Using the Kinect as a Navigation Sensor for Mobile Robotics

Ayrton Oliver
Dept. of Electrical and Computer Engineering
University of Auckland
Auckland, New Zealand
aoli009@aucklanduni.ac.nz

Steven Kang
Dept. of Electrical and Computer Engineering
University of Auckland
Auckland, New Zealand
skan070@aucklanduni.ac.nz

Burkhard C. Wünsche
Dept. of Computer Science
University of Auckland
Auckland, New Zealand
burkhard@cs.auckland.ac.nz

Bruce MacDonald
Dept. of Electrical and Computer Engineering
University of Auckland
Auckland, New Zealand
b.macdonald@auckland.ac.nz

ABSTRACT

Localisation and mapping are the key requirements in mobile robotics to accomplish navigation. Frequently laser scanners are used, but they are expensive and only provide 2D mapping capabilities. In this paper we investigate the suitability of the Xbox Kinect optical sensor for navigation and simultaneous localisation and mapping. We present a prototype which uses the Kinect to capture 3D point cloud data of the external environment. The data is used in a 3D SLAM to create 3D models of the environment and localise the robot in the environment. By projecting the 3D point cloud into a 2D plane, we then use the Kinect sensor data for a 2D SLAM algorithm. We compare the performance of Kinect-based 2D and 3D SLAM algorithm with traditional solutions and show that the use of the Kinect sensor is viable. However, its smaller field of view and depth range and the higher processing requirements for the resulting sensor data limit its range of applications in practice.

Categories and Subject Descriptors

I.2.10 [**Vision and Scene Understanding**]: 3D/stereo scene analysis; I.4.8 [**Scene Analysis**]: Range data; I.2.9 [**Robotics**]: Autonomous vehicles

Keywords

mobile robotics, robot navigation, Kinect, SLAM, scene reconstruction

1. INTRODUCTION

Navigation in mobile robotics involves three key questions: where am I, where am I going and how do I get there. The corresponding terms in robotics are *self-localisation*, *map-building and map-interpretation*, and *path planning* [18]. One of the most important problems in this field is *Simultaneous Localisation and Mapping* (SLAM) [19], where a robot autonomously explores and maps its environment with its sensors while localising itself at the same time. Despite considerable research over the past two decades numerous challenges remain, such as implementations in unstructured, difficult and large scale environments [3] and multi-robot SLAM [10].

The use of SLAM is constrained by currently available sensors. Suitable laser sensors can cost US$ 2000 and more and frequently only produce a 2D view of the environment [1]. Optical sensors are much cheaper, but have a limited resolution and require complex algorithms to determine depth values from the camera images [4].

The Kinect is a motion-sensing input device for the Microsoft Xbox 360 video game console. The sensor bar contains a traditional web cam and an infrared projector and sensing system, which can produce a depth map with up to 1 cm accuracy. The device combines some of the advantages of laser sensors (depth map) with those of cameras (scan volume rather than scan line). With a price of less than US$ 150 the Kinect is very affordable. As 3D SLAM can be performed without odometry, the Kinect sensor could be particular useful for underground exploration, autonomous mapping of buildings, or remote exploration and visualisation of dangerous areas.

In this paper we investigate the feasibility and reliability of using the Xbox Kinect as a sensor for navigation and SLAM. We explore both 2D and 3D SLAM applications and compare the Kinect to the laser scanner Hokuyo URG-04LX in terms of SLAM performance and suitability as a navigation sensor.

Section 2 reviews related work. Section 3 presents the design and implementation details of our mobile robotics system. Section 4 presents an evaluation of the laser and Kinect sensor for 2D and 3D SLAM. Section 5 concludes this paper and gives an outlook on future work.

2. LITERATURE REVIEW

IVCNZ '12, November 26 - 28 2012, Dunedin, New Zealand

The Kinect sensor is composed of a colour image CMOS sensor, an Infra-Red (IR) CMOS sensor, an IR light source and a PS1080 chip. The Infra-Red light source projects an IR *Light Coding* image into the scene. An image of the projected light pattern is received by the IR CMOS sensor, and processed on the PS1080 chip using an image-based 3D reconstruction algorithm [15]. This produces an 11-bit 640×480 depth map of the scene at 30Hz.

