Door Pose Estimation and Robot Positioning for Autonomous Door Opening

Ulrich Mitterhuber[1], Raimund Edlinger[1], Roman Froschauer[1] and Andreas Nüchter[2]

Abstract— **For autonomous robots to deliver value in human centered environments, they must be able to autonomously open doors. For doing so, they have to overcome multiple challenges, one of which is, to estimate the desired door's orientation and position. The information if the door handle is on the left or right side of the door must also be obtained. In this work, a novel method, solving the stated problems is proposed. It's perception is based on a sensor fusion of a monocular camera based state-of-the-art deep learning object detection algorithm with a 2D laser scan and subsequent line estimation. Additionally, a differential drive controller, using the advancement of continuous goal pose updating, is proposed. During real-world experimentation with a differential drive robot, the implemented system was able to position the robot in front of a door every time with sufficient accuracy and is thus found to solve the stated problem successfully.**

I. Introduction

Autonomous mobile robots can be used in many cases such as retail, industrial site monitoring or rescue robotics just to name a few. To enable them operating in human environments they must be capable to detect doors and door handles, position themselves in front of them, manipulate them and subsequently traverse through them. A door traversing behaviour could be utilised in a fully autonomous application such as in industrial site monitoring or as a semi autonomous assistance function in for example search and rescue applications.

The aim of this work is to tackle the above stated problem by solving the first tasks required for autonomous door opening. More specifically, a robot behaviour for detecting a previously unknown door, estimating it's pose, determining the side at which the door handle is located and then positioning the robot in front of the door at a specific distance, is implemented. The implemented methods aim to be executed in real time on embedded computing platforms commonly used on autonomous robots.

To fulfill the stated objective, multiple algorithms were developed which require the sensor input from a calibrated monocular RGB camera with a minimum resolution of 320x240 pixels, a 2D laser scanner covering the horizontal field of view (FOV) of the camera and an odometry source for robot state estimation. The robots also must provide a velocity controller which accepts a linear and angular velocity. For door pose estimation, a state-of-the-art deep learning based object detection algorithm, more specifically *YOLOV5*[1], is used to detect doors within camera images. This information is then fused with a laser scan to determine the pose of the door by using line estimation. For that, three different line estimation algorithms are implemented and compared experimentally. The side of the door handle in the door is estimated using the output of the object detection. Lastly, a differential drive controller using the contribution of continuous goal pose updating is implemented.

The method relies on the conditions that before starting, the robot is less than laser scanner's maximum measurement distance away from the door, there are no obstacles between the robot and the door and the door is within the FOV of the camera.

In the following pages first, in sec. II the current state-of-the-art in door opening is examined. After that, sec. III and sec. IV show the implemented methods with their postulated hypothesis and their experimental validation respectively. Sec. V sums up the findings and provides an outlook on possible fields of applications and further improvements of this work.

II. Related Work

The field of research in opening and detecting doors and estimating their pose has been very active for many years. However, especially since the rise of deep learning, many methods based on convolutional neural networks were proposed.

An example of an older approach for opening doors is the assistance system proposed in [9]. The human user of the system has to use green laser pointer to indicate where the door handle is. This point is then detected by the robots vision system and the door opening procedure is started. It is the only work discussed here, which needs a user to actively interact with the robot.

The works proposed in [17], [10] and [14] are not focused on door detection but rather include methods for door handle detection. [17] proposed a deep learning based algorithm while [14] used a template matching algorithm.

The authors of [8] proposed a mobile robot employing a manipulator for autonomous disinfection operations. They use a 3D model of a building and the Iterative Closest Point algorithm [5] for detecting door handles and then refining their pose after positioning.

