PRECISE EXTERNAL POSITIONING OF MACHINE TOOLS USING ANGULAR MEASUREMENTS FROM DIGITAL CAMERAS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

İLKIN EGE OKAY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
MECHANICAL ENGINEERING

JUNE 2022

Approval of the thesis:

PRECISE EXTERNAL POSITIONING OF MACHINE TOOLS USING ANGULAR MEASUREMENTS FROM DIGITAL CAMERAS

submitted by **İLKIN EGE OKAY** in partial fulfillment of the requirements for the degree of **Master of Science in Mechanical Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. M.A. Sahir Arıkan
Head of Department, **Mechanical Engineering** _____

Assoc. Prof. Dr. A. Buğra Koku
Supervisor, **Mechanical Engineering, METU** _____

Assist. Prof. Dr. Murat Durmaz
Co-supervisor, **Geomatics Engineering, Hacettepe University** _____

Examining Committee Members:

Assoc. Prof. Dr. Kıvanç Azgın
Mechanical Engineering, METU _____

Assoc. Prof. Dr. A.Buğra Koku
Mechanical Engineering, METU _____

Assoc. Prof. Dr. Mehmet Bülent Özer
Mechanical Engineering, METU _____

Assist. Prof. Dr. Ali Emre Turgut
Mechanical Engineering, METU _____

Assist. Prof. Dr. Murat Durmaz
Geomatics Engineering, Hacettepe University _____

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: İlkin Ege Okay

Signature :

ABSTRACT

PRECISE EXTERNAL POSITIONING OF MACHINE TOOLS USING ANGULAR MEASUREMENTS FROM DIGITAL CAMERAS

Okay, İlkin Ege
M.S., Department of Mechanical Engineering
Supervisor: Assoc. Prof. Dr. A. Buğra Koku
Co-Supervisor: Assist. Prof. Dr. Murat Durmaz

June 2022, 75 pages

Nowadays, robot arms are highly available for industrial applications such as welding and automation. The major problem is the inaccuracy at tip position of the robot arm due to several structural effects. This thesis aims to increase the positioning accuracy of robot arm tip position by developing an external positioning method via angular measurement from digital cameras. Two gimbal systems are configured as the steering system of cameras. Each gimbal provides the angular position of the target in real-time. A calibration procedure is developed to establish a common reference frame. A python-based simulation is implemented for error budgeting. Real-world measurements are analyzed and discussed. According to the results, position accuracies better than 0.15% can be achieved. The results further can be improved by better vision components and online estimation of calibration parameters.

Keywords: Gimbal, Robotic Machining, Real Time Control Systems, Image Processing, Multiangulation

ÖZ

DİJİTAL KAMERALARDAN AÇISAL ÖLÇÜMLER KULLANILARAK TAKIM TEZGAHLARININ HASSAS KONUMLANDIRILMASI

Okay, İlkin Ege
Yüksek Lisans, Makina Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. A. Buğra Koku
Ortak Tez Yöneticisi: Dr. Öğr. Üyesi. Murat Durmaz

Haziran 2022 , 75 sayfa

Günümüzde robot kolları kaynak ve otomasyon gibi çeşitli endüstriyel uygulamalarda sıkça kullanılmaktadır. Uygulamalardaki ana sorun çeşitli yapısal etkiler sonucunda robot kolun uç noktasının konumlandırma hatasıdır. Bu tez, dijital kameralardan açısal ölçüm yoluyla harici bir konumlandırma yöntemi geliştirerek robot kol ucu konumunun konumlandırma doğruluğunu artırmayı amaçlamaktadır. İki gimbal sistemi, kameraların hareket sistemi olarak yapılandırılmıştır. Her bir gimbal, hedefin açısal konumunu gerçek zamanlı olarak ölçmektedir. Ortak referans ekseni oluşturmak için bir kalibrasyon metodu geliştirilmiştir. Hata bütçelemesi için python tabanlı bir simülasyon yazılmıştır. Gerçek dünya ölçümleri analiz edilmiş ve tartışılmıştır. Sonuçlara göre %0.15'den daha iyi konum doğrulukları elde edilebilmektedir. Sonuçlar, daha iyi görü bileşenleri ve kalibrasyon parametrelerinin çevrimiçi kestirimi ile daha da iyileşebilir.

Anahtar Kelimeler: Gimbal, Robotik işleme, Gerçek Zamanlı Kontrol sistemleri, Görüntü İşleme, Nirengi

To my family

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor, Assoc. Prof. Dr. Buğra Koku for his guidance, suggestions, advice, and encouragement throughout my study. Without his observation skills, I would not be on this path.

I am grateful to Assist. Prof. Dr. Murat Durmaz as an advisor, as a role model. His multidisciplinary guidance throughout my carrier has helped me to reach success.

I would like to thank my boss Onur Yarkınoğlu for his unlimited and priceless support. Also, Modesis Machine Technologies family for their help, understanding, motivation, and toleration.

Special thanks to my parents and my twin brother Arda Efe Okay for their huge support.

Also, I would like to thank Assist Prof. Dr. Lütfi Taner Tunç, from Sabanci University, Faculty of Engineering and Natural Sciences. The studies are founded from Tübitak by 1003 Primary Subjects R&D Funding Program with project number 217M078.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

2D	Two-Dimensional
3D	Three-Dimensional
AoA	Angle of Arrival
AoV	Angle of View
AZ	Azimuth
CAD	Computer Aided Drawing
CAM	Computer Aided Manufacturing
cm	centimeter
CNC	Computer Numerical Control
CS	Coordinate System
DoF	Degree of Freedom
EDM	Electronic Distance Meter
EL	Elevation
EtherCAT	Ethernet for Control Automation Technology
FoD	Field of Depth
FoV	Field of View
G	Gimbal Coordinate System
GUI	Graphical User Interface
HSV	Hue, Saturation, Value
IP	Industrial Protocol
LTS	Laser Tracking System
LSM	Least Square Method
m	meter
OpenCV	Open Source Computer Vision Library

RF	Radio Frequency (Identification)
RG	Roteted Gimbal Coordinate System
RGB	Red, Green, Blue
RMS	Root Mean Square
RSS	Receiver Signal Strength (Indicator)
SWIG	Simplified Wrapper and Interface Generator
TDoA	Time Difference of Arrival
ToA	Time of Arrival
W	World Coordinate System
YUV	Luminance, Bandwith, Chrominance

CHAPTER 1

INTRODUCTION

1.1 Motivation and Problem Definition

CNC machining is a manufacturing operation in which numerically controlled computer software manages the movement of cutting tools and machines. Plenty of complex machines can be controlled with this operation. Since the Industrial Revolution, CNC technologies have been continuously developed based on human needs.

The rapid increase in technology, ensured by Industry 4.0, caused an expeditious rise in global competition. Companies worldwide try to optimize their production costs while keeping an indistinguishable precision level. This situation directs companies to the practical usage of the factory with faster, cheaper, and more precise production. High-volume workpieces such as wind turbine blades and ship and airplane bodies were machined in large gantry machining centers. Due to precision requirements and physical constraints, it is impossible to machine huge parts in a smaller machining center. However, giant gantry machines have very high installation and operating costs, which cause inefficient usage of the factory because a large area is only dedicated to specific processes and specific sizes of the workpiece.

Karim et al. [1] mentioned the drilling processes of the windmill blades, which height reaches up to 50 meters. They claimed that the big milling machines could not machine all the holes around the root end in one step. The root end has to be rotated too many times to drill all the required holes. This rotation requirement decreases precision, causes a slower process time, and increases the production cost of the workpiece.

The usage of robots in the industry started in the 60s. Apart from being used for pick

and place purposes, industrial robots are also utilized in assembly, welding, painting, and machining applications [2]. Robotic machining research was initially introduced to take over human workers in the manufacturing area [3]. They described a variety of robot applications, including grinding, deburring, and drilling.

Robotic machining is the best alternative for traditional gantry machining because a robot arm can reach all the spots with an access amount of degree of freedom (DoF). Instead of 50 meter-square CNC machine, a few robot arms can be used simultaneously. Using robotic arms increases flexibility and accessibility, decreases cost, and overcomes system layout issues with various robot configurations.

Although robotic machining has several advantages over classical CNC machining, there are some disadvantages [4]. The CNC machining machines are structurally durable; on the other hand, machining robots have insufficient rigidity due to their cantilever beam structures. During high material removal rate operations, drilling, and milling, the tip point of the tool touch and transfers reaction forces from the workpiece to the robot. This phenomenon causes poor accuracy. The CNC machining technology has been improved over a few decades through machine control programs, and CAM simulation tools have developed year by year. Since robotic machining was first suggested in the 2000s and is not widely used, robotic machining simulation and programming tools are not advanced compared to the classical CNC tools. As a result of robotic machining still being a topic with much room for development and potential improvement, there are research topics associated with robotic machining operations.

One of the most important research topics associated with robotic machining operations is robot stiffness and chattering since they are vital to achieving precisely machined products. That is why the researchers are trying to decrease the maleffect of robotic machining with different solutions. On the other hand, some research focuses on calibration and compensation strategies of robotic machining by measuring the tip point of the robot.

Both the rigidity of the robot and the vibration of the machining process are critical factors in the stability of robotic machining. Low rigidity is a crucial disadvantage for high material removal rate processes, such as robotic milling and drilling. Absolute and relative stiffness are two types of robot stiffness. The former may be enhanced by

improving robot components and optimizing control parameters, whereas the latter is determined by the workpiece position and posture.

There is some research about vibration and chattering problems at robotic machining in literature. Some of them focus on modeling the vibration and chattering; on the other hand, some of the researchers [5], design a semi-active magnetorheological elastomers absorber to suppress the machining chatter.

There are two different compensation strategies, offline and online calibration. The former measures the robot's hand positions in different locations and changes the robot parameters before the operation starts. In contrast, the latter measures the end point's position during machining with external sensors and closes the control loop with a sensor feedback.

Good positioning repeatability is essential on the end points of the robotic arms. Even minor geometric inaccuracies in the end effector might result in substantial positioning errors. This leads to a decrease in the end point accuracy.

The tip point position of the robot can be simply computed from the forward kinematics, which can be obtained from the robots' CAD model. The lengths of the links and the angle between those links are sufficient to calculate a theoretical end point pose. Chen et al. [6] mentioned that due to the gravitational loading, gear backlash, structural deformation, wear, manufacturing, and assembly imperfections, the robot arm should be calibrated, and actual values should be estimated.

In the literature, several researchers have examined the errors of the robots [6, 7] and showed that the calibration of a robot is beneficial in terms of increasing the accuracy of the robots. Researchers also prove that the system's accuracy and repeatability are increased via calibration measurements.

Lehmann et al. [8] offered a compensation procedure in which the end point's displacement is calculated using the feed rate and control cycle period of the robot. The deflection error is estimated utilizing the variation between measured and computed force which is estimated by the force sensor attached to the tool center point. As a result of this error being fed into the controller, the system's performance improved.

Brunete et al. [9] used laser profilers for offline compensation. 3D points could be created by combining 2D laser profiles. The compensation increases the final point accuracy from a few millimeters to less than a millimeter. Qin et al. [10] utilized two tri-axis accelerometers to decrease static deflection effects. Some other researchers use a laser tracker as a sensor. Posada et al. [11] proposed a laser tracker-based offline compensation method. On the contrary, Kothe et al. [12] benefit from laser tracker in the closed-loop control system, which is an online compensation method. Moeller et al. [13] used two laser trackers to compensate for instantaneous error in the robot machining arm.

The online compensation technique developed in this thesis will enable industrial robots to perform required machining applications to manufacture high value-added products. Currently, the laser tracking systems, which are extremely expensive compared to the cost of the robotic systems, are used for position correction purposes. The position tracking and correction subsystem developed in this thesis will be a cost-effective alternative to such expensive laser trackers. In this regard, the thesis will be an innovative step to increase the effectiveness of industrial robots by increasing their tool path contouring accuracy.

1.2 Outline

The thesis is structured as follows: Introduction is done in Chapter 1, and the motivation of the thesis is explained. Chapter 2 provides the state of art literature survey about positioning algorithms, methods, and sensor alternatives. Also, steering alternatives of vision and laser-based systems are explained briefly. Finally, the image processing and the camera calibration are explained briefly. Chapter 3 explains the suggested vision-based positioning method in the thesis. Chapter 4 starts with implementation details of the studies. Then, both simulation results and error analysis are explained. The algorithm is discussed and the real system results are given. The overall study is concluded together with future ideas mentioned in Chapter 5.

CHAPTER 2

LITERATURE SURVEY

Positioning is the determination of the position of a body; thus, it can be considered as a subset of navigation. In literature, positioning and localization are used interchangeably as synonyms, especially in short-range applications, where the range of sensors is generally less than three kilometers. According to Oxford Dictionary [14], localization is the act or process of finding out exactly where something is. The aforementioned definitions show that the position is expressed quantitatively, whereas the localization is qualitative.

Positioning can be classified into several categories with different perspectives [15]. In the first category, positioning can be divided into two subcategories based on position calculation instant as real-time and post-processed. The second category helps to distinguish whether the object is fixed or static. The last distinctive characteristic is how collecting the measurements and positioning calculations are done. If the position is calculated at the object, this method is called self-positioning. Otherwise, if the position is calculated elsewhere, called remote positioning.

Having given a brief introduction to positioning in the previous paragraphs, Section 2.1 introduces the Least Squares Method, which is frequently used in position estimation. In Section 2.2, positioning approaches and methods are explained. Section 2.3 accounts for indoor positioning sensors, short-range positioning sensors, and alternative position measurement and metrology methods. In Section 2.4, examples of different steering methods for laser and vision-based positioning alternatives are described. Then, the topics related to image processing are explained in Sections 2.5 and 2.6 . In Section 2.5, the necessity of camera calibration, pinhole camera geometry, and calibration steps are explained. Then, in Section 2.6, video processing subjects,

object detection, object classification, and object tracking are introduced.

2.1 The Least Squares Method (LSM)

The Least Squares Method is a mathematical optimization technique that is widely employed to estimate the unknowns in overdetermined systems in which the number of equations is greater than the number of unknowns. In LSM, the sum of the squares of the residuals is minimized in order to carry out optimization.

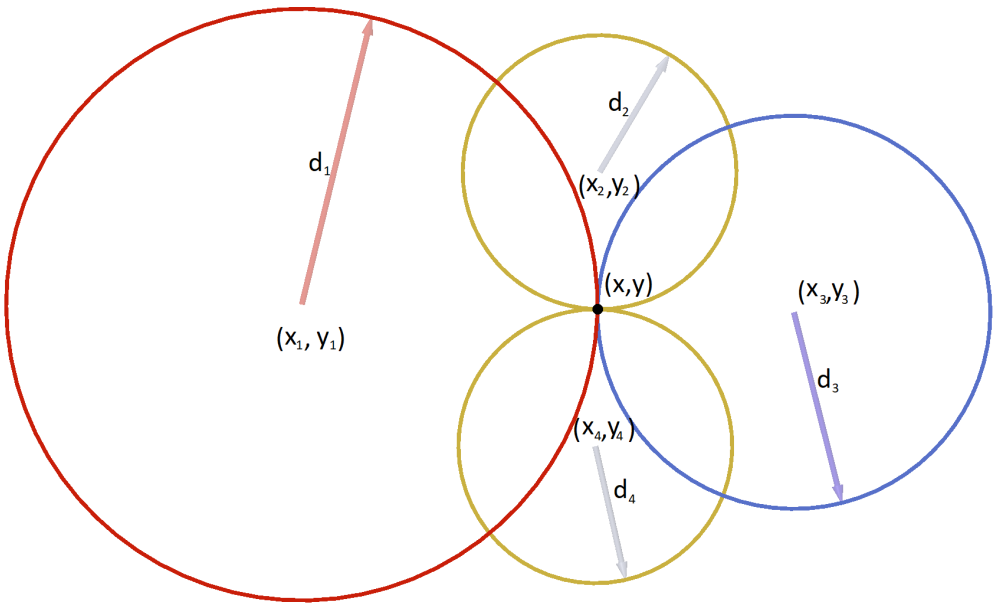

Figure 2.1: 2D Ideal Positioning with Four Measurement Nodes

Assume a planar problem with an unknown location (x, y) and n different measurement nodes with coordinates $(x_1, y_1), ..., (x_n, y_n)$. The distance between each node and the unknown location is measured as d_i. Ideally, the node should be at an intersection of n circle with d_i radius where $i = 1, 2, .., n$. The equation set is given in Eq. 2.1.

$$\begin{aligned}
(x_1 - x)^2 + (y_1 - y)^2 &= d_1^2 \\
(x_2 - x)^2 + (y_2 - y)^2 &= d_2^2 \\
\vdots \qquad\qquad & \vdots \qquad\qquad \vdots \\
(x_n - x)^2 + (y_n - y)^2 &= d_n^2
\end{aligned} \tag{2.1}$$

As mentioned above, the number of equations is greater than the number of variables; that is why the exact solution cannot be obtained. Therefore, the LSM can be employed to find out the estimated solution. The first step of the LSM is linearization since the equations are nonlinear with respect to x and y. The linearization is performed by subtracting the nth equation from each equation as given in Eq. 2.2.

$$\begin{aligned}
x_1^2 - 2x_1x - x_n^2 + 2x_nx + y_1^2 - 2y_1y - y_n^2 + 2y_ny &= d_1^2 - d_n^2 \\
x_2^2 - 2x_2x - x_n^2 + 2x_nx + y_2^2 - 2y_2y - y_n^2 + 2y_ny &= d_2^2 - d_n^2 \\
\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots & \\
x_{n-1}^2 - 2x_{n-1}x - x_n^2 + 2x_nx + y_{n-1}^2 - 2y_{n-1}y - y_n^2 + 2y_ny &= d_{n-1}^2 - d_n^2
\end{aligned} \tag{2.2}$$

As seen below in Eq.s 2.3, 2.4 and 2.5, the set of Eq.s 2.2, are converted into $A\gamma = B$ form by arranging the known and unknown terms to achieve a solution.

$$\mathbf{A} = \begin{bmatrix} 2(x_1 - x_n) & 2(y_1 - y_n) \\ 2(x_2 - x_n) & 2(y_2 - y_n) \\ \vdots & \vdots \\ 2(x_{n-1} - x_n) & 2(y_{n-1} - y_n) \end{bmatrix} \tag{2.3}$$

$$\mathbf{B} = \begin{bmatrix} x_1^2 - x_n^2 + y_1^2 - y_n^2 + d_n^2 - d_1^2 \\ x_2^2 - x_n^2 + y_2^2 - y_n^2 + d_n^2 - d_2^2 \\ \vdots \\ x_{n-1}^2 - x_n^2 + y_{n-1}^2 - y_n^2 + d_n^2 - d_{n-1}^2 \end{bmatrix} \tag{2.4}$$

$$\boldsymbol{\gamma} = \begin{bmatrix} x \\ y \end{bmatrix} \tag{2.5}$$

It is known that there is no unique solution; however the aim is to minimize the residuals.

$$\arg\min_{\gamma} (\mathbf{B} - \mathbf{A}\boldsymbol{\gamma}) \tag{2.6}$$

$$L(\gamma) = \left\| \mathbf{B} - \mathbf{A}\boldsymbol{\gamma} \right\|^2 = (\mathbf{B} - \mathbf{A}\boldsymbol{\gamma})^T (\mathbf{B} - \mathbf{A}\boldsymbol{\gamma}) \tag{2.7}$$

$$= \mathbf{B}^T\mathbf{B} - \mathbf{B}^T\mathbf{A}\boldsymbol{\gamma} - \boldsymbol{\gamma}^T\mathbf{A}^T\mathbf{B} + \boldsymbol{\gamma}^T\mathbf{A}^T\mathbf{A}\boldsymbol{\gamma} \tag{2.8}$$

The slope is the gradient of the function for a graph or function.

$$\frac{\partial L(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \frac{\partial(\mathbf{B}^T\mathbf{B} - \mathbf{B}^T\mathbf{A}\boldsymbol{\gamma} - \boldsymbol{\gamma}^T\mathbf{A}^T\mathbf{B} + \boldsymbol{\gamma}^T\mathbf{A}^T\mathbf{A}\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \tag{2.9}$$

$$= -2\mathbf{A}^T\mathbf{B} + 2\mathbf{A}^T\mathbf{A}\boldsymbol{\gamma} \tag{2.10}$$

Finally, the gradient of the function is set to zero to find the local maxima point. Also, note that the change of the slope is zero at the local maxima point.

$$-2\mathbf{A}^T\mathbf{B} + 2\mathbf{A}^T\mathbf{A}\boldsymbol{\gamma} = 0 \rightarrow \mathbf{A}^T\mathbf{B} = \mathbf{A}^T\mathbf{A}\boldsymbol{\gamma} \tag{2.11}$$

$$\boldsymbol{\gamma} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{B} \tag{2.12}$$

To find the unknown coordinates, the matrix equation is solved as represented in Eq. 2.12. The ordinary least squares method is explained between Eqs. 2.6 and 2.12. There are also weighted least squares and generalized least squares methods in the literature where both methods are used for localization problems [16, 17].

