Design and modelling a of bimodal camera system

Ahmad Zawawi Jamaluddin

SPIM

Thèse de Doctorat

UBFC UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ

école doctorale **sciences pour l'ingénieur et microtechniques**

UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ

Design and Modelling of a Bimodal Camera System

AHMAD ZAWAWI JAMALUDDIN

ImViA

cnrs

SPIM

Thèse de Doctorat

UBFC UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ

école doctorale **sciences pour l'ingénieur et microtechniques**

UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ

N° | X | X | X

THÈSE présentée par

AHMAD ZAWAWI JAMALUDDIN

pour obtenir le

Grade de Docteur de

l'Université Bourgogne Franche-Comté

Spécialité : **Instrumentation et Informatique de l'Image**

Design and Modelling of a Bimodal Camera System

Unité de Recherche :

Laboratoire Imagerie et Vision Artificielle(ImVia)

Soutenue publiquement le 20 Janvier 2021 devant le Jury composé de :

FRANCK MARZANI	Président du Jury	Professeur à l'Université de Bourgogne
MOHAMAD NAUFAL MOHAMAD SAAD	Rapporteur	Assoc. Professeur à l'Universiti Teknologi PETRONAS
SAMIA AINOUZ	Rapporteur	Professeur à INSA de Rouen
DAVID FOFI	Directeur de thèse Examinateur	Professeur à l'Université de Bourgogne
OLIVIER MOREL	Invité	Maître de Conférence à l'Université de Bourgogne
RALPH SEULIN	Invité	Ingenieur de Recherche Laboratoire ImVia

ACKNOWLEDGMENT

This manuscript allows me to express my sincere gratitude to my supervisors David Fofi, Olivier Morel and Ralph Seulin, for their continuous support throughout my PhD journey. Thanks to David Fofi for his patience, direction, motivation, continuous moral support and understanding, which is the primary key to maintain my working spirit to finish my research works. I am very grateful to Olivier Morel for his immense knowledge of research, including professionalism and encouragement. To Ralph Seulin for his support, helpful and relevant comments, as well as fruitful discussion and corrections.

My sincere thanks also go to the jury members who have agreed to evaluate this dissertation. I am honoured and thankful to Pr Samia Ainouz and Associate Professor Dr Mohammad Naufal Mohamad Saad for their interest in this research work and agreeing to be the thesis reporters. Thank you to Pr Franck Marzani for his acceptance to be my viva-voce examiner cums the chairman of the jury.

I want to thank the Universiti Kuala Lumpur (UniKL) and Majlis Amanah Rakyat (MARA) a Malaysian Government agency, for the research funding.

I express my appreciation to the staff, colleagues, and friends at the ImViA laboratory at IUT Le Creusot. In particular, Fabrice, Eric, Cristophe S, Desire, Olivier L, Olivie A, Raphael, Nathalie, Alice, Cristophe L, Cedric, Cansen, Mohamad, David S, Nathan P, Nathan C, Abir, Thibault, Yifei, Ashvaany, Mark etc. Thank you for your help, support, cooperation and of course, friendship during these years. Everybody worked hard and were giving such good ambience in the laboratory.

I am also grateful to all of my fellow Malaysians in France, especially who live in Besanson for giving me such a great and pleasant time throughout my stay in France. Special thanks to Abang Aiman and family who always help me during my difficult time in France.

Finally, I would like to convey my heartful appreciation to my families. To my mother Etah Binti Yit and Faridah Lebai Omar for their prayers, encouragement and blessing. My siblings for their support, motivation and wisdom. For my endless love wife,

Allahyarhamah Ismahana Binti Ismail, my children Naufal, Nina and Nadine. Thank you very much for all my friends and their families in Malaysia. I can't possibly mention all their names. It has been a long and challenging journey, with much memorable memory.

Ahmad Zawawi Jamaluddin
Le Creusot
France

Contents

Notations and Abbreviations

Notations

x	A scalar.
$\times$	Vector product.
ξ	Eccentricity.
D	Distortion.
I	Indentity matrix.
p	Point on the image plane.
P	3D point in the space.
P_s	Point on the sphere.
K	Instrinsic matrix of camera.
F	Fundamental matrix.
E	Essential matrice.
R	Rotation matrix.
t	Translation vector.
T	Rigid transformation matrix.

ABBREVIATIONS

FoV	Field of view
CdV	Champe de Vue
3D	Three-dimensional
2D	Two-dimensional
SVP	Single view point
NSVP	Non Single view point
IPOA	Interior Point Optimization Algorithm
CCD	Charge-Coupled Device.
SVD	Singular Vector Decomposition.
CNOC	Calibration Non Overlapping Camera.

1

INTRODUCTION

Nature offers numerous natural solutions tested by evolution, whereby accession can considerably enlarge the search space for designers. Hence, bio-inspired designs offer abstraction, transfer, and application of knowledge gained from biological models.

Bio-inspired design implies the transfer of analogies from the natural domain to the technical realm. In this light, bio-inspired design refers to an approach to design by analogy using distant analogues. Researchers have developed and applied several procedures for bio-inspired design. The technology pull (top-down or problem-driven) approaches start with a technical task that leads to discovering a biological analogue. As for biology push (bottom-up or solution-driven) approaches, they originate in knowledge about biological solutions and are applied to devise technical application [1] [2] [3] [4].

The life of nature has always been a reference for enhancing the quality of human life. Researchers and engineers have constantly observed nature's behaviour that may serve as an aide or to substitute human in accomplishing specific missions.

A robot refers to a machine that is built based on nature's form [5]. It is designed to function as a guide externally controlled by human or internally embedded in a computer controller. The camera is a vital component in a robotics application. Since the past five decades, many studies have embedded the vision system for numerous robotic applications and missions.

Mechatronic engineering denotes the combination of electrical, electronics, mechanical, and computer engineering domains. The term 'mechatronic' was coined in Japan in 1969 to describe an engineering discipline associated with automation and robotics, telecommunication, control system, and product engineering. Combining of various engineering fields enabled the generation of a simple, reliable, and versatile system

for the users. To date, the mechatronic system occupied a vital position in our life. The development of a mechatronic system creates scientific, ergonomic, and economic values to facilitate and enhance perfectly human life. The system imposes tasks that humans are incapable of performing, such as repetition and dangerous tasks, and tasks that demand high accuracy measurement. In the industrial sector, industrial robotics has rapidly replaced human for specific tasks, including painting, welding, printing circuit board, and labelling, which are repetitive tasks that require high precision and speed.

Both underwater and space explorations used mobile robot-guided manually for movement, in which the manipulator can be controlled from a distance for inaccessible mission by a human. The bio-inspired approach facilitated seabed localisation and mapping [6]. A robot that is capable of performing a mission on its own is classified as an intelligent robot. The term 'autonomous robot' denoted a robot that performed a high level of autonomy without external influence or human interference. A remotely operated system is neither an intelligent nor autonomous robot, as a human operator is required to control it. Meanwhile, space and ground robots guided by humans automatically perform tasks, such as localisation, object detection, and obstacle avoidance.

Once a robot changes its place, the environment also changed, and the robot has to adapt to the new environment. This ability is the question of perception, which reflects the robot's ability to identify and integrate the capability of the robot. It also adapts the robot to perceive the environment. An intelligent robot reckons the environment with its sensors and camera. However, an autonomous robot has a perception issue, such as for the purpose of localisation and cartography of the environment.

Human being has five senses to study and adapt to the environment. A robot equipped with sensors, range-finder, lidar, sonar, and the camera is called a smart or an intelligent robot. The purpose of placing sensing devices is to detect the environment condition by collecting information around it or placing a gyroscope to identify the body position relative to the environment.

For instance, a heavy or long vehicle faces more difficult when faced with the possibility of collision. In the past, the driver would take several seconds to verify and understand his surrounding environment before moving and manoeuvring the vehicle. Such verification heavily depended on the windows, mirrors, and cameras installed surrounding the vehicle. Blind-spot is another issue faced by the driver as it is harmful to other road users. Hence, a combination of several mirrors and cameras is an effective solution to

increase the driver's field of view (FoV).

Since the past decade, vision sensors have become popular due to multiple advantages. Vision is not only merely for viewing an object or the environment; but more broadly for finding, attending, and recognising the object [7] [8]. Cameras differ from other sensor types, cost-effective and offer detailed information (e.g., colour, density, movement, and three-dimensional (3D) view). One impressive instance of using a camera refers to the exploration of the planet Mars, in which the robot was equipped with cameras for visual odometry. NASA (National Aeronautics and Space Administration) had developed a vision system to be used by the next Rover explorer. This advanced camera system called Mastcam-Z is meant to explore the planet Mars [9]. Another remarkable example is the car navigation system linked with Google Street View. The system facilitates people around the world to move from one place to another.

Replicating human abilities to view things and observe the environment, especially for intelligent robotics application, has been the researcher's goal. The interconnection between the camera, computer, and actuators lead to the right decision to act. Similarly, the interconnection among eyes, brain, hands, and legs enabled human to estimate and predict before deciding. In the robotic, camera images served as inputs to act. The FoV of a camera is limited to 70°. However, a hybrid camera or omnidirectional camera solved these issues as it offered a large FoV. This large FoV fed more information to the processor, thus providing numerous advantages for various applications.

1.1/ HUMAN VISION ANATOMY

The eyes are a primary reference for developing a camera. There is a similarity in some points between the eye and a camera. The cornea reflects a window that allows light to enter the eye. Before reaching the retina, the light passes through the pupil and iris. The pupil is an open centre, which permits light to pass through. The amount of light that can enter the eye is controlled by the iris, which functions as a shutter. The iris is flexible and has the abilities to shrink and enlarge. The crystalline lens generates the focusing point on the retina. Figure1.1 shows the main components of the human eye.

Failure to focus on the shadow that falls on the retina is an eye defect, either myopic or hypermetropy. Llorente et al. (2004) [10] asserted that far-fetched or myopic occurs

when the eyepiece cannot squeeze the light as it should. Light from an infinite object defines and forms shadows in front of the retina, as shown in Figure1.2 (b). Serway et al. (2008) [11] claimed that near-sightedness or hypermetropy is an eye defect that happens when the eyepiece is not quite convex as it should. As a result, light from the object is focused and forms the shadow behind the retina, as portrayed in Figure1.2 (c).

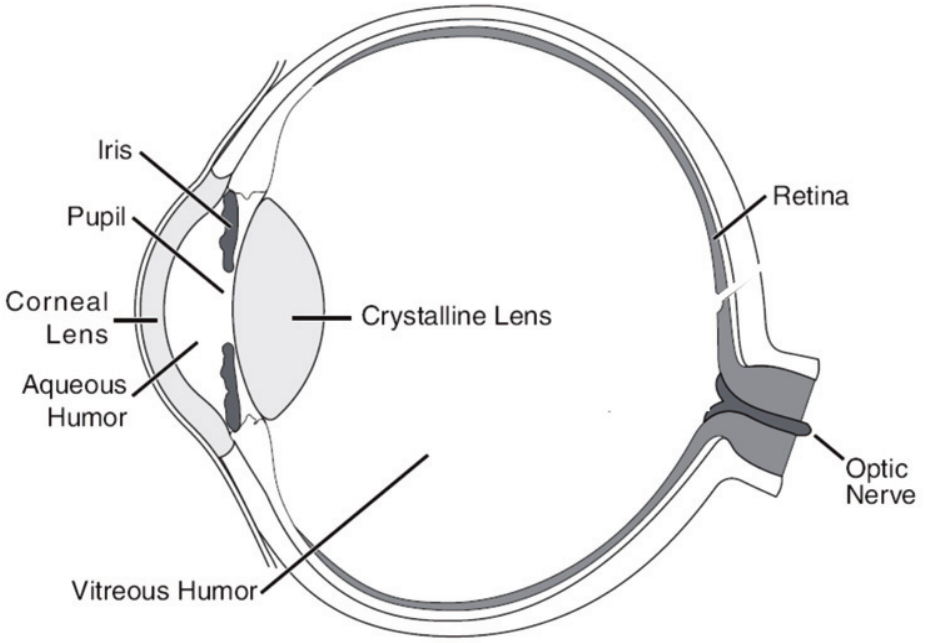


Figure 1.1: The human eye anatomy. The six main parts that form an image in the eye and sent to the brain are the cornea, pupil, iris, lens, retina, and optic nerve.

The function of the retina is similar to the film in a camera. It captures light by using cells called a photoreceptor. The photoreceptor processes the light signal with millions of nerves then transforms the light into electric impulses and sends them to the brain via the optic nerve. The two types of photoreceptors are cones and rods. The cones are less sensitive to light, particularly for day-night comparison. During daylight, the cones adapt more quickly in changing the level of light intensity. Hence, the cones are responsible for providing high-resolution vision and conferring colour vision.

The eyes move and try to continue to keep the light from the object on the fovea region. In contrary to the cones, the rods are more sensitive to light. They are insensitive to colour but confers to achromatic vision. The rods are higher in number than the cones

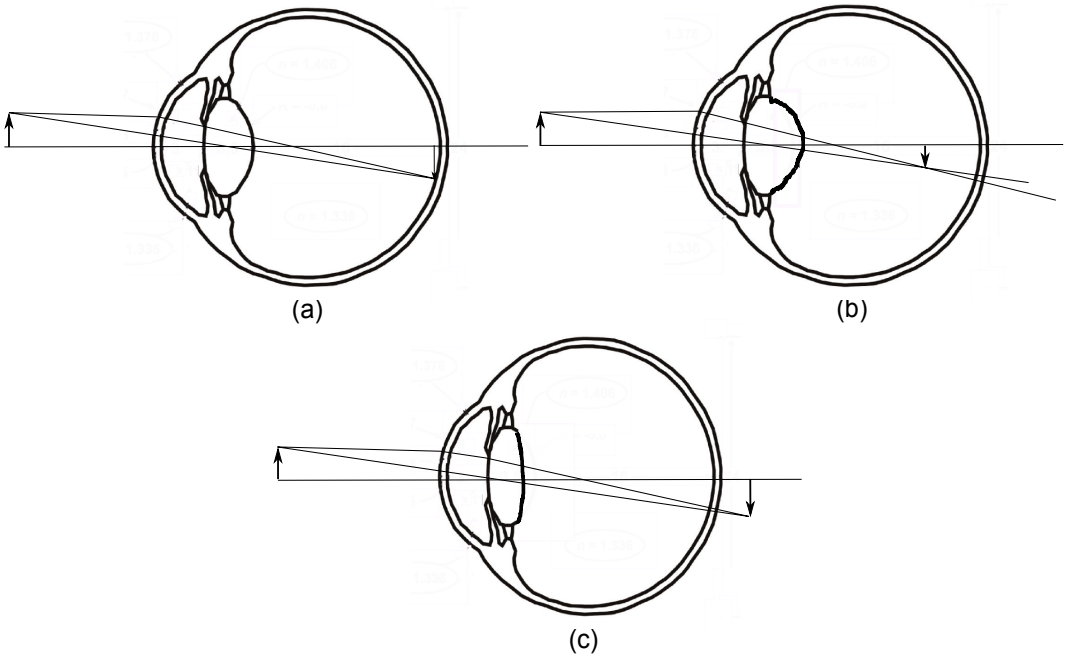


Figure 1.2: (a) Normal eyes. (b) Eye defect: Myopic (c) Eye defect: Hypermetropy.

and concentrate in the periphery region. The rods represent 95% of the total cells in the retina. It is responsible for scotopic vision or dark vision, which refers to vision with low intensity. While the cones offer a better visual resolution, the rods are better as motion detector. Figure1.3 illustrates the density of the rods and the cones in the retina.

1.2/ HUMAN VISION FIELD OF VIEW

The human FoV is composed of two monocular visions from the left and right eyes. Due to the small baseline between the left and right eyes, an overlapping region is present, and the two eyes share the same object. This region is called binocular vision or stereo vision. Healthy human eyes have FoV around 120° horizontally and vertically. The position and eyes and neck's ability to freely move helps the retina maintain the densities of light in the fovea and the peripheric region. Hence, one can observe the environment more efficiently and predict the position of an object more accurately when it moves.