*This work was not supported by any organization

[1]Ulrich Mitterhuber, Raimund Edlinger and Roman Froschauer are with University of Applied Sciences Upper Austria, 4600 Wels, Austria ulrich.mitterhuber@fh-wels.at, raimund.edlinger@fh-wels.at

[2]Andreas Nüchter is with Informatics VII - Robotics and Telematics, Julius-Maximilians University Würzburg andreas.nuechter@uni-wuerzburg.de

[1]https://github.com/ultralytics/yolov5

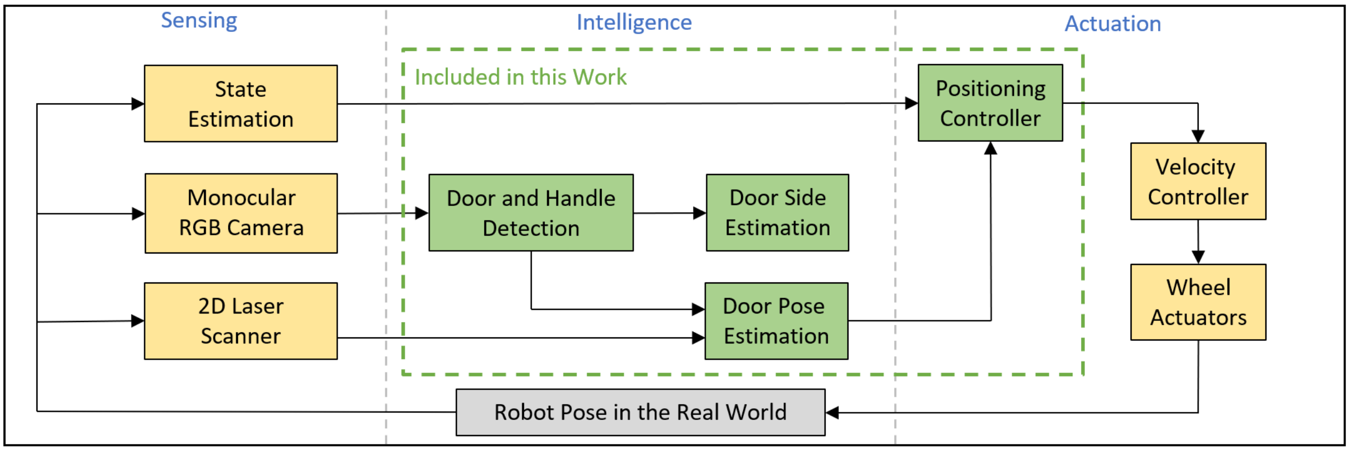

Fig. 1. System overview: The elements in green are developed in this work.

There were also some methods published in which the robot's objective only is to manipulate a door. Thereby the door poses are already coarsely marked in a map or included through prior information such as the robot's pose. [15], [23] and [12] for example each propose different methods of opening or even traversing doors, however do not detect them themselves. They all use the information provided through the map but also employ point cloud processing algorithms to refine the door pose. [3] on the other hand, proposes a semi-autonomous system, for which the robot must be pre-positioned well in front of the door. When the operator starts the opening procedure, the robot uses a 2D laser scanner, and the assumptions where it expects the door, to estimate it more accurately.

The available methods which are detecting doors without prior information can be further divided into two parts. First, methods relying not on deep learning and second, deep learning based ones.

Three examples for non deep learning algorithms are [21], [16] and [4]. The method proposed in [21] uses uncoloured 3D point clouds while the one proposed in [16] uses coloured 3D point clouds to detect doors and estimate their poses in 3D space. [4] uses classical machine vision techniques and the assumption that a door has many vertical and horizontal lines to detect doors within camera images.

The works proposed by [2] and [20] use the same deep learning based object detection algorithms to both detect doors and door handles. They use the obtained bounding box to choose a ROI in a depth map they obtain with a RGB-D camera. A plane is then fitted into the depth map. With that, the door pose becomes available. They estimate the pose of the door handle in a similar fashion as well. The authors of [11] also proposed a convolutional neural network to first estimate a ROI representing the door. They then use the point cloud and visual data from within the ROI to estimate the door plane and find the door handle with non deep learning methods. Lastly, [18] compares various deep learning based algorithms to estimate if a door is open, semi-open or closed. They do however not estimate the doors pose or detect door handles.