2.2 Positioning Approaches

All positioning approaches utilize two fundamental methods (1) position fixing, using external measurements to estimate position, and (2) dead-reckoning, which combines the initial position with instant velocity and direction.

This section explains (1) multilateration, and (2) multiangulation which are the standalone positioning fixing techniques, (3) dead-reckoning, and (4) hybrid positioning methods in which more than one standalone positioning method is employed.

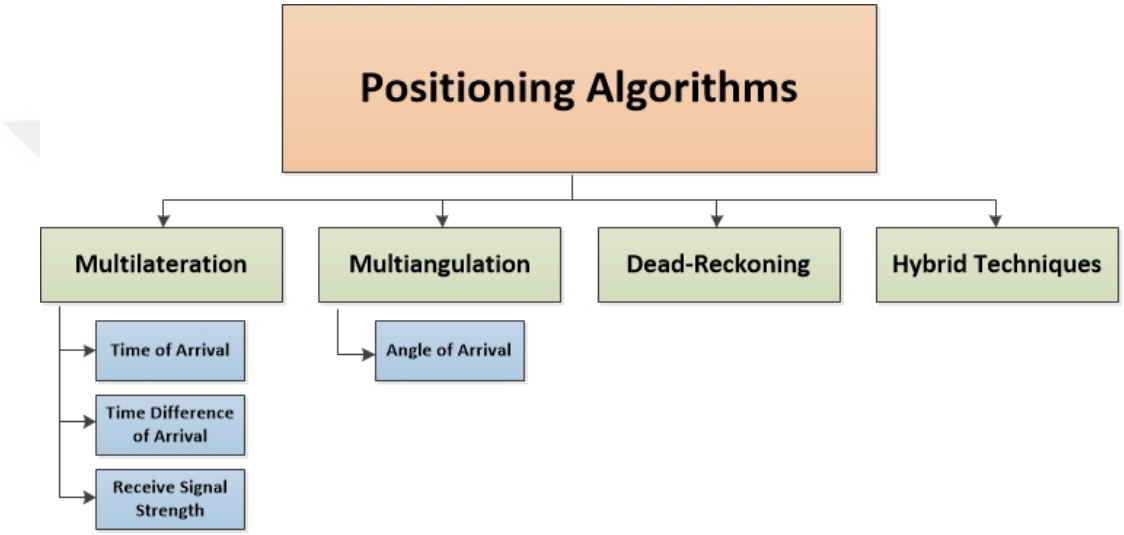

Figure 2.2: Positioning Methods and Algorithms

2.2.1 Multilateration

Multilateration is a technique for obtaining a target's position in space based on the distance or range, which can be defined as the distance between the object and the measurement device. In $2D$ space, all measurements denote a circle whose radius is identical to the measured distance. The peripheral of the circle indicates where the target is, whereas the center demonstrates the known position of measurement stations. Two measurement stations describe two different circles where the circles intersect at two points, creating ambiguity. At least three different measurement stations are required to find the target position.

In $3D$, one measurement station denotes a sphere, two measurements intersect circular regions, and three measurement stations intersect at two different points. The fourth range of information is needed to obtain a unique solution. On the contrary, if the rough position of the target is known, the ambiguity with three different measurement stations can be solved, and the fourth or more measurement stations are utilized to decrease measurement uncertainty.

Multilateration is commonly used in navigation and surveillance applications. Distance-based sensors and some laser tracker applications also use the multilateration method [18].

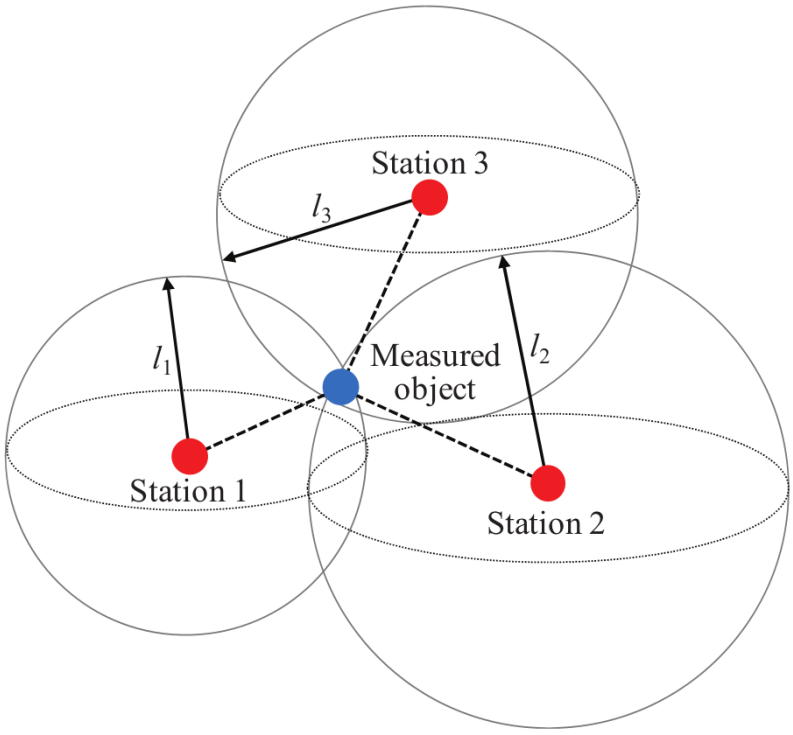

Figure 2.3: 3D Multilateration with Three Measurement Nodes [19]

$3D$ multilateration is demonstrated in Figure 2.3. The target $\mathbf{T}(x_t, y_t, z_t)$ is measured from three different measurement stations, $\mathbf{P}_1(x_1, y_1, z_1)$, $\mathbf{P}_2(x_2, y_2, z_2)$ and $\mathbf{P}_3(x_3, y_3, z_3)$. For each station, three equations can be written to express the relation between the point and the target. There are several methods to solve those equations. Norrdine [20] suggested the weighted least squares method by arranging equations.

Sertatil et al. [21] employed another method in which there are four different mea-

surements. In the method, the target position is calculated using three of four measurements with four different combinations. The average of all measurements is used directly with equal weights, given as in Eq. 2.13.

$$P(x_t, y_t, z_t) = P(\frac{1}{4}\sum_{i=1}^{4} x_i, \sum_{i=1}^{4} y_i, \sum_{i=1}^{4} z_i) \tag{2.13}$$

Takatsuji et al. [22] mentioned that the initial distance of each measurement station is denoted as $l_i, (i = 1, .., 4)$, and measurement change of i_{th} station at the instant j is expressed as m_{ij} . The interval between the i_{th} node and j_{th} node is described as c_{ij} where $i \neq j$.

$$R_j = \left(\sum_{i=1}^{4}[c_{ij} - (l_i + m_{ij})]^2\right)^{\frac{1}{2}} \tag{2.14}$$

The main goal here is to keep the R_{res} variable expressed as sums of squares of residuals to a minimum.

$$R_{res} = \sum_{j=1}^{4} R_j^2 \tag{2.15}$$

There are several different techniques to measure distance information. Time of Arrival, Time Difference of Arrival, and Received Signal Strength are explained in the upcoming subsections.

2.2.1.1 Time of Arrival (ToA)

Time of Arrival (ToA) is introduced as the time that any signal arrives at a receiver from a signal source. Multiplying the transition time by the speed of the light provides the distance. The formulation is given in Eq. 2.16 where R is the range between the signal source and signal sensor, t_t is the timestamp when the signal is generated from the transmitter, t_r is the timestamp when the signal reaches the receiver and c is speed of the light or speed of air [23].

$$R = (t_t - t_r) \times c \tag{2.16}$$

The crucial defect of the ToA approach is that calculating the air travel time requires synchronization between transmitters and receivers. To overcome this disadvantage, another server can be deployed for time-delay measurement, which raises the system's cost. Moreover, the variations in propagation speed, especially in acoustic systems, may create an error. To measure the distance information, the ToA system utilizes the multilateration approach.

2.2.1.2 Time Difference of Arrival (TDoA)

In this method, given the position of several pairs of reference points the time difference between them is estimated by utilizing the corresponding time data at each receiving node rather than the absolute one. To indicate the time and identify the position, synchronized time source of transmission is not required. TDoA entails several receiving nodes receiving a signal with an uncertain start time, with just the receivers requiring time synchronization [24]. In the localization space where the mobile node's position is placed, variations in arrival time measurement form hyperbolic trajectories. The intersection of many hyperbolic trajectories demonstrates the feasible target positions.

$$\frac{R}{c_1} - \frac{R}{c_2} = t_1 - t_2 \tag{2.17}$$

In Eq. 2.17, c_1 and c_2 represent the speed of two different type of signals, t_1 and t_2 indicate the time for those signals to travel from one measurement station to the another one, and R illustrates the range between the transmitter and the receiver.

2.2.1.3 Received Signal Strength (RSS) Based

RSS refers to the voltage detected by the receiver's received signal strength indicator (RSSI) circuit. RSS is calculated using measured power, which is inversely propor-

tional to the square of the distance between transmitter and receiver. RSS is capable of measuring acoustic, radio frequency (RF), and any other signals [25].

The offline training and runtime localization phases are the two steps of RSS-based lateration algorithms. RSS examples are gathered at several well-known positions using various measurement nodes during the offline training phase, and distances between the known locations and measurement nodes are determined. The signal propagation model is then matched to the obtained RSS measurements and distances, and the signal distance relation is computed as represented in Eq. 2.18.

$$P(d) = P(d_0) - 10\gamma \log 10(\frac{d}{d_0}) + X_\sigma \tag{2.18}$$

Where $P(d_0)$ denotes a wireless device's transmitting power at the initial distance d_0, d indicates the range between the wireless device and the measurement node, γ symbolizes the path loss exponent, and X_σ represents shadow fading, which follows a zero-mean Gaussian distribution with σ standard deviation [26].

2.2.2 Multiangulation

The location of a point is estimated by creating multiple triangles from known measurement points. Utilizing angular data is sufficient for this method; in other words, additional information, such as length, is not required. The main application areas include surveying and computer vision [27].

The summary of the multiangulation is shown in Figure 2.4. There are two different angles expressed as λ and θ. The azimuth angle, λ, is the angle of projection of the coordinate on the horizontal plane with respect to x-axis. The other angle θ, which represents elevation, is the angle between the line of sight of the object and the horizontal plane. In 3-D coordinate system, the target position is shown as **H**(x,y,z). The position of n different angular measurement devices are given in a common coordinate system as $\mathbf{G}_1(x_1, y_1, z_1)$, $\mathbf{G}_2(x_2, y_2, z_2)$, ... , $\mathbf{G}_n(x_n, y_n, z_n)$. The angular coordinates of the target can be measured as (λ_1, θ_1) , (λ_2, θ_2) , ... , (λ_i, θ_i) respectively where $i = 1, 2, ..., n$. The following equations can be written between the n

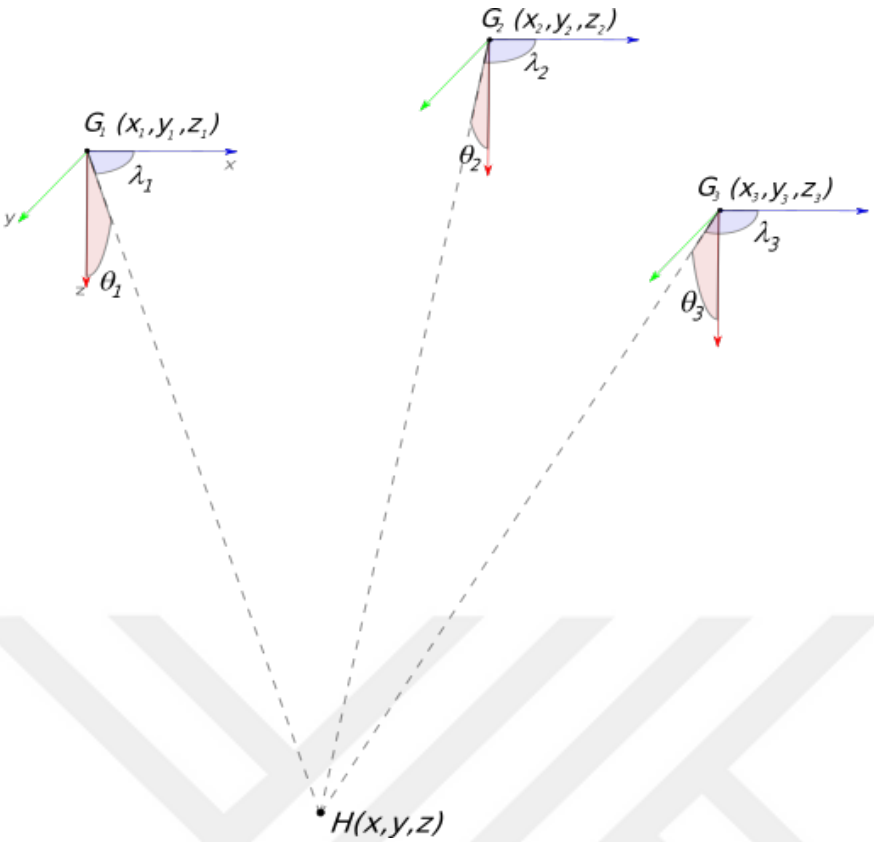

Figure 2.4: Geometry of Multiangulation

angular measurement device and the target position.

$$\theta_i = cos^{-1}\left(\frac{z - z_i}{\sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}}\right), i = 1, 2, ..., n \tag{2.19}$$

$$\lambda_i = tan^{-1}\left(\frac{y - y_i}{x - x_i}\right), i = 1, 2, ..., n \tag{2.20}$$

The nonlinear system of equations created by these measurements can be estimated by the nonlinear LSM. The only method to measure the angle between the measurement station and the target is the Angle of Arrival. Camera-based vision systems, laser scanners, and image sensors use the multiangulation method.

2.2.2.1 Angle of Arrival (AoA)

The angle at which the signal comes from the observer to the target is calculated using Angle of Arrival. Each observer has different lines of sight, where each line creates a line of position. The target position is an intersection of n different lines. Even though there are only two observers, there is no ambiguity since two lines intersect at one point in 3-D coordinates. If slight errors in the AoA calculation occur, the position estimation error might be significant. Each angular measurement error forms a cone around the centerline of sight. As the range increases, the position uncertainty also increases.

2.2.3 Dead-Reckoning

Dead-reckoning, or deduced-reckoning, is a navigation term which determines the present position of a moving object whose initial speed and direction are known. The estimation is done simply by Eq. 2.21, where R is the final position vector, and r is the initial position vector, v is the speed, and t is time.

$$R = r + vt \tag{2.21}$$

Dead-reckoning has been used throughout the history as a navigation method thanks to its reliability and simplicity [28]. In automated systems such as mobile phones, the target is equipped with motion sensors like accelerometers, gyroscopes, and magnetometers. During movement, location can be calculated by moving directions, velocity, and sensor sampling rate.

Dead-reckoning is generally used for pedestrian navigation applications with low-cost sensors such as mobile phone inherit sensors [29], and outdoor localization with inertial measurement units [30]. The positioning accuracy of dead-reckoning systems is up to a few centimeters.

2.2.4 Hybrid Positioning Techniques

Hybrid Positioning Techniques utilize the combination of two or more independent measurement methods. The researchers are trying to find out the optimum combi-

nation of methods (1) to increase the precision and accuracy and (2) to decrease the cost and computational time. In Figure 2.5, Yassine et al. [31] suggested an imaginary combination of ToA, AoA, and RSS-based methods. Venkatraman et al.[32] proposed a TOA and AOA-based wireless positioning system. Li et al.[33] offered a TDOA and AOA-based hybrid method.

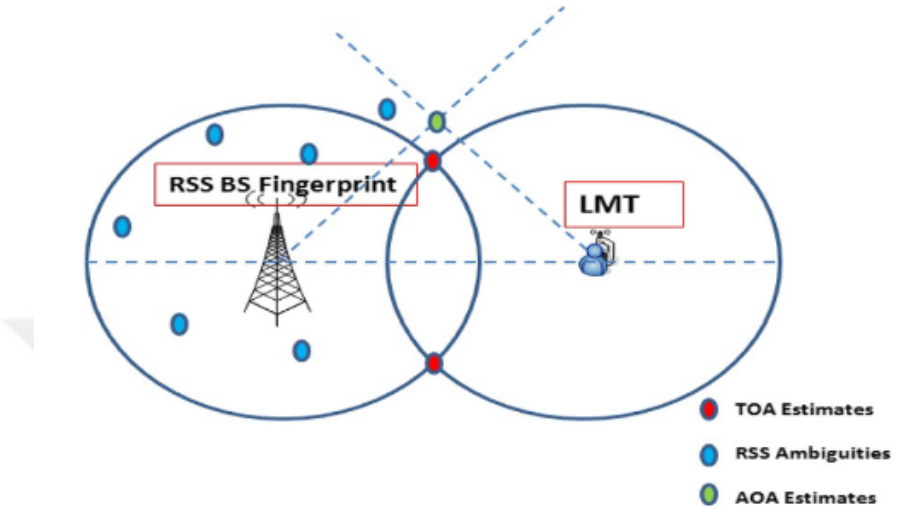

Figure 2.5: Example of Hybrid Technique [31]

The most common hybrid method is mono sight systems, which is explained briefly are the next subsection.

2.2.4.1 Mono Sighting Systems

The target's position is expressed in spherical coordinates R, θ, ϕ. The distance between the source and the target measured by distance-meter with the ToA method is R, the horizontal axis is x, and the vertical axis is z. The angles between the source and the target measured by the AoA method are θ in the x-axis and ϕ in the z-axis. The transformation from spherical coordinates to Cartesian coordinates can be obtained as represented in Eqs. 2.22, 2.23, and 2.24. Commercially available laser tracker systems and total stations use this principle.

$$\Delta x = R\sin(\theta)\cos(\phi) \tag{2.22}$$

$$\Delta y = R\sin(\theta)\sin(\phi) \tag{2.23}$$

$$\Delta z = R\cos(\theta) \tag{2.24}$$

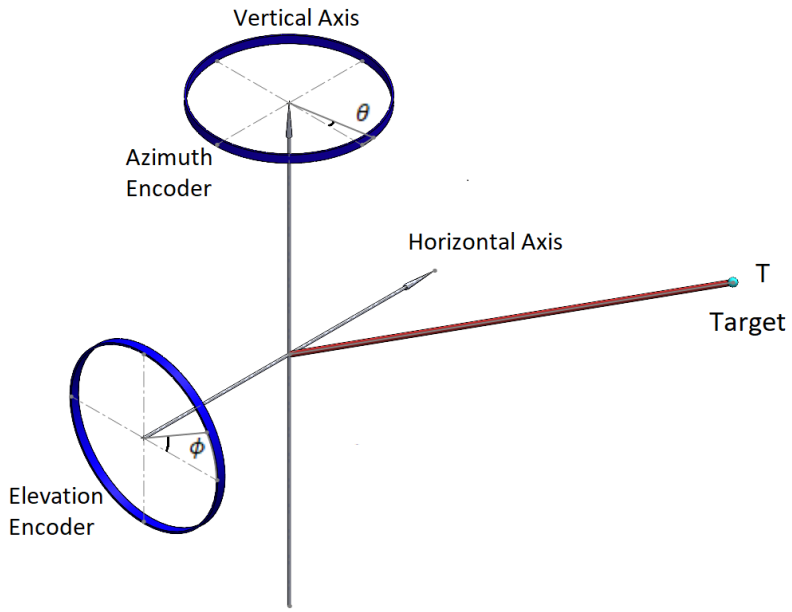

Figure 2.6: Mono-Sight System Geometry

2.3 Positioning Applications

There are several methods to measure the end point of the robot. Indoor and short-range positioning sensor alternatives such as WLAN-based, Ultrasonic-based, RF-based, Bluetooth-based, and metrological measurement devices such as Theodolite, Total Station, and Laser Tracker System are explained in the following subsections.

