

Multistage Localization for High Precision Mobile Manipulation Tasks

Christopher James Mobley

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Masters of Science
in
Mechanical Engineering

Tomonari Furukawa
Brian Lattimer
Kevin Kochersberger

December 14, 2016
Blacksburg, Virginia

Keywords: Autonomous Navigation, SLAM, Visual Servoing, Mobile Manipulation,
Computer Vision, State Machine
Copyright 2016, Christopher James Mobley

Multistage Localization for High Precision Mobile Manipulation Tasks

Christopher James Mobley

(ABSTRACT)

This paper will present a multistage localization approach for an autonomous industrial mobile manipulator (AIMM). This approach allows tasks with an operational scope outside the range of the robot's manipulator to be completed without having to recalibrate the position of the end-effector each time the robot's mobile base moves to another position. This is achieved by localizing the AIMM within its area of operation (AO) using adaptive Monte Carlo localization (AMCL), which relies on the fused odometry and sensor messages published by the robot, as well as a 2-D map of the AO, which is generated using an optimization-based smoothing simultaneous localization and mapping (SLAM) technique. The robot navigates to a predefined start location in the map incorporating obstacle avoidance through the use of a technique called trajectory rollout. Once there, the robot uses its RGB-D sensor to localize an augmented reality (AR) tag in the map frame. Once localized, the identity and the 3-D position and orientation, collectively known as pose, of the tag are used to generate a list of initial feature points and their locations based on *a priori* knowledge. After the end-effector moves to the approximate location of a feature point provided by the AR tag localization, the feature point's location, as well as the end-effector's pose are refined to within a user specified tolerance through the use of a control loop, which utilizes images from a calibrated machine vision camera and a laser pointer, simulating stereo vision, to localize the feature point in 3-D space using computer vision techniques and basic geometry. This approach was implemented on two different ROS enabled robots, the Clearpath Robotics' Husky and the Fetch Robotics' Fetch, in order to show the utility of the multistage localization approach in executing two tasks which are prevalent in both manufacturing and construction: drilling and sealant application. The proposed approach was able to achieve an average accuracy of ± 1 mm in these operations, verifying it's efficacy for tasks which have a larger operational scope than that of the range of the AIMM's manipulator and its robustness to general applications in manufacturing.

Acknowledgments

I would like to thank my Lord and Savior Jesus Christ. It is His grace alone that has sustained me throughout the completion of this degree despite life's twists and turns.

I would like to thank my advisor, Dr. Furukawa, for his guidance and mentorship throughout my graduate career. His constant feedback has allowed me to mature not only as a researcher, but as a person. I would also like to thank my committee, Dr. Kochersberger and Dr. Lattimer, for their time, counsel, and support.

I want to thank my corporate sponsors who helped to make this work possible. I would like to thank Matt Stremler, Tim Ward, and Eric Holterman at the Commonwealth Center of Advanced Manufacturing (CCAM) for their direction and support of this project.

I would like to thank members of every robotic lab on campus; without their depth of knowledge, guidance, and support of this project, it would not have been completed satisfactorily. In particular, I would like to thank Jason Ziglar, John Seminator, Jack Newton, and Matt Meeder of the Mechatronics lab; John Peterson, Evan Smith, and Danny Whitehurst of the Unmanned Systems Lab; Robert Griffin of the Terrestrial Robotics Engineering and Controls (TREC) Lab, and Adam Shoemaker of the Center for Dynamic Systems Modeling and Control (DySMaC). In addition, I would like to thank all my colleagues in the Computational Multi-physics Systems (CMS) Lab and the Mitchell Robotics Lab (MRL). Without their help, input, and the long nights, I would not have finished. Specifically, I would like to thank Kuya Takami, Boren Li, Rich Fedora, Orson Lin, Murat Ambarkutuk, Luan Doan, Hangxin Liu, Yazhe Hu, Tian Yi, Peter Amico, Cameron Ridgewell, Spencer Leamy, Jason Doyle, Urvi Desai, Jihong Cai, Mickey Cowden, Tom Ehrenzeller, Diya Li, George Kontoudis, Josiah Steckenrider, and Cong Chen, as well as Yoonchang Sung and Mengyu Song.

I would like to thank my twin brother, Steven, parents, Ken and Luany, as well as all of my extended family, for their love, encouragement, continued moral support, and advice on every difficult life decision. I would also like to thank all of my friends, some of whom are as close as family, for all of the support and encouragement they have given to me along the way. I would not be the person I am today without their love, support, and direction throughout my life.

Finally, I would like to thank Virginia Tech, the Virginia Tech Corps of Cadets, Air Force De-

tachment 875, and Denbigh Baptist Christian School. It was through your positive learning environment that I was able to grow as both a student and a person.

Contents

1	Introduction	1
1.1	Background	1
1.2	Objectives	3
1.3	Summary of Original Contributions	3
1.4	Outline	3
2	Literature Review	5
2.1	Localization and Mapping For Autonomous Mobile Manipulators in Manufacturing and Construction	5
2.2	Task Association and <i>A Priori</i> Knowledge for Mobile Manipulators	6
2.3	Feature Localization Techniques for Mobile Manipulators	7
2.4	Summary	7
3	Fundamentals of Autonomous Robotics	9
3.1	ROS Concepts	9
3.1.1	ROS Communication	10
3.1.2	Rigid Body Transformations	10
3.2	Simultaneous Localization and Mapping Concepts	12
3.3	Localization and Path Planning Concepts	14
3.4	Manipulation Concepts	21
3.5	Task Execution Concepts	25
3.6	Camera Concepts	27

3.7	Computer Vision Concepts	32
3.7.1	Color Spaces	33
3.7.2	Linear and Non-Linear Filters	33
3.7.3	Morphological Operations	35
3.7.4	Canny Edge Detector	36
3.7.5	Hough Circle Detector	38
3.7.6	Good Feature To Track	41
3.7.7	Optical Flow	44
3.7.8	Pose Estimation and Tracking Through Augmented Reality Tag Detection	46
3.8	Summary	48
4	Multistage Localization for High Precision Mobile Manipulation Tasks	49
4.1	Approach Overview	49
4.2	Global Map Creation and Task Location Specification	50
4.3	Autonomous Localization and Navigation	51
4.4	Initial Feature Localization Framework	53
4.5	Motion Planning	53
4.6	Feature Correction Framework	54
4.7	Approach Implementation	57
4.7.1	System Overview	57
4.7.2	Drilling Framework	58
4.7.3	Sealant Application Framework	59
4.8	Summary	60
5	Experiments and Results	62
5.1	Hardware Architecture	62
5.2	Software Architecture	64
5.3	Experiments	64

5.3.1	Camera Calibration Setup	64
5.3.2	Navigation System Experimental Setup	67
5.3.3	Augmented Reality Tag Detection, Identification, and Localization System Experimental Setup	68
5.3.4	Drill Operation Experimental Setup	68
5.3.5	Sealant Application Experimental Setup	69
5.4	Results and Discussion	70
5.4.1	Camera Calibration Accuracy Achieved	70
5.4.2	Navigation System Accuracy Achieved	71
5.4.3	Augmented Reality Tag Accuracy Achieved	75
5.4.4	Drilling Operation Accuracy Achieved	78
5.4.5	Sealant Application Accuracy Achieved	79
5.5	Summary	80
6	Conclusions and Future Work	81
Bibliography		83

List of Figures

1.1	Vision for future transformable production systems. Adapted from [10].	2
3.1	Simple ROS Node Flowchart. Adapted from [33].	10
3.2	Conversion from Coordinate Frame A to B.	11
3.3	Robot Model with TFs.	11
3.4	Overview of SLAM Framework.	13
3.5	Example of 2-D Occupancy Grid Map Produced.	13
3.6	Real World Performance Analysis of ROS Available SLAM Algorithms. Adapted from [54].	14
3.7	ROS Navigation Stack setup [55].	15
3.8	One-Dimensional Monte Carlo Localization Example. Adapted from [59].	17
3.9	Particle Filter Resampling Example	19
3.10	Visual of AMCL in RVIZ	20
3.11	Trajectory Rollout Path Planning Framework.	21
3.12	Moveit!'s System Architecture. Adapted from [66]	22
3.13	Forward and Inverse Kinematics Example	23
3.14	Graphical View of State Machine using SMACH	27
3.15	Photometric Image Formation. Adapted from [80].	28
3.16	Digital Camera Diagram. Adapted from [82].	29
3.17	Ground Sample Distance Effects on Image Quality. Adapted from [87].	31
3.18	Shutter Type Effects on Image Quality. Adapted from [90].	32
3.19	RGB and HSV color space models	33

3.20	Effects of common filters.	35
3.21	Effects of common binary operations.	36
3.22	Summary of Canny Edge Detection Framework.	38
3.23	Hough Circle Detector's 2-D Accumulator Array with Known Radius (Penny). Adapted from [108].	39
3.24	Summary of Hough Circle detector with known radius.	40
3.25	Summary of Hough Circle detector with unknown radius.	40
3.26	Summary of Good Feature detector. Adapted from [112].	41
3.27	Image Gradient Example.	42
3.28	Difference Between Good Feature (a) and Harris Corner (b) Scoring Functions.	44
3.29	Example of optical flow. Adapted from [114].	45
3.30	Summary of Kanade-Lucas-Tomasi feature tracker.	46
3.31	Summary of ALVAR AR Tag Detection Framework. Adapted from [118]. . .	48
4.1	Multistage Localization Approach Overview.	50
4.2	2-D map environment with specified start locations.	51
4.3	Robot Localization Framework.	52
4.4	Initial Feature Localization Framework.	53
4.5	Feature Detection and Tracking Pipeline.	56
4.6	Manipulator Correction Control Loop.	57
4.7	General system overview.	58
4.8	Drilling operation framework.	59
4.9	Sealant Application Framework.	60
5.1	Clearpath Robotics' Husky and the Fetch Robotics' Fetch Mobile Manipulator	63
5.2	Camera Calibration Setup.	65
5.3	Extracted Corners of Calibration Pattern.	66
5.4	Extracted Corners and Laser Center	67
5.5	Navigation System Experimental Setup.	67
5.6	Drilling Operations Experimental Setup.	69

5.7 Sealant Application Experimental Setup.	70
5.8 Global X and Y Localization Accuracy Given Distance From Start - Clearpath Robotics' Husky.	72
5.9 Global ψ Localization Accuracy Given Distance From Start - Clearpath Robotics' Husky.	73
5.10 Global X and Y Localization Accuracy Given Distance From Start - Fetch Robotics' Fetch.	74
5.11 Global ψ Localization Accuracy Given Distance From Start - Fetch Robotics' Fetch.	74
5.12 Augmented Reality Accuracy Given Specific Start Conditions - Primesense Carmine 1.09.	76
5.13 Augmented Reality Accuracy Given Specific Start Conditions - Microsoft Kinect V2.	77
5.14 Drilling Operation Accuracy Given Distance From Feature Point.	79
5.15 Sealant Application Accuracy Given Distance From Feature Point.	80

List of Tables

3.1	Real World Error Estimation for ROS Available SLAM Algorithm. Adapted from [54].	14
3.2	Comparison Between ROS Available Moveit! Inverse Kinematic Plugins. Adapted from [76].	25
5.1	Global Localization Accuracy Given Distance From Start - Clearpath Robotics' Husky.	72
5.2	Global Localization Accuracy Given Distance From Start - Fetch Robotics' Fetch.	73
5.3	Augmented Reality Accuracy Given Specific Start Conditions - Primesense Carmine 1.09.	75
5.4	Augmented Reality Accuracy Given Specific Start Conditions - Microsoft Kinect V2.	76
5.5	Drilling Operation Accuracy Given Distance From Work Surface.	78
5.6	Sealant Application Accuracy Given Distance From Work Surface.	80

Chapter 1

Introduction

1.1 Background

In order for manufacturing companies to remain competitive, their production systems need to be able to constantly evolve and to accommodate for the market's changing demands [1]. Currently, manufacturing is undergoing a paradigm shift from "mass production" to "mass customization" [2], [3], [4], [5]. As customers demand more customizable products, traditional "mass production" becomes untenable. While mass production focuses on producing large volumes of identical items with high efficiency, mass customization focuses on producing highly variable products, while maintaining the necessary production volume [6].

Robots are widely used in the manufacturing industry to perform tasks which are dumb, dangerous, dull and/or dirty. Consequently, industrial robotics forms an essential part of the manufacturing backbone [4], [7]. Industrial robots tend to be used as stationary units, which require significantly more time and capital to install and integrate than that of other machines due to the multiple aspects that must be configured. As a result, parts need to be produced continuously at the maximum rate possible in order to ensure an adequate Return of Investment (ROI). This contradicts the need for fluctuation of batch sizes due to product variability and the need to keep total production volume unchanged [2].

Unlike industrial robots, mobile robots have the capability of moving around in their environments. Thus, if these two types of robots were combined, they would have the dual advantage of the mobility offered by the mobile platform and the dexterity offered by the manipulator. Consequently, these robots, which are commonly referred to as autonomous industrial mobile manipulators (AIMMs), enhance the capability of traditional industrial robots and mobile robots [8]. AIMMs can be quickly moved and adapted to varying industrial needs in order to provide drastically new possibilities to manufacturing industries [7]. Contrary to the traditional stationary and pre-programmed industrial robots, AIMMs

can be used to provide assistance at multiple locations. They are able to provide highly flexible logistical opportunities and can improve productivity by providing assistance in time-consuming, dangerous, or straining situations. This constitutes a very different use case for AIMMs than the one for traditional fixed industrial robots [9]. Thus, AIMMs could be used to seek an optimum balance between traditional automation, which can only be reconfigured with great difficulty, and manual labor, which is very flexible but not economically viable for large scale production, with the benefits of a compromise between efficiency and flexibility [7], [1]. Thereby, creating a transformable production system, which retains a high degree of automation but also a high degree of flexibility as seen in Figure 1.1.

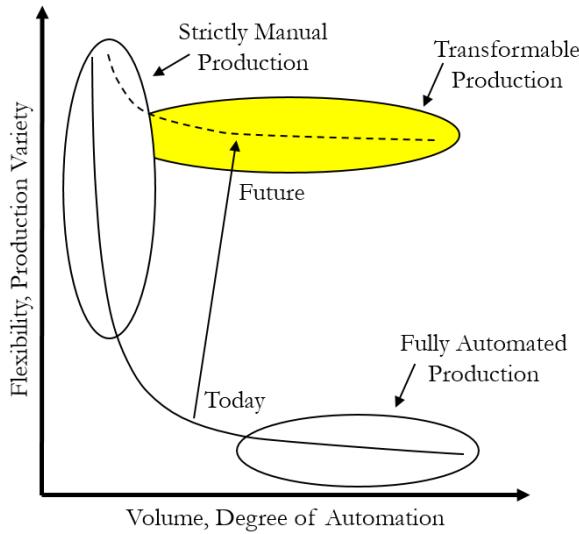


Figure 1.1: Vision for future transformable production systems. Adapted from [10].

The manufacturing industry is not the only industry that could benefit from the introduction of AIMMs. Unlike the substantial benefits seen in the manufacturing industry through the use of fixed industrial robotics, adoption of automation and robotics in construction (ARC) has lagged far behind [11]. Consequently, when compared with other industries, construction has seen a significant decrease in productivity, as well as an increase in workplace injuries/fatalities over the last several decades [12]. Several technical complexities inherent in construction have hindered the development and adoption of field construction robots [13]. However, through the capitalization of advances made in other industries, such as manufacturing, ARC can quickly close this gap. Through the adoption of these advances, dangerous and/or mundane repetitive tasks to be accomplished autonomously, thus increasing productivity and decreasing workplace injuries/fatalities [11]. However, ARC faces unique challenges when compared to other industries. Unlike manufacturing environments, which are tightly controlled, typical construction sites tend to lack structure and are continuously evolving. In addition, the reversal in relationship between the part and

AIMM's manipulator has dramatically increased the complexity of the problem to be solved [13].

1.2 Objectives

The objectives of the thesis are to:

1. Navigate autonomously to specified operation start location while avoiding static and dynamic obstacles.
2. Detect a fiducial marker and populate the specific operation's feature locations in the robot's map frame.
3. Refine feature locations to ± 1 mm.
4. Simulate drilling and sealant operations.

1.3 Summary of Original Contributions

The original contributions within this thesis are:

1. A multistage localization approach which allows for tasks with an operational scope outside the range of the robot's manipulator to be completed without having to recalibrate the position of the end-effector each time the robot's mobile base moves to another position.
2. Implementation of this multistage localization approach to accomplish two tasks which are prevalent in both manufacturing and construction industries: drilling and sealant application operations.

1.4 Outline

Chapter 2 presents where each component of the multistage localization technique fits into current literature, as well as why specific techniques were chosen. Chapter 3 briefly explains all concepts needed to understand the work presented in the chapters thereafter including the Robot Operating System (ROS), Simultaneous Localization and Mapping (SLAM), localization and path planning, manipulation, task execution using state machines, and the computer vision algorithms implemented. Chapter 4 thoroughly explains each

component of the multistage localization approach presented in this paper and expounds upon the implementation of this approach to accomplish two tasks which are prevalent in both manufacturing and construction: drilling and sealant application operations. Chapter 5 validates each stage of the multistage localization approach and system implementation presented in Chapter 4 through testing on two different ROS enabled robotic platforms: the Clearpath Robotics' Husky and Fetch Robotics' Fetch. In addition, it discusses methods of improvement for each stage of the approach. Chapter 6 summarizes the work presented and proposes future work.

Chapter 2

Literature Review

The purpose of this chapter is to present where each component of the multistage localization technique fits into current literature. In addition, it will explain why specific techniques were used over others.

2.1 Localization and Mapping For Autonomous Mobile Manipulators in Manufacturing and Construction

Several different approaches exist to localize a mobile manipulator's base within a large industrial or unstructured environment. Several current common techniques used within manufacturing facilities use the triangulation of artificial landmarks placed at specific locations throughout the facility. Ronzoni et al. and Loevsky et al. both used a laser sensor and retro-reflective stripes strategically placed in known locations throughout a facility in order to localize a mobile robot. They were able to achieve an accuracy of approximately 1.5 cm [14], [15]. Zhang et al. and Okuyama et al. proposed placing fiducial 2-D marker on the ceiling, which can be read by the robot's onboard camera system in order to determine its position to within approximately 6 cm [16], [17]. Another form of triangulation proposed is through the triangulation of radio frequencies, such as RFID and Wifi Localization. DiGiampaolo et al. used a Kalman filter to estimate the location of the robot using the robot's odometry and RFID tags located on the ceiling. This approach achieves an accuracy of approximately 4 cm [18]. Park et al. used a particle filter to estimate the location of the robot using the robot's odometry and RFID tags embedded in the floor. This approach was able to achieve an accuracy of approximately 2.5 cm [19]. Youssef presented the Horus WLAN localization, a RSSI-based fingerprinting scheme, while Gao presented a WLAN fingerprinting technique dubbed DeepFi which used Channel State Information (CSI) and deep learning. Horus was able to achieve an accuracy of approximately 1.55 m, while DeepFi was able to achieve an accuracy of .95 m [20], [21]. Other popular methods include

following magnetic or chemical tape or an inductive wire [22]. However, many of the methods mentioned require making significant modifications to the manufacturing facility. In order to limit the modification necessary, several simultaneous and localization (SLAM) techniques exist that allow a mobile robot to create and/or update a map of a facility while localizing itself within that environment. Kelly demonstrated the utility of applying SLAM to large manufacturing facilities in 2000 [23]. In recent years, optimization based smoothing SLAM techniques have proved more efficient, scalable, and robust than that of filtering SLAM techniques while still retaining a relatively high degree of accuracy, less than 8 cm [24]. Due to the fact that SLAM techniques require no modification of the environment and are able to either build or modify a 2-D or 3-D map of the environment, while retaining a high level of accuracy, they are the building block for the multistage localization presented in this paper. Specifically, a 2-D map will be made using an optimization based smoothing SLAM technique while localization within this map will be accomplished via adaptive Monte Carlo localization (AMCL).

2.2 Task Association and *A Priori* Knowledge for Mobile Manipulators

Fiducial makers have a wide variety of uses due to their tri-ability of being easy to detect under most circumstances, able to be precisely localized, and the ability to either associate data with a specific fiducial maker's identity or encode data. As mentioned above, Zhang et al. and Okuyama et al. placed fiducial 2-D markers on the ceiling, which were then read by the robot's onboard camera system in order to determine it's position to within approximately 6 cm [16], [17]. In order to allow social robots to permeate our daily lives, Limosani et al. proposed the use of a set of environmental tags, composed of AR Tags and QR codes, which encoded the specific address to a cloud service from which a mobile robot could download a map of the local areas and all other necessary information [25]. Feng et al. used April Tags to autonomously identify building components, such as block or brick; so that, a manipulator could be used to pick up specific ones and assemble them into a pre-designed modular structure [13]. In addition, they have been used in order to calibrate a mobile robot to workstations as in [26]. In addition to using a fiducial maker to calibrate a mobile robot to a workstation, they can also be used to specify feature locations at that workstation as they are in this paper.

2.3 Feature Localization Techniques for Mobile Manipulators

Depending on the complexity and required tolerances of a task, a mobile robot may or may not need to calibrate themselves with respect to the work surface. Simple pick and place operations, as seen in [27] do not require the robot to calibrate itself with respect to the work surface. Through the use of feedback from a camera or torque sensor, the robot is able to perform its purpose without calibration. However, many operations in manufacturing require high tolerances. Consequently, the mobile manipulator must calibrate itself with respect to the environment; so that, the robot implicitly knows where features, such as buttons or knobs, are located without having to explicitly detect these features with high tolerances [1]. Several methods exist to do this. Haptic calibration involves measuring location on the work surface in the x, y, and z directions by moving the end-effector until it comes in contact with prominent points on the work surface itself. This calibration can be done in approximately 30 seconds and has an accuracy of ± 1 mm [28]. Visual high-precision calibration involves taking multiple images of a calibration plate, which is fixed to the work surface, by a camera on the end-effector. This calibration can be done in approximately 60 seconds and has an accuracy of ± 0.1 mm [29]. Visual high-speed calibration involves taking a single image of a calibration plate, which is fixed to the work surface, as well as three depth measurements by a camera and laser sensor respectively. This calibration can be done in approximately 10 seconds and has an accuracy of ± 1 mm [29]. QR calibration involves capturing a RGB and depth image of a QR code, which is fixed to a work surface, using an RGB-D camera, whose location is known precisely in relation to the end-effector. This calibration method can be completed in approximately 1 second and has an accuracy of ± 4 mm [26]. While these calibration techniques allow for high tolerances, these calibrations are done with respect to the end-effector. Therefore, if the reach of the manipulator is insufficient to complete the entire operation, such as drilling and sealant application operations in particular, the robot must move and then recalibrate the manipulator with respect to the work surface. Consequently, calibration plates or fiducial markers must be placed at set distance along a project, which requires significant overhead. This paper proposes a multistage localization approach, which localizes the mobile robot within a map, uses a fiducial marker, which can easily be replaced with a calibration plate, to localize an operation's features in the map frame verses with respect to the end-effector. This significantly degrades the localization of the features. However, through the addition of robust feature detection, millimeter to sub-millimeter accuracy will be maintained while allowing for mobility of the mobile base.

2.4 Summary

The purpose of this chapter is to present where each component of the multistage localization technique fits into current literature, as well as why specific techniques were used over

others. The following chapter will present in depth all of the concepts needed to understand the multistage localization approach; so that, anyone can understand the work presented regardless of background.

Chapter 3

Fundamentals of Autonomous Robotics

The purpose of this chapter is to briefly explain all of the concepts needed to understand the work presented in the chapters thereafter. The following subsections will explain the basic concepts of the Robot Operating System (ROS), Simultaneous Localization and Mapping (SLAM), Localization and Path Planning, as well as Manipulation. In addition, Task Execution using State Machines and the specific Computer Vision algorithms implemented will be expounded upon.

3.1 ROS Concepts

The Robot Operating System (ROS) [30] is an open-source set of software libraries and tools that aim to simplify the task of creating robotics applications that can be used across a wide variety of platforms. ROS was originally developed in 2007 by Willow Garage as an extension of Switchyard, a collection of robotics software developed by the Stanford Artificial Intelligence Laboratory in support of the Stanford AI Robot (STAIR) and Personal Robotics (PR) projects. The first distribution of ROS, Box Turtle, was released in 2010. Currently, ROS has had ten major releases. The most current being Kinetic Kame Turtle, which was released on 23MAY16. In addition, ROS boasts tens of thousands of users around the world ranging from hobbyists and researchers to the commercial and industrial industries, as well as hundreds of packages which provide everything from hardware drivers to algorithms for autonomous navigation and manipulation [31]. ROS's large support base acts as a force multiplier, which allows individuals, labs, and company to concentrate on one particular aspect while capitalizing on work that has already been done.

3.1.1 ROS Communication

ROS uses a name server, called the ROS Master, to maintain a list of nodes and available topics. Nodes communicate with the master server using the XML-RPC protocol, while peer-to-peer communications generally use TCP/IP sockets through the TCPROS protocol. Figure 3.1 shows a simple diagram of two ROS nodes communicating with messages and service topics. In addition to the concepts of messages and services, ROS also employs actions. Actions are similar to service calls, but are designed for long duration tasks that are capable of providing feedback. These communication interfaces provide ROS a great deal of flexibility for robotic applications [32], [33].

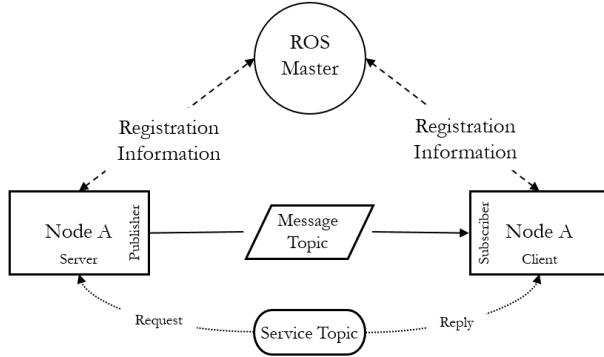


Figure 3.1: Simple ROS Node Flowchart. Adapted from [33].

3.1.2 Rigid Body Transformations

Figure 3.2 shows a purple dot, which represents a point in space. The dot's coordinates in frames a and b are different. A rigid body transform, which can be performed using Equation 3.1, can be used to convert one set of coordinates to another coordinate frame.

$$x_a = T_b^a x_b \quad (3.1)$$

T_b^a is equal to Equation 3.2.

$$\begin{bmatrix} R_b^a & t_b^a \\ 0^T & 1 \end{bmatrix} \quad (3.2)$$

R_b^a is the rotation matrix, which performs the rotation part of transforming frame b into alignment with frame a , while t_b^a is the translation matrix, which performs the translation part of transforming frame b origin to frame a [34].

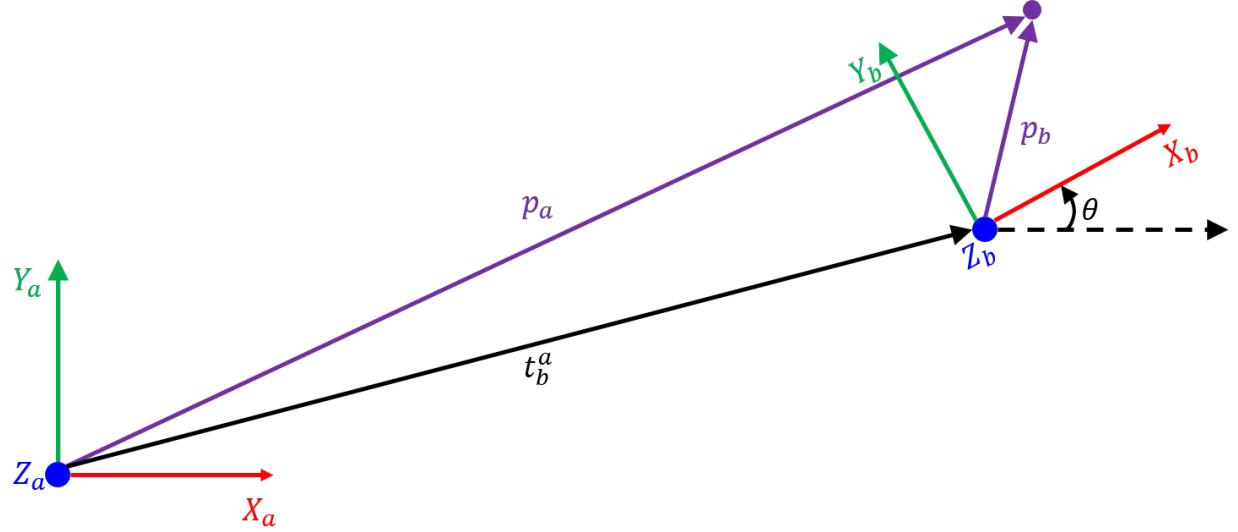


Figure 3.2: Conversion from Coordinate Frame A to B.

A robotic system, such as the Clearpath Robotics' Husky and the Fetch Robotics' Fetch shown in Figure 3.3, typically has many three dimensional coordinate frames that change over time as the robot performs different functions. ROS's TF [35] and TF2 [36] packages keep track of coordinate frames and allow for data to be easily converted between these coordinate frames using rigid body transforms. These coordinate frames, as well as the robot's links/joints they correspond to and the rigid transformations between them are set up in the Robot's Unified Robot Description Format (URDF) File [37].

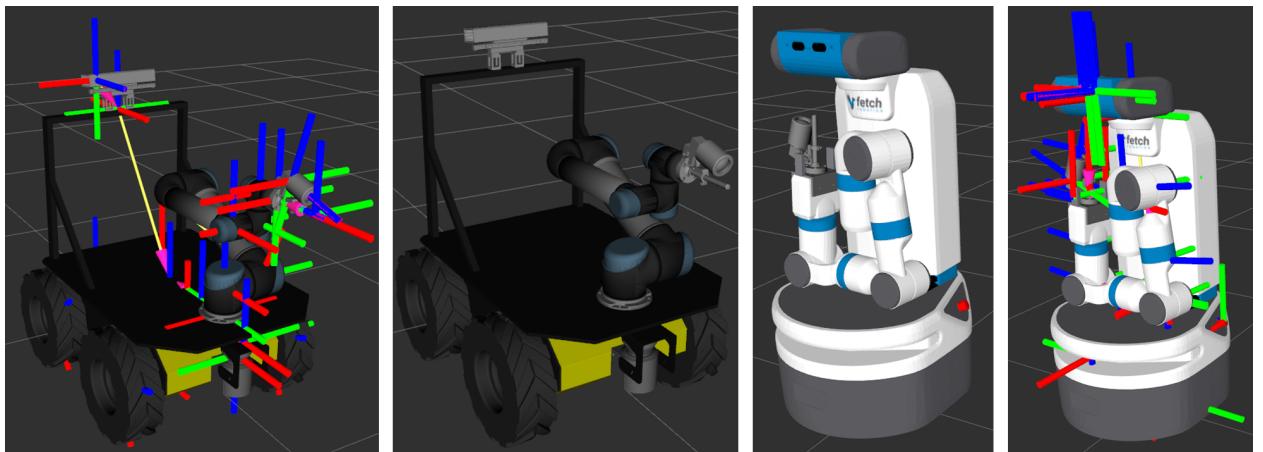


Figure 3.3: Robot Model with TFs.

