Robot@Home, a Robotic Dataset for Semantic Mapping of Home Environments

J.R. Ruiz-Sarmiento, C. Galindo, and J. Gonzalez-Jimenez

Machine Perception and Intelligent Robotics Group
System Engineering and Auto. Dept., Instituto de Investigación Biomédica de Málaga (IBIMA), University of Málaga, Campus de Teatinos, 29071, Málaga, Spain.

{jotaraul,cgalindo,jgonzalez}@uma.es

Abstract

This paper presents the *Robot-at-Home dataset* (Robot@Home), a collection of raw and processed sensory data from domestic settings aimed at serving as a benchmark for semantic mapping algorithms through the categorization of objects and/or rooms. The dataset contains 87,000+ time-stamped observations gathered by a mobile robot endowed with a rig of 4 RGB-D cameras and a 2D laser scanner. Raw observations have been processed to produce different outcomes also distributed with the dataset, including 3D reconstructions and 2D geometric maps of the inspected rooms, both annotated with the ground truth categories of the surveyed rooms and objects. The proposed dataset is particularly suited as a testbed for object and/or room categorization systems, but it can be also exploited for a variety of tasks, including robot localization, 3D map building, SLAM, and object segmentation. Robot@Home is publicly available for the research community at `http://mapir.isa.uma.es/work/robot-at-home-dataset`.

keywords: *Semantic mapping, object categorization/recognition, room categorization/recognition, contextual information, mobile robots, domestic robots, home environment, robotic dataset, benchmark*

1 INTRODUCTION

The extraction and representation of semantic knowledge of the world is a crucial step toward achieving intelligent robots (Pronobis et al. 2010). Semantic maps enrich traditional metric and topological maps with high-level information, which enables the robot to process commands like "go to the bedroom and stop the alarm" (Galindo & Saffiotti 2013). In this way, the robot has to create and manage its own internal representation of the world incorporating the needed semantic knowledge, e.g. this room is a *bedroom* and contains an *alarm clock* placed on a *night table*. Two major problems arise in the extraction of this information: *object categorization*, i.e. to label parts of the robot sensory data as belonging to a certain object class (bed, night table, alarm clock, etc.), and *room categorization*, i.e. to classify areas of the environment as rooms of a certain type (kitchen, bedroom, living room, etc.).

In order to cope with this categorization problem[1], a large number of sample data are needed to test, validate and compare different solutions. Considering this, the research community has released a number of public repositories on the internet, e.g. PASCAL (Everingham et al. 2010), NYUv2 (Silberman et al. 2012), ImageNet (Russakovsky et al. 2014), or SUN3D (Xiao et al. 2013). However, these datasets exhibit shortcomings when used by cutting-edge categorization techniques leveraging contextual information (Anand et al. 2013, Ruiz-Sarmiento et al. 2015*b*). Synthetic data could be used instead under specific circumstances (Ruiz-Sarmiento et al. 2015*a*), albeit real sensory datasets are preferred in most cases.

In this work we present the Robot-at-Home dataset (Robot@Home), a compilation of raw and processed data gathered by a mobile robot in different domestic settings. This dataset is unique in three aspects: the sensory system employed for its gathering, the diversity and amount of provided data, and the availability of dense ground truth information. Data collection followed a *place-centric* perspective (Xiao et al. 2013), and comprises 87,000+ timestamped observations as sequences of RGB-D images and 2D laser scans taken in 5 apartments. These raw data fully cover the common challenges to be faced by a robotic categorization system, like changing lighting conditions, occlusions, viewpoint variations, or cluttered room layouts. On the other hand, the processed data include:

- *Per-pixel labeling* (ground truth information) of every RGB-D observation, along with the *category of the room* containing them.
- *3D reconstructions* in the form of colored point maps and *2D geometric maps* of the inspected rooms.
- *Per-point object labeling* of the 3D reconstructed rooms along with their *room category*.
- Topology of each apartment, stating the connectivity of the rooms within them.

During the data collection, a total of 36 rooms were completely inspected, so the dataset is rich in contextual information of objects and rooms. This is a valuable feature, missing in most of the state-of-the-art datasets, which can be exploited by, for instance, semantic mapping systems that leverage relationships like *pillows are usually on beds* or *ovens are not in bathrooms*. Robot@Home is publicly available and is accompanied with the software application employed for its processing, named the Object Labeling Toolkit (OLT) (Ruiz-Sarmiento et al. 2015*c*).

The sensory system comprises a rig of 4 RGB-D cameras and a radial laser scanner (see Fig. 1). The rig covers ∼180° horizontally and ∼58° vertically, which permits the user to simulate the performance of sensors with different field of views, a valuable feature during the design of a robotic sensing system (de la Puente et al. 2014). Sensors have been intrinsically and extrinsically calibrated with state-of-the-art algorithms (Fernandez-Moral et al. 2014, Gómez-Ojeda et al. 2015, Teichman et al. 2013). It is worth mentioning that a number of distinctive patterns and objects have been strategically added to the apartments for possible exploitation of the dataset in robotic competitions, like those in RoboCup@Home (Almeida et al. 2016) or RobotVision (Martinez-Gomez et al. 2014), where robots need to detect predefined patterns in the environment to accomplish certain challenging missions: to explore specific areas, to efficiently find a particular object, etc. In summary, this dataset contributes a repository suitable for a variety of robotic tasks like object/room categorization or recognition[2], object segmentation, 2D/3D map building, and robot localization among others.

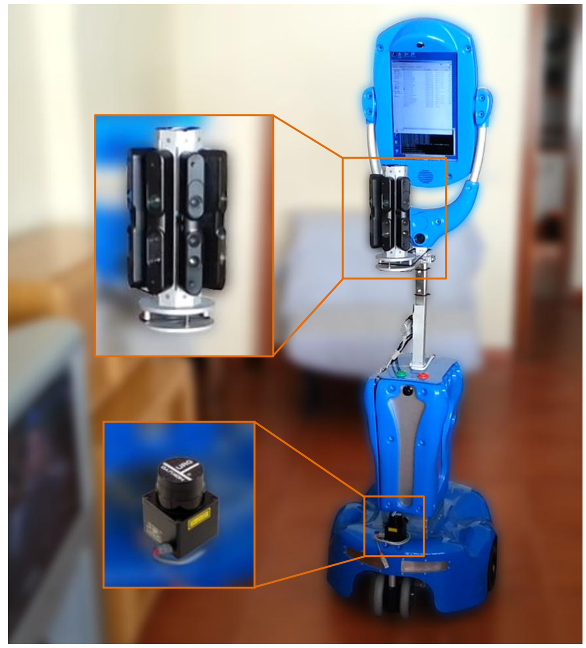

Figure 1: Robotic platform employed to collect the dataset along with details of the sensors mounted on it.