In the studies conducted by Maisano et al. [34], the system topology of the measurement systems is divided into two parts Distributed and Centralised. The Centralised ones can be used individually, and the geometry of the object can be measured as an independent module. To increase the accuracy, several of them can be used simultaneously. On the contrary, the distributed systems need to work co-operatively

to obtain the coordinates of the object; the measurement can be obtained by combining different sensor data from other measurement modules. Generally, standalone measurement methods are distributed systems. On the other hand, the systems which use hybrid methods and dead-reckoning are centralized systems. Laser Tracker and Theodolite are examples of centralized systems.

2.3.1 Wireless Local Area Network(WLAN) Based

WLAN is a group of collaborated computers or other devices that form a network based on high-frequency radio waves rather than wired connections. Indoor positioning systems based on WLAN use several WLAN access points to obtain the target position in an indoor space. WLAN runs in the 2.4 or 5 GHz frequency bands and is based on IEEE 802.11 standards [35]. Extra access points often needed to be added to the existing network to improve accuracy.

Evennou et al. [36] used dead reckoning and an RSS-based system for self-localization. Yang et al. [37] utilized ToA and AoA-based systems. Both researchers emphasize that the cost of the WLAN system is too low because most locations already have a WLAN system. The systems mentioned above, which measure this WLAN-based principle, have reached an accuracy of around 1 m.

2.3.2 Ultrasonic Based

Sound frequencies exceeding the highest audible limit of human hearing, which is roughly 40 kHz, are used in ultrasonic positioning systems [38]. By measuring the time it takes for a signal to travel from a transmitter to its receivers, the target's position may be determined. Low system cost, reliability, scalability, and excellent energy efficiency are the advantages of ultrasonic locating systems over other indoor position systems.

In their studies, Gonzalez et al. [39] offer an indoor broadband ultrasonic system that utilizes beacons to estimate the pose of the mobile device in 3D. They used a system based on ToA and AoA. In 95% of cases, the suggested technique had an accuracy

of less than 1cm. Sertatıl et al. [21] used acoustic signals by the ToA measurement technique with the Triangulation location estimation strategy. The method provides an estimation with a precision of 99 %, which is better than 2 cm.

2.3.3 Radio Frequency Based

Radio Frequency Identification (RFID), also known as Radio Frequency Proximity Sensors, is a communication system composed of two stations: a reader and a tag. During communication, radio waves are used to store and retrieve information between stations. RFID readers are responsible for perceiving the information distributed from the tag. Compared to other techniques, RFID is a cheap and flexible technology.

Randell et al. [40] suggested a low-cost RF and ultrasonics combined system which employs the ToA method. Wang et al. [41] used RFID sensors with RSS and AoA methods. Ni et al. [42] suggested an RSS-based approach by employing the concept of reference tags. The aforementioned studies show that the accuracy of the RF-based systems is in order of cm.

2.3.4 Bluetooth Based

Bluetooth is a short-range wireless technology norm for transferring data across short areas. 2.4 GHz ISM spectrum band, which allows an optimum balance between range and efficiency, is used by Bluetooth technology [43]. Depending on the receiver sensitivity and the usage environment, the Bluetooth range can be up to 100 meters. Thanks to the low price, Bluetooth is available on most mobile phones and smart devices.

The positioning approach recommended by Zhou et al. [44] is depended on the RSSI method. The RSSI method is used to determine the position and to collect the Bluetooth access point signal strength. RSSI data is collected via mobile phones. The data is processed using the Kalman filter and other statistical methods, and the positioning accuracy of the system is up to 1 meter.

By installing the Bluetooth gateway, Li et al. [45] presented an indoor location technique for dynamic RSSI correction. The Kalman filter and the least-squares estimation are implemented to determine the position in the related research. The final positioning accuracy of this work is in order of meters.

2.3.5 Theodolite

An optical device, employed to compute azimuth and elevation angle between the object of interest and the measurement device, is named a theodolite. Typically, theodolites are used in land surveying, building, infrastructure construction, and geometric applications.

The overall accuracy of the system depends on the resolution, repeatability, and misalignment between two axes. Although, from an angular point of view, the accuracy of the theodolites can be decreased to the resolution of angular encoders, the depth of the object cannot be measured because there is no sensor related to distance measurement. On the contrary, if the exact length of the object is known, the distance between theodolite and the object can be estimated.

2.3.6 Stereo Vision

At least two digital cameras, each with a distinct perspective of the same scene, are used in stereo vision. Visual features seem to be offset in their relative locations in several perspectives or images. In stereo vision, the vision cameras are fixed to different locations, and their offset between each other is known or estimate via calibration methods. During the measurements, there is no moving part in the system. The triangulation method is used to evaluate the position of the target. However, matching these images to a highly accurate point cloud is the challenging part of stereo vision. In the literature, there are several applications and research about this subject. This method is used in diagnostics, quality control, manufacturing, and tracking [46]. The precision of the suggested systems in studies [47, 48, 49] are in the order of millimeters.

2.3.7 Total Station

The total station is simply an addition of EDM (Electronic Distance Meter) to the theodolite. The EDM measures the optical path using an infrared-based pulse laser between the target and the instrument. The total stations are commonly used for land surveying and construction building in geomatics and civil engineering. With recent development in the area, the accuracy of EDM's increased significantly. Nowadays, commercially available Total Stations have similar accuracy to LTS systems, explained in Subsection 2.3.8. In recent studies, Paraforos et al. [50] used Total Station to measure the robot arm. The overall system performance is up to 4 mm. Luo et al. [51] showed that the accuracy of the total station is around 2 mm when the range is around 5 meters.

2.3.8 Laser Tracking System (LTS)

The target position is measured in 3D using a laser beam and a retro-reflective by the laser tracking system. The LTS measures two angles with two different encoders and the distance to the retro-reflector connected to the object of interest with a laser interferometer. The laser beam emitted by an interferometer for displacement measurement is directed by a laser tracker using a beam steering mechanism to a retro-reflector hooked to the advancing target. The target's position is calculated in 3D using two angular encoders where each encoder responds to one rotation axis, AZ and EL. Although different laser trackers may have various components, laser tracker systems generally consist of the following subsystems. Those subsystems work together not only to track the object but also to measure the object's coordinates [52].

- Laser beam source,
- Beam-steering mechanism,
- The distance measurement device,
- Beam splitter,
- Retro-reflector attached on the object of interest,

- Precise sensor for measuring the offset of the mirrored laser beam.

2.4 Beam Steering Mechanisms - Vision System Steering Mechanisms

It is known that the tracking system is the primary source of uncertainty in the laser tracker system [53]. Similarly, for other measurement devices with motion, the positioning system is the primary error source.

To adjust the orientation of the laser beam, the tracking system employs a mirror or similar equipment. To follow the retroreflector using the laser beam, the tracking mirror is capable of rotating in two axes. On the contrary, in vision systems, the camera is generally static [54].

2.4.1 Gimbal

As defined in Oxford Dictionary [55], gimbal is a device for keeping an instrument such as a compass or chronometer horizontal in a moving vessel or aircraft, typically consisting of rings pivoted at right angles.

In the literature, there are several applications in which the vision system is excited by the gimbal. Those studies mainly focus on the theory and the stability [56, 57]. It is also observed that the camera is arranged at the rotation center of the gimbals in those applications.

Gimbal is used in inertial guidance systems, measuring systems, photography, and filming. Commercially available Laser Tracker Systems, Theodolites, and Total Stations are used gimbals for moving and scanning purposes.

2.4.2 Galvo-Head

A galvo-head, also named as marking head, is an electromechanical system that uses rotatable mirrors to change the vision system's direction according to the user's preference. Galvo-head is composed of two mirrors, which are connected to two differ-

ent rotors. Mirrors are placed close to each other on orthogonal axes, and each one of them controls one axis. When mirrors rotate, the direction of the beam changes accordingly. Using electric motors is a common way to rotate galvo scanners. Galvo-heads can reach extremely high speed and acceleration compared to the other steering methods. The drawback of the Galvo-head is the small angular range.

Galvo-heads are mainly used for controlled laser motion. They are widely used in machining, polishing, and marking applications with a high-power laser source. The laser beam can be steered across the workpiece at exceptionally high speeds with excessive precision and repeatability [58]. Okumura et al. [59] utilized galvo-head in vision system. Instead of moving the camera, the researchers suggest a method in which the galvo-head mirrors are rotated to track the object. The study is mainly focused on the response time rather than the accuracy; that is why no accuracy value is presented.

2.4.3 Hemisphere Method

In laser-based tracking systems, an articulating hemisphere is an option for the steering mechanism for the laser beam. This method is made up of a single mirror attached to a hemisphere that is connected to a moving plane. Three little balls support the hemisphere. There is a mirror on the flat side, and a stick is connected to the round side of the hemisphere. The other side of the stick is guided by as X-Y stage; thus, the mirror can turn around two axes [60, 53].

On the other hand, this steering mechanism needs the mirror and the laser beam to intersect at the center of the sphere. This condition cannot be guaranteed due to manufacturing and assembly errors and imperfections. Therefore, a slight shift from the center, which causes measurement uncertainty, may occur.

2.5 Camera Calibration

Nowadays, the camera is one of the most common sensors for industrial applications and household products. To use the camera in applications where precision is essen-

tial, the mathematical relation between the real 3D world and the 2D camera sensor plane should be obtained. That relation is called extrinsic parameters. Moreover, due to the manufacturing imperfections and assembly misalignments of optical lenses, there are some distortions in the image. All of those parameters are called intrinsic parameters.

The camera can be represented by different mathematical models. The most common, basic, and useful representation is the pinhole camera model [61]. Other basic camera models are fisheye [62], weak perspective, and orthographic (or affine) projection models.

2.5.1 Camera Matrix

The pinhole camera model assumes a small opening that allows a single ray to move from the real world to the image sensor. For any 3D point in the world, this point can be modeled as a ray that passes through a camera lens and projects to the 2D image sensor. Figure 2.7 shows the projection where f is the focal length of the camera and Z is the distance from the object to the camera. Eq. 2.25 gives the relation between real-world point $\mathbf{Q}(X, Y, Z)$ and image point $\mathbf{q}(u, v)$ where f_x and f_y are focal length in x and y directions and c_x and c_y are optical center coordinates in x and y directions.

$$\begin{aligned} u &= f_x \frac{X}{Z} + c_x \\ v &= f_y \frac{Y}{Z} + c_y \end{aligned} \tag{2.25}$$

A projective transform is a relation that scales a point $\mathbf{Q}(X, Y, Z)$, which is located in the real world to point $\mathbf{q}(u, v)$, which is on the image screen. A $n + 1$ dimensional vector is used to express the homogeneous coordinates associated with a point in a projective space of dimension n. This helps to create a mathematical relation between $\mathbf{Q}(x, y, z)$ and $\mathbf{q}(u, v, w)$ [63].

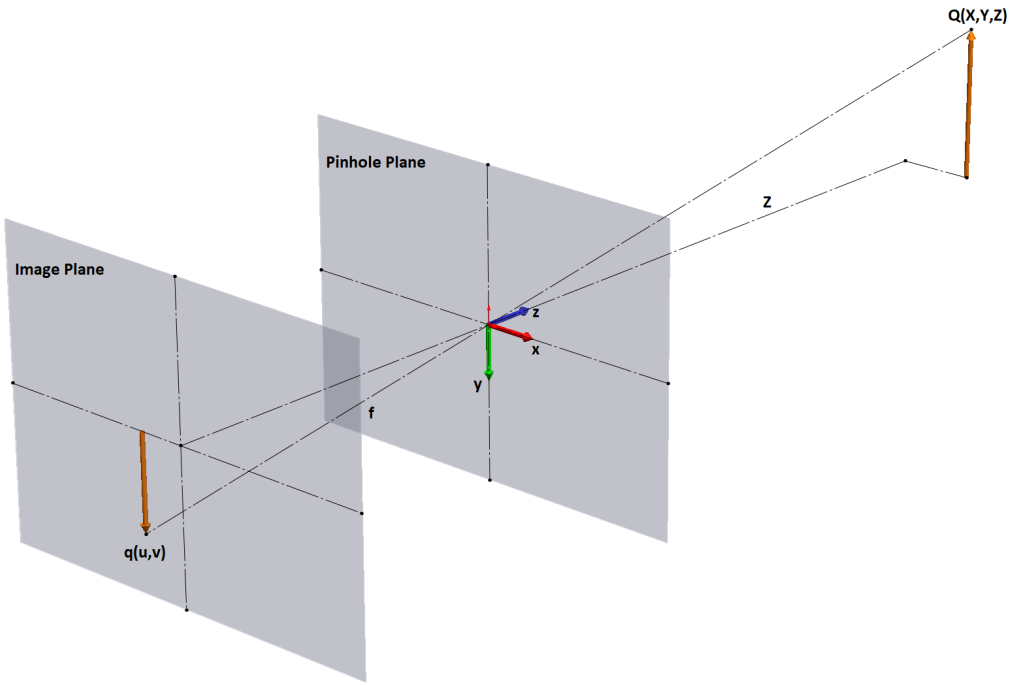

Figure 2.7: Pinhole Camera Model

$$\mathbf{q} = \mathbf{K}\,\mathbf{Q} \quad where \quad \mathbf{q} = \begin{bmatrix} u \\ v \\ w \end{bmatrix}, \mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{2.26}$$

2.5.2 Lens Distortion

Optical lenses have some distortions due to manufacturing and assembly errors. Those imperfections can be classified in different mathematical models: radial distortion, tangential distortion, thin prism model, and tilted model. The overall distortion formulation can be arranged with a single distortion or combinations of those distortions. Figure 2.8 represents the radial distortions; no distortion, positive radial distortion, and negative radial distortion, respectively.

The lenses are manufactured as spherical, not as parabolic. This is the main reason for radial distortions. The reflection increases while going from center of the lens to the edges. The radial distortions are zero at the image's optical center and can be

Figure 2.8: Distortion Examples [64]

expressed by the Taylor series expansion. The radial distortion can be modeled as Eq. 2.27.

$$\begin{aligned} x_{radial} &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ y_{radial} &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \end{aligned} \tag{2.27}$$

The tangential distortion occurs when the lens is mounted not perfectly parallel to the image plane. Two more parameters, p_1 and p_2 arise due to the tangential distortion.

$$\begin{aligned} x_{tangential} &= x + 2p_1 xy + p_2(r^2 + 2x^2) \\ y_{tangential} &= y + p_1(r^2) + 2y^2 + 2p_2 xy \end{aligned} \tag{2.28}$$

where $r^2 = x^2 + y^2$. As a result, there are five distortion coefficients k_1, k_2, p_1, p_2, and k_3. The other distortion models have less effect compared to the tangential and radial distortion.

2.5.3 Homography

Homography is the projection mapping from one plane to another plane. As a result, mapping a 3D point to a 2D image point is possible by using planar homography. Homogenous coordinates of target Q and image plane projection q are expressed in Eq.s 2.29, and 2.30. The homography can be written as Eq. 2.31.

$$\tilde{\boldsymbol{Q}} = \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T \tag{2.29}$$

$$\tilde{\boldsymbol{q}} = \begin{bmatrix} u & v & 1 \end{bmatrix}^T \tag{2.30}$$

$$\tilde{\boldsymbol{q}} = s\mathbf{H}\tilde{\boldsymbol{Q}} \tag{2.31}$$

Where s is the scale parameter that represents an arbitrary scale factor. **H** has two different parts, the physical transformation and the projection. The rotation **R** and translation **t** generate the first part, and the camera matrix represents the projection. Therefore, the homography matrix can be expressed as a rotation matrix **R**, translation vector **t**, and camera matrix **K**.

$$\mathbf{H} = \mathbf{K}\begin{bmatrix} \mathbf{R} \mid \mathbf{t} \end{bmatrix} \tag{2.32}$$

Then, by putting Eq. 2.32 into Eq. 2.31 and divided both side by s, Eq. 2.33 can be obtained.

$$\alpha\tilde{\boldsymbol{q}} = \mathbf{K}\begin{bmatrix} \mathbf{R} \mid \mathbf{t} \end{bmatrix}\tilde{\boldsymbol{Q}} \tag{2.33}$$

$$\alpha \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{2.34}$$

The camera matrix **K** and distortion coefficients are called intrinsic parameters, while rotation matrix **R** and translation vector **t** represent the extrinsic parameters.

2.5.4 Camera Calibration Process

To solve camera equations and to find the unknown intrinsic and extrinsic parameters, known world coordinate points are needed. A known geometric pattern like marker

corner, marker center, circular blobs, and chessboard patterns [65] are used for this purpose. A chessboard is the most common calibration pattern.

The chessboard represents a coordinate system at $z = 0$ plane. Coordinates of each point can be found on the image plane by image processing. On the other hand, the sides of all corners are known by chessboard geometry in 3D world coordinates. Different images should be taken to estimate lens distortions, which are distributed all around the image plane.

In Eq. 2.34, there are intrinsic parameters, extrinsic parameters, image and world points of the calibration board. There are four camera parameters (f_x, f_y, c_x, c_y) and five distortion parameters in which three of them (k_1,k_2,k_3) are radial and two of them (p_1,p_2) are tangential components. The Euler rotations ψ, ϕ, and θ are enough to construct 3×3 rotation matrix, t_x,t_y, and t_z represent the translation vector. Totally, there are nine intrinsic for camera and six extrinsic parameters for each image.

The image and world coordinates of the Eq. 2.34 are known. The equation can be represented as $\mathbf{A}\ \boldsymbol{\gamma} = \mathbf{B}$ form, which can be solved by LSM. Since distortion coefficients, camera matrix, and external coefficients matrix are in the multiplication form, there is a correlation between those terms. As the number of images and the number of corner points increase, these terms can be solved more precisely.

2.6 Real-Time Object Tracking

The process of computing the trajectory or path of an object in consecutive frames in real-time is expressed as real-time object tracking. The object tracking uses features of the object in the video stream. The real-time object tracking pipeline consists of three steps: object detection, object classification, and object tracking. Real-time object tracking may be utilized for several purposes, including security, surveillance, and human-computer interaction.

2.6.1 Object Detection

Object detection is utilized to obtain any item in the camera frame and identify them. As stated before, it is the first step in the object tracking pipeline. Detecting objects may be done in two ways. In the first approach, the user manually locates the object of interest in the video stream's initial image and allows the system to identify features such as edges, corners, or colors. The second approach is to use predefined features to perform automated detection. There are different approaches to classify object detection methods. With the development in the deep learning area, those methods can be classified as traditional computer vision-based methods and deep learning-based detection methods.

Viola-Jones Detectors [66] and Histogram of Gradient Detector [67] are common traditional detector examples. On the other hand, YOLO [68] and RCNN [69] are common deep learning-based object detection methods.

2.6.2 Object Classification

Object classification is a computer vision method used to classify or predict the class of a specific object in an image. The main goal of this method is to identify the features in an image accurately. In classical computer vision methods, those features are defined by researchers. Commonly used features are shape, motion (optical flow), color, texture, or combinations. In deep learning-based methods, those features are created by the neural network itself.

