Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2009) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

DSAI 2012

RoboGuideDog: Guiding blind users through physical environments with laser range scanners

Javier V. Gomez[a], Frode Eika Sandnes[b]*

[a]Robotics Lab, Carlos III University of Madrid, Spain
[b]Faculty of Technology, Art and Design, Oslo and Akershus University College of Applied Sciences, Norway

Abstract

In this paper we discuss initial concepts of the development of a fully automatic guide dog system for blind users. The physical scene is scanned using a laser range device, and the three dimensional point cloud measurements are analyzed and transformed into a description of the environment that is communicated to the user via synthetic speech and/or haptic feedback allowing the user to navigate around physical space.

Keywords: guide system: laser range finder; intelligent system; haptic feedback

1. Introduction

Recent advances in technology have greatly contributed to reducing barriers for blind people and allowing them to participate in society. The research on assistive technologies is increasing and valuable contributions are also made from non-assistive areas such as the car industry where the intension is to reduce the visual load on drivers [1-5].

Screen reader technology has become a huge enabler for visually impaired users. Synthetic speech combined with technologies for transforming web-sites and documents into streams of text allows blind users access to most of the information on the Internet. This technology has matured into mainstream products such as the Apple voice over interface for tablets and smartphones with synthetic voice [6-10].

The physical world still presents itself as a challenge. Some research has gone into sensing and interpreting the physical world and transforming this into audible speech. Technology includes image recognition systems that can recognize labels ad signs through character recognition, automatic scene analysis, auralizers that transform visual images into sound, and human guidance through mobile video links where the blind user is fitted with a camera and a online operator can interpret the scene and feed the person vital information about the scene [11-15].

In spite of these technologies, navigation is still a challenge for blind people who still are reliant on guide dogs or

* Corresponding author. Tel.: +47 22 45 30 23.
E-mail address: frode-eika.sandnes@hioa.no

white canes. They are able to learn how to navigate comfortably through the city or their own home and environments that they already know well. Navigation with canes requires experience and it is limited to detecting obstacles in the immediate vicinity of the person. Even for very experienced cane users, their behavior of walking is mostly reactive since they are not able to adapt their trajectories anticipatively to unknown obstacles such as stairs and steps. Guide dogs are expensive as the training of guide dogs takes a long time and require special competences. Moreover, guide dogs are limited in the sense that they are partially color blind and unable to interpret traffic lights and signposts.

A technology based on talking GPS navigators can be helpful when navigating in unknown areas but this solution has the same restrictions as GPS, that is, relatively high error when a precise path is needed, it only works outdoors and could fail on cloudy or rainy days. Furthermore, this solution is not effective in cluttered environments since the GPS navigation systems are based on predefined maps that do not take obstacles into account.

Although visually impaired individuals' navigation is made easier, it is still hard for a blind person to navigate physical space in the same way as individuals with vision. The conclusion is that the navigation aid technologies can be improved.

Much research gas gone into image based scene analysis, yet the results are still not sufficiently good as the analysis of scenes based on simple two dimensional images is hard. Some improvements have been made with stereoscopic images or IR-cameras that enable depth cues to be incorporated in the analysis. However, the analysis of stereoscopic images is computationally intensive [16]. Moreover, IR-based approaches are reliant on an IR-light source. There are also other approaches to artificial stereo vision where depth information is acquired, but often the 3D data acquisition is slow and inaccurate [17,18].

This work takes another approach as it involves developing a model of the surrounding environment of a blind individual a using range-finder device. The model initially consists of a 3-dimensional point cloud of the environment which is processed to obtain the most relevant details which then can be converted to a synthetic speech description of the environment that is transmitted to the user.

2. Hardware setup

The prototype is based on a Hokuyo range-finger attached to a Dynamixel servomotor (see Fig. 1). This concept prototype is fixed to a table. However, future work will address how to make the scanning range finder mobile. This attachment can be done in two ways: by tilting or spinning. In the former, the inclination of the range-finder measurement plane varies respecting the floor. In this configuration the sensor view is smallest and the computational cost is higher (since more complex operations are required to convert the sensor and servomotor data into Cartesian points). In the latter, the measurement plane of the sensor is rotated over the middle axis of the measurement plane, keeping this perpendicular to the floor. In this case the area reached by the sensor is 360º with the inconvenient that the point cloud data is not homogeneous. However, this difficulty can be overcome through filtering and down sampling of the dataset. The hardware employed is a Robotis Dynamixel EX-106+ attached to a Hokuyo UTM-30LX with a custom-designed piece (available online [20]). With this configuration, we obtain a system able to measure objects at more than 30 meters every 0.25º, with a maximum error of 3 centimeters for measurements lower than 10 meters, and 5 centimeters for ranges of 30 meters. The system is able to rotate 180º, with a precision of 0.07º in the rotation angle. Since the scanning range-finder has a field of view of 270º, 360º point clouds can be obtained.