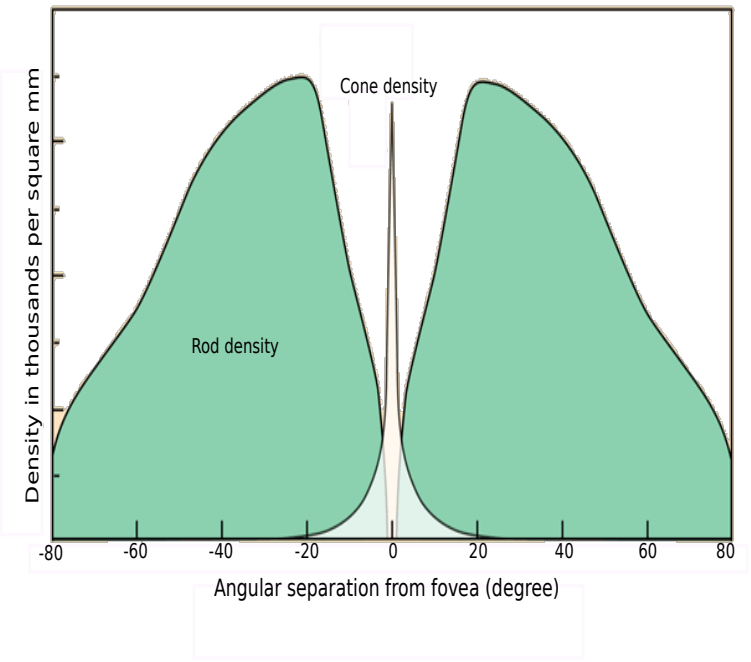


Figure 1.3: The density of rods and cones in the retina. The highest number of cones is found in the fovea centre (0°). At the centre point of the fovea, much light is pointed out to produce high-resolution vision with high colour sensitivity. Although no rod is present in the centre of the fovea, it is spread all over the retina (peripheral vision). The maximum density of rods is 20°. At this point, the density of cones is zero. This condition makes the rods responsible for low-intensity vision (dark vision) and very sensitive to movement.

1.3/ Context and Motivation

Reality and presence are the two aspects weighed in when technology is invented. Since a decade ago, robotics has become popular as it substitutes humans to simplify human life or perform dangerous or impossible jobs for a human to do. Advancement in science and technology has brought the research realm to its highest level, besides changing human life to be better. For example, the research on telecommunication undertaken by three prominent companies, namely Samsung, Apple, and Huawei, had enabled the smartphone technology in actual reality when touchscreen was only seen in science fiction movies in the 1980s. To date, the telephone is used to make or receive calls and serves as a camera, a voice recorder, and a mini-computer to surf the internet.

A camera is an artificial vision that replaces the human eyes and replaces human in the industry. A camera system is commonly integrated into a robot as a perception tool to survey the environment or carry out a specific task, such as object detection and recognition. Regardless of the autonomous or non-autonomous robot, robots

tend to suffer when performing some tasks due to limited FoV. Narrow FoV increases response time and consumes more energy as the robot needs to move around and use its actuators until it detects the target object. The perspective camera is typically used in robots because it captures a high-resolution image, has minor image deformation due to uniform lens magnification, and low cost. A full-view camera system may be the solution to the issue of limited FoV. The proposed hybrid camera is not only used for visualisation, but it is also capable of visualising vital information at the scene for image analysis, such as depth perception, 3D reconstruction, and detection of objects.

Observing nature implies studying the behaviour of animals or insects living in a challenging environment [12]. The ability of the animals and insects to recognise their environment and know their way home are interesting case studies [13] [14] [15] [16] [17]. There are two animal categories in wildlife with different vision capability – prey and predator. Prey animals have eyes on the lateral sides of their head. The position of the eyes increases the FoV and covers a large area of the environment. Such vision capability can detect the approach of predators and other dangerous animals. These animals do not have a high-resolution vision or unique eyes with large FoV because their main purposes are detecting any movement and running away to a safer place. The other animal category is the predator. Their intention is for hunting using their vision capability. The position of the eyes is mostly on the anterior of their head. The location of the eyes generates a more precise vision to estimate the location of the prey. The prey animals have a view with more massive depth information despite their limited FoV. For instance, a jumping spider has both vision abilities. This tiny creature has eyes at the anterior and the side of its head. Hence, the jumping spider acts as a predator while simultaneously observing the environment during hunting.

The eye structure of animals and insects depends on the environment where they live and their daily activities. Insects have two types of eyes. Their simple eyes consist of one lens, while their compound eyes are hemisphere in shape and have several lenses or thousand of light-sensitive cells [169] [170]. Due to its shape, the compound eye has more than 180° FoV. Figure 1.4 presents the compound eyes of an insect.

Studies on bio-inspiration are increasingly essential, along with the initiation of autonomous robotics to replace operating systems [171] [172]. Marefat et al., [17] developed an omnidirectional vision system by referring to an arthropods insect's compound eye that has a wide FoV. However, compound eyes are beneficial in detecting

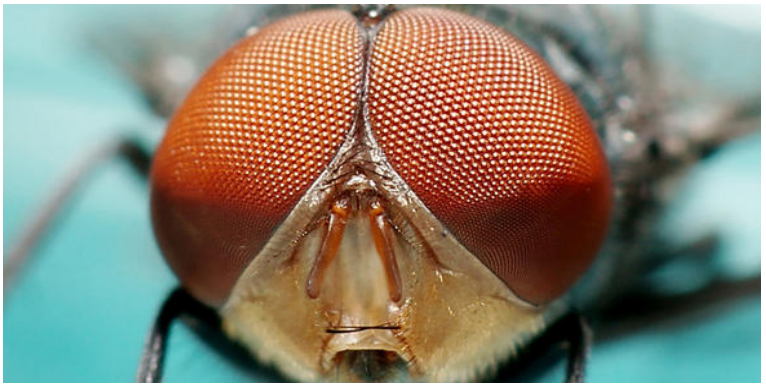


Figure 1.4: Compound eyes of a fly which is similar to omnidirectional sensor

fast motion, its low-resolution categories, this type of eyes behind a vertebrate's eye.

Lee et al. [173] introduced a compound-like vision system and enhanced the sensor's resolution. Meanwhile, Maddern et al. [174] built a camera system by taking advantage of the compound eye for six degrees of freedom ego-motion estimation.

The fabrication of a wide-FoV camera implemented in computer vision technology and integrated with a robotic system highlights this study's novelty. To date, the camera is not only used for visualisation but also for manoeuvring and analysing the current state of the robot [175].

Bio-inspiration has always been the primary resource of the design and development process. The proposed system derived from the observation of two animal categories; prey and predator. Both the eyes' geometry and location indicate if an animal is a hunter (predator) or the hunted (prey).

Prey: The prey animals vary from the smallest insect to the enormous moose. A prey or a grazing type animal is always the hunted animal. Most prey animals are herbivores (plant-eater), while some are omnivores (plant and meat-eater). The eyes of prey animal are generally located at the side of their head to give a panoramic view of their surrounding. The structure of the big ball eye with horizontal-slit pupils increases the FoV. However, prey animals, such as horse and sheep, have a blind spot area in front of the head. Therefore, a horse often moves its head left and right to observe the ground with a minimal blind spot. While grazing, the horizontal-slit pupil is maintained horizontal and parallel to the ground to detect the predator's presence. This type of animal's first critical visual is to detect approaching animals with a minimal blind spot. The second critical visual, when a predator is detected, is for the prey animal to know where it is running.

Predator: A predator hunts for food. The two eyes of the predator are located in front of their face. Their side-by-side eyes' location produces a binocular view (stereovision) for predators to aim the target accurately. The circular or vertical-slit of the predator's pupil provides depth information, FoV, and the ability to estimate the distance to prey. Figure 1.5 and 1.6 display the different locations and slits of eyes.

Figure 1.5: The vertebrate eye with different eyes location

Figure 1.6: The different eye slits of prey and predator. The horse (prey) has a horizontal slit, while the tiger and crocodile (predator) has a vertical slit.

The combination of prey and predator views offers plenty of advantages for robotic application. The proposed system provides a full surrounding view with stereo vision view. Such abilities can be found in a jumping spider.

The study of jumping spider is rather impressive. The jumping spider has four sets of eyes to see both the anterior and the surrounding areas. The four sets of eyes enabled the jumping spider to detect a range of colours, recognise ultraviolet (UV) light, perceive the world in 3D, and estimate distance using motion parallax. Figure 1.7 display the eyes of jumping spider. The four sets of eyes are as follows:

- Posterior median eyes - motion detector and sensitive to the blue and UV light.
- Posterior lateral eyes (PLE) - wide angle motion detector which can detect motion surround the body.
- Anterior laterals eyes (ALE) - visual-based vision system. The ALE is a second eye of the spider.
- Anterior median eyes (AME) - primary eyes. The AME is the principle and refers to movable eyes than the above which are fixed.

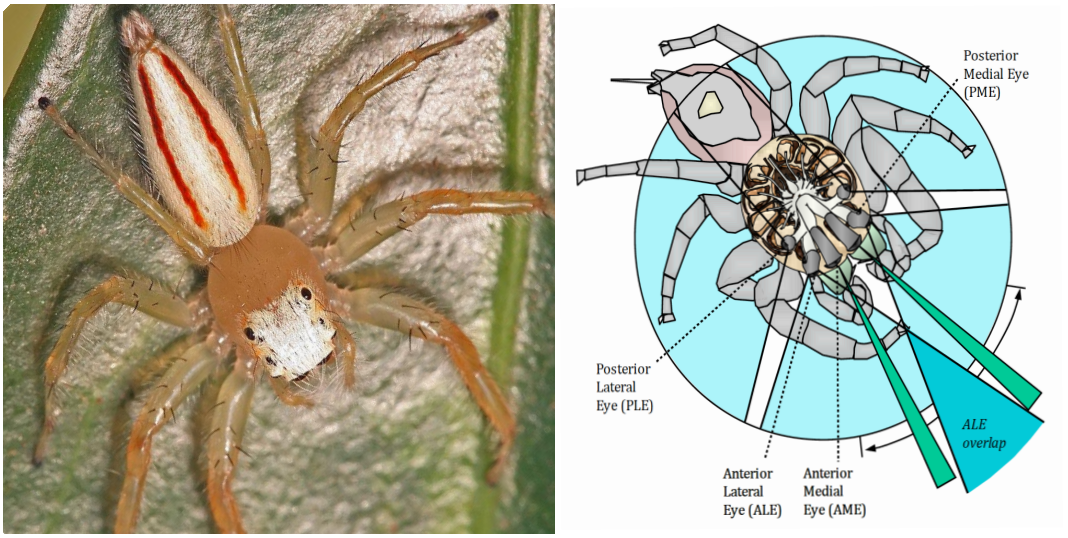


Figure 1.7: The two sets of eyes used by the jumping spider to observe its surrounding area and the other two eyes for hunting.

A hybrid camera is the replication of the animal or insect vision capability to see the world. The combination of predator and prey types of eyes gave a perfect vision system; the scene details are analysed using predator vision, and prey vision is used for surveying and observing. An omnidirectional camera referred to a hybrid camera that offers a wide FoV of a scene. The idea is to combine several cameras to produce a full-view vision for environment observation and high-resolution windows to capture and estimate the scene details, such as depth information.

This present study probed into the ability of each camera (perspective and omnidirectional) to perform 360° FoV, the image quality, and the coverage area (vertically and horizontally). Rigidity was considered an essential element. The form and the structure of the proposed hybrid camera should be ergonomic and suitable for robotic applications.

The existing panoramic vision using a combination of fisheye cameras served as reference [18] [5] [19]. The proposed vision system is called an Omni-Vision camera. It enhanced and maximised the FoV with a minimum number of cameras and cost-efficient.

1.4/ Problem Statement

A hybrid vision system with wide FoV is beneficial to the security industry and for robotic applications, regardless of the autonomous or non-autonomous robot. Vision ability is vital to increase a robot's capability to navigate in unknown territory and hinder from unsafe or dangerous situations, such as boiling and cold temperatures, radiation, and unpredictable weather condition. The autonomous robot performed a challenging tasks as the camera influences its movement for manoeuvring or detecting obstacles instead of mere surveillance. Another application is the surveillance system, which is integral in restricted areas to identify and control undesired activities. Hence, the vision ability should fulfil the assigned tasks' demand, which is not limited to navigation, but also for other intricate missions such as depth information analyses, robot localisation, and 3D reconstruction.

Vision for robotic referred to the active research area, which aimed to enhance the robot-environment interaction. It is the extension of a computer vision technique for robotics application. Computer vision remained the fundamental to understand the characteristic of still image or a single fixed view of the camera. On the contrary, the robotic vision applied the computer vision technique, and it is tailored to suit the specific application or research problem. The input images for robotics vision are more dynamic and require a different perspective for each application level. The computer vision struggles to determine a suitable algorithm to ensure that the visual system does comprehend the scene and the objects within the environment. The present invention of a high-speed processor enhanced both the processing speed and the robot's cognitive process. The progressive computer vision methods became trending in robotic vision within the robotic industry.

Most of the presently available vision systems seem to fail to fulfil specific tasks due to limited FoV. Among the several solutions proposed, one refers to installing several cameras to increase coverage area, but this is cost-ineffective due to costly cameras.

Besides, synchronisation between the cameras made images handling processing tasks more complex. The appropriate specifications are vital to be considered before devising a hybrid vision system. All aspects, including form, structure, weight, rigidity, and ergonomic, of the vision system are weighed in for better fitting during implementation and to maintain system performance. In response to these circumstances, a technique to increase the FoV is studied and explored in the literature. For instance, the omnidirectional camera is a hybrid camera with a wide FoV. This type of camera is suitable for surveying purposes and robotics application.

This study developed an algorithm to combine a high-resolution predator-like vision system with a 360° FoV prey-like vision system. The resulting vision system increasing the robot's ability to view its surrounding environment at once and adapting with a high-resolution camera to analyse the details and estimate the depth perception of the scene. The combination of the two vision systems enhances the hybrid camera FoV, thus increasing the robot's efficiency. For example, in the following scenario for a search-and-found object, the robot detects an object instantly, even it located behind the robot, except if it is behind an obstacle. Detailed information of the object can be estimated by using the high-resolution camera.

A system becomes perfect if it is free from error and completes all assigned tasks without error. The system is optimised to minimise the range of errors. A novel hybrid vision technology is the desired objective to support the industry, especially one that is related to robotic application and environment surveillance. This present study is the continuity of the vast development in robotics that demands high-end computer vision technique to guide robots for exploration, navigation, and surveillance in order to comprehend a scene better.

1.5/ RESEARCH OBJECTIVES

The research objectives aimed to support the problem statement and answer the 4W and 1H research questions toward its success. Some of the research questions are pointed out and blast the idea to solve the difficulties along the research periods. Examples of questions such as " Can we produce a 360-degree FoV vertically and horizontally vision system, including a binocular view?", "How to analyse the quality of the output im-

age?", the research objectives are important for the generation of the solutions and idea.

The first objective of this study is to investigate the literature by examining and analysing current issues concerning the wide field of view camera. The study aimed to understand the camera's basic knowledge, including all the associated essential parameters on how it produces an image. The technique to increase the field of view, fusion, and combination is vital to comprehend as it guides the research's goal. It included the related mathematic approach to optimise the parameters.

The second objective is to evaluate and optimise the parameter. This point identifies the novel method to estimate the optimum result and shape the camera system to be more applicable. The evaluation is measured by both qualitative and quantitative.

The third objective is to implement the system for 3D reconstruction. This objective is also to explore this study's implications, particularly in robotics, industry, and surveillance system in which an ultra-wide field of view camera is needed.