Histogram of gradient is a widely used descriptor that uses shape information. It might be used for classification as well as for detection. Another feature that is beneficial for classification is motion. It is possible to determine a person's identity based on its gait or classify the vehicles using an optical flow.

Color is another feature that is useful for classification. The color information is not limited to grayscale or RGB(red, green, blue). There are various color spaces in literature like HSV(hue, saturation, value), YUV(Luminance, bandwidth, chrominance), and LUV. For example, Vandenbroucke et al. [70] combined some color

spaces to design an accurate pixel classifier. Lastly, texture analysis is a handy tool in some computer vision applications like classification and segmentation based on local changes of color or intensity values. There are several ways to represent the texture information. Scalar numbers, discrete histograms and empirical distributions are some of those representations. They characterize the textural information in the images like disparity, smoothness and orientation.

2.6.3 Object Tracking

Object tracking is an important problem with various application areas such as surveillance, traffic and crowd monitoring, human-computer interaction, robotics, and autonomous vehicles. Object tracking aims to estimate the trajectory of a moving object using data sequence from single or multiple sensors. The sensors that are used in this work are cameras, and the problem becomes visual object tracking. Several issues make object tracking challenging. Some of them are occlusion, illumination change, complicated object motion, image noise, scale, and pose changes. To overcome the mentioned issues, the tracking method should be robust. There are several ways to design an object tracker.

According to Fiaz et al. [71], visual trackers may be classified as the number of objects, context-awareness level, and algorithm learning instant. Single object trackers focus only on the single target in the whole image, while multiple object trackers may follow multiple targets simultaneously. Generative models search the image for the best-matched window. On the other hand, discriminative models separate the target patch from the background.

Online learning algorithms tune the hyper-parameters while tracking the object, whereas offline algorithms use predefined parameters and do not update them. One can also classify the visual object trackers as deep learning-based vs. classical computer vision-based trackers. Although deep learning methods are more successful in most computer vision problems, they do not apply to this work for several reasons, like lack of labeled training data and long training times.

CHAPTER 3

MULTIANGULATION WITH DIGITAL CAMERAS

There are several measurement techniques mentioned in Chapter 2. Although the most accurate compensation method is the online compensation using LTS, it is a very expensive method due to laser interferometers. Therefore, a cheaper and comparable in terms of performance alternative is suggested in this chapter.

Section 3.1 briefly explains the suggested vision-based multiangulation method from several perspectives. In Section 3.2, positioning approaches and methods are explained. In Section 3.3, the calibration route in which the mathematical relation between the gimbal center and the target is estimated. Section 3.4 explains the Parallax effect, a source of error in vision systems, which can cause an unexpected miscalculation. The mathematical formulations to combine camera measurements used in multiangulation are described in Section 3.5. Finally, Section 3.6 demonstrates the method to obtain AZ and EL angles from camera measurements in the simulation environment.

3.1 System Design

There are two gimbals in the measurement system, where each gimbal carries a camera. The cameras are mounted eccentrically with respect to the rotation axis of the gimbal. Each gimbal behaves as an angular measurement unit, and multiangulation is selected as the position estimation method. The suggested system design shown in Figure 3.1 composes of two measurement gimbals and a CNC Machining Center frame with the probe.

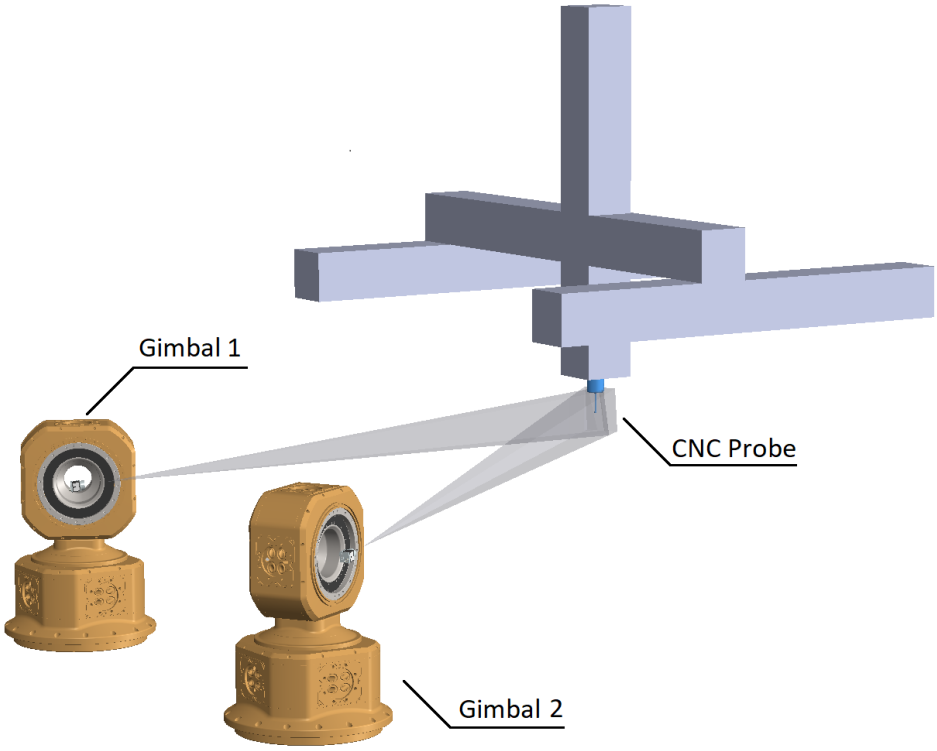

Figure 3.1: System Architecture

Each camera tracks the end of the machining center with real-time image processing methods. The tracking is initialized manually by clicking on the object of interest from the user interface. A tracking window appears around the object. Due to the Parallax effect, explained in Section 3.4 and lens distortions, described in Subsection 2.5.2, the object of interest is kept at the center of the image. The main controller calculates the offset from the center, and angular commands are fed into the corresponding gimbal controller to compensate for the offset. The two DoF gimbal is rotated in two angles during tracking. The angular encoders of the gimbals, azimuth and elevation, are determined through the angular measurements. The unit vectors from the camera to the target are measured from camera measurements in the pixels. The current angle read from the encoder, and calculated unit vector are combined by geometrical relation, explained in Section 3.5.

There is a lens in front of the camera sensor for vision systems. Due to the lens geometry and other imperfections, there are deflections in the rays passing through the

camera lens from the camera sensor to the object. Through the camera calibration process, focal length, intrinsic parameters, and lens distortions are estimated in advance; the process is explained comprehensively in Section 2.5. Due to the radial distortion, the imperfections increase in the image by moving the center from the outside. Also, after the camera calibration, there will be a slight but non-zero offset between calibrated points and original geometry. Therefore, to minimize the aforementioned errors, only the central region of the camera is used during the procedure.

Figure 3.2: Position Estimation from Camera Measurements

The second calibration routine estimates the relative position of the gimbal with respect to the target CS. Each gimbal is utilized as an angular measurement unit. The multiangulation is used to estimate the position of the target. Angular measurement of each gimbal creates a line in the space. If there are no errors, the line of sight of two cameras intersects at a point, which is called the target point. Due to the imperfections and other error accumulations, those lines do not intersect in the real world. Geometrically, the closest points of each line to the other one should be found, and a line is drawn between the two closest points. The middle point of this line is obtained as the target point. Analytically, the LSM is used to determine the target point precisely. The process is given in Figure 3.2 and described extensively in Section 3.5.

There are two parts to the studies. First of all, a virtual simulation environment used the validate the psychical formulations and determine the error budget. In the second part, the real system is used. The measurements are collected from DMG Mori CMX 70 U 5-axis CNC Machining Center. The position accuracy of the CNC Machining

center is 2 μm, so it is assumed as ground truth.

3.2 Coordinate System (CS) Relations

The suggested architecture has four different CS; (1) World CS (**W**), (2) Gimbal CS (**G**), (3) Rotated Gimbal CS (**RG**), and (4) Camera CS (**C**). The **W** and **G** are shown in Figure 3.3, whereas, Figure 3.4 displays **G**, **RG** and **C**. Note that the W point is the origin of **W**, and the G point is the origin of **G**. Also, all CSs are right-handed.

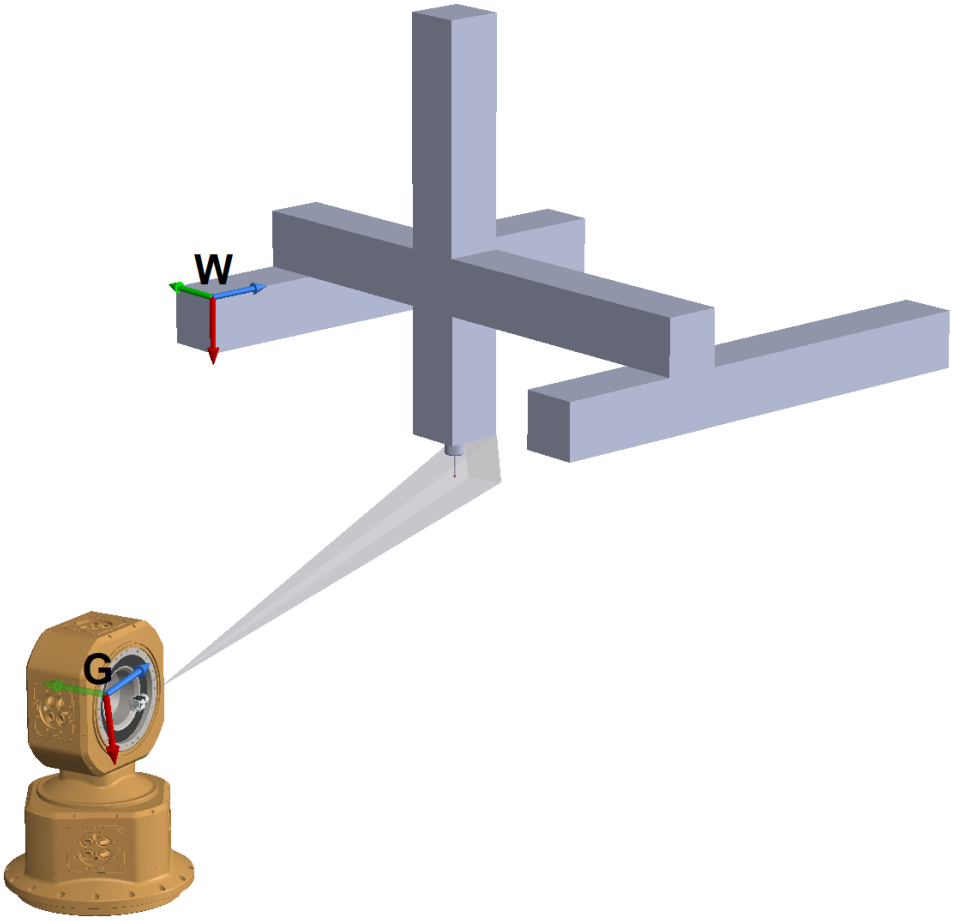

Figure 3.3: World CS and Gimbal CS

Since the gimbals are driven by their coordinate system, the relation between **G** and **C** should be found. This relation will be used to convert the image-based measurements taken from **C** to **W**.

To obtain an incremental precise measurement device, the rotation center of the gimbal in **W** should be estimated. The process called individual gimbal calibration finds the geometric relation between **W** and **G**. Both rotation and translation matrices between two CSs, are estimated by a calibration routine.

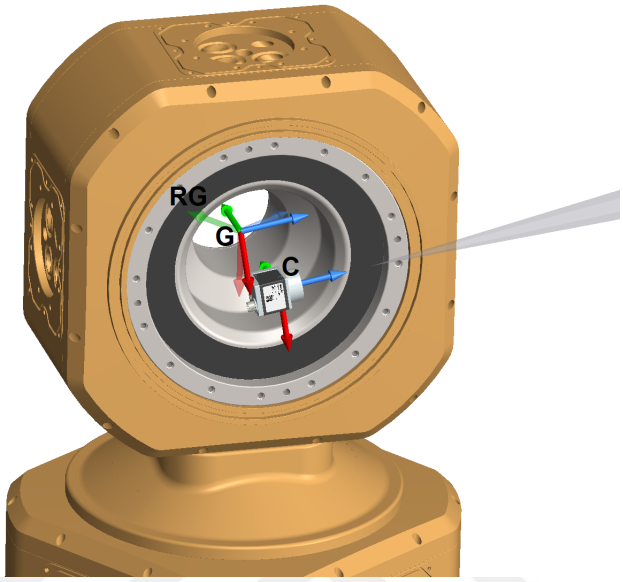

Figure 3.4: Gimbal CS, Rotated Gimbal CS and Camera CS

The relation between **G** and **RG** is measured by AZ and EL encoders' measurements. These two CSs are pivoted at the rotation center of the gimbal, so there is no translation between them. 3×3 rotation matrix is represented as $\mathbf{R}^{z}_{x,y}$, is the rotation operations of frame x with respect to frame y, expressed in frame z. $\mathbf{n}^{x}$ represents unit vector in CS x. In the Eq. 3.1, the unit vector **n** is transformed from **G** to **RG** where $\mathbf{n}^{g}$ represents unit vector at **G** and $\mathbf{n}^{rg}$ indicates the unit vector at **RG**.

$$\mathbf{n}^{g} = \mathbf{R}^{g}_{rg,g}\mathbf{n}^{rg} \tag{3.1}$$

The CS is placed in a way that the AZ-axis is the angle from the line of sight to the x-z plane, and the EL-axis is the angle from the line of sight to the y-z plane. Each rotation matrix can be expressed as a combination of three basic rotation matrices, where basic rotation matrix is a rotation around one single axis. Basic rotation matrices are expressed as $\mathbf{R}_{a}$, where a is the rotation axis. When the geometry of the gimbal is

analyzed, it is clarified that first, the gimbal is rotated AZ angle, then EL angle.

$$\mathbf{R}^{g}_{rg,g} = \mathbf{R}_{y}(\text{-EL})\mathbf{R}_{x}(\text{-AZ}) = \begin{bmatrix} \cos(\text{-EL}) & 0 & -\sin(\text{-EL}) \\ 0 & 1 & \\ \sin(\text{-EL}) & 0 & \cos(\text{-EL}) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\text{-AZ}) & \sin(\text{-AZ}) \\ 0 & -\sin(\text{-AZ}) & \cos(\text{-AZ}) \end{bmatrix} \tag{3.2}$$

As shown in Figure 3.4, the camera is attached at the side of the EL axis of the gimbal. There is both rotation and translation between **RG** and **C**. The relation is given in Eq. 3.3.

$$\mathbf{n}^{rg} = \mathbf{R}^{rg}_{c,rg}\mathbf{n}^{c} + \mathbf{T}^{rg}_{c} \tag{3.3}$$

The camera connection flange is machined with very high accuracy, so the designed CAD model of the camera position is directly used as a translation vector of the camera. The camera placed in a way that the image plane is on the y-axis. The image is created behind the camera and there is a distance f which is focal length of the camera, can be calculated by camera calibration. The manufacturing and assembly imperfections are expected in the order of micrometers, and they are assumed as zero.

$$\begin{aligned} \mathbf{T}^{rg}_{c} &= \begin{bmatrix} 0 & -124.5mm & -f \end{bmatrix}^{T} \\ \mathbf{R}^{rg}_{c,rg} &= \mathbf{I}_{3} \end{aligned} \tag{3.4}$$

The relation between the **W** and **G** is given in Eq. 3.5.

$$\mathbf{n}^{w} = \mathbf{R}^{w}_{g,w}\mathbf{n}^{g} + \mathbf{T}^{w}_{g} \tag{3.5}$$

The relation between the **W** and **C** are given in Eq. 3.6, and the combination of Eqs. 3.1, 3.3, and 3.5 are given in Eq. 3.7.

$$\mathbf{n}^{w} = \mathbf{R}^{w}_{c,w}\mathbf{n}^{c} + \mathbf{T}^{w}_{c} \tag{3.6}$$

$$\mathbf{n}^w = \mathbf{R}^w_{g,w}\mathbf{R}^g_{rg,g}\mathbf{R}^{rg}_{c,rg}\mathbf{n}^c + \mathbf{R}^w_{g,w}\mathbf{R}^g_{rg,g}\mathbf{T}^{rg}_c + \mathbf{T}^w_g \tag{3.7}$$

The known terms shown in Eqs. 3.2 and 3.4 are put in Eq. 3.7 and Eq. 3.8 is obtained.

$$\mathbf{n}^w = \mathbf{R}^w_{g,w}\mathbf{R}_y(\text{-EL})\mathbf{R}_x(\text{-AZ})\mathbf{n}^c + \mathbf{R}^w_{g,w}\mathbf{R}_y(\text{-EL})\mathbf{R}_x(\text{-AZ})\mathbf{T}^{rg}_c + \mathbf{T}^w_g \tag{3.8}$$

There are 2 unknown notations in Eq. 3.7, which are $\mathbf{R}^w_{g,w}$ and $\mathbf{T}^w_g$; overall, there are 6 unknowns.

3.3 Individual Gimbal Calibration

The relation between **C** and **W** can be estimated using Open Source Computer Vision Library (OpenCV) method named **solvePnp**, which finds the transformation parameters, $\mathbf{R}^w_{c,w}$ and $\mathbf{T}^w_c$ from 3D to 2D point matching using intrinsic camera parameters.

$$\begin{aligned} \mathbf{R}^w_{c,w} &= \mathbf{R}^w_{g,w}\mathbf{R}^g_{rg,g}\mathbf{R}^{rg}_{c,rg} \\ \mathbf{T}^w_c &= \mathbf{R}^w_{g,w}\mathbf{R}^g_{rg,g}\mathbf{T}^{rg}_c + \mathbf{T}^w_g \end{aligned} \tag{3.9}$$

Since the gimbal is rotating around the rotation axis, the position of the camera changes during the process. The geometric relation between the camera and the gimbal is known, so the estimated camera position can be transformed to the gimbal center. Rearrange variables in Eq. 3.9 to isolate unknown terms as given in Eq. 3.10.

$$\begin{aligned} \mathbf{R}^w_{g,w} &= \mathbf{R}^w_{c,w}\mathbf{R}^{rg^{-1}}_{c,rg}\mathbf{R}^{g^{-1}}_{rg,g} \\ \mathbf{T}^w_g &= \mathbf{T}^w_c - \mathbf{R}^w_{g,w}\mathbf{R}^g_{rg,g}\mathbf{T}^{rg}_c \end{aligned} \tag{3.10}$$

3.4 The Parallax Effect

The parallax effect is the phenomenon in which the location or direction of a target appears to change when seen from multiple perspectives, such as through a lens of a camera or a viewfinder.

In Figure 3.5, the parallax effect is shown. The first measurement is taken from point A, in which the optical axis and the target are coincident. After the rotation of the camera to point B, around the gimbal center, the angle between the target and the camera consists of two different components. There is angle α for rotation and β for the parallax effect. As a result, the object should be kept at the center of the image to overcome the parallax effect, or the camera position should be calculated in each instant.