A program controls the hardware needed which ensure correct data acquisition. This control program also processes the obtained data in order to create the 3-dimensional point cloud. Fig. 2 shows the room and the corresponding three dimensional point cloud, respectively.

Fig. 1. Prototype based on Hokuyo range-finder attached to a Dynamixel servomotor.

Fig. 2. Actual room and its corresponding point cloud. The laser range finder is positioned in the front edge of the table.

Another effective, cheaper hardware option is the widely used Microsoft Kinect (or ASUS Xtion Pro Live). The refresh rate of these devices is high (24-30 frames per second) with a VGA resolution (640×480 pixels) and also depth information for each pixel. The main disadvantage of these devices is their limited field of view (horizontal 57º, vertical 43º) and their limited range of less than 7 meters (and closer to 0.5 meters there is also no data). However, these sensors provide useful RGB color information and have lower energy requirements. An interesting comparison of Kinect and a Hokuyo range finder can be found in [19].

This system can also be used outdoors. Outdoor operation is not recommended for the Kinect cameras because its infrared-based working principle interferes with sunlight and heat sources, providing erroneous data. The hardware in this case is limited to range-finders.

Therefore, the spinning range finder may be a better option because of its long range, and complete field of view. The comparatively limited refresh rate is not critical since an update each 4-5 seconds is enough as the preferred walking speed for humans is approximately 1.4 m/s.

3. Obtaining the point cloud

In this section we assume that the scanning range-finder is in spinning configuration, with a 180º rotation angle. Sensor data is gathered while the servomotor is rotating. Each one of the points obtained by the range-finder is defined by the rotation angle of the servomotor φ, the angle within the range-sensor measurement plane θ and the distance in that direction r (schema shown in fig. 3).

In order to convert this data into a 3-dimensional point cloud the following equations are applied:

$$x = r \cdot \cos(\theta) \cdot \cos(\varphi)$$
$$y = r \cdot \cos(\theta) \cdot \sin(\varphi) \qquad (1)$$
$$z = r \cdot \sin(\theta)$$

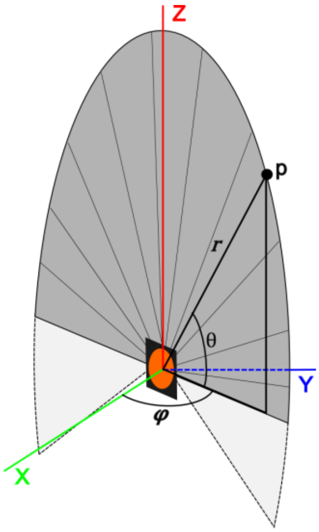


Fig. 3. Representation of the point *p* obtained by the laser and its characterization in Cartesian coordinates. The point *p* is within the measurement plane (grey).

4. Scene analysis

Based on the three dimensional point clouds one can determine what constitutes the floor of a room, the walls, the roof and objects in the room. The floor, roof and the walls are characterized by flat planes. In order to avoid the use of inertial measurement units, if it is possible to ensure that the range-finder measurement plane will be perpendicular (or nearly perpendicular) to the floor, this can be detected mathematically since this is usually going to be one of the largest planes with lower Z coordinate.

When analyzing the floor it is of particular interest to determine if there are stairs in the room or other uneven elements including steps, ramps, skirting boards, wires, etc, which may cause the user to trip and fall. Moreover, the analysis of the floor will also reveal the location of other physical objects in the scene that the user may want to avoid or use. In fact, an analysis of the floor is fundamental to providing navigational aid to the user. An example of floor analysis is shown in Fig. 5 where the floor projection is shown with obstacles in black, cleared areas for walking in white and undetermined regions in gray. The generation of probabilistic visibility maps [21] from this can be very useful to obtain a complete environment representation. These probabilistic-based maps can help also in deciding what to communicate to the user.