While originally developed as a game controller, the Kinect is now used in a wide variety of applications ranging from art and advertisement to healthcare and business [9]. In 2011 the *KinectFusion* algorithm was presented, which uses the Kinect depth data to reconstruct a 3D scene using the Kinect sensor like a handheld laser scanner [11]. An open source implementation of this algorithm is available within the *Point Cloud Library* [14].

Since the Kinect's release in November 2010 a large number of projects have explored its use in mobile robotics. Applications range from navigation tasks where the sensor data is used for obstacle detection and path planning [5], to complete 3D mappings of environments [8].

Ruhnke et al. optimise the Kinect's pose and the positions of the surface points measured with it in order to create more accurate 3D reconstructions [17]. Endres et al. present a system performing SLAM and 3D reconstruction using the Kinect and evaluate it using a recently introduced benchmark for RGBD SLAM systems [6]. The authors report a RMSE of $9.7cm$ and 3.95° in a typical office environment.

Bachrach et al. use the Kinect for enabling an Unmanned Aerial Vehicles (UAV) to explore indoor environments, where GPS data is not available [2]. The authors use a Belief Roadmap algorithm to minimise the positional error of the helicopter by incorporating a predictive model of sensing.

Viager evaluates the use of the Kinect in mobile robotics [20]. The author reports that resolution is best at short distances, but that a minimum distance of 0.4 must be observed. Other reported problems are interference when using multiple Kinect sensors, and issues with robustness.

In this paper we compare the Kinect to the laser scanner Hokuyo URG-04LX in terms of 2D and 3D SLAM performance.

3. DESIGN & IMPLEMENTATION

3.1 System Architecture

Figure 1 illustrates our system architecture, which consists of a *Robot System* and a *Base System*.

The Robot System's hardware consists of the Kinect sensor, the Hokuyo laser scanner and the robot, a Pioneer III (P3-DX) mobile robot from Adept MobileRobots. A laptop was mounted onto the Pioneer robot. The robot has 2 differential drive wheels and is equipped with a USB laser scanner (Hokuyo URG-04LX) and sonar sensors. The Kinect is connected to the laptop. Its power adapter was cut off and a socket plug was attached that can be attached to the Pioneer robot's power supply.

The Robot System's software is based on the Robot Operating System (ROS), a software framework for robot software development, which is installed on the laptop. The ROS provides libraries and tools including drivers for the Kinect, robot base and laser scanner [16]. We used the following key utilities from ROS:

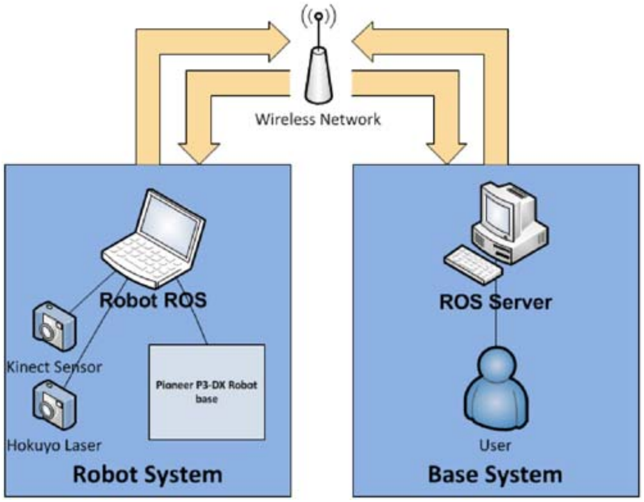


Figure 1: The system architecture of our mobile robotics system performing SLAM using a laser scanner and the Kinect sensor.