The door pose estimation algorithm developed in this work is similar to the one proposed in [2] and [20]. However, it is based on a simpler sensor design which enables it to employ model estimation methods with lower computational requirements.

III. Method

Fig. 1 shows the methods developed in this work embedded in their context of an autonomous mobile robot. The elements in yellow and gray are required for the methods to work and are therefore not part of this study while the ones in green represent the methods developed in this work.

A. Door and Door Handle Detection

The developed door and door handle detection algorithm is based on a state-of-the-art object detection method, more specific *YOLOV5* which is an improvement over the original *YOLO* architecture proposed in [19]. It was chosen due to it's very good implementation, vast community support and the ability to be executed in real-time on an embedded platform.

In order to enable the model to recognise doors and door handles, a data set was created which consists of images and labels of these objects. For its creation, own images were taken and labeled. However, to increase its size and diversity, images of publicly available datasets provided by the authors of [18] and [2] were added as well. They had to be relabeled partly but added a lot of diversity in terms of geographical bias and camera parameters. In total, the dataset consists of 1022 images and labels with a class instance count 1031 for doors and 1103 for door handles. A sample of the dataset is shown in Fig. 2.

The exact dataset split for training and evaluation, the specific hyper-parameter choices, the training results and the model deployment are further explained in section IV.

B. Door Pose Estimation

As already noted in sec. I, this method relies on sensor fusion. More specifically it uses a cooperative sensor fusion

Fig. 2. Object detection dataset sample: Strong red indicates door and light red indicates door handle.

approach whereby the data of the exteroceptive monocular camera and the 2D laser scanner are combined. [1]

The motivation for fusing those sensors is that this enables the method to leverage the recent developments in deep learning machine vision algorithms which provide great performance-efficiency trade-offs such as shown in works as [19] and combine that with the efficiency of 2D laser scan processing. This combination makes the method perfectly suited for autonomous robotics. Furthermore, assumptions which simplify the problem drastically are made. First, a door is always seen as a plane perpendicular to the ground plane and second, the often in robotics made simplification that the world in which the robot's base navigates is represented as a 2D plane, where an object only has the three degrees of freedom $x, y, theta$, is used. A visual representation of the method, which is often referred to in the following algorithm explanations, is shown in Fig. 3.

The input of the algorithm is comprised of door and door handle bounding boxes delivered by the object detection algorithm, the camera's calibration parameters and a laser scan. Since the different input artefacts are not delivered at exactly the same time, they are approximately synchronised with each other. It is important to note that the camera's and laser scanner's optical centers must not be at the same position. For all calculations, the points obtained by the laser scanner are transformed into the camera's frame.

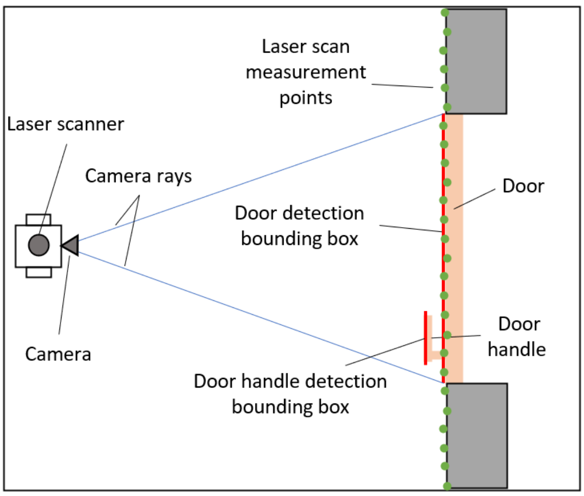

Fig. 3. Door pose estimation method overview.

The algorithm's first step is to evaluate if a valid door was detected and chose one if multiple were detected. Choosing one of multiple is done by finding the one with the highest corresponding confidence score. The selected door must however obey to a geometric rule which is that the box is not allowed to go all the way out to the left or right image end. This is done to ensure that the door is fully covered in the image and therefore a bounding box representing only the half of a door is not chosen.