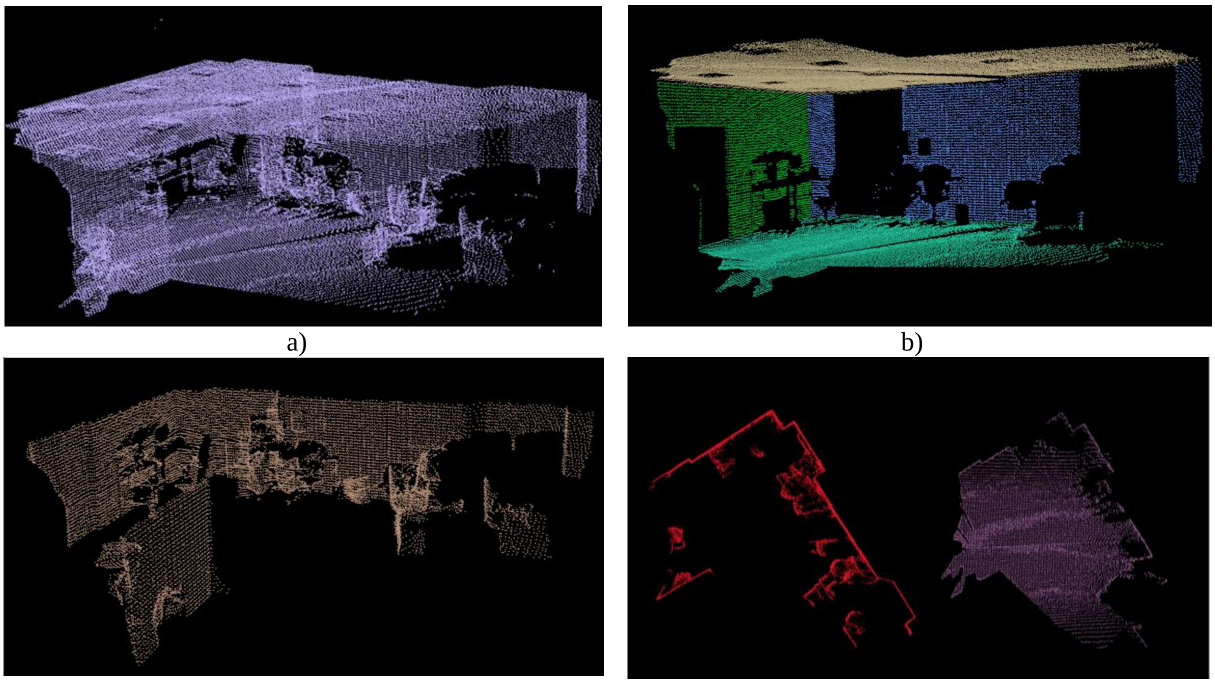
The proposed algorithm to detect the walkable zones is as follows:

1. Downsampling: Initially the point cloud comprises more points than needed and hence the processing is slower. This downsampling is carried out by using a voxel grid approximation. It consists of dividing the whole space in voxels (3-dimensional pixels) with a fixed size. All the points inside the same voxel are substituted by its centroid. By doing this with a proper voxel size, the number of points reduced to less than 30% of the original number of points without losing any important information. Moreover, noise is reduced due to the centroid operation. Another advantage of this step is that the point cloud density becomes more uniform.
2. Plane detection: main planes are detected by means of the RANSAC algorithm [22]. This algorithm detect planes by randomly choosing 3 points and evaluating the plane those 3 points generating in relation to the total number of points which are in that plane in the entire point cloud. After a fixed

number of iterations, the best plane is chosen. Thus, the biggest planes are selected first. Using RANSAC several times (usually 6 times) is enough to detect the roof (which is usually the biggest plane detected), floor and walls.

3. Floor selection: Selecting the floor is a trivial task: choosing that horizontal (or near horizontal) with lower Z value or if a gravity measurement is available, the floor will that plane which is in the gravity direction.
   In the case that there is a large ramp instead of a large floor, the floor can be confused with the ceiling. This needs to be taken into account during the development, allowing the algorithm to identify ramps as walkable if no floor is detected. Also, in the case of outdoor maps, no roof will be detected, being easier to detect the floor.
4. Projection: The last step consists of projecting all points in the point cloud which can be dangerous when navigating to the floor and to label them as obstacles. All those points with $Z_i>Z_{max}$ are not projected, since they are not reachable and thus not dangerous. Zones where the range-finder can detect floor underneath but where there is an obstacle such tables or shelves are not classified as walkable because the blind user will collide, for instance, with the top of the table. The floor points (with no obstacles) are labeled as walkable.