Next section contrasts Robot@Home with other datasets also applicable to the categorization problem. Section 3 presents the robotic platform used and the methodology followed for gathering the raw data, while section 4 describes the dataset content and some use cases. Finally, section 5 summarizes the paper.

2 Related Datasets

Mobile robots have traditionally resorted to intensity images to categorize objects and/or rooms, which motivated the collection of datasets providing this kind of information (Everingham et al. 2010, Russell et al. 2008, Russakovsky et al. 2014). Nowadays, the tendency is for the datasets to also include depth information (Janoch et al. 2011, Anand et al. 2013, Lai et al. 2011), given the proved benefits of exploiting morphological and spatial information in assisting categorization methods (Ruiz-Sarmiento et al. 2014). These datasets can be roughly classified as: *object-centric*, *view-centric*, and *place-centric*.

Object-centric datasets, like ACCV (Hinterstoisser et al. 2013), RGBD Dataset (Lai et al. 2011, 2014), KIT object models (Kasper et al. 2012), or BigBIRD (Singh et al. 2014), provide RGB-D observations in which a

Table 1: Summary of related datasets (CR: Collected by a robot, DT: Dataset type, EOC: Enables object context exploitation, ERC: Enables room categorization).

Dataset	CR	DT	EOC	ERC	#obs / size
ACCV Hinterstoisser et al. (2013)		*object-centric*			18,000 / 3.6GB
Berkeley-3D Janoch et al. (2011)		*view-centric*	✓(local)	✓(limited)	849 / 0.8GB
UMA-Offices Ruiz-Sarmiento et al. (2015*a*)		*view-centric*	✓(local)	✓(limited)	25 / 0.01GB
BigBIRD Singh et al. (2014)		*object-centric*			150,000 / 2,625GB
Cornell-RGBD Anand et al. (2013)	✓	*view-centric*	✓(local)	✓(limited)	207 / 0.1GB
KIT object models Kasper et al. (2012)		*object-centric*			163,188 / –
Multi-sensor 3D Object Dat. Garcia-Garcia et al. (2016)		*object-centric*			1,792 / 0.84GB
NYUv1 Silberman & Fergus (2011)		*view-centric*	✓(local)	✓(limited)	51,000 / 90GB
NYUv2 Silberman et al. (2012)		*view-centric*	✓(local)	✓(limited)	408,000 / 428GB
RGBD Dataset Lai et al. (2011)		*object-centric*			– / 84GB
RGBD Dataset 2 Lai et al. (2014)		*view-centric*			11,427 / 5.5GB
TUW Aldoma et al. (2014)	✓	*view-centric*	✓(local)	✓(limited)	124 / 0.43GB
SUN3D Xiao et al. (2013)		*place-centric*	✓	✓	– / –
UBC VRS Meger & Little (2012)	✓	*view-centric*	✓(local)		1,082 / –
Robot@Home	✓	*place-centric*	✓	✓	87,891 / 9.6GB

unique object spans over each image. The exploitation of these images for categorization exhibits some drawbacks: (i) they are not representative of the typical images gathered by a robot at a real environment, (ii) they prevent the utilization of valuable contextual information of objects, and (iii) they are not suitable for the room categorization problem.

On the other hand, *view-centric* datasets as Berkeley-3D (Janoch et al. 2011), Cornell-RGBD (Anand et al. 2013), UMA-Offices (Ruiz-Sarmiento et al. 2015*a*), NYU (Silberman & Fergus 2011, Silberman et al. 2012), TUW (Aldoma et al. 2014), or UBC VRS (Meger & Little 2012), consist of isolated RGB-D images, or a sequence of them, which cover a partial view of the working environment. This information permits the exploitation of contextual relations but only from a local, reduced perspective, since information of the entire scene is not collected. Therefore, their use for the categorization problem is still limited.

Finally, *place-centric* datasets like SUN3D (Xiao et al. 2013) provide comprehensive information from the inspected room, or even the entire work environment, typically through the registration of RGB-D images. This type of dataset provides the best option to take advantage of both depth and contextual information in the categorization problem, albeit, unfortunately its number is quite limited. A dataset worth mentioning at this point is ViDRILO (Martinez-Gomez et al. 2015), which comprises 5 sequences of RGB-D observations of two office buildings collected by a robot combining *object* and *environment-centric* perspectives. This dataset annotates each observation with its room type and the objects found within it, although this labeling is not per-pixel and the number of object categories is reduced. Tab. 1 shows a summary of datasets applicable to the categorization problem and their characteristics, including the novel, *place-centric* Robot@Home dataset.

3 Data Collection

3.1 Robotic platform

The Robot@Home dataset has been collected using the commercial robot Giraff (Giraff Technologies AB 2015), which consists of a motorized wheeled platform endowed with a videoconferencing set. The robot is controlled by a low-cost onboard computer running Windows 7, with a CPU Intel® Core™2 T7200 at 2Ghz., 1GB. of RAM and a 160 GB. hard disk. This platform has been enhanced with the following sensors:

- Four Asus XTion Pro Live RGB-D cameras (ASUS 2015) with a 58°x45° field of view (FOV). These devices can provide synchronized intensity and depth images at VGA (640x480) or QVGA (320x240) resolutions.
- A Hokuyo laser scanner model URG-04LX-UG01 (Hokuyo Automatic Co. 2015), a device that surveys 2D planes with a FOV of 240° and 0.352° of angular resolution.