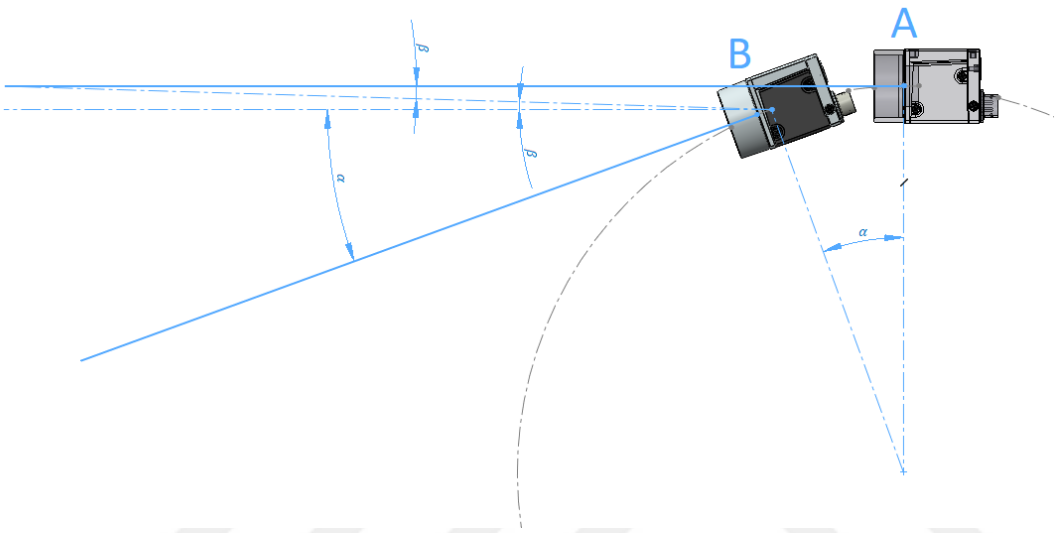

Figure 3.5: Demonstration of the Parallax Effect

3.5 Position Estimation from Angular Measurements

A method based on the perpendicular distances of the target position to the direction vectors measured by the camera was considered more suitable for thesis purposes. The sum of the squares of the distances between the unknown target point and the line equations constructed using the gimbal locations and direction vectors is determined to be the minimum using this technique. In this approach, gimbal positions a_i are expressed in Eq.3.11.

$$\mathbf{a}_i = \begin{bmatrix} a_x & a_y & a_z \end{bmatrix}^T \tag{3.11}$$

Using the angles obtained through each camera, the unit direction vector n_i can be

expressed as shown in Eq. 3.12.

$$\mathbf{n}_i = \begin{bmatrix} n_x & n_y & n_z \end{bmatrix}^T, ||\mathbf{n}||_2 = \mathbf{n}^T\mathbf{n} = 1 \tag{3.12}$$

The equation of the camera direction from the camera to the target is given in Eq. 3.13, where t is a scalar value that represents the depth of the object and $\mathbf{p}$ is the target position.

$$\mathbf{p} = \mathbf{a} + t\mathbf{n} \quad , \quad -\infty < t < \infty \tag{3.13}$$

The sum of the distances of the target position to the line equation obtained with K different measurements ($K \geq 2$) can be written as Eq. 3.14.

$$D(\mathbf{p}; \mathbf{A}, \mathbf{N}) = \sum_{j=1}^{K} D(\mathbf{p}; \mathbf{a}_j, \mathbf{n}_j) = \sum_{j=1}^{K} (\mathbf{a}_j - \mathbf{p})^T (\mathbf{I} - \mathbf{n}_j\mathbf{n}_j^T)(\mathbf{a}_j - \mathbf{p}) \tag{3.14}$$

The location that provides the smallest value of this sum can be set as the target location. Taking the derivative of the above term with respect to $\mathbf{p}$ and setting it equal to zero:

$$\frac{\partial D}{\partial \mathbf{p}} = \sum_{j=1}^{K} -2(\mathbf{I} - \mathbf{n}_j\mathbf{n}_j^T)(\mathbf{a}_j - \mathbf{p}) = 0 \tag{3.15}$$

Convert Eq. 3.15 in $\mathbf{Ap} = \mathbf{B}$ form to use LSM where,

$$\mathbf{A} = \sum_{j=1}^{K} (\mathbf{I} - \mathbf{n}_j\mathbf{n}_j^T) \quad , \quad \mathbf{B} = \sum_{j=1}^{K} (\mathbf{I} - \mathbf{n}_j\mathbf{n}_j^T)\mathbf{a}_j \tag{3.16}$$

A system of linear equations is obtained in Eq. 3.16. Solving this system of equations gives the target position estimation. $\mathbf{p}$ vector can be estimated by multiplication of pseudo-inverse of $\mathbf{A}$ and $\mathbf{B}$, shown in Eq. 3.17.

$$\mathbf{p} = \mathbf{A}^{+}\mathbf{B} \tag{3.17}$$

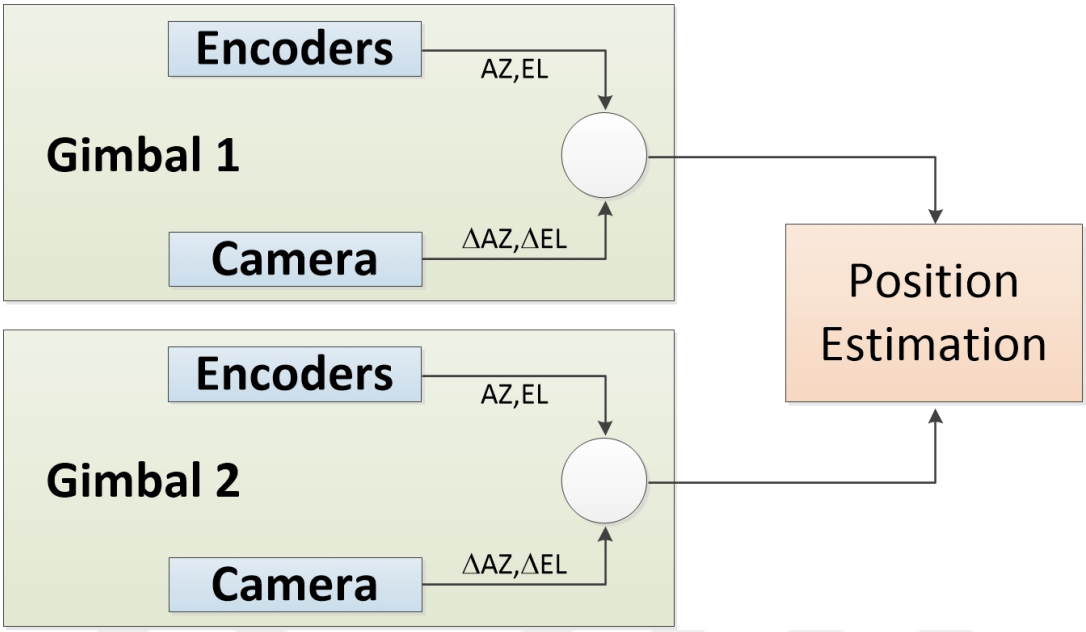

Figure 3.6: Measurements Usage

3.6 Obtaining Angular Measurement From Camera

The measurements are completed via image processing in the real world, and u and v terms are obtained in the image plane. The usage of camera directions is explained in Section 3.5, where the gimbal angles are measured from AZ and EL encoders. On the contrary, the ideal gimbal angles should be calculated in the simulation. The image coordinates u and v are used to obtain gimbal angles. An imaginary plane is assumed at $z = 1$. The ray casting from the camera center with AZ and EL angles crosses the $z = 1$ plane at point (u, v). The gimbal angles are obtained from the image plane by Eqs. 3.18 and 3.19 with the given geometry in Figure 3.7.

$$EL = -\tan^{-1}(u) \tag{3.18}$$

$$AZ = \tan^{-1}(\frac{v}{u^2 + 1}) \tag{3.19}$$

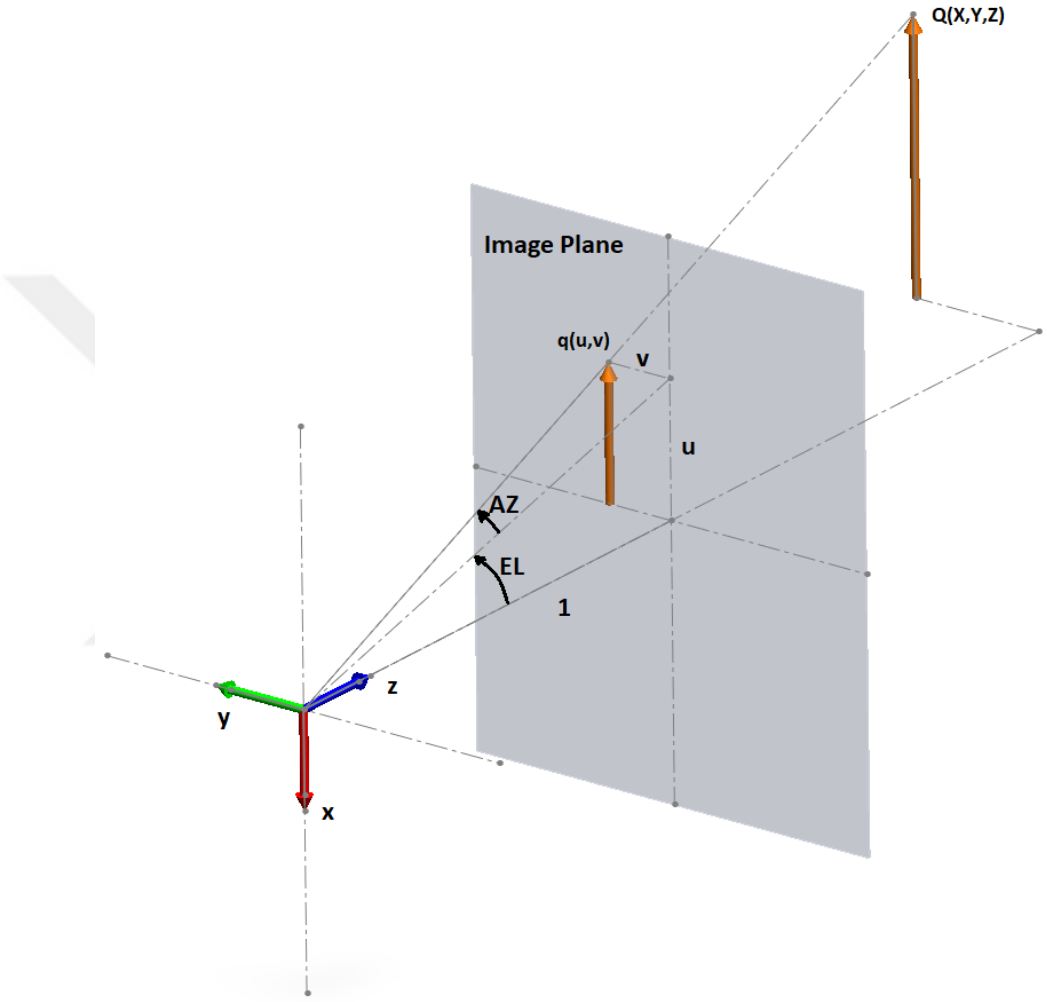

Figure 3.7: Angular Measurements in Imaginary Camera Coordinate System

CHAPTER 4

IMPLEMENTATIONS, RESULTS AND DISCUSSION

In this chapter, implementation details are explained in Section 4.1. Then, the Python Simulation is explained comprehensively in Section 4.2. The simulation results are given and discussed in Section 4.3. The real system results are explained in Section 4.4, including camera calibration results, circle fit results, gimbal calibration results, and overall system performance. Finally, we comment on the results and discuss the essential points in Section 4.5.

4.1 Implementations Details

As mentioned in Chapter 3, the system used gimbals. In the thesis, the gimbals are borrowed from Modesis Machine Technologies. According to the datasheet of the product, the repeatability is 4 arcsecond("), and the positioning accuracy is 2" where arcsecond is a unit of angular measurement equal to $\frac{1}{3600}$ of one degree denoted by the symbol ". In Figure 4.1, the gimbal is shown with attached camera.

There are stand-alone Gimbal Controllers written in C++, filtering the input data and creating instantaneous command data for each axis motor driver. The electronic and software bases of the product are configured according to thesis requirements.

The Control and Communication diagram is given in Figure 4.2. Both motor drivers of gimbals are controlled by their individual Gimbal Controllers. Each control loop between the Gimbal Controller and the motor drivers can be closed in 250 microseconds using EtherCAT technology, where EtherCAT is a real-time Industrial Ethernet-based field system invented by Beckhoff Automation [72]. The controller is con-

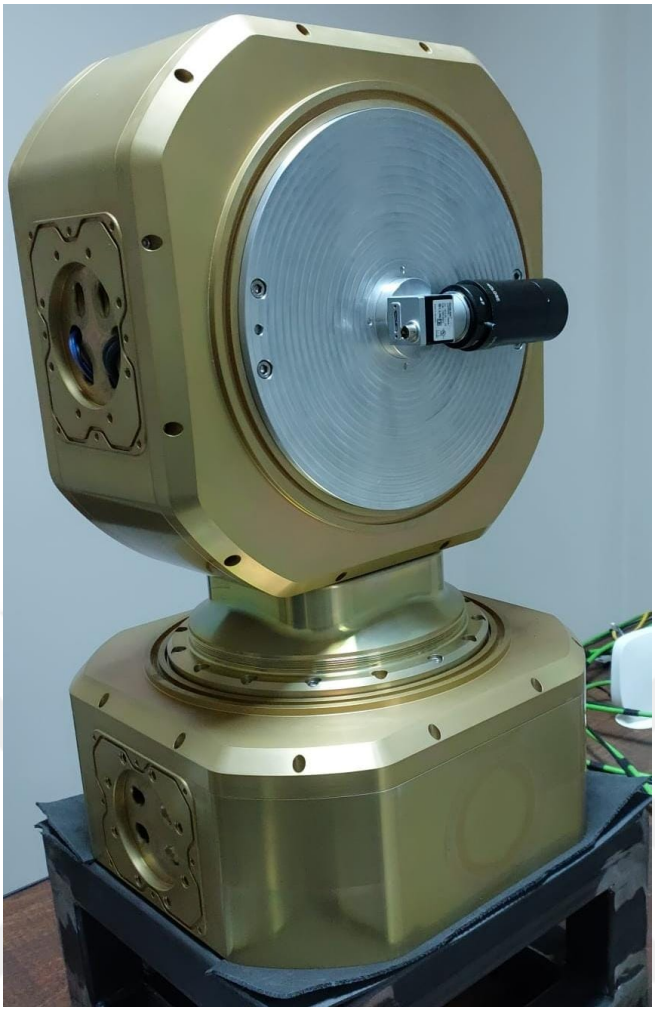

Figure 4.1: Gimbal with Camera

figured as an EtherCAT master, and the motor drivers act as a slave. This control loop measures the precise angular position of the target position by using the encoder information received from the gimbals in real-time with EtherCAT technology.

The camera of each gimbal is connected to Main Controller directly via USB3.0, and image processing is handled in Main Controller. Filtered commands are sent from Main Controller to each gimbal position controller via EtherNet/IP technology.

The graphical user interface is written in Java. The GUI has several different sub-modules such as Object Tracking, Camera Calibration, and Log recording.

In the thesis, monochrome Basler acA2040-55um cameras are used. The cameras manufactured by Basler are working on Pylon software. For the developers, there is

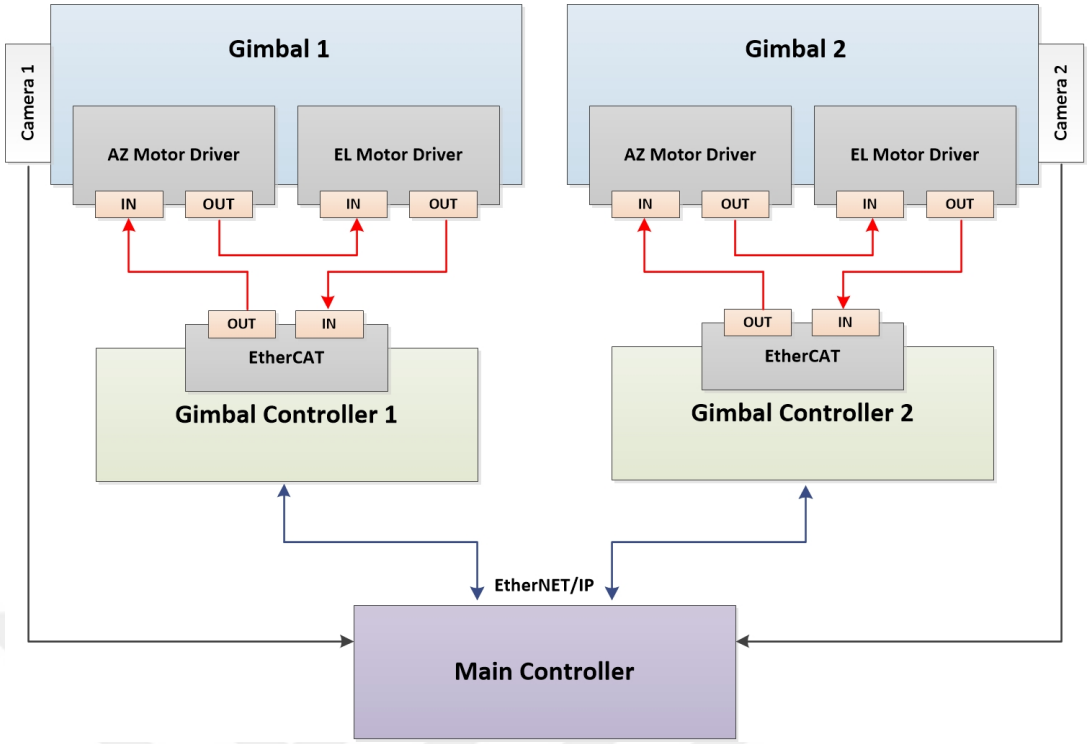

Figure 4.2: Communication and Control Cycle

only C++ and python support from Pylon. Since the GUI is written in Java, the C++ libraries of Pylon should be converted to Java.

A software development tool named Simplified Wrapper and Interface Generator (SWIG) links the C and C++ programs to other programming languages [73]. SWIG creates the interface between two different programming languages so Java code can access C++ functions and use them.

For image processing, camera calibration, and real-time object tracking, Open Source Computer Vision (OpenCV) libraries are used. The up-to-date OpenCV and OpenCV-Contrib, in which there are extra OpenCV modules, repositories are compiled via CMake.

The camera lens is selected with several concerns. First of all, the mount type of the camera and the lens should be the same. Secondly, since the Depth of Field changes according to the working environment, the lens has an adjustable focal length, focal aperture, and adjustable focus. On the other hand, the camera was used with high focus so, the all area of camera sensor is not used. Normally, sensor format of camera

and lens should be same, otherwise an obstruction occurs in front of the camera. But in this application since all camera sensor is not used, the format is not essential. A varifocal manual zoom lens is selected.

As mentioned in Section 3.1, the thesis consists of two parts. In the first part, a python-based simulation environment is implemented for two purposes. Formulations explained in the Chapter 3 are corrected, and error budgets of several different error sources are analyzed.

SolidWorks CAD program was used to draw the overall system. 3D CAD model of the system was used not only for manufacturing but also for verification of python simulation. During the implementation of the simulation, the fundamental geometric relationship between objects and CSs is justified via the SolidWorks CAD model.

4.2 Python Simulation

The simulation is implemented using Python 3.9 . The real geometry of the designed system and the formulations given in Chapter 3 is verified via this simulation.

In the simulation, n different gimbals are placed in a given position with a given orientation in 321-Euler angle representation. The Euler angle representation shows that the gimbal is rotating around the z-axis, y-axis, and x-axis, respectively.

At each target point, the gimbal starts the tracking again. The camera position is calculated with updated pose and current AZ and EL angles of gimbal. Also, the camera direction $\mathbf{n}_c$ is converted to $\mathbf{n}_w$ using $\mathbf{R}^{rg}_{c,rg}$, $\mathbf{R}^{g}_{rg,g}$, and $\mathbf{R}^{w}_{g,w}$. Using the method explained in Section 3.5, the target position is estimated.

In Figure 4.3, the overall simulation view is given. The red and green spheres represent the gimbals, black points represent the current camera positions, and calibration points are demonstrated as blue.