1.6/ CONTRIBUTION OF RESEARCH

This research work has developed an algorithm based on a multi-view camera set-up. This set-up offered a full-view vision using an omnidirectional camera and a high-resolution perspective camera to gather all the scenes in details. The original idea derived from animals' behaviour on how they see the world to hunt for food and observe the environment. The biological visions of prey animals and predators served as a standard reference to the initial idea of this study.

This study combined both abilities to develop a 360° or a panorama vision system, along with stereo vision windows, to suit multiple applications. This thesis presented the development of an algorithm to unify a set of cameras. It integrated and harmonised images from both perspectives and omnidirectional cameras.
The contributions of this research work are listed in the following:

1. A compact Omni-Vision system, along with a stereo-camera, to offer plenty of information from 360° view of the environment and detailed depth information.

2. A new camera calibration method that takes the advantages of Unified Camera Model representation to outperform the other state-of-the-art methods.

3. An Interior Point Optimisation Algorithm (IPOA) based on a pure rotation matrix estimation approach is proposed to fuse several images with different distortion degrees, which offers seamless image stitching results.

4. An affine projective distortion is integrated into the perspective image before projecting onto the unit sphere, thus reducing the registration error between the matching points and enhancing the overlapping image quality..

Along the way, this research work has been presented and published at three international conferences and one national-level conference. The list of conferences is listed in the Appendices.

1.7/ THESIS ORGANIZATION

This thesis is composed of six chapters. The chapters correspond to the flow of research works, from identifying the research gap to reporting the research conclusion.

Chapter 1: Introduction: This chapter starts with the topics' general background and ends with the points related to the research topic, including the bio-inspired contributions to enhance human life. The chapter essentially focused on the anatomy of the human eyes and described the similarity with the camera. Next, the study's context and motivation are defined, including the problem statement that depicted the problem investigated in this study. This chapter ends with the contribution of this thesis.

Chapter 2: State of the Art: The discovery of a camera by famous philosophers is unfolded. The mathematical aspects of each camera model are described. It starts with the geometric projection and pinhole camera model, besides introducing the camera's basic parameters. The types of image distortions are reported in this chapter. It explains the omnidirectional types of cameras and stereo vision and the development of the resulting camera and image. Single viewpoint (SVP) and non-single viewpoint (NSVP) are explained. They involve many types of mirrors and lenses to produce omnidirectional images of the camera. Next, the omnidirectional projection model is described. The chapter ends with multiple view geometry, where fundamental and vital matrices are depicted for use in the next chapter.

Chapter 3: Proposed System: Omni-Vision: In this chapter, the proposed hybrid camera design and development are described. The design is following the problem state-

ment and relates to bio-inspired as the primary natural reference. This chapter focused on studying and matching a suitable vision sensor, minimising the number of sensors, and installing ergonomics to the robot. The limitation of the design in terms of blind-spot is explained. The chapter ends by proposing a solution to minimise the blind-spot.

Chapter 4: Multi-Camera Calibration and Fusion: This study aims to develop an algorithm to combine several cameras into a single vision system. The calibration process for all types of camera is described. The calibration included estimating intrinsic and extrinsic parameters and calculating the rigid transformation from the two images. In the fusion process, a unified spherical projection model is used. The images from all cameras are projected onto the unit-sphere. Next, the parameters and the fusion results were analysed and optimised using the proposed projection method and Interior Point Optimisation algorithm.

Chapter 5: System Implementation: The proposed system implements to construct the 3D points. The epipolar geometry of the omnidirectional camera is explained, and the matching features suited the fisheye image. The 3D points were triangulated using the linear least square triangulation method. The algorithm performance is tested with synthetic data before using an actual image from a fisheye camera. Next, the method will perform a 3D reconstruction from the Omni-Vision camera's full view.

Chapter 6: Conclusion: This thesis is concluded by discussing and summarising the necessary observation from all chapters. The chapter ends with a list of recommendations for future endeavour.

2

STATE OF THE ART

2.1/ INTRODUCTION

An image is defined as the imitation of a real object formed when the light rays originating from a particular object unite. The ray light projected the image on a screen as it sets as the image's plane. The object emitted back the light in all directions. Some of this light (represented by rays) reached the shining surface mirror and then reflected off following the reflection law.

It is pleasant to know the short history of the origin of optics before advancing to the modern optic era. The evolution of optic technology evolved during the era of ancient Egyptian. Many artefacts related to optics were discovered at the ruined ancient pyramid (ca. 1900 B.C.E), such as a mirror made from polished bronze or copper.

In the ancient Greeks, Euclid (300 B.C.E), in his book entitled *" Catoprics "*, wrote the rectilinear propagation of light and the Law of Reflection. He embraced the theories from the famous Greek philosophers, such as Pythagoras, Aristotle, and Plato, which described the nature of light. One of Plato's theories is about the bending of an object if the object immersed itself in the water.

During the Roman empire, the burning of glass was the early application of the lens. The Roman empire ruins discovered the proof and found plenty of crystal spheres, glass, and convex lens. Seneca (3 B.C.E-65 C.E), a Roman philosopher, claimed that an object is magnified if seen through a glass globe filled with water. A book entitled *" On the Burning Instrument "* written by an Arab philosopher, Ibn Sahl (940-1000C.E), studied the details of refraction.

The Chinese writing script called Mozi, is recorded as the first known scripture

writing of the camera. He wrote to explain that the intersection point (pinhole) is the inverted point that collects light. The inverted image is called the collecting point of all light through the pinhole.

After several centuries, Anthemius of Tralles, a Greek mathematician, assessed the effect of light through the pinhole or camera obscura. His understanding of optics is explained in his famous diagram of light (see Figure 2.1(a))

Amongst all the books written by Ibn Al-Haytam, an Arab physicist, 14 books explained the aspects of the optic. In the *" Book of Optics "* , Ibn Al-Haytam (965-1039 C.E) explained that the rays of light travel on a straight line and the object influence the ray of light by reflecting it in another direction. He described the function of camera obscura (in Latin), which means 'a dark room'. The famous experiment of light from three sources passing through different pinhole sizes is expressed in Figure 2.1(b). He studied the *" Law of Reflection "* and explained how the light reflected in detail. Furthermore, he studied the spherical mirror. He also described the structure and the function of human eyes. The work and books of Ibn Al-Haytam served as a reference in Europe, especially in the domain of optic.

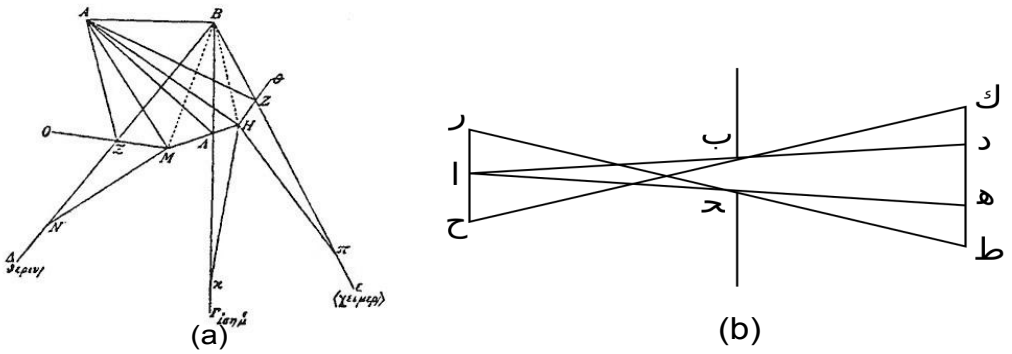


Figure 2.1: (a) Light invented by a Greek mathematician, Anthemius of Tralles, who observed the effect of light through a pinhole. (b) The experiment carried out by Ibn Al-Haytam observed light from three sources through a pinhole.

An Italian painter, draftsman, sculptor, architect, and engineer, Leonardo da Vinci (1452-1519), worked and explained the camera obscura. Another Italian, Battista Della Porta, also wrote on the same topics and has been considered the inventor of the camera obscura working principle through his book entitled *" Camera Obscura "*. This era was

when the period of optic had begun in Europe.

At the beginning of the 17^{th} century, the *" Law of Refraction "* was emphasised within the optic world. Many optical apparatus, such as concave mirror, telescope, and microscope, were generated due to the previous findings. Rene Descartes (1596-1650) had elaborated the *" Law of Refraction "* and published a familiar formula.

Grimaldi (1618-1663) observed the phenomenon of diffraction and its effect. The diffraction is the divergence from the rectilinear propagation that happens when the light advanced exceeds the interference. Studies of light and electromagnetic waves are pretty successful in that era.

Isaac Newton (1642-1727) explained that white light is the composition of several independent lights. He tried to remove the effect of chromatic aberration from his 6-inch long refracting telescope. Many scientists were involved indirectly in the optic research domain. Their philosophies, brilliant ideas, and mind-blowing proposals have been used to advance optics research. The value-added and cross knowledge in various research fields has led to the rapid advancement in optics technology.

After the second half of the 20^{th} century, optics became more interesting. Numerous studies had begun, teaching and emphasising optics with mathematical techniques and the theory of communication. For example, mathematics can translate an image in the form of matrices. The concepts of mathematics and spatial frequency offer a luxurious new way to appreciate the optical domain. Active research works on camera and advancement in the digital computer area have given birth to the field of computer vision. In contrast, the mathematical technique and optic have been integrated to produce the desired outcome in the robotic application's . More details about the history of the optic are described in [20] [21] [22]. Various optical instruments, Figure 2.2, have been discovered and become a valuable reference for optic researchers.

The term 'state-of-the-art' aims to review the current knowledge's and critical points, including substantive findings and theoretical-methodological contribution to a particular topic. The ultimate goal of reviewing the literature is to update readers with the current literature on a topic from the basis for another goal, such as future research sought in the area. The logical flow of an idea characterizes a structured literature review. The current and relevant references with consistent referencing type, proper use of terminology, and an unbiased view of prior research work on the topics are reviewed.

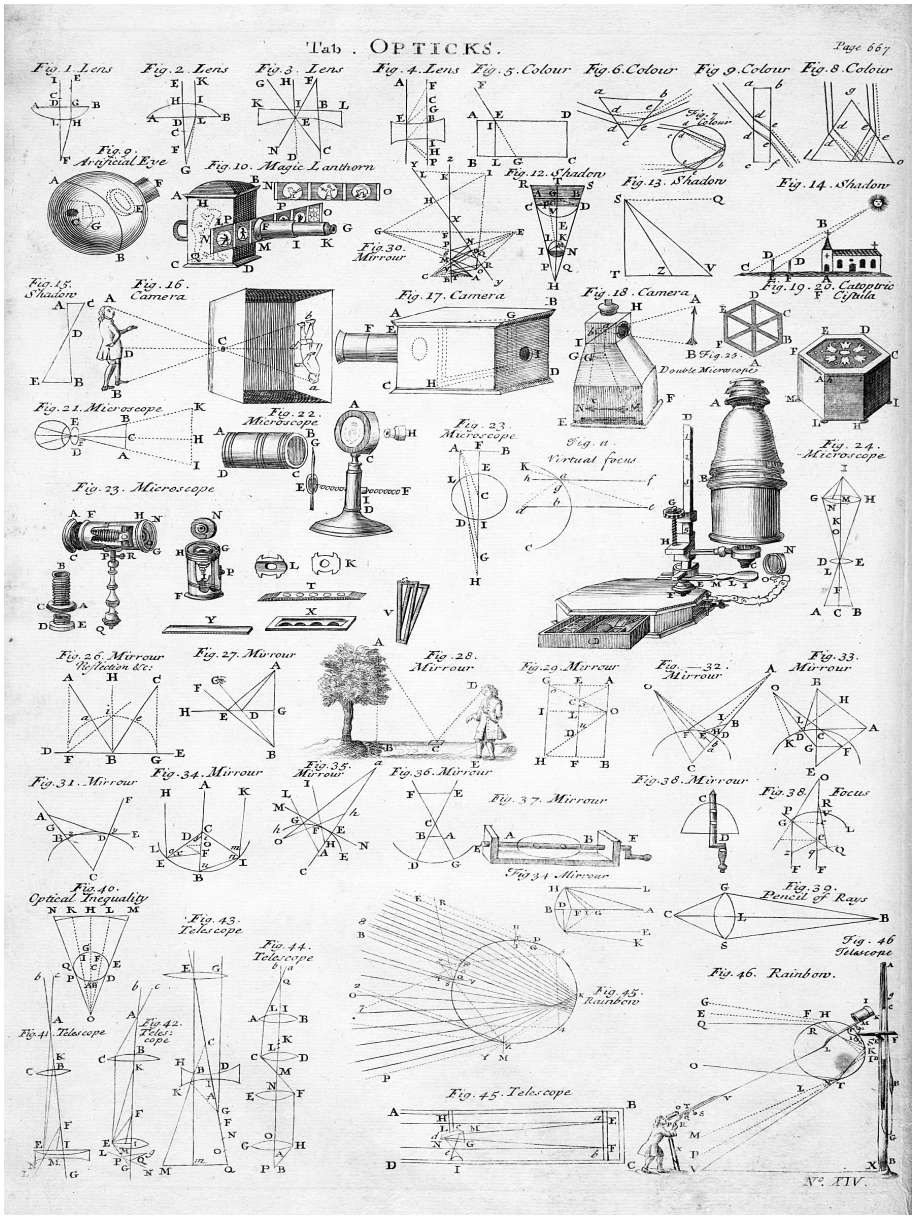


Figure 2.2: The table of Opticks from 1728 *" Cyclopaedia "* Volume 2 that shows various types of optical instruments invented since 1728.

2.2/ PROJECTIVE GEOMETRY

Projective geometry was initiated during the Renaissance period from 14^{th} to 16^{th} century. It started when a painter tried a new technique to express what they see (the 3D world) on a two-dimension (2D) canvas. This idea led to the discovery of geometry, although it is not a mathematical approach.

The contributions created the flow of how 3D offers a sense of depth. For example, A painting or drawing will modify an object's movement and structure to make it more

realistic. Psychologically, looking at a crowd in the distance, the people looked very small. Even though, in reality, their size is approximately the same. The same applies to buildings and mountains. This phenomenon is called the perspective view. (see Figure 2.3).

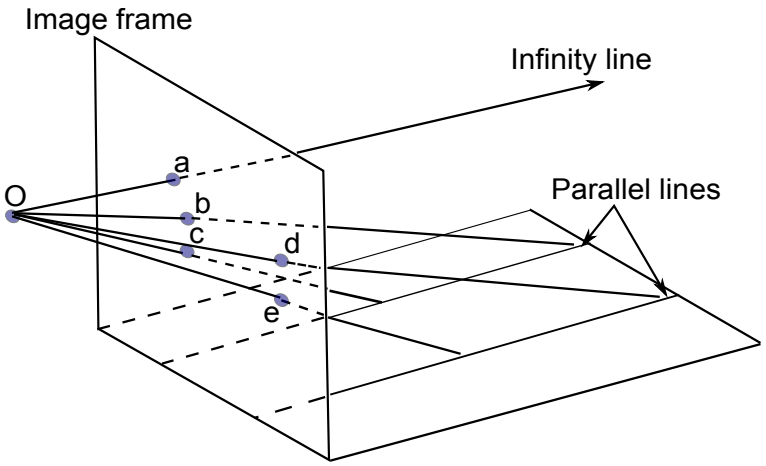


Figure 2.3: The construction of perspective viewing. a, b, c, d and e are the points on the image plane.

A famous example to explain the perspective phenomena is a straight railway or a straight road. Referring to Figure 2.4, the rail and side of the road will intersect at one point called infinity, with the assumption that the earth is plane (flat surface) and wide.

Figure 2.4: The railway and side of the road will meet at the point infinity or vanishing point on the horizon. In reality, the distance between the two parallel lines is the same until the horizon.