- **ROSBAG:** a tool that facilitates recording and play back of data being published by any ROS node, for example the Kinect's depth and colour images and the robot odometry. It saves the data to disk in a "bag" file for later use.
- **ROSTOPIC:** a command-line tool for displaying debug information, e.g., about received sensor data.
- **ROSLAUNCH:** a tool for initiating execution of a sequence of processing nodes.

The robot is controlled with a joystick and is connected to the laptop running ROS with an USB-to-Serial converter RS-232 cable. The ROS receives the sensor data and the robot's current status such as odometry, battery level, motor status and markers.

3.2 SLAM Implementation

3.2.1 2D SLAM

2D SLAM algorithms utilise 2D scan data and odometry information (the position of the robot relative to the initial position). In an unknown environment the algorithm will use the sensor data to create a map with estimates of obstacle positions. When the robot is moving the odometry information is used to update the map by integrating new sensor data into it, and to localise the robot within this map. This "step-correct process" will result in a more accurate position estimate than odometry alone, and it improves the 2D map of the environment at the same time.

In order to use the Kinect as a sensor for the 2D SLAM algorithm, the 3D depth data is transformed into 2D by projecting it onto the ground plane and using for each 2D direction the closest projected point. The Kinect and laser were calibrated using a simple test scene in order to align the depth data.

The 2D SLAM algorithm is performed using an ROS implementation of GMapping from OpenSLAM [13].

3.2.2 3D SLAM

3D SLAM is achieved by feature matching. The algorithm utilises depth and monotone images to extract features from the scene. These features are then matched with features extracted from previous frames. The relative position of features with respect to each other makes it possible to estimate the position of the robot, and to integrate the point cloud data of the current frame into the 3D scene geometry obtained from previous frames.

As 3D SLAM does not use odometry, it can be performed using any 3D (depth) camera, whether it be on a flying, ground based, or hand held robot. However, the absence of odometry means that the position estimate relies solely on the 3D feature matching stage, which can create large errors in the position estimate. This problem is most likely to occur if the algorithm cannot find any features, or the robot moves too fast and incoming frames can not be processed fast enough.

As 3D SLAM implementation we use RGBD SLAM [7], which performs pure visual SLAM using a feature matching algorithms. We found that the algorithm is very slow and requires a large amount of memory and CPU power. In order for the algorithm to work in real-time, the robot had to move extremely slowly. If not, consecutive frames varied significantly and feature matching failed. In our evaluation we hence used recorded sensor data (see subsection 4.2).

3.3 Real-Time Visualisation

Visualising the point cloud data across the network is useful for debugging, analysis, and robot operation. Initial testing indicated that this would not be a feasible as the RGB point cloud data from the Kinect is produced at 76.6 MB/s. The max transfer rate for Wi-Fi 802.11G is 54 Mbit/s. Therefore the update rate of the point cloud would fall from 30 Hz to 4 Hz due to the network's bandwidth limitation. Sending only depth points uses 37 MB/s, but is still not sufficient for the Kinect's 30Hz frame rate.

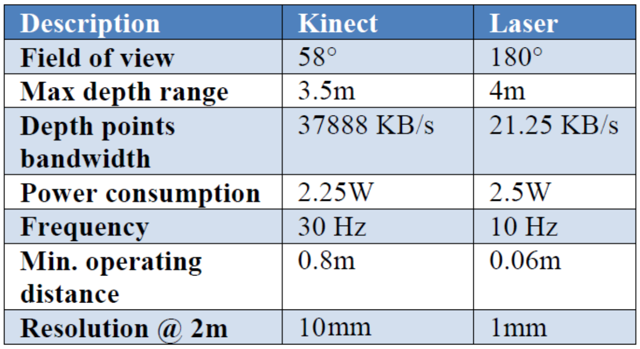
To overcome the limitations due to the network and large amount of data, the Kinect's resolution was reduced to QVGA (320 × 240). This meant that visualising the depth points could be achieved with relatively low latency. However, RGB depth points were still a problem. The bandwidth limitations were a key factor for the decision to run all SLAM and image processing nodes onboard the robot laptop. This enabled us to minimize delays and retain the high refresh rate (30 Hz).

