Interactive Perception based on Gaussian Process Classification for House-Hold objects Recognition & Sorting

Aamir Khan[1], Li Sun[2], Gerardo Aragon-Camarasa[1], J. Paul Siebert[1]

***Abstract*— We present an interactive perception model for object sorting based on Gaussian Process (GP) classification that is capable of recognizing objects categories from point cloud data. In our approach, FPFH features are extracted from point clouds to describe the local 3D shape of objects and a Bag-of-Words coding method is used to obtain an object-level vocabulary representation. Multi-class Gaussian Process classification is employed to provide and probable estimation of the identity of the object and serves a key role in the interactive perception cycle – modelling perception confidence. We show results from simulated input data on both SVM and GP based multi-class classifiers to validate the recognition accuracy of our proposed perception model. Our results demonstrate that by using a GP-based classifier, we obtain true positive classification rates of up to 80%. Our semi-autonomous object sorting experiments show that the proposed GP based interactive sorting approach outperforms random sorting by up to 30% when applied to scenes comprising configurations of household objects.**

I. INTRODUCTION

It is essential for service robots to have the ability to recognise objects in their immediate vicinity when working in dynamically evolving human environments. Ideally, these robots should be capable of detecting and classifying objects within their environment and then interacting with these objects without need for supervision. In this paper, we present an interactive perception system which is able to sort everyday household objects into their respective categories through direct visual observation (typically when objects are not occluded), and then by means of active object manipulation where objects are occluded or "difficult to recognise". Our proposed framework does not require prior knowledge about the environment or scene. We also present a visually assisted object sorting system which is capable of segmenting a set of household objects lying directly on the robot's workspace table and categorising those objects into their respective object classes (ie.g. juices bottles, mugs, etc.). The system has been pre-trained on a subset of these object instances, while a proportion of the objects we investigated have not been used to pre-train the system. A high-level model of our classification model is presented in Figure 1.

[1]Aamir Khan is with the School of Computing Science, University of Glasgow, G12 8QQ, UK a.khan.4.@research.gla.ac.uk

[2]Li Sun is with the Intelligent Robotics Lab, University of Birmingham, B15 2TT, UK lisunsir@gmail.com

[1]Gerardo Aragon-Camarasa is with the School of Computing Science, University of Glasgow, G12 8QQ, UK gerardo.aragoncamarasa@glasgow.ac.uk

[1]J. Paul Siebert is with the School of Computing Science, University of Glasgow, G12 8QQ, UK paul.siebert@glasgow.ac.uk

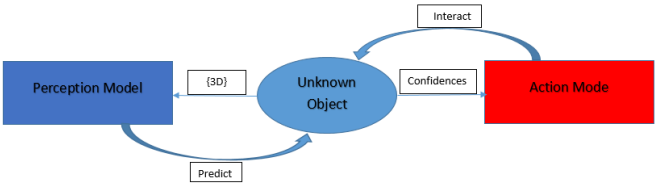

Fig. 1. Our Gaussian Process Classification based Interactive Perception Model (GP-IPM).

Our system is portable, invariant to 6 DOF pose changes and operates close to real-time. The pipeline comprises the following: object segmentation, visual representation, classification, semantic visualization and finally in our current experiments, a human operator responsible for removing the object from the scene following its correct classification by the system. Our Gaussian Process classification based Interactive Perception Model (GP-IPM) has been cross-validated by comparing the categorisations results obtained when using an SVM multi-class classifier.

The operating scenario we have adopted comprises a visual search for an object, such as a bottle or pen, potentially partially occluded. If an initial search fails to locate the object, each object is re-positioned with a prescribed minimum level of confidence. Based on the above scenario, we present results obtained from experiments where the objects are sorted by the combination of class predictions obtained from the GP based multi-class classifier with highest probability. The authors claim that the above GP classification based interactive perception model makes the following contributions to the state-of-the-art in object sorting:

1) We present the first example of research to adapt non-parametric multi-class probabilistic classification (via Gaussian Processes) to the house hold object recognition problem.
2) We demonstrate that the proposed GP-IPM approach applied to a semi autonomous sorting task yields substantially improved performance over non-interactive alternatives.

This paper is organised as follows: Section 2 presents a literature review of current interactive perception technologies. Section 3 presents our GP classification based interactive perception model. Finally, Section 4 and Section 5 comprise preliminary experimental validation of the system and concluding remarks of this paper, respectively.