With the door bounding box being chosen, in the next step, it's width and the camera's parameters are used to calculate vectors representing camera rays from image pixels. For the left ray's image pixel x coordinate, the x value of the left vertical edge of the bounding box is taken and vice versa for the right ray. The y coordinates for the ray calculation are set to a constant value representing the height at which the camera is mounted on the robot. The calculated rays are then projected onto the 2D x, y plane. Fig. 3 shows the door and door handle bounding boxes in red and the camera rays in blue.

The 2D ray vectors are then used to calculate angles representing the left and right edge of the bounding box. To find the laser scan points which are within the ray angles, first, a x, y point of each laser scan range point is calculated which is then transformed from the scan frame into the camera frame. After that, each point's angle from the camera center is calculated. Every angle that is between the range of the left and right ray angle is assumed to be located on the door. Laser scan range points with invalid values are discarded. Fig. 3 shows the laser scan measurement points in green.

The following step in the algorithm is to fit a 2D line, representing the door pose, into the selected laser scan points. For that, three methods to choose from were implemented. The first one, which is further referred to as naive, connects the leftmost to the rightmost point to construct the line. It ignores all other points within the range and is therefore expected to be very unstable when outliers are occurring at the door edges. The second one uses the commonly known ordinary least squares estimation to fit a line into all available door-points. If there are no outliers at all, this method is expected to deliver optimal results. Due to the fact that it is expected that the algorithm has to cope with many outliers because of already expected bounding box inaccuracies and sensor errors, the third method chosen is the *Random Sample Consensus* (RANSAC) proposed in [7]. It is a probabilistic method commonly known in robotics for its ability to very efficiently find models in noisy data. It is expected that the RANSAC method yields the most accurate results in real-world scenarios. To verify this hypothesis and thus determine the best method for the task at hand, thorough experiments, shown in sec. IV, were conducted.

Once a door line is estimated, the left and right edge points of the door represented in 2D space are obtained. This is done by calculating an intersection each for the left and right camera rays with the estimated door line. Using those

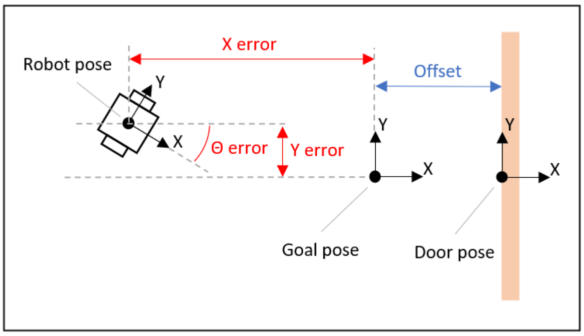

Fig. 4. Robot positioning control visualisation.

two points, the doors center coordinate x, y and its rotation $theta$ are calculated. Lastly, to obtain the goal pose to which should be navigated in front of the door, a variable offset is added normal to the door. It's value needs to be chosen while contemplating the workspace size of the manipulator used on the mobile robot. Fig. 4 shows an estimated door and goal pose.

Parallel to the steps above, the estimation of the door handle side is executed. If there is a door handle bounding box found to be within the selected door bounding box it is determined if the center of the handle is in the left or right half of the door. The information of left or right is provided for further manipulation methods. With it, it is clear on which side the door is to be opened and in which direction the handle needs to be turned.

C. *Differential Drive Positioning Controller*

The controller implemented to move the robot from its starting pose to the goal pose is based on the well known differential drive controller proposed in [22]. Fig. 4 shows the robot in its starting pose, a goal pose and the error which is minimised by the controller.

Another hypothesis made here is that the final positioning accuracy can be improved, if during the robot is moving towards a goal pose, the goal pose estimation goes on and the old estimation is updated continuously, instead of only setting it once at the beginning. This is based on the assumptions that first, the door pose estimation is more reliable than the pose estimation of the robot and second, that if the door is at the beginning of the positioning partly occluded, the view of the door will get better during positioning and with that the goal pose estimation also gets more accurate. It is expected that the door detection is only reliable if the robot is far away from the door. For that, the euclidean distance between the robot's current pose estimation and the goal pose is used to determine when to stop updating. The validity of this hypothesis is evaluated during experimentation shown in section IV.