The four RGB-D devices have been mounted vertically on an octagonal rig, which sets a radial configuration of camera's optical axes, with an angular difference of 45° (see Fig. 1). The rig is placed in the front part

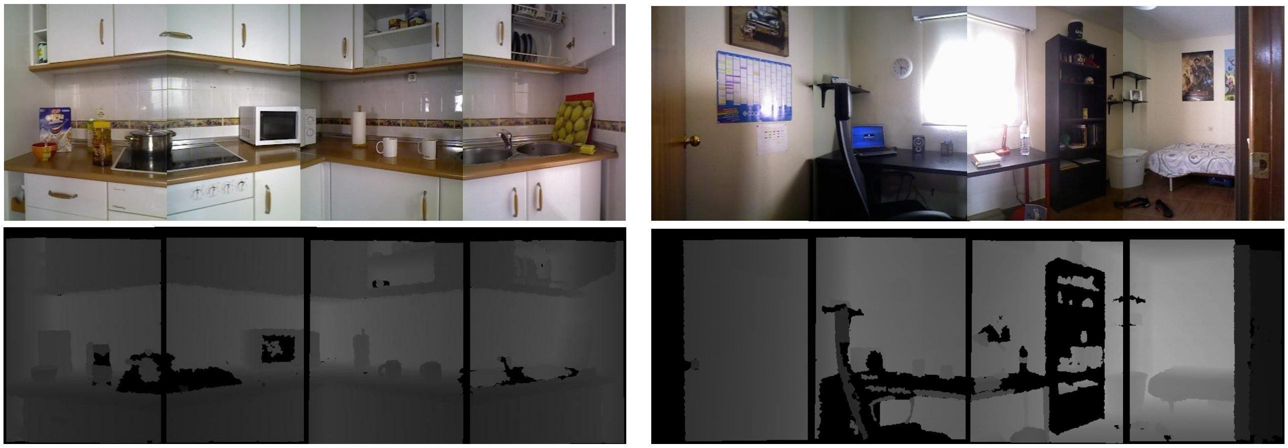

Figure 2: RGB and depth images from the 4 RGB-D devices mounted on the robot in two locations: a kitchen and a bedroom.

of the robot, at a height of ∼0.92m.[3], and the devices are connected to the onboard computer using a PCIe card with 4 USB 2.0 ports. Notice that the rig could hold up to 8 RGB-D cameras, but we considered that the utilization of 4 slots was enough for the purposes of this dataset. This setup yields two important advantages: first, there is no overlap among the FOV of the four units, avoiding in this way possible sensor interferences, and second, the combination of the output of the devices produces RGB-D observations with ∼180° of horizontal FOV (see Fig. 2).

Concerning the 2D laser scanner, it is mounted at the front part of the robot base (see Fig. 1), at a height of ∼0.31m. In this position the sensor cannot perceive any part of the robot while it surveys a plane horizontal to the floor at its maximum FOV.

3.2 Sensors calibration

In order to provide accurate information within the Robot@Home dataset, the sensors mounted on the robot must be calibrated both intrinsically and extrinsically. The locations of the devices mounted on the robot, i.e. their extrinsic parameters w.r.t. the robot frame[4], have been computed in a three-steps process. First, the RGB-D devices were calibrated between them following the technique in Fernandez-Moral et al. (2014). Then, the relative pose between the RGB-D devices and the laser scanner is obtained by the procedure presented in Gómez-Ojeda et al. (2015). Finally, the position of the RGB-D rig in the robot frame is computed by minimizing the error of fitting planes to the walls and the floor of a room using RANSAC (Fischler & Bolles 1981) while the robot is turning on the spot, and imposing vertical and horizontal conditions respectively. At this point every sensor is accurately related to the robot frame.

Regarding the sensors' internal parameters, for the correction of the depth images from the RGB-D devices we have resorted to the CLAMS framework (Teichman et al. 2013), while for the RGB and the laser scanner data we have relied on the factory values given their good outcome.

3.3 Software for the collection of data

Data streams coming from the five devices, i.e. 4 RGB-D cameras and the laser scanner, must be conveniently managed and stored. For that, in this work we have opted for the *rawlog-grabber* application from the Mobile Robot Programming Toolkit project (J.L. Blanco Claraco 2015), which provides mechanisms to collect and save sensory data from different sources into a file. In a nutshell, this software launches a dedicated thread for each sensor that time-stamps and saves the collected data to a compressed binary file in the *Rawlog* common robotic dataset format[5], which is automatically translated to human-readable information (plain text files and *PNG* images). Sensory observations have been saved at a frequency ranging from 1.25Hz up to 10Hz for the 2D laser scanner, and from 1Hz up to 11Hz for each RGB-D camera. These values are limited by the computational performance of the onboard computer, which have been compensated by reducing the robot speed (maximum of 0.1 m/s and 10 deg/s for linear and angular speeds respectively), ensuring in this

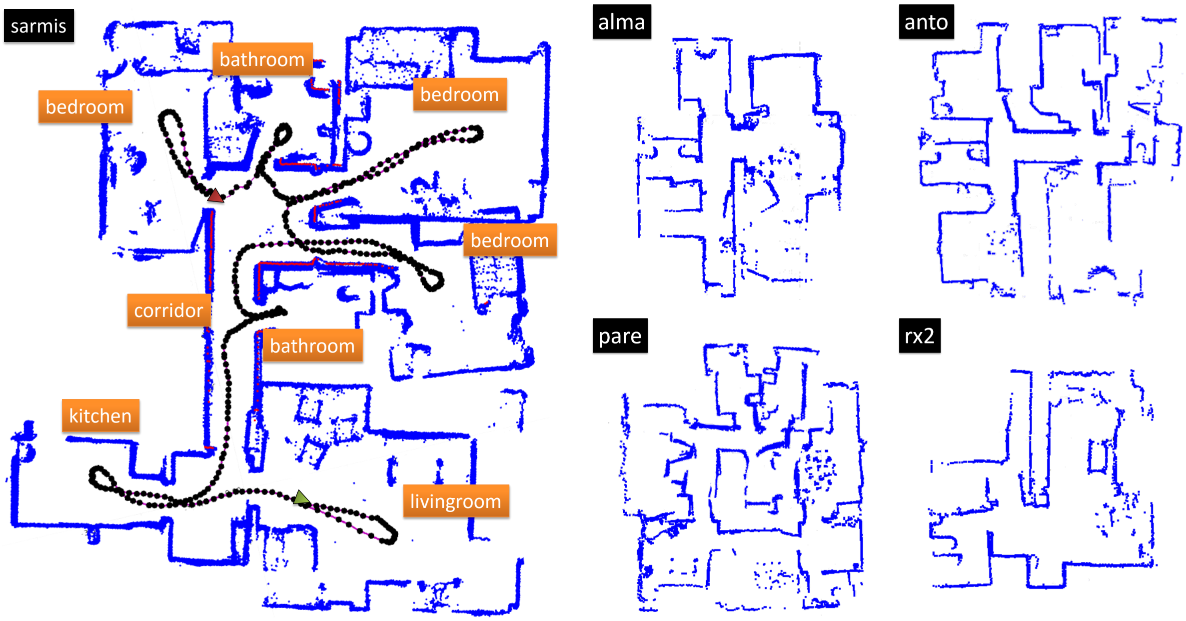

Figure 3: Left, example of 2D geometric map of the *sarmis* house, annotated with the type of the inspected rooms (orange boxes). The black dots represent the path followed by the robot during the inspection of the house, starting at the green triangle (livingroom) and ending up at the red one (corridor). Right, examples of geometric maps of the remaining domestic settings. For a better understanding of the descriptions resorting to color the reader is referred to the online version of this work.

way a good coverage of the inspected areas.