4. RESULTS

4.1 Comparison of Specifications

Figure 2 shows a comparison of the specifications of the Kinect sensor and the Hokuyo URG-04LX laser scanner. Three key differences can be observed: (1) The laser scanner has a ten times higher depth resolution, which makes it much easier to match features across multiple views. (2) The laser scanner has a three times higher field-of-view, which enables it to detect much more features than the Kinect sensor. (3) The minimum operating distance of the Kinect is more than ten times higher, which means that a close range examination of features is often not possible.

4.2 Methodology



Description	Kinect	Laser
Field of view	58°	180°
Max depth range	3.5m	4m
Depth points bandwidth	37888 KB/s	21.25 KB/s
Power consumption	2.25W	2.5W
Frequency	30 Hz	10 Hz
Min. operating distance	0.8m	0.06m
Resolution @ 2m	10mm	1mm

Figure 2: Comparison of the specifications of the Kinect sensor and the Hokuyo URG-04LX laser scanner.

Figure 3: Testing environment with the robot at its starting position.

A small testing environment, displayed in figure 3, was set up. It included many features and differently sized objects to enable a diverse range of sensor readings to be recorded. Path markings on the ground were laid out to ensure that the robot would at some point face each object or wall from a distance of about 2 m, since this is the optimal distance for the Kinect's depth camera.

In order to get the same views of the environment for different sensors and SLAM algorithms we used a joy stick to drive the robot remotely around the environment. Data from the robot and sensors were collected and recorded using ROSBAG, namely:

- Kinect - using OpenNI drivers [12]
    - Depth images
    - Colour images
    - Monotone images
    - Camera information messages
- Laser scanner - using Hokuyo drivers
    - Scans
- Robot - using P2OS drivers

