3D Place Recognition and Object Detection using a Small-sized Quadrotor

Slawomir Grzonka Bastian Steder Wolfram Burgard
University of Freiburg, Department of Computer Science, 79110 Freiburg, Germany

***Abstract*— We present a system for 3D place recognition and object detection using a small-sized quadrotor. The robot is equipped with a horizontally scanning 2D range-scanner and occasionally acquires 3D scans by hovering on the spot and changing its altitude. Our approach is able to accurately and robustly recognize previously seen places of the environment. Additionally, our system can be applied to match the current observations to models stored in a database which allows the robot to perform object detection. We evaluate our approach in real-world experiments to demonstrate the robustness and reliability of our algorithms.**

I. Introduction

Place recognition, meaning the detection that a robot revisited an already known area, is a crucial part in key navigation tasks including localization and SLAM. The majority of state-of-the-art place recognition techniques have been developed for vision- or two dimensional range data. Relatively few approaches work on three-dimensional laser range scans and can efficiently calculate the similarity or the relative transformation between two scans. Even more, most place recognition and object detection algorithms have been specially suited for ground robots. However, 3D scans gathered with aerial vehicles equipped with a 2D laser scanner typically have a substantial higher noise in the measurements than in the case of ground robots. The main reason for this is that the pose and orientation of the flying robot can be affected by high variations during the measurement acquisition.

We present a robust place recognition system operating on 3D range data which can be used with data gathered by a wheeled as well as a flying robot. The flying quadrotor robot is equipped with a horizontally scanning 2D laser range scanner, an IMU, and a laser mirror which is used to deflect some of the laser beams along the z direction, i.e., providing measurements about the robot's altitude. Our navigation system estimates the current pose of the quadrotor by projecting the current measurement on a 2D plane (using measurements about roll and pitch from the IMU). This allows us to use efficient 2D scan-matching algorithms for pose estimation. Together with the estimate of the global height and the robot's attitude (IMU), we then project the 2D measurements into 3D.

The quadrotor records a 3D scene by hovering around the spot while changing it's altitude. Our approach transforms this 3D range scan into a range image and uses matches between point-features to estimate relative poses, that are then individually scored. The same principle can be used for object recognition. Here, we first acquire high density object models using a wheeled robot equipped with laser scanner mounted on a pan-tilt unit. We then match the features of the current 3D scan acquired by the flying robot with features stored in a database (representing the objects). Figure 1 shows an example of our system used for object detection. Again, the quadrotor acquired a 3D scan by hovering around the spot and changing it's altitude. The image shows the accumulated 3D scan together with the outcome of our object detection approach. In this example, we search in the scene for the object "chair". The two chairs present in the scene are overlayed by the objects detected by our approach. Since we estimate the location of the corresponding objects as well, this allow us to add the high density models from the database to the current 3D scan.

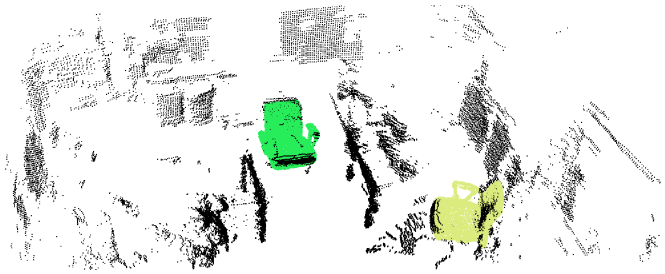

Fig. 1. Object detection using 3D data acquired by our quadrotor. The robot acquired a 3D scan by hovering around the spot while changing its altitude. Two chairs were found and their estimated 6DoF locations are visualized.

II. Related Work

In the past, the problem of place recognition has been addressed by several researchers and a wide variety of approaches for different types of sensors have been developed. Cameras are often the first choice. Compared to 3D data, vision features are typically very descriptive and unique. On the other hand, spacial verification is naturally easier in 3D data. One very successful approach using vision is the Feature Appearance Based MAPping algorithm (FABMAP) proposed by Cummins and Newman [4]. This algorithm uses a bag-of-words approach based on SURFs [2] extracted from omni-directional camera images and was shown to work reliably even on extremely large-scale datasets. We would like to refer the reader to this paper for a detailed discussion of vision-based place recognition approaches.

Laser scanners, either 2D or 3D have been also employed for object and place recognition purposes [9, 3, 14, 6, 5, 8], but we would like to refer the reader to [11, 12] for a detailed discussion about related work wrt. laser scanners.