The above explanation supports the meaning of projective geometry, which is the study of geometry properties. The studied geometry properties is referred to as the invariant properties, besides distance and angle. Projective geometry is not part of Euclidean geometry but Euclidean geometry is the part of projective geometry. The Euclidean geometry has parallel lines on the same plane that would never intersect each other. The distance between the two lines remains the same as far as the range is expanded. Logically, there is a contradiction that what we see in everyday life, such as the railway scenario explained above. Girard Desargues (1591-1661) revealed this theory, a French architect in his book *" Bruillon Project "* (1639). He explained that the parallel lines have an endpoint called infinity point (vanishing point).

The term infinity does not exist in the Euclidean geometry. Meanwhile, the projective geometry can expand without limit. The first geometry properties of the projection were discovered in the 3^{rd} century by Pappus, a Greek mathematician.

The projective geometry has a role in the vast expansion of technology. The 3D images on the screen and the calculations that produce 3D realistic images applied the perspective geometry formula. *" The Rise and Fall of Projective Geometry "* is the right source if one seeks the history of geometry [23].

Homogeneous Coordinates: The homogeneous coordinates system or perspective projection is an essential tool when modelling projective geometry. It included all common vectors operation, such as rotation, translation, and scaling. The homogeneous system is essential if the research dealing with computer vision. It was introduced by August Ferdinand Möbius in 1827. His book entitled *" Der barycentriche Calcül "*. He explains the use of a coordinate system in projective geometry.

The idea put forward by Möbius helped to explain in detail the projection geometry. It becomes more critical when dealing with the term infinity. Let's say the 3D point $P = (x, y, w)$, reflect homogeneous coordinates at any point on the projective plane. A scale factor is introduced, α, as the non-zero real number. So, $\alpha P = (\alpha x, \alpha y, \alpha w)$ precisely represented the same point. The projection points move towards infinity when the scale factor, α, increases towards infinity. A more straightforward representation of a homogeneous coordinate is dividing each coordinate with the third coordinate, w. This equation is as follows:

$$(x/w, y/w, 1), \tag{2.1}$$

This equation is similar to $(x/w, y/w)$, in Euclidean coordinates. Based on Equation 2.1, if w goes towards 0, the point $(x/w, y/w)$ goes to infinity. The fact is that all the points with 0 in the third coordinate lied on the infinity lines. Further explanation is depicted in Figure 2.3 if the third coordinate, $w = 0$, thus, the point lies on the infinity line or a projection line. Otherwise, if $w = 1$, this is a projection point.

2.3/ PINHOLE CAMERA

The goal of a camera is to bring 3D space to a 2D plane. The geometry explained the transformation from a world scene (3D) to an image plane (2D). Several projection models is used to describe the dimensions graphically.

The pinhole is the optical system that used natural phenomenon. It applied as a reference to all model of camera. This pinhole camera has plenty of names, such as *Sténopé (France)*, *Utta Fotocamera Contorostenopieco (Italy)* and *Obscura (Scandinavia)*. The pinhole camera is the simplest existing camera. It contains a box (black box) with a small hole (at the camera centre) to allow ray light enter into the box. It stamps the collected ray light on the opposite plane (image plane) of the pinhole. The image is formed on the image plane as all objects seen by the camera reflected the ray.

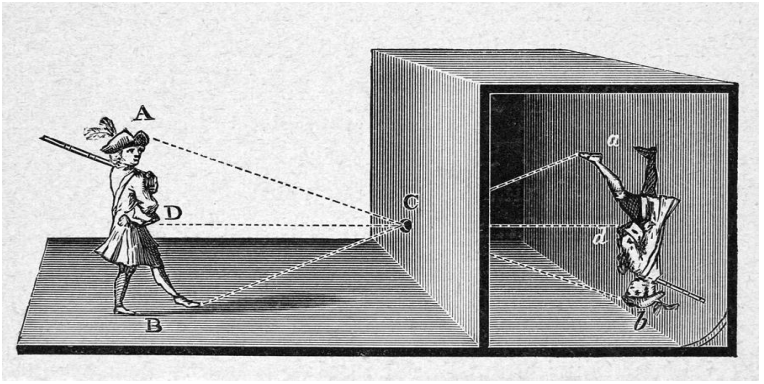


Figure 2.5: The basic principle of the pinhole model used in camera obscura. The image of a soldier carved on the plane-wood opposite the pinhole.

The reflected ray then passed through the pinhole and reconstructs the same image, as seen by the camera in 2D. Compared to the modern camera, the small hole functions as an optical device responsible for forming the quality of image. Figures 2.5

and 2.6 display the pinhole model. Figure 2.7 illustrates the detail formation of the image in the pinhole camera.

The distance between the pinhole plane and the opposite plane is called the focal length (**f**) of the camera. The sharpest image on the plane image depended on the amount of ray that enters through the hole. The smallest hole will result in the sharpest image. For example, if the hole is too big, a square shape object will look like a circle in the pinhole camera (blur image). On the contrary, if the hole is small, less light passes through, making the image sharper. The camera geometrical parameters projected the point 3D (X, Y, Z) in space to 2D, (x, y), on the image plane. This projection is made by extracting and estimating the two parameters; which are the intrinsic and extrinsic parameters. These two parameters modelled the perspective projection of the pinhole camera. The intrinsic parameter (**K**) is the internal parameters of the camera. The parameters are written as 3×3 matrix. The extrinsic parameter is the external parameters of a camera that described the orientation and camera position towards the scene.

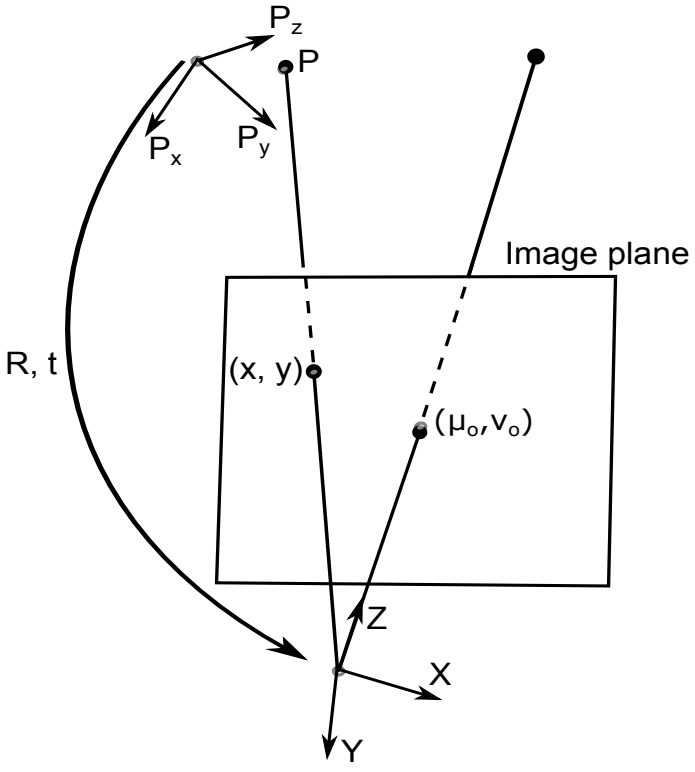


Figure 2.6: The pinhole projection model.

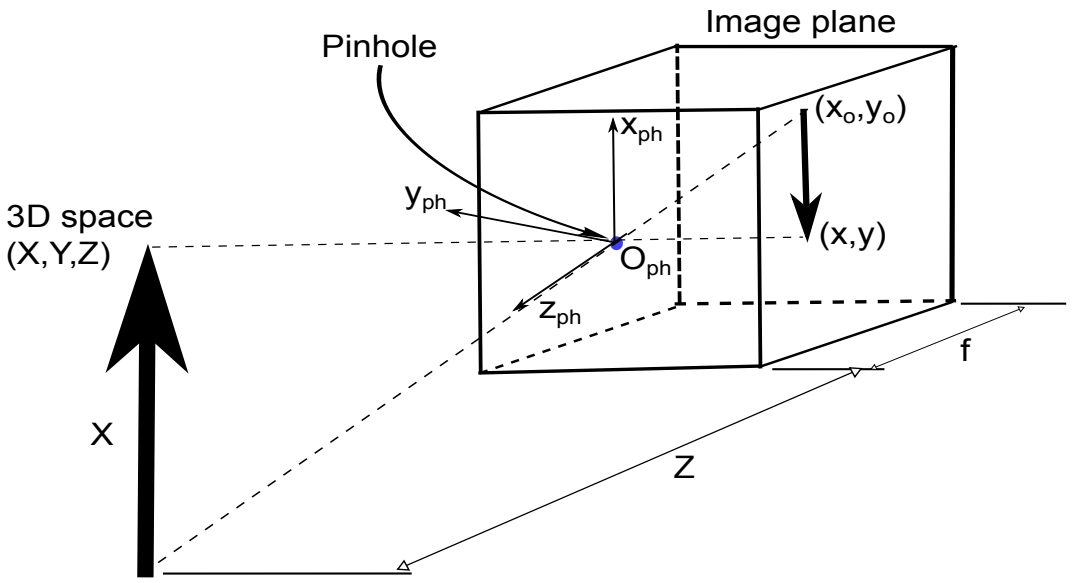


Figure 2.7: The formation of an upside-down arrow image in the pinhole camera. Every point on the object reflects the ray of light that passes through the pinhole concerning to its position. The ray from the top of the arrow passes through the pinhole to the opposite position.

The combination of both parameters can be written as a product of 3×3 rotation matrix ($\mathbf{R}$) and a 3×1 translation vector ($\mathbf{t}$). Combining these two parameters called a projection matrix, (3×4), which brought in a linear method to transform the 3D point to 2D. Due to the camera's rapid evaluation, the pinhole camera has been replaced with a lens and the adjustable pinhole. This technique controls the amount of ray light that passed through the hole to get a clear and sharp image. The lens allowed the camera to focus the ray light and brought it perfectly on the image plane in any condition.

Ryberg et al, [24] developed a general projection model perspective or omnidirectional camera. It included non-central cameras and compensations for decentring distortion. Kanala et al, [25], developed a generic camera model and a calibration method for an omnidirectional camera.

2.3.1/ INTRINSIC PARAMETERS

Every camera has its digital characteristics that described its optical model. In computer vision, this is called an intrinsic parameters. The intrinsic parameters is defined

as the internal parameters of cameras. It described their optical geometry's distinctive nature, and allowed the mapping between camera coordinates and pixel coordinates in the image frame, which is the transformation from 3D coordinate to 2D coordinate or, in general terms, a mapping from world point to image coordinates. The parameters are focal length, principal point, frame size, and geometry distortion.

A perspective projective or focal length is the distance between the centre of the camera frame and the image frame. A line perpendicular from the centre of the camera to the centre of the image plane called the principal line. This line passed through the principal point, $\mathbf{O}$, or the centre coordinate of the image frame $(\upsilon_{\mathbf{0}}, \nu_{\mathbf{0}})$. A world point $\mathbf{P}$ that passed through the image frame at coordinate (υ, ν) is called image coordinate. Figure 2.8 shows the world point from Euclidean 3D-$\mathbb{R}^3$ space $(X, Y, Z)^T$ mapped to the image plane Euclidean 2D-$\mathbb{R}^2$ point $\left(\frac{fX}{Z}, \frac{fY}{Z}\right)^T$. Meanwhile, there is another parameter which is involved in the develpment of wide angle camera The parameter called eccentricity. The eccentricity ξ, is a parameter that estimates the level of distortion and it influenced by the lenses.

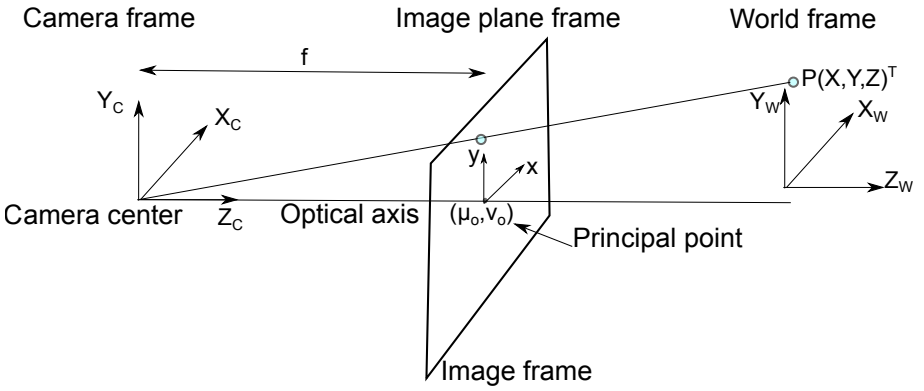


Figure 2.8: The camera model and the coordinate system.

The point $\mathbf{P}$ is expressed as follows:

$$(X, Y, Z)^T \mapsto \left(\frac{fX}{Z}, \frac{fY}{Z}\right)^T, \tag{2.2}$$

Equation 2.2 becomes non-linear if it is transformed directly into Euclidean space. In order to make the equation linear, it should be re-written in the form of homogeneous coordinate. The linear equation can be written in homogeneous coordinate, as given

below:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{2.3}$$

Generally, the matrix representation can be expressed as $\mathbf{diag(f, f, 1)[I|0]}$, where $\mathbf{diag(f, f, 1)}$ is a diagonal matrix and $\mathbf{[I|0]}$ represent a $\mathbf{3 \times 3}$ identity matrix with a column vector (zero vector).

The expression above is a general expression by assuming the origin coordinate is at the image plane's centre. In reality, this is not the case, mainly because the image plane's central origin is not always at the centre of the image plane. It is commonly situated at the corner of the image plane. The matrix representation for mapping from 3D to 2D is generalized as follows:

$$\mathbf{P} \mapsto \mathbf{p}, \tag{2.4}$$

$$(X, Y, Z)^T \mapsto \left(\frac{fX}{Z} + \mu_0, \frac{fY}{Z} + \upsilon_0 \right)^T, \tag{2.5}$$

where $(\mu_\mathbf{0}, \upsilon_\mathbf{0})^T$ is the coordinate of the principal point. Equation 2.5 expressed in homogeneous coordinate as:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} fX + Z\mu_0 \\ fY + Z\upsilon_0 \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & \mu_0 & 0 \\ 0 & f & \upsilon_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{2.6}$$

Another parameter that should be considered is the orthogonality of the camera frame. A frame is orthogonal if axes x and y are perpendicular to each other (see Figure 2.9). Previously, the camera's internal parameters are considered orthogonal with similar scales on both axial directions. In some cases, the CCD camera has a different pixel size between both axials. Thus, the camera frame is not square. In light of this possibility, another two parameters needed to be modelled and added to the camera matrix (Equation 2.7). The two parameters are λ (scale factor of the width and length of the camera frame) and s (a parameter of non-orthogonality of the camera frame). If the camera frame is orthogonal, so $s = 0$.

After considering the scale factor and the skew, the camera matrices are written

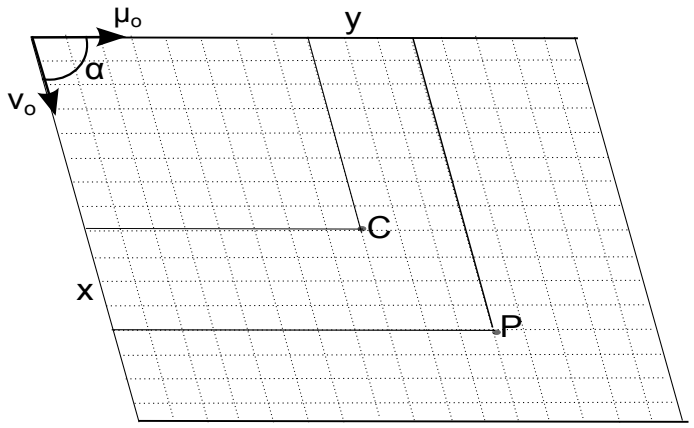


Figure 2.9: In the camera frame, the skew is derived from the tangent of an angle, α and the focal length, $\mathbf{f}$, where $\mathbf{s} = \mathbf{f}\tan(\alpha)$, the function that defines the orthogonality of a pixel.

as follows:

$$K = \begin{bmatrix} f & s & \mu_0 \\ 0 & \lambda f & \upsilon_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.7}$$

where,

- λ is the ratio between the width and the length of the pixel in the image frame. This ratio is to make the pixel perfectly square. If the pixel is square, λ=1.
- f is the focal length, which is the distance between the camera centre and the image plane. This parameter has a role in zooming the image.
- μ_0 and υ_0 are the principal point of the image.
- s is the skew, a parameter that corresponds to the non-orthogonality of the image frame.