II. RELATED WORK

A variety of techniques for scene understanding aided by interaction, also known as interactive perception techniques, are directed at improving the performance of segmentation through interaction [7], [8], [9]. Recently, interactive object sorting/recognition systems have been reported to achieve viable levels of performance [1], [2]. Interactive perception models have also been adopted for tasks such as learning, manipulation, etc. For example, in [10], active curiosity-driven manipulation is used when learning the appearance of objects. Chang[4] presents a framework that can singulate objects contained in a pile through multiple interactions with the pile by *pushing* objects using a robot manipulator. In this framework, a decision module distinguishes between a single-item pile and a multiple-item pile, while a perception module combined with pushing actions accumulates evidence of items over multiple interactions to reduce the number of grasp errors.

Motivated by human infant development, a perceptual system is proposed by Lyubova[5] to allow a robot to learn about physical entities in its surrounding workspace in two stages. In the first stage, a human partner moves a workspace element and the robot learns the object's appearance of this moving element. In a second stage, the robot interacts with the objects to learn its appearance. Although, this system requires only very limited prior knowledge, it has to have the knowledge of parts of its own body, parts of a human partner, and what constitutes a manipulable object.

Katz[6] reports an approach to interactive perception that by focussing on object function, rather than object appearance, is able to manipulate unknown objects which possess inherent degrees of freedom, such as scissors, pliers, but also door handles, drawers, etc., without requiring a priori model. This interactive perception system is able to manipulate articulated objects successfully by acquiring a model of the objects kinematic structure.

Gupta et al. [1] report an investigation on sorting objects in clutter. In particular, small objects (Duplo bricks) are sorted on a tabletop by segmenting the scene into regions: uncluttered, cluttered and pile. For the uncluttered case, every object is picked and placed in its respective bin to clear the table. In the cluttered case, objects are first separated and then picked and placed in bins. For the piled scene, the pile is knocked over by the manipulator in order to decompose it so that the objects lie directly on the table. However, this framework only works for small objects of single and homogeneous colour.

Krishnanend et al. [11] presents an interactive perception based approach that addresses perception uncertainty in order to reduce failure rates in a robotic bin-picking task. Human intervention is required when the uncertainty in the part detection leads to perception failure. The automated perception system provides the part match and this approach estimates the confidence in the part match. Thereafter, a sensor-less fine-positioning planner is used to correct the part placement errors.

However, current techniques are yet to compromise over general object manipulation tasks such as an autonomous sorting system involving interaction for general house hold objects. Our proposed GP classification based interactive perception model is an initial attempt to demonstrate that the proposed GP-IPM approach applied towards autonomous sorting task for house hold objects yields substantially improved performance over non-interactive alternatives.

III. METHODOLOGY

Our approach consists of interleaving five stages: segmentation, visual representation, category classification, semantic visualization, and a human user clearing the object into the object's class bin. The proposed visually assisted vision system for objects sorting is implemented and an initial evaluation is presented from cross validation simulation and demonstration on a dual-arm industrial robot. The dual arm robot is equipped with stereo cameras, a pair of Xtion pro camera and tactile sensing grippers. Our system utilizes the depth images from one of the Xtion pro cameras mounted on each arm of the robot which serves as input and triggers the rest of the pipeline of the system. The visual system pins out the category of the objects lying on the table directly and the system recognizes the category of the object and produce a color coded semantic representation of the scnece contents. We point out that our visual system is implemented using the Point Cloud Library(PCL) [17] and integrated into ROS. In the following sections, we briefly describe the components of our pipeline.

A. Segmentation

An input point cloud is first filtered for points within a defined range by a pass through filter. The latter filter consists of filtering points that are far and outside the table. Initially, segmentation is carried out by detecting the table plane solely from the depth information. The operating table is segmented by using Random Consensus Sampling (RANSAC) [18] which estimates all the points in a point cloud for a model plane. Then, objects lying on the table are segmented by computing a convex hull from the plane coefficients, it will extrude it a certain height to create a prism, and give back all points that are lying inside.