IV. Experimental Results

To validate the hypothesis made above and verify the functionality of the developed methods, real-world experiments were conducted. The differential drive robot platform used for testing provided an odometry solely based on wheel velocities, a wheel velocity controller, a 360 degree 2D laser scanner with 1 degree angular resolution and a calibrated RGB camera with a horizontal FOV of 60 degrees delivering images with a size of 640x480.

A. *Door and Door Handle Detection*

For training the object detection model, the dataset was split randomly into 715 train, 153 validation and 154 test images. Training was executed on a PC comprising of an AMD Ryzen 9 3950X CPU, 128 GB DDR 4 RAM and two NVIDIA RTX 2080 Ti graphics cards running on Ubuntu 18.04. The batch size was set to 90 images with a size of 640x640 each. For data augmentation random horizontal flipping with a probability of 0.5, linear translation with a factor of 0.1 and scaling with a factor of 0.5 were used. *YOLOV5* provides a few different model architecture types from which the lightweight *YOLOV5S* was chosen. Training was done from scratch and stopped after 154 epochs because no improvements were made over the previous 30 epochs. For model evaluation the mean average precision metric (mAP), introduced for the PASCAL VOC challenge in [6], was employed. When evaluating the trained model on the test dataset, it achieved a mAP 0.5 of 90.1% and mAP 0.5:0.95 of 59.9%. Therefore it is concluded, that the model is able to reliably detect previously unseen doors and door handles. Some qualitative inference results are shown in Fig. 5. Both images were acquired in the real world and are not included in the training dataset. It can be observed that all doors and door handles present in the images are detected.

For inference, the model was deployed within a ROS node implemented in plain Python and PyTorch on the *NVIDIA Jetson AGX Xavier*. It must be noted that no optimization frameworks such as *NVIDIA TensorRT* were employed. Nonetheless, the model is able to perform inference with an input image size of 640x640 with a speed of above 30 frames per second.

B. *Door Pose Estimation*

To validate the hypothesis that the RANSAC based door line estimation method is the most accurate one for this task, a measurement series of 20 samples comparing each method to a ground truth was taken. For each sample, the robot was placed in a new scenario in which the door or

Fig. 5. Object detection: Qualitative inference examples.

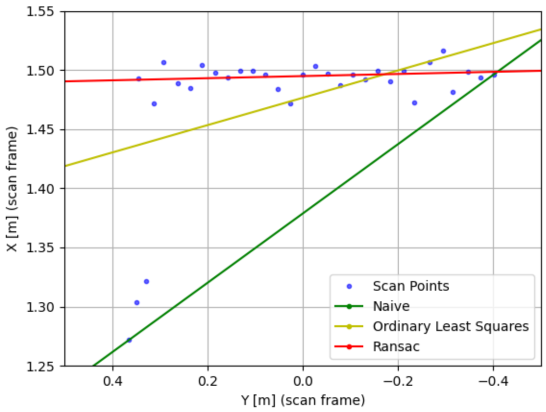

Fig. 6. Door line estimation methods comparison with an obstacle at the left side.

an obstacle placement was changed. During experimentation with the RANSAC method, the max trials parameter was set to 1000 and the residual threshold to 0.01. Figure 6 shows one sample measurement with strong outliers at the left side resulting from an obstacle. It can be observed that the RANSAC method is the most and the naive method the least accurate. Each estimated door line is subsequently used to calculate the door's estimated pose. The estimated and ground truth door poses were then used to calculate an absolute error for every dimension, e.g. $x, y, theta$, of each estimated pose, for each sample. The three pose dimension errors are then summed up resulting in a single error metric for a given sample. With that, for each method, 20 samples of total absolute pose errors representing their accuracy were obtained. Tab. I shows the mean and standard deviations calculated for each methods error data. To determine if the collected data supports the hypothesis that the method based on RANSAC is the most accurate, i.e. yields the lowest error, two Z-Tests with known variances such as described in [13] were executed. The significance threshold was set to 5% . In both tests, the null-hypothesis was that both methods are equally good while the alternative-hypothesis was that the error of the naive and least squares method is higher than the RANSAC method's error. In both tests, the null-hypothesis was reject and thus the data significantly supports the hypothesis that the RANSAC method is the most accurate one.

