AprilSAM: Real-time smoothing and mapping," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2018.
- [15] J. Neira and J. D. Tardós, "Data association in stochastic mapping using the joint compatibility test," *IEEE Transactions on Robotics and Automation*, 2001.
- [16] E. Olson and Y. Li, "IPJC: The incremental posterior joint compatibility test for fast feature cloud matching," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012.
- [17] E. Olson and P. Agarwal, "Inference on networks of mixtures for robust robot mapping," *International Journal of Robotics Research*, 2013.
- [18] E. Olson, M. Walter, J. Leonard, and S. Teller, "Single cluster graph partitioning for robotics applications," in *Proceedings of the Robotics: Science & Systems Conference*, 2005.
- [19] E. Olson, "M3RSM: Many-to-Many Multi-Resolution Scan Matching," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2015.