The MIT Stata Center dataset

Citation for published version:
Fallon, M, Johannsson, H, Kaess, M & Leonard, JJ 2013, 'The MIT Stata Center dataset', *International Journal of Robotics Research*, vol. 32, no. 14, pp. 1695-1699. https://doi.org/10.1177/0278364913509035

Digital Object Identifier (DOI):
10.1177/0278364913509035

Published In:
International Journal of Robotics Research

The MIT Stata Center Dataset

Maurice Fallon, Hordur Johannsson, Michael Kaess and John J. Leonard

DRAFT: February 14, 2014

Abstract

This paper presents a large scale dataset of vision (stereo and RGB-D), laser and proprioceptive data collected over an extended duration by a Willow Garage PR2 robot in the 10 story MIT Stata Center. As of September 2012 the dataset comprises over 2.3TB, 38 hours and 42 kilometers (the length of a marathon). The dataset is of particular interest to robotics and computer vision researchers interested in long-term autonomy. It is expected to be useful in a variety of research areas - robotic mapping (long-term, visual, RGB-D or laser), change detection in indoor environments, human pattern analysis, long-term path planning. For ease of use the original ROS 'bag' log files are provided and also a derivative version combining human readable data and imagery in standard formats. Of particular importance, this dataset also includes ground-truth position estimates of the robot at every instance (to typical accuracy of 2cm) using as-built floor-plans - which were carefully extracted using our software tools. The provision of ground-truth for such a large dataset enables more meaningful comparison between algorithms than has previously been possible.

1 Overview

In this paper we present a vast scale multi-sensor dataset of interest to the robotics and computer vision research communities. The dataset was collected by a Willow Garage PR2 robot over an extended duration beginning in January 2011 and as of September 2012 comprises 2.3TB, 38 hours and 42 kilometers of exploration of the 10 story MIT Stata Center.
http://projects.csail.mit.edu/stata/

This dataset was collected as part of a project to develop a visually-driven real-time Simultaneous Localization and Mapping system which could navigate a robot within a large building and over multi-year timescales (Johannsson et al. (2013); Johannsson (2013)). As such that work provides an ideal demonstration of where we would find this dataset to be most useful.

The remainder of this paper is as follows: Section 2 describes the components of the robotic platform which are of most interest to the reader. Section 3 describes the typical mode of operation of the robot during exploration and describes a section of the dataset in detail. Section 4 describes our ground-truthing procedure while Section 5 describes tools for accessing and using the dataset.

Draft manuscript, February 14, 2014.

The authors are with the Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts 02139, USA {mfallon,hordurj,kaess,jleonard}@mit.edu.

Fig. 1: Image Courtesy of Willow Garage

2 Willow Garage PR2 Robot

The PR2 is a research and development robotic designed and built by Willow Garage. With two compliant arms the robot is typically used for research into robotic manipulation, however the on-board sensing suite also makes for an excellent source of sensor data for mapping and perception.

The following sensors were logged during the exper-

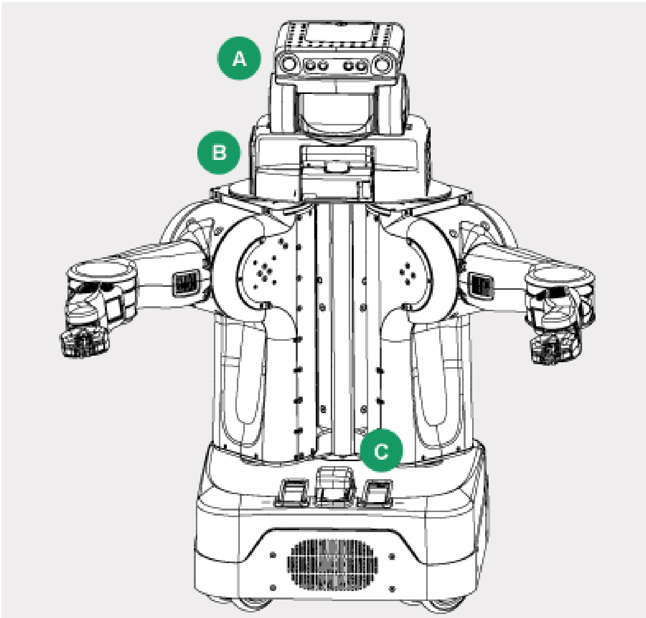

Fig. 2: A wire-frame model of the PR2. The location of sensors are indicated A-C and described in Sec. 2.

