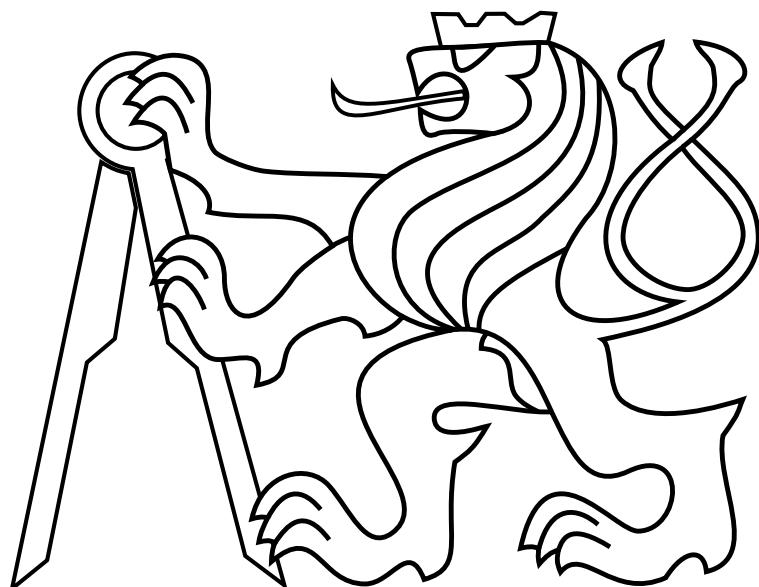


CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Electrical Engineering

MASTER'S THESIS



Zdeněk Rozsypálek

Brick Detection for MBZIRC Competition

Department of Control Engineering

Thesis supervisor: RNDr. Petr Štěpán Ph.D.

Prohlášení autora práce

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne
.....
podpis autora práce

Author statement for undergraduate thesis

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

Prague, date
.....
signature

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Rozsypálek** Jméno: **Zdeněk** Osobní číslo: **457216**
Fakulta/ústav: **Fakulta elektrotechnická**
Zadávající katedra/ústav: **Katedra řídicí techniky**
Studijní program: **Kybernetika a robotika**
Studijní obor: **Kybernetika a robotika**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Detekce cihel pro soutěž MBZIRC

Název diplomové práce anglicky:

Brick detection for MBZIRC competition

Pokyny pro vypracování:

Seznam doporučené literatury:

- [1] Himmelsbach, Michael, et al. "LIDAR-based 3D object perception." Proceedings of 1st international workshop on cognition for technical systems. Vol. 1. 2008.
- [2] Dou M., Guan L., Frahm JM., Fuchs H. (2013) Exploring High-Level Plane Primitives for Indoor 3D Reconstruction with a Hand-held RGB-D Camera. In: Park JL., Kim J. (eds) Computer Vision - ACCV 2012 Workshops. ACCV 2012. Lecture Notes in Computer Science, vol 7729. Springer, Berlin, Heidelberg
- [3] Ma, L., Kerl, C., Stückler, J., & Cremers, D. (2016, May). CPA-SLAM: Consistent plane-model alignment for direct RGB-D SLAM. In 2016 IEEE International Conference on Robotics and Automation (ICRA) (pp. 1285-1291). IEEE.

Jméno a pracoviště vedoucí(ho) diplomové práce:

RNDr. Petr Štěpán, Ph.D., Multirobotické systémy FEL

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **14.01.2020**

Termín odevzdání diplomové práce: _____

Platnost zadání diplomové práce:

do konce letního semestru 2020/2021

RNDr. Petr Štěpán, Ph.D.
podpis vedoucí(ho) práce

prof. Ing. Michael Šebek, DrSc.
podpis vedoucí(ho) ústavu/katedry

prof. Mgr. Petr Páta, Ph.D.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací.
Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

Datum převzetí zadání

Podpis studenta

Acknowledgements

!!!OUTDATED!!! I would like to express my appreciation to Ing. Tomáš Petříček for his valuable and constructive suggestions during the planning and development of this thesis. I would also like to thank to the Department of Cybernetics of the Czech Technical University and to Michal Němec for the provided hardware. Finally, I wish to thank my family for support throughout my study.