Fig. 4 shows the main steps of the algorithm applied to the cloud A, and Fig. 5 shows the results. The PointCloud Library (PCL) [23] is used to rapidly conduct these steps. The initial point cloud comprises 317,685 points. After downsampling which takes 148 milliseconds, the cloud is reduced to 52,392 points (less than 17% of the initial cloud size). In this case, there are 4 main planes: ceiling (139 milliseconds to process 18,427 points), floor (129 milliseconds to process 10,792 points), first wall (122 milliseconds to process 6,029 points) and the second wall (185 milliseconds to process 3,534 points). The algorithm stops when there are not any planes which contains at least 30% of the points not already included in a plane. Extracting the floor parameters is done in less than 1 millisecond and the projection takes 3 milliseconds. At the end, the algorithm takes 726 milliseconds to produce the map shown in figure 5. Figure 6 shows the results obtained in a full-room point cloud (point cloud B) and Table I lists processing statistics including computation times and number of points.



a) b) c) d)

Fig. 4. Algorithm steps for point cloud A. a) Downsampled cloud. b) Main planes extracted. c) Points to project. d) Obstacles projected (left) and walkable floor (right).

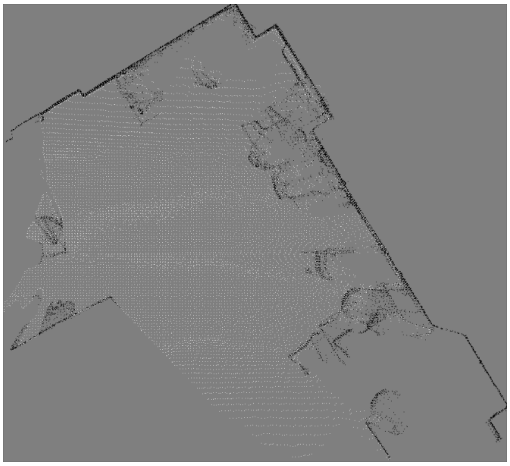


Fig. 5. Example of floor analysis based on a two-dimensional projection with obstacles (black), walkable (white) and unknown (gray) zones.

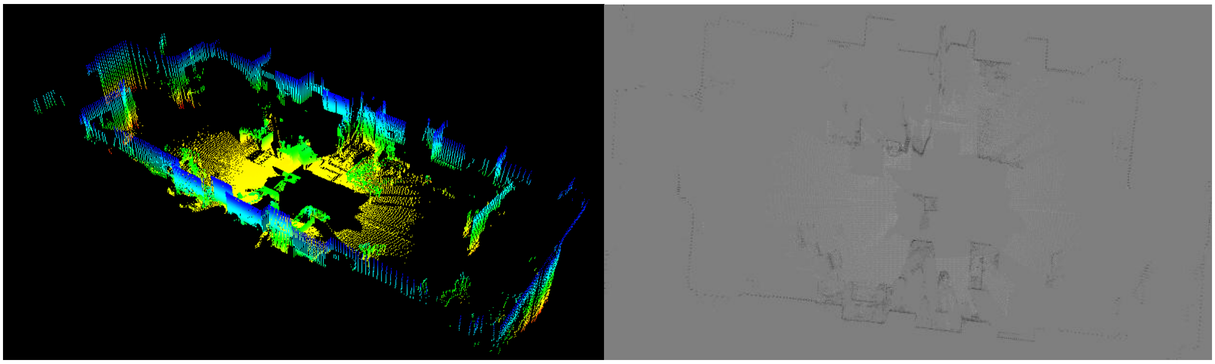


Fig. 6. Full-room point cloud (point cloud B) and the final result of the proposed algorithm. The scanning range finder was placed at a height of 75 cm above the floor (z=0.75m). If positioned higher, the algorithm will be able to identify a larger walkable zone.

TABLE 1: Number of points and computation times during the different steps of the algorithm.