3.2 Simultaneous Localization and Mapping Concepts

In order for a mobile robot to operate and perform intricate tasks in a complex, GPS-denied environment, such as that of manufacturing facility or construction site, without modifying said environment, the robot must be able to create an accurate map of its environment while simultaneously localizing itself within this map using only on-board sensors. Simultaneous localization and mapping (SLAM) is the problem of building and/or updating a map of an unknown environment while simultaneously localizing the robot within this map [38]. SLAM was pioneered in the early 1990s by Hugh F. Durrant-Whyte and John J. Leonard [39], who based their work on research done by Smith and P. Cheeseman in the mid to late-1980s [40], [41].

Several techniques exist to solve the SLAM problem. Most of these techniques can be categorized into two main paradigms: filtering and optimization-based smoothing [42], [24]. Filtering techniques model the SLAM problem as an incremental state estimation, where the state of the system is composed of the robot's current pose and the map. These estimates are refined at each step by incorporating current sensor measurements. Due to their incremental nature, these SLAM techniques are typically referred to as on-line SLAM approaches. Popular filtering SLAM techniques include the extended Kalman filter, particle filter, and information filters. Filtering SLAM techniques have been used widely in past years due to their ability to model different sources of noise and their effect on sensor measurements. However, in recent years optimization-based smoothing techniques have proven to be more efficient, scalable, and robust than that of filtering techniques [24]. Unlike filtering techniques, optimization-based smoothing techniques estimate the robot's entire trajectory and the map. Due to the fact that the final map is based off the robot's entire trajectory and world features instead of the most recent pose and map, these techniques are known as full SLAM approaches. These SLAM techniques incorporate a graph-based structure, where graph nodes represent the robot's pose and world features, while edges represent a spatial constraint relation between two robot poses given by sensor measurements [43]. The graph is optimized using error minimization techniques, such as least-squares, in order to refine the robot's trajectory and map.

In addition to a variety of techniques which can be used to solve the SLAM problem, a wide range of sensors can also be used. Typically, sensors used include LIDARs, stereo cameras, monocular camera, and RGB-D sensors.

Figure 3.4 shows the generic framework for solving the SLAM problem. In the front-end, raw sensor inputs are processed in order to extract features and perform scan matching; so that, necessary parameters and/or constraints, as well as the system's state can be estimated. The system's state and necessary parameters and/or constraints are sent to the back-end of the SLAM algorithm where the system's state is refined and returned based on the parameters and/or constraints.

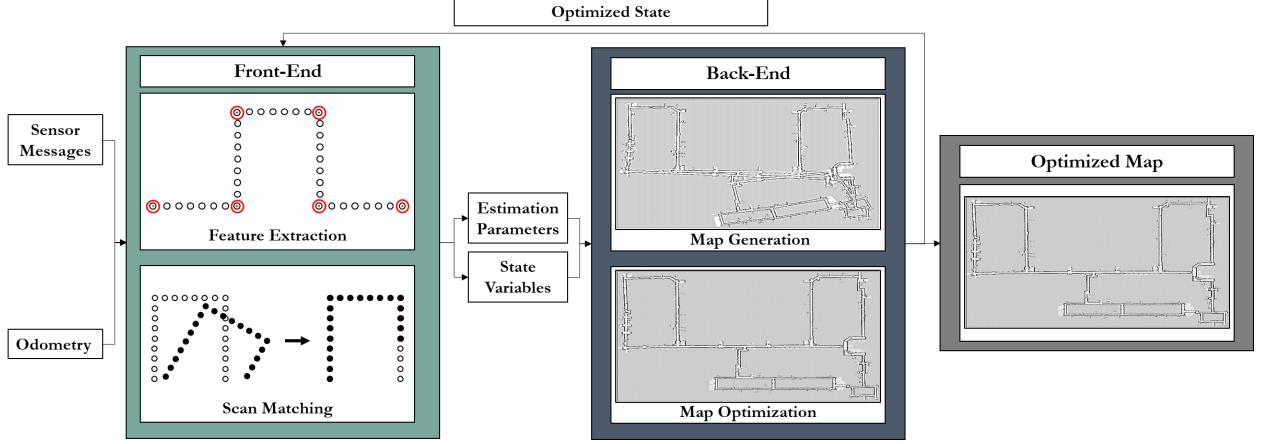


Figure 3.4: Overview of SLAM Framework.

ROS includes several 2-D SLAM packages. These packages are used to build accurate 2-D occupancy grid maps, an example of which is shown in Figure 3.5. These maps are then used to localize the robot within its environment, as well as to plan and execute appropriate trajectories in order for the robot to reach its destination.

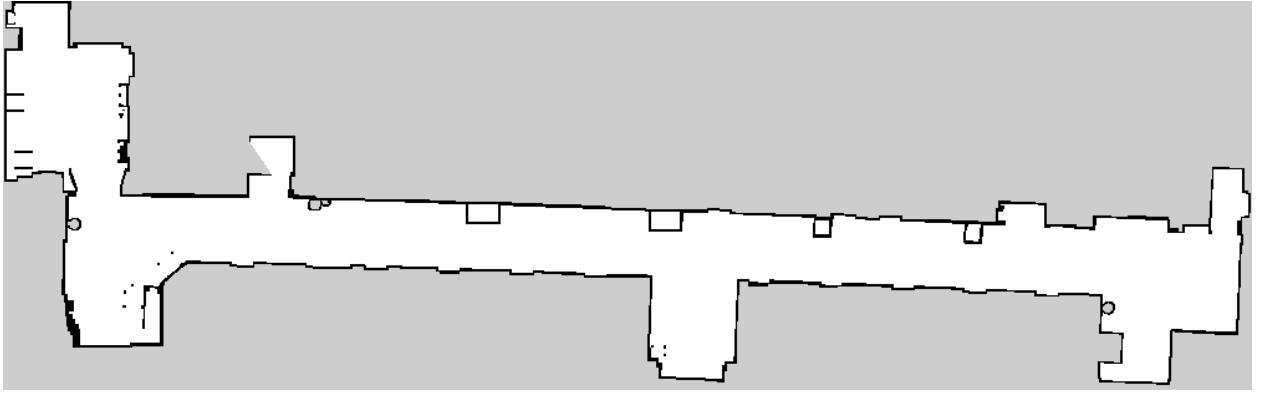


Figure 3.5: Example of 2-D Occupancy Grid Map Produced.

Five of the most common 2-D SLAM ROS packages are HectorSLAM, GMapping, KartoS LAM, CoreSLAM, and LagoSLAM. HectorSLAM [44] is neither a filtering nor an optimization-based smoothing technique, but rather relies solely on robust scan matching [45]. CoreSLAM [46] and GMapping [47] are filtering SLAM techniques. Both CoreSLAM and GMapping utilize a particle filter. While CoreSLAM employs a very simple particle filter [48], GMapping uses a more complex and efficient Rao-Blackwellized particle filter [49]. Both KartoS LAM [50] and LagoSLAM [51] are optimization-based smoothing SLAM techniques.

However, KartSLAM, which was developed by SRI robotics, uses a highly-optimized and non-iterative Cholesky matrix decomposition for sparse linear systems, known as Sparse Pose Adjustment (SPA) [43], [52], while LagoSLAM uses LAGO optimizer developed by Carlone et al. [53].

Figure 3.6 and Table 3.1 show the aforementioned 2-D SLAM ROS packages' real world performance based on testing done by Santos et al. [54]. KartSLAM achieved the smallest error demonstrating the robustness of its sparse pose adjustment (SPA) and that of full SLAM techniques in general. As a result, KartSLAM was used to generate Figure 3.5 as well as the maps used during testing of the multistage localization approach presented in this paper.

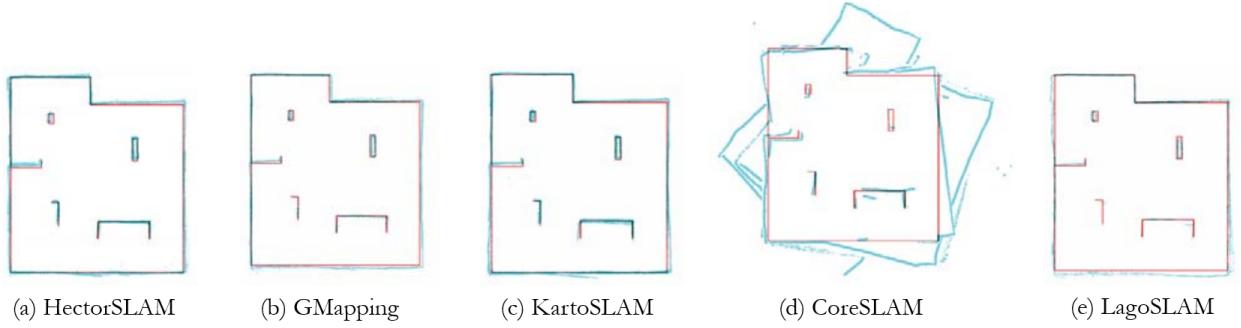


Figure 3.6: Real World Performance Analysis of ROS Available SLAM Algorithms. Adapted from [54].

Table 3.1: Real World Error Estimation for ROS Available SLAM Algorithm. Adapted from [54].

Real World Experiments				
HectorSLAM	GMapping	KartoSLAM	CoreSLAM	LagoSLAM
1.1972	2.1716	1.0318	14.75333	3.0264
0.5094	0.6945	0.3742	7.9463	0.8181
1.0656	1.6354	0.9080	7.5824	2.5236

3.3 Localization and Path Planning Concepts

ROS's Navigation Stack [55] is a collection of packages, which uses odometry and laser scan data, as well as a goal position and orientation in order to output the velocity commands

needed to reach the specified goal. Figure 3.7 shows an overview of how the individual packages work together to achieve this objective. The Map_Server node [56] loads a previously generated two-dimensional grid map. Once the AMCL [57] node receives the map, odometry, and laser scan data, it is able to localize the robot within the provided map, using the Adaptive Monte Carlo Localization technique from which it gets its name. The Move_Base [58] node maintains both global and local planners and costmaps. Information about obstacles in the world are stored in these costmaps. The global costmap is used for long-term planning, while the local costmap is used for short-term planning and obstacle avoidance. The global planner computes an optimal path to the goal given the starting state of the robot and the global costmap, while the local planner computes shorter trajectories given the current state of the robot and the local costmap. Once a path is developed, the Move_Base node outputs the necessary velocity commands needed to reach the specified destination [55].

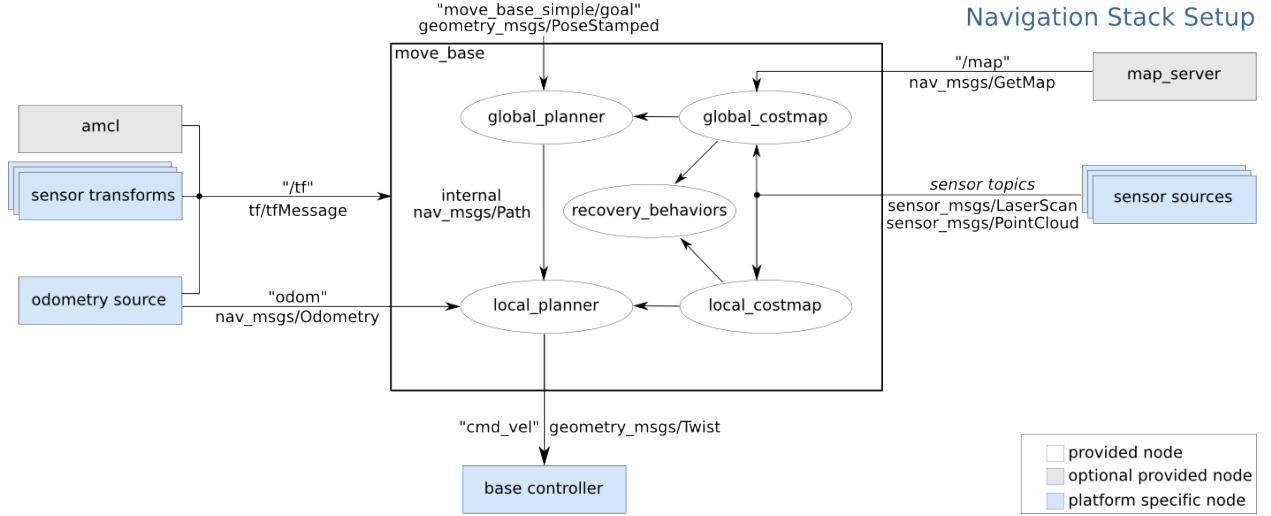


Figure 3.7: ROS Navigation Stack setup [55].

After using SLAM to build a two-dimensional occupancy grid map, shown in Figure 3.5, it becomes crucial to accurately localize the robot within this predefined map; so that, the robot can both plan and execute appropriate trajectories to reach its destination. Localization involves estimating the position and orientation of the robot, known collectively as pose, while the robot moves throughout its environment. One routine localization technique used involves tracking the robot from an initial known starting pose. Through the measurement of wheel rotation and the integration of accelerations provided by an inertial measurement unit (IMU), the distance traveled by the robot from the initial position can be calculated and the robot's pose in the map estimated with some certainty. However, these methods do not account for wheel slippage or measurement error. As a result, the accuracy of the pose estimate will degrade over time. Consequently, a solution which can compensate for

the accumulated odometry error and inaccuracies in the initial starting pose is needed. One accepted solution to this problem is Monte Carlo Localization (MCL), which utilizes a particle filter to keep track of the robot's pose. However, additional options include Kalman Filters and Markov Localization, which employ Gaussian distributions and histograms respectively.

Figure 3.8 depicts MCL using a one-dimensional corridor with a few doors. The robot initially has no information about where it is in this corridor. As a result, the graph of the robot's belief states, which defines the probability of the robot being at a particular position, is drawn from a uniform distribution of discretely sampled positions along the corridor. A measurement update is performed at each step. A measurement update involves convolving the measurement model, the probability of receiving a specific sensor measurement in the corridor, with the belief states to get an updated belief state. The updated belief state is the same as the previous belief state; however, the weight of each particle have been updated based on the sensor reading. At step $k = 1$ the robot senses a door; so, the weights of the particles around each of the three door are increased. At the next step, a motion model update is performed. The odometry indicated that the robot moved forward a specified distance d . As a result, the belief state is updated by moving the particles forward by distance d with noise added to account for the aforementioned odometry errors. It should be noted that the particles at this stage in Figure 3.8 were also resampled, which will be covered in the following paragraphs. The motion model update is followed by a measurement update. The robot again senses a door. As a result, the measurement model is the same as the previous time step. After convolving the current measurement model with the current belief state, the cumulative probability mass is centered at door two indicating that the robot is likely at this location [59].

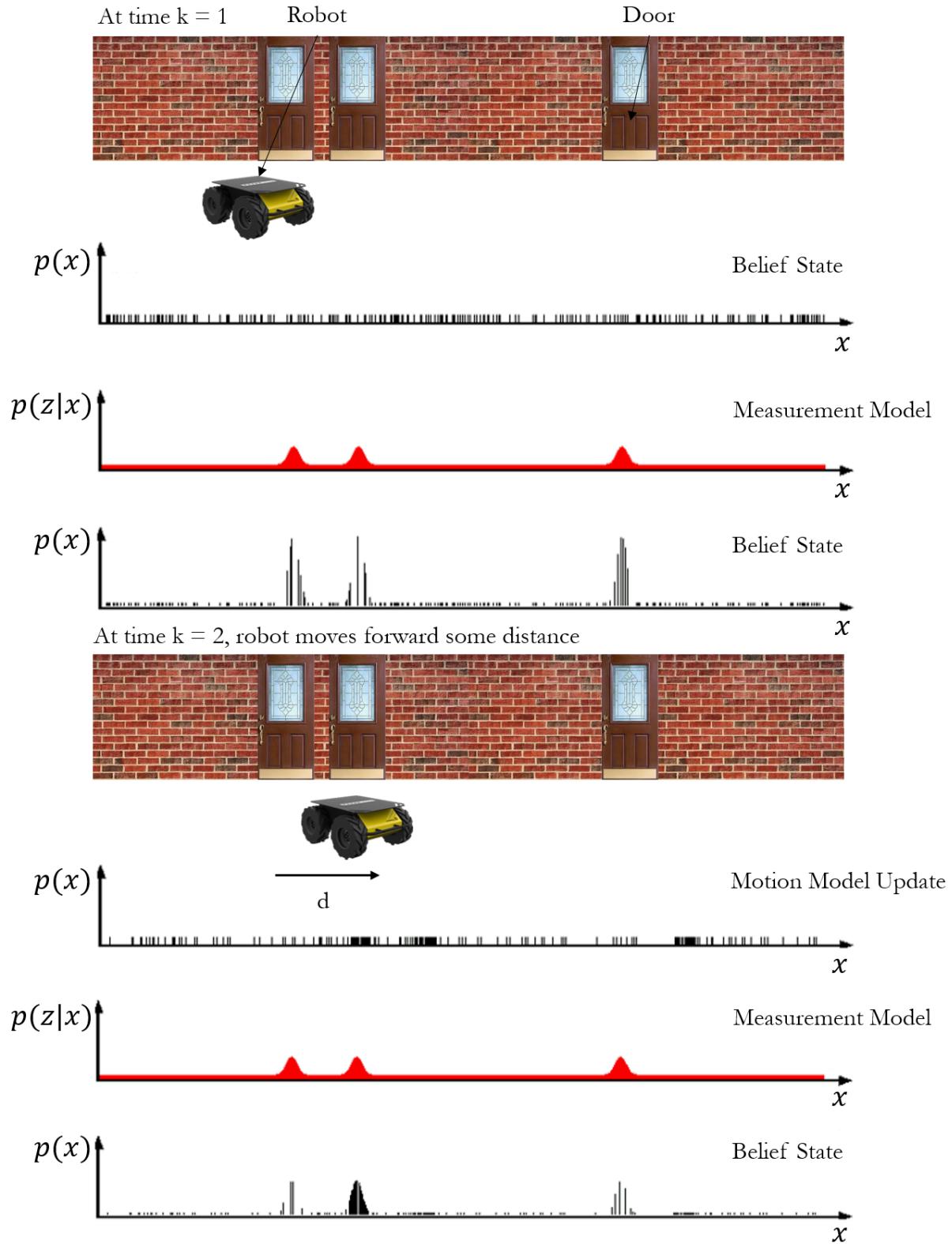


Figure 3.8: One-Dimensional Monte Carlo Localization Example. Adapted from [59].

Two-dimensional MCL follows the same format of a motion model and then measurement update. The initial particles are drawn from the current odometry with added noise. At each step, the particles are updated via the odometry and then corrected via a measurement update. For each particle, the correlation between the two-dimensional occupancy grid map, seen in Figure 3.5, and laser scan is calculated using Equation 3.3, where A is the predefined map, B is the map created by the current laser scan, and \bar{A} and \bar{B} are the mean values of the pixels of both maps respectively. Note that while obstacle pixels are black and have a value of 1, free space pixels are grey and have a value of 0. The x and y values of each pixel are represented by m and n . The particle (pose) with the highest correlations score is chosen as the pose for the current step. The weight of each particle at step k is found by multiplying the particle's weight at step $k - 1$ by its normalized correlation score at step k , as seen in Equation 3.4.

$$s = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \quad (3.3)$$

$$W_k \leftarrow W_{k-1}s \quad (3.4)$$

As weights are multiplied over steps, the particles with consecutive small correlation scores are reduced to very small weight values. As a result, the particle filter eventually has very few effective particles to ensure that good results are produced. Consequently, re-sampling, shown in Figure 3.9, is performed when the number of effective particles becomes too small. This is done by drawing samples close to the particles that have higher weights, indicated by their size. Thus the new sample has a higher density near the positions where the particles with higher weights existed. Each particle after resampling has the same weight. As a result, the particle filter begins from scratch.

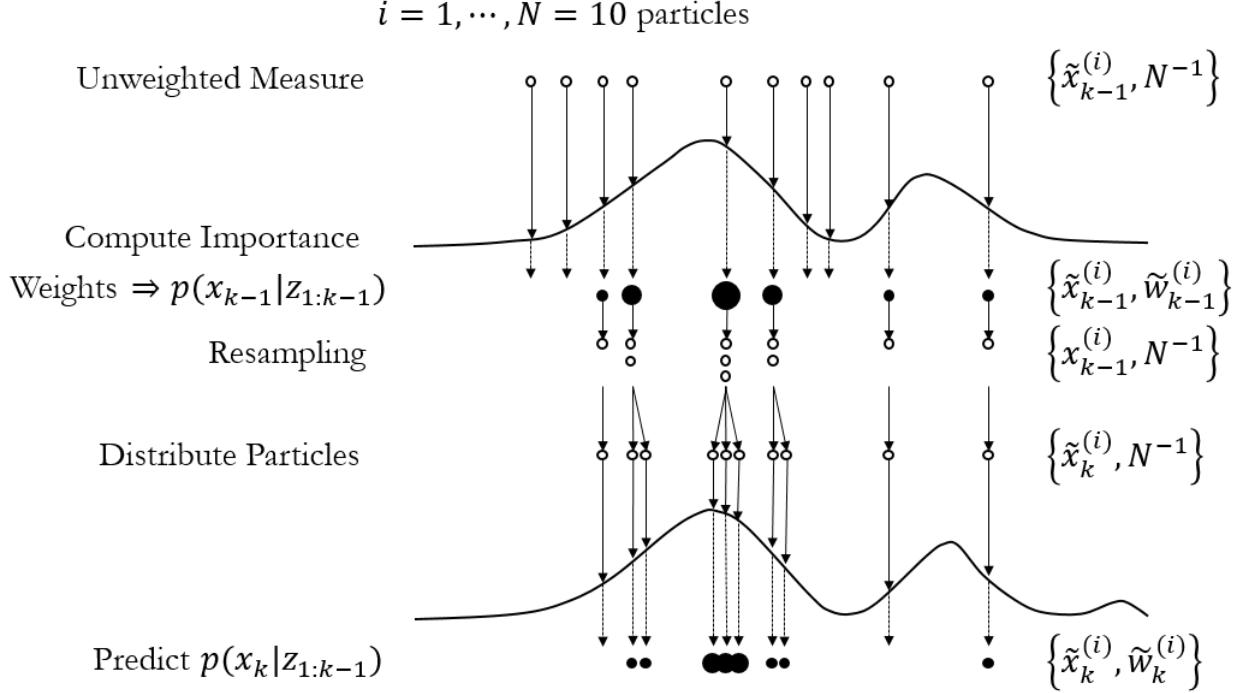


Figure 3.9: Particle Filter Resampling Example

Due to the computational complexity inherent in iteratively calculating the correlation score for each particle, an optimization technique known as *KLD-sampling* is used. *KLD-sampling*, derived from *Kullback-Leibler divergence*, is a technique that determines the number of particles needed such that the error between the sample and true posterior is less than ϵ [59]. KLD-sampling basically controls the number of particles based on the difference in odometry and particle base location.

Figure 3.10 shows how KLD-sampling works. Initially when the position is unknown, the particle cloud is large due to the uncertainty in the position and orientation of the robot. However, as the robot moves, the particles converge and the particle cloud size reduces as KLD-sampling removes the redundant particles and improves computational performance.

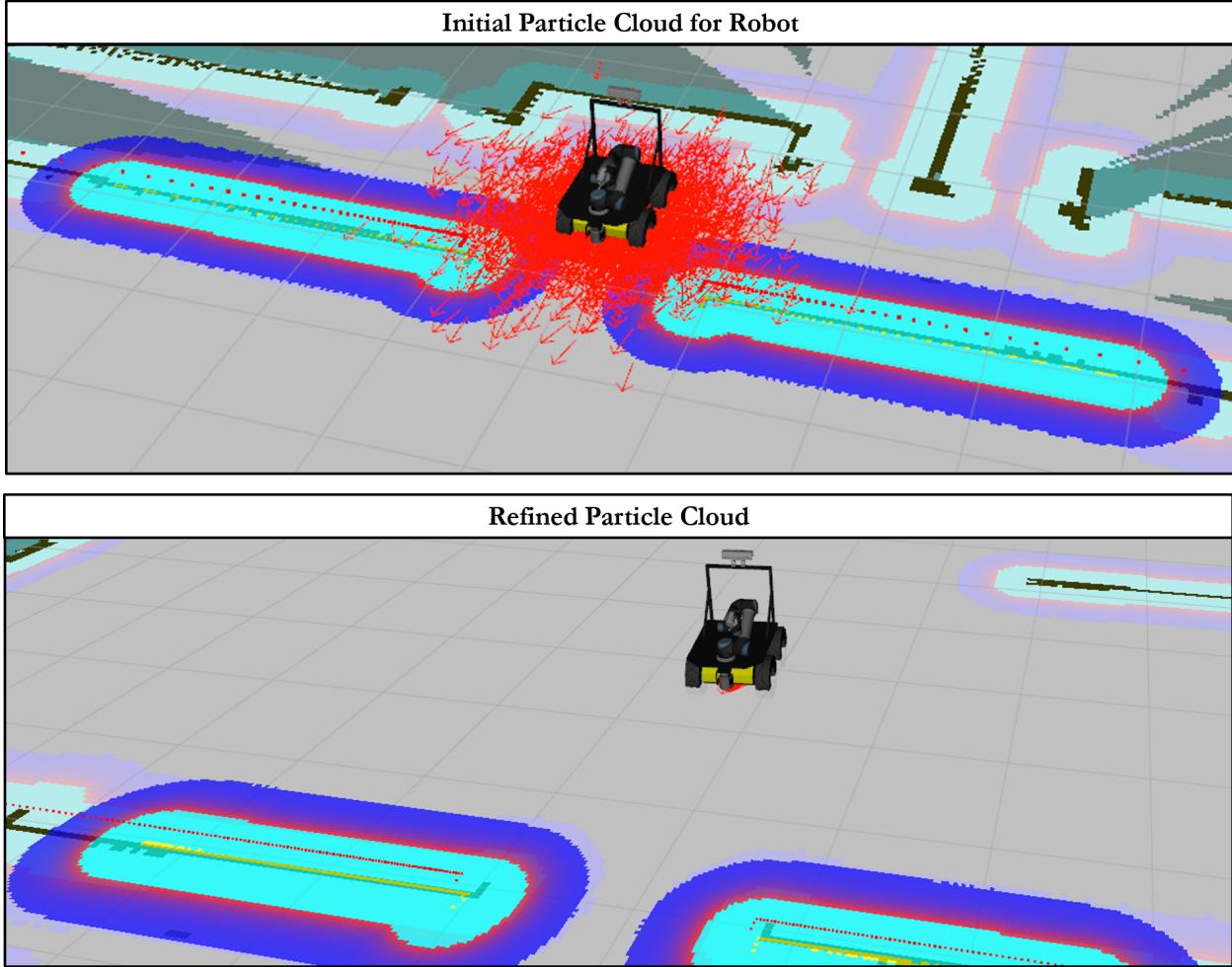


Figure 3.10: Visual of AMCL in RVIZ

After calculating the local and global costmaps, as well as the location of the robot within the given map, the robot must now both plan and execute appropriate trajectories to its destination. Two frequently used techniques are the Dynamic Window Approach (DWA) and Trajectory Rollout. Both sample the space of feasible controls. For a differential drive robot, such as the Clearpath Robotics' Husky and the Fetch Robotics' Fetch, this controls space is 2D and consists of translational and rotational velocities, $\dot{x}, \dot{\theta}$, which are limited by the robot's capabilities. Each sampled velocity is forward simulated from the robot's current position for a short period of time in order to generate simulated trajectories as shown in Figure 3.11. These simulated trajectories are then scored using the cost function in Equation 4.1,

$$C(k) = \alpha Obs + \beta Gdist + \gamma Pdist + \delta \frac{1}{\dot{x}^2} \quad (3.5)$$

where Obs is the sum of grid cell cost through which the trajectory passes (taking into account the robot's actual footprint in the grid); $Gdist$ and $Pdist$ are the estimated shortest distance from the endpoint of the trajectory to the goal and the optimal path, respectively; and \dot{x} is the translation component of the velocity command that produces the trajectory.

The simulated trajectory that minimizes this cost function is chosen. As a result, chosen trajectories tend to keep obstacles at a distance, proceed towards the goal, remain near the optimal path, and have higher velocities [60]. DWA and Trajectory Rollout differ in that Trajectory Rollout samples achievable velocities over the entire forward simulation, while DWA samples only from achievable velocities for just one simulation step [61].

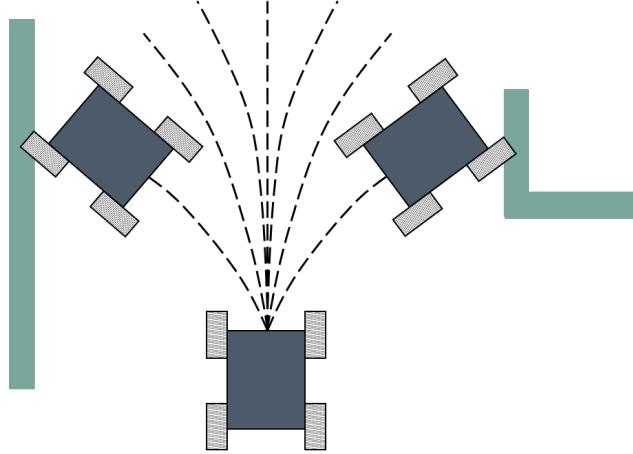


Figure 3.11: Trajectory Rollout Path Planning Framework.

3.4 Manipulation Concepts

ROS's MoveIt! [62] is a collection of packages for mobile manipulation, which incorporate the latest advances in motion planning, manipulation, 3-D perception, kinematics, control, and navigation. Figure 3.12 shows Moveit's overall system architecture. The move_group node integrates all individual packages together in order to provide the user a set of ROS actions and services. The user interface allows the user to interact with the move_group node through the ROS interface by using C++, Python, or RVIZ [63], which is ROS's graphical user interface and 3-D visualization tool. ROS's Parameter Server [64] provides the move_group node the robot's URDF and Semantic Robot Description Format (SRDF) files, as well as Moveit!'s specific configuration files. The SRDF is used to represent information about the robot that is not included in the URDF file, such as a set of links or joints, known

collectively as a group, that make up the manipulator, predefined group states, and a list of links between which collision checking should be disabled [65]. Moveit!'s configuration files set joint limits, as well as kinematics, motion planning and perception parameters. The SRDF file and Moveit!'s configuration files are set up through Moveit!'s Setup Assistant, which is Moveit!'s graphical user interface for configuring any robot for use with Moveit!. The robot interface allows Moveit! to send commands to and receive feedback from the robot. Moveit!'s Planning Scene Monitor, which monitors information from the robot's sensors, is used to maintain a planning scene, which represents the world around the robot and stores the state of the robot itself. This information is used by Moveit!'s Flexible Collision Library, as well as the motion and kinematic planning plugins to plan and execute obstacle free paths for the manipulator [66].