There are 12 different points for both camera calibration and individual gimbal calibration. A tracker is inside the simulation, implemented with a Proportional control strategy. At the beginning of each simulation, the gimbals are getting measure-

Algorithm 1 Simulation

Initialize p different calibration points
Initialize m different measurement points
Initialize g different gimbal with random starting pose
Initialize WP array which represents 3D points
Initialize IP array which represents 2D points
for each gimbal G **do**
    for each calibration point P **do**
        Calculate n^c using G origin, camera offset, P and $R^w_{g,w}$
        Insert P and n^c into WP and IP
    end for
    Use SolvePnp function to obtain rotation and translation vectors from WP and IP
    Transform rotation vector to rotation matrix
    Recover Euler-321 Angles from rotation matrix
    Update Gimbal pose with using Euler angles and translation vector.
end for
for each measurement point M **do**
    Initialize A and B matrices
    for each gimbal G **do**
        Calculate camera position using $R^g_{rg,g}$, $R^w_{g,w}$, camera offset and G origin.
        Calculate n_w using current AZ and EL angles, $R^{rg}_{c,rg}$, and $R^w_{g,w}$.
        Calculate A and B matrices in Eq. 3.16
    end for
    Solve the matrices as in Eq. 3.17 and obtain point M
end for

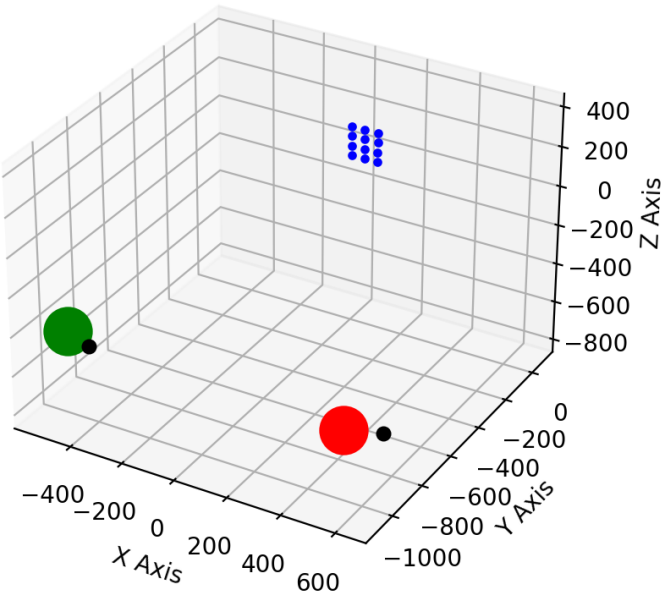

Figure 4.3: Python Simulation Overall View

ments from calibration points using the tracker. Each camera's intrinsic parameters are found by using OpenCV **calibrateCamera** function. When the measurements from each camera is calculated, the camera matrix is used, so the actual values of camera matrix are known. On the other hand, The **calibrateCamera** function returns the estimated camera matrix and RMS of re-projection error. So, the first set of the output of the simulation is camera matrix errors and RMS of re-projection errors. Then, **solvePnp** function is used with estimated camera matrix and calibration points in 2D and 3D. The outputs of **solvePnp** function are rotation and translation vectors of that image instant. The rotation vector is converted to rotation matrix. The rotation matrix is resolved into Euler angles by Euler-321 rotation. These estimated Euler angles represent the camera sensor plane with respect to target. The gimbal center and orientation can be calculated through given formulation in Chapter 3. The pose error of the gimbal is the second set of the output of the simulation.

Using the estimated gimbal pose, each gimbal tracks the measurement points and

calculates the **A** and **B** matrices given in Eq. 3.16. The third set of output is the position error of measurement points.

4.3 Simulation Results

In the real system, there are two different error sources. Firstly, the measurements taken from the camera may be noisy. Secondly, although the encoders of the gimbals have ultra high precision, there is a small angular error from the encoders. The encoders are 32 bits, so one turn, 360° is divided into 2^{32}. As a result, 1" arc corresponds to ~ 3314 counts. The simulator is designed to measure the effect of two error sources with the distance between gimbals.

In the first set, the gimbals are placed with a 1000 mm distance between them, and they are mirrored with respect to $y = 0$ plane. Gimbal 1 placed in $(-500, 500, -1000)$ coordinates with $(0, 0, 0)$ orientation, Gimbal 2 placed in $(-500, -500, -1000)$ coordinates with $(0, 0, 0)$ orientation.

Table 4.1: Noise Levels and Test Cases

Noise Type	Encoder Noise (°)	Camera Noise (px)	Test Case
Noise Level	0.0	0.0	A
	0.00279	0.0	B
	0.01	0.0	C
	0.1	0.0	D
	0.0	0.03125	E
	0.0	0.1	F
	0.0	0.25	G
	0.0	0.5	H

A Monte-Carlo simulation is designed for error budgeting. First of all, the simulation is run 1000 times without noise to verify the formulations, which is represented as Case A in Table 4.1. Then, 10", 0.01°, and 0.1° noise is added to both AZ and EL encoders of gimbals in Case B, C and D. Then, $1/32$, 0.1, 0.25, and 0.5 normally distributed pixel errors are added to camera measurements in Case E, F, G and H.

Table 4.2: Test Cases and Camera Calibration Errors

Test Case	Re-projection	f_x (px)	f_y (px)	c_x (px)	c_y (px)
A	2.86E-05	0.00527	0.00361	0.00229	0.00507
B	2.86E-05	0.00527	0.00361	0.00229	0.00507
C	2.86E-05	0.00527	0.00361	0.00229	0.00507
D	2.86E-05	0.00527	0.00361	0.00229	0.00507
E	0.01097	0.83568	0.85866	0.64519	0.80733
F	0.03511	2.97540	2.92388	2.14214	2.69909
G	0.08737	7.61401	7.96542	6.29677	6.87458
H	0.17526	15.25009	14.77996	11.03916	13.85074

The RMS value of re-projection error and error in the camera matrix is given in Table 4.2. This RMS value represents the quality of the camera calibration process. As the RMS decreases, the quality increases. The encoder noises do not affect the re-projection RMS value of **calibrateCamera**. Even if the encoder value is wrong, there is no error in the image so that Case A, B, C, and D represent perfect calibration for both re-projection error and camera matrix elements (f_x, f_y, c_x, and c_y). As a result, encoder noise does not affect the calibration quality. On the other hand, even a small

Table 4.3: Test Cases and Gimbal Pose Errors

Test Case	Orientation Error (°)			Translation Error (mm)			
	x	y	z	x	y	z	d
A	0.0001	0.0001	0.0000	0.0020	0.0007	0.0033	0.0039
B	0.0023	0.0020	0.0010	0.0020	0.0007	0.0033	0.0039
C	0.0081	0.0070	0.0033	0.0020	0.0007	0.0033	0.0039
D	0.0812	0.0694	0.0327	0.0020	0.0007	0.0033	0.0039
E	0.0133	0.0170	0.0058	0.3668	0.3632	0.5779	0.7749
F	0.0481	0.0603	0.0181	1.1892	1.1848	1.9911	2.6043
G	0.1207	0.1600	0.0528	3.0998	3.4125	5.3398	7.0546
H	0.2420	0.3102	0.0944	6.2578	5.7298	10.2339	13.2937

pixel error affects the calibration quality. The pixel error is directly proportional to the re-projection RMS error. As camera pixel error increases, the camera matrix error increases.

The estimated gimbal pose errors are given in Table 4.3. Encoder noise does not create any translation error but creates a small orientation error. This orientation error will create an effect on the final estimation. Moreover, as the camera noise increases, angular and translational error of gimbal pose increase.

Table 4.4: Test Cases and Measurement Errors

Test Case	Translation Error (mm)			
	x	y	z	d
A	0.0024	0.0005	0.0030	0.0039
B	0.0574	0.0360	0.0780	0.1033
C	0.2154	0.1321	0.2789	0.3763
D	2.0270	1.3814	2.7293	3.6696
E	0.5500	0.2262	0.7350	0.9455
F	1.7515	0.7065	2.2712	2.9539
G	5.4134	1.9625	6.8965	8.9843
H	8.8922	3.6392	11.6716	15.1176

The estimated point errors are given in Table 4.4. Although encoder errors do not create a significant error in the camera matrix and gimbal pose, they affect the position accuracy of the system. On the other hand, The encoder noise is directly proportional to final position accuracy.

The simulations are repeated with a 500 mm distance between gimbals. This analysis shows us whether the gimbals should stay closer or place as far as possible.

The camera calibration results are given in Table 4.5. The distance between gimbals does not affect image properties, so there is no significant change in the camera calibration quality and camera matrix elements with respect to the distance. The encoder noise and camera pixel noise relation are similar to the previous case. Encoder noise does not affect the camera matrix elements, and the re-projection error and the error

Table 4.5: Test Cases and Camera Calibration Errors

Test Case	Re-projection RMS	f_x (px)	f_y (px)	c_x (px)	c_y (px)
A	2.89E-05	0.02889	0.00410	0.00500	0.00049
B	2.89E-05	0.02889	0.00410	0.00500	0.00049
C	2.89E-05	0.02889	0.00410	0.00500	0.00049
D	2.89E-05	0.02889	0.00410	0.00500	0.00049
E	0.01095	0.86590	1.18131	1.22740	0.31009
F	0.03502	2.76254	3.60420	3.75199	1.03613
G	0.08761	6.84553	9.77469	9.44449	2.46075
H	0.17519	13.60810	18.87543	20.83641	4.47615

of camera matrix elements (f_x, f_y, c_x, c_y) are proportional to the camera pixel noise.

Table 4.6: Test Cases and Gimbal Pose Errors

Test Case	Angular Error (°)			Translation Error (mm)			
	x	y	z	x	y	z	d
A	0.0000	0.0001	0.0000	0.0007	0.0007	0.0032	0.0033
B	0.0023	0.0021	0.0006	0.0007	0.0007	0.0032	0.0033
C	0.0077	0.0075	0.0023	0.0007	0.0007	0.0032	0.0033
D	0.0800	0.0745	0.0229	0.0007	0.0007	0.0032	0.0033
E	0.0089	0.0204	0.0039	0.0089	0.0204	0.7839	0.8624
F	0.0274	0.0643	0.0126	0.8157	0.8458	2.3899	2.6631
G	0.0698	0.1602	0.0306	1.8900	2.1189	6.3038	6.9137
H	0.1286	0.3241	0.0634	3.9014	4.3747	12.8289	14.1046

The gimbal pose analysis are summarized in Table 4.6. Similar to the first result set, the encoder noise does not create any translation error but creates a small orientation error. The orientation error will create an effect on the final estimation. Moreover, as the camera noise increases, angular and translational error of gimbal pose increase.

The estimation errors are given in Table 4.7. As camera noise and encoder noise increase, estimation errors increase in both 3-axes. On the other hand, the final error of

Table 4.7: Test Cases and Measurement Errors

Test Case	Translation Error (mm)			
	x	y	z	d
A	0.0023	0.0006	0.0009	0.0025
B	0.0044	0.0005	0.1339	0.1677
C	0.0044	0.0005	0.4655	0.5835
D	0.0201	0.0016	4.5204	5.6546
E	0.4691	0.1064	0.5794	0.7530
F	1.6743	0.3495	1.9293	2.5783
G	3.6816	0.8675	4.5612	5.9255
H	8.4034	1.8531	9.3574	12.7127

the systems is different from the distance between gimbals. As the distance increases, the effect of encoder noise increases, but the effect of camera noise decreases.

4.4 Real System Results

In this section, real system test results are explained. The measurement are taken by two gimbals with two different focus levels. The gimbals are fixed during the tests. Only the focal length and zoom level of the lenses are changed. First, the camera calibration process is explained in Section 4.4.1, and results are given. Then, the performance of the circle fit algorithm is analyzed in Section 4.4.2. The individual gimbal calibration results are given with optimization methods in Section 4.4.3. Finally, overall system performance is given in Section 4.4.4 for two different focal lengths.

4.4.1 Camera Calibration Result

The camera calibration process is explained in Section 2.5. The calibration target is the chessboard pattern, the most reliable and most widely used calibration pattern. The calibration pattern used is given in Figure 4.4. This pattern is named using num-

ber of inner corners. The used one is a 6×9 pattern with a 10 mm checker width.

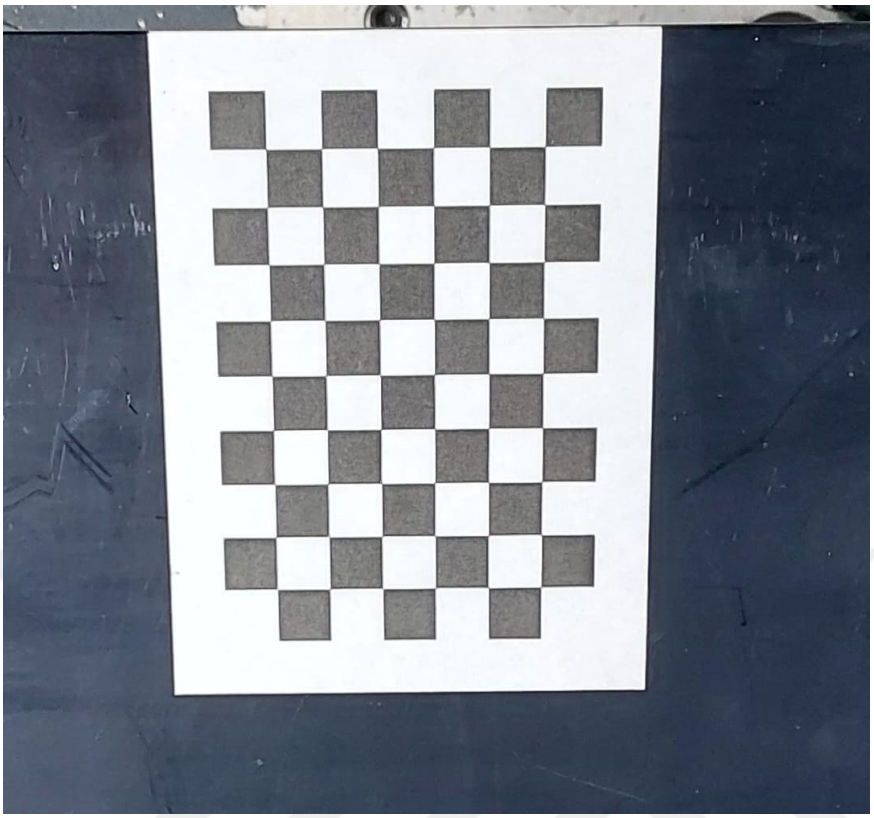

Figure 4.4: Chessboard Pattern for Camera Calibration

The intrinsic parameters found by the OpenCV **cameraCalibration** function are tabulated for both cameras in Table 4.8. In each calibration, 18 different sets of images were taken where each set consists of 10 different images, so there are 180 images. The chessboard is 6×9, so in each image, there are 54 calibration points. Totally, 9720 different points were used in the calibration process.

Table 4.8: Calibration Intrinsic Results

Gimbal	Focal	RMS	Camera Matrix				Distortion				
			f_x	f_y	c_x	c_y	p_1	p_2	k_1	k_2	k_3
1	1	0.4174	31078.41	33738.59	1017.34	2830.49	7.17	101.98	0.57	-0.0019	-31.16
1	2	0.2528	18391.47	18437.33	868.87	2035.55	-0.60	78.58	0.041	-0.0058	-8.39
2	1	0.5388	31756.40	31640.28	1526.62	529.58	7.25	1105.57	-0.062	0.1455	-1.45
2	2	0.2914	19288.55	19494.19	928.05	2058.37	-0.44	94.99	0.073	-0.0033	-2.74

The first and the third cases have similar f_x and f_y values. Moreover, the second and

the fourth cases also have similar f_x and f_y values. Moreover, the focal length can be calculated with Eq. 4.1 where f_x and f_y values are focal lengths in x and y axis in pixel units, p is sensor pixel size in mm and f is focal length in mm. The calculated focal length are tabulated in Table 4.9.

$$\frac{(f_x + f_y)}{2} p = f \tag{4.1}$$

Table 4.9: Estimated Focal Lengths

Case	1	2	3	4
Focal Length	111.81 mm	63.53 mm	109.36 mm	66.90 mm

In the system, zoom lens camera is used. Due to zoom lenses, pinhole camera model is not a good option for that type of cameras. Due to the incoming beam is refracted by zoom, the focal length of the lens-camera system is not a real focal length. The estimated results also proves that because the used lens has 5-100 mm focal length range but in 1st and 3rd cases, the estimated focal lengths are above maximum limit. Although the pinhole camera model is not suitable, since the zoom level is constant during the measurement and the calibration procedure is repeated for different zoom levels, the results are logical. The re-projection of the corners points on the image proves the quality of the procedure.(Some aspects of zoom lens camera calibration)

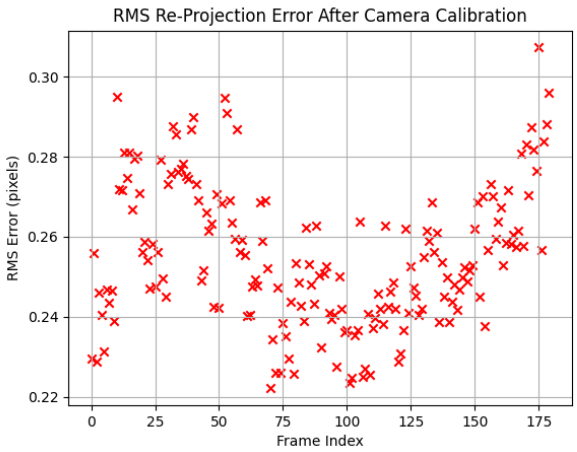

Figure 4.5: RMS Re-projection Error After Camera Calibration

After calibration, the corners points in 3D are re-projected back by using camera parameters on the image. The distance between each corners point and re-projected point are calculated for each images. The RMS error for each image are shown for Case 3 in Figure 4.5. The figure show us that all images have similar RMS value, there is no outlier in the dataset.

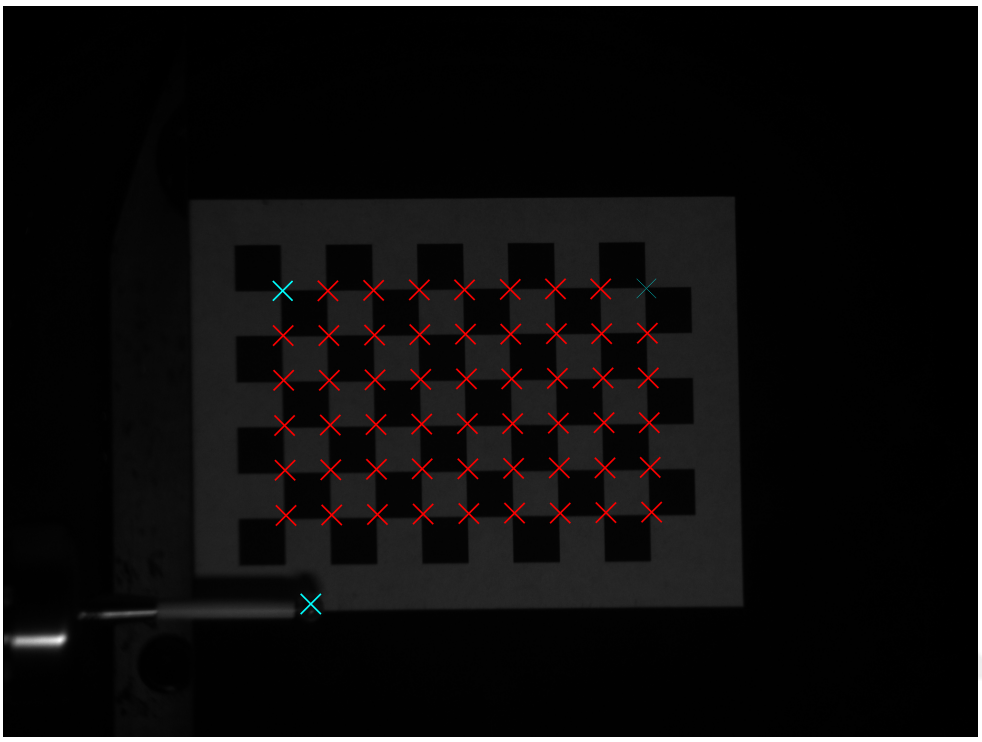

Figure 4.6: Re-projection of Corner Points on to the Image

4.4.2 Circle Fit

The probe of the CNC Machining Center is in sphere shape. To identify the center of the probe, a circle fit algorithm is used. There is a circle detector written in Java using OpenCV library. To estimate the repeatability of the detector, 500 images are taken where there is 1 second between two consecutive frames. The center point of each circle is found individually by subpixel precision. The repeatability values are tabulated in Table 4.10, where the overall performance is 1.6 pixels.