Abstract

BLAH BLAH

Abstrakt

BLAH BLAH

Keywords: Keywords

Contents

1	Introduction	1
2	MBZIRC Contest	1
2.1	Second Challenge	1
3	Equipment	4
3.1	Velodyne VLP-16	5
4	STH ABOUT ROS???	6
5	Lidar data processing	7
6	EM algorithm	8
6.1	Maximization	8
6.2	Expectation	9
6.3	Algorithm	9
7	RANSAC	9
7.1	Tentative Correspondences	10
8	Lidar to camera registration	11
9	Global model and transformations	11
9.1	Symbolic map	12
10	Lidar detection range analysis	13
11	Program pipeline	13
11.1	Line segmentation	13
11.2	Pile detection	13
11.3	Pattern fitting	13
12	Conclusion	14
12.1	Future work	14

Appendix A CD Content 16

Appendix B List of abbreviations 17

List of Figures

1	Bricks definition	2
2	Initial brick layout	3
3	Brick destinations	4
4	UGV robot setup	5

1 Introduction

Multirobot, autonomous, motivation, BLAH BLAH BLAH

2 MBZIRC Contest

The contest took place in March in Abu Dhabi. Whole competition consisted of three challenges and the grand challenge which connected all challenges together. The first challenge was the only one which was focused solely on UAVs. The goal of the first challenge was to pop multiple big color balloons and to catch small ball carried by the organizer's drone. The other two challenges was designed for both UGVs and UAVs. The second challenge was about building a wall using the robots. Multiple polystyrene bricks were placed in the arena and the robots should have move these bricks to the destination area and stack them on top of each other to build the wall. Lastly the third challenge was to extinguish fire on the surface of the building model. This task was motivated by inability of firefighters to extinguish fire inside high-rise buildings in UAE. UAVs and UGVs carried tanks full of water and squirt it into the fire dummy. Every team had three rehearsals in before the contest and then two competition attempts for each one of three challenges. Only the best teams from all three challenges was nominated into the grand challenge which was limited to just one attempt.

2.1 Second Challenge

This thesis is focused on the second challenge and more specifically on the ground robot section of the second challenge, so that we provide more detailed description of this challenge. Each team is given thirty minutes to explore the arena (40×60 meters), localize all interest areas and build the wall. There are four types of the bricks with different colors. All bricks must be very light to enable the UAVs to pick them up. The dimension and colors of the bricks can be seen in the figure 1. Team obtains points for every placed brick. Bricks with different colors are rewarded by the different number of points. Placing bigger bricks means higher rewards. In addition the UAV bricks are rewarded by twice as many points as the UGV bricks.

INTRODUCTION

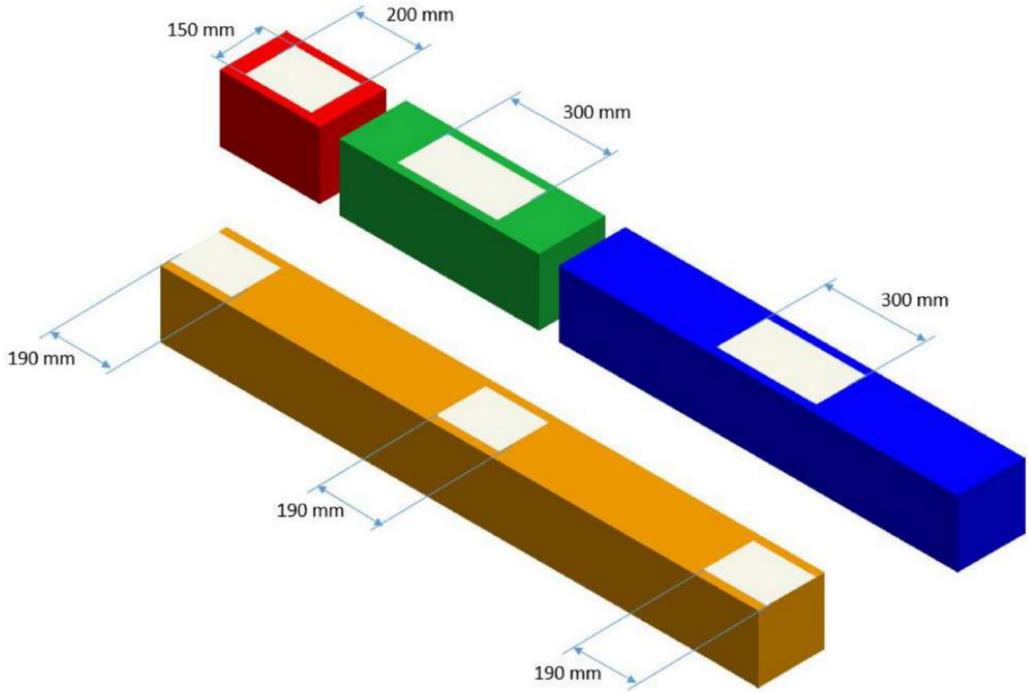


Figure 1: Colors and dimensions of the bricks provided by the organizer.