Cloud	Initial points	Downsampling time (ms)	Downsampling (points/%)	Planes segmented
A	317,685	148	52,392 / 16.5%	4
B	331,868	81	88,010 / 26.5%	5

Cloud	Plane segmentation times (ms)		Plane segmentation (points)	Projection time (ms)	Total time (ms)
A	Ceiling:	139	18,427	3	726
	Floor:	129	10,792		
	1st Wall:	122	6,029		
	2nd Wall:	185	3,534		
B	Ceiling:	56	40,129	7	1,544
	Floor:	199	11,442		
	1st Wall:	452	4,243		
	2nd Wall:	397	3,906		
	3rd Wall:	352	3,248		

Walls can be further analyzed to determine the location of doors, windows, cupboards and light switches. Doors, cupboards and windows may be open or closed. A user may want to use a door to exit a room and move into another location, or may wish to open or close a particular window. Although light switches may not seem relevant to a blind user, they may still want to control the lighting for the sake of other people who are in the room or expected to enter the room as it may feel awkward to join a meeting with someone in a dark room. Moreover, switches may also control other facilities such as air-conditioners, fans, motorized blinds and marquees.

The three dimensional point clouds can also be used to analyze the objects in the scene, that is, objects that are not part of the walls, roof or the floor. At a very basic level one may identify obstacles without any attempts at determining what the obstacles are. Moreover, the obstacles can be further analyzed according to their size and overall shape to perform simple object classifications such as determining common objects such as furniture including chairs, sofas, tables, desks, lamps, vases, dustbins, etc. In this case, the Kinect RGB color data can be very helpful.

5. Synthetic speech

Information gathered about the scene may be presented through one of the other non-visual modalities such as audio. Synthetic speech is a cost effective way to convey information. First, it is crucial to warn the user of upcoming obstacles such as stairs and walls etc with feedback such as “step down ahead”, “chair ahead”. Next, the user may need to know about objects at some distance, for instance a door on the other side of the room, windows, a particular chair, etc. This information may be played exhaustively or the user may be given the control to explore the contents of the scene in a similar way to which a user explores a web page using web browser.

6. Haptic feedback

A problem with speech is that is can be disturbing if the sound is played out loud and competes with the sounds of the environment. Moreover, a user may miss an important detail if some vital information is only read aloud once.

Another strategy is to employ haptic feedback akin of how the information about the environment is communicated from a guide dog to the user via the dog leash. This may for instance be done using a force feedback joystick. If it is ok to move the joystick may give no resistance. If there is an obstacle coming up ahead the joystick may react with an opposite force towards the user and the strength of the force may be proportional to the distance. The joystick may respond with left and right motions of various degrees to direct the user left and right. Force feedback joysticks are usually large and bulky, but one can imagine a small force feedback joystick which can be held inside the hand and for instance controlled with the thumb.

Although haptic feedback may be more responsive than synthetic speech allowing the user to more quickly navigate the environment, it is harder to communicate information about the environment such as the type of object, etc. Therefore, one may employ a hybrid strategy based on both haptic feedback and synthetic speech.

7. Conclusions

This paper is reporting on the initial stages of a project with the objective of implementing a virtual guide dog. The realization of this project supposes a novel assistive system for blind individuals which can drastically improve the physical freedom of blind users. Even more, if the system incorporates an option to establish the destination point, path planning techniques can be applied in order to prove a safe, collision-free trajectory for blind people. This path is also communicated using also synthetic speech instructions.

Experimental tests show the reliability of the proposed system. Even more, the results can be improved in computation time, stopping the algorithm once the floor is detected, avoiding wall segmentation which is the most computationally expensive part of the algorithm.