3.4 Collection methodology

The data provided by Robot@Home have been collected within 5 dwelling apartments, named *anto*, *alma*, *pare*, *rx2*, and *sarmis*. For illustrative purposes, Fig. 3 depicts their geometric maps, showing the annotations for the room categories in one of them. Raw data were collected in different sessions, each one containing a number of sequences of RGB-D observations and laser scans. These sequences were gathered by teleoperating the robot to fully inspect each individual room. Fig. 3 shows an example of the path followed by the robot while collecting a sequence of the *sarmis* house.

A total of seven sessions were conducted, three in the *sarmis* house and one in each of the remaining settings. During the data collection, special attention was paid to conveniently steer the robot in order to provide different viewpoints of the objects in the scene, so they can appear partially or totally occluded. As an example, the Fig. 4 shows a pencil case that is fully visible in the first and third images, although showing a different pose, while it is partially occluded in the second one, and totally disappears in the fourth image.

Moreover, a number of particular characteristics have been intentionally included in each scenario to provide additional data for testing different object recognition algorithms and techniques. Concretely,

- *Inclusion of distinctive objects.* A number of patterns/objects have been placed at different rooms within these houses, concretely: teddies in *alma*, fruits in *anto*, numerical patterns in *pare* (see top row of Fig. 5), and geometric patterns in *rx2* (see bottom row in Fig. 5).
- *Varying lighting conditions.* Each of the three sessions in *sarmis* house was conducted at a different time of the day, which means that the objects were visualized under different lighting conditions.
- *Varying sets of objects.* In those three sessions, the set of objects placed in each room from session to session differs, with objects dis/appearing as well as being moved (see Fig. 6).

4 Dataset description

4.1 Raw data

The Robot@Home dataset comprises ∼75 minutes of recorded data from a total of 83 sequences collected in

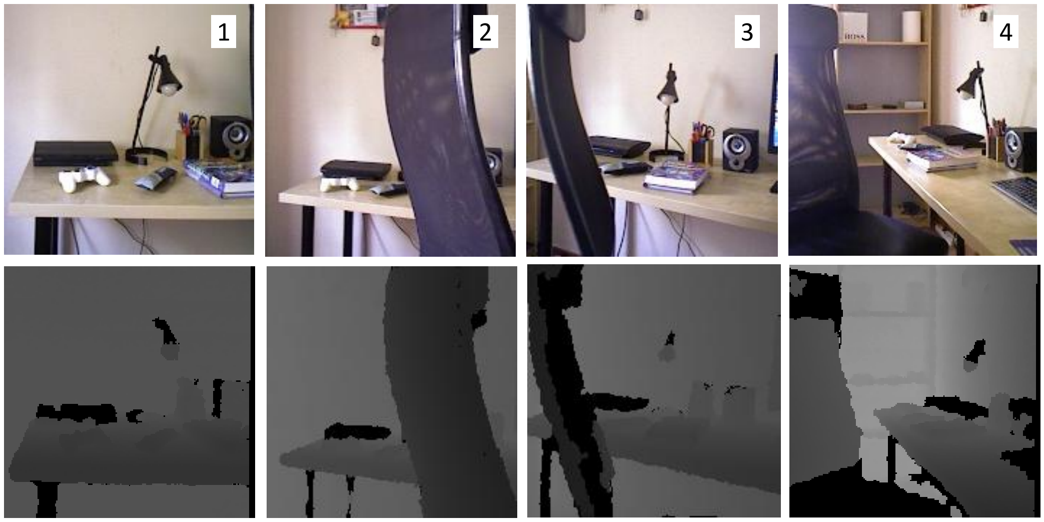

Figure 4: Top row, different viewpoints from a sequence of cropped intensity images of the same set of objects, and bottom, their associated depth images. Notice that throughout the sequence some objects are totally or partially occluded by others. Numbers indicate the order of the viewpoint within the sequence.

Table 2: Number of sequences, rooms, and observations per house and time spent collecting them.

	alma	anto	pare	rx2	Sarmi-house 1^{st} S.	Sarmi-house 2^{nd} S.	Sarmi-house 3^{rd} S.	Dataset
# Sequences	6	10	11	5	17	17	15	81
# Rooms	10	18	20	8	25	25	23	129
# Observations	15,535	22,301	26,506	9,906	4,939	4,218	4,486	87,891
# Laser scans	3,100	4,407	5,291	2,016	1,311	1,146	1,224	18,310
# RGBD obs.	12,435	17,894	21,215	7,890	3,866	3,276	3,519	69,581
Time (min)	5.12	7.53	8.48	3.22	17.47	15.25	16.30	74.57

the aforementioned sessions. These raw data include:

- ***Laser scanner data***: 2D observations from the laser scanner (see first row in Fig. 8) captured in the inspected rooms.
- ***RGB-D data***: Observations from the four RGB-D cameras, including intensity images, depth images, and 3D point clouds (see second and third row in Fig. 8).
- **Topological information** of the rooms connectivity, stating the rooms that are reachable by the robot from a certain location.

Table 2 shows a summary of the information gathered from each apartment, including the number of sequences, rooms inspected, number of 2D laser scans and RGB-D observations, as well as the time spent in their collection.

The surveyed scenarios include a total of 36 rooms (some of them visited several times), divided into 8 categories, that contain ~1,900 object instances belonging to 157 categories. An exhaustive list of the categories of the objects and rooms appearing in the dataset can be consulted at the dataset website.