Each brick has thin metal plate on top of it, so that the robots are able to pick them up using electromagnets. At the beginning of challenge are all bricks placed in beforehand unknown position. Initial position of bricks is unknown but there is a predefined pattern in which are the bricks put together. There are different patterns for the UGV piles of bricks and the UAV piles. The UGV bricks are stacked into the multiple height levels whereas the UAV bricks are stacked into the width and all are put on the ground. Due to the low weight of the bricks it is necessary to put UAV bricks into the rails, otherwise the bricks could be easily blown away by the propellers of the drones. Since the UAV bricks are all on the ground level (in purely horizontal pattern), it is much easier to detect them with the UAV bottom camera than using the lidar. That is why we are further concerned only about UGV bricks. These bricks are stacked in the positions displayed in the figure 2.

INTRODUCTION

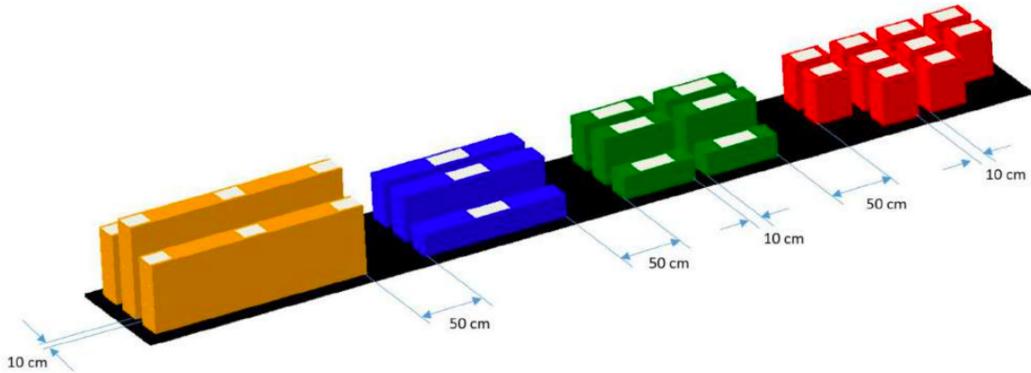
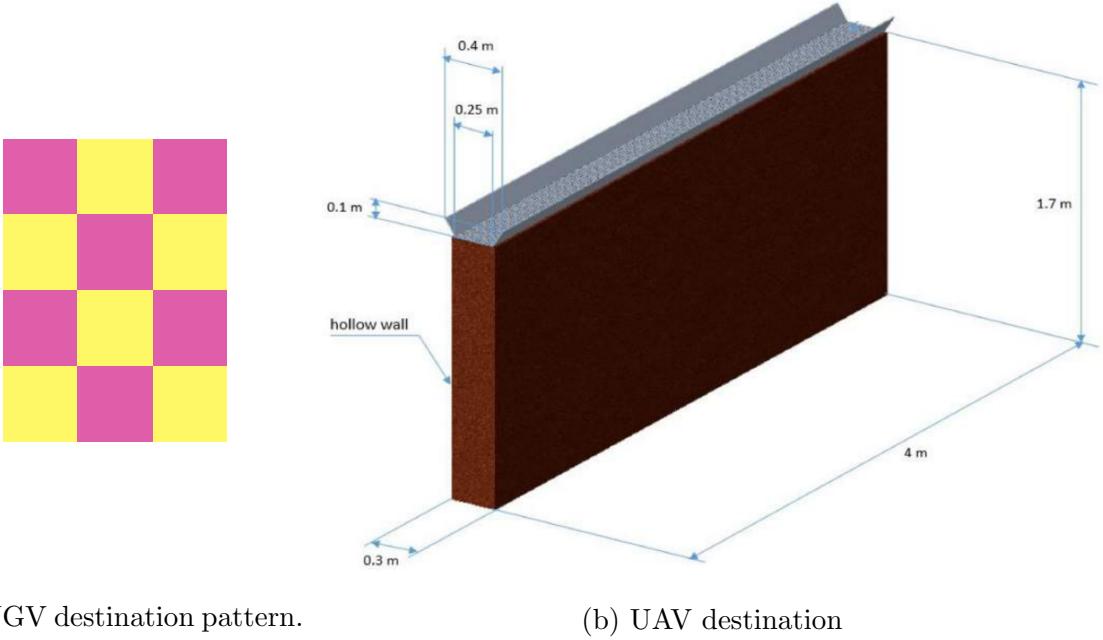