With parameters $(\lambda, f, \mu_0, \upsilon_0, s)$, the camera matrix has 5 degrees of freedom, and finally, Equation 2.6 is extended to:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} fX + Z\mu_0 \\ fY + Z\upsilon_0 \\ Z \end{bmatrix} = \begin{bmatrix} f & s & \mu_0 & 0 \\ 0 & \lambda f & \upsilon_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{2.8}$$

2.3.2/ EXTRINSIC PARAMETERS

The extrinsic parameters reflected the transformation between the unknown camera reference frame and the known world reference frame. In precise, extrinsic parameters determined the location and the orientation of a camera toward the world frame.

Referring to Figure 2.10, lets say a 3D point $\mathbf{P_m}$, has a coordinate $(\mathbf{X_m}, \mathbf{Y_m}, \mathbf{Z_m})$ for world axes $\mathbf{O_m}$. In order to change from world axes to camera axes, a rigid transformation matrix, $\mathbf{T}$, is applied. It consists the rotation, $\mathbf{R}$ matrix, and vector translation $\mathbf{t} = [\mathbf{t_x}, \mathbf{t_y}, \mathbf{t_z}]^\mathbf{T}$. Using the extrinsic parameters, the coordinates of point $\mathbf{P}$ in world $\mathbf{P_W}$ and camera $\mathbf{P_c}$ can be estimated and finally connected at their correlation.

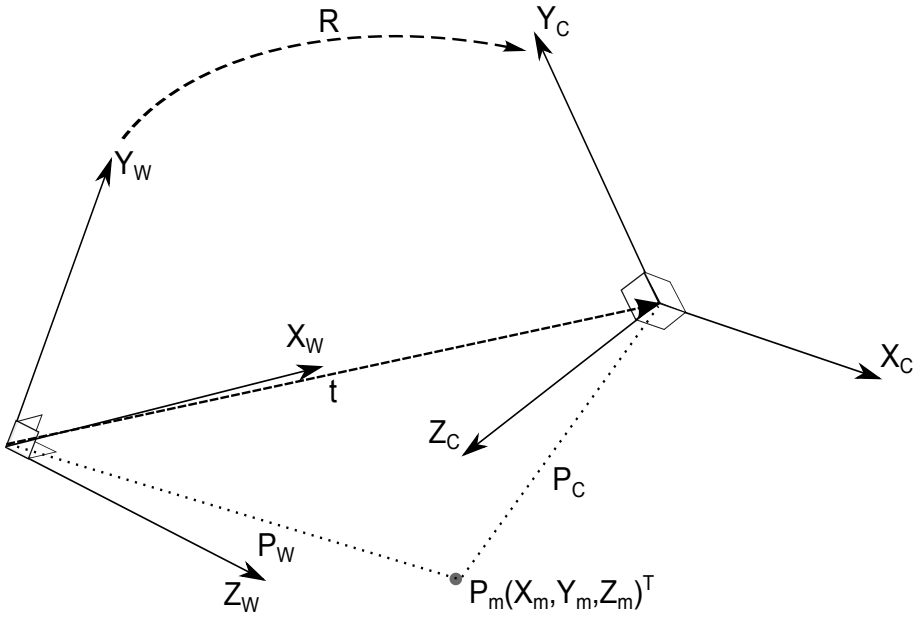


Figure 2.10: The extrinsic parameters are used to compute transformation, $\mathbf{T}$. The translation vector $\mathbf{t}$ estimate the relative positions between two reference frames, where $\mathbf{t}$ has 3 degrees of freedom. The rotation matrix, $\mathbf{R}$, is a parameter that aligns the corresponding axes of two frames, where $\mathbf{R}$ also has 3 degrees of freedom. These parameters identify the transformation between an unknown camera reference and a known world reference frame.

Translation: The translation, $\mathbf{t}$, is a geometry transformation that moves all the related points without changing the form or the orientation of object or space. The 3D translation considered all three directions in the Euclidean coordinates. In equation 2.12, let's consider a 3D point $\mathbf{P_m}$ with coordinate $(\mathbf{X_m}, \mathbf{Y_m}, \mathbf{Z_m})$ while the world coordinate and the camera coordinate are $(\mathbf{X_W}, \mathbf{Y_W}, \mathbf{Z_W})$ and $(\mathbf{X_c}, \mathbf{Y_c}, \mathbf{Z_c})$ respectively. The translation from world axes to camera axes performed by adding the translation $(\mathbf{t_x}, \mathbf{t_y}, \mathbf{t_z})$ vector to the

respective direction. The equation should become:

$$\mathbf{X_W} = \mathbf{X_c} + \mathbf{t_x}, \tag{2.9}$$

$$\mathbf{Y_W} = \mathbf{Y_c} + \mathbf{t_y}, \tag{2.10}$$

$$\mathbf{Z_W} = \mathbf{Z_c} + \mathbf{t_z}, \tag{2.11}$$

or

$$\begin{bmatrix} \mathbf{X_W} \\ \mathbf{Y_W} \\ \mathbf{Z_W} \end{bmatrix} = \begin{bmatrix} \mathbf{X_c} \\ \mathbf{Y_c} \\ \mathbf{Z_c} \end{bmatrix} + \begin{bmatrix} \mathbf{t_x} \\ \mathbf{t_y} \\ \mathbf{t_z} \end{bmatrix} \tag{2.12}$$

Rotation: Another parameter in transformation matrix ($\mathbf{T}$) is the rotation matrix, ($\mathbf{R}$). The rotation matrix is used to perform a rotation in Euclidean space. It is aligned with the orientation of the world axis to the camera place axis. Upon reckoning that the axes of the image plane are $\mathbf{x}$ and $\mathbf{y}$, the camera's orientation is that the world axes should be perpendicular to the image plane. For this purpose, a 3×3 matrix $\mathbf{R}$ is used to estimate the rotation on $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ axes. The matrix $\mathbf{R}$ is the product of a matrix at all axes. The rotation matrix at the respective axis is estimated as follows:

$$\mathbf{R_x}(\theta_\mathbf{x}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\theta_x) & -sin(\theta_x) \\ 0 & sin(\theta_x) & cos(\theta_x) \end{bmatrix}, \tag{2.13}$$

$$\mathbf{R_y}(\theta_\mathbf{x}) = \begin{bmatrix} cos(\theta_y) & 0 & sin(\theta_y) \\ 0 & 1 & 0 \\ -sin(\theta_y) & 0 & cos(\theta_y) \end{bmatrix}, \tag{2.14}$$

$$\mathbf{R_z}(\theta_\mathbf{z}) = \begin{bmatrix} cos(\theta_z) & -sin(\theta_z) & 0 \\ sin(\theta_z) & cos(\theta_z & 0 \\ 0 & 0 & 1) \end{bmatrix}, \tag{2.15}$$

where θ is an angle of rotation in radians.

The simultaneous rotation of the three-axis represents the Euler representation, which is:

$$\mathbf{R} = \mathbf{R}(\theta_\mathbf{x}, \theta_\mathbf{y}, \theta_\mathbf{z}) = \mathbf{R_x}(\theta_\mathbf{x})\mathbf{R_y}(\theta_\mathbf{y})\mathbf{R_z}(\theta_\mathbf{z}), \tag{2.16}$$

Matrix, $\mathbf{R}$ is called a particular orthogonal matrix with determinant one and has the following properties:

$$\mathbf{AA^T = I}, \tag{2.17}$$

where $\mathbf{A^T}$ is the transpose of $\mathbf{A}$ and $\mathbf{I}$ is an identity matrix.

Referring to Figure 2.11, in the case of pure rotation, the orientation of the camera is changed, and the formulation is as follows:

$$\mathbf{P_c = RP_W}, \tag{2.18}$$

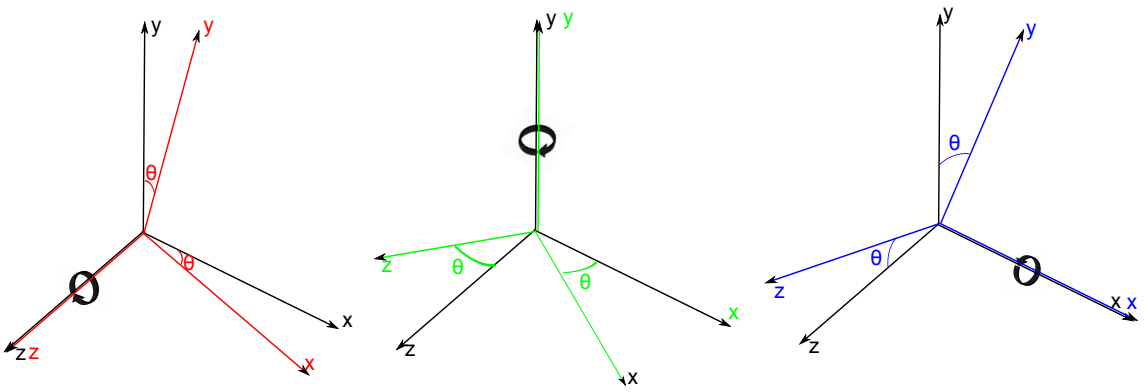


Figure 2.11: The rotation of X, Y, and Z axes. The rotation matrix, $\mathbf{R}$ has the rotation of all axis.

Finally, the matrix of extrinsic parameters that contain rotation $\mathbf{R}$ and translation $\mathbf{t}$ is:

$$[\mathbf{R}|\mathbf{t}] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}, \tag{2.19}$$

where $(\mathbf{r_{11}.....r_{33}})$ is a parameter of general rotation matrix, while $(\mathbf{t_x, t_y, t_z})$ is a parameter of translation vector

Projection Matrix: After estimating the intrinsic and extrinsic parameters, the combination of that two parameters can estimate the global matrix called a projection matrix. This combination estimated the linear equation that projected the points of the scene on the image. By taking into account the intrinsic and extrinsic parameters, Equations 2.18 and 2.19 are combined to estimate the global equation. This general equation, when written

in homogeneous coordinate, is as follows:

$$\mathbf{P_c} = \mathbf{K[R|t]P_W}, \tag{2.20}$$

Where $\mathbf{K[R|t]}$ is a projection matrix with the dimension $\mathbf{3 \times 4}$, that contained the rotation matrix and the translation vector to allow transformation geometry of points on the scene. The final equation is:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} & & & t_x \\ & R_{3\times3} & & t_y \\ & & & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{2.21}$$

2.3.3/ Distortion: Modelling and Parameters.

The distortion optics is the divergence from the rectilinear projection. The standard projection projected the straight line on the scene and should also be straight on an image. It is called rectilinear due to the consequence of using the lens and the property's existence in the optical system called **aberration**.

The lens is part of the optical systems. The use of a lens causes the image to be blurred or distorted. Looking back to the projection model in Equation 2.20 for a pinhole camera, the projection model is good if and only if there is no lens or a thin lens applied. It does not change the form of the image or the straightness of the straight line on the image. A fair approximation of distortion on the pinhole camera by adding a lens in front of the camera is inadequate and unsuitable for a fisheye image, or wide-angle camera [26] [27] [28]. However, psychologically, sometimes looking at a moving object through lens or binoculars disabled the prediction [29] [30].

A camera equipped with a special lens to increase the FoV with a short focal length is not considered as a pure pinhole model projection. This camera known as an omnidirectional camera, and it will discuss in the following section. The omnidirectional camera equipped with a lens to increase FoV. The lens generated the radial and tangential distortions, whereby these two distortion parameters should be considered in the projection model.

A projection of one point on the image plane should be realistic with its actual projection point. If those points are not realistic (between projection plane and real),

the camera system considered as imperfect. This imperfection is due to the presence of distortion. Some studies have modelled the distortion of a fisheye camera. Brauer et al. [31] corrected such distortion by only using a single image and imaged objects' linearity. Devenay et al., [32] used the fundamental properties to identify distortion on the camera. They used a logical theory that any straight line on the lens should be straight in reality. However, no unique solution is proposed to remove distortion on the lens.

Meanwhile, Basu et al. [33], introduced fisheye transform (FET) to describe the variable resolution mapping function. A sensor model was developed to acquire image according to the transform. Miyamoto et al. [34] explained the existence of distortion on lens or image. The distortion existed and it is proportional to the change of field of view. It can be modelled and added to the projection model. Let's say, if the projection of image plane is $\mathbf{p_p(x, y)}$), the real projection is $\mathbf{p_r(x_r, y_r)}$ and the model of distortion is $\mathbf{D(p)}$. So, the projection points should be corrected using:

$$\mathbf{p_p = p_r + D(p),} \tag{2.22}$$

In photography, the two types of distortion are perspective distortion and optical distortion. The effect can been seen on the formation of an image on the image plane. The perspective distortion and the effect of aberration are not discussed in detail in this thesis. The distortion generated by omnidirectional cameras is more visible and should be addressed an the critical issues in this thesis. In general, the optical distortion on the lens and mirror are called radial distortion and it divides into three. The three types of radial distortion are a barrel, pincushion, and wavy. A special lens is designed to reduce the spherical effect. In other words, the lens ascertained that all light is pointed on a common point.

The illumination is also one of the cause in generating the distortion together on the formation of an image on the image plane. However, it contributed only minor effect [35]. The correct distortions level should be modelled and able to be used in precise application such as in the scene analysis.

Radial distortion: The radial distortion is split into three types: barrel, pincushion, and wave or moustache distortions. Figure 2.12 illustrates the three types of radial distortion. Among them, barrel and pincushion distortions are the most visible on the image. The presence of distortion is detected by observation if a straight line in the world becomes a

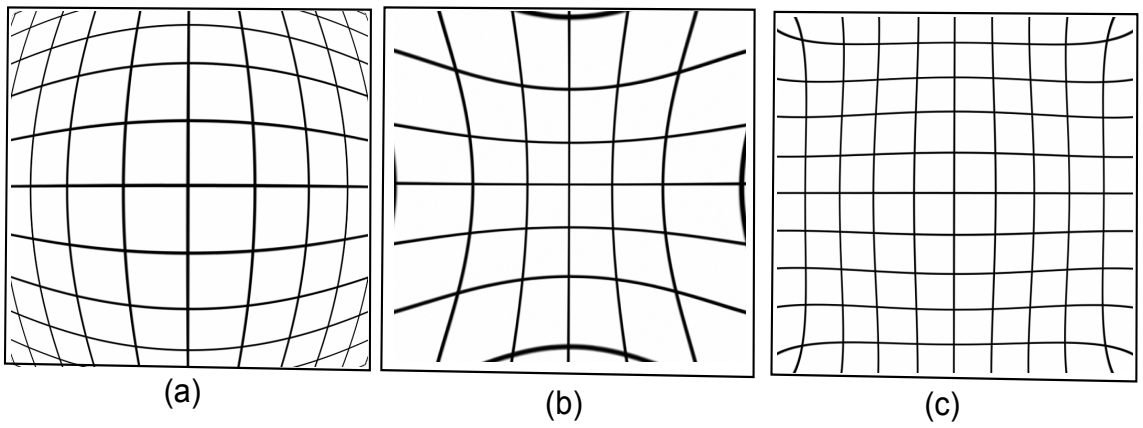


Figure 2.12: Radial distortion developed by the lens. The types of radial distortion are barrel distortion (a), pincushion distortion (b), and moustache distortion (c).

curve. The straight line is bent outward from the centre due to barrel distortion or radial distortion positive from the fisheye lens. It typically existed on the most wide-angle lens with a short focal length. The straight line bent inward toward the centre of the image due to pincushion distortion or radial distortion negative, which is the exact opposite of barrel distortion.