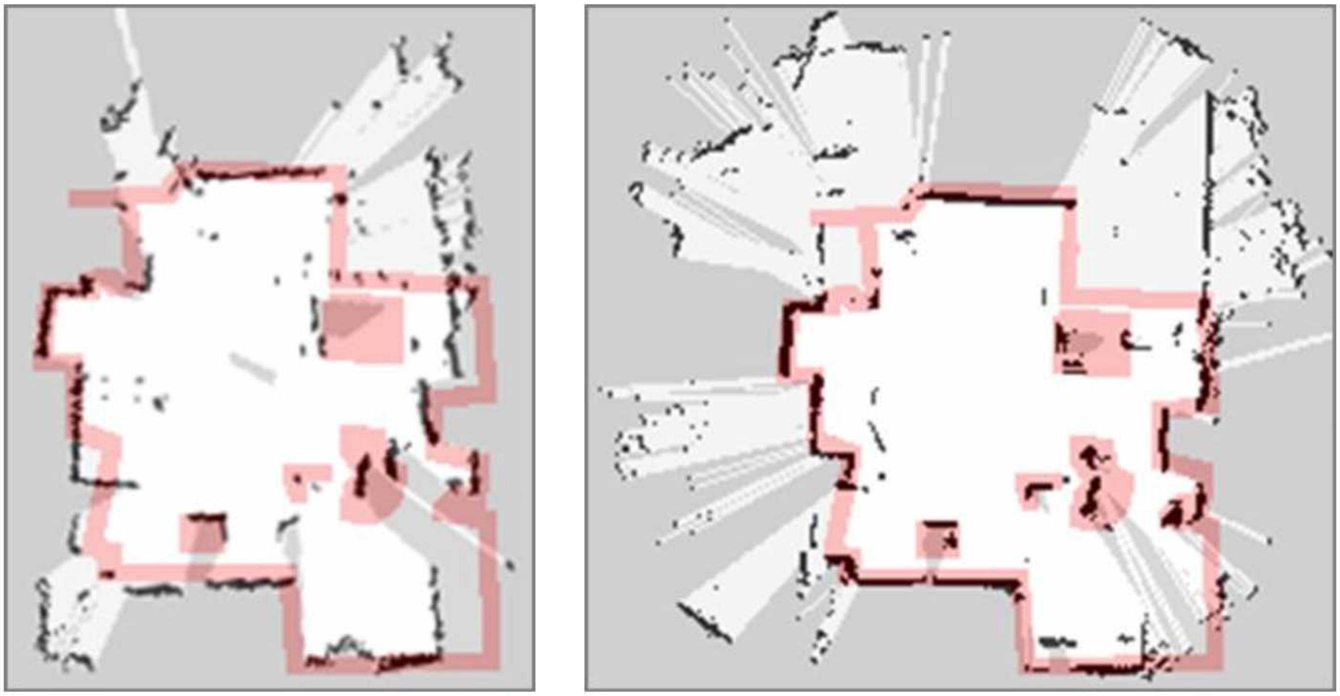


Figure 4: Map of the testing environment generated using the 2D SLAM with the Kinect data projected onto a ground plane (left) and the laser scan data (right). The overlaid red lines illustrate the actual room layout.

  - Odometry
- Transformations
  - Kinect's relative position to the robot
  - Laser's relative position to the robot

The RGBD point clouds were produced at 76.6 MB/s, which exceeded the disk writing speed. Writing to RAM was not feasible due to the limited amount of onboard memory available. To overcome this, the OpenNI Kinect driver was modified to launch only the processes necessary to capture the raw depth, colour and mono images, which were then saved to disk at 4.6 MB/s each.

The recorded data was played back and depth images were converted to point clouds using OpenNI. The resulting 3D point clouds, the corresponding 2D depth images, and the 2D laser scan data was used for the 2D and 3D SLAM.

4.3 Results 2D SLAM

Figure 4 shows the map of the testing environment generated using the 2D SLAM with the Kinect data projected onto a ground plane (left) and the laser scan data (right). The overlaid red lines illustrate the actual room layout.

The error of the estimated wall positions is considerable higher when using the Kinect sensor. In addition there are more gray regions behind obstacles, which indicates that no sensor data was available for these regions. This is due to the lower horizontal field-of-view of the Kinect (58° vs. 180°). In order to obtain the same measurements as with the laser sensor, the robot would have had to take detours behind the obstacles.

However, the results for the laser sensor show many measurement points outside the room boundaries. This is due to reflections off shiny objects, and the fact that the laser only acquires data within a plane, i.e., for obstacles such as tables only its legs were observed.

4.4 Results 3D SLAM

The 3D point cloud data from the Kinect sensor was processed with RGBD SLAM and visualised using a point cloud viewer. Figure 5 shows a model of the resulting 3D environment. Overall the room layout and major obstacles are correctly represented. However, there is a lot of noise in the data. Figure 6 shows that the point cloud for a single measurement location corresponds well with the room map. The major source of errors in the 3D reconstructions seems to be the limited range, field-of-view, and horizontal resolution of the Kinect sensor, which in many instances prevents feature matching between multiple views.

5. CONCLUSIONS AND FUTURE WORK

Our research indicates that the Kinect is a viable option for use as a sensor for mobile robotic navigation and SLAM. It offers significant advantages over conventional laser scanners, such as 3D model building, pure visual SLAM, a considerably lower price, and the inclusion of colour into the maps. However, the laser is more precise and accurate in terms of 2D mapping. The laser has a consistent resolution across its operational range, a greater depth range, a 210% wider field of view, and it can detect objects at close range.

The Kinect will be most valuable to low budget projects, robots with unreliable odometry data or no such information (e.g., unmanned aerial vehicles), and applications where a 3D map is required and precise measurements are not necessary.