Figure 2: The positions of the bricks at the beginning of the second challenge.

Another objects of interest are destinations where the bricks should be placed. The robots must during the exploration look for them too. UGV bricks destination is marked by checker pattern. Detecting/pattern was very challenging because exact shape was not known till second rehearsal. The final form of the pattern is shown in the figure 3a. Although we are not concerned about the UAV bricks, the destination of the UAV bricks is a vertical object, so it is much easier to detect it from the ground using the lidar. The UAV brick destination is basically a wall as can be seen in the picture 3b. Bricks should be placed on top of this wall. The metal plate on top of each brick shifts the center of mass to the top and make the brick very susceptible to rolling. That is why are on the top of UAV destination auxiliary handles to help place the bricks properly. At the beginning of the challenge is each team given the instructions which describe how the wall should look like at the end. When the built wall does not fit the instructions, the team gets penalty and gains less points for inaccurately placed bricks.

INTRODUCTION



(a) UGV destination pattern.

(b) UAV destination

Figure 3: Description of target places for UAVs and UGVs. Each square in the UGV pattern has 10cm. Pattern consists of two 4×0.4 meter segments which are connected into the L shape. Whole UAV destination consists of five similar segments arranged into the M shape with right angles.

3 Equipment

For the sake of completeness is necessary to describe what exact equipment was available. We used **Clearpath Husky A200** which is wheeled robot designed for outside robotics. The robot is equipped with many additional devices. As a computer running all code controlling the robot is used **Intel NUC**. To manipulate the bricks we mounted **Kinova robotic arm** on top of the Husky robot. Two **12V electromagnets** are attached to the end-effector to enable the arm to grip the bricks. It would be very hard to grip the bricks without any feedback loop to the hand. For visual servoing and proper gripping we placed **Intel Realsense** camera close to the end of the arm. It is also possible to obtain feedback from electromagnets thanks to hall effect sensors and decide whether the brick is gripped correctly. For the localization, collision avoidance and detection is used **Velodyne VLP-16** lidar sensor. Lastly for the moving the bricks around the arena we created a handmade cargo area which can contain up to six bricks and attached it to the rear bumper. It was not possible to carry more bricks mainly because of restrictions on robots size and also due to the limited range of Kinova arm. Whole setup is captured in the figure 4.