Fig. 3: Using the dataset comparison between different algorithms and sensor configurations is straightforward. In this figure we present an illustration of the ground-truth for one log-file compared to wheel odometry and visual odometry (Huang et al. (2011)). The occasions where the visual odometry algorithm failed to produce a reliable output are also shown, as is the building floor plan.

iments that make up this dataset:

1. Willow Garage Wide-Angle Global Shutter Color Stereo Ethernet Camera (Location A)
2. Microsoft Kinect Active Infrared RGB-D Camera (Location A)
3. Tilting Hokuyo UTM-30LX Laser Scanner (Location B)
4. Microstrain 3DM-GX2 IMU (Location B)
5. Base Hokuyo UTM-30LX Laser Scanner (Location C)
6. Wheel odometry (both raw and integrated, see Figure 3)
7. Internal joint readings

The broad location of these sensors are illustrated in Figure 2. The robot also contains a 5-Megapixel monocular camera and a narrow-angle monochrome stereo camera. The former was occasionally logged at low rates and the latter was never logged.

More details about the PR2 platform itself can be found on the PR2 web-page **`http://www.willowgarage.com/pages/pr2/overview`**

2.1 Coordinate Frames

The robot has a spine extension mechanism, a pan-and-tilt head and 4 caster wheels in addition to each of the 8 degree of freedom arms. In total 100 relative transforms describe the system and the transform tree can be visually inspected here: **`http://projects.csail.mit.edu/stata/pr2_frames.pdf`**

The joint-to-joint coordinate transform tree (created and updated by the ROS **tf** package) was logged at a typical rate of 500Hz, however beyond the tilting laser scanner, the relative transforms of the robot were typically fixed throughout a particular experiment. Transform rotations are maintained as quaternions.

The arms were kept in a tucked position at all times and are unlikely to be of interest to the reader. Finally, the robot was teleoperated by a series of volunteers who accompanied the robot at all times. The building contains about a dozen elevators, however during the experiments the robot traveled in only two of them to maintain uniformity.

3 Dataset Overview

Given the extent of the dataset, we will not provide details of each individual log here. The dataset website indicates which floors were explored and the timestamps at which elevator transitions began. Additionally we provide overview maps of the trajectory of the robot on particular floors, as discussed in Section 4. Instead we will provide an overview of the entire dataset.

Some figures of merit of the overall dataset are as follows:

1. Period covered: January 2011 to September 2012.
2. Total distance traveled: 42km over 10 floors.
3. Total file-size: about 2.3 Terabytes.
4. Total time taken: about 38 hours, 84 sessions.

To the best of our knowledge this dataset is the largest such dataset of repeated operation in a single location. We hope that it will provide the basis for future research tackling the scaling limitations of robotics research problems such as Simultaneous Localization

ROS, Robotics Operating System, is a software framework providing hardware abstraction, message passing and package management developed by Willow Garage and used by the PR2.

On occasion the height of the sensing shoulder, the pitch angle of the head varied. This is captured in the transformation tree.

Fig. 4: Ray and Maria Stata Center designed by Frank O. Gehry and completed in 2004. The 10 floor, 67,000 m^2 building is the home to the Computer Science and Artificial Intelligence Laboratory.

and Mapping (SLAM) as well as more generally contributing to long-term understanding and autonomy for robotic systems.

This dataset is also diverse and varied: In addition to capturing people moving and furniture being repositioned; lighting conditions change between seasons and decorations and renovations have been carried out to the building structure. Some of these changes are detected by the robot's sensors while others are only apparent upon revisiting the specific location days afterwards.

3.1 Why such a Large Dataset?

There has been interest in long term challenges for robotic autonomy and perception — as evidenced by a series of workshops at the ICRA, IROS and RSS robotics conferences as well as several academic journal special editions. Nonetheless long term operation is still typically defined in the order of minutes and hours (and limited by battery life), however only at extended durations do many of the scalability problems of current approaches become apparent.

While some notable demonstrations of long term robotic operation do exist, including the Mars Exploration Rovers, the Google Autonomous Car and Willow Garage's own marathon PR2 demonstration, Marder-Eppstein et al. (2010), the raw sensor data collected by these systems has not been made available to the research community.