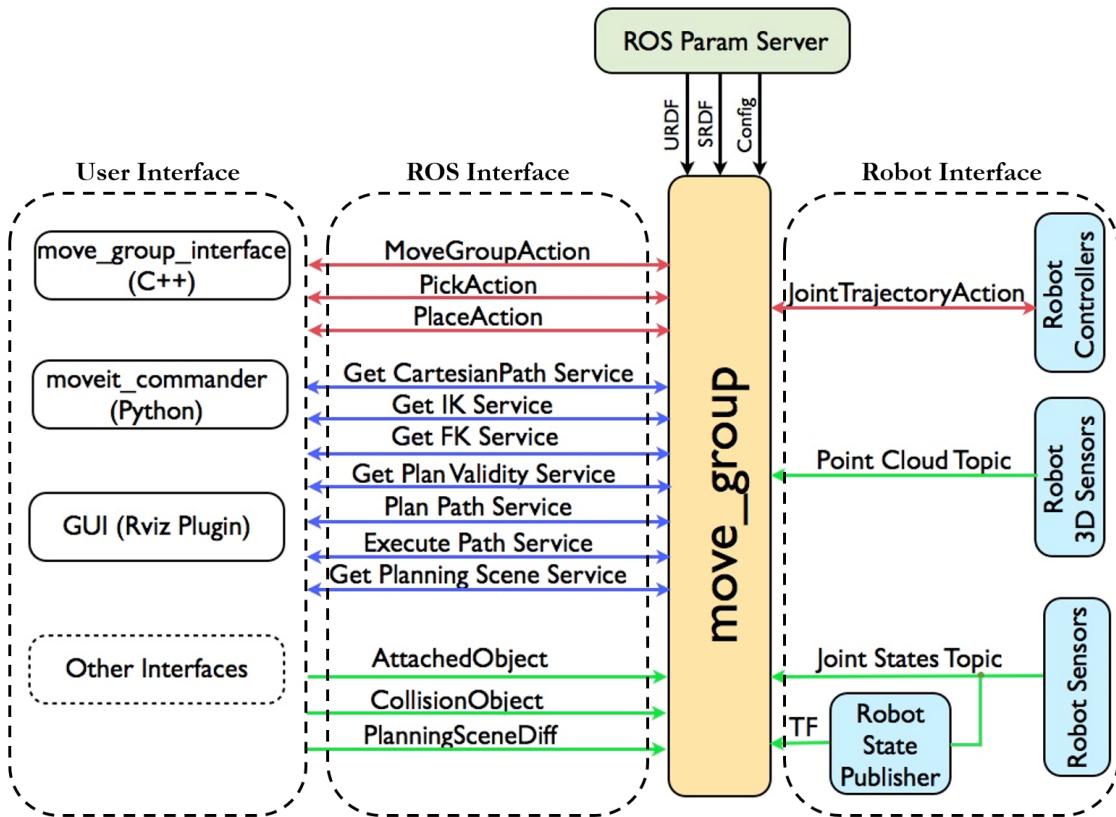


Figure 3.12: Moveit!'s System Architecture. Adapted from [66]

Moveit! allows the user to plan trajectories in either joint space, forward kinematic (FK),

or Cartesian space, inverse kinematics (IK). Forward kinematics uses the manipulator's kinematic equations to compute the position of the end-effector given specific predetermined joint angles [67]; whereas, inverse kinematics uses the manipulator's kinematic equation to determine the joint angles necessary for the end-effector to arrive at the desired position [68]. Figure 3.13 illustrates the difference between forward and inverse kinematics. In most cases, including the multistage localization approach presented in this paper, the target joint angles are not known ahead of time. As a result, a target pose is specified for the manipulator's end-effector in Cartesian space and Moveit!'s IK solver will determine the appropriate joint angles needed in order to reach the desired pose [69].

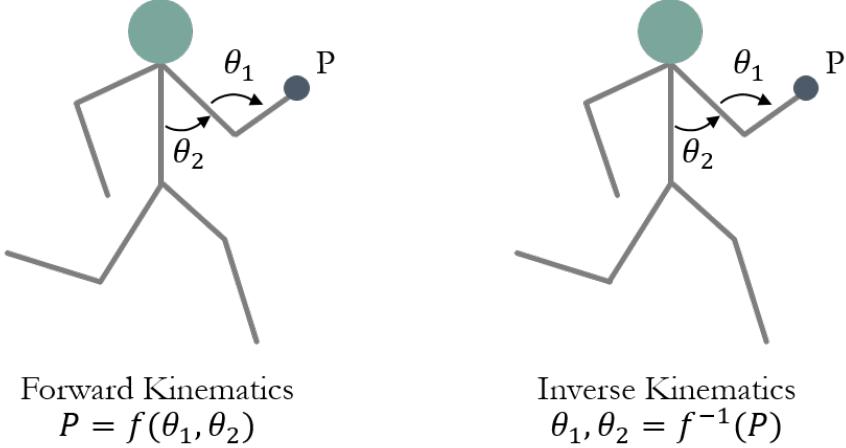


Figure 3.13: Forward and Inverse Kinematics Example

There are two main approaches for solving IK problems: numerical and analytical (closed form) solutions. Due to the recursive nature of numerical solvers, they tend to be slower than analytical solvers and are prone to getting trapped in local minima [70]. However, analytical solvers require that a separate kinematic solver be created for each manipulator and manipulator configuration. In order to ensure cross platform functionality with minimum overhead, only available numerical kinematic solvers were examined. However, OpenRAVE (Open Robotics Automation Virtual Environment) [71], which is software for simulating and deploying planning algorithm, provides a tool called IKFast [72], which allows users to generate custom analytical kinematic solvers, which can produce solutions within approximately four microseconds. Table 3.2 shows three common ROS Moveit! numerical kinematic solver implementations, Orococos KDL [73], as well as the TRACLab's [74] KDL-RR and TRAC_IK [75]. Orococos Kinematics and Dynamics Library (KDL) is a joint-limit-constrained pseudoinverse Jacobian solver, whose convergence algorithms are based on Newton's method. However, KDL suffers from the following issues, which results in the high failure rates and slow average speed seen in Table 3.2.

1. Frequent convergence failures due to manipulator joint limits

2. No action is taken when the search gets trapped in a local minima
3. Lack of tolerance support and/or utilization in the solver itself

The TRACLab’s KDL-RR kinematic solver implementation attempts to solve issues two and three by altering Orocosp’s KDL implementation to detect and “untick” the iterative convergence algorithms, as well as to loop for a maximum time versus a maximum number of iterations. KDL-RR is able to detect local minima by monitoring when the difference between joint angles across iterations becomes approximately zero. When this situation occurs, random seed angles are introduced, which “unstick” the convergence algorithms. In addition, KDL-RR loops for a maximum time versus a maximum number of iterations due to the fact that the maximum number of iterations is an arbitrary number, while the maximum time can be computed based on the size and complexity of each manipulator. In addition, using maximum time makes comparing implementations easier. Table 3.2 shows that KDL-RR had a dramatic increase in solve rate percentage, as well as a modest decrease in runtime compared to KDL. However, KDL-RR doesn’t solve issue one. Consequently, the failure rate is still unacceptably high. TRACLab’s TRAC-IK kinematic solver attempts to solve issue one by using two separate IK solvers concurrently. It uses KLD-RR, as well as a sequential quadratic programming (SQP) IK implementation, which is an iterative algorithm for nonlinear optimization. The SQP implementation, which incorporates the same local minimum detection, is able to deal with issue one. However, SQP can have a much longer solve rate than KDL or KDL-RR. As a result, TRAC-IK implements both of these IK method concurrently and waits for either to converge. As a result, its solve rate ranges between 99.1% and 99.9 % and its average convergence times are well below one millisecond [76].

Table 3.2: Comparison Between ROS Available Moveit! Inverse Kinematic Plugins. Adapted from [76].

Robot	DOFs	IK Technique					
		Orocoss' KDL		KDL-RR		TRAC-IK	
		Solve Rate (%)	Avg Time (ms)	Solve Rate (%)	Avg Time (ms)	Solve Rate (%)	Avg Time (ms)
Atlas 2013 Arm	6	75.54	1.35	97.13	0.39	99.97	0.33
Atlas 2015 Arm	7	75.71	1.50	93.13	0.81	99.18	0.48
Baxter Arm	7	61.07	2.21	89.52	1.02	99.17	0.60
Denso VS-068	6	27.92	3.69	98.13	0.42	99.78	0.38
Fanuc M-430iA/2F	5	21.07	3.99	88.34	0.92	99.16	0.58
Fetch Arm	7	92.49	0.73	93.82	0.72	99.96	0.44
Jaco2	6	26.23	3.79	97.66	0.58	99.51	0.58
LBR iiwa 14 R820	7	37.71	3.37	94.02	0.73	99.63	0.56
KUKA LWR 4+	7	67.80	1.88	95.40	0.62	99.95	0.38
PR2 Arm	7	83.14	1.37	86.96	1.27	99.84	0.59
NASA Robonaut2 'Grasping Leg'	7	61.27	2.29	87.57	1.10	99.31	0.67
NASA Robonaut2 'Leg + Waist + Arm'	15	97.99	0.80	98.00	0.84	99.86	0.79
NASA Robonaut2 Arm	7	86.28	1.02	94.26	0.73	99.25	0.50
NASA Robosimian Arm	7	61.74	2.44	99.87	0.36	99.93	0.44
TRACLabs Modular Arm	7	79.11	1.35	95.12	0.63	99.80	0.53
UR10	6	36.16	3.29	88.05	0.82	99.47	0.49
UR5	6	35.88	3.30	88.69	0.78	99.55	0.42
NASA Valkyrie Arm	7	45.18	3.01	90.05	1.29	99.63	0.61

3.5 Task Execution Concepts

It is relatively straightforward to program a robot to execute an individual action. However, programming a fully autonomous robot that will be expected to select which individual action it will perform depending upon the task at hand and the current conditions isn't so straight forward. The task execution system must be able to rank tasks by priority, break down tasks into subtasks, execute tasks in parallel, monitor conditions and react according, as well as pause and resume tasks at a later time if necessary [69]. A commonly used construct to achieve the aforementioned requirements is a hierarchical state machine.

Finite state machines are used to model control and sequencing in a system [77]. In a state machine, the robot occupies one state, which has set behaviors or tasks associated with it. As long as the robot remains in that state, it will continue to carry out the same behavior or tasks. States are connected together by transitions. When certain predefined conditions are met, a transition is triggered and the system changes from its current state to the target state. State machines are a powerful tool; however, it can be difficult to express some behaviors, such as "alarm behaviors". Imagine an autonomous industrial mobile manipulator (AIMM) that is responsible for drilling specific patterns on different sections of plane wings. This can be easily implemented using a normal state machine. The AIMM will need to navigate to the specific starting location of each wing, determine the appropriate drill locations based

on the associated pattern, and finally carry out the individual drilling operations. This can be easily implemented using a normal state machine. Unfortunately, the AIMM's battery does not provide power indefinitely. When the AIMM's power level drops to a certain level, it will need to stop and navigate to the nearest charging location to be recharged, regardless of what it is doing at the time. When it is fully recharged, it will need to pick up exactly where it left off. The recharging period is an alarm mechanics, which is something that interrupts normal behavior to respond to something important. Representing this in a state machine leads to a doubling the number of states. With only one level of "alarm behaviors", this is no problem; however, if we add levels of "alarm behaviors", the number of states will increase exponentially. So, rather than combining all the logic into a single state machine, the logic can be separated it into several. Each alarm mechanism should have its own state machine, along with the original behavior. These individual state machines are arrange into a hierarchy, so the next state machine down is only considered when the higher level state machine is not responding to its alarm. The nesting of state machines inside another makes what is called a hierarchical state machine [78].

ROS's SMACH [79] package includes a standalone python library for programming hierarchical state machines, as well as a ROS wrapper for integrating the library with ROS topics, services, and actions. Figure 3.2 shows a graphical view of a hierarchical state machine set up using SMACH. This hierarchical state machine simulates a robot performing tasks such as drilling and sealing in a manufacturing environment while having the option to change its tool or recharge if necessary.

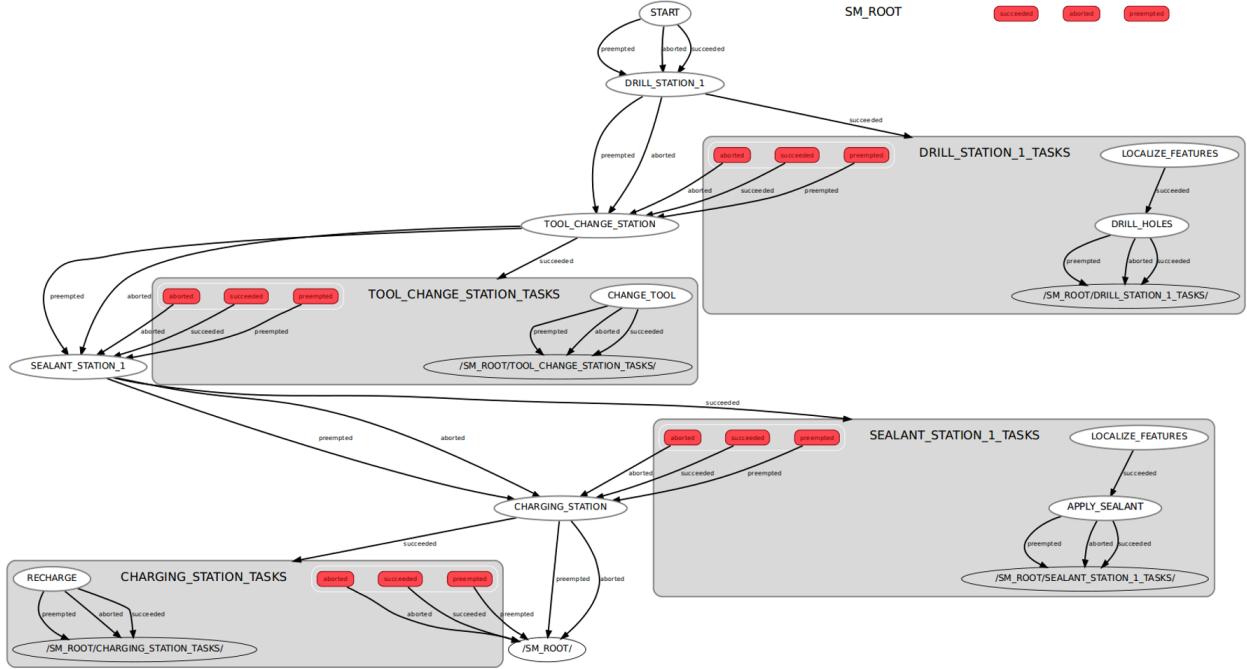


Figure 3.14: Graphical View of State Machine using SMACH

3.6 Camera Concepts

An image is the optical representation of an object illuminated by a radiating source. Figure 3.15 shows a simplified model of how photometric images are formed.

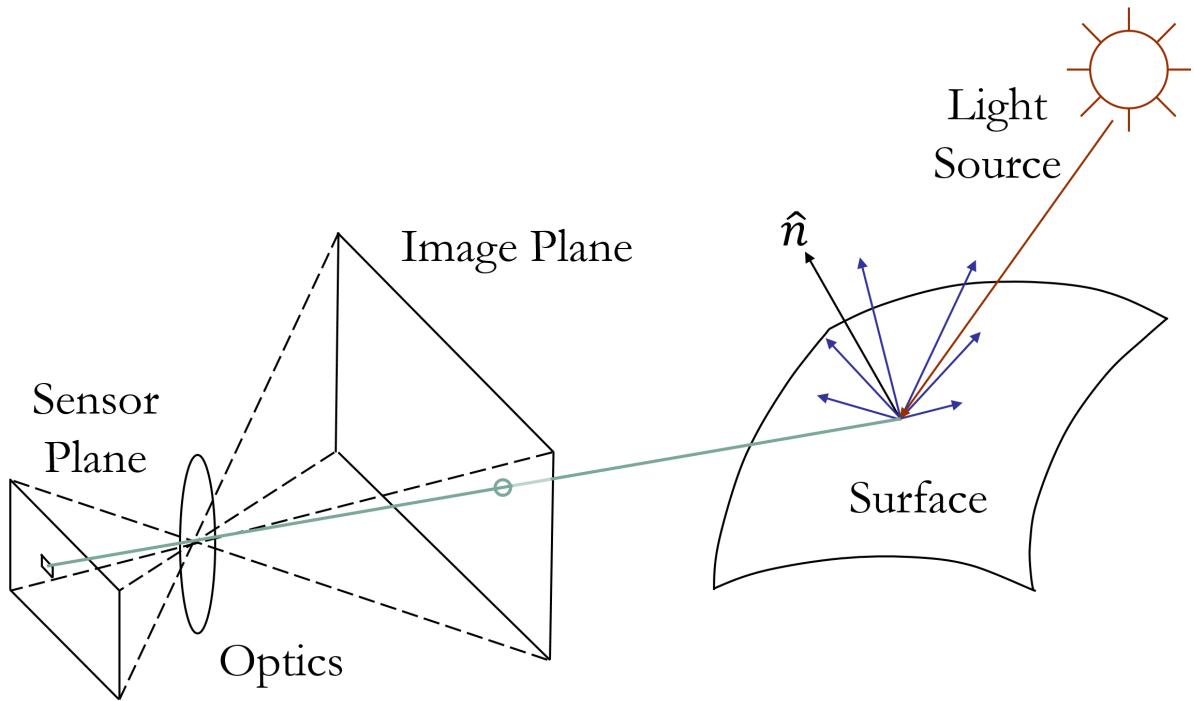


Figure 3.15: Photometric Image Formation. Adapted from [80].

Visible light is emitted by one or more sources, which is then reflected off an object's surface. This reflected light, is the optical image which is the input of a digital image formation systems [81], such as a digital camera, depicted in Figure 3.16.

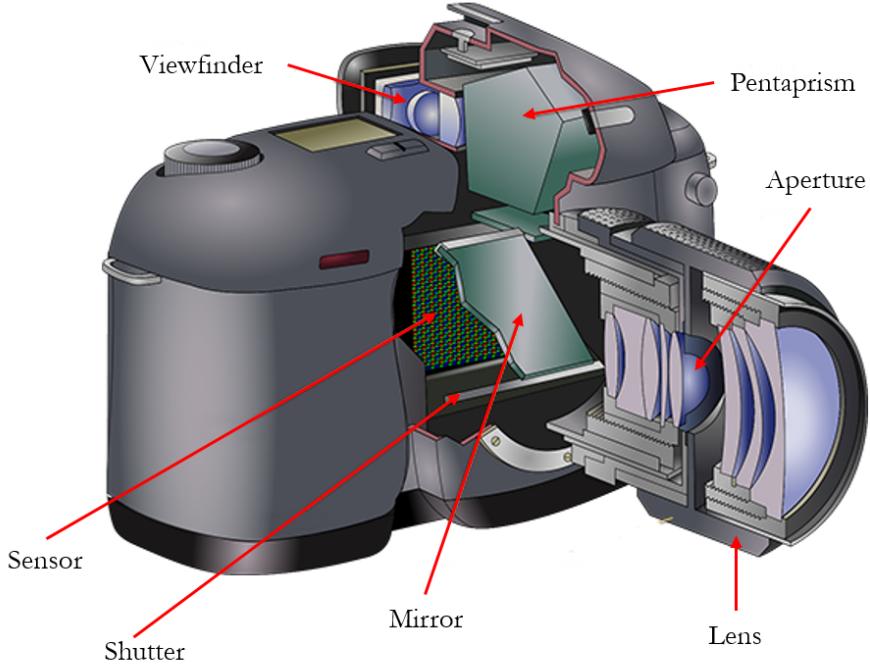


Figure 3.16: Digital Camera Diagram. Adapted from [82].

The light passes through a convex lens which focuses all of the rays through the optical center point of the lens and onto a silicon sensor. This sensor, also referred to as the sensor plane, is essentially a 2-D array, which consists of thousands to millions of solar cells that convert the light into electrons. The accumulated charge of each cell is transmitted to an onboard computer which digitizes this information; so that, a computer can render the digital image of the world [83]. Since the sensor plane is vertically rotated 180°, due to the principle of the pinhole camera model [84], the image plane is introduced for mathematically convenience. The image plane, also known as the focal plane, is a virtual plane that is located in front of the optical center of the camera lens and is equivalent to the sensor plane, in that it represents a projection of the world that is mapped onto a plane. In addition, it accounts for the rotation of the image in the sensor plane.

In order for the multistage localization technique presented in this paper to achieve the desired accuracy, a digital RGB camera with the appropriate sensor size, resolution, focal length, frame rate, and shutter type must be chosen. The focal length, which is the distance between the optical center point of the camera lens to the image plane [85], and sensor size of the RGB camera are used to determine both the angular horizontal and vertical field of view (FOV), which can be calculated using Equations 3.6 and 3.7 respectively,

$$HFOV_{angular} = 2 \tan^{-1}\left(\frac{w_s}{2f}\right) \quad (3.6)$$

$$VFOV_{angular} = 2 \tan^{-1}\left(\frac{h_s}{2f}\right) \quad (3.7)$$

where w_s and h_s are the weight and heights of the sensor, and f is the focal length. These angular FOVs can be converted from angular to distance measurements, using Equations 3.8 and 3.9,

$$HFOV_{distance} = 2d \tan\left(\frac{HFOV_{angular}}{2}\right) \quad (3.8)$$

$$VFOV_{distance} = 2d \tan\left(\frac{VFOV_{angular}}{2}\right) \quad (3.9)$$

where d is distance from camera to work surface. In addition, Equations 3.10 and 3.11, where r_h and r_v are the horizontal and vertical resolution respectively, can be used to calculate the horizontal and vertical ground sample distance (GSD). GSD is the distance between two consecutive pixels centers. Whereas spatial resolution is the area of each pixel, which is a function of the horizontal and vertical GSD [86]. The ground sample distance is the highest accuracy a localization system can achieve. Figure 3.17 shows the effect that GSD can have on the detail of an image and consequently the accuracy of a localization system.

$$HGSD = \frac{w_s d}{f * r_h} \quad (3.10)$$

$$VGSD = \frac{h_s d}{f * r_v} \quad (3.11)$$

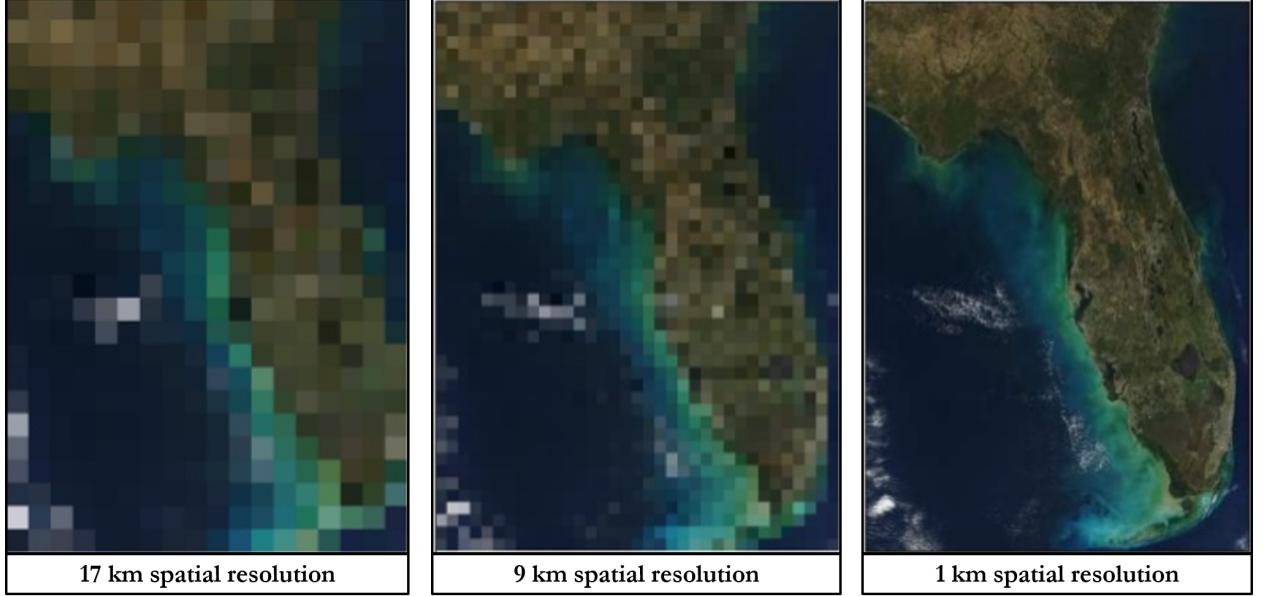


Figure 3.17: Ground Sample Distance Effects on Image Quality. Adapted from [87].

Equations 3.8 and 3.11, all depend on d , the distance from the work surface. Knowing the desired accuracy, as well as horizontal and vertical FOV of a certain application, an appropriate d can be calculated by rearranging the equations above.

In order for the localization technique to be able to update in real time, even as the system moves, two additional camera characteristics must be considered, frame rate and shutter type. An appropriate frame rate is necessary to ensure an adequate image overlap; so that, feature points can be tracked using optical flow, which will be discussed in Section 3.7.7, instead of having to redetect the feature each frame. The required frame rate can be calculated using Equation 3.12,

$$FR = \frac{v}{FOV_{distance}(1 - o)} \quad (3.12)$$

where v is velocity and o is percent overlap. However, due to the fact that while localizing the system may move in both the horizontal and vertical planes, Equation 3.13, where v_h and v_v are horizontal and vertical velocity respectively and o_h is horizontal percent overlap, while o_v is vertical percent overlap, should be used to ensure an adequate frame rate in both planes [88].

$$FR = \sqrt{\frac{v_h}{HFOV_{distance}(1 - o_h)} + \frac{v_v}{VFOV_{distance}(1 - o_v)}} \quad (3.13)$$

There are two types of camera shutters, a rolling shutter (RS), in which the horizontal rows of the sensor array are scanned at different times, and a global shutter (GS), in which all pixels are exposed at the same time. Due to the RS not exposing all pixels simultaneously, they are susceptible to motion blur [89] as seen in Figure 3.18. Consequently, in order to ensure features are not lost between frames as the system moves, a digital RGB camera with a GS is needed.

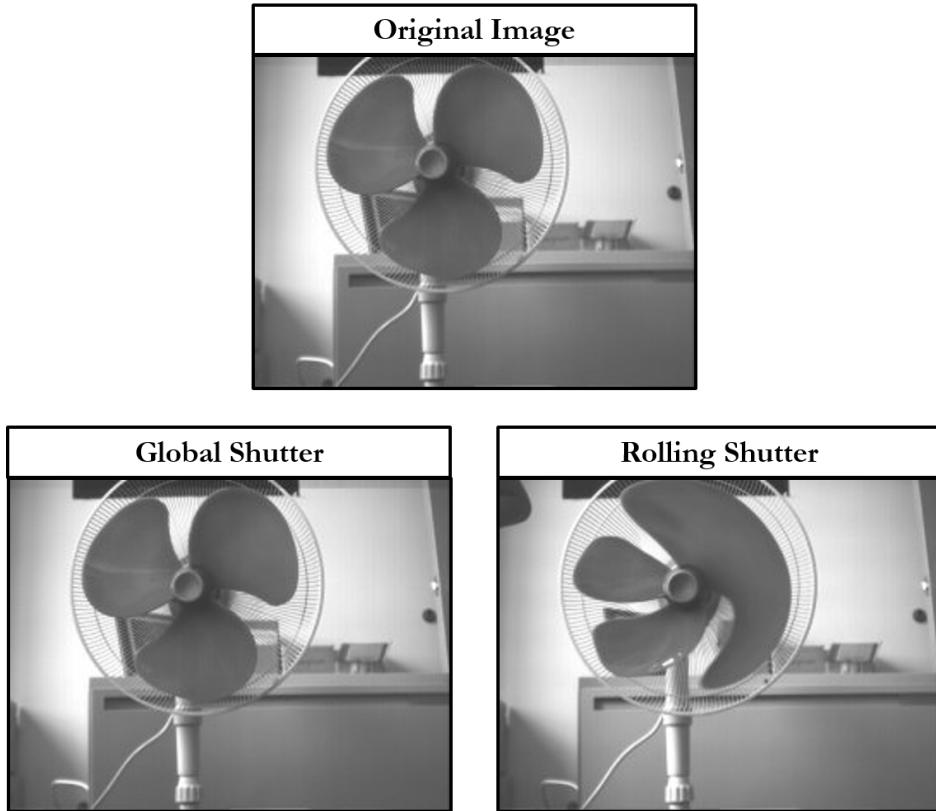


Figure 3.18: Shutter Type Effects on Image Quality. Adapted from [90].

3.7 Computer Vision Concepts

The purpose of this subsection is to briefly explain all the computer vision concepts needed to understand the multistage localization approach presented in the subsequent chapters. The following subsections will explain the basic concepts and rational behind the use of certain color spaces, filters, and binary operations. In addition, Canny Edge Detection, Hough Circle Detection, Good Features to Track, and Optical Flow, as well as Augmented Reality tag detection and pose estimation will be expounded upon.

3.7.1 Color Spaces

While the RGB color space is the primary color space used to describe spectral content of color signals, a variety of other representations have been developed [80]. Each color space has unique advantages and disadvantages that must be considered before using a specific color space for a given application. The two color spaces used in the multistage localization approach presented in this paper are the RGB and the HSV color spaces, depicted in Figure 3.19. The RGB color space is an additive color space based on the RGB color model, in which red, green, and blue light is added together to produce any color that is the triangle defined by those three primary colors [91], [92]. The HSV color space, which was developed by Alvy Ray Smith in 1978, is a cylindrical-coordinate representation of points in the RGB color model [93]. In some situations, such as color picking, the HSV model is more intuitive, as it mirrors traditional color mixing methods [94]. HSV stands for hue, saturation, and value. Hue represent angular position on the color wheel, where red starts at 0° , green at 120° , and blue at 240° . Saturation, or distance from the center axis, indicates the amount of grey in a color, where 0 is the color grey and 1 is the primary color. Value, or distance along the center axis, represents the brightness of a color, where 0 is black and 1 is white [93].

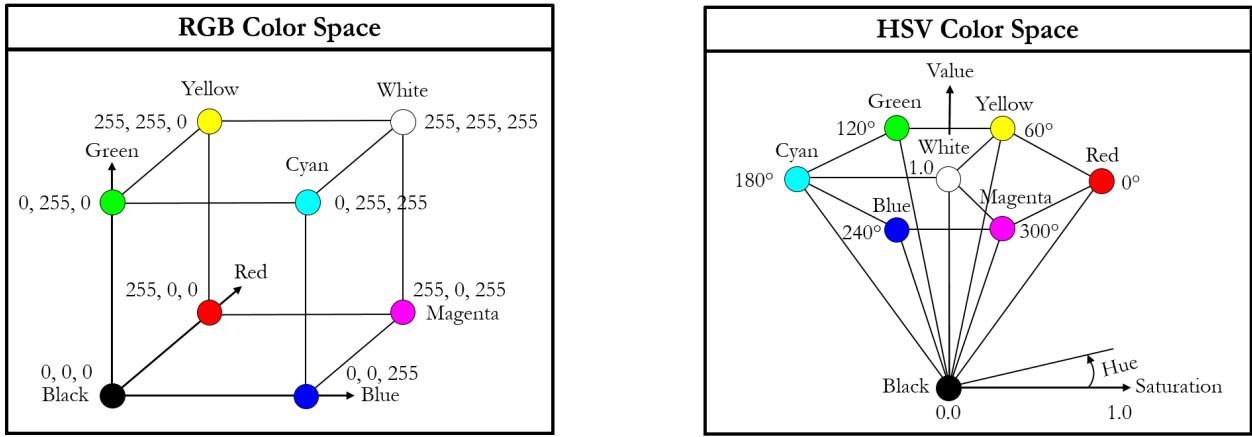


Figure 3.19: RGB and HSV color space models

3.7.2 Linear and Non-Linear Filters

Images are often corrupted by random variations in intensity, illumination, as well poor contrast that must be dealt with in the early stages of image processing before higher level computer vision techniques, such as feature detection and tracking, can be reliably used [95]. Random variation in the intensity of an image is called noise. Common types of image noise include salt and pepper, impulse, and Gaussian noise. While salt and pepper noise is the random occurrence of both black and white pixels, impulse noise is merely the random

occurrence of white pixels. Gaussian noise, on the other hand, contains variation in intensities drawn from a Gaussian distribution, which can be used to model many kinds of sensor noise [96]. In most cases, convolving an image with a linear low-pass filter, a filter which replaces each pixel value with the weighted sum of all pixel values in its neighborhood, or a spatially invariant non-linear filter, which does not use a weighted sum but rather performs the same operation at each pixel, is able to remove these types of noise. Figure 3.20 shows that a Gaussian filter is apt at removing noise drawn from a Gaussian distribution, such as white noise. On the contrary, a median filter, which replaces each pixel with the median value in its neighborhood, is suitable for removing salt and pepper, as well as impulse noise. In addition to removing noise, filters can be used to enhance an image, such as sharpen it as seen in Figure 3.20, extract information, like edges and texture, as well as detect patterns through template matching [96].