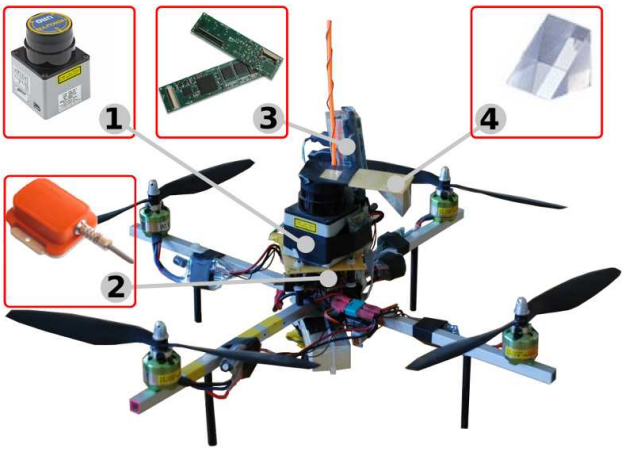

Fig. 2. Our quadrotor robot used in this experiments. The platform is based on a Mikrokopter [1] and is equipped with a Hokuyo URG (1), an XSens MTi IMU (2), a Gumstix embedded computer (3), and a laser mirror (4).

III. Technical Section

We use the quadrotor platform shown in Figure 2 to acquire 3D scans and use them to recognize previously seen parts of the environment. In this section we discuss how the scans are obtained and present our algorithm for place and object recognition.

A. Acquiring a 3D Scan of the Environment

The quadrotor acquires a 3D scan by hovering around a desired spot and changing it's altitude. To obtain a full 360° scan, we turn by 180° and repeat the process. The choice of the place where to acquire a scan can be triggered manually (e.g., by pressing a button), or by incorporating an exploration behavior into the robots navigation system. However, our focus is on how to extract 3D data from the robot and on the algorithms for place/object recognition. We therefore acquired all measurements by manual flights.

The robots navigation system estimates a full 3D pose (i.e., $(x, y, z, \phi, \theta, \psi)$) by means of 2D laser scan matching. In more detail, we use the current measurement about roll (ϕ)) and pitch (θ) from the IMU to project the current laser measurement onto a 2D plane. We then employ hierarchical correlative scan matching similar to the one proposed by Olson *et al.*[10] to estimate the incremental 2D transformation ((x, y, ψ)). To get an accurate estimate of the robot's pose, we employ a grid resolution of $0.01\,\mathrm{m} \times 0.01\,\mathrm{m}$ at the finest resolution of the grid map. The laser beams deflected by the laser mirror are used to estimate the current global altitude (even in the presence of obstacles underneath, see [7] for more details). Together with the IMU measurements of roll and pitch we now obtain an estimate of the full 3D pose of the robot. A 3D measurement consists of all measurements taken while the robot hovered around a spot (i.e., within a region). Additionally, we require a substantial variety in the altitude during the scan (i.e., having at least a variation of 1 m). An example of such a scan is shown in Figure 3. The left image depicts the 2D map build from the measurements. This map was used to determine the pose of the robot. The right image shows the corresponding measurements projected into 3D, given the 3D pose estimate of our navigation system. In the remainder of this section, we will call such a set of measurements acquired while hovering around a spot a 3D scan of the environment. During the quadrotor's mission, the robot continues to acquire such 3D scans. These are then stored in a database. However, the database can also contain high density models of single objects like chairs recorded with a different platform not necessarily a flying one.

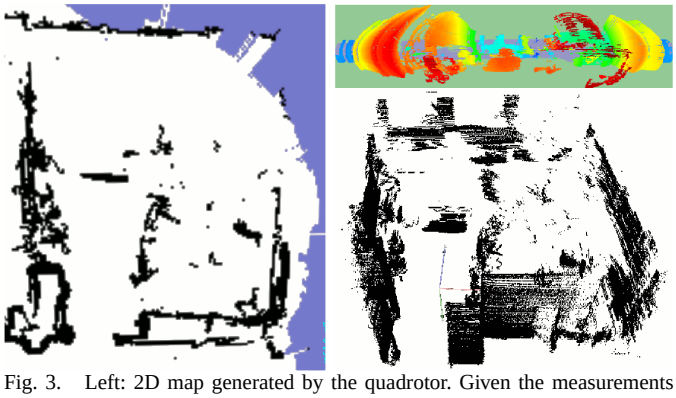

Fig. 3. Left: 2D map generated by the quadrotor. Given the measurements about roll, pitch and the estimated altitude we then project the measurements into 3D to create the 3D point cloud shown in the bottom right image. The top right image shows the corresponding range image.