INTRODUCTION

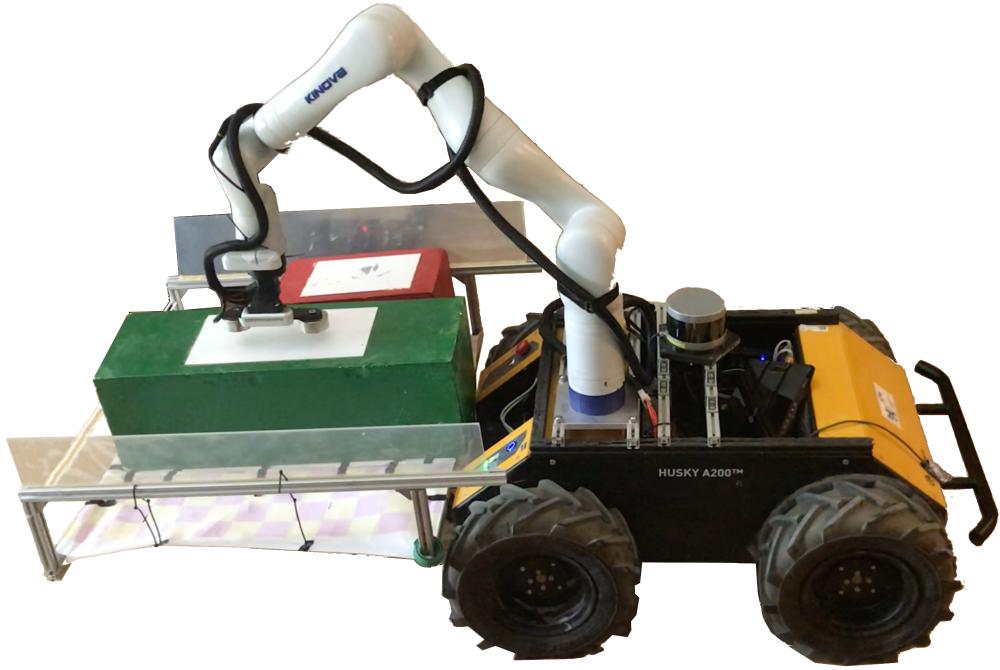


Figure 4: Clearpath Husky A200 adjusted for the second challenge.

3.1 Velodyne VLP-16

This thesis deals mainly with lidar data so following subsection will provide more detailed description of lidar sensor. Inside the VLP-16 puck is rotating class one infrared laser which measures the distance using the time of flight principle. Lidar is powered by 12V power supply and the data are transferred via UDP packets over the ethernet. Parameters of the Velodyne lidar are listed in the table 1.

Layers	16
Range (m)	100
Vertical FOV (°)	±20
Vertical resolution (°)	2
Horizontal FOV (°)	360
Horizontal resolution (°)	0.1
Frequency (Hz)	5
Precision (m)	±0.03

Table 1: Parameters of VLP-16 lidar sensor.

INTRODUCTION

4 STH ABOUT ROS???

5 Lidar data processing

To detect individual bricks would be handy to extract straight lines from lidar data. Several methods can be used to achieve this goal. One of the most popular algorithms for lines extraction is currently split and merge algorithm. Initially was this algorithm proposed for image segmentation by Horowitz and Pavlidis [1]. Simple version of this algorithm for pointcloud processing is described in algorithm 1. There are many implementations of this algorithm which differs mainly in a way how they compute some particular steps of the algorithm. For example just the method of fitting a line to cluster can vary a lot. Very often is used the least squares method, but as simple method as connecting endpoints of cluster could be used. When the latter method is applied, algorithm is usually referred as Iterative End Point Fit (IEPF) [2]. For the cluster creation are the points iterated in each layer one by one. When the distance of subsequent points is too high we split the cluster. Every cluster is then further recursively split based on the most distant point from the fitted line. In a comparison to other line extraction algorithms is the split and merge algorithm one of the best performing in terms of precision and computational complexity [3].

```
Data: pointcloud
Result: line_segments
initialize constants C, S;
clusters = find_clusters(pointcloud, C);
while clusters is not empty do
    cluster = clusters.pop();
    line = fit_line(cluster);
    point = most_distant_point(cluster, line);
    if distance(point, line) > S then
        c1, c2 = split_cluster(cluster, point);
        clusters.push_back(c1, c2);
    else
        line_segments.push_back(cluster[start], cluster[end]);
    end
end
merge_colinear(line_segments);
```

Algorithm 1: Lidar data segmentation using split and merge algorithm. C is clustering distance and S is splitting distance.

6 EM algorithm

Expectation-maximization (EM) algorithm is iterative process which can find parameters of certain statistical model based on incomplete data. One of the most used statistical description for the EM algorithm is the Gaussian mixture model. This model is particularly useful because it emerges in many real world situations and it is easy to maximize. As the name of algorithm suggest it repeats expectation and maximization step. Each iteration of the algorithm should improve the likelihood of the model until the terminating criterion is met. |Termination criterion could be simply number of iterations or the algorithm can be stopped when the model is not improving anymore. Although we are discussing mainly the Gaussian distribution, EM algorithm can be also used for other distributions from exponential family [4].