B. Visual Representation

Local statistical 3D shape features are extracted for each of the object obtained from segmentation. These features are then encoded to create a dictionary containing codes for all the trained objects. For each extracted object, a visual representation is acquired by means of the Bag-Of-Words model. For low level features, local 3D curvature features are computed by means of the Fast Point Feature Histogram(FPFH) [3]. For each query point p_q and its neighbours, a set of tuples α, ϕ, θ are computed that results in Simplified Point Feature Histograms (SPFH). Subsequently, for each point, its k neighbours are re-evaluated, and the neighbouring SPFH features are used to

weight the final histogram p_q (called FPFH) as follows:

$$FPFH(P_q) = SPFH(P_q)\frac{1}{k}\sum_{i=1}^{k}\frac{1}{w_k}.SPFH(P_k)$$

At the training stage, after features are extracted for all the views and for all the objects, we perform K-means clustering over all the vectors to obtain feature codes from the centres of the computed clusters. We set 256 K clusters as the size of our codebook. A putative visual representation for each object is then obtained by comparing all the features in the codebook generated by finding the minimum euclidean distance, max pooling and l2 normalization, resulting in 256 dimensional long vectors for each pose of an object. At test phase, features extracted for all the objects are encoded using the codebook generated, and fed to the classifier.

C. Classification

SVM-like classifiers achieve outstanding classification performance on small-scale datasets due to its max-margin and kernel function mechanisms, whereas for the multi-class classification problem, they cannot generate true predictive probabilities through one-vs-one or one-vs-all voting. In order to model the distribution of predictive probabilities among all object categories, multi-class Gaussian Process Classification [12] is employed.

In our approach, RBF kernel is used as the kernel function, the posterior of the Gaussian Process is estimated by straight forward Laplace approximation. The RBF kernel we use has two hyper-parameters:

$$kernel_(x_1, x_2) = \alpha^2 \exp^{-\frac{1}{2}(x_1-x_2)^T diag(\frac{1}{\beta^2},\ldots,\frac{1}{\beta^2})(x_1-x_2)}, \quad (1)$$

where x_1 and x_2 are two examples, and α and β are the two hyper-parameters controlling the scale and shape of the exponential function. We optimize those hyper-parameters by the log marginal likelihood maximization. A more detailed description can be found in our previous work [13].

A pre-trained Gaussian process based classifier is used to work on the resultant encoded visual representation for each object from the stage above in order to predict the category of the object along with probability confidences. These confidences are then used to make a decision of the identity of the object. The visual system described in this paper has been tested with two multi-class classifiers, Support Vector Machines and Multi-class Gaussian Process models [12].

D. Semantic Visualization

This provides a semantic representation of the results from the classifier by colour coding each object according to their respective categories. A colour coded visual representation is used for objects via the RVIZ tool available in ROS (http://wiki.ros.org/rviz), where each colour corresponds to a category and probabilities are used to either place the object in the respective bin or the object with lower confidence probabilities are further investigated by placing the object in a different uncluttered location on the operating table.

Algorithm 1 Interactive Objects Sorting

```
procedure INTERACTIVE OBJECTS SORTING
    segmentation into objects ← point cloud
    visual representation ← point cloud representing each object
    classification:confidences,labels ← visual representation
top:
    if no object on the table then return false
loop:
    if confidence(i) > threshold then
        color code each object according to its class.
        pick and place object into bin with highest confidence.
        goto loop.
    goto top.
```

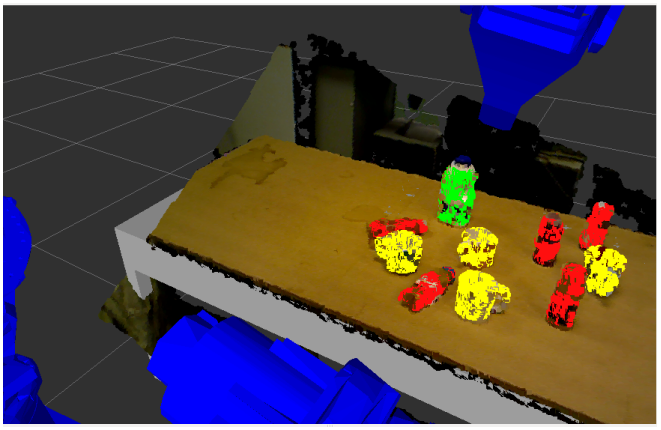

Fig. 2. Robot recognising object categories according to colours.