4.2 Processed data

The raw data have been processed in order to enrich the dataset with the following information:

- ***2D geometric maps*** of each inspected room/house, built by registering the observations from the laser scanner.
- ***3D colored point maps***, 3D reconstructions of rooms based on the registration of the collected RGB-D data.

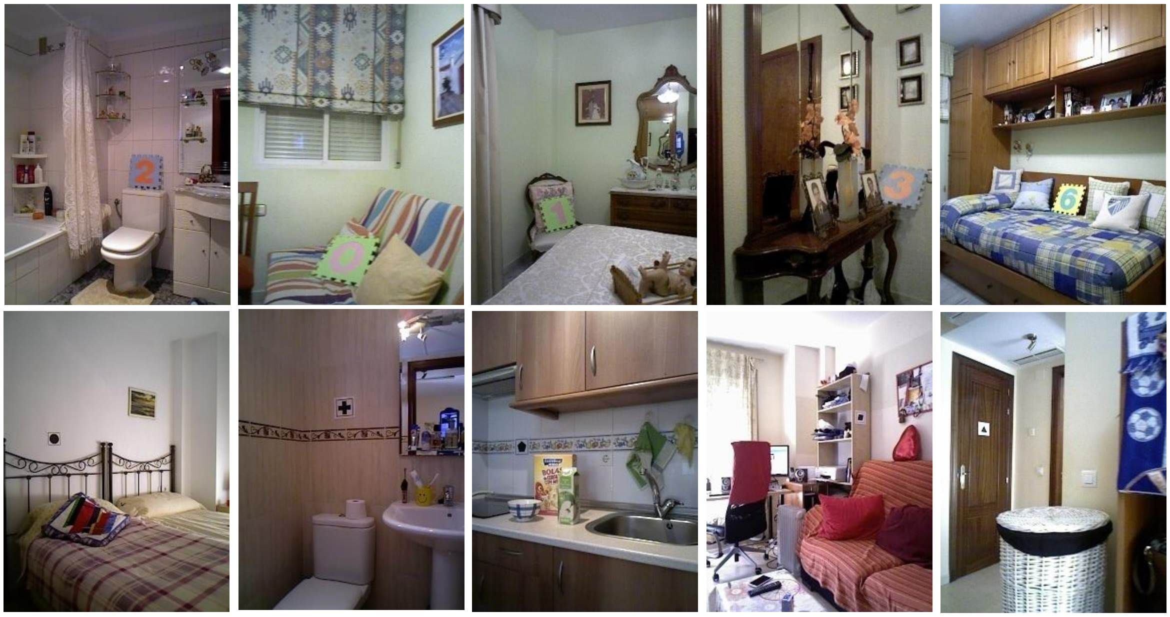

Figure 5: Top row, numerical patterns in *pare* house. Bottom row, geometric patterns in *rx2* house.

Figure 6: Intensity images of a bedroom from the same RGB-D sensor illustrating the change of lighting conditions during the three conducted sessions at *sarmis* house. The overlapped numbers represent the identifier of the session. It can be also observed how the set of visible objects differs from session to session.

- ***Labeled 3D point maps***, including per-point object and room labels (category and instance) within the reconstructed rooms.

- ***Labeled RGB-D observations***, including per-pixel object labels (category and instance) within each RGB-D observation, i.e., both intensity and depth images, and per-point labels within their respective point clouds.

Processed data have been produced employing two software tools, namely the aforementioned Mobile Robot Programming Toolkit (MRPT), and the Object Labeling Toolkit (OLT) (Ruiz-Sarmiento et al. 2015*c*). OLT comprises a set of public tools[6] aimed to help in the management and labeling of sequential RGB-D observations. Next sections describe with more detail the applications and methodologies followed to process the raw data.

4.2.1 2D geometric maps.

The *ICP-slam* application within MRPT has been used to register sequences of laser scans for building 2D geometric maps. Thereby, the Robot@Home dataset contains a total of 41 geometric maps, one per inspected room and a global map for each house (see Fig. 3 and Fig. 8, fourth row). These maps are distributed along with the *logs* produced during the SLAM process, which include additional information like the estimated path followed by the robot, snapshots of the scans' registration over time, etc.

4.2.2 3D colored point maps.

We have used the *Mapping* tool from OLT in order to produce aligned 3D representations of the recorded RGB-D data. This software registers sequences of RGB-D observations using the *Generalized Iterative Closest Point* technique (GICP) (Segal et al. 2009).

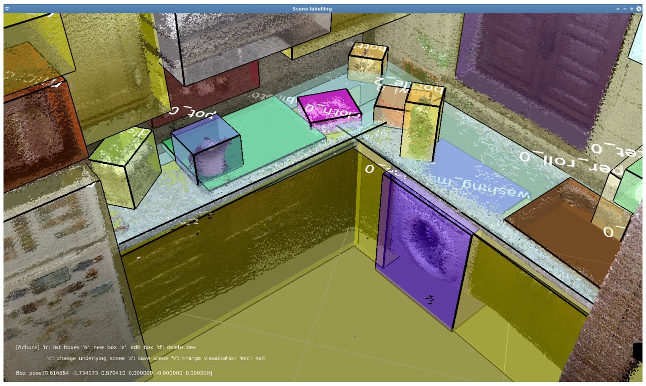

Figure 7: Snapshot of a kitchen from the *alma* house during its labeling process through the *Label scene* OLT component.

This ICP variant requires an initial pose estimation to accurately align RGB-D observations, which in our case is obtained using visual odometry (Jaimez & González-Jiménez 2015). Some examples of the provided reconstructions of rooms are shown in Fig. 8 (fifth row).

4.2.3 Labeled 3D point maps.

Each reconstructed room has been labeled with the *Label scene* tool from OLT. This tool allows us to easily set bounding boxes to the objects appearing in the point cloud reconstruction, and include annotations with the ground truth information about their category, e.g. *counter*, *book*, *couch*, *shelf*, as well as an object *id* to identify the particular instance, i.e. *counter-1*, *book-3*, etc. Fig. 8 (sixth row) illustrates some examples of annotations, while Fig. 7 shows a snapshot of the labeling process.