TABLE I
COMPARISON OF THE IMPLEMENTED LINE ESTIMATION ALGORITHMS WITH REAL WORLD DOORS.

Method	**Mean [1]**	**Standard Deviation [1]**
Naive	0.3212	0.1774
Least Squares	0.1167	0.0667
Ransac	0.0311	0.0062

TABLE II
COMPARISON OF INITIAL AND CONTINUOUS GOAL POSE UPDATING IN THE CONTROLLER.

Goal Updating	**Mean [1]**	**Standard Deviation [1]**
Initial	0.2492	0.0948
Continuous	0.0650	0.0361

C. Differential Drive Positioning Controller

To validate the hypothesis that continuously updating the goal pose during moving the robot to the goal yields a more accurate final placement than updating only once in the beginning, an experiment was conducted. The robot was placed at 18 different start poses from which it was already able to estimate the door pose, i.e. the goal-pose. From each start pose, positioning was once done with and without continuous updating. The euclidean distance parameter for stopping continuous updating was set to 0.2m. For estimating the goal pose, the RANSAC method with the maximum trials parameter set to 1000 and the residual threshold set to 0.01, was used. Lastly, derived from the employed robot's manipulator workspace size, the goal offset was set to 0.5m.

After each positioning procedure, the error between the reached and the ground truth pose was measured. To represent the positioning error within one metric, again the absolute $x, y, theta$ errors were summed up to form a total absolute pose error. Tab. II shows the mean and standard deviation of both method's errors. To determine if the collected data supports the hypothesis that continuous goal pose updating yields a more accurate final pose, i.e. a lower positioning error, than only estimating the goal once, a Z-test, such as used before, comparing both method's measurements, was executed. The significance threshold was again set to 5%. The null-hypothesis was that both methods perform equally good while the alternative-hypothesis was that non continuous updating yields a higher error. When executing the test, the null-hypothesis was clearly rejected and with that the data significantly supports the hypothesis that continuous goal pose updating improves the positioning accuracy.

V. SUMMARY AND OUTLOOK

This work aimed at solving the problem of detecting doors, estimating their pose and controlling a differential drive robot to a specified goal in front of the door. The objective was successfully fulfilled by developing a novel door pose estimation algorithm based on deep learning object detection and 2D laser scan processing. It was also shown that the implemented controller achieves a higher positioning accuracy when the goal pose is estimated and updated continuously while the robot is moving towards the goal pose. The fields of application of the proposed method do not only lie at autonomous door opening, but also in areas such as environment exploration where the method could be used to detect various objects autonomously, estimate their pose and integrate them into an environment map. Future work will include the creation of a more robust and diversified dataset for the training of the neural network and deploying the algorithm on other types of mobile robots.

References

[1] M. B. Alatise and G. P. Hancke, "A review on challenges of autonomous mobile robot and sensor fusion methods," *IEEE Access*, vol. 8, pp. 39 830–39 846, 2020.

[2] M. Arduengo, C. Torras, and L. Sentis, "Robust and adaptive door operation with a mobile robot," *Intelligent Service Robotics*, vol. 14, no. 3, pp. 409–425, 2021.

[3] B. Axelrod and W. H. Huang, "Autonomous door opening and traversal," in *2015 IEEE International Conference on Technologies for Practical Robot Applications (TePRA)*. IEEE, 2015, pp. 1–6.