B. Place Recognition

Given a database of 3D scans and a scan as input query, our algorithm returns a set of scans which are potential matches with the input. The database consists of measurements of the environment previously recorded by the flying robot. These measurements could have been recorded during the same mission or in a previous one. Additionally, our approach calculates for every returned scan pair (i.e., query scan and matched scan from the database) a transformation and a score reflecting how certain the system is, that the two scans actually match.

More formally, let D denote the database of 3D range measurements and z^* a query scan. The goal of our approach is to calculate a set of candidate pairs, $C(z^*) = (\langle z_1, T_1, s_1 \rangle, \ldots, \langle z_n, T_n, s_n \rangle)$. Here, $z_i \in D, i \in \{1, \ldots, n\}, n = |D|$, are the potential measurement candidates from the database which are similar to the current query z^*. Whereas T_i denotes the estimated transformation from z^* to z_i, s_i is a score reflecting the confidence about the match. Our algorithm for calculating $C(z^*)$ mainly consists of the following steps.

1) Generate a list of possible scan-pairs. This list could be obtained using for example a bag-of-words (BoW) approach for a fast pre-selection [12]. However, in our case the database D of previously acquired scans is typically small. In such scenarios, using BoW is computationally more intensive than checking all possible candidates. We therefore calculate a list of all pairs, $\langle z^*, \hat{z}_k \rangle, \hat{z}_k \in \hat{D}(z^*)$.
2) For each pair $\langle z^*, \hat{z}_k \rangle, \hat{z}_k \in D, k = 1, \ldots, |D|$, calculate a set of possible transformations between z^* and $\hat{z}_k$ by matching point features of the corresponding scans.

3) Score each of the possible transformations and get the transformation T_k with the highest score s_k. If this score is above an acceptance threshold then $\langle \hat{z}_k, T_k, s_k \rangle$ is a candidate for a recognized place, i.e., it is added to $C(z^*)$.

Although we work with a database of 3D range scans, we do not use this data directly. We rather represent each three-dimensional range scan by its dual, namely a range image (see Figure 3 (top right)). If the 3D scan is captured from one point in space, i.e., the sensor does not move while the 3D points are generated, the range image contains the same information as the scan. Although this assumption is violated to some degree when using flying vehicles, we will still use the range image, as they allow us to model unknown areas as well as maximum range readings more efficiently.

We will now briefly describe the individual components of our approach. More details about specific parts can be found in [12].

C. Feature Extraction

To calculate a similarity between two scans, we first calculate a set of features representing the scan. Our approach uses the so-called NARFs (Normal-Aligned Radial Features) [13] recently developed for robust object recognition based on 3D scans. These point features are used to find corresponding regions between two 3D measurements. The descriptors of the features can be compared using standard norms like the Manhattan distance. The resulting measure (the *descriptor distance*) describes the similarity between the described regions. Here, a high value reflects a low similarity.

D. Determining Candidate Transformations

Each NARF encodes a full 3D transformation. Therefore, the knowledge about a single feature correspondence between two scans enables us to retrieve all six degrees of freedom of the relative transformation between them (i.e., by calculating the difference between the two poses). To obtain the candidate transformations, we order the feature pairs according to increasing descriptor distance and evaluate the transformations in this order. In our experiments we stop after a maximum number of 2000 evaluated transformations for computational reasons.

E. Scoring of Candidate Transformations

The result of the feature matching is a list of relative poses $\hat{T}_k = \{\hat{T}_{k_1}, \ldots, \hat{T}_{k_n}\}$ for the candidate pair $\langle z^*, \hat{z}_k \rangle, \hat{z}_k \in \hat{D}(z^*)$. In the next step, we evaluate those candidate transformations and calculate a score (likelihood) for each $\hat{T} \in \hat{T}_k$ reflecting the confidence of the transformation given a model of our sensor. Since we use 3D range data, i.e., each measurement z is a set of 3D points, we evaluate the candidate transformation $\hat{T}$ on a point-by-point basis (i.e., we assume the points are mutually independent).