The current trend in photography is to couple a lens in front of the camera to increase the FoV. Thus, it generated the undesirable distortion on the image. The value of distortion varies non-linearly from the centre of the lens to the periphery. This effect of distortion is due to the concave or convex form of the lens. Generally, the centre of the lens is less affected if compared to the periphery, in the case of convex mirror and opposite for the concave mirror. The distortion is modelled and expressed in polynomial symmetrically by considering the symmetrical model of the lens [36]. If the coordinate of distorted point is $\mathbf{p(x_d, y_d)}$ and the centre of camera or distortion is $\mathbf{c(\mu_0, \upsilon_0)}$, the distance $\mathbf{r}$ is defined as follows:

$$\mathbf{r} = \sqrt{(\mathbf{x_d} - \mu_0)^2 + (\mathbf{y_d} - \upsilon_0)^2}, \tag{2.23}$$

From equation 2.23, the radial distortion can be modelled as:

$$\mathbf{D_r} = \begin{pmatrix} (\mathbf{x_d} - \mu_0)\left(\mathbf{k_1 r^2} + \mathbf{k_2 r^4} + ...\right) \\ (\mathbf{y_d} - \upsilon_0)\left(\mathbf{k_1 r^2} + \mathbf{k_2 r^4} + ...\right) \end{pmatrix}, \tag{2.24}$$

where,

- $\mathbf{D_r}$ is radial distortion.
- $\mathbf{(x_d, y_d)}$ is the coordinate of distorted point.
- $\mathbf{(\mu_0, \upsilon_0)}$ is the coordinate of camera center.
- $\mathbf{k_n}$ is the parameters of distortion to be estimated during calibration.
- $\mathbf{r}$ is the distance from the distorted coordinate to the center of distortion.

Tangential distortion: The tangential or decentring distortion is common when there is a misalignment between the number of a lens and the camera system's sensor plane (see Figure 2.13). This means that, the centre of all lenses does not unite in a straight line [37]. It is the linear shift of points on the photo in the direction of normal towards the radial line. The effect of this decentring distortion can cause a hyperbolic impression on the photo captured by the camera [26]. Defect lens and camera also can cause a tangential distortion.

Two polynomial equations usually describe this distortion in the directions of $\mathbf{x(dx)}$ and $\mathbf{y(dy)}$. The coefficient of tangential distortion is always denoted with $\mathbf{P_1}$ and $\mathbf{P_2}$. The tangential distortion cannot be avoided as it occurs during the production of the camera. It existed in a small scale. However, the error is still based on the tolerance allowed, which does not qualitatively affect the photo's quality. Hence, tangential distortion is typically omitted in the production of optical distortion [38] [39] [40].

2.4/ OMNIDIRECTIONAL CAMERA

A vision system with wide FoV offered a lot of advantages. A conventional camera or perspective system perceived insufficient information due to limited FoV [41]. A vision system with many modalities is more beneficial in carrying out specific tasks, such as surveillance and more sophisticated robotic application. The application of a camera system for video surveillance needs a wide range of view to optimize observation outcomes and minimizes the number of camera for the same coverage area. A simple example is seen in all direction while providing situational awareness in surveillance and performing autonomous navigation tasks. Some applications of omnidirectional camera are robot localization and mapping [42] [43] [44] [45] as well as robot navigation [46] [47] [48] [49].

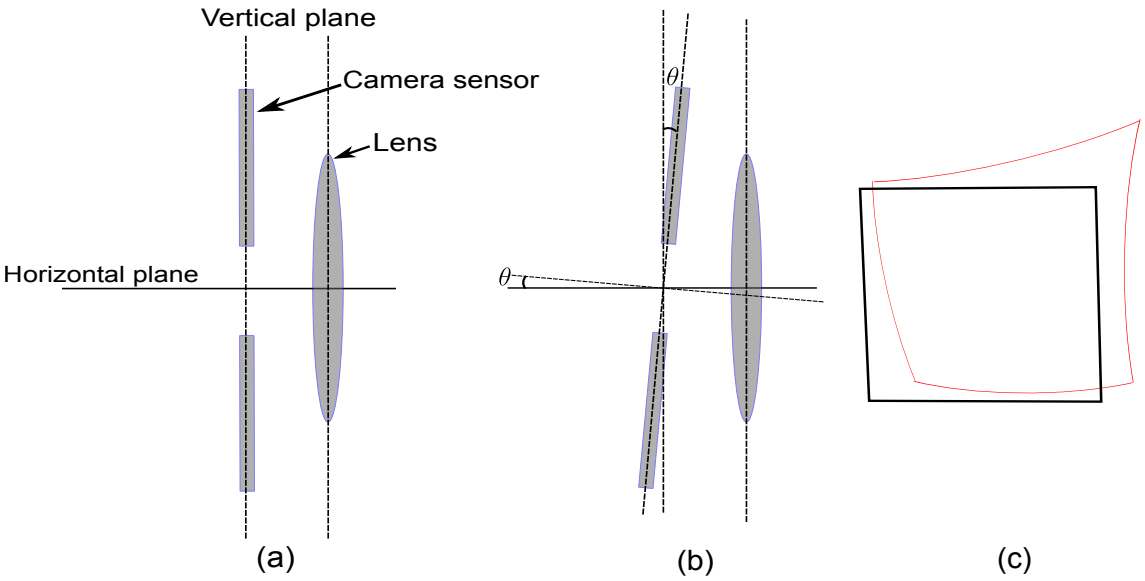


Figure 2.13: Tangential distortion is due to misalignment between the camera plane and lens plane. The value of tangential distortion is zero if the vertical plane camera sensor and lens are in parallel. Illustration (a) shows the camera sensor plane and lens plane are perfectly perpendicular to the horizontal plane (Error=0). In illustration (b), the camera planes are not parallel with the lens plane (Error≠0). In illustration (c), the red line is the image affected by tangential distortion.

A camera system with wide FoV is also used for object detection, recognition or tracking [50] [51], and visual servoing [52] [53] [54] [55]. Other interesting applications of omnidirectional camera are the structure from motion (SFM) [56] [57] [58] [59] and virtual-reality [60] [61] [62]. The Google map is a great application that offers web-mapping service. The integration with 360° satellite imagery offers immense tools for navigation [63]. The word “Omni”, which derived from Latin, means ‘all or every’. An omnidirectional camera is a camera system with a wide FoV about 360°. This means that this camera system can see the scene in all directions vertically and horizontally simultaneously.

Compared to the traditional perspective camera with limited FoV, the omnidirectional camera is widely applied in computer vision either for research or real application, especially for video surveillance. In light of robotics application, the omnidirectional camera reduced the response time for object detection and minimizes the extensive vision coverage costs. The direct estimation of the scenes, such as 3D reconstruction and depth estimation, is the famous research field to define the sensor’s maximum capability. Despite the vast advantages of wide FoV, the drawbacks of the omnidirectional camera are the low resolution and low readability images. This condition due to the presence of

several distortions that can deform the images.

Several techniques are proposed to increase the FoV. For example, a dioptric that uses a special lens with a short focal length [64] [65]. Huang et al. [66] compared the performance of the existing omnidirectional cameras. They looked into issues that revolved around the omnidirectional cameras by focussing on the mathematical point of view. It led to the combination of reflecting mirror and perspective camera. It is called catadioptric [67]. The combination of several cameras called polydioptric camera, and the integration of a perspective camera with a rotary motor to acquire panoramic images. Among the techniques, some presented the advantages and disadvantages, such as the quality and resolution of vision, rigidity, ergonomic, speed, synchronization issue, and cost.

To date, there are several examples of such modalities that have emerged with the development of portable and computational devices, including mobile phones and wireless communication systems. Ricoh Theta, Nikon KeyMission 360°, Bublcam, Nokia OZO, Samsung Gear 360, and Ladybug5 are among the latest polydioptric cameras available in the market. Several research articles have incorporated Ricoh Theta [68] and earlier versions of Ladybug5 [69] [70] [71].

2.4.1/ DIOPTRIC

A dioptric type camera or fisheye camera is a combination of perspective camera and fisheye lens. It called a fisheye camera because of the lens's form, which is based on how a fish sees its surrounding environment with a wide-angle view. The fisheye camera is known as ultra or a super wide-view camera. Due to their short focal length, this camera can view more than 180° vertically and horizontally concurrently.

The hemisphere image or wide panoramic image type tends to deform the raw image. Such as a building's wall or any straight features, to be a curve due to radial distortion. This camera's resolution varied from the centre to the peripheral of the far edge of the lens. The centre of this camera has a higher resolution when compared to low resolution at the periphery due to several distortion levels. Figure 2.14 shows the example of a fisheye camera and the image taken from the fisheye camera. Figure 2.15 shows the direction of ray light through the fisheye lens, which produced a wide FoV.

Figure 2.14: (a) image taken from fisheye camera. (b) conventional camera attached with the fisheye lens.

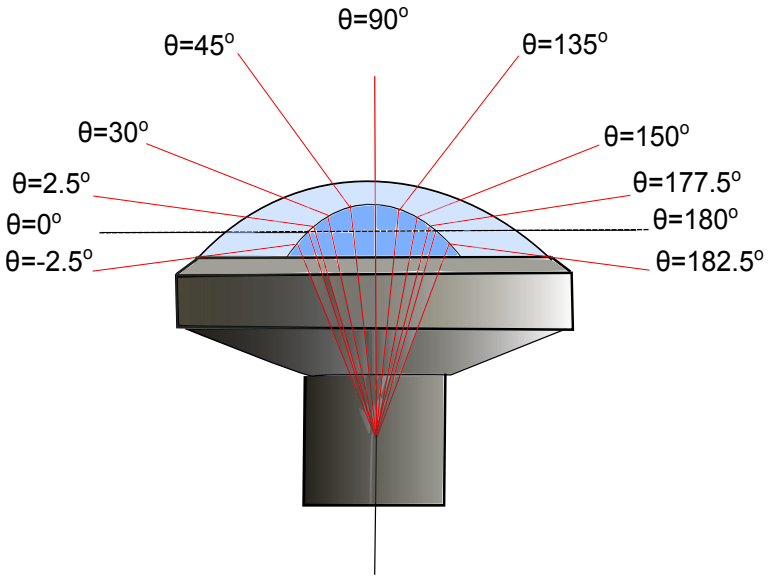


Figure 2.15: Lines of ray light enters into the fisheye camera. The FoV of this fisheye lens is 185° which are 2.5° degrees below the 0°-180° line on the left and right of hemisphere.

2.4.2/ CATADIOPTRIC

A catadioptric camera is a combination of a single perspective camera coupled with a conic section reflective mirror (parabola shape). The perspective camera is placed

towards or perpendicular to the conic mirror so that 360° reflected images can capture 360° panoramic images horizontally [41] [72].

The images obtained by the camera are reflected image. It means; the image's quality is reduced by one step before the camera captured the image due to the quality of the mirror itself. The reflected mirror reduced light energy from the world point to the image plane. Same as the dioptric camera, the image resolution changed from the centre of the camera to the conic mirror's periphery. The presence of a fix blind-spot at the centre of the reflected image is referred to the camera's images itself. This point placed the catadioptric camera behind the dioptric camera for better coverage area. Figure 2.16 shows the image was taken and the example construction of a catadioptric camera.

(a) (b)

Figure 2.16: (a) the image taken fron the catadioptric camera and the illustration (b) the camera face on the dome shape reflected mirror.

At the central projection camera (single viewpoint), all light from the image is projected on the same point simultaneously. In the case of a catadioptric camera, the conic mirror's shape characteristic if followed so that all the reflected rays from the image are concentrated on the same point. Otherwise, it becomes a non-central projection camera model with more than one projection point (see Figure 2.17).

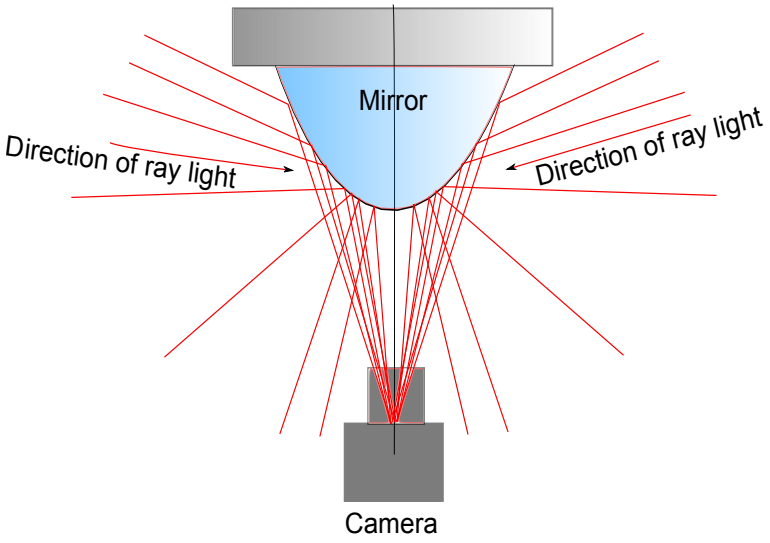


Figure 2.17: The direction of light reflecting on the conic mirror to the camera. The centre of the conic mirror considered a blind-spot that contains the image of the camera.

2.4.3/ POLYDIOPTRIC

A polydioptric camera combined several conventional or perspective cameras. All the cameras placed in all direction to capture and visualize a panorama image a round a body (see Figura 2.18). This method is another solution to increase FoV. The cameras are placed in specific positions to cover 360° FoV, in which all the images are fused to form a panoramic image. The position between cameras is estimated during the calibration process. The polydioptric camera used high-resolution perspective cameras to capture high-resolution images. This system offers advantages in front of the dioptric and catadioptric cameras to produce the best quality images and provide a high-resolution panoramic image as the final output (see Figure 2.19).

The problem of synchronization between cameras and the high cost of development are the main disadvantages of this system. The different intensity between all the cameras causes features matching between two overlapping images to more complex. A system called Omnipolydioptric camera is similar to a polydioptric camera. This system combined several omnidirectional and polydioptric system cameras instead of using a perspective camera. As mentioned in Section 2.3.1 and Section 2.3.2, the use of omnidirectional cameras will minimize the number of cameras due to their large FoV, but the disadvantage of omnidirectional camera is the low-resolution images.

(a) (b)

Figure 2.18: The polydioptric camera consists of six cameras. Five cameras are used to capture horizontal images and produce 360° horizontally, with one capturing the upper image. (a) shows the panorama image stitched from five images. (b) The image is the commercial polydioptric camera, LadyBugs from Point Grey.

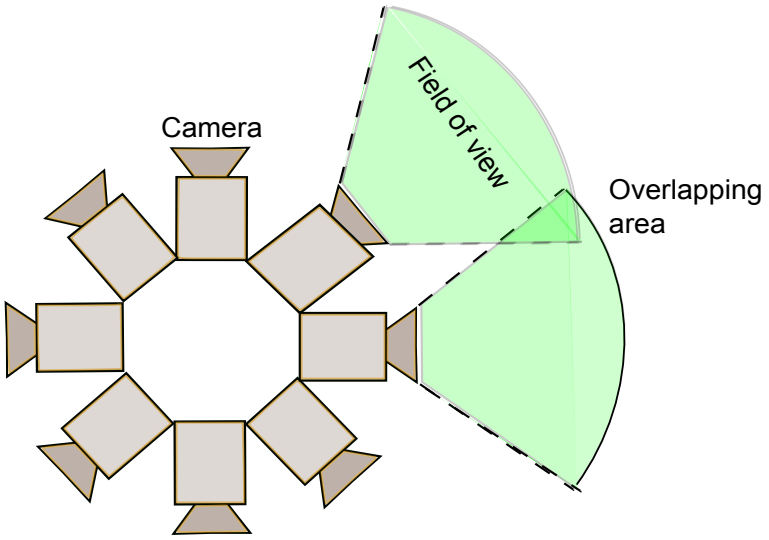


Figure 2.19: An example of a polydioptric camera that has eight perspective cameras. Each camera has less than 90° FoV. Between the two cameras is an overlapping view, where each camera views the same area. These mutual points are used to stitch the images.