This dataset aims to provide a platform for open comparison between approaches supporting long term robotic operation and their direct comparison using quantitative metrics. The authors plan to continue to expand the dataset, and it is intended to eventually become a living document of the experiences that a typical indoor mobile robot would detect during the course of a lifetime of operation.

In brief we suggest that a robotic perception system which can maintain a coherent understanding of its environment and can gradually update, renew and indeed improve that understanding when processing a dataset so substantial as this — including starting from any random starting point — can be truly said to be robust, reliable and scalable.

3.2 Comparison to other datasets

Within the vision community the Middlebury Evaluation System, Scharstein and Szeliski (2002), is an example of a particular dataset which has been widely used to provide **quantitative** comparison between different algorithms and implementations. The dataset presented in this paper is intended to mimic this approach for robotic mapping (and Visual SLAM in particular) by the provision of an extensive, rich multi-sensor dataset.

Simulated and real-world datasets such as the Victoria Park and Intel datasets have been widely used to compare LIDAR-based SLAM algorithms. More recently the New College and Rawseeds datasets have been while also been used of vision research. However both datasets contain only a few of loops and were collected at a single time. For ground-truthing, only GPS was available in the case of the former, while the latter used a mix of GPS and manual laser scan-matching.

For RGB-D cameras specifically, such as the Microsoft Kinect, Sturm et al. (2011) have developed an automated system for the comparison of 3D RGB-D SLAM systems using a motion capture system to provide ground-truth — the Freiburg dataset. However, due to the constraints of the motion capture system the Freiburg dataset is limited to a small environment and duration (36 minutes). Our dataset is intended to be symbiotic with the Freiburg dataset, but at a larger scale. In particular it is likely that inter-frame accuracy is likely to be higher for the Freiburg dataset than our dataset.

4 Ground-truthing the Dataset

As mentioned above, access to reliable ground-truth has typically been lacking in robotic mapping — making end-to-end comparison between SLAM systems qualitative rather than quantitative. Additionally, we recognize that the utility of any mapping database is vastly increased by providing access to ground-truth robot and sensor pose measurements.

For these reasons we have made considerable effort to estimate the position of the robot at each instant of the data logs via a ground-truthing system which we believe has a typical accuracy of 2–3cm. While we recognize that error minimization algorithms for SLAM can

This would enable queries such as 'provide all log segments when the robot was in a Room 2-213 in June 2012'.

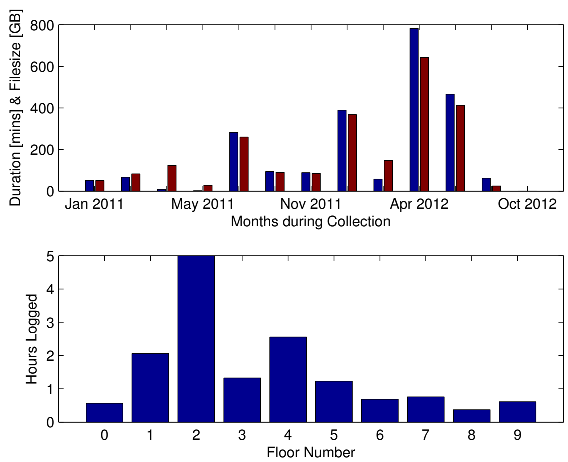

Fig. 5: Top: Collection rate of the dataset per month over time in duration (mins, blue) and file-size (GB, red). In some months no data was collected, these are excluded. Bottom: Distribution of the data for different floors (for a 17hr portion of the full dataset. The robot was based first on the second floor and later on the fourth. Floor 0 is the building's basement and Floor 1 the ground floor.

report residual measurement errors in this order of magnitude, this 2–3cm figure is in a single global coordinate frame common to all logs.

The Stata Center is a relatively new building (finished in 2004). MIT's building services department maintains accurate and reliable 2D building plans of each floor of the building, as illustrated in Figure 6. Our ground-truthing procedure involved creating alignments between the LIDAR scans from the PR2's base laser scanner and these floor-plans.