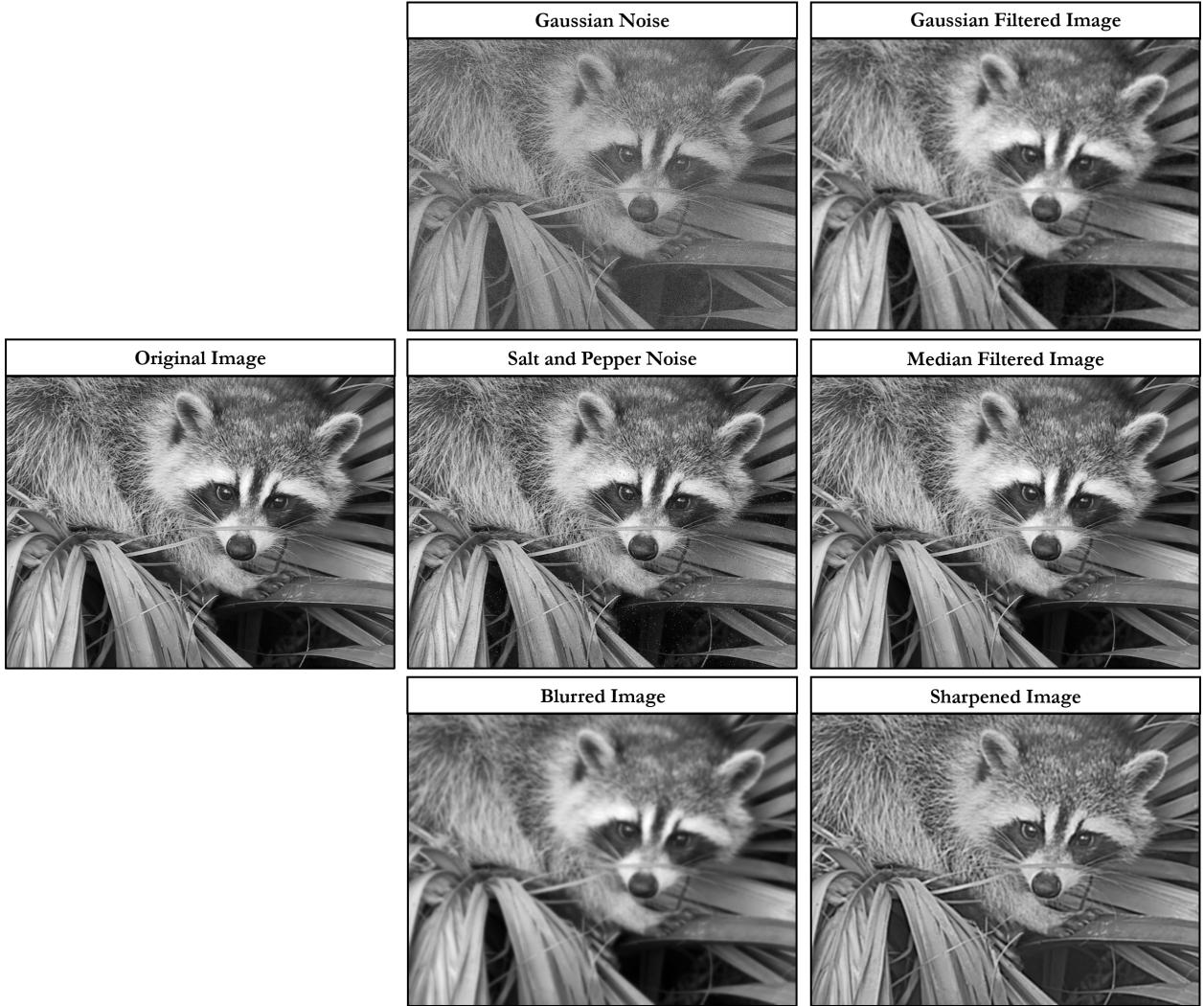


Figure 3.20: Effects of common filters.

3.7.3 Morphological Operations

In addition, to enhancing images, non-linear filters are also used extensively to process binary images [80]. A binary image is an image with only two possible values for each pixel, typically black, background pixels with a value of 0, and white, foreground pixels with a value of 1. Binary images often result after operations such as segmentation, thresholding, and dithering [97]. The most common binary image operations are called morphological operations, since they change the shape of the underlying binary objects [80]. Whether or not a given foreground or background pixel value changes depend on the image, as well as the morphological operations and the structuring element chosen. The structuring element

is a rectangular array of pixels, which contain values of either 1 or 0. When the structuring element is centered on the pixel under consideration, that pixel's neighborhood is determined by the pixels in the structuring element that have a value of 1 [98]. The two fundamental morphological operations are dilation, which thickens foreground pixels, and erosion, which thins foreground pixels. These two operations can be applied in different combinations in order to obtain more sophisticated operations, such as closing, dilation then erosion, and opening, erosion then dilation [99]. The morphological operations, depicted in Figure 3.21, are typically used to clean up binary images; so that, higher level computer vision techniques can be used robustly [80].

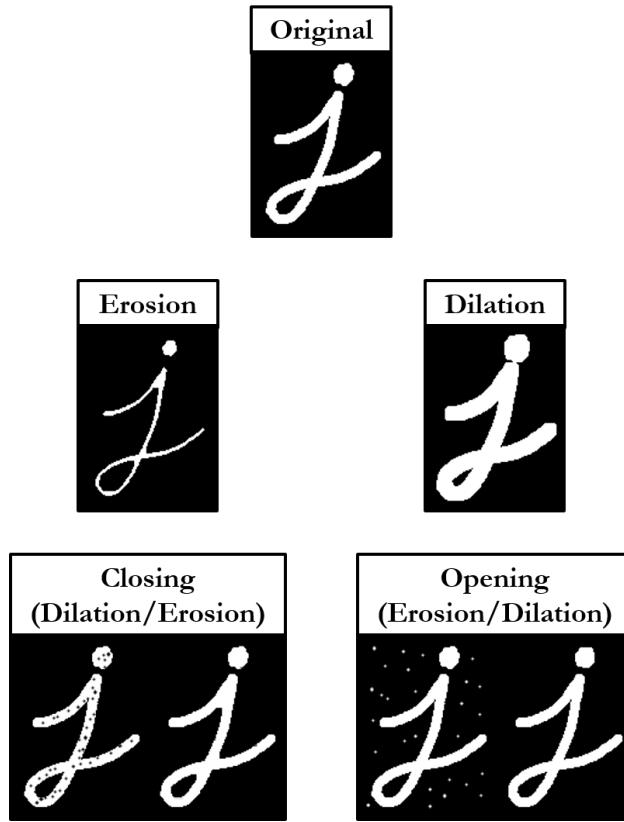


Figure 3.21: Effects of common binary operations.

3.7.4 Canny Edge Detector

The Canny edge detector, which is arguably the most used edge detector to date, was developed in 1986 by John Canny, from whom it derives its name [100]. Mark Nixon notes in [101] that John Canny developed the Canny edge detector to meet three main objectives, two of which are indicated below.

1. *optimal* detection with no spurious responses
2. *good* localization with minimal distances between detected and true edge positions

Figure 3.22 shows the Canny edge detector's framework. The detector first acquires a greyscale image, either directly or through conversion from another image format. The first objective Nixon notes aims to reduce noise. Consequently, the detector applies a Gaussian filter to smooth the image, thereby reducing noise. After which, the Sobel edge detection operators, G_x , which is equal to Equation 3.14, and G_y , which is equal to Equation 3.15, are used to find the edge gradient, G , and direction, θ , for each pixel using Equation 3.16 and 3.17 respectively.

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \quad (3.14)$$

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \quad (3.15)$$

$$G = \sqrt{G_x^2 + G_y^2} \quad (3.16)$$

$$\theta = \tan^{-1}\left(\frac{G_y}{G_x}\right) \quad (3.17)$$

The second objective Nixon notes aims for accuracy. This accuracy is achieved through non-maximum suppression, which uses the calculated edge gradient and direction to maintain only those edge points at the top of the hill created by the gradient, thus thinning the edges. Hysteresis thresholding is then used to connect the edge points. It does this by having a double threshold, one to start an edge and a lower one to continue an edge [101], [102].

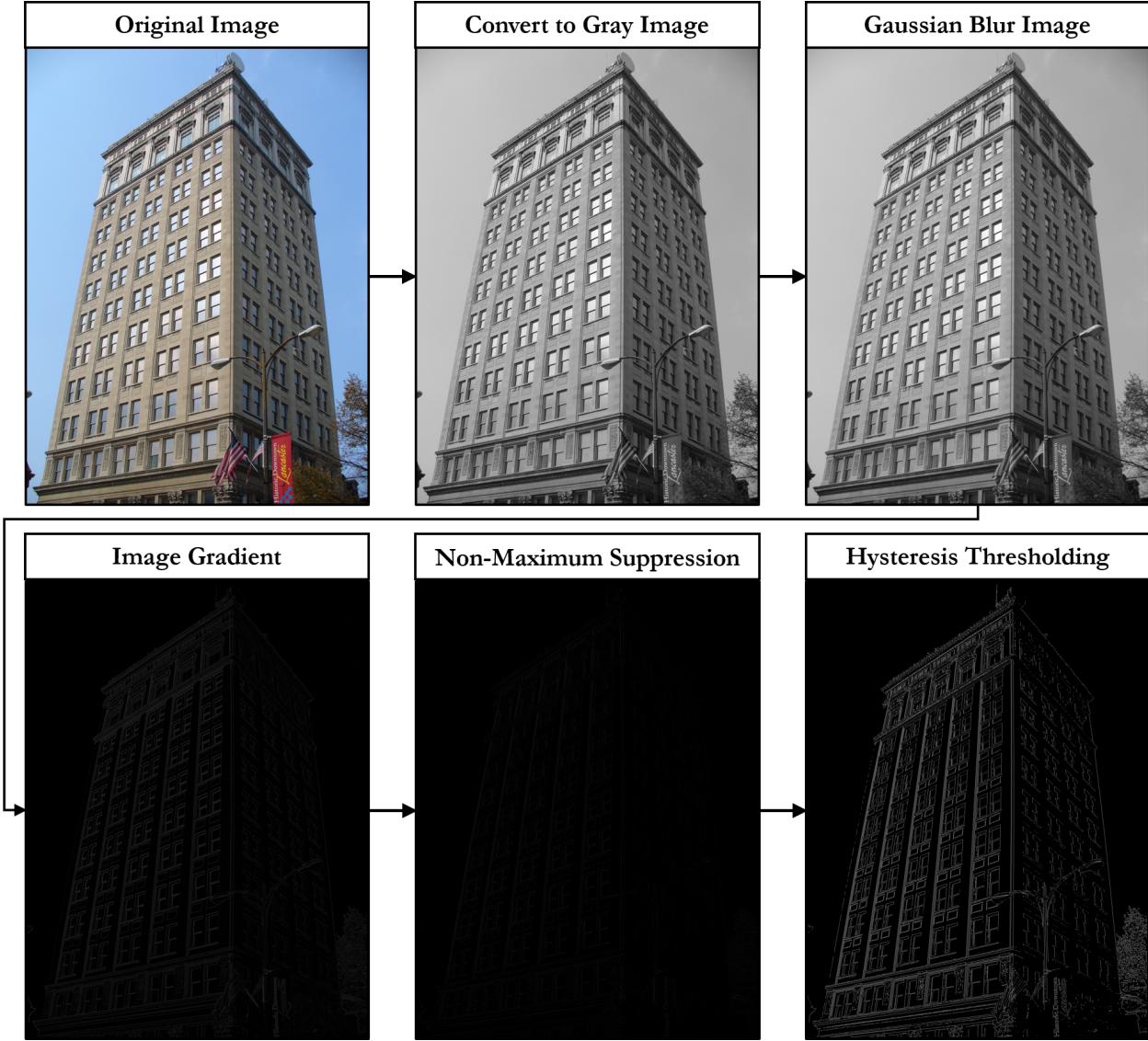


Figure 3.22: Summary of Canny Edge Detection Framework.

3.7.5 Hough Circle Detector

In order to implement the multistage localization technique presented in this paper to drilling operations, our detection algorithm must be able to both reliably and accurately detect circles on a hole template. The Hough circle detector is a circle detector which can be used to robustly and accurately determine the three parameters that define a circle on the coordinate plane, the radius of the circle and the center of the circle in x and y coordinates [103]. The Hough Transform, named after its original inventor, is a popular technique for having edges

“vote” for plausible line locations [80]. In 1981, Ballard extended the generalized version of the Hough Transform developed by Duda and Hart in [104] to be used for the detection of arbitrary shapes, specifically lines, circles, and parabolas [105].

In general, the Hough circle detector uses the Canny edge detector to produce a binary image of edges. After which, the x-y coordinates of each edge point in the image space are mapped to a two-dimensional or three-dimensional accumulator array, the Hough space, as shown in Figure 3.23. After all of the edge points have been processed, a thresholding technique is applied to determine the locations and radii of circles that meet a user specified criteria [106], [107].

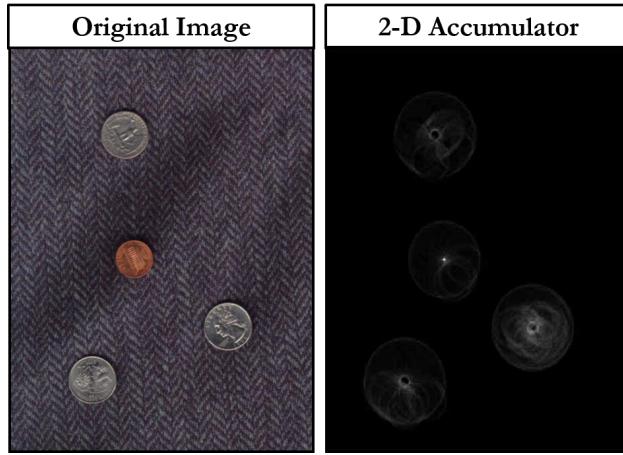


Figure 3.23: Hough Circle Detector’s 2-D Accumulator Array with Known Radius (Penny). Adapted from [108].

If the circles in an image have a known radius, R , then the search can be reduced to two dimensions. Figure 3.24 shows that each edge point in the image space (left) generates a circle in the Hough space (right) through the sweeping of the angle θ through a full 360 degree rotation in Equations 3.18 and 3.19,

$$x = a + R \cos(\theta) \quad (3.18)$$

$$y = b + R \sin(\theta) \quad (3.19)$$

where R is radius and (a, b) is the circle center in x and y coordinates. The points in the two-dimensional Hough space with peaks above a certain threshold are determined to be the circle center locations in x and y coordinates [108].

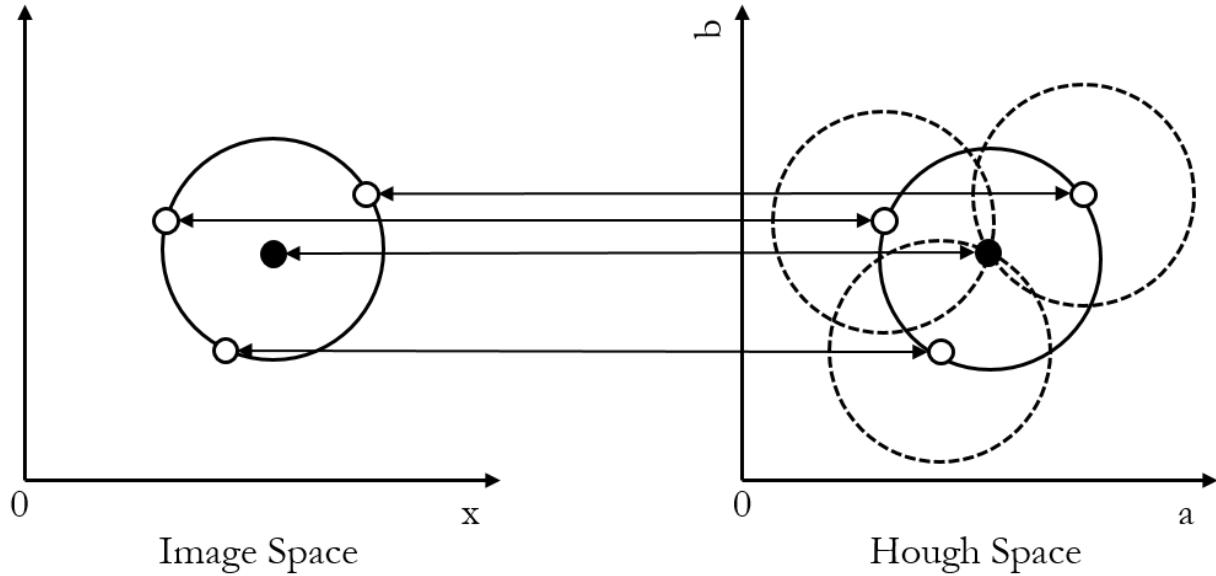


Figure 3.24: Summary of Hough Circle detector with known radius.

If the radius is not known, then the search must be conducted in three dimensions. Each edge point in the image space produces a cone surface, instead of a circle, in the Hough space as seen in Figure 3.25. After all edge points have voted, then the points with a peaks above a certain threshold will be chosen as the circle center locations and radii [108].

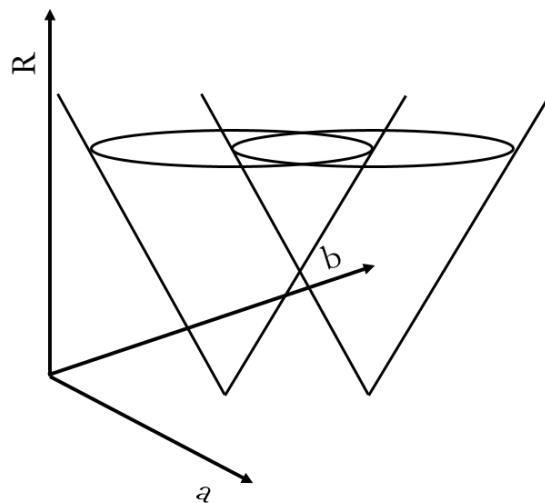


Figure 3.25: Summary of Hough Circle detector with unknown radius.

3.7.6 Good Feature To Track

In order to implement the multistage localization technique presented in this paper to sealant application operations, our detection algorithm must be able to both reliably and accurately detect corners. Good Feature to Track is a corner detection method which can be used to robustly and accurately determine the position of a corner in x and y coordinates [109]. In 1988, Harris and Stephens developed the Harris corner detector in [110]. Six years later, Shi and Tomasi developed Good Feature to Track in [109] by slightly modifying the Harris corner detector's scoring function, which significantly improved the overall reliability of the detector [111].

Figure 3.26 shows that if one were to examine the change in intensity of a small image patch as it was shifted around, they would be able to determine if the region was “flat” or if it contained an “edge” or a “corner”. A “flat” region would have no significant change in intensity regardless of direction shifted, while an “edge” would have a significant change in intensity if shifted orthogonal to the edge’s direction and a “corner” would have a significant change in intensity if shifted in any direction [112].

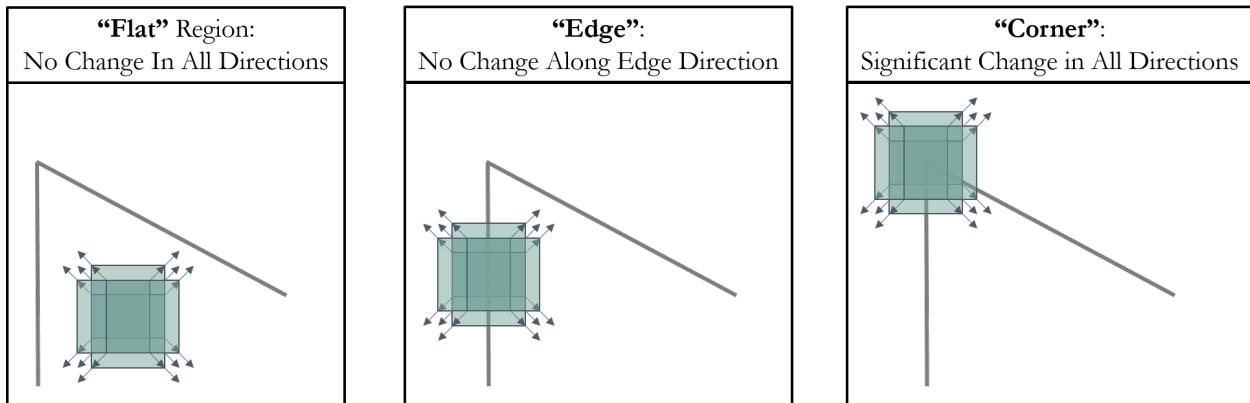


Figure 3.26: Summary of Good Feature detector. Adapted from [112].

The change in intensity of each pixel in an image can quickly and efficiently be found by taking the derivative of the image in the x and y directions, which produces gradient images [112]. Gradient images for each of the aforementioned cases are shown in Figure 3.27.

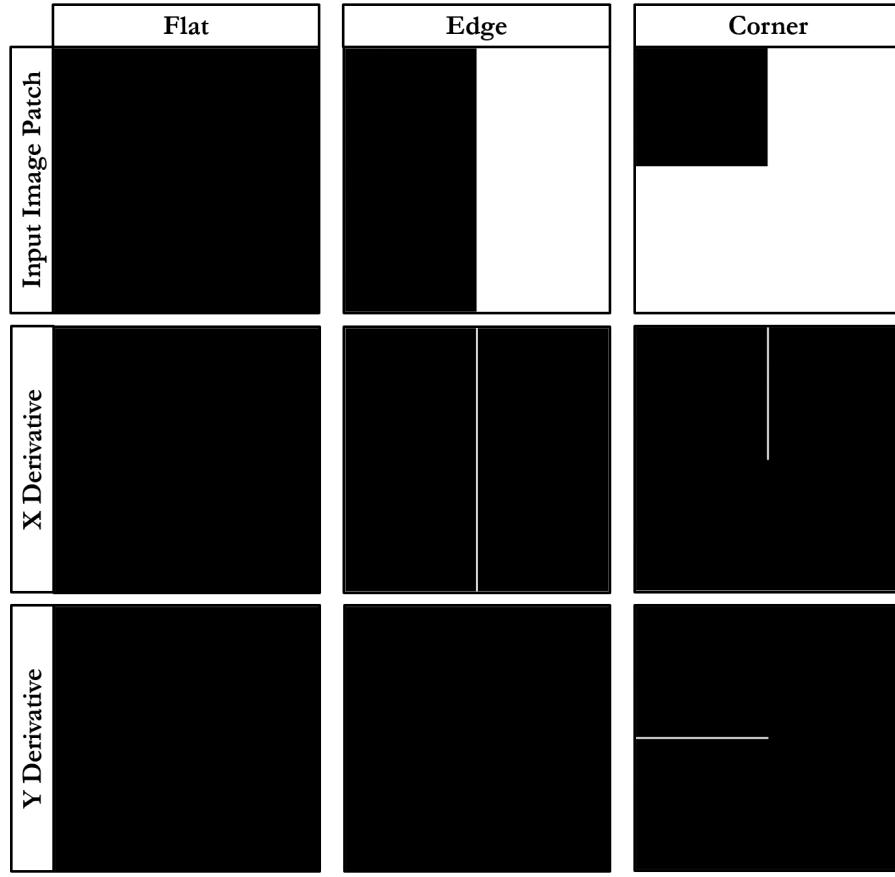


Figure 3.27: Image Gradient Example.

Equation 3.20 mathematically expresses the change of intensity for the shift, $[u, v]$ [112].

$$E(u, v) = \sum_{x,y} w(x, y)[I(x + u, y + v) - I(x, y)]^2 \quad (3.20)$$

$w(x, y)$ is the window function, while $I(x + u, y + v)$ and $I(x, y)$ are the intensity values at the shifted and original locations respectively.

Equation 3.21 evaluates to approximately zero for “flat” regions; however, for an “edge” or a “corner” the value will be significantly larger.

$$\sum_{x,y} [I(x + u, y + v) - I(x, y)]^2 \quad (3.21)$$

Through the use of Taylor Series Expansion, Equation 3.21 becomes Equation 3.22, which

can be simplified to Equation 3.23.

$$E(u, v) \approx \sum_{x,y} [I(x, y) + uI_x + vI_y - I(x, y)]^2 \quad (3.22)$$

$$E(u, v) \approx \sum_{x,y} u^2 I_x^2 + 2uv I_x I_y + v^2 I_y^2 \quad (3.23)$$

Re-writing Equation 3.23 in matrix form results in Equation 3.24. Equation 3.25 is the *bilinear* approximation of Equation 3.24 for small shifts in $[u, v]$,

$$E(u, v) \approx [u \ v] \left(\sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \right) \begin{bmatrix} u \\ v \end{bmatrix} \quad (3.24)$$

$$E(u, v) \cong [u \ v] M \begin{bmatrix} u \\ v \end{bmatrix} \quad (3.25)$$

where $M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$. Equation 3.26 is the Harris corner scoring function,

$$R = \det(M) - k(\text{trace}(M))^2 \quad (3.26)$$

where $\det(M) = \lambda_1 \lambda_2$ and $\text{trace}(M) = \lambda_1 + \lambda_2$. So, Equation 3.26 can be rewritten as Equation 3.27.

$$R = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \quad (3.27)$$

Equation 3.28, on the other hand is Shi and Tomasi scoring function for Good Feature to Track. Figure 3.28 shows the difference that this slight change in the Harris corner detector's scoring function makes. However, the premise remains the same. If the eigenvalues, λ_1 and λ_2 , are both lower than a certain threshold, then the region is determined to be "flat". If $\lambda_1 \gg \lambda_2$ or vice versa, then the region is determined to be an "edge" and if λ_1 and λ_2 are both larger than a certain threshold then the region is determined to be a "corner" [113].

$$R = \min(\lambda_1, \lambda_2) \quad (3.28)$$

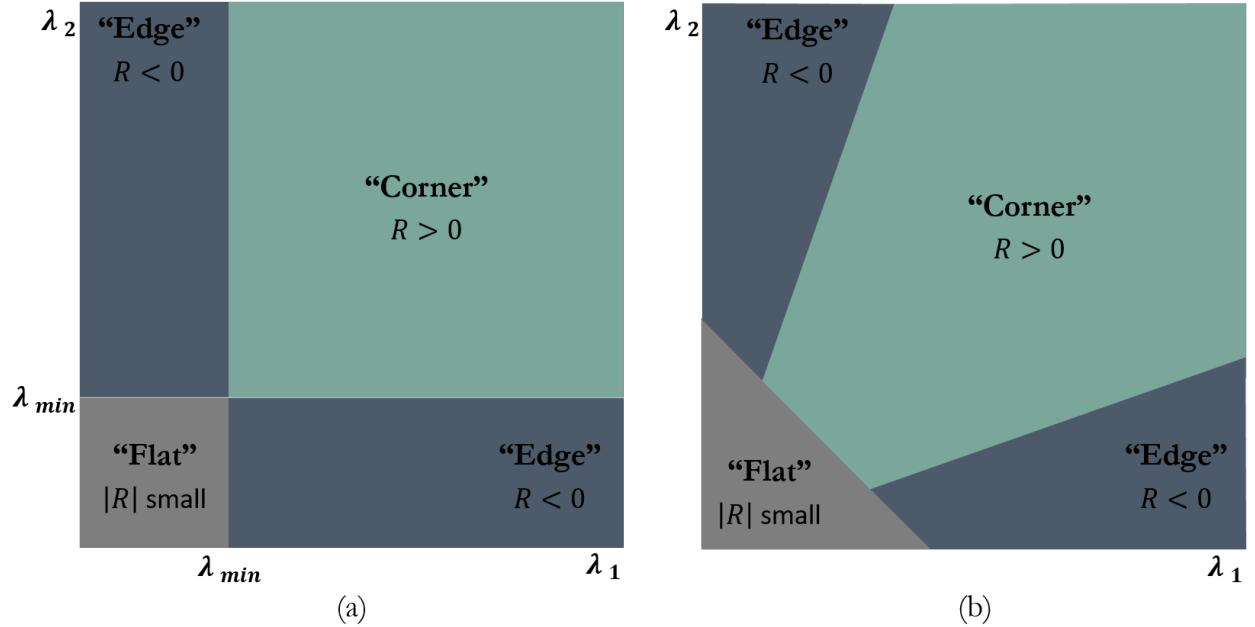


Figure 3.28: Difference Between Good Feature (a) and Harris Corner (b) Scoring Functions.

3.7.7 Optical Flow

In order to reduce the computational complexity inherent in running feature detection algorithms on every frame, optical flow, a method of computing an independent motion estimate for each pixel in an image [80], can be a computational inexpensive way to track features between frames. Figure 3.29 shows the computed flow vectors between two frames. The length of the vector is dependent on the motion of the pixel or patch between frames.



Figure 3.29: Example of optical flow. Adapted from [114].

Kanade-Lucas-Tomasi (KLT) tracker is a widely accepted and used method to track image point features between frames using optical flow [115]. The goal of KLT is to align a template image, $T(\mathbf{x})$, to an input image $I(\mathbf{x})$, where $\mathbf{x} = (x, y)^T$ is a column vector containing the pixel location of the image point features. Consequently, KLT's inverse compositional algorithm seeks to minimize Equation 3.29,

$$\sum_x [T(\mathbf{W}(\mathbf{x}; \Delta \mathbf{p})) - I(\mathbf{W}(\mathbf{x}; \mathbf{p}))]^2 \quad (3.29)$$

where $\mathbf{p} = (p_1, \dots, p_n)^T$ is a vector of parameters, the warp $\mathbf{W}(\mathbf{x}; \mathbf{p})$ takes all of the pixels, \mathbf{x} , in the template coordinate frame and maps it to image coordinate frame, while $\mathbf{W}(\mathbf{x}; \Delta \mathbf{p})$ denotes the incremental warp between the same frames [116].