E. Human Operator

As a final step, the object is removed by a human operator from the table and placed into its respective bin by observing the colour coded semantic representation.

A pseudo code Algorithm of our proposed method is show in Algorithm 1.

IV. Experiments

These experiments are designed to demonstrate the validity of our Gaussian Process based perception model, capable of improving the recognition rate with aid of the interaction with objects in order to carry out the task of semi autonomous object sorting. We carried out experiments for five house hold object categories i.e Juice Bottles, Milk Packs, Bowls, Mugs and Juice Boxes. We arranged scenes based on combinations of known and unknown object instances, of 5 different object classes. Objects were placed in arbitrary poses and locations as shown in figure 2. In figure 2, red colour is for juice bottles, green for milk boxes, blue for bowls and yellow for mugs. Our training dataset consists of the above 5 object classes – we created this dataset by capturing point clouds of each object at angular intervals of approximately 20 degrees.

In this section, our GP based perception model is evaluated for objects sorting into their respective bins. Note that the objects are placed directly on the table. An input point cloud is obtained from the ASUS Xtion pro camera from one view. The point cloud is subsequently passed on to our pipeline.

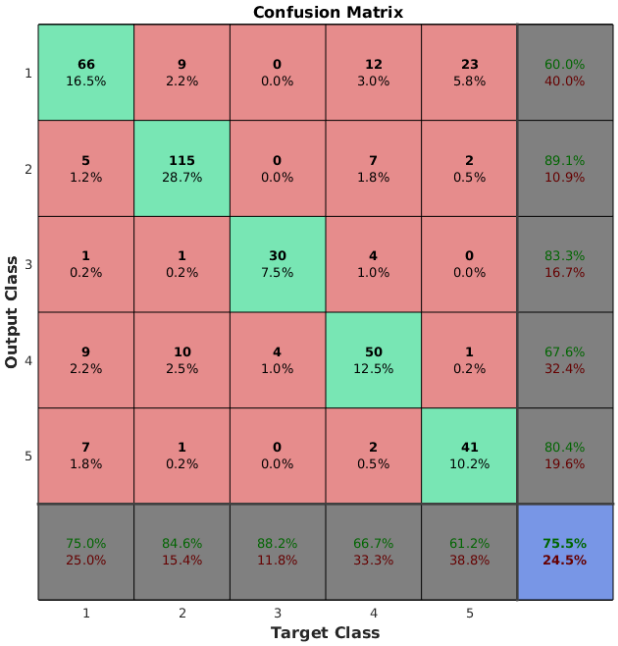

Fig. 3. Confusion matrix shown obtained from multi-class svm classifier.

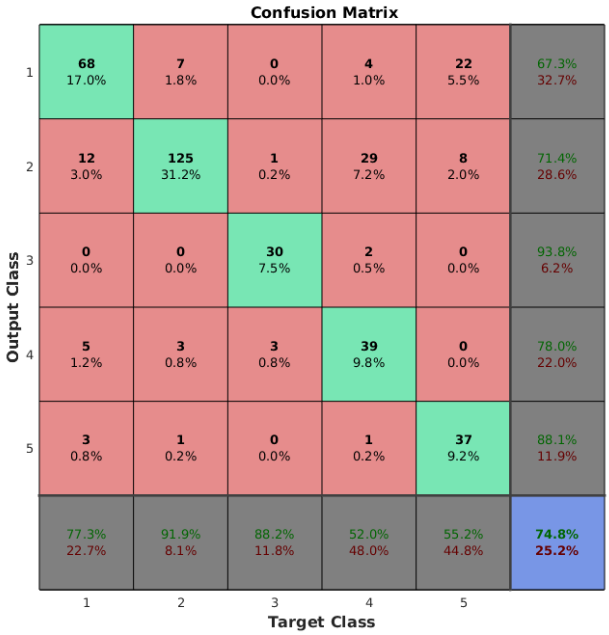

Fig. 4. Confusion matrix shown obtained from multi-class gp classifier for 5 object categories.

As stated before, our experiments consisted of evaluating a SVM multi-class classifier and a GP based classifier as shown in figure 3 and figure 4, respectively. The diagonal of each confusion matrix, coloured in green, shows the true positives and the bottom row gives the recognition rate for each category. Figure 5 shows the prediction confidences, where each vertical line corresponds to an object category i.e red, blue, black, green, yellow, respectively, different colour marks outside its group refer to incorrect predictions.