The procedure involved repeated solution of small optimization problems for approximately 3 seconds of data:

- First, the alignment of a scan at the start of this period with the floor-plan was determined. This scan was chosen such that a reliable and accurate alignment could be observed. (This was typically verified manually).
- Second, the corresponding alignment for a scan 3 seconds subsequent to this was determined in the same manner.
- Incremental consecutive alignment of the intermediate 120 scans to one another was determined using a dense scan-matcher. This approach was more accurate than continuous alignment to the floor-plan as in certain locations only part of the floor-plan was unobstructed by furniture.
- The marginal drift in the scan-matcher over this period results in a small inconsistency which we smoothed using iSAM, Kaess et al. (2008). The smoothing problem consisted of a graph of two fixed poses connected by the chain of incremental alignments.

The scan-matching algorithm used is that presented by Bachrach et al. (2011). An example of the accuracy of the alignment is illustrated in Figure 7.

We have taken particular care during the alignment process that the poses estimated are in no way correlated across the log files — aiming to provide instantaneous measurements of error for each pose to an accuracy limited only by the LIDAR sensor. We estimate this accuracy to be approximately 2–3 cm, based on the clarity of the reprojected returns.

Finally, the PR2's ROS coordinate frame manager, **tf**, maintains the internal relative calibration of each sensor. This allows us to derive the 3D positions of its tilting Hokuyo, Microsoft Kinect and stereo camera by combining the relative position from the base laser to these sensors with the ground-truth position of the base laser.

The tools used to generate the ground-truth results and to carry out SLAM evaluation are available to download from:
`http://projects.csail.mit.edu/stata/tools.php`

5 Tools for Accessing the Data

The most straightforward way to access the data is simply by replaying the individual log-files using ROS. Various tools such as transform caching, stereo image debayering and rectification are provided by ROS as well as visualization tools.

However, as some readers may not be interested in using the ROS infrastructure and so as to support long term archival of the data, we have also made a human readable version of the logs available in an easy to use format described on our website. The individual camera images are provided in PNG format in a folder system.

For example the following folder path: /2012-01-28-12-38-24/wide_stereo/left/image_raw/1327783200/1327783203593310687.png represents a specific image from this log in January 2012, from the ROS message channel 'wide_stereo/left/image_raw'.

Finally we also provide a straightforward wrapper program (in C++) to the FOVIS visual odometry library, Huang et al. (2011), which we ourselves use within our research. This enables researchers to begin developing vision-based mapping out 'out of the box' using reliable visual odometry. The library supports both the stereo camera and the RGB-D/Kinect camera and we provide calibration for both cameras. The library's original source code and documentation is available from:
`http://code.google.com/p/fovis/`

Additionally we also provide scripts for automated error metric analysis, see Figure 3 for an example.

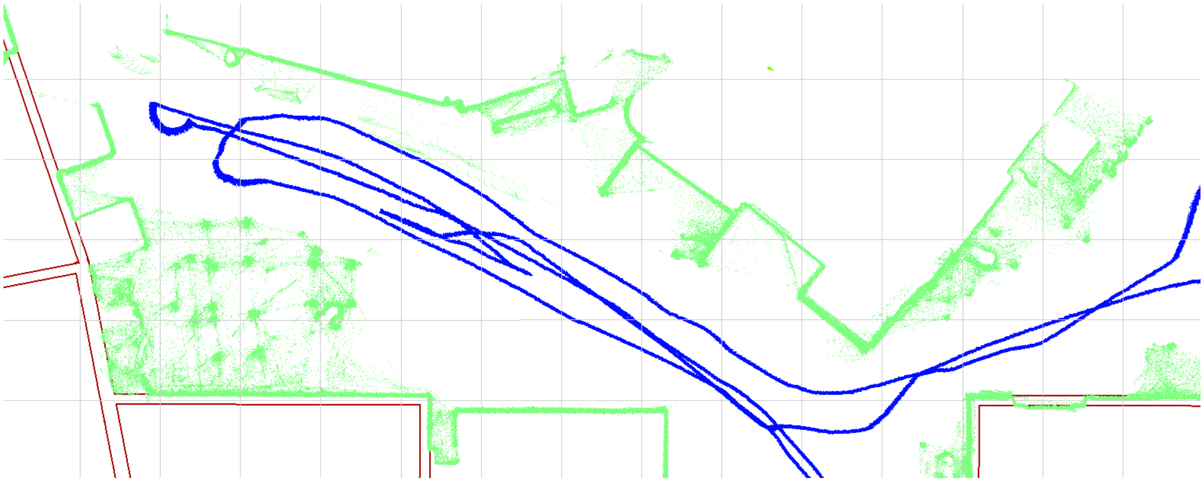

Fig. 7: Example of a portion of the ground-truthed PR2 trajectory (blue) and the reprojected LIDAR returns (green) aligned with the floor-plan. Each box is 0.5m square. Alignment accuracy is indicated by distribution of range measurements along the existing walls, despite the location containing furniture and clutter as well as door recesses. We estimate typical pose accuracy to be approximately 2-3 cm.