In future work we want to test different 2D and 3D SLAM implementations, and quantify the effect of different param-

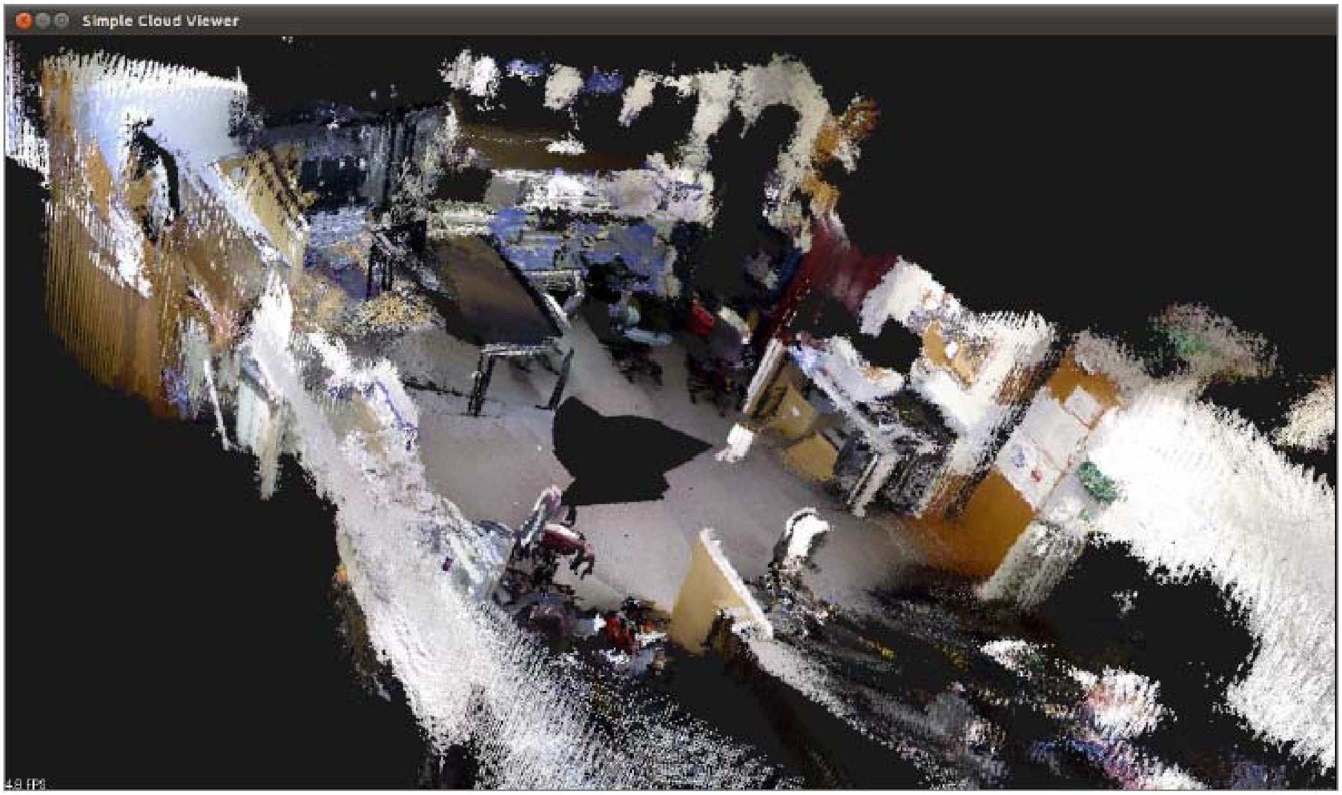


Figure 5: Testing environment reconstructed using the Kinect point cloud data and a 3D SLAM algorithm.

eters on SLAM performance. Furthermore we would like to run all algorithms in real-time without pre-recording sensor data.

6. REFERENCES

[1] Acroname Robotics. Hokuyo URG-04LX Laser. `http://www.acroname.com/robotics/parts/R283-HOKUYO-LASER1.html`, retrieved 21st September 2012.

[2] A. Bachrach, S. Prentice, R. He, P. Henry, A. S. Huang, M. Krainin, D. Maturana, D. Fox, and N. Roy. Estimation, planning and mapping for autonomous flight using an RGB-D camera in GPS-denied environments. *International Journal of Robotics Research*, 2012. (to appear), `http://groups.csail.mit.edu/rrg/papers/ijrr11-brm.pdf`.