6.1 Maximization

For maximization step is used maximal likelihood estimate weighted by α from expectation step. For parameters of Gaussian distribution $\mathcal{N}(\mu, \sigma)$ and number of samples N looks maximization as follows:

$$\mu = \frac{1}{N} \sum_{n=1}^N \alpha_n x_n \quad (1)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{n=1}^N \alpha_n (x_n - \mu)^2}. \quad (2)$$

The most important assumption for the usage of the algorithm is that its likelihood with respect to estimated parameter must be concave. This can be easily proved by computing the second derivative of likelihood function. For example for mean value μ it is easy to show that second derivative of likelihood is always negative.

$$\mathcal{L} = \prod_{n=1}^N \mathcal{N}(x_n, \mu, \sigma) \quad (3)$$

$$\frac{\partial \log \mathcal{L}}{\partial \mu} = \frac{1}{\sigma^2} \sum_{n=1}^N (x_n - \mu) \quad (4)$$

$$\frac{\partial^2 \log \mathcal{L}}{\partial \mu^2} = -N. \quad (5)$$

6.2 Expectation

The expectation step is done simply by evaluating probability density function of Gaussian distribution with the parameters from maximization step.

$$\alpha_n = \mathcal{N}(x_n, \mu, \sigma), \quad (6)$$

where n is index of the data sample. If the prior probabilities of observed random variable are known, it is also possible to exploit the Bayes theorem to compute α .

6.3 Algorithm

How to implement general version of EM algorithm on sampled data is shown in algorithm 2. All important calculations for Gaussian distribution are described in previous subsections. It is not clear where to start iterating. It is possible to start both with the expectation and with the maximization step, but both parts are dependent on the result of the other one. Here we start with the maximization step so during the initialization we set $\alpha_n = 1$. If some prior information about parameters of the model is available, they can be set during the initialization and the algorithm can be started with expectation step. This informed initialization can highly reduce the number of iterations and sometimes even an outcome of the algorithm.

```
Data: x  
Result: parameters  $\theta$   
set all  $\alpha_n = 1$ ;  
while not stopping_criterion do  
    |  $\theta = \text{maximization}(x, \alpha)$ ;  
    |  $\alpha = \text{expectation}(x, \theta)$ ;  
end
```

Algorithm 2: Pseudocode shows how to implement the EM algorithm. x is the observed data.

7 RANSAC

Random sample consensus (RANSAC) is an iterative method which can estimate parameters of hypothesis given the data. It was first presented by Fischler [5] with application in scene and image analysis, but it can be used for fitting arbitrary hypothesis. The biggest advantage of this algorithm is its robustness to outliers. Major drawback of this method

METHODS

```
Data: x
Result: best parameters  $\theta^*$ 
initialize  $\theta, \theta^*, C, C^*$ ;
while not stopping_criterion do
    samples = draw_samples(x);
     $\theta$  = find_parameters(samples);
     $C$  = compute_cost(x,  $\theta$ );
    if  $C > C^*$  then
         $C^* = C$ ;
         $\theta^* = \theta$  ;
    end
end
```

Algorithm 3: Pseudocode shows how to implement the RANSAC algorithm. x is the observed data, C is the maximized cost and θ are the parameters of the hypothesis.

is very high time complexity when fitting hypothesis to noisy data with large number of samples. Whole iterative process is described in algorithm 3.

Number of drawn samples in **draw_samples** must be equal or higher than the number of degrees of freedom of the hypothesis. After drawing the samples method **find_parameters** assigns the correspondences between sampled data and the hypothesis. The correspondences are used to obtain the parameters of the hypothesis. Then is the algorithm evaluating quality of the hypothesis by applying the hypothesis to whole dataset. This can be done by arbitrary cost function. Common practice is to define some metrics in our domain and use a threshold value to obtain the number of samples which fits the hypothesis. These samples are often referred as inliers. Stopping criterion is usually met when the probability of sampling better hypothesis is lower than a specified threshold. This section describes just the basic version of the algorithm. Many improvements to RANSAC algorithm was proposed since 1981 such as [6] or [7].

7.1 Tentative Correspondences

The tentative correspondences can help us to choose better samples from the data to generate the better hypothesis. It is necessary to define some function which measure the similarity between data and hypothesis. The data which has higher similarity to the hypothesis are then chosen with higher probability. It is also possible to completely ban correspondences with low similarity. Given the typical application in scene analysis is similarity usually computed by comparing keypoint descriptors.

8 Lidar to camera registration

As can be seen in the figure 1 (where the bricks are defined) besides the dimensions another important feature of the bricks is their color. Although, the color manifests itself little bit in reflectivity of the surface which can be detected by lidar, it is not possible reliably distinguish the colors using only the lidar sensor. Until there is a gap between the individual bricks, it is possible to detect brick using just the spatial data. Ideally the robot should be stacking bricks next to each other without any significant gap. Without any information about the color is then impossible to decide whether is the robot detecting one large brick or several small bricks put together. So for this part of detection is necessary to color the pointcloud. This can be done by using the image from Intel RealSense camera and projecting 3D lidar pointcloud to the camera plane. For this purpose is used the pinhole camera model. To describe such a model is used intrinsic camera matrix K which consists of intristic camera parameters [8]

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \quad (7)$$

where f_x, f_y are focal lengths and c_x, c_y stands for optical center of the camera.