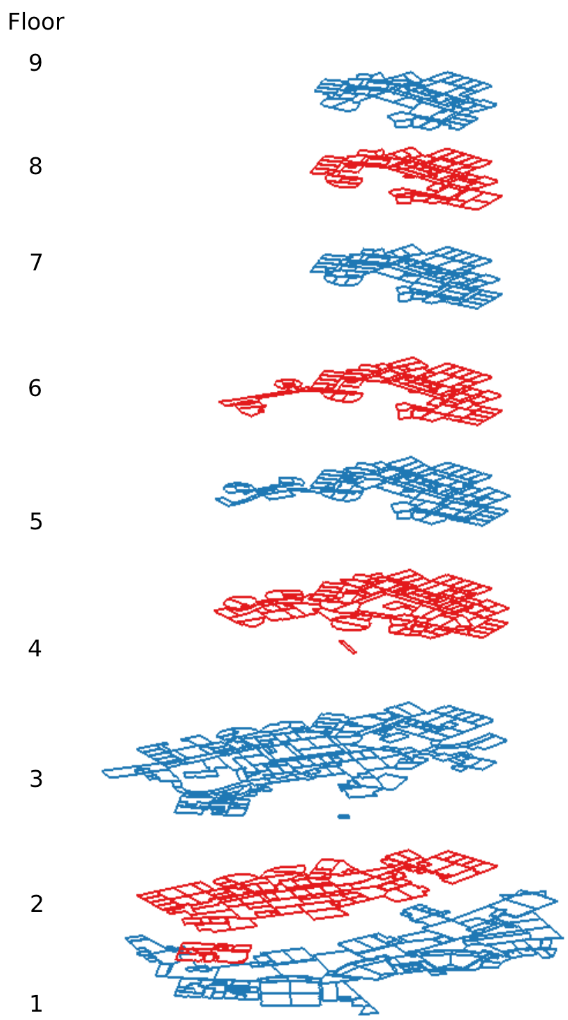

Fig. 6: Aligned floor-plans of each of the Stata Center's floors.

Acknowledgements

This dataset is released under the Creative Commons 3.0 license, as described on the aforementioned website.

We would like to acknowledge Alper Aydemir who made available the floor plans described in Section 4. Models of MIT and KTH are available from: **`http://www.csc.kth.se/~aydemir/floorplans.html`**

References

Bachrach, A., Prentice, S., He, R., and Roy, N. (2011). RANGE - robust autonomous navigation in GPS-denied environments. *J. of Field Robotics*, 28(5):644–666.

Huang, A., Bachrach, A., Henry, P., Krainin, M., Maturana, D., Fox, D., and Roy, N. (2011). Visual odometry and mapping for autonomous flight using an RGB-D camera. In *Proc. of the Intl. Symp. of Robotics Research (ISRR)*, Flagstaff, USA.

Johannsson, H. (2013). *Toward Lifelong Visual Localization and Mapping*. PhD thesis, Massachusetts Institute of Technology.

Johannsson, H., Kaess, M., Fallon, M., and Leonard, J. (2013). Temporally scalable visual SLAM using a reduced pose graph. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, Karlsruhe, Germany.

Kaess, M., Ranganathan, A., and Dellaert, F. (2008). iSAM: Incremental smoothing and mapping. *IEEE Trans. Robotics*, 24(6):1365–1378.

Marder-Eppstein, E., Berger, E., Foote, T., Gerkey, B., and Konolige, K. (2010). The office marathon: Robust navigation in an indoor office environment. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*.

Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42.

Sturm, J., Magnenat, S., Engelhard, N., Pomerleau, F., Colas, F., Burgard, W., Cremers, D., and Siegwart, R. (2011). Towards a benchmark for RGB-D SLAM evaluation. In *Proc. of the RGB-D Workshop on Advanced Reasoning with Depth Cameras at Robotics: Science and Systems Conf. (RSS)*, Los Angeles, USA.