[3] J. Blanco, J. Fernández-Madrigal, and J. Gonzalez. A new approach for large-scale localization and mapping: Hybrid metric-topological SLAM. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '07)*, pages 2061–2067, 2007.

[4] F. Bonin-Font, A. Ortiz, and G. Oliver. Visual navigation for mobile robots: A survey. *J. Intell. Robotics Syst.*, 53(3):263–296, Nov. 2008.

[5] D. S. O. Correa, D. F. Sciotti, M. G. Prado, D. O. Sales, D. F. Wolf, and F. S. Osorio. Mobile robots navigation in indoor environments using kinect sensor. In *Proceedings of the 2012 Second Brazilian Conference on Critical Embedded Systems (CBSEC '12)*, pages 36–41, Washington, DC, USA, 2012. IEEE Computer Society.

[6] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard. An evaluation of the RGB-D SLAM system. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA '12)*, May 2012.

[7] N. Engelhard, F. Endres, J. Hess, J. Sturm, and W. Burgard. Real-time 3d visual slam with a hand-held camera. In *Proc. of the RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum*, Apr. 2011.

[8] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments. *Int. J.*

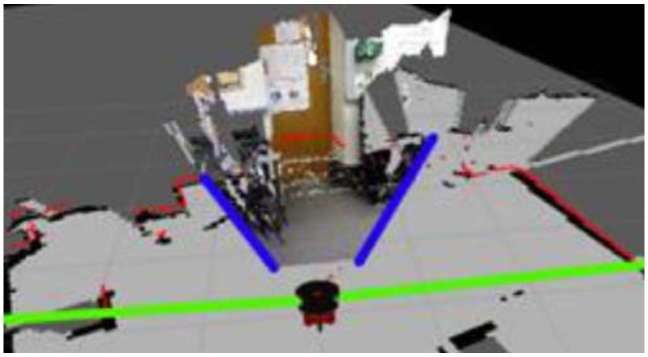


Figure 6: 3D point cloud for one measurement location overlaid onto the 2D room map.

Rob. Res., 31(5):647–663, Apr. 2012.

[9] Microsoft. The Kinect Effect - How The World is Using Kinect. `http://www.xbox.com/en-GB/kinect/kinect-effect`, retrieved 21st September 2012.

[10] J. Neira, J. D. Tardós, and J. A. Castellanos. Linear time vehicle relocation in SLAM. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '03)*, pages 427–433, 2003.

[11] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR '11)*, pages 127–136, Washington, DC, USA, 2011. IEEE Computer Society.

[12] OpenNI. Introducing OpenNI. `http://openni.org/`, retrieved 21st September 2012.

[13] OpenSLAM. GMapping. `http://openslam.org/gmapping.html`, retrieved 21st September 2012.

[14] Point Cloud Library. An open source implementation of KinectFusion. `http://pointclouds.org/news/kinectfusion-open-source.html`, retrieved 21st September 2012.

[15] Primesense Ltd. Our full 3D sensing solution. `http://www.primesense.com/en/technology/115-the-primesense-3d-sensing-solution`, retrieved 21st September 2012.

[16] ROS.org. ROS Wiki. `http://www.ros.org/wiki`, retrieved 21st September 2012.

[17] M. Ruhnke, R. Kümmerle, G. Grisetti, and W. Burgard. Highly accurate 3d surface models by sparse surface adjustment. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '12)*, pages 751–757, 2012.

[18] R. Siegwart and I. R. Nourbakhsh. *Introduction to Autonomous Mobile Robots*. MIT Press, 2004.

[19] R. Smith, M. Self, and P. Cheeseman. Estimating uncertain spatial relationships in robotics. In *Autonomous robot vehicles*, pages 167–193. Springer-Verlag, New York, NY, USA, 1990.

[20] M. Viager. Analysis of Kinect for mobile robots. Technical report, Technical University of Denmark, Lyngby, Denmark, Mar. 2011. `http://www.scribd.com/doc/56470872/Analysis-of-Kinect-for-Mobile-Robots-unofficial-Kinect-data-sheet-on-page-27`.