Table 4.10: Circle Fit Repeatability Values

	x	y	d
σ	1.4514	1.2009	1.6354

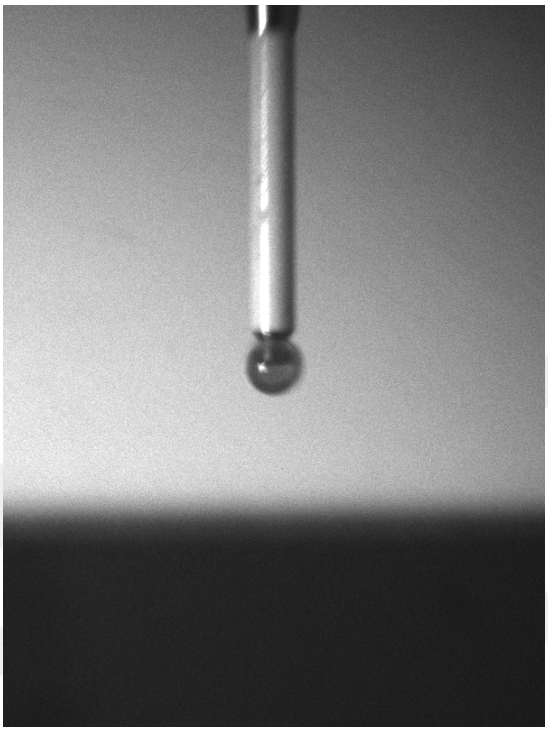

Figure 4.7: Probe of CNC Machining Center

Table 4.11: FoV and AoV Results

AoV		FoV	
x (°)	y (°)	x (mm)	y (mm)
4	3	99	74

During the studies, the depth is around 1.5 meters. The camera is calculated via Basler Lens Selector [74]. The camera sensor size, the focal length of the lens, and working distance are given as input. Angle of View (AoV) and Field of View (FoV) are calculated and tabulated in Table 4.11. 2048 pixels correspond to 99 mm, which leads to 1 pixel $\approx$ 50 μm. So, 1.6 pixels leads to 80 μm. In Figure 4.7, one of the

probe images is shown.

4.4.3 Individual Gimbal Calibration Results

In this section, the rotation and the orientation of each gimbal with respect to the **W** are estimated. The CS of the CNC Machining Center is assumed as **W**. Another CS is assumed at the left-top inner corner of the calibration pattern. The former step is making these two CSs parallel. The calibration target is fixed by clamps onto CNC Machining Center. The angles between **W** and the pattern are corrected by rotating the CNC Machining Center in two rotation axes. The other axis is measured by the dial indicator shown in Figure 4.8. As the CNC Machine head moves in the x-direction, the end point of the dial indicator measures the contact points in μm precision. As a result, the rotation between two CSs is set to zero.

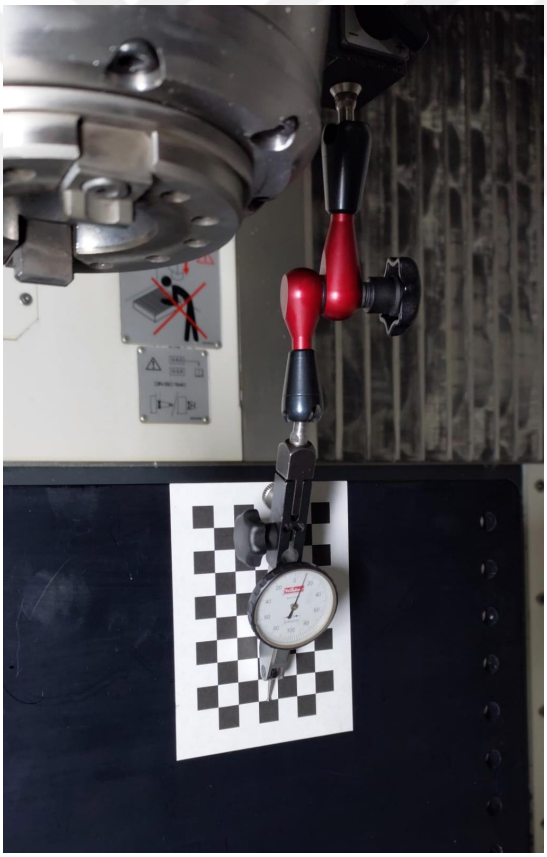

Figure 4.8: The Dial Gauge with Calibration Pattern

The latter step is arranging the center of CSs. In the z-axis (depth), simply the probe

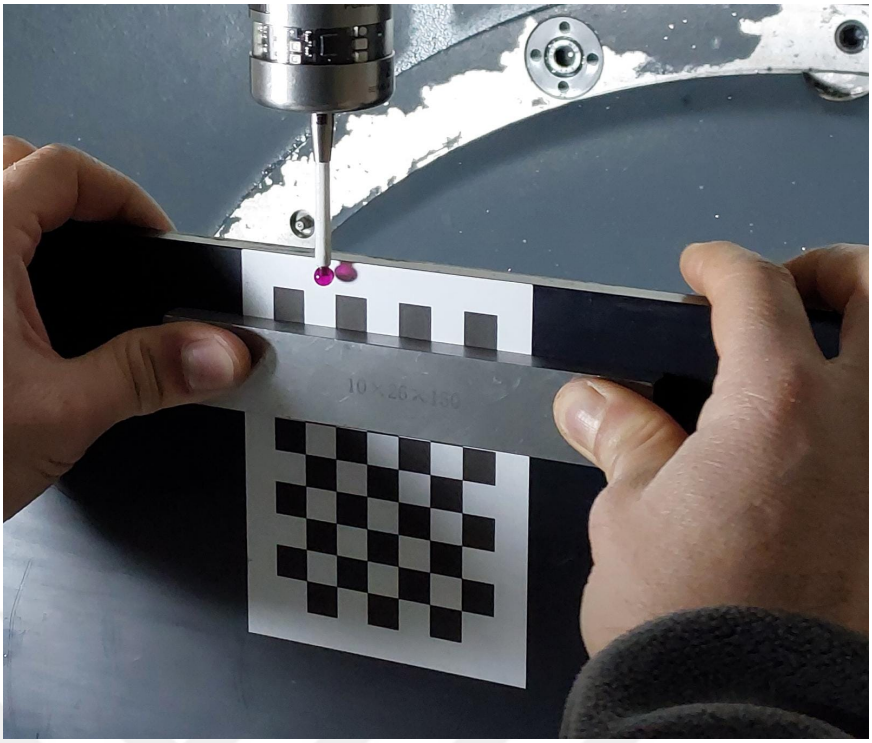

Figure 4.9: Johansson Gauge with Calibration Pattern

is touched to the calibration pattern and this x-y plane is set as $z = 0$. The chessboard pattern is stuck on a surface, so there might be an angle between chessboard CS and top side of the workpiece. The top or side surface of the workpiece can not be used for measurement purposes. The calibration of the other two axes is completed using an ultra-precise Johansson gauges block, which has been precision ground. The CNC operator holds the gauge block on the upper row of the chessboard in Figure 4.9. This line is assumed as $x = 0$, and the probe is touched to gauge the block. A similar process is repeated for the other axis. As a result of the two-step process, the center of chessboard CS and CNC machine CS are coincident at the top-left inner corner of the chessboard.

The gimbal and the camera arranged in a way that all chessboard can be seen from the camera. In each view, one set of images is taken, and the gimbal is rotating with chessboard constraints. Another image set is saved. This process was repeated with several different AZ and EL angle combinations.

Table 4.12: Individual Gimbal Calibration Results

Gimbal	Focal	Position		
		x	y	z
1	1	-110.08 ± 4.93 mm	-748.70 ± 7.47 mm	-2055.62 ± 2.28 mm
1	2	-92.00 ± 6.58 mm	-709.91 ± 1.84 mm	-1902.38 ± 0.68 mm
2	1	-73.50 ± 8.90 mm	397.15 ± 17.79 mm	-2229.24 ± 5.42 mm
2	2	-90.69 ± 4.80 mm	353.11 ± 6.58 mm	-1899.90 ± 1.28 mm
Gimbal	**Focal**	**Orientation**		
		x	y	z
1	1	-20.60 ± 0.22 °	36.11 ± 0.22 °	-20.53 ± 0.44 °
1	2	-21.20 ± 0.10 °	35.19 ± 0.20 °	-20.68 ± 0.36 °
2	1	-10.31 ± 0.94 °	40.17 ± 0.60 °	7.48 ± 0.65 °
2	2	-16.11 ± 0.18 °	40.32 ± 0.11 °	3.81 ± 0.32 °

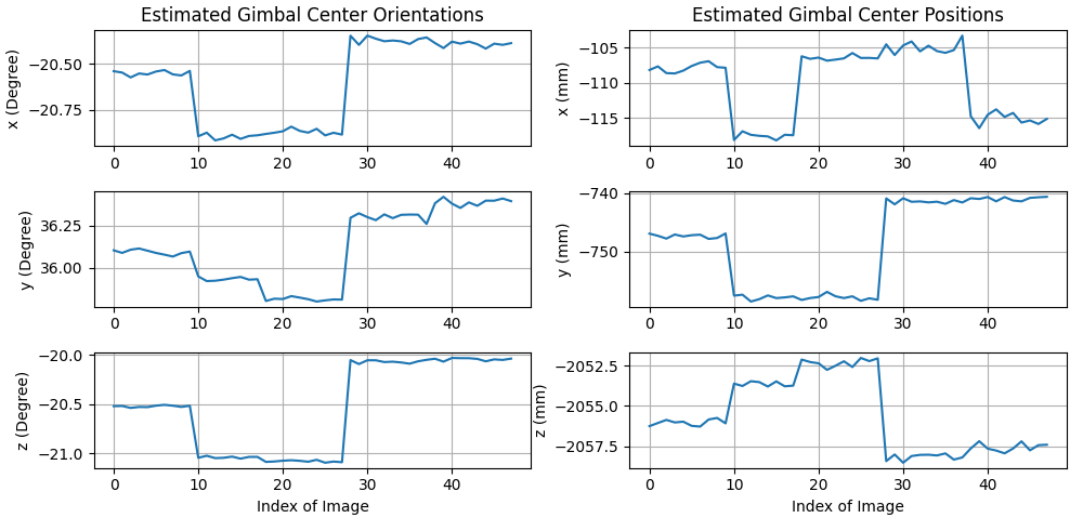

Figure 4.10: Position and Orientation of Gimbal 1 for Focal 1

In each image, position parameters are estimated via **solvePnp** function. The relation between **G** and **W** is estimated using recorded position, AZ, and EL angles. 50 images are used to calculate the center of each gimbal. The calibration results are tabulated in mean ± standard deviation form in Table 4.12. The gimbal orientations are in 321-

Euler angle form. The results show us that the standard deviation of the orientation is much smaller than the position. The gimbal centers are not as accurate as the orientation of the gimbal.

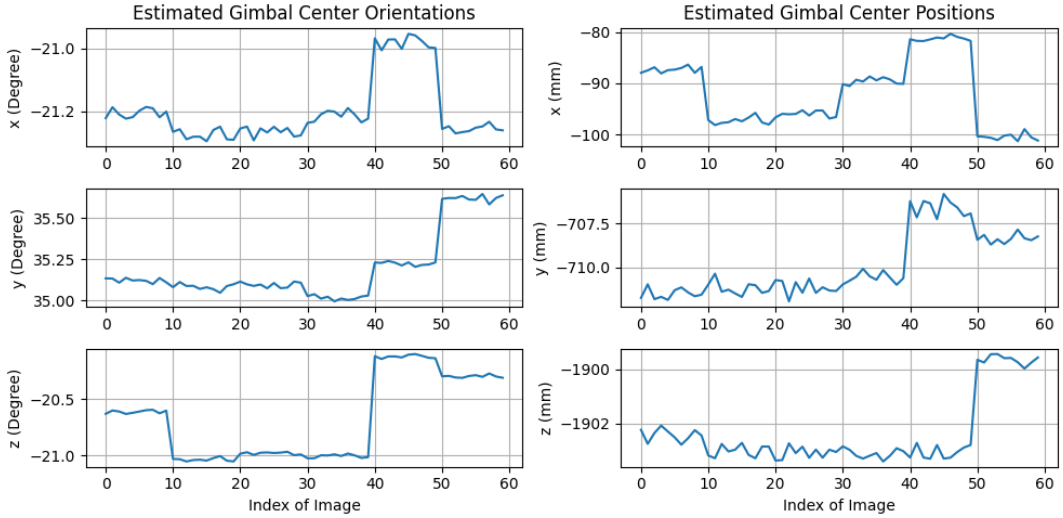

Figure 4.11: Position and Orientation of Gimbal 1 for Focal 2

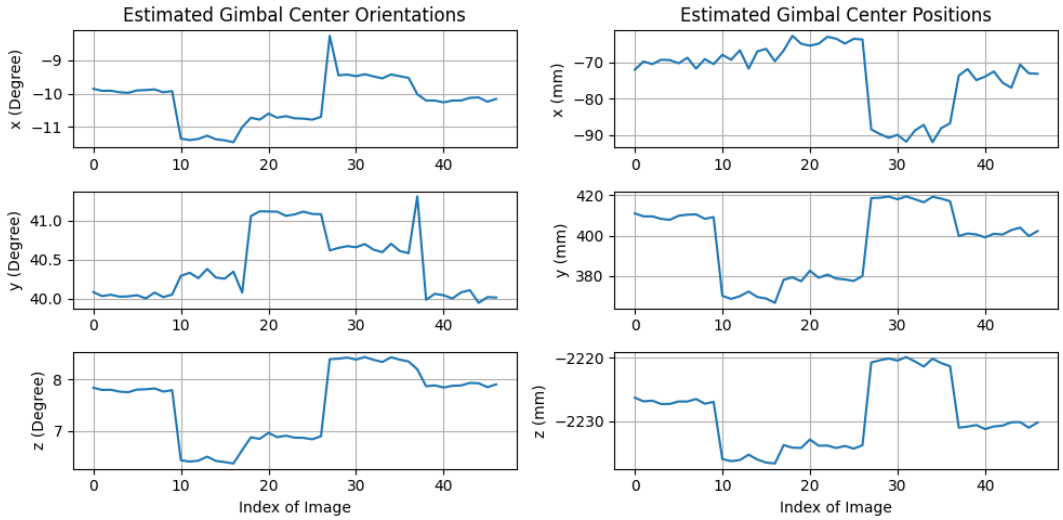

Figure 4.12: Position and Orientation of Gimbal 2 for Focal 1

In Section 4.4.1, it is mentioned that using the pinhole model causes the unreal focal length and extrinsic parameters. This phenomenon causes the estimated gimbal centers not to be the same for different zoom and focal lengths given in Table 4.12. Estimated Gimbal poses are given in Tables 4.10, 4.11, 4.12, and 4.13 for each case

separately. Although the position values are not real, there are convenient for that specific zoom level.

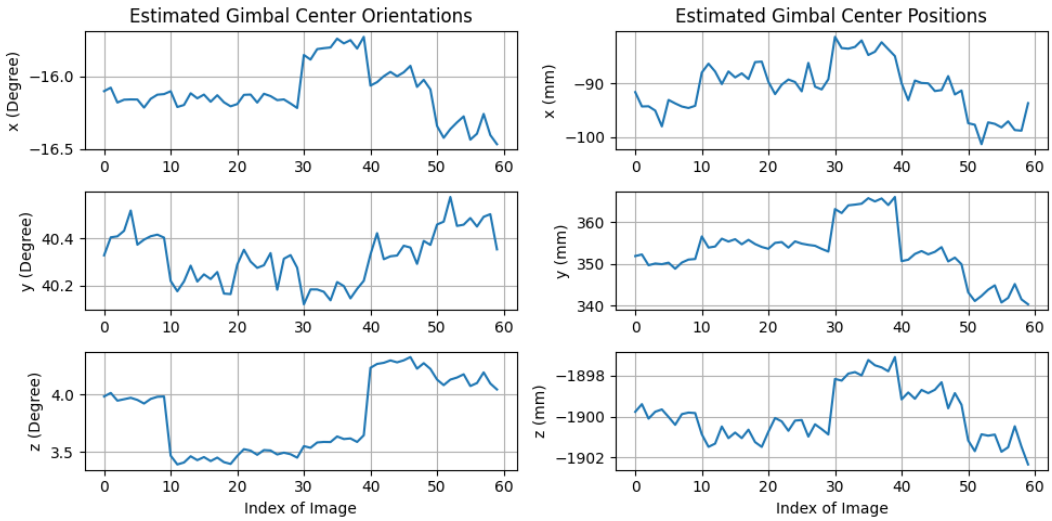

Figure 4.13: Position and Orientation of Gimbal 2 for Focal 2

4.4.4 Overall System Performance

The image given in Figure 4.16 shows the measurement system with two gimbals, the calibration pattern, and CNC Machining Center. Both gimbals track the probe by rotating in their AZ and EL axes. The images are taken from each gimbal at the same view. Those images are post-processed, and measurement point is found and compared with the real results taken from CNC Machining Center. The probe point is found on the image by the Circle Fit algorithm expressed in Section 4.4.2. The formulations, expressed in Section 3.5 and validated in Section 4.2, are used with gimbal centers estimated in Section 4.4.3 to obtain the estimated position of the center of the probe. The position of 50 measurement points is estimated, and measurement errors are tabulated in Table 4.13. X, Y and Z columns give the statistical value of the difference between actual and measured values. In the last column, the distance between actual and measured values is calculated, and statistical values of those are given.