2.4.4/ ROTARY

A rotative camera is another way to capture a panoramic image. A perspective camera (CCD) is placed on a rotating vertical bar to capture several images in the vertical direction. The output image is the image from the perspective image with high resolution. The generation of panoramic images using this type of camera depends on the motor's rotation speed and the shutter speed of the camera. The two parameters

must be synchronized to produce an excellent panoramic photo. Usually, a single rotation axis will produce a cylindrical panoramic image. It concerning pure rotation estimation on respective axis. As mentioned in Section 2.3.3, the different light intensities between all images effected the fusion process on the overlapping area (see Figure 2.20 and 2.21). The same problem occurred for the rotary system during mosaicking or stitching. The non-uniform light will affects the production of a good panorama image.

An image processing technique called mosaicking is used to fuse all the images to produce a wide FoV with a high-resolution image. There is a possibility of extending a large mosaicking image at all degrees of freedom and with different zooming levels. For example, the rotative surveillance camera called Pan-Tilt-Zoom (PTZ). The result of the mosaicking image is a high resolution with a wide FoV image.

(a) (b)

Figure 2.20: The high-resolution camera takes four images in one rotation. They are indexed as N(north), E(east), S(south) and W(west). Mean that all the photos are at the right angle with one another, respecting the flow of rotation. The combination of photos N→ E→S→E→N, produce a high resolution 360° panorama image. Picture (b) is the rotary camera.

2.4.5/ Pan-Tilt-Zoom (PTZ) Camera

A PTZ camera is a camera system that combined mechanical movement and a high-resolution vision perspective camera. The combination helped the camera to increase vision coverage. From the definition of 'pan', the camera can swivel horizontally at its fixed position. The tilt refers to the vertical movement of the camera from its rigid position.

The zoom is the capability of the camera to zoom out and zoom in by adjusting

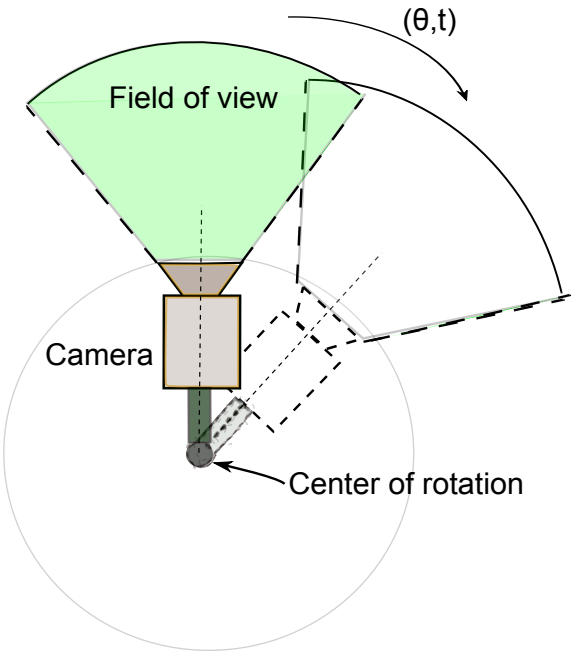


Figure 2.21: The detail of rotary camera. The camera rotates at the center of rotation with $\mathbf{f}(\theta, \mathbf{t})$. Where θ is the angle for camera to capture the image.

its optical lens. The PTZ camera moved in two degrees of freedom in its fixed position. Hence, this camera offers flexibility in its movement with a high-resolution image. The best application of the PTZ camera is in the high-end video surveillance system [73]. Chen et al. [74] asserted that the current algorithm demands the projection model's knowledge to solve the non-linear spatial correspondences between PTZ and omnidirectional camera.

The PTZ camera addressed the drawbacks of using a geometry and homography. This camera (slave: high-resolution image) coupled with the omnidirectional camera (master: low-resolution image). As the omnidirectional camera's role is to survey the environment, the purpose of the PTZ camera is to monitor specific target object activities. Figure 2.22 portrays the camera PTZ (a) and the application of the PTZ camera with an omnidirectional camera (b)..

2.5/ STEREO-VISION CAMERA

Stereo vision is a technique used to extract the 3D description of a scene observed from different viewpoints [75]. In biology, this process called stereopsis. The stereopsis was discovered and explained by Charles Wheatstome in 1838. He claimed, "*..the mind perceives an object of three dimensions by using two dissimilar pictures projected by it*

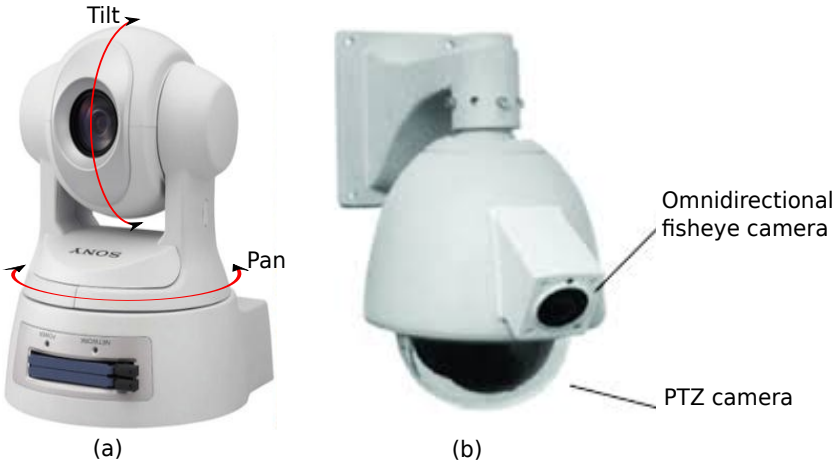


Figure 2.22: The PTZ camera. (a) presents the two degrees of movement in horizontal and vertical directions. Photo (b) is the combination between a PTZ camera and an omnidirectional camera for video surveillance.

on the two retinæ".

An example of a natural stereo vision is our vision system. The left and right eyes see at the same point and created the illusion of 3D from the raw images. The interpretation between image acquisition and information processed by the brain estimates the relative distance from the human/viewer to the object or scene's depth dimension.
Some cues and signals sent to the brain are as follows:

- The stereopsis is the necessary visual information extended by human eyes to estimate the perception of depth and 3D structure.
- The eye's ability changes the optical power to maintain a clear vision of the object when the distance varies. This can be done by changing the lens form naturally or changing the distance between the rigid lens and the retina by muscle.
- Linear perspective is the eyes' capability to create the illusion of depth on a flat surface. In linear perspective, the orthogonal line joins at infinity.
- The capability of the eyes to estimate the distance from the object. The vertical position of the object on the horizon is always perceived as too far by the eyes.
- The eyes' ability to estimate the distance due to the change of visibility of view, saturation, and modification of colour..
- The ability to estimate the distance due to the change of appearance and texture size.

In computer vision, the two images that can acquire by using stereo vision are:

1. Using two different cameras, the cameras are usually placed side by side horizontally with a baseline between them, similar to human vision. The two cameras can also be placed horizontally from each other and observe the same point in space to obtain two different view on the same object.

2. Using a single camera. The same camera takes pictures of the same scene from two different positions.

(a) (b)

Figure 2.23: The ZED (top) and Bumblebee (below) stereo vision cameras. The new generation stereo vision camera captures high-resolution images. On the left are images captured from the ZED camera that display high-resolution from both left and right cameras, the 3D point clouds, and disparity image.

Returning to the stereo vision system, if no additional lighting is required, this technique is known as passive stereo vision [76]. Otherwise, it is an active stereo vision. This technique is used to perceive depth information by generating disparity maps to detect obstacles in the environment (see Figure 2.24). It is a classical technique that facilitates robot for localisation, navigation, and obstacles detection for the past several decades [77]. An omnidirectional camera with low resolution can be used as a stereo vision camera [78]. The development of high-resolution optical sensors and dedicated processing units helped the engineers and researchers to design and develop better stereo vision camera for robotics and other relevant applications. Lately, the release of stereo vision camera Bumblebee2 and ZED has enabled researchers to gain high-resolution 3D depth-sensing [79] (see Figure 2.23).

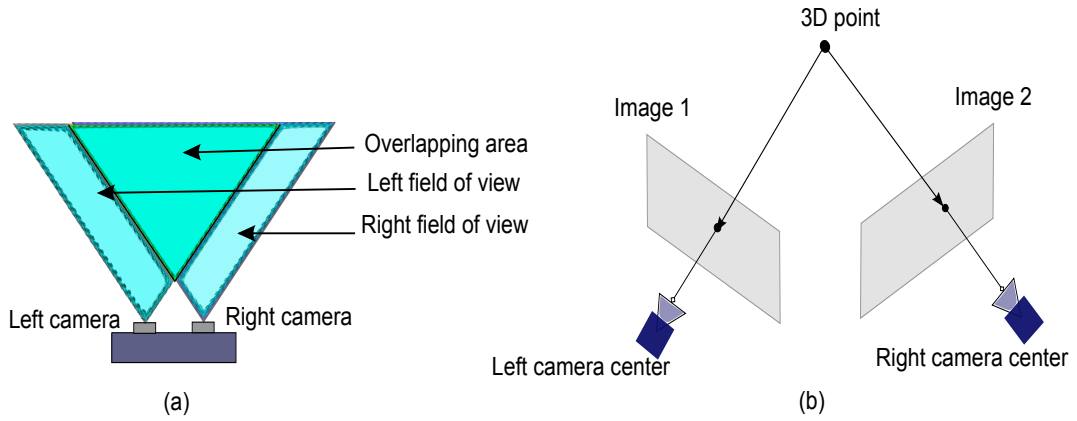


Figure 2.24: View from stereo vision camera. The superimpose view is binocular view that can perceive 3D objects and depth accurately.

2.6/ SINGLE VIEWPOINT

2.6.1/ PERSPECTIVE CAMERA

Perspective Camera: The cameras have a single viewpoint (SVP) if all the incoming light intersect at one single point. This term is also used for combining several cameras, where all the cameras share a common intersection point. The pinhole model or perspective model is a classical model in the state-of-the-art in forming an image. It has a geometry characteristic and image plan, where the image is formed by the perspective projection. It was discussed in Section 2.3.

This principle is depicted in the case of the pinhole camera model. As presented in Figure 2.6, the ray from pinhole projected on the image plane. Adding a lens in front of a pinhole camera can improve the precise position of the ray light at the intersect point. This intersection point also known as the focal point of the lens or the focus distance from the lens's centre to the intersection point.

2.6.2/ OMNIDIRECTIONAL CAMERA

Omnidirectional Camera: The omnidirectional camera has SVP if it can gather the rays light to a single point. The rule is the same as the perspective camera. The omnidirectional images are formed by a mirror with a conic section or lens. In the case of a catadioptric camera, the conic mirror placed in front of the camera. This placement specified to the SVP characteristic. In 1974, E. Hect and A. Zajac [22] explained that catadioptric is an optical vision system combined with several mirrors and lenses. The following subsection discusses the shape of the mirror invented in the early research of the omnidirectional camera.

2.6.2.1/ PLANAR MIRROR

Planar mirror: The planar mirror is a feasible method to capture 360° images. The 360° images can be obtained by placing the planar mirror at the correct distance and angle from the normal of the camera. Baker and Nayar [80] found the solution, although the FoV was not enhanced. This work was continued by Nalwa [81] who constructed a mirror in the form of a pyramid. The idea presented by Nalwa is portrayed in Figure 2.25. It consists of four planar mirrors to form a pyramid with four perspective cameras for each section of planar mirror.

The cameras captured a high-resolution 360° FoV with SVP. Currently, the idea is used by FullView Inc. This company manufactures cameras that provide unrivalled, live, and up to 360° panoramic video.

Kawanishi et al. [82]., in 1998, reviewed the work by Nalwa and proposed a stereo vision from an omnidirectional camera. He used two pyramid mirrors in the form of a hexagon and placed them back to back so that each pyramid can see the same scene. Hexagon means each pyramid faced the six calibrated cameras with unique SVP. The system captured the video sequence with synchronisation between the cameras. Stereo vision is obtained from the overlapping FoV between the two hexagon pyramids. The wide vertical FoV proposed by Hua and Ahuja [83] is the other extended work by Nalwa. A double mirror pyramid was placed back to back to acquire a 360° FoV horizontally, besides increasing the vertical FoV. Another researcher proposed another modification on the pyramid mirror by producing multiple viewpoints using a single pyramid mirror. The

idea proposed by Kar Han [84]. It retrieved depth information from a single pyramid mirror.

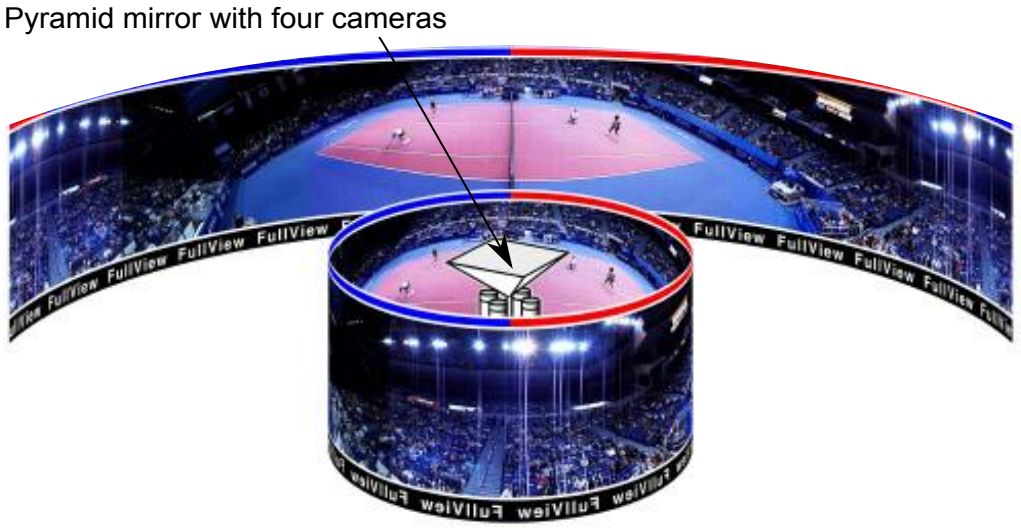


Figure 2.25: The mirror pyramid system. Four perspective cameras are placed upright, where the cameras' lens is placed in a vertical position towards the inclined mirror surface. The combination of images from all four individual cameras produces a panoramic image. The image is courtesy of *www.fullview.com*.

To reduce the cost, reducing the number of cameras should be the right decision. Besides, it made the system more compact and practical for robotic application. A conic mirror coupled with a conventional camera is the solution. Nene and Nayar [85] designed several forms of a mirror, including a planar mirror and used only one conventional camera instead of multiple cameras, to retrieve a 360° FoV (see Figure 2.26).
The advantages of this technique reduced the number of cameras and decreased data acquisition. These advantages outlined by comparing with the multiple cameras vision system. Gluckman and Nayar [86] in 1998 used reflection from two mirrors to increase FoV. The system simplified capturing stereo vision images from a single conventional camera. Since the system used only one conventional camera, the internal parameters were similar for the two views.