4.2.4 Labeled RGB-D observations.

Each RGB-D observation within the collected sequences has been also labeled with the category/instance of their contained objects through the *Label rawlog* application within OLT. This tool is fed with both the recorded sequence and the labeled, reconstructed map (obtained as described in the previous section) in order to automatically propagate the ground truth information to the RGB-D observations. The outcome of this process is a per-pixel labeling of the intensity and depth images within each observation, as well as a per-point labeling of its point cloud data (please refer to Ruiz-Sarmiento et al. (2015*c*) for further information). The last row of Fig. 8 depicts depth images colored according to the propagated ground truth labels.

4.3 Usage

All the raw and processed data within Robot@Home have been conveniently structured into data types and sessions at its site, so the interested user can download chunks of information according to his/her needs (see Fig. 9). The data are available in (human readable) plain text files[7] and *PNG* images. Some of their immediate applications are listed below.

Semantic mapping. The Robot@Home dataset is specially suited as a benchmark for algorithms aimed at robotic semantic mapping through the categorization of objects and/or rooms, given its collection by a mobile robot and the inclusion of annotated 3D reconstructions and sequences of RGB-D observations (Ruiz-Sarmiento et al. 2016, Oliveira et al. 2015). It can be also considered for testing recognition algorithms (Bo et al. 2013), since the provided ground truth information also includes the instance of the object/room to which it belongs to, e.g. *sofa_1*, *bottle_3*, *bathroom_1*, etc.

Robot@Home also enables the benchmarking of categorization systems relying on different kinds of information, namely: i) exclusively using laser scans, RGB, depth, or RGB-D observations, ii) employing a stream of data from a sequence, iii) resorting to partial registrations of such a sequence, or iv) exploiting the resultant whole registered scene.

From a semantic point of view, the compiled data are useful as input for modern categorization systems leveraging contextual information within domestic settings (Ruiz-Sarmiento et al. 2015*a*, Anand et al. 2013). This enables, for example, the exploitation of typical objects' and rooms' configurations like *beds are in bedrooms*, *microwaves are not in bathrooms*, or *cushions are on couches*, in order to enhance the categorization performance.

An additional feature worth mentioning is that Robot@Home is ready to be used by the *Benchmark rawlog* application from OLT. This software compares two sequences of labeled RGB-D observations and computes the similarity of their annotations. In other words, it permits us to compare a sequence from the dataset including ground truth annotations, with the same sequence labeled by a categorization algorithm, retrieving information about the performance of such algorithm. Thereby, a common benchmarking frame for the comparison of algorithms that exploit the Robot@Home dataset can be easily set.

In order to standardize the dataset usage for categorization/recognition purposes, we encourage the utilization of a leave-one-out cross-validation procedure (Arlot

Figure 8: Excerpts of the information provided by Robot@Home. From top to bottom, examples of 2D laser scans from three different rooms, RGB and depth images gathered from them, their built 2D geometric maps and 3D reconstructions, the labels in such reconstructions as boxes where colors stand for different object categories and, finally, the labeled depth information.

& Celisse 2010), where the data from one apartment are employed for testing and those from the remaining ones for training. This process is repeated 5 times, changing the testing home, and the individual results are finally averaged.

Object/room segmentation. Segmentation or clustering algorithms (Mura et al. 2014, Carreira & Sminchisescu 2012) can be also benchmarked given the per-pixel and per-point labeling of its reconstructed rooms and RGB-D sequences, which sets the extension and boundaries of the objects and rooms appearing in the dataset.

Simulation of virtual sensors. The coverage provided by the rig of RGB-D sensors, i.e., $\sim 180°$ horizontally and $\sim 58°$ vertically, enables the simulation of virtual sensors with different field of views. This is a valuable feature in the design phase of a robotic sensing system (de la Puente et al. 2014), since it permits the dataset user to test different sensing configurations in order to find the most convenient one for his/her purposes.

Data compression/transmission. Many robotic platforms have limited resources, which are typically shared among a number of software processes. In these cases efficient compression/transmission algorithms for dense sensory information are a plus (Kammerl et al. 2012, Mekuria & Cesar 2016), for which the amount of data provided within Robot@Home can be a useful testbed for checking their performance.

Other. Finally, the provided data can be also exploited for addressing typical robotic problems like 3D map building, localization (Castellanos & Tardos 2012) or SLAM (Cadena et al. 2016), since the robot's poses can be accurately estimated from the sequence of 2D scans.

5 Summary

In this work we have presented the Robot@Home dataset, a collection of data gathered by a mobile robot in domestic settings, publicly available at `http://mapir.isa.uma.es/work/robot-at-home-dataset`, which main purpose is to serve as a testbed for semantic mapping algorithms through the categorization of objects and/or rooms . Such a robot has been endowed with a rig of 4 RGB-D devices and a 2D laser scanner, which have been extrinsically and intrinsically calibrated employing state-of-the-art algorithms. Robot@Home comprises i) sequences of RGB-D observations and 2D laser scans from five home environments, ii) topological information about the connectivity of the rooms in those homes, iii) 2D geometric maps of the inspected rooms/homes, and iv) 3D reconstructions. Ground truth information about the categories of the observed objects and rooms is available in the form of v) annotated bounding boxes over the reconstructed rooms, and vi) labeled sequences of RGB-D observations.

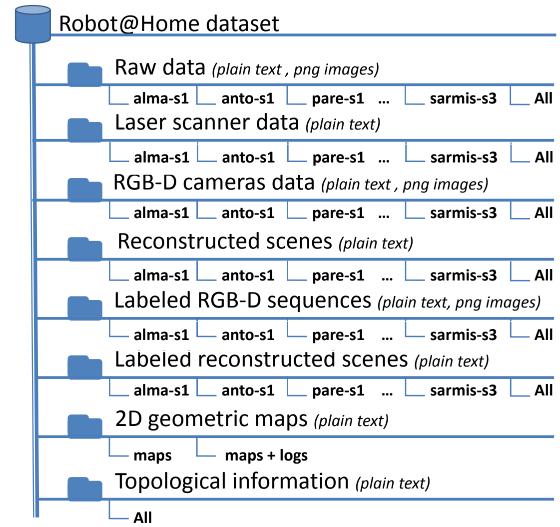

Figure 9: Tree structure of the data provided in the dataset webpage. Notice that the different types of data are available to the user both, split in sessions, or all together. The topology of the houses and the 2D geometric maps have particular, more convenient download options.