F. Object Detection

Our approach for object detection is similar to the one of place recognition. Again, we have a database D containing models of specific objects. These models were previously acquired by a different robot by merging multiple scans of the object from different perspectives.

To detect an object in the current measurement we match NARF features against the object features from models in the database D similar to the case of place recognition. However, the main difference between our algorithms for object detection and place recognition comes from having a full 3D model of the corresponding object. Given a candidate transformation, i.e., the expected position and orientation of the object obtained from the matched NARF features, we can calculate an expected range image of the object. This allows us to compare this range image to the one obtained from the current scan pixel by pixel.

Again, based on our observation model, we calculate a score for each candidate (object and transformation). An object is detected in the environment, if the corresponding score is above a given threshold.

IV. Experiments

This section provides our experimental results. We will first show our results for place recognition. Subsequently, we will demonstrate our first results on object detection.

A. Place Recognition

In the following experiment we manually flew the quadrotor several times in an office environment. The whole set consists of 23 scans obtained in four distinct runs. The overall 3D map using our navigation system is shown in Figure 4. The individual places where the quadrotor recorded a full 3D scan are highlighted by the labels $1, \ldots, 23$. The blue part of the rectangles in Figure 4 are pointing forwards wrt. to the quadrotor. Figure 4 (top right) shows the ground truth confusion matrix of the individual scans. Here, dark areas reflect a high similarity between the corresponding scans. The confusion matrix estimated by our approach is shown in the bottom right together with a plot of the recall rate versus the distance of individual scans. For example, for all scan-pairs which were recorded at most 1 m from each other, our approach correctly recognized all loop closures. Subsequently, for all scan-pairs which were at most 2 m apart from each other, our approach still was able to correctly recognize about 73% of all potential loop closures. Note that the recall rate drops quickly starting from 3 m on. This originates from the fact that the laser range scanner has a maximum range of 5.6 m. In this case, scans which have been acquired 3 m apart from each other only have a very limited overlap.

B. Object Detection

To get some initial results for the object recognition procedure, we used the quadrotor platform to capture five 3D scans in an office environment, where several instances of a chair of which we obtained a full 3D model were present.