References

1. A. Stanton Neville, A. Dunoyer and A. Leatherland, Detection of new in-path targets by drivers using Stop & Go Adaptive Cruise Control, Applied Ergonomics, Volume 42, Issue 4, pp. 592-601. May, 2011.
2. V. Milanes, E. Onieva, J. Perez, et al. Making transport safer: V2V-based automated emergency braking system, Transport, Volume 26, Issue 3, pp. 290-305, 2011.
3. V. Milanes, F. D. Llorca, J. Villarga, et al. Intelligent automatic overtaking system using vision for vehicle detection, Expert Systems With Applications, Volume 39, Issue 3, pp. 3362-3373, Feb 2012.
4. J. M. Flores, J. M. Armingol and A. de la Escalera, Driver drowsiness detection system under infrared illumination for an intelligent vehicle, IET Intelligent Transport Systems, Volume 5, Issue 4, pp. 241-251, Dec 2011.
5. G. C. Keller, T. Dang, H. Fritz, et al. Active pedestrian safety by automatic braking and evasive steering, IEEE Transactions on Intelligent Transport Systems, Volume 12, Issue 4, pp. 1292-1304, Dec 2011.
6. A. Stefik, C. Hundhausen and R. Patterson, An empirical investigation into the design of auditory cues to enhance computer program comprehension, International Journal of Human-Computer Studies, Volume 69 Issue 12, pp. 820-830, Dec 2011.
7. Y. Shimomura, E. T. Hvannberg and H. Hafsteinsson, Accessibility of audio and tactile interfaces for young blind people performing everyday tasks, Universal Access in the Information Society, Volume 9, Issue 4, pp. 297-310, Nov 2010.
8. H. Hochheiser and J. Lazar, Revisiting breadth vs. depth in menu structures for blind users of screen readers, Interacting With Computers, Volume 22, Issue 5, pp. 389-398, Sep 2010.
9. A. Chalamandris, S. Karabestos, P. Tsiakoulis, et al. A unit selection text-to-speech synthesis system optimized for use with screen readers, IEEE Transactions on Consumer Electronics, Volume 56, Issue 3, pp. 189-1897, Aug 2010.
10. K. Kiyota, N. Exaki, K. Itou, et al. Development of pen-based note-taking system for persons with visually disabilities, International Journal of Innovative Computing Information and Control, Volume 5, Issue 3, pp. 653-659, Mar 2009.
11. L. K. Bouman, G Abdollahian, M. Boutin, et al. A low complexity sign detection and text localization method for mobile applications, IEEE Transactions on Multimedia, Volume 13 Issue 5, pp. 922-934, Oct 2011.
12. A. Amedi, W. Stern, E. Striem, et al. A what/where visual-to-auditory sensory substitution fMRI study: Can blind and sighted hear shapes and locations in the visual cortex?, 31st European Conference on Visual Perception, Aug 2008. [Online]: http://www.artificialvision.com/
13. J. Faria, S. Lopes, H. Fernandes, et al. Electronic white cane for blind people navigation assitance, World Automation Congress (WAC), Sep 2010.
14. M. Bousbia-Salah, M. Bettayed and A. Larbi, A navigation aid for blind people, Journal of Intelligent and Robotic Systems, Volume 64, Issue 3-4, pp. 387-400, Dec 2011.
15. S. Fazli, M. D. Hajar and P. Moallen. A robust negative obstacle detection method using seed-growing and dynamic programming for visually-impaired/blind persons, Optical Review, Volume 18, Issue 6, pp. 415-422, Nov 2011.
16. B. Porr, B. Nurenberg, F. Worgotter A VLSI-compatible computer vision algorithm for stereoscopic depth analysis in real-time, International Journal of Computer Vision. Volume 49, Issue 1, pp 39-55.
17. S. Meers. A vision system for providing 3D perception of the environment via transcutaneous electro-neural stimulation, Proc. of the 8th International Conference on Information Visualization, pp 546 - 552.July 2004.
18. V. Pradeep., G. Medioni, J. Weiland. Robot vision for the visually impared. IEEE Computer Society Conference on Computer Vision and Patern Recognition Workshops (CVPRW), pp 15-22. June 2010.
19. C. Park, S. Kim, D. Kim, et al. Comparison of plane extraction performance using laser scanner and Kinect, Ubiquitous Robots and Ambient Intelligence (URAI) 2011, pp. 153-155, Nov 2011.
20. Javier V. Gómez personal Website. [Online]: http://www.sgpsproject.org/JavierVGomez/index.php?title=3D_Mapping
21. K. Haraguchi, N. Shimada, Y. Shirai, et al. Probabilistic map building considering sensor visibility for mobile robot, Intelligent Robots and Systems (IROS) 2007, pp. 4115-4120, Oct 2007.
22. M. A. Fischler and R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM, Volume 24, Issue 6, pp. 381-395, 1981
23. R. B. Rusu and S. Cousins, 3D is here: Point Cloud Library (PCL), International Conference on Robotics and Automation (ICRA) 2011, May 2011.