For the SVM multi-class classifier, we obtained an average recognition rate of 75% and for GP based multi-class classifier, 79%. We must point out that the bowl's category has the lowest recognition rate which in turns affects the overall recognition rate. This is due to the bowl's intrinsic shape as it contains less visual information as well as consisting of a highly reflective surface, to which depth sensing is very sensitive (i.e. limitation of the ASUS Xtion pro camera).

To demonstrate the usability of our GP based interactive perception model for improving the recognition rate towards the task of object sorting, we constructed two scenes; (1) without any occlusion and simple objects, and (2) with occlusion among objects and challenging objects such as bowls. Also, for each scene, we consider two types of sorting. First random sorting is considered, where the output class of the classifier along with the highest object is first removed and placed in to the respective bin. Second, by using GP based interactive perception model, we remove the object from the table based on confidence probability. Figure 2 show the scene without any occlusion and hence the output of our system is shown with the help of colours representing categories. Figure 6 shows the second scene described above.

After all the objects have successfully been recognised and classified accordingly, objects were manually placed into

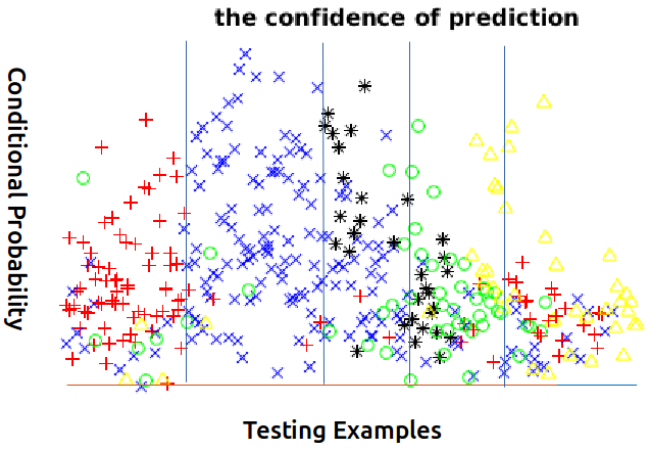

Fig. 5. The classification performance under different confidences for 5 object categories.

their respective bin, Table I shows the success rate for the objects placed into the respective bin guided by our proposed GP-IPM. For complex scenes, containing objects having challenging shapes and occlusion, our proposed method outperforms random sorting by a factor of 30%.

V. Conclusion and Future Work

We have described an interactive perception based object sorting system, incorporating depth sensing, 3D feature extraction, object representation and classification. Our visual perception system achieves real-time object category identification with associated perception confidence estimates, which serves to establish whether an observation is sufficient to make a classification or to determine if it is necessary

Fig. 6. Scene containing complex objects and occlusion.

TABLE I

TABLE SHOWING TRUE POSITIVE FOR RANDOM AND GP-IPM

Scene Complexity	Method of Sorting	True Positives					Total
		Juice Bottles	Milk Cartons	Bowls	Mugs	Juice Boxes	
Uncluttered	Random	5/5	1/1	-	4/4	-	9/9
	GP-IPM	5/5	1/1	-	4/4	-	9/9
Cluttered	Random	2/4	1/2	2/3	2/3	1/2	8/14
							57%
	GP-IPM	4/4	2/2	2/3	3/3	1/2	12/14
							86%

to interact with the object to achieve a sufficiently reliable classification of the instance contained in the observation.

We have also established a dataset of household objects used in the evaluation in this paper, and the experimental results show that our proposed perception pipeline achieves 80% recognition accuracy for 5 object categories. As the autonomous robot manipulation phase is currently work in progress, this paper focused on the visual perception components of our system and we presented sorting results obtained using a human operator to manipulate the objects. From these semi-autonomous sorting experiments, the proposed Gaussian Process based interactive sorting system outperformed random sorting by upto 30% in terms of sorting accuracy. We also observed that our interactive perception strategy not only mitigates segmentation failures prevalent in single-shot sorting, but it also increased perception performance in terms of recognition rate, thereby facilitating the sorting decision.