Firstly is necessary to transform whole pointcloud from lidar coordinate frame to camera frame. For this purpose are used so called extrinsic camera parameters, which describes where in lidar coordinate frame is camera placed. This is more discussed in the next section of the thesis. Secondly we can use the camera intrinsic parameters to calculate the projection. Note that it is needed to work in homogenous 2D coordinate system.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \quad (8)$$

Using this matrix equality is possible to decide which coordinates u, v on image plane corresponds to 3D point x, y, z from pointcloud and assign color of a certain pixel to the point. However this works only for the simplest pinhole camera model without any distortion of image. If the lens has non-negligible distortion, this distortion must be included in the camera model. For description of distortion are used the distortion coefficients.

9 Global model and transformations

One of the goals of this thesis is develop a global model which can efficiently store and update the positions of interest points. This global model is in the map coordinate frame. The localization of the robot is absolutely essential for precise global model. Every

METHODS

detection must be transformed from the coordinate frame of the sensor to coordinate frame of the map. Transformation can be easily done using matrix multiplication:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R \begin{bmatrix} x \\ y \\ z \end{bmatrix} + t, \quad (9)$$

where R is 3×3 the rotation matrix and t is a 1×3 translation vector between coordinate frames. Similarly can be the transformation done in homogenous coordinates by merging translation and rotation into one matrix T :

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}. \quad (10)$$

Different framework for computing the transformations is using the quaternions. Quaternions are currently the standard for transformations in computer graphics and robotics. The biggest advantage of quaternions is that they are more efficient and does not suffer from gymbal lock and ambiguity of rotation. Arbitrary rotation and scaling can be expressed as quadruple of numbers in quaternion framework. The rotation between coordinate frames $B \rightarrow A$ is computed with quaternions as:

$$q_A = q_T q_B q_T^*, \quad (11)$$

where q_A is quaternion in coordinate frame A , q_B is quaternion in coordinate frame B , q_T is quaternion representing the transformation between these coordinate frames and q_T^* is its conjugate. There is available a library within the ROS which can handle all these transformations in different forms [9].

9.1 Symbolic map

When are all detections transformed into the map frame we can add them into a symbolic map. The symbolic map is storing the positions of all interest points and makes up the global model of the arena.

10 Lidar detection range analysis

11 Program pipeline

11.1 Line segmentation

11.2 Pile detection

11.3 Pattern fitting

12 Conclusion

BLAH BLAH

12.1 Future work

BLAH BLAH

References

- [1] Steven L. Horowitz and Theodosios Pavlidis. Picture segmentation by a tree traversal algorithm. 1974.
- [2] Ali Syadat and Axel Kaske. An optimized segmentation method for a 2d laser-scanner applied to mobile robot navigation. 1997.
- [3] Viet Nguyen and Stefan Gachter. A comparison of line extraction algorithms using 2d range data for indoor mobile robotics. 2006.
- [4] A. P. Dempster and N. M. Laird. Maximum likelihood from incomplete data via the em algorithm. 1977.
- [5] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. 1981.
- [6] Ondrej Chum and Jan Matas. Locally optimized ransac. 2003.
- [7] Ondrej Chum and Jan Matas. Optimal randomized ransac. 2008.
- [8] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2017.
- [9] Tully Foote. tf: The transform library. In *Technologies for Practical Robot Applications (TePRA), 2013 IEEE International Conference on*, Open-Source Software workshop, pages 1–6, April 2013.

APPENDIX

Appendix A CD Content

In Table 2 are listed names of all root directories on CD.

Directory name	Description
thesis	the thesis in pdf format
ctu_thesis	latex source codes
lidar-gym	OpenAI gym environment

Table 2: CD Content

APPENDIX

Appendix B List of abbreviations

In Table 3 are listed abbreviations used in this thesis.

Abbreviation	Meaning
EM	Expectation maximization

Table 3: Lists of abbreviations

APPENDIX