Table 4.13: Overall Measurement Results

Focal		x	y	z	d
1	μ	-1.7612 mm	-0.1848 mm	0.1642 mm	2.9677 mm
	σ	2.762 mm	0.4029 mm	0.7982 mm	1.6684 mm
2	μ	-0.4218 mm	0.1495 mm	0.1014 mm	2.7625 mm
	σ	1.9815 mm	1.8733 mm	1.0391 mm	1.0463 mm

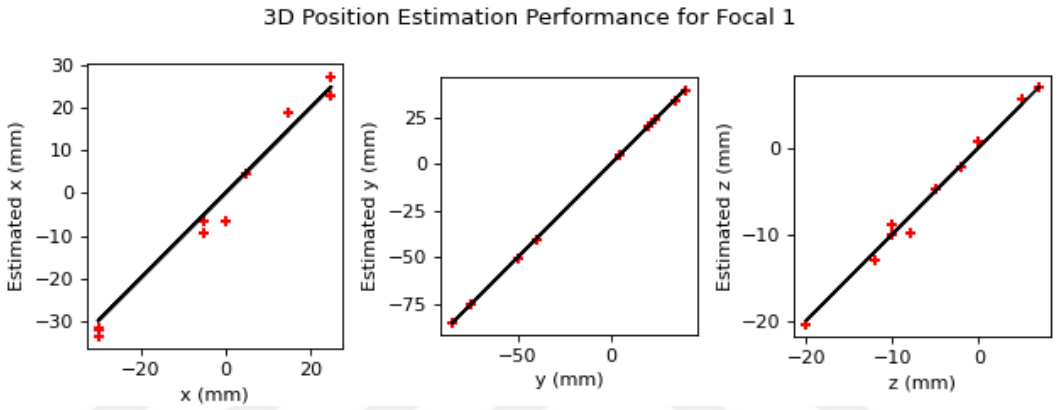

Figure 4.14: Estimated Position vs. Real Position for Focal 1

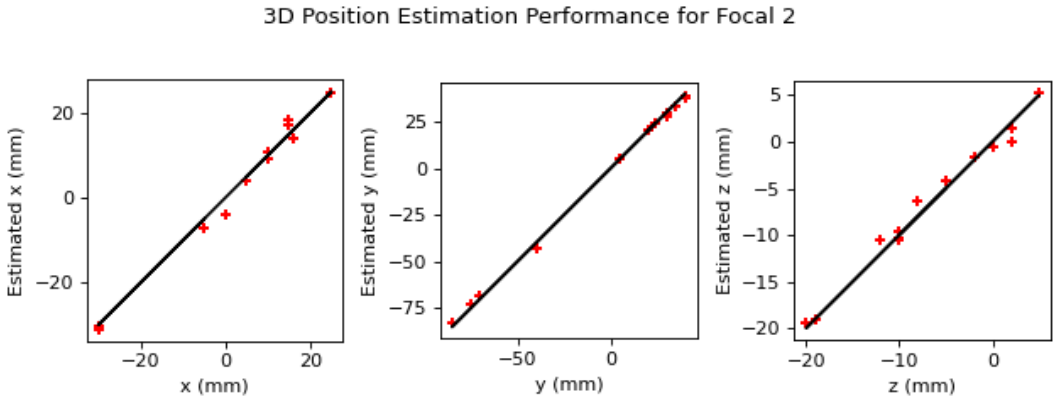

Figure 4.15: Estimated Position vs. Real Position for Focal 2

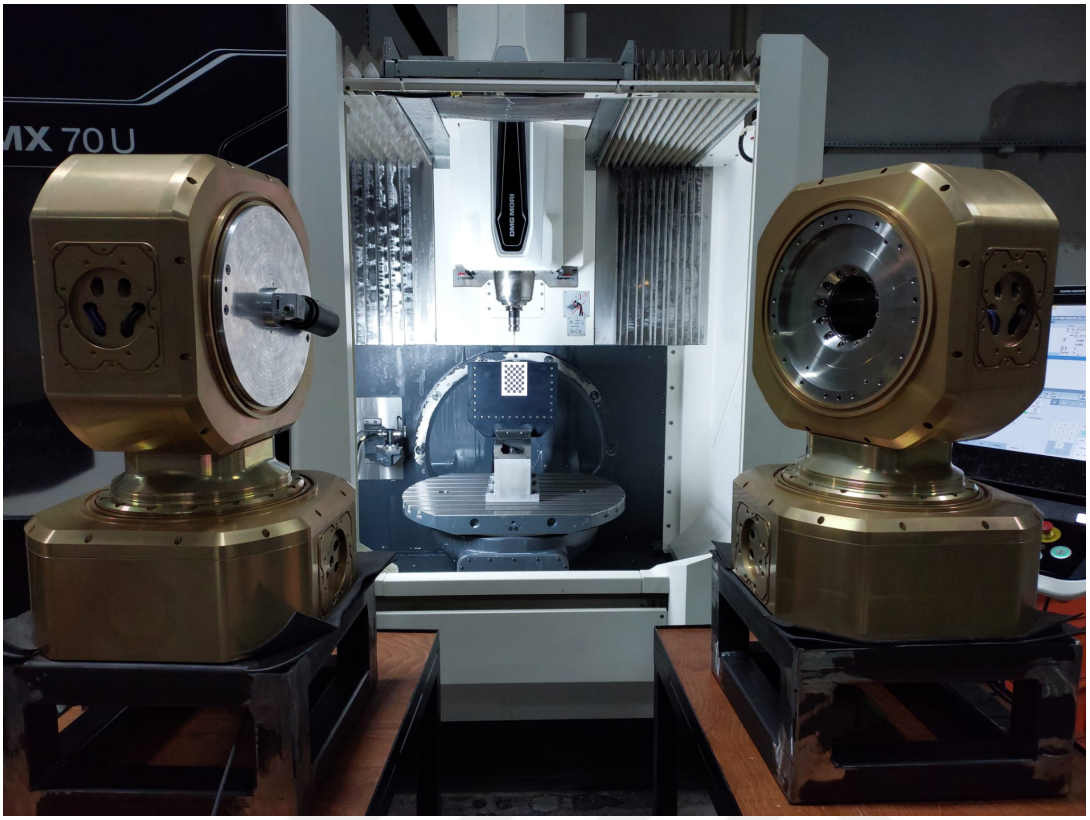

Figure 4.16: Overall System

4.5 Discussion

In this thesis, a vision-based positioning system is explained. The results shown in Table 4.13 prove that the suggested system can be used for positioning applications. The final results are similar to up-to-date stereo-vision applications.

The suggested system has several advantages and disadvantages compared to state-of-art alternative systems. The most accurate positing system is Laser Tracker Systems (LTS), discussed in Section 2.3.8. The suggested system is not cheap because of ultra-high precise gimbal systems. Using cheaper gimbals makes this system much cheaper compared to commercially available LTS systems. Considering the optics to be used in the LTS and the required positioning accuracy for the optics, the vision-based system has ease of assembly and calibration.

This thesis proves that the suggested vision-based system can be used as a concept and can be developed further. When the tracked object is moving faster, the frame

rate of the camera should be increased. For that type of scenario, LTS might be the better alternative.

The vision applications are effected from lighting and illumination. With the position of the camera and the light source, there is reflected light on the calibration target. This reflected light decreases the performance of the corner detection algorithm. Instead of a traditional chessboard pattern, an array of LED lights can be used where a coordinate measurement machine can measure each LED position. The LED pattern may decrease the undesirable effect of lighting.

The calibration routine of the individual can be changed. In the calibration procedure, firstly, the camera images are used to find the camera position, then encoder values are used to calculate the gimbal center. This means that the initial measurement is done using the camera. After that, encoder angles are used. Since the accuracy and the precision of the encoders are better than the cameras, using encoder measurements can be increased the overall system performance. The **calibrateCamera** function of OpenCV should be re-implemented by putting angular constraints between frames. The optimization problem is assumed as non-linear and we shall apply bundle block adjustment.

The measurement procedure be changed to the real-time and a Kalman filter can be implemented with state space estimation.

The resolution of the camera and the quality of the lens can be increased. As the camera resolution increases, the object detection accuracy will increase. There are low-distortion lenses for vision applications. In the suggested system, there is an adjustable lens which is not low distortion. High quality vision equipments may increase the performance of the overall system.

The detected object is the probe of the CNC Machining Center. Since the center of the probe is positioned precisely with CNC Center, the position of the probe is known accurately. On the other hand, the probe is red and half-transparent. For object detection, this probe is not a suitable object. We could have used brighter or more detectable object to increase object detection performance. This may increase the overall system performance.

CHAPTER 5

CONCLUSIONS

An alternative vision-based positioning system is proposed in this thesis. Two gimbals and eccentric cameras are used to measure the object by offline methods. A calibration procedure is developed to establish a common reference frame and estimate the center of gimbals. A python-based simulation is implemented for error budgeting. The real-world system performance is measured in a precise CNC Machining Center, which is assumed to be as ground truth. The post-processed results are obtained, and overall system performance is found.

In the literature, camera-based systems are configured with two different geometry. First of all, in stereo vision applications, two cameras in the system are not moving with respect to each other. On the other hand, when a gimbal is used for vision application, the camera center and the rotation axis of the gimbal intersect. This positioning eliminates all possible parallax effects and makes the calculations easier.

The results given in Chapter 4 show that the total system performance is not as good as what the LTS system can achieve. This study proves that it is possible to use two eccentric cameras for precise positioning applications as a concept. Also, it is shown that the eccentric camera, which is mounted on a gimbal, calibrated to be used as an angular measurement unit.

5.1 Future Work

The system performance is measured by the offline image processing method. To use this study in a real process, first of all, the calibration routine and measurement

should be changed in a way that the system can measure the position of the object of interest in real-time. Moreover, the controller of the system should be connected to the controller of the robot and fed a compensation value to the corresponding register.

The overall system can be used to measure a robotic machining arm. Moreover, the system controller can connect to the controller of the robot arm and send compensation commands. The compensated system performance can be measured by LTS or another sensor to observe the affect of the correction.

In an ideal system, there will be some measurement gimbals, and the tooltip will move between those measurement stations. The area will be around 5x5 meters. The Depth of Field of the suggested system is not enough for this range. A motorized adjustable lens can be used in front of the camera so that the depth of field can be changed by adjusting the lens's position during the tooltip's movement. On the other hand, adjusting the focal length changes the camera matrix and the distortion parameters mentioned in Section 2.5, so those parameters can be estimated online during object tracking.

The Gimbal configuration can be modified so that the camera can be placed at the rotation center of the gimbal. This process decreases the assumptions and the complexity of the individual gimbal calibration. As a result, as the uncertainty decreases, it is expected that the total system precision also increases.

REFERENCES

[1] A. Karim and A. Verl, "Challenges and obstacles in robot-machining," 2013.

[2] W. Ji and L. Wang, "Industrial robotic machining: a review," *International Journal of Advanced Manufacturing Technology*, vol. 103, 2019.

[3] E. Appleton and W. DJ, *Industrial Robot Applications.* HALSTED PRESS, 1987.

[4] J. Pandremenos, C. Doukas, P. Stavropoulos, and G. Chryssolouris, "Machining with robots: A critical review," 2011.

[5] L. Yuan, S. Sun, Z. Pan, D. Ding, O. Gienke, and W. Li, "Mode coupling chatter suppression for robotic machining using semi-active magnetorheological elastomers absorber," *Mechanical Systems and Signal Processing*, vol. 117, 2019.

[6] J. Chen and L. M. Chao, "Positioning error analysis for robot manipulators with all rotary joints," *IEEE Journal on Robotics and Automation*, vol. 3, 1987.

[7] R. P. Judd and A. B. Knasinski, "A technique to calibrate industrial robots with experimental verification," 1990.

[8] C. Lehmann, M. Halbauer, D. Euhus, and D. Overbeck, "Milling with industrial robots: Strategies to reduce and compensate process force induced accuracy influences," 2012.

[9] A. Brunete, E. Gambao, J. Koskinen, T. Heikkilä, K. B. Kaldestad, I. Tyapin, G. Hovland, D. Surdilovic, M. Hernando, A. Bottero, and S. Anton, "Hard material small-batch industrial machining robot," *Robotics and Computer-Integrated Manufacturing*, vol. 54, 2018.

[10] C. Qin, J. Tao, M. Wang, and C. Liu, "A novel approach for the acquisition of vibration signals of the end effector in robotic drilling," 2016.

[11] J. R. D. Posada, U. Schneider, S. Pidan, M. Geravand, P. Stelzer, and A. Verl, "High accurate robotic drilling with external sensor and compliance model-based compensation," vol. 2016-June, 2016.

[12] S. Kothe, S. P. V. Stürmer, H. C. Schmidt, C. Boehlmann, J. Wollnack, and W. Hintze, "Accuracy analysis and error source identification for optimization of robot based machining systems for aerospace production," vol. 2016-October, 2016.

[13] C. Moeller, H. C. Schmidt, P. Koch, C. Boehlmann, S. Kothe, J. Wollnack, and W. Hintze, "Real time pose control of an industrial robotic system for machining of large scale components in aerospace industry using laser tracker system," *SAE International Journal of Aerospace*, vol. 10, 2017.

[14] Localization, *Oxford Learners Dictionaries*. Oxford University Press, 2022.

[15] P. D. Groves, *Principles of GNSS Inertial and Multi-Sensor Integrated Navigation Systems - GNSS Technology and Applications*. 2008.

[16] Z. G. Jun, L. Xin, X. Z. Long, and L. H. Chao, "Weighted least square localization algorithm based on rssi values," 2016.

[17] M. Wax, J. Sheinvald, and A. J. Weiss, "Detection and localization in colored noise via generalized least squares," *IEEE Transactions on Signal Processing*, vol. 44, 1996.

[18] D. H. Stojanović and N. M. Stojanović, "Indoor localization and tracking: Methods, technologies and research challenges," *Facta Universitatis, Series: Automatic Control and Robotics*, vol. 13, 2014.

[19] W. Gao, S. W. Kim, H. Bosse, H. Haitjema, Y. L. Chen, X. D. Lu, W. Knapp, A. Weckenmann, W. T. Estler, and H. Kunzmann, "Measurement technologies for precision positioning," *CIRP Annals - Manufacturing Technology*, vol. 64, 2015.

[20] A. Norrdine, "An algebraic solution to the multilateration problem," *2012 Int. Conf. on Indoor Positioning and Indoor Navigation (IPIN)*, 2012.

[21] C. Sertatil, M. A. Altinkaya, and K. Raoof, “A novel acoustic indoor localization system employing cdma,” *Digital Signal Processing: A Review Journal*, vol. 22, 2012.

[22] T. Takatsuji, Y. Koseki, M. Goto, T. Kurosawa, and Y. Tanimura, “Laser-tracking interferometer system based on trilateration and a restriction on the position of its laser trackers,” vol. 3479, 1998.

[23] A. Yassin, Y. Nasser, M. Awad, A. Al-Dubai, R. Liu, C. Yuen, R. Raulefs, and E. Aboutanios, “Recent advances in indoor localization: A survey on theoretical approaches and applications,” 2017.

[24] D. Zhang, F. Xia, Z. Yang, L. Yao, and W. Zhao, “Localization technologies for indoor human tracking,” 2010.

[25] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, “Locating the nodes: Cooperative localization in wireless sensor networks,” *IEEE Signal Processing Magazine*, vol. 22, 2005.

[26] J. Yang and Y. Chen, “Indoor localization using improved rss-based lateration methods,” 2009.

[27] G. H. Kaplan, “Angles-only navigation: Position and velocity solution from absolute triangulation,” *Navigation, Journal of the Institute of Navigation*, vol. 58, 2011.

[28] N. Bowditch, *The new American practical navigator*. 1802.

[29] U. Steinhoff and B. Schiele, “Dead reckoning from the pocket - an experimental study,” 2010.

[30] M. Brossard, A. Barrau, and S. Bonnabel, “Ai-imu dead-reckoning,” *IEEE Transactions on Intelligent Vehicles*, 2020.

[31] A. Yassine, Y. Nasser, M. Awad, and B. Uguen, “Hybrid positioning data fusion in heterogeneous networks with critical hearability,” *Eurasip Journal on Wireless Communications and Networking*, vol. 2014, 2014.

[32] S. Venkatraman and J. Caffery, “Hybrid toa/aoa techniques for mobile location in non-line-of-sight environments,” vol. 1, 2004.

[33] C. Li and Z. Weihua, "Hybrid tdoa/aoa mobile user location for wideband cdma cellular systems," *IEEE Transactions on Wireless Communications*, vol. 1, 2002.

[34] D. A. Maisano, J. Jamshidi, F. Franceschini, P. G. Maropoulos, L. Mastrogiacomo, A. R. Mileham, and G. W. Owen, "A comparison of two distributed large-volume measurement systems: The mobile spatial co-ordinate measuring system and the indoor global positioning system," vol. 223, 2009.

[35] I. C. Society, "Ieee standard 802.11: Wireless lan medium access control (mac) and physical layer (phy) specifications ieee computer society," 2012.

[36] F. Evennou and F. Marx, "Advanced integration of wifi and inertial navigation systems for indoor mobile positioning," *Eurasip Journal on Applied Signal Processing*, vol. 2006, 2006.

[37] C. Yang and H. R. Shao, "Wifi-based indoor positioning," *IEEE Communications Magazine*, vol. 53, 2015.

[38] A. Carullo and M. Parvis, "An ultrasonic sensor for distance measurement in automotive applications," *IEEE Sensors Journal*, vol. 1, 2001.

[39] J. R. Gonzalez and C. J. Bleakley, "High-precision robust broadband ultrasonic location and orientation estimation," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, 2009.

[40] C. Randell and H. Muller, "Low cost indoor positioning system," vol. 2201, 2001.

[41] J. Wang and D. Katabi, "Dude, where's my card? rfid positioning that works with multipath and non-line of sight," vol. 43, 2013.

[42] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "Landmarc: Indoor location sensing using active rfid," 2003.

[43] Bluetooth®Technology, "Bluetooth technology overview." `http://www.bluetooth.com/`. Accessed: 2022-01-05.

[44] C. Zhou, J. Yuan, H. Liu, and J. Qiu, "Bluetooth indoor positioning based on rssi and kalman filter," *Wireless Personal Communications*, vol. 96, 2017.

[45] G. Li, E. Geng, Z. Ye, Y. Xu, J. Lin, and Y. Pang, "Indoor positioning algorithm based on the improved rssi distance model," *Sensors (Switzerland)*, vol. 18, 2018.

[46] A. Orriordan, T. Newe, G. Dooly, and D. Toal, "Stereo vision sensing: Review of existing systems," vol. 2018-December, 2019.

[47] H. J. Chien and R. Klette, "Substantial improvement of stereo visual odometry by multi-path feature tracking," vol. 2017-December, 2018.

[48] S. Y. Park and M. Subbarao, "A multiview 3d modeling system based on stereo vision techniques," *Machine Vision and Applications*, vol. 16, 2005.

[49] W. yu, M. Yao, and B. xu, "3-d surface reconstruction and evaluation of wrinkled fabrics by stereo vision," *Textile Research Journal*, vol. 79, 2009.

[50] D. S. Paraforos, M. Reutemann, G. Sharipov, R. Werner, and H. W. Griepentrog, "Total station data assessment using an industrial robotic arm for dynamic 3d in-field positioning with sub-centimetre accuracy," *Computers and Electronics in Agriculture*, vol. 136, 2017.

[51] Y. Luo, J. Chen, W. Xi, P. Zhao, X. Qiao, X. Deng, and Q. Liu, "Analysis of tunnel displacement accuracy with total station," *Measurement: Journal of the International Measurement Confederation*, vol. 83, 2016.

[52] B. Dörband, H. Müller, and H. Gross, *Handbook of Optical Systems Volume 5: Metrology of Optical Components and Systems*, vol. 5. 2012.

[53] S. Osawa, "High-performance laser tracker using an articulating mirror for the calibration of coordinate measuring machine," *Optical Engineering*, vol. 41, 2002.

[54] X. Zou, H. Zou, and J. Lu, "Virtual manipulator-based binocular stereo vision positioning system and errors modelling," *Machine Vision and Applications*, vol. 23, 2012.

[55] Gimbal, *Lexico, Powered by Oxford*. Lexico, 2022.

[56] R. J. Rajesh and C. M. Ananda, "Pso tuned pid controller for controlling camera position in uav using 2-axis gimbal," 2015.

[57] A. Altan and R. Hacıoğlu, "Model predictive control of three-axis gimbal system mounted on uav for real-time target tracking under external disturbances," *Mechanical Systems and Signal Processing*, vol. 138, 2020.

[58] Thorlabs, "Galvanometers." `https://www.thorlabs.com/navigation.cfm?guide_id=2269`. Accessed: 2022-01-05.

[59] K. Okumura, H. Oku, and M. Ishikawa, "High-speed gaze controller for millisecond-order pan/tilt camera," 2011.

[60] K. Umetsu, R. Furutnani, S. Osawa, T. Takatsuji, and T. Kurosawa, "Geometric calibration of a coordinate measuring machine using a laser tracking system," *Measurement Science and Technology*, vol. 16, 2005.

[61] J. Kannala and S. S. Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, 2006.

[62] K. M. Dawson-Howe and D. Vernon, "Simple pinhole camera calibration," *International Journal of Imaging Systems and Technology*, vol. 5, 1994.

[63] G. Bradski and A. Kaehler, *Learning OpenCV*. 2008.

[64] OpenCV, "Camera calibration and 3d reconstruction." `https://docs.opencv.org/4.x/d9/d0c/group__calib3d.html`, 2022. Accessed: 2022-01-05.

[65] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," vol. 1, 1999.

[66] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," vol. 1, 2001.

[67] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," vol. I, 2005.

[68] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, 2017.

[69] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," vol. 2016-December, 2016.

[70] N. Vandenbroucke, L. Macaire, and J.-G. Postaire, "Color pixels classification in an hybrid color space," vol. 1, pp. 176–180, 1998.

[71] M. Fiaz, A. Mahmood, and S. K. Jung, "Tracking noisy targets: A review of recent object tracking approaches," *arXiv preprint arXiv:1802.03098*, 2018.

[72] E. T. Group, "Ethercat - technology overview." `https://www.ethercat.org/en/technology.html`, 2022. Accessed: 2022-01-24.

[73] SWIG, "Executive summary." `http://www.swig.org/`. Accessed: 2022-01-05.

[74] B. the Power of Sight, "Basler - lens selector." `https://www.baslerweb.com/en/products/tools/lens-selector/`, 2022. Accessed: 2022-01-31.