Taking the first order Taylor Expansion of Equation 3.29 results in Equation 3.30.

$$\sum_x \left[T(\mathbf{W}(\mathbf{x}; \mathbf{0})) + \nabla T \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \Delta \mathbf{p} - I(\mathbf{W}(\mathbf{x}; \mathbf{p})) \right]^2 \quad (3.30)$$

Assuming that $\mathbf{W}(\mathbf{x}; \mathbf{0})$ is the identity warp, the solution to this least-square problem is Equation 3.31,

$$\Delta \mathbf{p} = H^{-1} \sum_x \left[\nabla \mathbf{T} \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right]^T [I(\mathbf{W}(\mathbf{x}; \mathbf{p})) - T(\mathbf{x})] \quad (3.31)$$

where the Hessian matrix, H , is equal to Equation 3.32 [116].

$$H = \sum_x \left[\nabla \mathbf{T} \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right]^T \left[\nabla \mathbf{T} \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right] \quad (3.32)$$

Figure 3.29 is a graphical representation of this optimization for a two parameter model, translation. However, the 6 parameter model, affine, is used in the multistage localization approach presented in this paper.

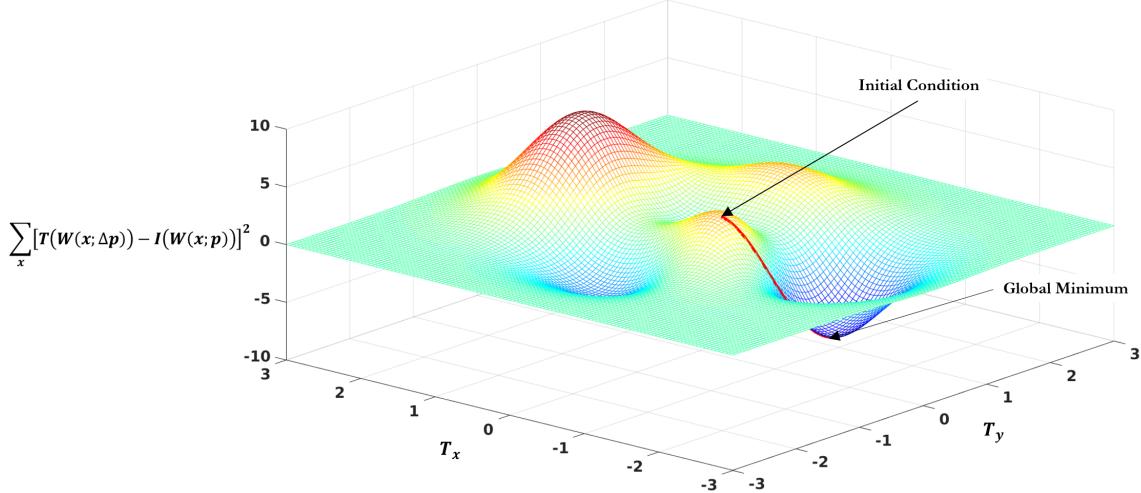


Figure 3.30: Summary of Kanade-Lucas-Tomasi feature tracker.

3.7.8 Pose Estimation and Tracking Through Augmented Reality Tag Detection

AR tags, which are also known as fiducial markers, are artificial landmarks which are designed to be easy to detect and identify under most circumstances [117]. Consequently, the detection method must be robust to changes in luminance (brightness), chrominance (color), and resolution, as well as be scale and rotationally invariant. Once detected and identified, AR tags can be used to accurately calculate the pose (location and orientation) of a camera, as well as to encode information or associate it with each tag's specific identity.

[118]. Using an AR tag to calculate the relative pose of a camera in real-time is called marker-based tracking. Several straightforward and robust marker-based tracking toolkits exist, such as ARToolKit [119], ALVAR [120], and AprilTags [117]. The multistage localization approach presented in this paper will use a ROS implementation of ALVAR [121] to both calculate the pose of an AR tag in the camera's frame and associate information with that tag's specific identity.

Figure 3.31 shows ALVAR's marker-based tracking framework. The system first acquires a greyscale image, either directly or through conversion from another image format. After this, adaptive thresholding, which can handle local changes in luminance, is used to create a binary image. A binary image consists of a background, black pixels, and objects, sections of white pixels. Subsequently, the contours of each object are found. Due to the fact that AR systems, like ALVAR, aim for real-time processing and fast performance, time cannot be wasted processing non-markers. As a result, fast acceptance/rejection tests are run. ALVAR uses two fast accept/reject tests including size and the four-corner test. The number of pixels in an object's perimeter can be efficiently used to estimate the object's area. Objects with areas that are either too large or too small can be rejected. Even if an object, whose area is too small, is a marker, it is too far from the camera to either be correctly identified or the pose of the camera accurately calculated. Objects, whose area is too large, can be rejected given some background knowledge about the upper and lower range of the camera from the marker, as well as the marker's size. The contours of all remaining markers are fitted with lines. A quadrilateral has exactly four straight lines and four corners. The number of lines and corners of each object are calculated. Those that fail the four-corner test are dropped. After determining that an object is a marker, it is identified and the corner locations are optimized to sub-pixel accuracy for further calculations. Even small errors in the detected 2-D locations of edges and corners can significantly affect the calculated pose of the camera [118]. The pose of a calibrated camera can be uniquely determined from a minimum of four coplanar but non-collinear points [84]. Thus, the system can calculate a marker's pose (relative to camera) in 3-D coordinates using the four corner points of the marker in image coordinates.

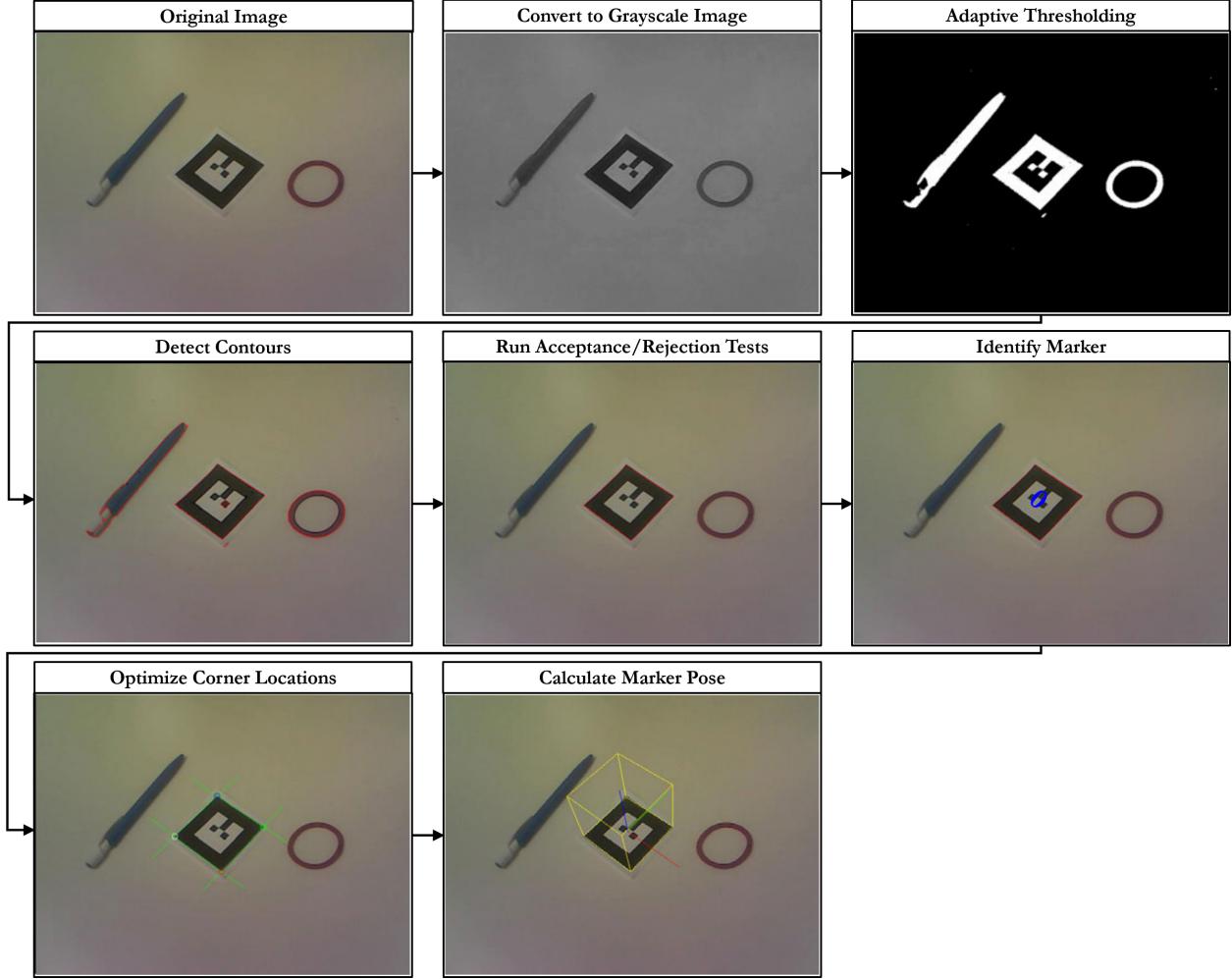


Figure 3.31: Summary of ALVAR AR Tag Detection Framework. Adapted from [118].

3.8 Summary

The purpose of this chapter was to briefly explain all of the concepts needed to understand the work presented in the chapters thereafter. The basic concepts of the Robot Operating System (ROS), Simultaneous Localization and Mapping (SLAM), Localization and Path Planning, Manipulation, Task Execution using State Machines, and the specific Computer Vision algorithms implemented were explained in depth. The following chapter will explain in detail each component of the multistage localization approach presented in this paper, as well as its implementation to two tasks which are prevalent in both the manufacturing and construction industries: drilling and sealant application operations.

Chapter 4

Multistage Localization for High Precision Mobile Manipulation Tasks

The purpose of this chapter is to thoroughly explain each component of the multistage localization approach presented in this paper, as well as how these individual components fit together in order to enable high precision 3-D feature point localization without having to recalibrate the position of the end-effector each time the robot's mobile bases moves. In addition, this chapter will expound upon the implementation of this approach to two tasks which are prevalent in both manufacturing and construction industries: drilling and sealant application operations.

4.1 Approach Overview

Figure 4.1 summarizes the multistage localization approach's framework. The mobile manipulator first localizes itself within the area of operation (AO) using adaptive Monte Carlo localization, which relies on the fused odometry and sensor messages published by the robot, as well as a 2-D map of the AO, normally generated using an optimization-based smoothing simultaneous localization and mapping (SLAM) technique. The robot navigates to a predefined start location in the map using a technique called trajectory rollout. Once there, the robot uses an RGB-D sensor to localize an augmented reality (AR) tag in the map frame. Once localized, the identity and the 3-D position and orientation, collectively known as pose, of the tag are used to generate a list of initial feature points and their locations based off of *a priori* knowledge. After the end-effector moves to the rough location of a feature point provided by the AR localization, the feature point's location, as well as the end-effector position and orientation are refined to within a user specified tolerance through the use of a control loop, which utilizes images from a calibrated machine vision camera and a laser pointer, simulating stereo vision, to localize the feature point in 3-D space using

computer vision techniques.

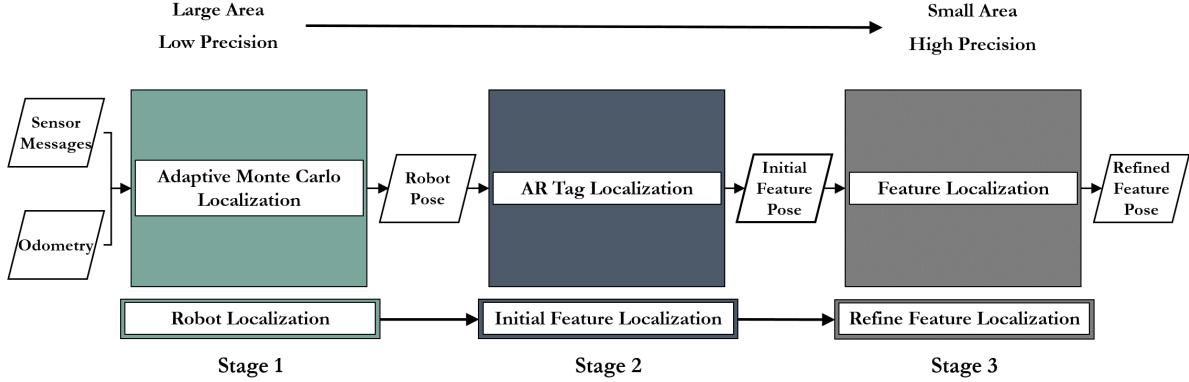


Figure 4.1: Multistage Localization Approach Overview.

4.2 Global Map Creation and Task Location Specification

In order for a mobile manipulator to operate and perform intricate tasks in a complex, GPS-denied environment, such as that of manufacturing facility or construction site, without modifying said environment, the robot must be able to create an accurate map of its environment. Simultaneous localization and mapping (SLAM) is the process of building or updating a map of an unknown environment while simultaneously localizing the robot within this map. Most SLAM techniques can be categorized into two main paradigms: filtering and optimization-based smoothing. While filtering SLAM techniques have been widely used in the past due to their ability to model different sources of sensor noise, in recent years optimization-based smoothing techniques have proven to be more efficient, scalable, and robust than that of filtering techniques [24]. ROS [30], through which the multistage localization approach was implemented, includes several 2-D SLAM packages. Santos et al. performed a thorough real world evaluation of these available 2-D SLAM packages in [54], in which KartSLAM [50], [43], an optimization-based smoothing SLAM technique, outperformed the other ROS implemented SLAM algorithms in terms of accuracy and computational efficiency. This not only demonstrates the robustness of its sparse pose adjustment (SPA), but that of full SLAM techniques in general. As a result, optimization-based smoothing techniques, such as KartSLAM, should be used to generate 2-D occupancy grid maps of the robot's environment. After which, start positions and orientations of specific operations are defined in the map as seen in Figure 4.2.

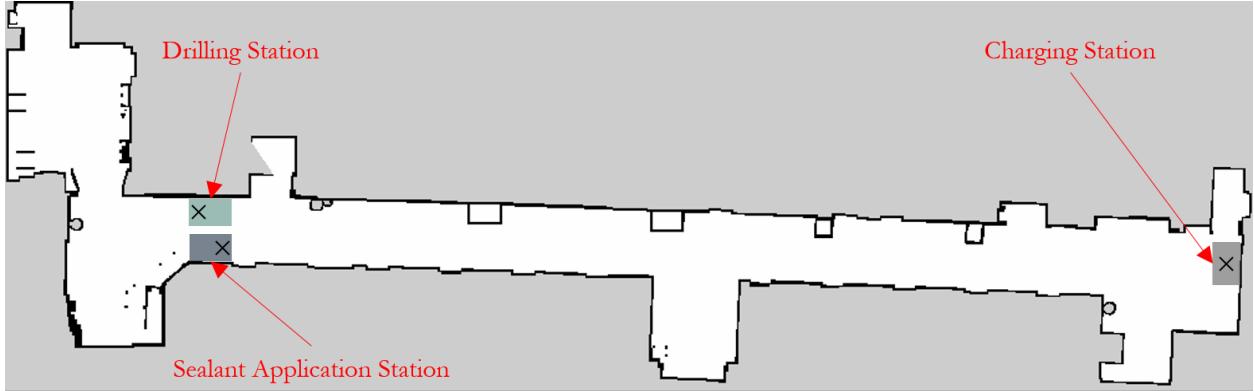


Figure 4.2: 2-D map environment with specified start locations.

4.3 Autonomous Localization and Navigation

After building a 2-D occupancy grid map, shown in Figure 4.2, using an optimization-based smoothing SLAM technique, it becomes crucial to accurately localize the robot within this predefined map; so that, the robot can both plan and execute appropriate trajectories to reach its destination. Figure 4.3 shows how the multistage localization approach localizes the robot within the map frame. Through the measurement of wheel rotation and the integration of accelerations provided by an inertial measurement unit (IMU), the distance traveled by the robot from its initial position can be calculated and the robot's pose in the map estimated. However, these methods do not account for wheel slippage or measurement error. As a result, the accuracy of the pose estimation will degrade over time. Consequently, the odometry measurements from these two sources are fused together through an Extended Kalman filter to provide an accurate pose estimate [122]. The corrected odometry, as well as laser scan measurements from a LIDAR and the prebuilt map are used by Adaptive Monte Carlo Localization (AMCL) [57] to localize the robot within this prebuilt map. AMCL utilizes a particle filter to keep track of the robot's pose and *KLD-sampling* to improve computational efficiency by removing redundant particles.

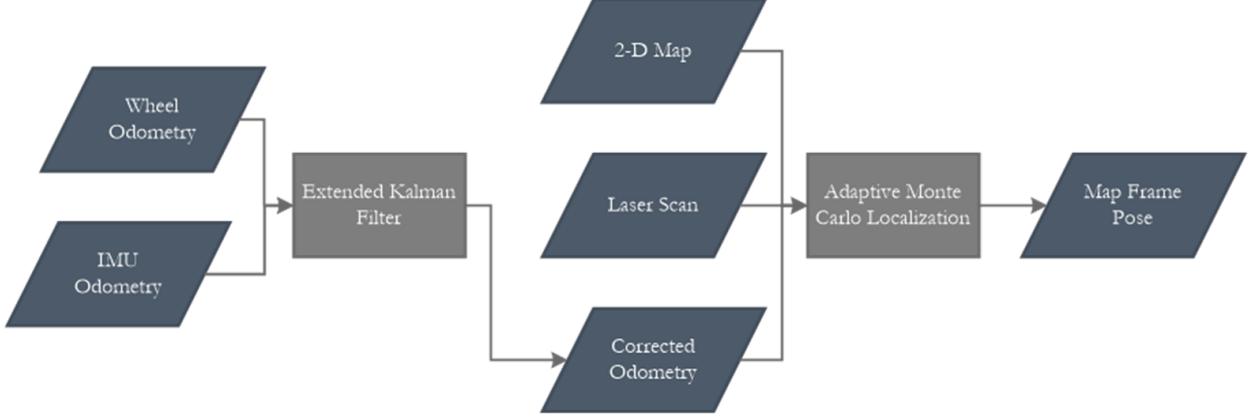


Figure 4.3: Robot Localization Framework.

After localizing the robot within the predefined map, and computing local and global costmaps, in which obstacles and a specified distance around them represent a cost, the robot must now both plan and execute appropriate trajectories to reach its destination. Trajectory Rollout [61] samples the space of feasible controls. For a differential drive robot, this controls space is 2D and consists of translations and rotational velocities, $\dot{x}, \dot{\theta}$, which are limited by the robot's capabilities. This sampled velocity is forward simulated from the robot's current position for a short period of time in order to generate simulated trajectories. These simulated trajectories are then scored using the cost function in Equation 4.1,

$$C(k) = \alpha Obs + \beta Gdist + \gamma Pdist + \delta \frac{1}{\dot{x}^2} \quad (4.1)$$

where Obs is the sum of grid cell costs through which the trajectory passes, $Gdist$ and $Pdist$ are the estimated shortest distance from the endpoint of the trajectory to the goal and the optimal path, respectively, and \dot{x} is the translation component of the velocity command that produces the trajectory. The simulated trajectory that minimizes this cost function is chosen. As a result, chosen trajectories tend to keep obstacles at a distance, proceed towards the goal, remain near the optimal path, as well as have higher velocities [60].

ROS's Navigation Stack [55] is a collection of packages, which the aforementioned techniques can be implemented in, in order to localize the robot within a prebuild map, as well as plan and execute trajectories by outputting the velocity commands needed; so that, the robot can reach the specified goal.

4.4 Initial Feature Localization Framework

Each Augmented Reality (AR) tag has a unique identity. As a result, relevant information, such as feature locations relative to the AR tag's pose for a specific operation, can be associated with a predetermined AR tag identity. So, once the robot has arrived at the operation's specified start positions and orientation, the robot uses the ALVAR package [121] and an RGB-D sensor to detect, identify, and calculate the Augmented Reality (AR) tag pose. Upon identification of the AR tag at the operation start position, the feature locations for that operation are populated and then localized on the work surface given the AR tag's pose in the map frame. Figure 4.4 shows this initial feature localization framework.

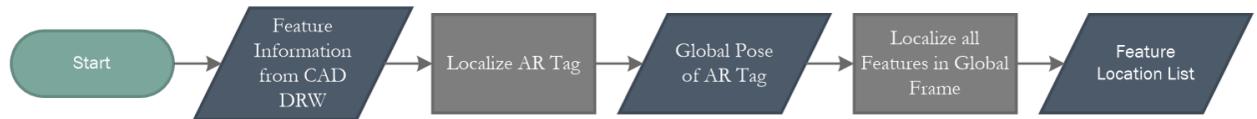


Figure 4.4: Initial Feature Localization Framework.

4.5 Motion Planning

Once the operation's feature locations have been populated and localized on the work surface in the map frame, the manipulator must be able to move the end-effector to these locations. The manipulator's motion planning is done through ROS's Moveit! package [62], which was set up to use the TRACLabs' numerical inverse kinematic (IK) solver, TRAC_IK [75]. Due to the recursive nature of numerical solvers, they tend to be slower than that of analytical solvers and are prone to get trapped in local minima. However, analytical solvers require a separate kinematic solver be created for each manipulator and manipulator configuration. Consequently, in order to ensure cross platform functionality with minimum overhead, numerical solvers were used. Orococos Kinematics and Dynamics Library (KDL) [73], Moveit! default numerical solver, is arguably the most widely-used generic numerical solver worldwide. However, it suffers from an unacceptably high failure rate and slow average speed. Beeson et al. determined in [76] the causes behind the high failure rates and slow average speed, as well as developed TRAC-IK to overcome them. Beeson et al. revealed that KDL had the following flaws:

1. Frequent convergence failures due to manipulator joint limits
2. No action is taken when the search gets trapped in a local minima
3. Lack of tolerance support and/or utilization in the solver itself

TRAC-IK solves these issues by using two separate IK solvers concurrently, as well as incorporating local minimum detection in both. It uses KDL-RR, a rewrite of KDL which incorporates local minimum detection in order to solve problems two and three by “unsticking” the convergence algorithm through the introduction of random seed angles when a local minimum is detected, as well as a sequential quadratic programming (SQP) IK implementation, which is an iterative algorithm for nonlinear optimization. The SQP implementation, which incorporates the same local minimum detection as KLD-RR, is able to deal with issue one. However, SQP can have a much longer solve time than KDL or KDL-RR. As a result, TRAC-IK implements both of these IK methods concurrently and waits for either IK to converge. As a result, the solve rate ranges between 99.1% and 99.9 % and the average convergence times are well below one millisecond.

4.6 Feature Correction Framework

Unfortunately, localizing features in the map frame based on the AR tag localization is not sufficient for tasks that require very high precision due to errors associated in localizing the AR tag in the map frame. Sources of error include, error in the RGB-D camera’s extrinsic calibration, error associated in localizing the AR tag itself, error in the location estimate of the camera with respect to the robot’s base link, and finally error in the location estimate of robot’s base link in the map. Consequently, a secondary localization method is required to refine the feature pose estimates; so that, they can be used for high precision tasks, such as drilling and sealant application operations in the manufacturing and construction industries.

In order to achieve high precision 3-D localization, the feature must be localized with a high degree of accuracy in the end-effector frame of reference. Consequently, a machine vision camera and a laser pointer, simulating stereo vision, were added to the end-effector. Figure 4.5 shows the 3-D localization pipeline developed for high precision feature localization. The 3-D feature localization pipeline runs two separate feature detection algorithms concurrently. The first feature detection algorithm is responsible for locating the center pixel of the laser point produced by the laser pointer. It does this by first converting each RGB frame to the HSV color space. The laser pointer is affixed so that the center of the laser point is located within the center region of each frame; so, the outside edges of each frame is masked. After which, the image is thresholded for the specific color range of the laser point. This produces a binary image, in which only pixels within the specific color range are considered foreground. Morphological filters are used to remove noise and fill in any holes in the laser point created by the laser pointer; so that, its center can be found. The center of the laser point is found by calculating the centroid of the largest blob in the binary image. The second feature detection algorithm is responsible for locating the center pixel of the operation’s specific feature. It does this by first converting each RGB frame to gray scale, after which, preprocessing operations, such as a Gaussian and Median filter, are used to minimize the noise in the image without degrading the features. Then Canny Edge Detection [100] is used to find appropriate edges

in the image, as well as to produce a binary image, which is used to reduce the computational complexity of the feature detection algorithm used. At which point, the outside edges of the image are masked, in order to decrease the likelihood that features other than the one desired are detected. After which, the appropriate feature detection algorithm, such as Hough Circle Detector [105] or Good Feature to Track [109], is run in order to find the center pixel of the feature closest to the laser point produced by the laser pointer. Once these features' pixel locations have been found, they are tracked using the Kanade-Lucas-Tomasi (KLT) tracker [116], based on the optical flow of the image, in order to reduce complexation complexity that would be associated with running the aforementioned feature detection algorithms on each frame. The location of the center of the laser point can be found by converting the pixel to 3-D camera coordinates using the camera's intrinsic parameters. However, in order to find the depth, the line-line intersection between the 3-D ray of the laser pointer in the camera frame, which was found during calibration, and the laser point's 3-D ray, which is calculated by finding the line connecting the camera's center point and the location of laser point in camera coordinates on the image plane, must be found. If the two lines do not intersect, the midpoint of the shortest line segment between the two lines should be used. The camera calibration procedure and the calculation of the laser pointer's 3-D ray equation in the camera frame are detailed in Section 5.3.1. Given the 3-D location of the laser point in the camera frame, as well as the surface normal of the laser pointer, which was calculated during calibration, the plane that the laser point lies on is calculated. Due to the proximity of the feature to the laser point, they are assumed to lie on the same plane. Consequently, the 3-D location of the feature point in the camera frame is calculated by finding the line-plane intersection between the 3-D ray of the feature, which is calculated by finding the line connecting the camera's center point and the location of feature point in camera coordinates on the image plane, and the plane of the laser point.

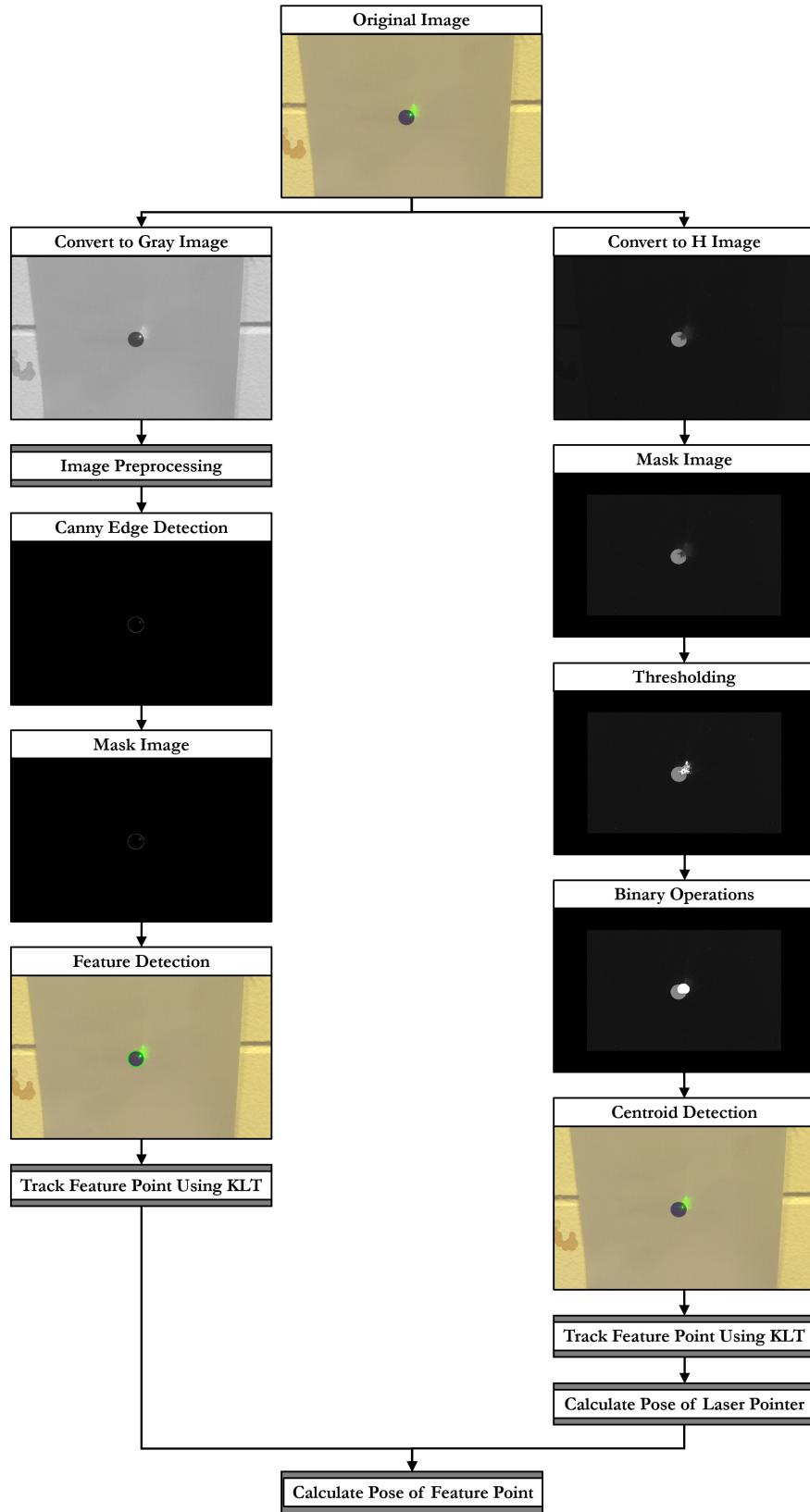


Figure 4.5: Feature Detection and Tracking Pipeline.

After the feature has been localized in the camera frame, the feature is transformed into the end-effectors frame. The manipulator then updates the position of the end-effector, while remaining a set distance off the work surface, through the use of the control loop depicted in Figure 4.6. Once the end-effector is within the reference tolerance set, the control loop terminates and the specified operation begins. This reference tolerance should be set no lower than the maximum accuracy achievable by either the detection algorithm or the end-effector itself, whichever is higher.

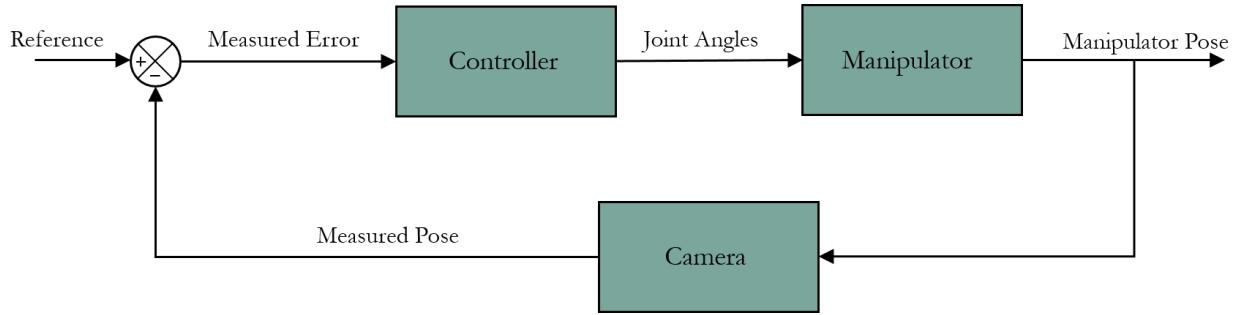


Figure 4.6: Manipulator Correction Control Loop.