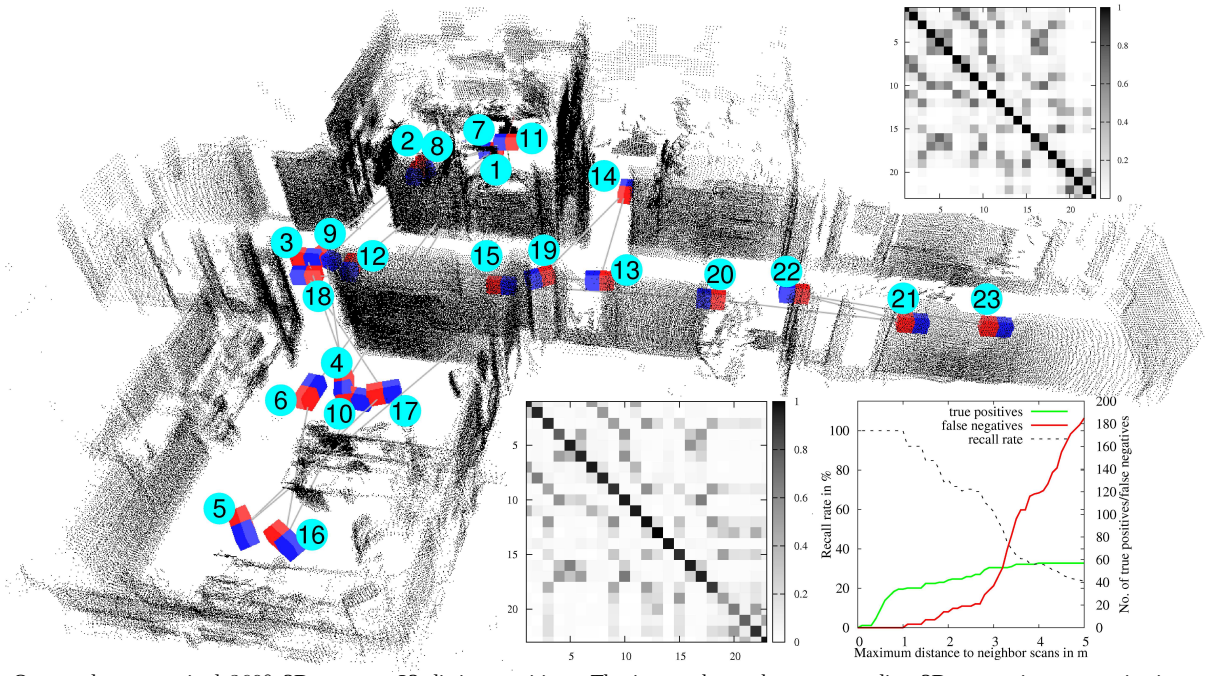

Fig. 4. Our quadrotor acquired 360° 3D scans at 23 distinct positions. The image shows the corresponding 3D map using our navigation system. The positions where the quadrotor acquired are labeled $1, \ldots, 23$. Note that Figure 3 shows the scan taken at position 14. Here, the blue part of the rectangle is headed towards the front wrt. to the quadrotor. The ground truth confusion matrix is shown in the top right, the one computed by our system is in the bottom right, together with a graph showing the recognition rate for different maximum distances between scans.

Using a minimum acceptance threshold that returned zero false positives, we were able to find about 73% of those chairs. See Figure 1 for an example. For lower acceptance thresholds there are a number of false positives where the system assumes that the chair was seen from the back. From this perspective it presents mostly a flat surface of a certain size, which is hard to distinguish from other flat structures in the environment. Our plan is to achieve higher recognition rates by using an active exploration strategy and tracking found objects until they were seen from more then one perspective to achieve higher recognition rates and remove false positives resulting from views that provide little distinctive structure.

V. Conclusions

We presented a novel approach for robust place recognition using data acquired from a flying vehicle. Additionally, we present a variation of our approach used for object detection. Our quadrotor acquires 3D scans of the environment by hovering around a spot while changing its altitude. Those scans are matched against a database containing previously acquired scans of the environment as well as models of different objects. Our system has been implemented and successfully tested. The experimental results demonstrate that our approach is able to robustly recognize previously seen parts of the environment. Our first results also imply, that our approach can reliably detect known objects in the environment. In future work, we aim to combine the object detection approach with an active exploration technique.

VI. Acknowledgments

This work has partly been supported by the German Research Foundation (DFG) under contract number SFB/TR-8.

References

[1] Mikrokopter, http://www.mikrokopter.de/.
[2] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In *Proc. of the Europ. Conf. on Comp. Vision (ECCV)*, 2006.
[3] M. Bosse and R. Zlot. Map matching and data association for large-scale two-dimensional laser scan-based slam. *International Journal of Robotics Research*, 27(6):667–691, 2008.
[4] M. Cummins and P. Newman. Appearance-only SLAM at large scale with FAB-MAP 2.0. *Int. Journal of Robotics Research*, Nov 2010.
[5] N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann. Robust global registration. In *Proceedings of the third Eurographics symposium on Geometry processing (SGP)*, page 197, Aire-la-Ville, Switzerland, Switzerland, 2005. Eurographics Association.
[6] K. Granström, J. Callmer, F. Ramos, and J. Nieto. Learning to detect loop closure from range data. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2009.
[7] S. Grzonka, G. Grisetti, and W. Burgard. Towards a navigation system for autonomous indoor flying. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2009.
[8] A.E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(5):433–449, 1999.
[9] Y. Li and E. Olson. A general purpose feature extractor for light detection and ranging data. *Sensors*, 10(11):10356–10375, 2010.
[10] E. Olson. Real-time correlative scan matching. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 4387–4393, Kobe, Japan, June 2009.
[11] B. Steder, G. Grisetti, M. Van Loock, and W. Burgard. Robust on-line model-based object detection from range images. In *Proc. of the Int. Conf. on Intelligent Robots and Systems (IROS)*, 2009.
[12] B. Steder, M. Ruhnke, S. Grzonka, and W. Burgard. Place recognition in 3D scans using a combination of bag of words and point feature based relative pose estimation. 2011. UNDER REVIEW.
[13] B. Steder, R. B. Rusu, K. Konolige, and W. Burgard. Point feature extraction on 3D range scans taking into account object boundaries. In *Proc. of the IEEE Int. Conf. on Rob. & Automation (ICRA)*, 2011.
[14] G. D. Tipaldi, M Braun, and K. O. Arras. FLIRT: Interest regions for 2D range data with applications to robot navigation. In *Proceedings of the International Symposium on Experimental Robotics (ISER)*, 2010.