Our future work will integrate the proposed visual perception approach with autonomous manipulation skills onto our robot testbed, which has been successfully achieved several visually-guided manipulations tasks [14], [13]. Eventually, a fully autonomous category-based visually-guided household object sorting is expected to be completed. We propose to expand our dataset by including more object categories and by increasing the variety of object instances it contains. We also propose to extend the system to segment objects in order to learn their appearance by interaction and to capture their visual representation from different views. In addition, we propose to integrate a high-level reasoning engine, based on a natural language processing tool, into the system in order to direct the robot to carry out an object sorting task or direct the system to learn the appearance of a specific object example.

REFERENCES

[1] M. Gupta and J. Muller and G. S. Sukhatme, IEEE Transactions on Automation Science and Engineering, Using Manipulation Primitives for Object Sorting in Cluttered Environments, 2015, pp 608-614.

[2] Guglielmo Gemignani, Roberto Capobianco, Emanuele Bastianelli, Domenico Daniele Bloisi, Luca Iocchi and Daniele Nardi, Living with robots: Interactive environmental knowledge acquisition, Robotics and Autonomous Systems, 2016.

[3] R. B. Rusu, N. Blodow and M. Beetz, Robotics and Automation, 2009. ICRA '09. IEEE International Conference on, Fast Point Feature Histograms (FPFH) for 3D registration, 2009, pp 3212-3217.

[4] L. Chang and J. R. Smith and D. Fox, IEEE International Conference on Robotics and Automation (ICRA), 2012 , Interactive singulation of objects from a pile, 2012, pp 3875-3882.

[5] Lyubova, Natalia, Ivaldi, Serena and Filliat, David, From passive to interactive object learning and recognition through self-identification on a humanoid robot, Autonomous Robots, 2016.

[6] D. Katz and O. Brock, Manipulating articulated objects with interactive perception, IEEE International Conference on Robotics and Automation, 2008. ICRA 2008.

[7] P. Fitzpatrick, First contact: an active vision approach to segmentation, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003.

[8] Li, Wai, Ho and Kleeman, Lindsay, Segmentation and modeling of visually symmetric objects by robot actions, The International Journal of Robotics Research, pp 1124-1142, 2011.

[9] D. Schiebener and A. Ude and J. Morimoto and T. Asfour and R. Dillmann, 11th IEEE-RAS International Conference on Humanoid Robots (Humanoids), 2011.

[10] S. M. Nguyen and S. Ivaldi, N. Lyubova, A. Droniou, D. Gerardeaux-Viret, D. Filliat, V. Padois, O. Sigaud and P. Y. Oudeyer, Learning to recognize objects through curiosity-driven manipulation with the iCub humanoid robot, 2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL).

[11] Krishnanand N. Kaipa, Akshaya S. Kankanhalli-Nagendra, Nithyananda B. Kumbla, Shaurya Shriyam, Srudeep Somnaath Thevendria-Karthic, Jeremy A. Marvel , Satyandr K. Gupta, Addressing perception uncertainty induced failure modes in robotic bin-picking, Robotics and Computer-Integrated Manufacturing, 2016

[12] Rasmussen, Carl Edward, Gaussian processes for machine learning, 2006.

[13] Sun, Li and Rogers, Simon and Aragon-Camarasa, Gerardo and Siebert, J Paul, Recognising the clothing categories from free-configuration using Gaussian-Process-based interactive perception, 2016 IEEE International Conference on Robotics and Automation (ICRA).

[14] Li Sun, and Gerardo, Aragon-Camarasa, and Simon, Rogers and J. Paul Siebert, Accurate Garment Surface Analysis using an Active Stereo Robot Head with Application to Dual-Arm Flattening, 2015 IEEE International Conference on Robotics and Automation (ICRA).

[15] Sun, Li and Camarasa, Gerardo Aragon and Khan, Aamir and Rogers, Simon and Siebert, Paul, A Precise Method for Cloth Configuration Parsing Applied to Single-arm Flattening, International Journal of Advanced Robotic Systems, 2016.

[16] Sun, Li and Aragon-Camarasa, Gerarado and Cockshott, Paul and Rogers, Simon and Paul, J, A Heuristic-Based Approach for Flattening Wrinkled Clothes, 2013 Conference Towards Autonomous Robotic Systems.

[17] Point Cloud Library `http://http://pointclouds.org/`, note = Accessed: 2016-07-15

[18] Fischler, Martin A. and Bolles, Robert C., Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, ACM Comm., 1981,