4.7 Approach Implementation

The purpose of this section is to briefly explain how the multistage localization approach presented in Sections 4.1 through 4.6 can be implemented on a robotic system, as well as to provide an expandable framework; so that, the mobile manipulator can be easily programmed to conduct multiple operations within a manufacturing facility or construction site. Two operations, drilling and sealing, will be explored in depth.

4.7.1 System Overview

The generic framework of the system is depicted in Figure 4.7. The system allows the user to input a number of predefined tasks. Given *a priori* knowledge of each tasks, as well as their global start locations, the mobile manipulator will navigate to and perform the requested operation(s) while monitoring its battery level to ensure mission completion. The framework shown was implemented using a hierarchical state machine and was written using the Python library SMACH [79] in order to ensure compatibility with ROS; thereby, ensuring seamless integration with the aforementioned packages and ease of use across differing robotic platforms.

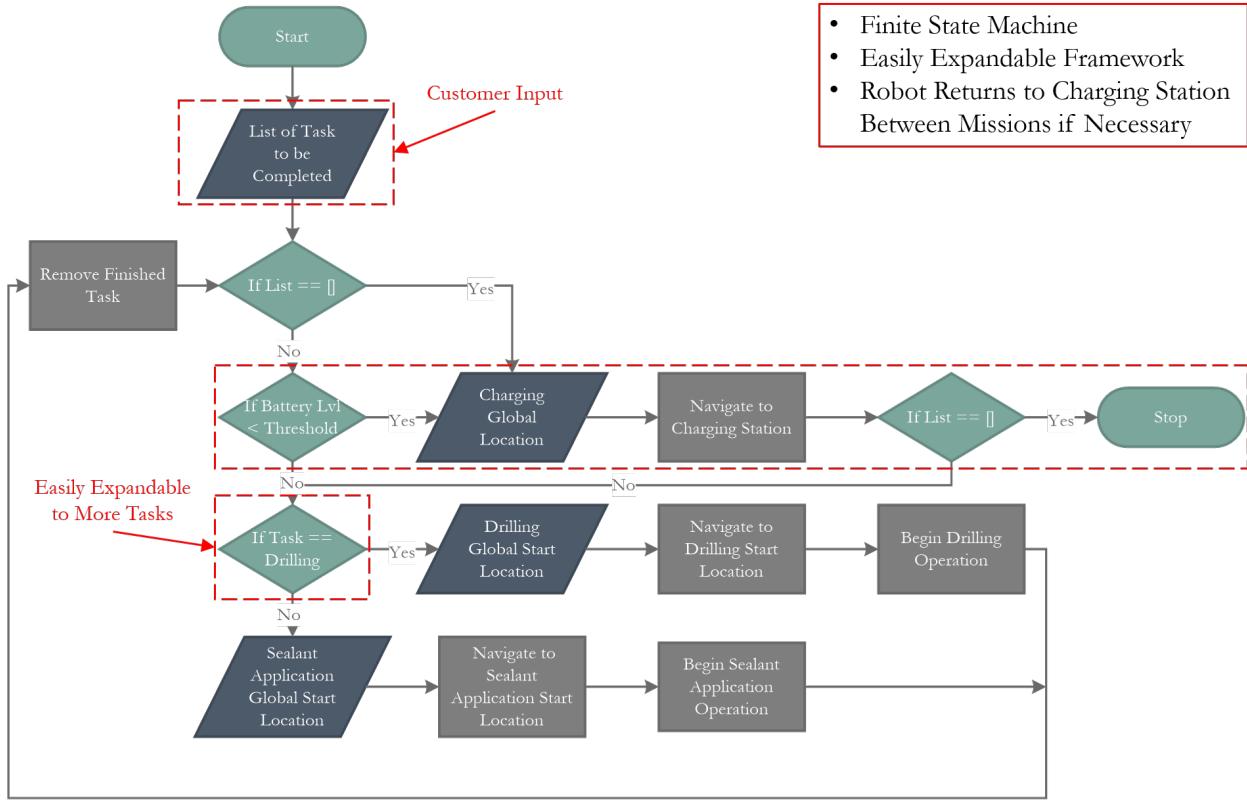


Figure 4.7: General system overview.

4.7.2 Drilling Framework

Figure 4.8 shows the basic drilling operation and error handling framework. Once the mobile manipulator has arrived at the drilling operation's start location and orientated itself in the prescribed direction, its RGB-D camera will detect, identify, and localize the AR tag located at the work station. Using this AR tag's location and identity, the specific features associated with that work station will be populated and localized on the work surface in the map frame. After these initial feature locations have been populated, the robot will move within range of the first feature location in the list. The manipulator will move the end-effector to the specified Cartesian location specified using Moveit! and the TRAC-IK numerical solver. Currently, many companies use hole templates to ensure that objects are drilled within necessary tolerances. Consequently, the feature correction framework will use these features as landmarks and detect them using Hough Circle Detector [105]. Using these landmarks, the feature locations are refined to within a user specified accuracy. After which, the manipulator will drill the hole to the specified depth and at the specified feed rate requested by the user. This will continue until all features have been drilled. However, if the

robot due to obstruction or another reason is unable to navigate within range of a feature, the machine vision camera is unable to detect the landmark, or the manipulator is unable to reach the location, these specified features will be added to a flag list and the customer notified after the operation.

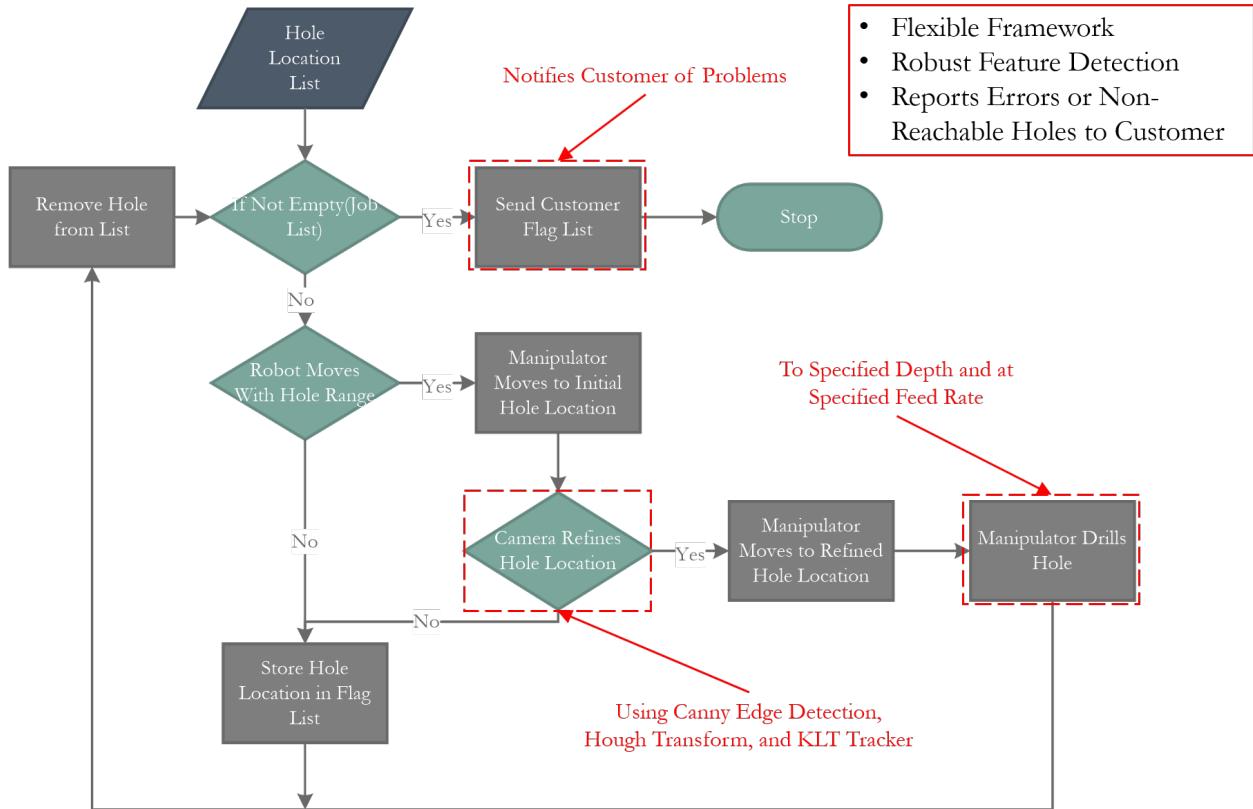


Figure 4.8: Drilling operation framework.

4.7.3 Sealant Application Framework

Figure 4.9 shows the basic sealant application operation and error handling framework. Once the mobile manipulator has arrived at the sealant applicant operation's start location and orientated itself in the prescribed direction, its RGB-D camera will detect, identify, and localize the AR tag located at the work station. Using this AR tag's location and identity, the specific features associated with that work station will be populated and localized on the work surface in the map frame. After these initial feature locations have been populated, the robot will move within range of the first feature location in the list. The manipulator will move the end-effector to specified Cartesian location using Moveit! and the TRAC-IK numerical solver. Most sealant applicant operations, such as sealing a window frame, require

the application of sealant from one corner to another. Consequently, the feature correction framework will use these features, corners, as landmarks and detect them using Good Feature to Track [109]. Using these landmarks, the feature locations are refined to within a user specified accuracy. After which, the manipulator will move to the next corner location in the list. After this corner's location has been refined, the manipulator will transition back to the previous corner's location and seal from that start corner to the end corner. This will continue until all connected corners have sealant applied between them. However, if the robot due to obstruction or another reason is unable to navigate within range of a feature, the machine vision camera is unable to detect the landmark, or the manipulator is unable to reach the location, these specified features will be added to a flag list and the customer notified after the operation.

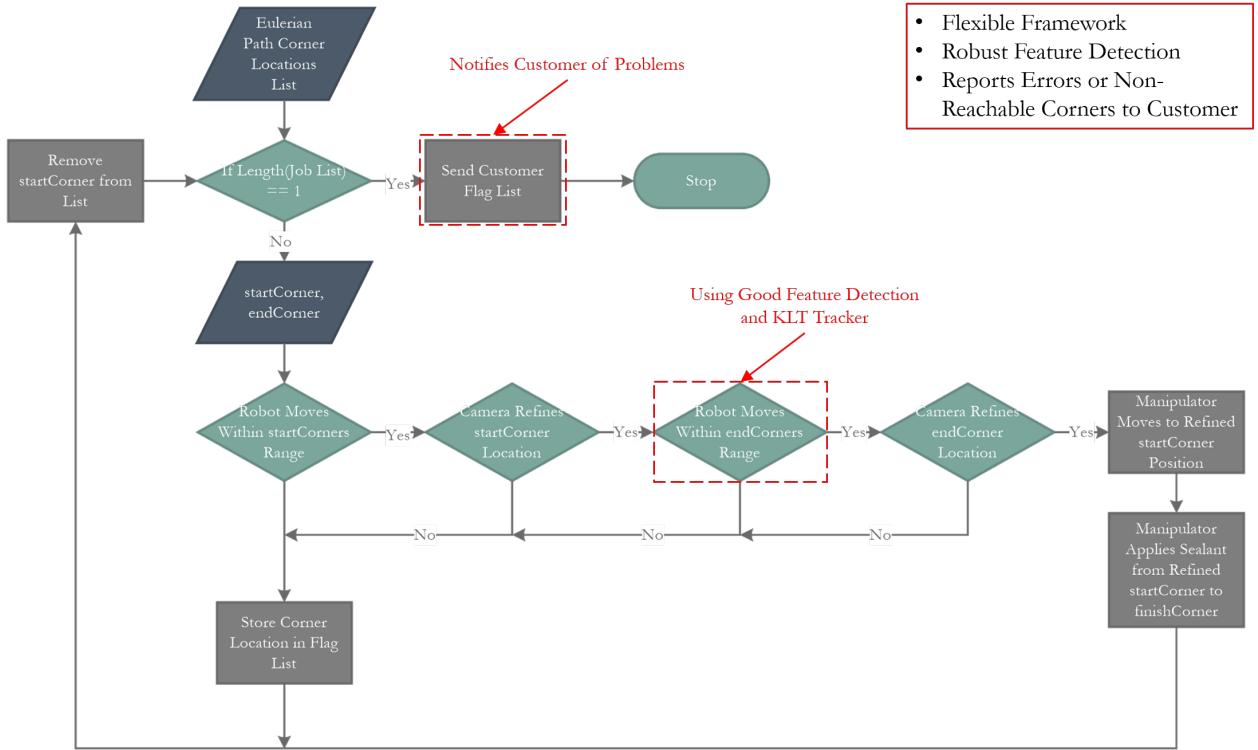


Figure 4.9: Sealant Application Framework.

4.8 Summary

The purpose of this chapter was to thoroughly explain each component of the multistage localization approach presented in this paper, as well as how these individual components fit together in order to enable high precision 3-D feature point localization without having

to recalibrate the positon of the end-effector each time the robot's mobile bases moves. In addition, this chapter expounded upon the implementation of this approach to two tasks which are prevalent in both manufacturing and construction industries: drilling and sealant application operations. The following chapter will discuss the experiment conducted in order to validate the approach using the two aforementioned operations.

Chapter 5

Experiments and Results

The purpose of this chapter is to validate each stage of multistage localization approach and the system implementation presented in Chapter 4 through testing. The accuracy of each component of the multistage localization technique will be determined on two different ROS enabled robotic platforms, the Clearpath Robotics' Husky and the Fetch Robotics' Fetch. In addition, these accuracies will be discussed and methods to improve them will be presented.

5.1 Hardware Architecture

Figure 5.1 shows the Clearpath Robotics' Husky and the Fetch Robotics' Fetch with the necessary equipment to simulate high precision operations, such as drilling and sealing. Each robot is equipped with an onboard computer, as well as an RGB-D and RGB camera, a laser pointer, LIDAR, IMU, wheel encoders, and a manipulator.

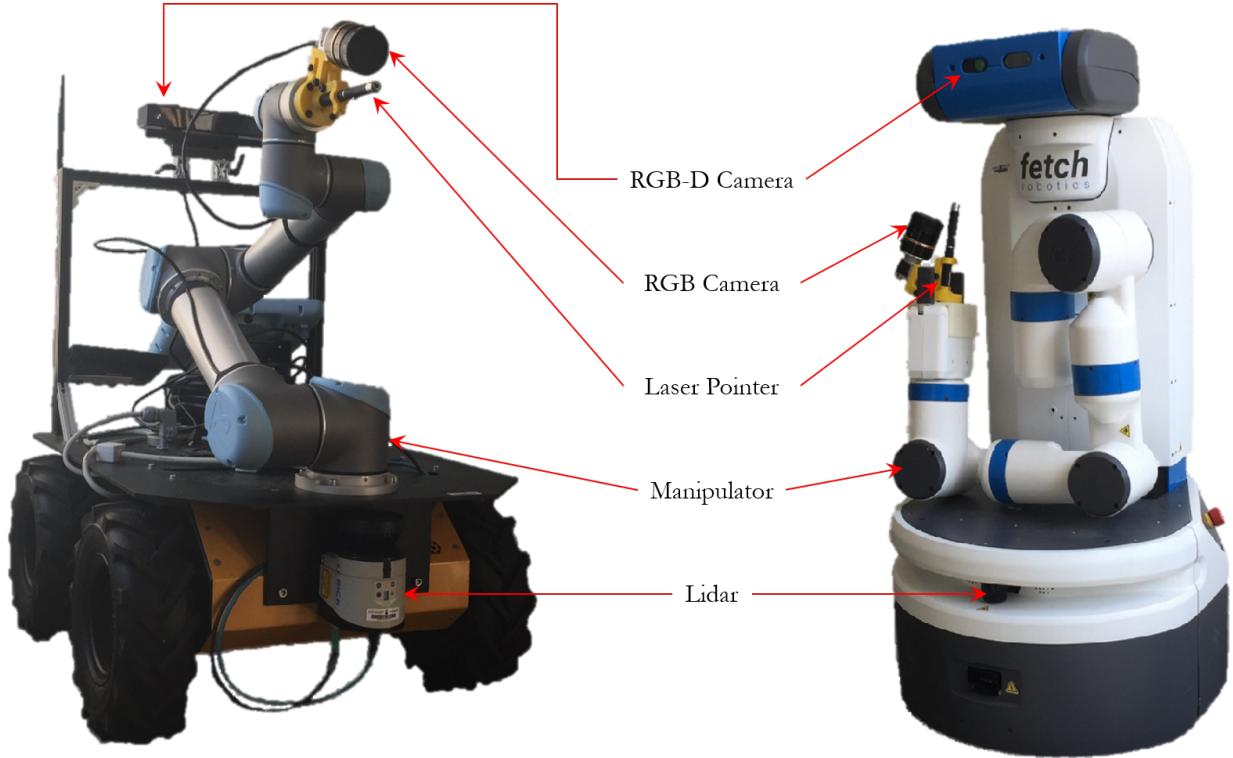


Figure 5.1: Clearpath Robotics' Husky and the Fetch Robotics' Fetch Mobile Manipulator

The Clearpath Robotics' Husky is equipped with a Microsoft Kinect V2, Baser ACA1920-150uc machine vision camera, Sick LMS-151 LIDAR, MicroStrain 3DM GX3-25 IMU, four high precision wheel encoders, and the Universal Robotics UR5 6 DOF manipulator with a 5kg payload capacity. In regards to the specific computer hardware, the Clearpath Robotics' Husky is equipped with a Gigabyte BRIX GB-BXi7G3-760 which includes an Intel i7 processor, NVIDIA GeForce GTX 760 GPU, 16 GB DDR3 RAM, 250 GB SSD, and two USB3.0 ports.

The Fetch Robotics' Fetch is equipped with a Primesense Carmine 1.09, Baser ACA1920-150uc machine vision camera, Sick TIM571 LIDAR, two 6 DOF IMUs, two high precision wheel encoders, and the Fetch Robot's proprietary 7 DOF manipulator with a 6 kg payload capacity. In regards to the specific computer hardware, the Fetch Robotics' Fetch is equipped with a standard Mini-ITX motherboard, which includes an Intel i5 processor, 16 GB DDR3 RAM, 120 GB SSD, and three USB 2.0 ports.

5.2 Software Architecture

The Clearpath Robotics' Husky and the Fetch Robotics' Fetch both used Ubuntu 14.04 and ROS Indigo for the experiments. The Clearpath Robotics' Husky used an extended Kalman filter through ROS's robot_localization package [123] to fuse its wheel encoder and IMU measurements together, while the Fetch Robotics' Fetch used an Unscented Kalman Filter. Both platforms used this corrected odometry, as well as laser scans from their respective LIDAR to produce a map of the area of operation (AO) using ROS's KartosLAM package [50]. Both robot's use ROS's Navigation Stack [55] to localize themselves in the map using the AMCL package [57] and implement trajectory rollout through the base_local_planner package [61] for path planning and obstacle avoidance. Both robots used ROS's ar_track_alvar package [124] to identify and localize augmented reality (AR) tags. For motion planning, both robot's used ROS's Moveit! [62], which was setup to use the TRAC-IK [75] numerical solver. For perception, both robot's used OpenCV 2.4 [125] to perform necessary operations on the images provided by the Basler machine vision camera.

5.3 Experiments

The purpose of this section is to explain the experimental setup of each test, as well as how the data presented in Section 5.4 was collected. In addition, the camera calibration setup and the software used for the calibration will be presented.

5.3.1 Camera Calibration Setup

Figure 5.2 shows the experimental setup used to calibrate both the Microsoft Kinect V2, the Baser ACA1920-150uc, as well as the laser pointer in the Baser ACA1920-150uc's camera frame. For each calibration the camera and checkerboard pattern were affixed to individual tri-pods. The Microsoft Kinect V2 was calibrated using the kinect2_calibration package [126], while the Baser ACA1920-150 was calibrated using Jean-Yves Bouguet camera calibration toolbox for Matlab [127].



Figure 5.2: Camera Calibration Setup.

The `kinect2_calibration` package and Bouguet's camera calibration toolbox both follow the calibration procedure proposed by Dr. Zhang in [128]. First, images of a checkerboard pattern affixed to a planar surface are taken by the camera at different distances and orientations. In these images, the feature points in the checkerboard pattern are detected, as depicted in Figure 5.3. These features are used to estimate the five intrinsic parameter of the camera, as well as all the extrinsic parameters using the closed-form equations developed by Dr. Zhang in [128]. The coefficients of radial distortion are then solved using linear least-square. After which, all parameters are refined through nonlinear minimization.

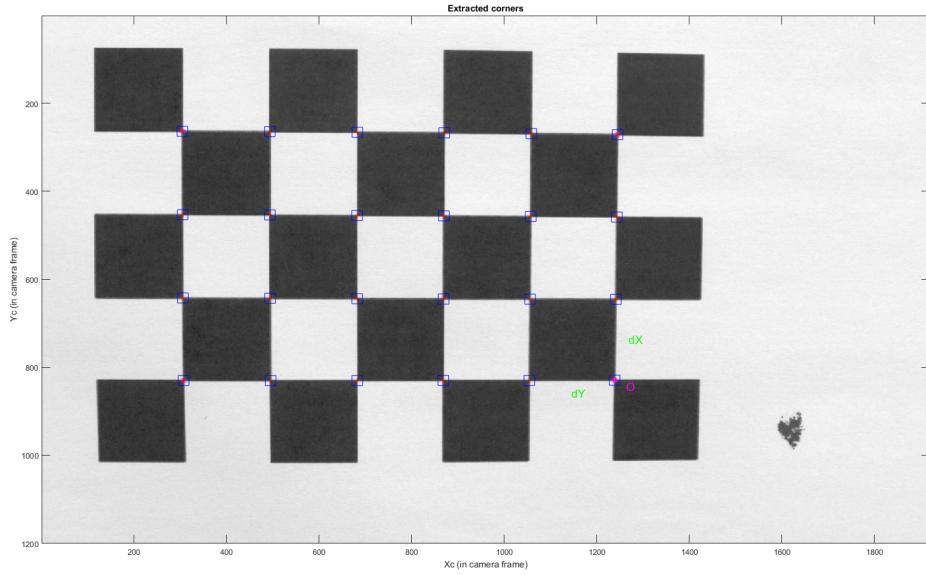


Figure 5.3: Extracted Corners of Calibration Pattern.

After the Baser ACA1920-150 camera was calibrated, the camera and the laser pointer were affixed to the mount shown in Figure 5.1. After which, the laser pointer 3-D ray equation in the camera frame can be calculated by minimizing the error between at least 2 known 3-D coordinates in the camera frame. Thirty points were used to calibrate the 3-D rays equation of the laser pointer for testing. Figure 5.4 shows a checkboard, along with a laser point produced by the laser pointer. The extrinsic parameters of each image are calculated using the features in the checkerboard. Using the camera's extrinsics, the plane equation of the checkerboard was calculated in camera coordinates. These 3-D camera coordinates points were found by finding line-plane intersection between the plane of the checkerboard and 3-D ray of the laser point, which is calculated by finding the line connecting the camera's center point and the locations of laser point in camera coordinates. The same procedure presented in Section 4.6 was used to find the center pixel of the laser point.

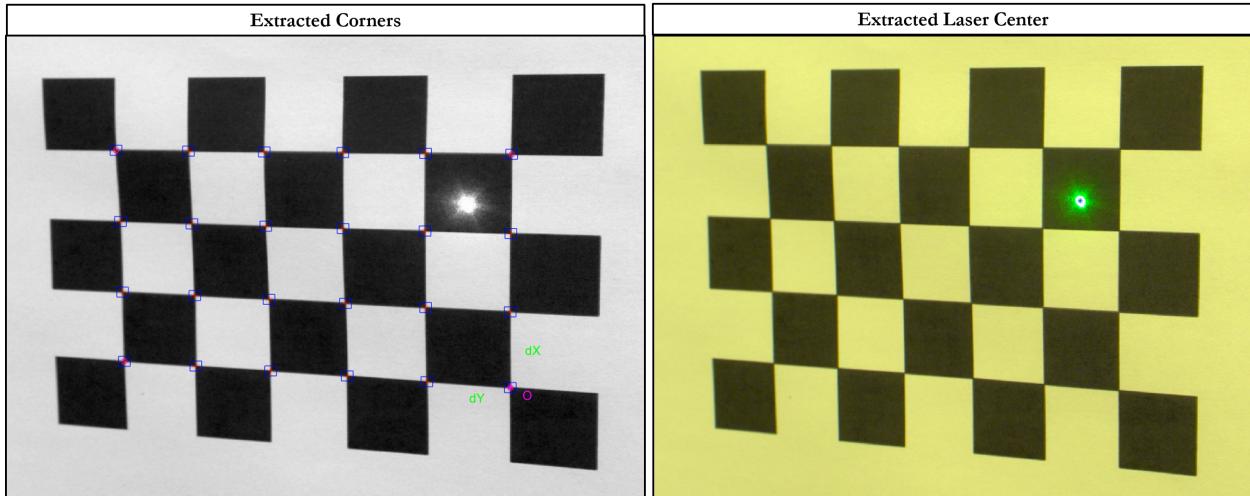


Figure 5.4: Extracted Corners and Laser Center

5.3.2 Navigation System Experimental Setup

Figure 5.5 shows the experimental setup for the navigation system test. Both the Clearpath Robotics' Husky, as well as the Fetch Robotics' Fetch navigated to specified positions 5, 10, and 20 meters away from a specified start position. The absolute difference in the x and y directions, as well as the angular yaw between the robot's final positons and the set reference point were measured. This test was run continually as the robot proceed to each 5, 10, and 20 meter location, 20 times each.

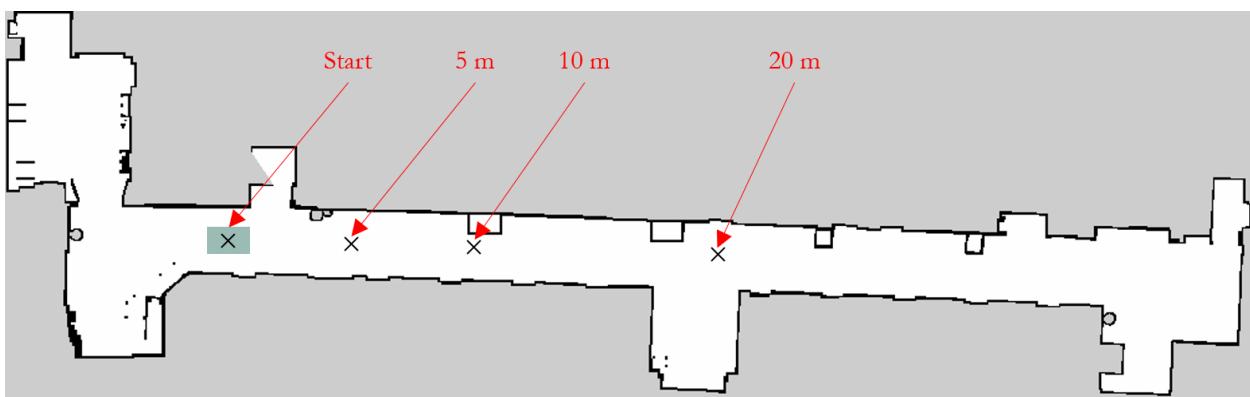


Figure 5.5: Navigation System Experimental Setup.

5.3.3 Augmented Reality Tag Detection, Identification, and Localization System Experimental Setup

Figure 5.6 shows the experimental setup for the augmented reality (AR) tag detection, identification, and localization tests. Both the Clearpath Robotics' Husky, as well as the Fetch Robotics' Fetch were placed at locations a specific x and y distance away, as well as at a set angle from the drilling station's augmented reality (AR) tag, given the maximum error from the navigation test. Each robot's RGB-D sensor was used to detect, identify and localize the specific AR tag. After the feature points for that operation were populated, the absolute difference in x, y, and z directions were measured. For each location, the experiment was conducted five times with 50 features, including the AR tag.

5.3.4 Drill Operation Experimental Setup

Figure 5.6 shows the experimental setup for the drilling operation test. First the sub-millimeter accuracy of the Universal Robotics' 6 DOF UR5 and the Fetch Robotics' 7DOF manipulator was verified. After which, the accuracy of the feature detection, tracking, and localization pipeline was tested using the Baser ACA1920-150uc camera and laser pointer. This accuracy was determined by calculating the difference between the distance measured between the laser point and hole center produced by the feature detection, tracking and localization framework and the ground truth distance between the laser center and hole center measured by hand. This test was performed with the end-effector a set distance off the work surface for 20 locations with the following distances between the laser point and corner: less than 5 mm, between 5 and 10 mm, between 10 and 15 mm, and between 15 and 20 mm.

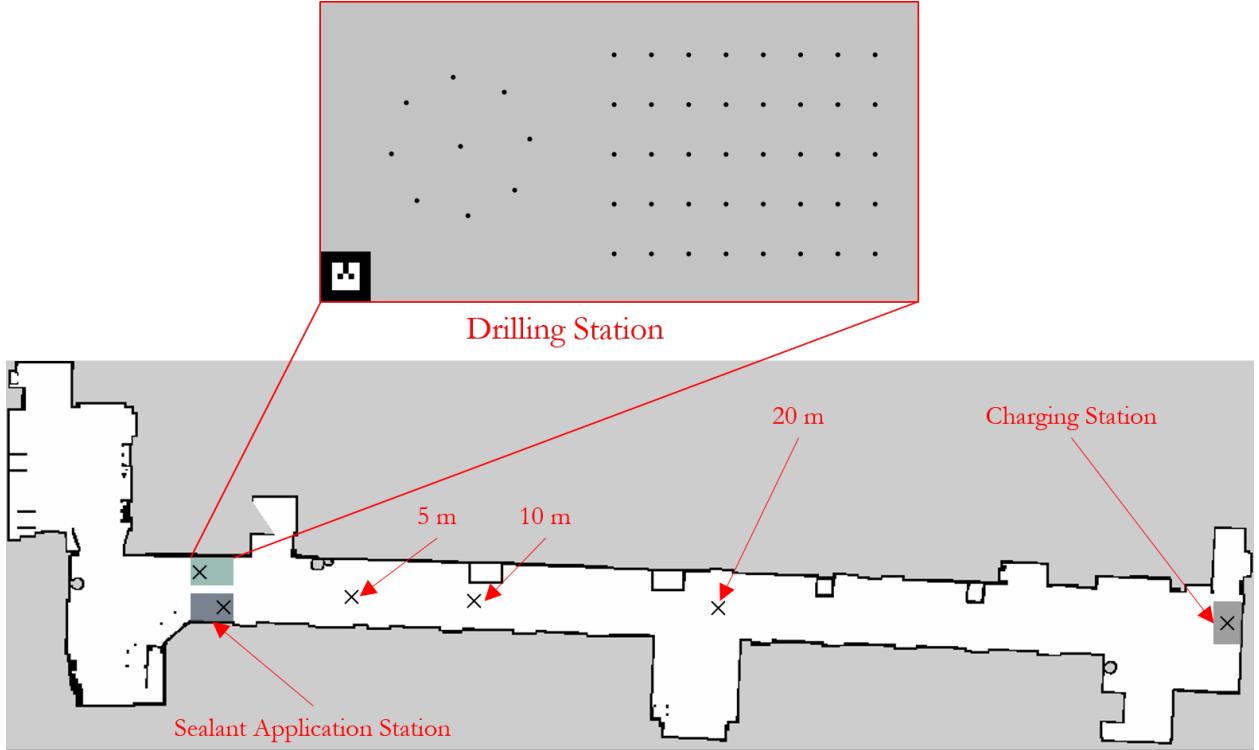


Figure 5.6: Drilling Operations Experimental Setup.