[4] N. Banerjee, X. Long, R. Du, F. Polido, S. Feng, C. G. Atkeson, M. Gennert, and T. Padir, "Human-supervised control of the atlas humanoid robot for traversing doors," in *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2015, pp. 722–729.

[5] A. Censi, "An icp variant using a point-to-line metric," in *2008 IEEE International Conference on Robotics and Automation*. Ieee, 2008, pp. 19–25.

[6] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.

[7] 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*, vol. 24, no. 6, pp. 381–395, 1981.

[8] A. Giusti, V. Magnago, D. Siegele, M. Terzer, C. Follini, S. Garbin, C. Marcher, D. Steiner, A. Schweigkofler, and M. Riedl, "Balto: A bim-integrated mobile robot manipulator for precise and autonomous disinfection in buildings against covid-19," in *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2021, pp. 1730–1737.

[9] A. Jain and C. C. Kemp, "Behaviors for robust door opening and doorway traversal with a force-sensing mobile manipulator." Georgia Institute of Technology, 2008.

[10] E. Klingbeil, A. Saxena, and A. Y. Ng, "Learning to open new doors," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 2751–2757.

[11] A. Llopart, O. Ravn, and N. A. Andersen, "Door and cabinet recognition using convolutional neural nets and real-time method fusion for handle detection and grasping," in *2017 3rd International Conference on Control, Automation and Robotics (ICCAR)*. IEEE, 2017, pp. 144–149.

[12] W. Meeussen, M. Wise, S. Glaser, S. Chitta, C. McGann, P. Mihelich, E. Marder-Eppstein, M. Muja, V. Eruhimov, T. Foote, *et al.*, "Autonomous door opening and plugging in with a personal robot," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 729–736.

[13] D. C. Montgomery and G. C. Runger, *Applied statistics and probability for engineers*. John Wiley & Sons, 2010.

[14] C. Ott, B. Bäuml, C. Borst, and G. Hirzinger, "Employing cartesian impedance control for the opening of a door: A case study in mobile manipulation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, Workshop on mobile manipulators: Basic techniques, new trends & applications*, 2005.

[15] S. A. Prieto, A. Adán, A. S. Vázquez, and B. Quintana, "Passing through open/closed doors: A solution for 3d scanning robots," *Sensors*, vol. 19, no. 21, p. 4740, 2019.

[16] B. Quintana, S. A. Prieto, A. Adán, and F. Bosché, "Door detection in 3d colored laser scans for autonomous indoor navigation," in *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2016, pp. 1–8.

[17] B. Ramalingam, J. Yin, M. Rajesh Elara, Y. K. Tamilselvam, M. Mohan Rayguru, M. Muthugala, and B. Félix Gómez, "A human support robot for the cleaning and maintenance of door handles using a deep-learning framework," *Sensors*, vol. 20, no. 12, p. 3543, 2020.

[18] J. G. Ramôa, V. Lopes, L. A. Alexandre, and S. Mogo, "Real-time 2d–3d door detection and state classification on a low-power device," *SN Applied Sciences*, vol. 3, no. 5, pp. 1–13, 2021.

[19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[20] S. Roelofs, B. Landry, M. K. Jalil, A. Martin, S. Koppaka, S. K. Tang, and M. Pavone, "Vision-based autonomous disinfection of high-touch surfaces in indoor environments," in *2021 21st International Conference on Control, Automation and Systems (ICCAS)*. IEEE, 2021, pp. 263–270.

[21] R. B. Rusu, W. Meeussen, S. Chitta, and M. Beetz, "Laser-based perception for door and handle identification," in *2009 International Conference on Advanced Robotics*. IEEE, 2009, pp. 1–8.

[22] R. Siegwart, I. R. Nourbakhsh, and D. Scaramuzza, *Introduction to autonomous mobile robots*. MIT press, 2011.

[23] M. Stuede, K. Nuelle, S. Tappe, and T. Ortmaier, "Door opening and traversal with an industrial cartesian impedance controlled mobile robot," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 966–972.