The surveyed scenarios include characteristics that turn the dataset into a sandbox to test robotic categorization systems dealing with issues like changing lighting conditions, cluttered room layouts, occlusions, or changing viewpoints. Additionally, a number of distinctive patterns and objects have been intentionally placed in these scenarios to enable their exploitation in robotic competitions. Although Robot@Home is specially suited as a benchmark tool for object and/or room categorization systems taking advantage of contextual relations among objects and rooms, its possible usages are diverse, e.g. object/room instance recognition, object segmentation, data compression/transmission, etc.

Acknowledgments. We are very grateful to our colleague E. Fernandez-Moral for his support for the extrinsic calibration of the RGB-D devices, and to Mariano Tarifa Jaimez for his valuable advices.

Funding. This work has been funded by the Spanish projects "IRO: Improvement of the sensorial and autonomous capability of Robots through Olfaction" [2012-TEP-530] and "PROMOVE: Advances in mobile robotics for promoting independent life of elders" [DPI2014-55826-R], and the European project "MoveCare: Multiple-actOrs Virtual Empathic CARgiver for the Elder" [Call: H2020-ICT-2016-1, contract number: 732158].

Notes

[1]For short, we use the term categorization to refer to the categorization of objects and rooms.

[2]Categorization methods yield information about the category of an object or room, e.g. *table* or *bedroom*, while the outcome of recognition methods refers to a particular instance, e.g. *table_1* or *bedroom_john*.

[3]The Giraff robotic platform was used in previous works like Jaimez et al. (2015), Melendez-Fernandez et al. (2016), Kiselev et al. (2015), and from such experiences we concluded that a height 0.92 is optimal for being able to *see* at a convenient distance both, the objects on surfaces like tables or counters, and the objects on the floor.

[4]In this work we have considered the origin of the robot frame as the center of the robot base.

[5]http://www.mrpt.org/Rawlog_Format.

[6]http://mapir.isa.uma.es/work/object-labeling-toolkit

[7]Each plain text file contains a header explaining its content.

References

Aldoma, A., Faulhammer, T. & Vincze, M. (2014), Automation of "ground truth" annotation for multi-view RGB-D object instance recognition datasets, *in* 'Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on', pp. 5016–5023.

Almeida, L., Ji, J., Steinbauer, G. & Luke, S. (2016), *RoboCup 2015: Robot World Cup XIX*, Lecture Notes in Computer Science.

Anand, A., Koppula, H. S., Joachims, T. & Saxena, A. (2013), 'Contextually guided semantic labeling and search for three-dimensional point clouds', *In the International Journal of Robotics Research* **32**(1), 19–34.

Arlot, S. & Celisse, A. (2010), 'A survey of cross-validation procedures for model selection', *Statistics Surveys* **4**, 40–79.

ASUS (2015), 'Xtion PRO LIVE', `http://www.asus.com/Multimedia/Xtion\_PRO\_LIVE/`. [Online; accessed 13-December-2016].

Bo, L., Ren, X. & Fox, D. (2013), Unsupervised feature learning for RGB-D based object recognition, *in* 'Experimental Robotics', Springer, pp. 387–402.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I. D. & Leonard, J. J. (2016), 'Simultaneous localization and mapping: Present, future, and the robust-perception age', *arXiv preprint arXiv:1606.05830* .

Carreira, J. & Sminchisescu, C. (2012), 'CPMC: Automatic object segmentation using constrained parametric min-cuts', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(7), 1312–1328.

Castellanos, J. A. & Tardos, J. D. (2012), *Mobile robot localization and map building: A multisensor fusion approach*, Springer Science & Business Media.

de la Puente, P., Bajones, M., Einramhof, P., Wolf, D., Fischinger, D. & Vincze, M. (2014), RGB-D sensor setup for multiple tasks of home robots and experimental results, *in* 'Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on', pp. 2587–2594.

Everingham, M., van Gool, L., Williams, C., Winn, J. & Zisserman, A. (2010), 'The PASCAL Visual Object Classes (VOC) challenge', *International Journal of Computer Vision* **88**(2), 303–338.

Fernandez-Moral, E., González-Jiménez, J., Rives, P. & Arévalo, V. (2014), Extrinsic calibration of a set of range cameras in 5 seconds without pattern, *in* '2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014)', Chicago, USA.

Fischler, M. A. & Bolles, R. C. (1981), 'Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography', *Communications of the ACM* **24**(6), 381–395.

Galindo, C. & Saffiotti, A. (2013), 'Inferring robot goals from violations of semantic knowledge', *Robotics and Autonomous Systems* **61**(10), 1131–1143.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Garcia-Rodriguez, J., Azorin-Lopez, J., Saval-Calvo, M. & Cazorla, M. (2016), 'Multi-sensor 3D object dataset for object recognition with full pose estimation', *Neural Computing and Applications* pp. 1–12.

Giraff Technologies AB (2015), 'Giraff robot', `http://www.giraff.org/`. [Online; accessed 13-December-2015].

Gómez-Ojeda, R., Briales, J., Fernndez-Moral, E. & Gonzlez-Jimnez, J. (2015), Extrinsic calibration of a 2D laser-rangefinder and a camera based on scene corners, *in* 'IEEE International Conference on Robotics and Automation (ICRA)', Seattle, USA.

Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G., Konolige, K. & Navab, N. (2013), Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes, *in* 'Proceedings of the 11th Asian Conference on Computer Vision - Volume Part I', ACCV'12, Springer-Verlag, Berlin, Heidelberg, pp. 548–562.

Hokuyo Automatic Co. (2015), 'Hokuyo URG-04LX-UG01', `http://www.hokuyo-aut.jp`. [Online; accessed 06-April-2015].

Jaimez, M., Blanco, J.-L. & González-Jiménez, J. (2015), 'Efficient reactive navigation with exact collision determination for 3D robot shapes', *International Journal of Advanced Robotic Systems* **12**(63).

Jaimez, M. & González-Jiménez, J. (2015), 'Fast visual odometry for 3-D range sensors', *IEEE Transactions on Robotics* **31**(4), 809–822.

Janoch, A., Karayev, S., Jia, Y., Barron, J. T., Fritz, M., Saenko, K. & Darrell, T. (2011), A category-level 3-D object dataset: Putting the Kinect to work, *in* '1st Workshop on Consumer Depth Cameras for Computer Vision (ICCV workshop)'.

J.L. Blanco Claraco (2015), 'Mobile Robot Programming Toolkit (MRPT)', `http://www.mrpt.org`. [Online; accessed 28-April-2015].