5.3.5 Sealant Application Experimental Setup

Figure 5.7 shows the experimental setup for the sealant operation test. First the sub-millimeter accuracy of the Universal Robotics' 6 DOF UR5 and the Fetch Robotics' 7DOF manipulator was verified. After which, the accuracy of the feature detection, tracking, and localization pipeline was tested using the Baser ACA1920-150uc camera and laser pointer. This accuracy was determined by calculating the difference between the distance measured between the laser point and corner produced by the feature detection, tracking and localization framework and the ground truth distance between the laser center and corner measured by hand. This test was performed with the end-effector a set distance off the work surface for 20 locations with the following distances between the laser point and corner: less than 5 mm, between 5 and 10 mm, between 10 and 15 mm, and between 15 and 20 mm.

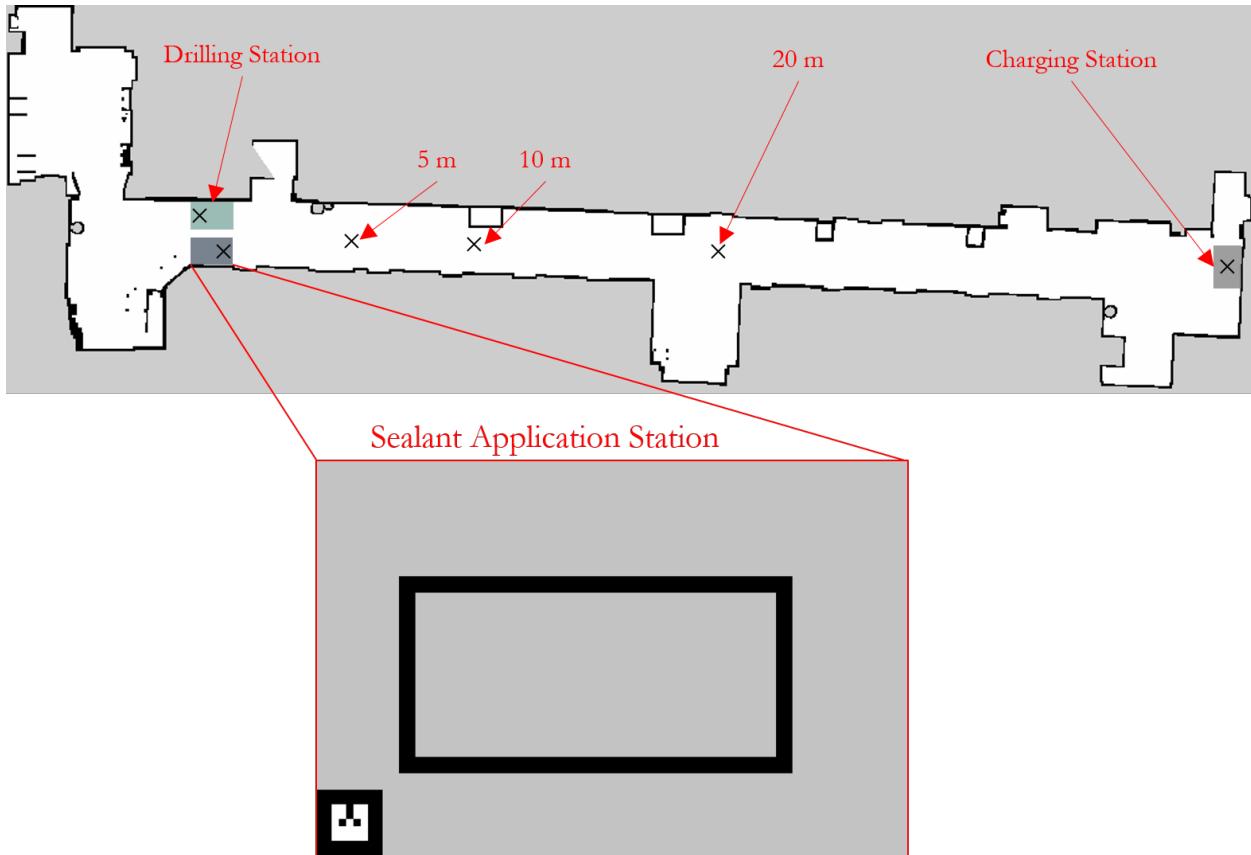


Figure 5.7: Sealant Application Experimental Setup.

5.4 Results and Discussion

The purpose of this section is to present the accuracy achieved by each stage of the multistage localization approach and system implementation presented in Chapter 4 through the tests presented in Section 5.3. In addition, steps to improve these accuracies will be presented. It should be noted that the coordinate frame used in these tests have positive x as forward, positive y as left, and positive z as up with respect to the robot.

5.4.1 Camera Calibration Accuracy Achieved

The Microsoft Kinect V2 incorporates both an RGB and an IR camera. Each camera's intrinsic parameters and distortion coefficients were calculated through individual calibration. After which, each calibrated camera took pictures of a checkerboard at the same distance and orientation. The extrinsic parameters computed for each image were used to calculate

the pose of each camera in the other's camera frame. The Microsoft Kinect V2's RGB, IR, and Pose pixel error from the calibrations were 0.376 pixels, 0.126 pixels, and .284719 pixels respectively. The Baser ACA1920-150uc pixel error from calibration was .418; while, the laser pointer's 3-D ray in the Baser ACA1920-150uc camera frame minimum mean square orthogonal distance to the line was 0.6468 mm.

5.4.2 Navigation System Accuracy Achieved

Table 5.1 and Figures 5.8 and 5.9 show the x, y and ψ error associated with the Clearpath Robotics' Husky in reaching specified positions 5, 10, and 20 meters away from a designated start position. Table 5.2 and Figures 5.10 and 5.11 show the x, y and yaw error associated with the Fetch Robotics' Fetch in reaching specified positions 5, 10, and 20 meters away from a designated start position. The following Tables and Figures show that neither the Husky's nor the Fetch's average error per test exceeded 6.4 cm in the x, 4.5 cm in the y, and .03 radians in the ψ . Given that these tests were done consecutively, equating to nearly a kilometer, and that y and ψ deviation decreased over time, while the error in x seems to follow a more sinusoidal pattern, which is probably due to the resampling of the particle filter's particles, the error is fairly low and did not show any significant effect on the steps in the multistage localization that followed it. However, due to the fact that these robots may need to recharge mid-task or the task's object(s) may be placed in a static location, the features populated by the AR tag localization may need to be used multiple times or there could be a desire to use them continuously. Consequently, the accuracy of the navigation system needs to be improved to mm level. The tests conducted were done in a long hallway with very few artificial features. A manufacturing facility or construction site tend to be filled with artificial features, thus improvement in the accuracy of the adaptive Monte Carlo localization should be seen in these environments. Additional performance would be seen by increasing the number of sensors fused together by the Kalman filter to improve the filtered odometry sent to the AMCL as seen in [122]. Rowekamper et al. in [129] was able to achieve an accuracy of \pm 5 mm through the addition of a second LIDAR sensor. It should be noted that the increased error seen by the Husky over the Fetch was most likely due to the Husky's IMU being unresponsive during the tests conducted, which supports the above research. However, much of this error could also be attributed to the difference in the steering techniques used by the two robots. The Fetch uses differential steering, due to its two driven wheels, while the Husky uses skid steering, due to its four wheels being driven by two motors, which results in a larger turn radius and therefore makes the Husky less maneuverable.

Table 5.1: Global Localization Accuracy Given Distance From Start - Clearpath Robotics' Husky.

	x (cm)	x std (cm)	y (cm)	y std (cm)	ψ (rad)	ψ std (rad)
5 m	6.29	4.20	4.50	3.58	0.03	0.02
10 m	4.16	3.16	2.88	2.97	0.03	0.03
20 m	6.35	4.55	2.37	1.89	0.02	0.01
Mean	5.60	3.97	3.25	2.81	0.03	0.02

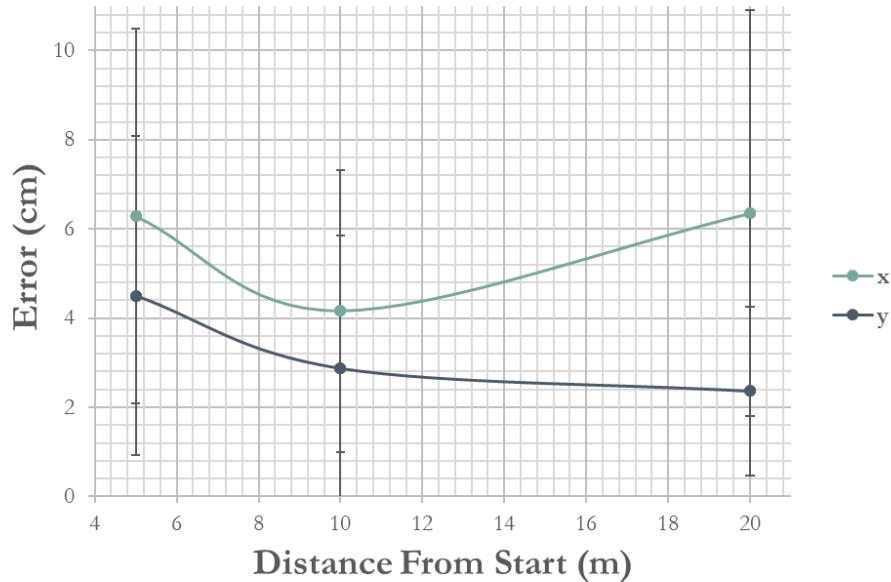


Figure 5.8: Global X and Y Localization Accuracy Given Distance From Start - Clearpath Robotics' Husky.

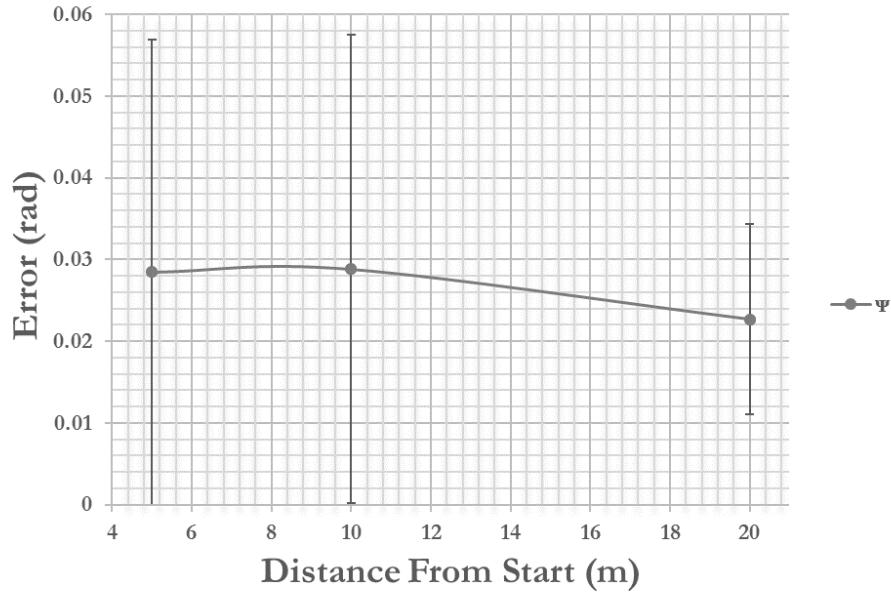


Figure 5.9: Global ψ Localization Accuracy Given Distance From Start - Clearpath Robotics' Husky.

Table 5.2: Global Localization Accuracy Given Distance From Start - Fetch Robotics' Fetch.

	x (cm)	x std (cm)	y (cm)	y std (cm)	ψ (rad)	ψ std (rad)
5 m	4.25	2.50	1.48	1.23	0.03	0.02
10 m	3.93	2.62	2.57	1.66	0.03	0.02
20 m	5.82	3.43	1.52	1.11	0.03	0.02
Mean	4.67	2.85	1.85	1.33	0.03	0.02

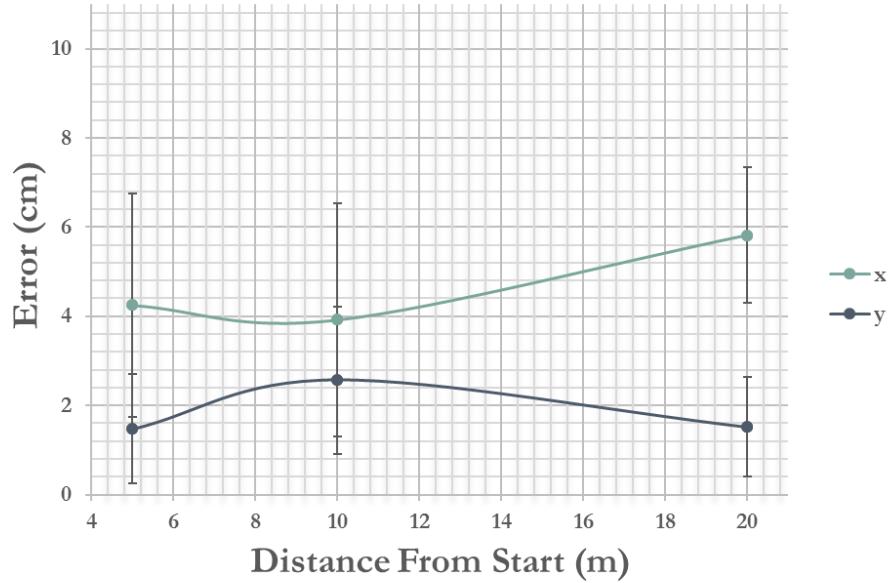


Figure 5.10: Global X and Y Localization Accuracy Given Distance From Start - Fetch Robotics' Fetch.

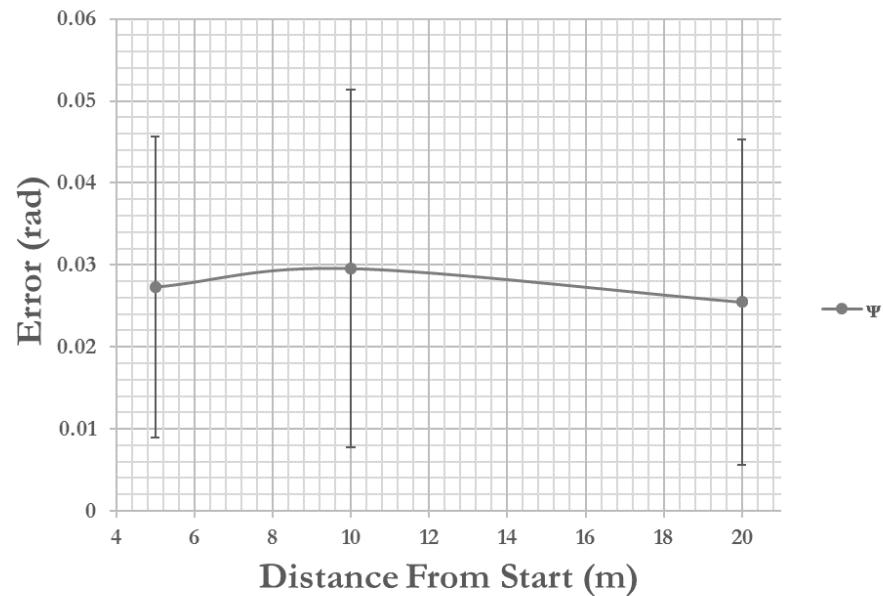


Figure 5.11: Global ψ Localization Accuracy Given Distance From Start - Fetch Robotics' Fetch.

5.4.3 Augmented Reality Tag Accuracy Achieved

Table 5.3 and Figure 5.12 show the augmented reality (AR) tag average localization error in the x, y and z directions using 250 features for each of six different tests using the Primesense Carmine 1.09. Test 1 and 2, had the RGB-D sensor centered on the AR tag but were offset in the negative x direction (backwards) 1.0 m and 0.75 m respectively. Test 3 and 5, as well as 4 and 6 were offsets of Test 1 and 2 respectively. Test 3 and 4 were offset 16.1 cm in the negative x direction, 7.5 cm in the positive y direction (left), and 0.09 radian in the negative ψ direction. Test 5 and 6 were offset 16.1 cm in the negative x direction, 7.5 cm in the negative y direction (right), and 0.09 radians in the positive ψ direction. These distance offsets were determined by taking the average error from the Fetch's navigation tests and adding three standard deviations. The tests show an average error of approximately 3.3 cm in the x direction, 12.6 cm in the y direction, and 11.8 cm in the z. The difference in error between Tests 1 and 2, 3 and 4, and 5 and 6, are either approximately the same or lower as the x distance from the AR tag is decreased. This makes mathematical sense in that as the sensor gets closer to the AR Tag as it is in Tests 2, 4, and 6, the ground sample distance (GSD) is lower equating to a higher spatial resolution. Also, it appears that moving the RGB-D sensor slightly in the positive y direction (left) and negative ψ seems to decrease overall error as well.

Table 5.3: Augmented Reality Accuracy Given Specific Start Conditions - Primesense Carmine 1.09.

	x (mm)	x std (mm)	y (mm)	y std (mm)	z (mm)	z std (mm)
Test 1	50.41	22.81	148.87	12.33	150.96	12.61
Test 2	23.71	16.06	151.82	6.86	117.57	12.81
Test 3	23.06	14.85	128.39	11.48	173.24	23.24
Test 4	24.36	17.47	138.66	6.77	42.80	16.37
Test 5	37.19	14.85	183.21	7.30	115.85	16.27
Test 6	39.16	23.52	7.27	3.70	105.24	9.16
Mean	32.98	18.26	126.37	8.07	117.61	15.08

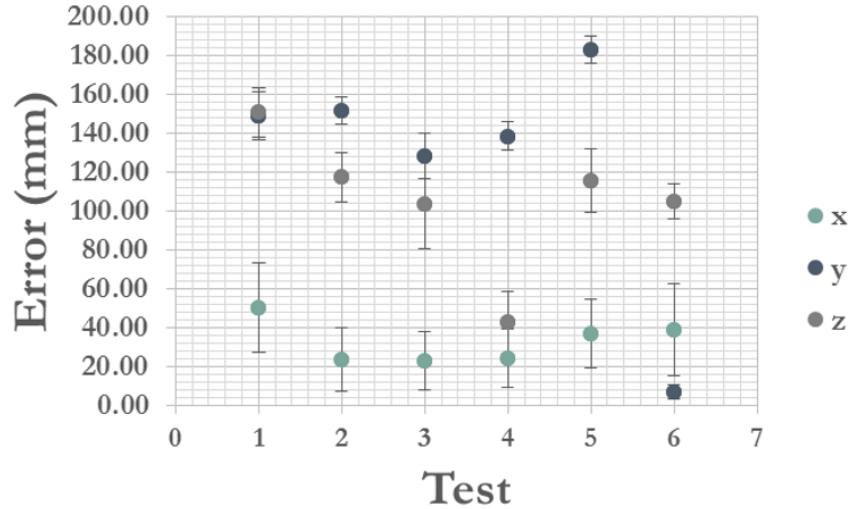


Figure 5.12: Augmented Reality Accuracy Given Specific Start Conditions - Primesense Carmine 1.09.

Table 5.4 and Figure 5.13 show the augmented reality (AR) tag average localization error in the x, y and z, directions using 250 features for each of three different tests using the Microsoft Kinect V2. Test 1 had the RGB-D sensor centered on the AR tag but offset in the negative x direction (backwards) 1.0 m. Test 2 and 3 were offsets of Test 1. Test 2 was offset 20.0 cm in the negative x direction, 15.2 cm in the positive y direction (left), and 0.115 radian in the negative ψ direction. Test 3 was offset 20.0 cm in the negative x direction, 15.2 cm in the negative y direction (right), and .115 radians in the positive ψ direction. These distance offsets were determined by taking the average error from the Husky's navigation tests and adding three standard deviations. The tests show an average error of approximately 8.8 cm in the x direction, 3.0 cm in the y direction, and 5.5 cm in the z. Also, it appears that moving the RGB-D sensor slightly in the positive y direction (left) and negative ψ seems to decrease error overall error in the y and z as seen with the Primesense Carmine 1.09.

Table 5.4: Augmented Reality Accuracy Given Specific Start Conditions - Microsoft Kinect V2.

	x (mm)	x std (mm)	y (mm)	y std (mm)	z (mm)	z std (mm)
Test 1	14.32	9.75	4.04	1.65	84.77	2.50
Test 2	163.85	57.58	36.32	9.65	25.31	12.46
Test 3	84.33	44.37	50.31	5.48	55.24	8.57
Mean	87.50	37.24	30.22	5.59	55.10	7.84

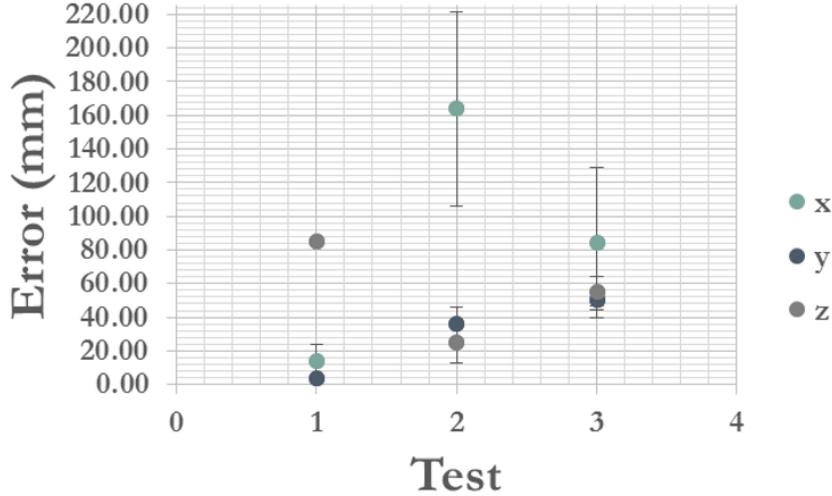


Figure 5.13: Augmented Reality Accuracy Given Specific Start Conditions - Microsoft Kinect V2.

Overall, the Microsoft Kinect V2 performs significantly better than the Primesense Carmine 1.09. This is due to the fact that the Microsoft Kinect V2 is a newer sensor, which has greater resolution, horizontal and vertical field of view (FOV), and onboard processing. However, neither the Primesense Carmine 1.09 nor the Microsoft Kinect V2 performed exceptionally. Sources of error include, error in the RGB-D camera's intrinsic calibration, error associated in localizing the AR tag itself, error in the location estimate of the camera with respect to the robot's base link, and error in the location estimate of the robot's base link in the map. In order to reduce these sources of error, the RGB-D camera would need to be calibrated on a rig which was certified to achieve sub-millimeter to millimeter level accuracy. In order to address the accuracy error inherent in the AR tag localization, different AR tag detection libraries could be explored, such as AprilTags [117] and ARToolKit [119], QR codes could be used, or the visual high-precision technique presented in Section 2.3, which uses a calibration plate and can achieve an accuracy of ± 0.1 mm, could be used. The positions of the RGB-D sensor's with respect to robot's base would need to be measured to within sub-millimeter to millimeter level accuracy. This can be achieved through the use of ROS's robot_calibration package [130], which would use the calibrated RGB-D camera and a checkerboard to adjust the position of RGB-D camera in the robot's URDF. In addition, newer RGB-D sensors such as the Orbbec Astra, Intel's RealSense R200, and Structure's Sensor or a stereo vision camera such as ZED's Stereo Vision Camera, Point Grey's Bumblebee2, and the DUO MLX could be used. In addition to providing depth information, stereo cameras have the ability to work both indoors and outdoors, which could be extremely advantageous depending on the environment. While millimeter level accuracy could be achieved, attaining sub-millimeter accuracy at a distance of 1 m or more would be nearly impossible with currently technology.

Consequently, the multistage localization approach would still be necessary for high precision operations.

5.4.4 Drilling Operation Accuracy Achieved

Table 5.5 and Figure 5.14 show the x, y and z error associated with the feature localization, detection, and tracking pipeline using 20 features for each of four different tests done using the Basler ACA1920-150uc and laser pointer, which were approximately 20 cm away from the work surface. This setup gives the Basler ACA1920-150uc with a 16 mm focal length lens, a horizontal and vertical field of view (FOV) of 11.9 and 7.6 cm respectively, as well as a ground sample distance (GSD) of .0062 cm. While the y accuracy increases slightly as the distance away from the feature is increased, the x accuracy decreases slightly and the z accuracy decreases significantly. However, due to the fact that z and x have decreasing curves as the laser point gets closer to the feature, the system is able to achieve sub-millimeter level accuracy in the y and z, and very low millimeter level accuracy in the x due to the method's control loop. The Basler ACA1920-150uc with a 16 mm focal length lens should be able to achieve localization accuracies close to its GSD of .006 mm, which exceeds the capability of both manipulators used. In order to improve the accuracy to this level, the camera and laser should be calibrated on rig certified to this level of accuracy. Alternatively, a stereo camera could be developed through the purchase of another machine vision camera and those cameras calibrated on a rigs certified to the accuracy mentioned above. In addition, the detection and localization pipeline would need to be improved by using more recent and robust computer vision techniques in order to find the laser point and feature to within pixel, if not sub-pixel resolution.

Table 5.5: Drilling Operation Accuracy Given Distance From Work Surface.

	x (mm)	x std (mm)	y (mm)	y std (mm)	z (mm)	z std (mm)
Distance Less Than 5 mm from Hole Center	1.44	1.25	0.70	0.43	0.82	0.56
Distance Between 5 and 10 mm from Hole Center	1.89	1.56	0.66	0.39	1.89	0.46
Distance Between 10 and 20 mm from Hole Center	2.15	1.83	0.53	0.30	4.06	0.70
Distance Between 20 and 30 mm from Hole Center	2.71	2.71	0.31	0.21	10.25	1.38
Mean	2.05	1.83	0.55	0.33	4.25	0.77

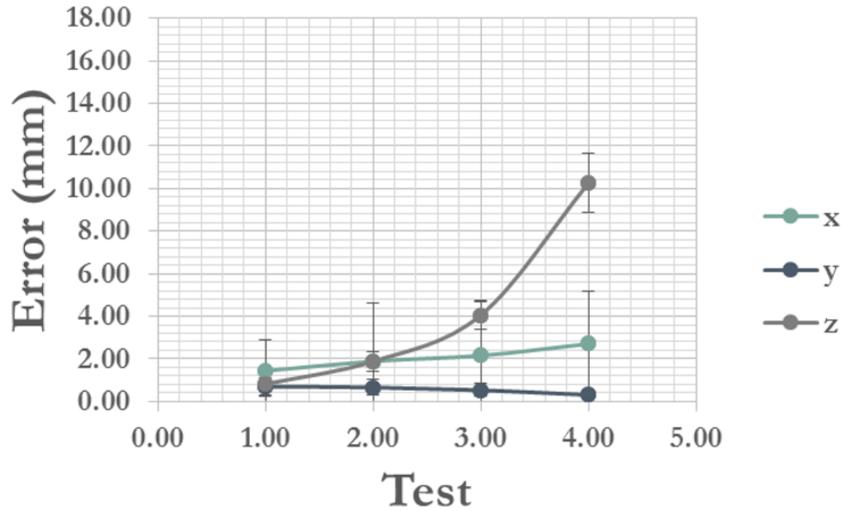


Figure 5.14: Drilling Operation Accuracy Given Distance From Feature Point.

5.4.5 Sealant Application Accuracy Achieved

Table 5.6 and Figure 5.15 show the x, y and z, error associated with the feature localization, detection, and tracking pipeline using 20 features for each of 4 different tests done using the Basler ACA1920-150uc and laser pointer, which were approximately 20 cm away from the work surface. This setup gives the Basler ACA1920-150uc with a 16 mm focal length lens, a horizontal and vertical field of view (FOV) of 11.9 and 7.6 cm respectively, as well as a ground sample distance (GSD) of .0062 cm. While the y accuracy increases slightly as the distance away from the feature is increased, the x accuracy decreases slightly and the z accuracy decreases significantly. However, due to the fact that z and x have decreasing curves as the laser point gets closer to the feature, the system is able to achieve very low millimeter level accuracy in the x, y, and z due to the method's control loop. The Basler ACA1920-150uc with a 16 mm focal length lens should be able to achieve localization accuracies close to its GSD of .006 mm, which exceeds the capability of both manipulators used. In order to improve the accuracy to this level, the camera and laser should be calibrated on rigs certified to this level of accuracy. Alternatively, a stereo camera rig could be developed through the purchase of another machine vision camera and those cameras calibrated on a rig certified to the accuracy mentioned above. In addition, the detection and localization pipeline would need to be improved by using more recent and robust computer vision techniques in order to find the laser point and feature to within pixel, if not sub-pixel resolution.

Table 5.6: Sealant Application Accuracy Given Distance From Work Surface.

	x (mm)	x std (mm)	y (mm)	y std (mm)	z (mm)	z std (mm)
Distance Less Than 5 mm from Corner	2.34	1.29	1.46	0.21	1.17	0.71
Distance Between 5 and 10 mm from Corner	1.60	2.22	1.27	0.22	1.48	0.91
Distance Between 10 and 20 mm from Corner	4.38	2.46	0.78	0.34	3.81	0.32
Distance Between 20 and 30 mm from Corner	3.32	1.95	0.79	0.60	8.21	0.40
Mean	2.91	1.98	1.08	0.34	3.67	0.59

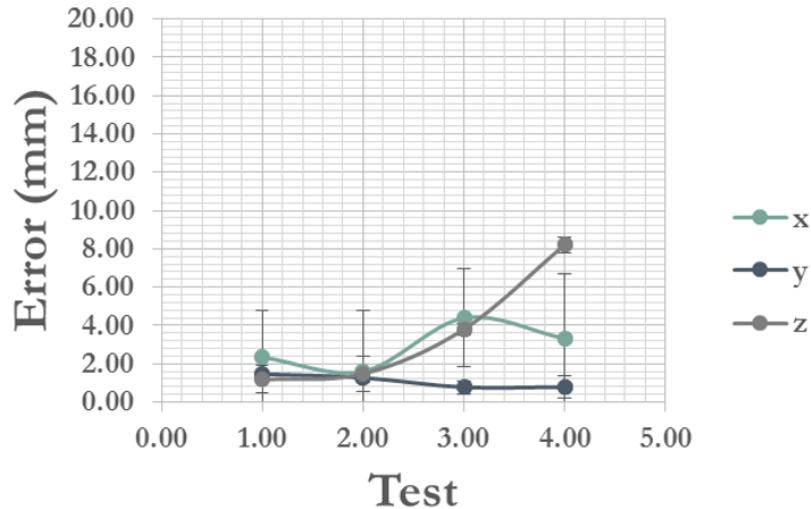


Figure 5.15: Sealant Application Accuracy Given Distance From Feature Point.

5.5 Summary

The purpose of this chapter was to validate each stage of the multistage localization approach and system implementation presented in Chapter 4 through testing. The accuracy of each component of the multistage localization technique was determined on two different ROS enabled robotic platforms, the Clearpath Robotics' Husky and the Fetch Robotics' Fetch. In addition, methods to improve each stage's accuracy was presented. Chapter 6 will present conclusions and suggest future work.

Chapter 6

Conclusions and Future Work

The purpose of this paper was to present a multistage localization approach for an autonomous industrial mobile manipulator (AIMM). This approach allows tasks with an operational scope outside the range of the robot's manipulator to be completed without recalibrating the positon of the end-effector each time the robot's mobile base moves to another position in order to complete said operation. This is achieved by localizing the AIMM within its area of operation (AO) using adaptive Monte Carlo localization (AMCL), which relies on the fused odometry and sensor messages published by the robot, as well as a 2-D map of the AO, which is normally generated using an optimization-based smoothing simultaneous localization and mapping (SLAM) technique. The robot navigates to a predefined start location in the map incorporating obstacle avoidance through the use of a technique called trajectory rollout. Once there, the robot uses its RGB-D sensor to localize an augmented reality (AR) tag in the map frame. Once localized, the identity and the 3-D position and orientation, collectively known as pose, of the tag are used to generate a list of initial feature points and their locations based off of *a priori* knowledge. After the end-effector moves to the rough location of a feature point provided by the AR tag localization, the feature point's location, as well as the end-effector's positon and orientation are refined to within a user specified tolerance through the use of a control loop, which utilizes images from a calibrated machine vision camera and a laser pointer, simulating stereo vision, to localize the feature point in 3-D space using computer vision techniques and basic geometry. This approach was implemented on two different ROS enabled robots, the Clearpath Robotics' Husky and the Fetch Robotics' Fetch, in order to show the utility of the multistage localization approach in executing two tasks which are prevalent in both manufacturing and construction: drilling and sealant application. The proposed approach was able to achieve an average accuracy of ± 1 mm in these operations, verifying it's efficacy for tasks which have a larger operational scope than that of the range of the AIMM's manipulator and its robustness to general applications in manufacturing.

Through the addition of further sensors, such as an extra IMU, LIDAR, or RGB-D camera,

the accuracy of the AMCL can be improved from approximately ± 6 cm to ± 5 mm as shown in [129]. In order to improve the accuracy of the initial feature localization, sensors can be calibrated on certified rigs, additional fiducial marker localization systems can be explored, or the visual high-precision technique presented in Section 2.3, which uses a calibration plate and can achieve an accuracy of ± 0.1 mm, can be used. Further improvements in the initial feature localization could be achieved through calibrating the RGB-D and RGB cameras' position with respect to the base-link to within ± 1 mm through the use of packages like ROS's robot_calibration, as well as upgrading to a newer RGB-D sensor with better resolution, such as the ZED stereo vision camera. The accuracy of the refined feature positions could conceivably be brought down to the ground sample distance (GSD) of the camera, $/pm .006$ mm at 20 cm away from the work surface for the Basler ACA1920-150uc machine vision camera used, which exceeds the capability of the manipulators used on both robot platforms tested. In order to achieve this, the calibration of the camera and laser must be done on a certified rig or a calibrated stereo camera rig used instead. In addition, more recent computer vision and machine learning techniques should be used to localize the feature points to subpixel accuracy.

Before the proposed system can be implemented inside a manufacturing or construction environment, additional future work besides that of increasing the accuracy of the system, must be done. In order to make AIMMs financially viable, they must be able to collaborate together in order to complete a specified task, as well as be easy to reprogram through methods like those suggested in [1] and [6]. Other difficult hurdles exist, such as ensuring that the robot is 100% safe to use around humans. Many methods exist or have been proposed to ensure this; however, options such as disabling the robot when a human enters its area of operation as proposed by Sbanca et al. in [131] will ensure 100% safety, but will not produce the transformable production system needed. AIMMs must be co-Robot, able to interact with humans, while ensuring safety. This can be achieved through the use of gesture controls, which if given act as alarm behavior stopping whatever the robot is doing in favor of the desired action as presented by Camprodon in [132]. Additionally, this could be achieved through speech recognition as presented by Mericli in [133].

Bibliography

- [1] Mikkel Rath Pedersen et al. “Robot skills for manufacturing: From concept to industrial deployment”. In: *Robotics and Computer-Integrated Manufacturing* 37 (2016), pp. 282–291. ISSN: 07365845. DOI: 10.1016/j.rcim.2015.04.002. URL: <http://dx.doi.org/10.1016/j.rcim.2015.04.002>.
- [2] George Michalos et al. “Performance Assessment of Production Systems with Mobile Robots”. In: *Procedia CIRP* 41 (2016), pp. 195–200. ISSN: 22128271. DOI: 10.1016/j.procir.2015.12.097.
- [3] Victor M. Cedeno-Campos et al. “Highly flexible self-reconfigurable systems for rapid layout formation to offer manufacturing services”. In: *Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013*. 2013, pp. 4819–4824. ISBN: 9780769551548. DOI: 10.1109/SMC.2013.820.
- [4] Mads Hvilsted et al. “Autonomous industrial mobile manipulation (AIMM): past, present and future”. In: *Industrial Robot: An International Journal* 39.2 (2012), pp. 120–135. ISSN: 0143-991X. DOI: 10.1108/01439911211201582.
- [5] Simon Bøgh et al. “Does your Robot have Skills?” In: *Proceedings of the 43rd International Symposium on Robotics*. 2012, p. 6. URL: <http://forskningsbasen.deff.dk/Share.external?sp=S9f184d42-459a-4d30-871c-bab8082cfbba%7B%5C%7Dsp=Saa>.
- [6] Mikkel Rath Pedersen, Dennis Levin Herzog, and Volker Krüger. “Intuitive skill-level programming of industrial handling tasks on a mobile manipulator”. In: *IEEE International Conference on Intelligent Robots and Systems*. IROS. 2014, pp. 4523–4530. ISBN: 9781479969340. DOI: 10.1109/IROS.2014.6943203.
- [7] Simon Boøgh et al. “Identifying and evaluating suitable tasks for autonomous industrial mobile manipulators (AIMM)”. In: *International Journal of Advanced Manufacturing Technology* 61.5-8 (2012), pp. 713–726. ISSN: 02683768. DOI: 10.1007/s00170-011-3718-3.
- [8] Hongtai Cheng, Heping Chen, and Yong Liu. “Object handling using autonomous industrial mobile manipulator”. In: *2013 IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems, IEEE-CYBER 2013*. 2013, pp. 36–41. ISBN: 9781479906109. DOI: 10.1109/CYBER.2013.6705416.

- [9] Ole Madsen et al. “Integration of mobile manipulators in an industrial production”. In: *Industrial Robot: An International Journal* 42.1 (2015), pp. 11–18. ISSN: 0143-991X. DOI: 10.1108/IR-09-2014-0390. URL: <http://www.emeraldinsight.com/doi/abs/10.1108/IR-09-2014-0390>.
- [10] R. Bischoff. *Robotics-enabled logistics and assistive services for the transformable factory of the future (TAPAS)*. Tech. rep. European Commission, 2010.
- [11] C Balaguer. “NOWADAYS TRENDS IN ROBOTICS AND AUTOMATION IN CONSTRUCTION INDUSTRY: TRANSITION FROM HARD TO SOFT ROBOTICS”. In: *Proceedings of International Symposium on Automation and Robotics in Construction*. 2004.
- [12] Eddy M. Rojas and Peerapong Aramvareekul. “Is Construction Labor Productivity Really Declining?” In: *Journal of Construction Engineering and Management* 129.1 (2003), pp. 41–46. ISSN: 0733-9364. DOI: 10.1061/(ASCE)0733-9364(2003)129:1(41).
- [13] Chen Feng et al. “Vision guided autonomous robotic assembly and as-built scanning on unstructured construction sites”. In: *Automation in Construction* (2015). ISSN: 09265805. DOI: 10.1016/j.autcon.2015.06.002.
- [14] Davide Ronzoni et al. “AGV Global Localization Using Indistinguishable Artificial Landmarks”. In: *IEEE International Conference on Robotics and Automation*. 2011. ISBN: 9781612843858.
- [15] I Loevsky and I Shimshoni. “Reliable and efficient landmark-based localization for mobile robots”. In: *Robotics and Autonomous Systems* 58 (2010), pp. 520–528. DOI: 10.1016/j.robot.2010.01.006.
- [16] Huijuan Zhang et al. “Localization and navigation using QR code for mobile robot in indoor environment”. In: *IEEE Conference on Robotics and Biomimetics*. 2015, pp. 2501–2506. ISBN: 9781467396752.
- [17] Kimihiro Okuyama, Tohru Kawasaki, and Valeri Kroumov. “Localization and Position Correction for Mobile Robot Using Artificial Visual Landmarks”. In: *Proceedings of the 2011 International Conference on Advanced Mechatronic Systems*. 2011, pp. 414–418.
- [18] Emidio Digiampaolo and Francesco Martinelli. “Mobile Robot Localization Using the Phase of Passive UHF RFID Signals”. In: *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS* 61.1 (2014), pp. 365–376.
- [19] Sangdo Park and Hongchul Lee. “Self-Recognition of Vehicle Position Using UHF Passive RFID Tags”. In: *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS* 60.1 (2013), pp. 226–234.
- [20] Moustafa Youssef and Ashok Agrawala. “The Horus location determination system”. In: *Wireless Networks* 14.3 (2008), pp. 357–374. DOI: 10.1007/s11276-006-0725-7.

- [21] Lingjun Gao. "Channel State Information Fingerprinting Based Indoor Localization: a Deep Learning Approach". Master of Science. Auburn University, 2015.
- [22] Michael Shneier and Roger Bostelman. *Literature Review of Mobile Robots for Manufacturing Literature Review of Mobile Robots for Manufacturing*. Tech. rep. National Institute of Standards and Technology, 2015.
- [23] Alonzo Kelly. "Localization from Appearance". In: *International Journal of Robotics Research* 19.11 (2000), pp. 1104–1125.
- [24] Doaa M. a Latif et al. "Comparison of Optimization Techniques for 3D Graph-based". In: *Proceedings of the 4th European Conference of Computer Science (ECCS '13) Recent Advances in Information Science*. October. 2013, p. 288.
- [25] Raffaele Limosani et al. "Enabling Global Robot Navigation Based on a Cloud Robotics Approach". In: *International Journal of Social Robotics* 8.3 (2016), pp. 371–380. ISSN: 1875-4805. DOI: 10.1007/s12369-016-0349-8.
- [26] Rasmus S Andersen et al. "Fast calibration of industrial mobile robots to workstations using QR codes". In: *International Symposium on Robotics (ISR)*. 2013, pp. 1–6. DOI: 10.1109/ISR.2013.6695636.
- [27] Hongtai Cheng and Heping Chen. "Object Handling Using Autonomous Industrial Mobile Manipulator". In: *Proceedings of the 2013 IEEE Conference on Cyber Technology in Automation, Control and Intelligent Systems*. 2013, pp. 36–41. ISBN: 9781479906123.
- [28] Mikkel Rath Pedersen. "Integration of the KUKA Light Weight Robot in a mobile manipulator". Master of Science. Aalborg University, 2011.
- [29] Mads Hvilsted et al. "Calibration Techniques for Industrial Mobile Manipulators: Theoretical configurations and Best practices". In: *41st International Symposium on and 2010 6th German Conference on Robotics (ROBOTIK)*. 2010. ISBN: 9783800732739.
- [30] Morgan Quigley et al. "ROS: an open-source Robot Operating System". In: *ICRA Workshop on Open Source Software*. 2009.
- [31] *ROS.org - Powering the world's robots*. URL: <http://www.ros.org/> (visited on 10/18/2016).
- [32] *ROS Concepts - ROS Wiki*. URL: <http://wiki.ros.org/ROS/Concepts> (visited on 10/18/2016).
- [33] James David Burton. "Development and Characterization of an Interprocess Communications Interface and Controller for Bipedal Robots". Masters of Science. Virginia Polytechnic Institute and State University, 2015.
- [34] Alonzo Kelly. *Mobile Robotics: Mathematics, Models, and Methods*. New York, NY, USA: Cambridge University Press, 2013.
- [35] *TF - ROS Wiki*. URL: <http://wiki.ros.org/tf> (visited on 10/19/2016).

- [36] *TF2 - ROS Wiki*. URL: <http://wiki.ros.org/tf2> (visited on 10/19/2016).
- [37] *URDF - ROS Wiki*. URL: <http://wiki.ros.org/urdf> (visited on 10/19/2016).
- [38] *OpenSLAM - What is SLAM?* URL: <http://openslam.org/> (visited on 10/28/2016).
- [39] J J Leonard and H F Durrant-Whyte. “Simultaneous map building and localization for an autonomous mobile robot”. In: *Intelligent Robots and Systems '91. 'Intelligence for Mechanical Systems, Proceedings IROS '91. IEEE/RSJ International Workshop on.* 91. 1991, 1442–1447 vol.3.
- [40] Randall C Smith and Peter Cheeseman. “On the Representation and Estimation of Spatial Uncertainty”. In: *The International Journal of Robotics Research* 5.4 (1986), pp. 56–68.
- [41] R. Smith, M. Self, and P. Cheeseman. “Estimating Uncertain Spatial Relationships in Robotics”. In: *Autonomous Robot Vehicles* 4.April (1990), pp. 167–193.
- [42] Giorgio Grisetti et al. “A tutorial on graph-based SLAM”. In: *IEEE Intelligent Transportation Systems Magazine* 2.4 (2010), pp. 31–43.
- [43] Kurt Konolige et al. “Efficient sparse pose adjustment for 2D mapping”. In: *IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings.* 2010, pp. 22–29.
- [44] *HectorSLAM - ROS Wiki*. URL: http://wiki.ros.org/hector%7B%5C_%7Dslam (visited on 10/28/2016).
- [45] Stefan Kohlbrecher et al. “A flexible and scalable SLAM system with full 3D motion estimation”. In: *9th IEEE International Symposium on Safety, Security, and Rescue Robotics, SSRR 2011.* 2011, pp. 155–160.
- [46] *CoreSLAM - ROS Wiki*. URL: <http://wiki.ros.org/coreslam> (visited on 10/28/2016).
- [47] *GMapping - ROS Wiki*. URL: <http://wiki.ros.org/gmapping?distro=kinetic> (visited on 10/28/2016).
- [48] Bruno Steux and Oussama El Hamzaoui. “tinySLAM: A SLAM algorithm in less than 200 lines C-language program”. In: *11th International Conference on Control, Automation, Robotics and Vision, ICARCV 2010.* December. 2010, pp. 1975–1979.
- [49] Giorgio Grisetti, Cyrill Stachniss, and Wolfram Burgard. “Improved techniques for grid mapping with Rao-Blackwellized particle filters”. In: *IEEE Transactions on Robotics* 23.1 (2007), pp. 34–46.
- [50] *KartoSLAM - ROS Wiki*. URL: http://wiki.ros.org/slam%7B%5C_%7Dkarto?distro=groovy (visited on 10/28/2016).
- [51] *LagoSLAM - ROS Wiki*. URL: <https://github.com/rrg-polito/rrg-polito-ros-pkg> (visited on 10/28/2016).

- [52] Regis Vincent, Benson Limketkai, and Michael Eriksen. "Comparison of indoor robot localization techniques in the absence of GPS". In: *SPIE 7664, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XV*. Ed. by Russell S. Harmon, John H. Holloway, Jr., and J. Thomas Broach. International Society for Optics and Photonics, Apr. 2010, 76641Z.
- [53] Luca Carlone and Rosario Aragues. "A linear approximation for graph-based simultaneous localization and mapping". In: *Int. Conf. Robotics: Science and Systems*. 2011, p. 8.
- [54] João Machado Santos, David Portugal, and Rui P Rocha. "An Evaluation of 2D SLAM Techniques Available in Robot Operating System". In: *2013 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. 2013.
- [55] *Navigation - ROS Wiki*. URL: <http://wiki.ros.org/navigation> (visited on 10/18/2016).
- [56] *Map Server - ROS Wiki*. URL: http://wiki.ros.org/map%7B%5C_%7Dserver (visited on 11/04/2016).
- [57] *AMCL - ROS Wiki*. URL: <http://wiki.ros.org/amcl> (visited on 11/04/2016).
- [58] *Move Base - ROS Wiki*. URL: http://wiki.ros.org/move%7B%5C_%7Dbase (visited on 11/04/2016).
- [59] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.
- [60] Brian P Gerkey and Kurt Konolige. "Planning and Control in Unstructured Terrain". In: *ICRA Workshop on Path Planning on Costmaps*. 2008.
- [61] *Base Local Planner - ROS Wiki*. URL: http://wiki.ros.org/base%7B%5C_%7Dlocal%7B%5C_%7Dplanner (visited on 10/17/2016).
- [62] *Moveit Home*. URL: <http://moveit.ros.org/> (visited on 10/30/2016).
- [63] *RVIZ - ROS Wiki*. URL: <http://wiki.ros.org/rviz> (visited on 10/30/2016).
- [64] *Parameter Server - ROS Wiki*. URL: <http://wiki.ros.org/Parameter%20Server> (visited on 10/30/2016).
- [65] *SRDF - ROS Wiki*. URL: <http://wiki.ros.org/srdf> (visited on 10/30/2016).
- [66] *MoveIt Concepts - MoveIt Documentation*. URL: <http://moveit.ros.org/documentation/concepts/> (visited on 10/30/2016).
- [67] *Forward Kinematics - Wikipedia*. URL: https://en.wikipedia.org/wiki/Forward%7B%5C_%7Dkinematics (visited on 10/31/2016).
- [68] *Inverse Kinematics - Wikipedia*. URL: https://en.wikipedia.org/wiki/Inverse%7B%5C_%7Dkinematics (visited on 10/31/2016).
- [69] Patrick Goebel. *ROS By Example Volume 2: Packages and Programs for Advanced Robot Behaviors*. Lulu Enterprises, Inc., 2015.

- [70] Behnam Asadi. "Single and Dual Arm Manipulator Motion Planning Library". In: *Workshop on Task Planning for Intelligent Robots in Service and Manufacturing. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Workshop on Task Planning for Intelligent Robots in Service and Manufacturing, located at IROS 2015*. 2015.
- [71] Rosen Diankov and James Kuffner. *OpenRAVE : A Planning Architecture for Autonomous Robotics*. Tech. rep. July. Pittsburgh, PA: Robotics Institute, Carnegie Mellon University, 2008. DOI: CMU-RI-TR-08-34.
- [72] *IKFast Module - OpenRAVE Documentation*. URL: http://openrave.org/docs/latest%5C_%7Dstable/openravepy/ikfast/%7B%5C%7Dikfast-the-robot-kinematics-compiler, (visited on 10/31/2016).
- [73] *Orocos KDL - ROS Wiki*. URL: http://wiki.ros.org/orocos%7B%5C_%7Dkdl (visited on 10/31/2016).
- [74] *TRACLab - An Exploratory Research Lab Driven by a Curious, Pioneering Spirit*. URL: <https://traclabs.com/> (visited on 10/31/2016).
- [75] *Trac IK - ROS Wiki*. URL: http://wiki.ros.org/trac%7B%5C_%7Dik (visited on 10/31/2016).
- [76] Patrick Beeson and Barrett Ames. "TRAC-IK: An open-source library for improved solving of generic inverse kinematics". In: *IEEE-RAS International Conference on Humanoid Robots*. Vol. 2015-Decem. 2015, pp. 928–935. ISBN: 9781479968855. DOI: 10.1109/HUMANOIDS.2015.7363472.
- [77] Hassan Gomaa. *Real-Time Software Design for Embedded Systems*. Cambridge University Press, 2016. ISBN: 9781107041097.
- [78] Ian Millington and John Funge. *Artificial Intelligence for Games*. CRC Press, 2016. ISBN: 9781498785815.
- [79] *SMACH - ROS Wiki*. URL: <http://wiki.ros.org/smach> (visited on 10/31/2016).
- [80] Richard Szeliski. *Computer Vision: Algorithms and Applications*. 1st. New York, NY, USA: Springer-Verlag New York, Inc., 2010. ISBN: 1848829345, 9781848829343.
- [81] Ioannis Pitas. *Digital Image Processing Algorithms and Applications*. John Wiley & Sons, 2000. ISBN: 9780471377399.
- [82] *How Does a Digital Camera Work?* URL: <https://www.google.com/url?sa=i%7B%5C%7Drct=j%7B%5C%7Dq=%7B%5C%7Desrc=s%7B%5C%7Dsource=images%7B%5C%7Dcd=%7B%5C%7Dved=0ahUKEwjy3uuhwovQAhUE24MKHe6DCMIQjxwIAw%7B%5C%7Durl=http%7B%5C%7D3A%7B%5C%7D2F%7B%5C%7D2Fwww.noteey.com%7B%5C%7D2Fblogs%7B%5C%7D2Fadc-travels%7B%5C%7D3Fpage%7B%5C%7D3D2%7B%5C%7Dpsig=AFQjCNFBhPHyArfz8DKZ4GolmgBgRXlyTQ%7B%5C%7Dust=1478225767235938> (visited on 11/02/2016).

- [83] Karim Nice, Tracy Wilson, and Gerald Gurevich. *How Digital Cameras Work*. URL: <http://electronics.howstuffworks.com/cameras-photography/digital/digital-camera.htm> (visited on 11/03/2016).
- [84] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Second. Cambridge University Press, 2004. ISBN: 0521540518.
- [85] Corinna Jacobs and I Parrish. *Interactive Panoramas: Techniques for Digital Panoramic Photography*. Springer Science & Business Media, 2012. ISBN: 9783642186653.
- [86] Robert Schowengerdt. *Remote Sensing: Models and Methods for Image Processing*. Elsevier Science, 2012. ISBN: 9780080516103.
- [87] S.A. Quadri. *What is Spatial Resolution?* URL: <http://www.slideshare.net/reachquadri/what-is-spatial-resolution> (visited on 11/02/2016).
- [88] Evan Mclean Smith. “A Collection of Computer Vision Algorithms Capable of Detecting Linear Infrastructure for the Purpose of UAV Control A Collection of Computer Vision Algorithms Capable of Detecting Linear Infrastructure for the Purpose of UAV Control”. Masters of Science. Virginia Polytechnic Institute and State University, 2016.
- [89] David Fleet et al. *Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part 7*. Springer International Publishing, 2014. ISBN: 9783319105840.
- [90] *Key Differences Between Rolling Shutter and Frame (Global) Shutter - Point Grey Knowledge Base*. URL: <https://www.ptgrey.com/KB/10028> (visited on 11/02/2016).
- [91] *RGB Color Space - Wikipedia*. URL: https://en.wikipedia.org/wiki/RGB%7B%5C_%7Dcolor%7B%5C_%7Dspace (visited on 10/30/2016).
- [92] *RGB Color Model - Wikipedia*. URL: https://en.wikipedia.org/wiki/RGB%7B%5C_%7Dcolor%7B%5C_%7Dmodel (visited on 10/30/2016).
- [93] *HSL and HSV - Wikipedia*. URL: https://en.wikipedia.org/wiki/HSL%7B%5C_%7Dand%7B%5C_%7DHSV (visited on 10/30/2016).
- [94] Alvy Ray Smith. “Color gamut transform pairs”. In: *ACM SIGGRAPH Computer Graphics* 12.3 (1978), pp. 12–19. ISSN: 00978930. DOI: 10.1145/965139.807361.
- [95] Ramesh Jain, Rangachar Kasturi, and Brian Schunck. *Machine Vision*. McGraw-Hill, 1995. ISBN: 9780070320185.
- [96] Devi Parikh. *Linear Filters [PowerPoint Presentation] - ECE 5554: Computer Vision*. 2015. URL: https://filebox.ece.vt.edu/%7B~%7DF15ECE5554ECE4984/lectures/parikh%7B%5C_%7Dlecture2%7B%5C_%7Dfilters.pptx (visited on 08/27/2015).
- [97] *Binary Image - Wikipedia*. URL: https://en.wikipedia.org/wiki/Binary%7B%5C_%7Dimage (visited on 11/03/2016).

- [98] Chris Solomon and Toby Breckon. *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. John Wiley & Sons, 2011. ISBN: 9781119957003.
- [99] Joseph Wilson and Gerhard Ritter. *Handbook of Computer Vision Algorithms in Image Algebra*. Second. CRC Press, 2000. ISBN: 9781420042382.
- [100] John Canny. “A Computational Approach to Edge Detection”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8.6* (1986), pp. 679–698. ISSN: 01628828.
- [101] Mark Nixon. *Feature Extraction & Image Processing*. Academic Press, 2008. ISBN: 9780080556727.
- [102] Lee Middleton and Jayanthi Sivaswamy. *Hexagonal Image Processing: A Practical Approach*. Springer Science & Business Media, 2005. ISBN: 9781852339142.
- [103] Cheng Xing and Jianqiang Wang. “An Improved Method for Circle Detection”. In: Springer Berlin Heidelberg, 2008, pp. 463–468. DOI: 10.1007/978-3-540-88513-9_50. URL: http://link.springer.com/10.1007/978-3-540-88513-9%7B%5C_%7D50.
- [104] Richard Duda and Peter Hart. “Use of the Hough Transformation to Detect Lines and Curves in Pictures”. In: *Comm. ACM* 15.1 (1972), pp. 11–15.
- [105] D Ballard. “Generalizing the Hough Transform to Detect Arbitrary Shapes*”. In: *Pattern Recognition* 13.2 (1981), pp. 111–122.
- [106] Samarth Brahmbhatt. *Practical OpenCV*. Apress, 2013. ISBN: 9781430260806.
- [107] Wen-Yen Wu. “An Improved Method for Circle Detection”. In: *Proceedings of the International Computer Symposium (ICS)*. IOS Press, 2015. DOI: 10.3233/978-1-61499-484-8-1097.
- [108] Devi Parikh. *Fitting: Voting and the Hough Transform [PowerPoint Presentation] - ECE 5554: Computer Vision*. 2015. URL: https://filebox.ece.vt.edu/%7B~%7DF15ECE5554ECE4984/lectures/parikh%7B%5C_%7Dlecture8%7B%5C_%7Dhough.pptx (visited on 08/27/2015).
- [109] Jianbo Shi and Carlo Tomasi. “Good Features To Track”. In: *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference*. 1994, pp. 593–600. ISBN: 0-8186-5825-8. DOI: 10.1109/CVPR.1994.323794.
- [110] Chris Harris and Mike Stephens. “A Combined Corner and Edge Detector”. In: *4th Alvey Vision Conference*. 1988, pp. 147–151. DOI: 10.5244/C.2.23.
- [111] Prateek Joshi, David Escriva, and Vinicius Godoy. *OpenCV By Example*. Packt Publishing Ltd, 2016. ISBN: 9781785287077.
- [112] Robert Collins. *Harris Corner Detector [PowerPoint Presentation] - CSE486: Computer Vision*. 2012. URL: <http://www.cse.psu.edu/%7B~%7Drtc12/CSE486/lecture06.pdf> (visited on 12/14/2016).

- [113] Devi Parikh. *Local Features and Image Matching [PowerPoint Presentation] - ECE 5554: Computer Vision.* 2015. URL: https://filebox.ece.vt.edu/%7B~%7DF15ECE5554ECE4984/lectures/parikh%7B%5C_%7Dlecture12%7B%5C_%7Dlocalfeats.pptx (visited on 08/27/2015).
- [114] Paul Sastrasinh. *Dense Realtime Optical Flow on the GPU.* URL: <http://cs.brown.edu/courses/csci1290/2011/results/final/psastras/> (visited on 12/14/2016).
- [115] Michael Trummer, Joachim Denzler, and Christoph Munkelt. “Guided KLT Tracking Using Camera Parameters in Consideration of Uncertainty”. In: *Computer Vision and Computer Graphics - Theory and Applications*. Springer Berlin Heidelberg, 2009, pp. 252–261. DOI: 10.1007/978-3-642-10226-4_20. URL: http://link.springer.com/10.1007/978-3-642-10226-4%7B%5C_%7D20.
- [116] Simon Baker and Iain Matthews. “Lucas-Kanade 20 Years On: A Unifying Framework”. In: *International Journal of Computer Vision* 56.3 (2004), pp. 221–255.
- [117] Edwin Olson. “AprilTag: A robust and flexible visual fiducial system”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA).* 2011, pp. 3400–3407. URL: <http://april.eecs.umich.edu>.
- [118] Sanni Siltanen. “Theory and applications of marker-based augmented reality”. PhD thesis. 2012.
- [119] *ARToolKit Website.* URL: <http://www.hitl.washington.edu/artoolkit/> (visited on 10/29/2016).
- [120] *ALVAR Toolkit Website.* URL: <http://virtual.vtt.fi/virtual/proj2/multimedia/alvar/index.html> (visited on 10/29/2016).
- [121] *AR Track Alvar - ROS Wiki.* URL: http://wiki.ros.org/ar%7B%5C_%7Dtrack%7B%5C_%7Dalvar (visited on 10/29/2016).
- [122] Thomas Moore and Daniel Stouch. “A Generalized Extended Kalman Filter Implementation for the Robot Operating System”. In: *Proceedings of the 13th International Conference on Intelligent Autonomous Systems (IAS-13)*. Springer, 2014.
- [123] *Robot Localization - ROS Wiki.* URL: http://wiki.ros.org/robot%7B%5C_%7Dlocalization (visited on 11/06/2016).
- [124] *AR Track ALVAR - ROS Wiki.* URL: http://wiki.ros.org/ar%7B%5C_%7Dtrack%7B%5C_%7Dalvar (visited on 11/06/2016).
- [125] Itseez. *Open Source Computer Vision Library.* 2015. URL: <https://github.com/itseez/opencv>.
- [126] Thiemo Wiedemeyer. *IAI Kinect2.* \url{https://github.com/code-iai/iai__kinect2}. University Bremen.
- [127] Jean-Yves Bouguet. *Camera Calibration Toolbox For Matlab.* 2015. URL: https://www.vision.caltech.edu/bouguetj/calib%7B%5C_%7Ddoc/index.html (visited on 11/06/2016).

- [128] Zhengyou Zhang Zhengyou Zhang. “Flexible camera calibration by viewing a plane from unknown orientations”. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Vol. 1. c. 1999, pp. –7. ISBN: 0-7695-0164-8. DOI: 10.1109/ICCV.1999.791289.
- [129] Jorg Rowekamper et al. “On the position accuracy of mobile robot localization based on particle filters combined with scan matching”. In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, Oct. 2012, pp. 3158–3164. ISBN: 978-1-4673-1736-8. DOI: 10.1109/IROS.2012.6385988. URL: <http://ieeexplore.ieee.org/document/6385988/>.
- [130] *Robot Calibration - ROS Wiki*. URL: <http://wiki.ros.org/robot%20%7Dcalibration> (visited on 11/07/2016).
- [131] Mădălin Petru Sbanca and Gheorghe Leonte Mogan. “Advances in Robot Design and Intelligent Control”. In: *Advances in Intelligent Systems and Computing* 371 (2016), pp. 47–57. ISSN: 21945357. DOI: 10.1007/978-3-319-21290-6. URL: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84951320510%7B%5C%7DpartnerID=tZ0tx3y1>.
- [132] Gerard Canal Camprodon. “Human Multi-Robot Interaction based on Gesture Recognition”. Master of Science. Polytechnic University of Catalonia, University of Barcelona, and Rovira i Virgili University, 2015.
- [133] C Meriçli et al. “An Interactive Approach for Situated Task Teaching through Verbal Instructions”. In: *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. AAMAS ’14. International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1069–1076. ISBN: 9781634391313. URL: <http://www.aaai.org/ocs/index.php/WS/AAAIW13/paper/viewPDFInterstitial/7148/6534>.