Kammerl, J., Blodow, N., Rusu, R. B., Gedikli, S., Beetz, M. & Steinbach, E. (2012), Real-time compression of point cloud streams, *in* '2012 IEEE International Conference on Robotics and Automation', pp. 778–785.

Kasper, A., Xue, Z. & Dillmann, R. (2012), 'The KIT object models database: An object model database for object recognition, localization and manipulation in service robotics', *The International Journal of Robotics Research* **31**(8), 927–934.

Kiselev, A., Kristoffersson, A., Melendez-Fernandez, F., Galindo, C., Loufti, A., González-Jiménez, J. & Coradeschi, S. (2015), Evaluation of using semi-autonomy features in mobile robotic telepresence systems, *in* '7th IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and the 7th IEEE International Conference on Robotics, Automation and Mechatronics (RAM)'.

Lai, K., Bo, L. & Fox, D. (2014), Unsupervised feature learning for 3D scene labeling, *in* 'Robotics and Automation (ICRA), 2014 IEEE International Conference on', pp. 3050–3057.

Lai, K., Bo, L., Ren, X. & Fox, D. (2011), A large-scale hierarchical multi-view RGB-D object dataset, *in* 'Robotics and Automation (ICRA), 2011 IEEE International Conference on', pp. 1817–1824.

Martinez-Gomez, J., Cazorla, M., Garcia-Varea, I. & Morell, V. (2014), Overview of the ImageCLEF 2014 Robot Vision Task, *in* 'CLEF 2014 Evaluation Labs and Workshop, Online Working Notes'.

Martinez-Gomez, J., Cazorla, M., Garcia-Varea, I. & Morell, V. (2015), 'ViDRILO: The visual and depth robot indoor localization with objects information dataset', *International Journal of Robotics Research* .

Meger, D. & Little, J. J. (2012), The UBC visual robot survey: A benchmark for robot category recognition, *in* 'Experimental Robotics - The 13th International Symposium on Experimental Robotics, ISER 2012, June 18-21, 2012, Québec City, Canada', pp. 979–991.

Mekuria, R. & Cesar, P. (2016), MP3DG-PCC, open source software framework for implementation and evaluation of point cloud compression, *in* 'Proceedings of the 2016 ACM on Multimedia Conference', MM '16, pp. 1222–1226.

Melendez-Fernandez, F., Galindo, C. & González-Jiménez, J. (2016), An assisted navigation method for telepresence robots, *in* '10th International Conference on Ubiquitous Computing and Ambient Intelligence'.

Mura, C., Mattausch, O., Villanueva, A. J., Gobbetti, E. & Pajarola, R. (2014), 'Automatic room detection and reconstruction in cluttered indoor environments with complex room layouts', *Computers & Graphics* **44**, 20 – 32.

Oliveira, M., Lopes, L. S., Lim, G. H., Kasaei, S. H., Sappa, A. D. & Tom, A. M. (2015), Concurrent learning of visual codebooks and object categories in open-ended domains, *in* '2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)', pp. 2488–2495.

Pronobis, A., Jensfelt, P., Sjöö, K., Zender, H., Kruijff, G.-J. M., Mozos, O. M. & Burgard, W. (2010), Semantic modelling of space, *in* H. I. Christensen, G.-J. M. Kruijff & J. L. Wyatt, eds, 'Cognitive Systems', Vol. 8 of *Cognitive Systems Monographs*, Springer Berlin Heidelberg, pp. 165–221.

Ruiz-Sarmiento, J., Galindo, C. & González-Jiménez, J. (2014), Mobile robot object recognition through the synergy of probabilistic graphical models and semantic knowledge, *in* 'European Conf. on Artificial Intelligence. Workshop on Cognitive Robotics'.

Ruiz-Sarmiento, J. R., Galindo, C. & González-Jiménez, J. (2015*a*), 'Exploiting semantic knowledge for robot object recognition', *Knowledge-Based Systems* **86**, 131–142.

Ruiz-Sarmiento, J. R., Galindo, C. & González-Jiménez, J. (2015*b*), Joint categorization of objects and rooms for mobile robots, *in* 'IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)'.

Ruiz-Sarmiento, J. R., Galindo, C. & González-Jiménez, J. (2015*c*), OLT: A Toolkit for Object Labeling Applied to Robotic RGB-D Datasets, *in* 'European Conference on Mobile Robots'.

Ruiz-Sarmiento, J. R., Galindo, C. & González-Jiménez, J. (2016), 'Building Multiversal Semantic Maps for Mobile Robot Operation', *Knowledge-Based Systems* . [Accepted, to appear].

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. S., Berg, A. C. & Fei-Fei, L. (2014), 'ImageNet large scale visual recognition challenge', *CoRR* **abs/1409.0575**.

Russell, B. C., Torralba, A., Murphy, K. P. & Freeman, W. T. (2008), 'LabelMe: A database and web-based tool for image annotation', *Int. J. Comput. Vision* **77**(1-3), 157–173.

Segal, A., Haehnel, D. & Thrun, S. (2009), Generalized-ICP, *in* 'Proceedings of Robotics: Science and Systems', Seattle, USA.

Silberman, N. & Fergus, R. (2011), Indoor scene segmentation using a structured light sensor, *in* 'Proceedings of the International Conf. on Computer Vision - Workshop on 3D Representation and Recognition'.

Silberman, N., Hoiem, D., Kohli, P. & Fergus, R. (2012), Indoor Segmentation and Support Inference from RGBD Images, *in* 'Proc. of the 12th European Conference on Computer Vision (ECCV 2012)', pp. 746–760.

Singh, A., Sha, J., Narayan, K., Achim, T. & Abbeel, P. (2014), BigBIRD: A large-scale 3D database of object instances, *in* 'Robotics and Automation (ICRA), 2014 IEEE International Conference on', pp. 509–516.

Teichman, A., Miller, S. & Thrun, S. (2013), Unsupervised intrinsic calibration of depth sensors via SLAM, *in* 'Proceedings of Robotics: Science and Systems', Berlin, Germany.

Xiao, J., Owens, A. & Torralba, A. (2013), SUN3D: A database of big spaces reconstructed using SfM and object labels, *in* 'Computer Vision (ICCV), 2013 IEEE International Conference on', pp. 1625–1632.