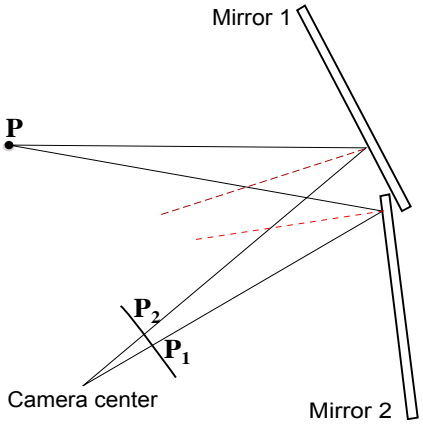


Figure 2.26: The formation of omnidirectional image from a single camera and two planar mirrors. Point $\mathbf{P}$ is reflected to points $\mathbf{P_1}$ and $\mathbf{P_2}$ by mirrors 1 and 2.

A curved mirror is considered to have SVP if the characteristic is approach to the conic section. A conic section is a curve generated by the intersection of a plane to the nappes of cones. The nappes of cones are described as the combination of two cones placed apex to apex. If the plane is perpendicular to the cone axis, a circle is produced. If the plane is not perpendicular to the axis and intersects only one cone, a parabola or ellipse is produced. If the plane is parallel to the cone axis and intersects both cones, a hyperbola is produced (see Figure 2.27, Figure 2.28, Figure 2.29 and Table 2.1).

The conic section can be defined by the value constant $\mathbf{e}$, eccentricity, which is the parameter associated with conic section.

Conic section		
Type of mirror	*Eccentricity*	*Range of essentricity*
Circle	0	0
Ellipse	$\sqrt{1-\frac{a^2}{b^2}}$	0 < e < 1
Parabola	1	1
Hyperbola	$\sqrt{1+\frac{a^2}{b^2}}$	> 1

Table 2.1: Conic section with the associate parameter $\mathbf{e}$ (eccentricity). Where $\mathbf{a}$ is half of the minor axis of the shorter diameter of an ellipse. $\mathbf{b}$ is half of the major axis or the longest diameter of ellipses (see Figure 2.28)

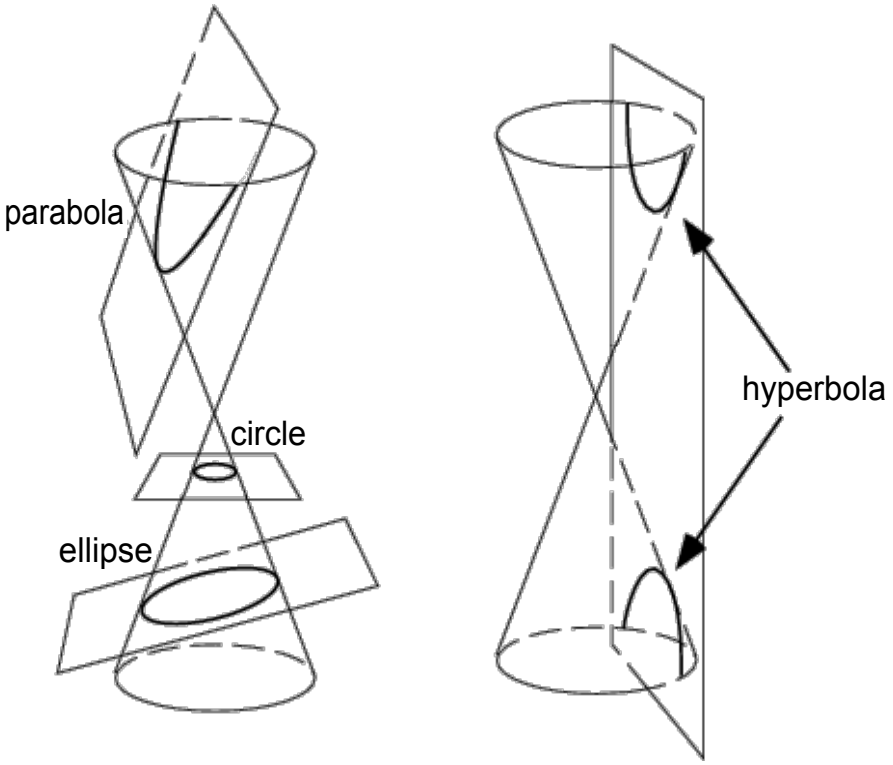


Figure 2.27: Double cones placed apex to apex. A rectangular plane intersects one of the cones and produces an ellipse or a parabola. A hyperbola mirror is produced when the rectangular plane intersects both cones in the direction of vertical. The image is courtesy of *WolfframMathWorld*

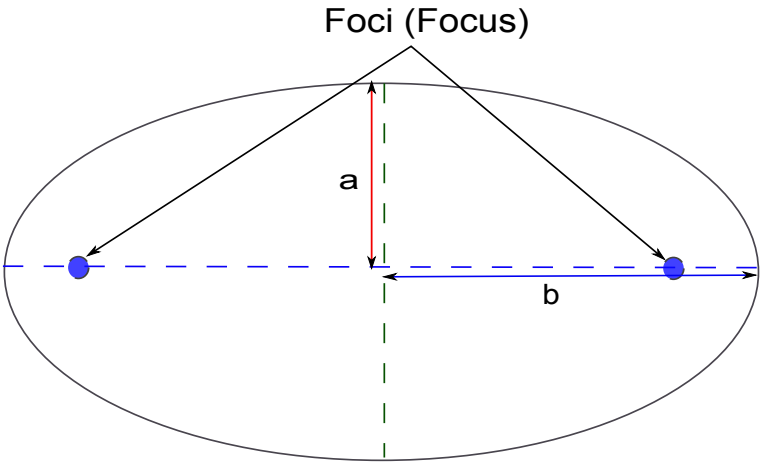


Figure 2.28: The conic parameters in the case of ellipses to a parabola that rotates on symmetrical axis.

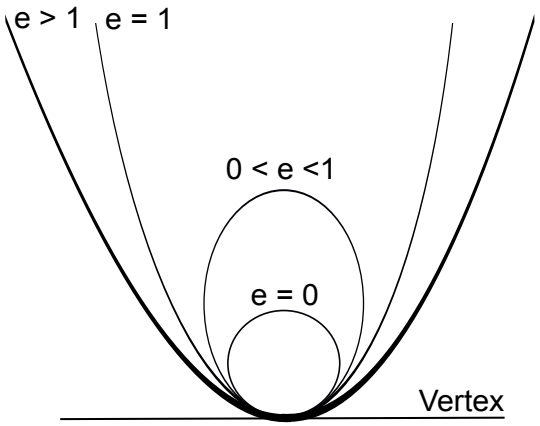


Figure 2.29: Conic-section in function of the varied values of the parameter eccentricity. It commences with a circle, where the value of eccentricity, $\mathbf{e = 0}$. The location of focal points changed with the change of the value of eccentricity.

2.6.2.2/ PARABOLIC MIRROR

Parabolic mirror: A parabola is the arrangement of all points situated at an equal distance from the directrix. The directrix is a fixed-line that describes any form of a curved mirror. The distance between directrix ($\mathbf{L}$) and the vertex is equal to the distance between vertex and foci. The focus parameter is the distance between directrix and foci or focal point, where $\mathbf{f = 2a}$ (see Figure 2.30). The term paraboloid referred to a parabola which rotates on the symmetrical axis.

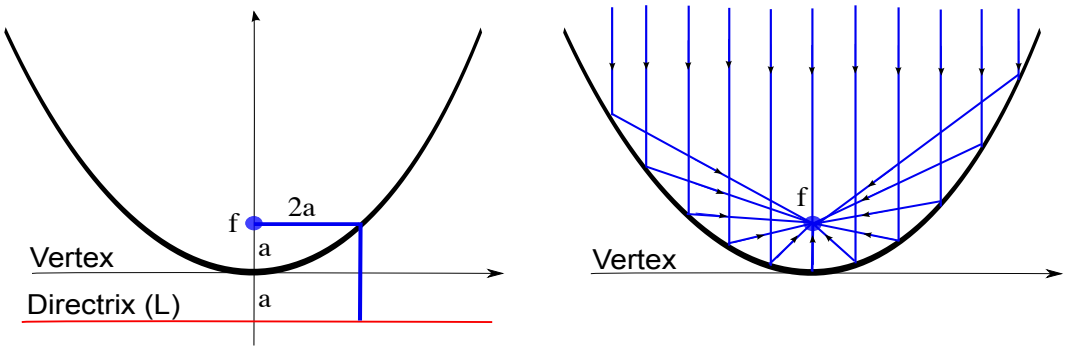


Figure 2.30: The parabolic mirror parameters. The directrix ($\boldsymbol{L}$) is a straight line perpendicular to the parabolic axis, which is used to differentiate the mirror type. Vertex is the highest point of a parabola or a meet point for the left and right curves. The parallel ray of light is reflected from the parabola and concentrates on point $\mathbf{f}$.

The principal of parabolic is explained in the beginning of Chapter 2, in an ancient book of ***"On the Burning Instrument"***. Due to the geometric properties of parabolic, it reflected the incoming light parallel to the axis, to the concentrated point or the focus point. The function is the same as the parabola of a satellite television signal, by receiving the signals and gathering all the signals on the receiver (focus).

This perspective view is depicted in Section 2.2. The perspective is the default view where one sees a distant object as small. The orthographic view means representing the 3D to 2D. In the orthographic view, the object's scale is preserved, and all objects appear on the same scale independent of object position in space. The orthographic view offers advantages in front of the perspective view. The geometrical mapping within the image on the mirror and the world point is independent of the mirror's translation concerning the camera system. The reduction in parameter simplified the perspective calibration process using orthographic projection.

Several methods have been proposed to obtain orthographic projections. This method, such as in the commercial of the telecentric lens by Edmund Optic [87] [88]. Watanabe et al. [89] [90] [91], used optical configuration to solve magnification problem by referring to telecentric optic. The solution is by adding an aperture to arrive at a precise displacement from the lens. The author noted that the telecentric had been initiated a couple of years back by referring [92] [88]. However, it is still underexplored within the computer vision domain.

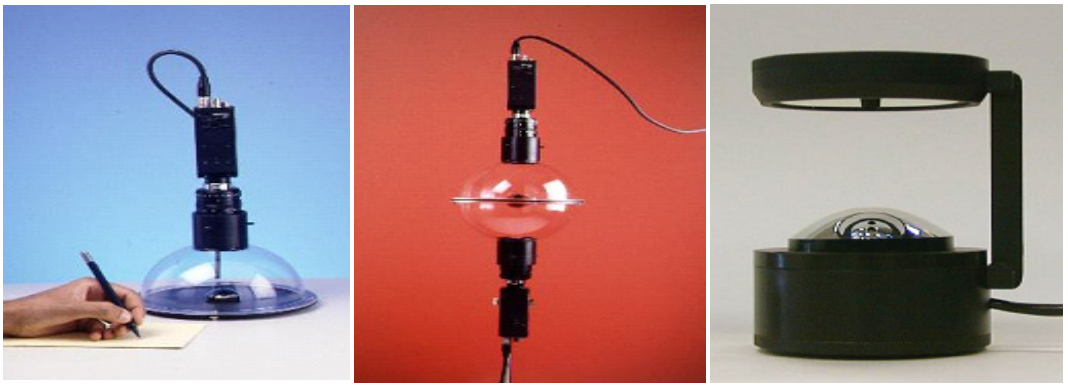


Figure 2.31: Use of a parabolic mirror to increase FoV. A single-camera produces 180° or hemisphere and 360° on the horizontal axis. The combination of dual-camera produces 360° vertical and horizontal or sphere. The distance between the cameras is short, which is ideal for an orthogonal camera. The image is courtesy of S. K. Nayar.

The ego-motion is defined as the 3D motion of a camera within the rigid environment. The ego-motion algorithm is existed for the planar perspective camera by exploiting

the motion field equation. Gluckman et al. [67] used the same algorithm for an omnidirectional camera. It based on the parabolic mirror to perform the ego-motion estimation. They used the pipeline by mapping the optical flow field on the sphere. In [93], the same authors used an omnidirectional camera as a compact panoramic stereo based on the parabolic mirror. The telecentric optics is used to attain orthographic projection while keeping the camera close to the mirror. The two omnidirectional cameras are placed vertically and both camera axes were aligned. The cameras captured the same panorama image with varied heights, which is used for stereo calculation.

Nayar and Peri 1999 [94] used the parabolic mirror to construct a folded catadioptric camera with SVP (see Figure 2.31). The paper explained the several parabolic mirrors to reduce the constraint of the size sensor. It discarded the useless optical effect (chromatic aberration, coma, and astigmatism) due to a large mirror's deflection. The system produced a more compact camera system with wide FoV, which proved to have the geometrical equivalent to a system using a single conic mirror.

The number of mirrors reduced the production of undesired rays of light, which can cause a reflection from the mirror, thus fail to converge on a single point. The unwanted light, which is viewed as the tangent of the surface, is called caustic. For this purpose, the second mirror is used to reflect this ray and unite at the SVP. Previously, the telescope and microwave devices implemented the system with several mirrors to reduce the unwanted radiation [95]. The folded panoramic system used two parabolic mirrors, and the method is described in the provided conic dictionary.

Sturm et al. [96] applied a strategy to achieve 3D reproduction from a single panoramic image. It procured from an omnidirectional sensor formed by a parabolic mirror and an orthographic camera. It placed the camera perpendicular towards the parabolic mirror. The view of the camera is parallel to the axis of the parabolic mirror. Nayar et al. further explained the geometry of the paraboloid mirror.

The orthographic projection can be simplified by placing the viewpoint on the focus point in the case of a paraboloid mirror. The method can interpreted the calibration and perspective calculations. Looking back at Sturm et al., the 3D reconstructions is performed using geometrical restrictions, such as coplanarity of points, perpendicularity of planes, as well as lines and parallelism of planes and lines. The calibration methods were applied by placing a circle to the border of the reflected image, and the reconstruction was obtained by using the properties of SVP of a catadioptric mirror. It allowed the

reprojection of 2D image along the circle line. The study proposed a stable method to obtain a 3D point using a parabolic mirror. The algorithm can be used with other types of a mirror within the range of the conic section.

2.6.2.3/ HYPERBOLIC MIRROR

Hyperbolic mirror: Reef [97] had invented a 360° viewing device using hyperboloidal ellipsoidal reflecting optic in 1970. The design, which was called a panoramic television view system, was initiated to be installed in vehicles. The goal is to increase the view for the vehicle operator. The patented device offers 360° FoV in real-time to the surrounding area of a vehicle, which can access a display unit inside the vehicle or a remote room. It became the first-ever application of 360° camera that was applied in the armoured vehicle.

The hyperbolic mirror is a conic section. It is defined as the locus of all points on the curve in the plane. The difference in the distance between any point on a hyperbola and two fixed points is constant. The two points are called the foci of a hyperbola, which are the same as other mirror types. The constraint of SVP should be valid once the camera and the mirror viewpoints are placed on the two foci. The hyperbolic mirror is used widely for the catadioptric camera. Svoboda et al. [98] used a system that combined a perspective camera with a hyperbolic mirror.

(a) (b)

Figure 2.32: Conic mirror. Hyperbolic mirror (a) and spherical mirror (b) manufactured using aluminium metal. They are formed using high precision CNC lathe machine. High spindle rotation and low feed speed can produce a smooth and shiny surface. The image is courtesy of Grassi et al.

The research goals were to give a foundation 360° FoV for stereo vision. The authors had proven that the hyperbolic mirror is conic with SVP. The study propelled by seeking improvement of motion estimation from two images. Yet, the outcomes may be used for structure reconstruction from panoramic stereo images. Grassi et al. [99] combined lenses and mirror to achieve a 360° FoV. The authors used two types of mirrors; a spherical mirror for the primary test and a hyperbolic mirror for the real tasks. The hyperbolic mirror was made of aluminium metal. It was formed using high precision Computer Numerical Control (CNC) machine (see Figure 2.32).

The Hyper Omni Vision [100] [101] [102] [103] is a multiple vision system that consists of a catadioptric camera to capture panoramic images and two PTZ cameras to capture stereo vision image. The catadioptric camera combines a CCD camera and a hyperbolic mirror (see Figure 2.33). The authors highlighted the low resolution of the camera system. They solved the problem by adding a high-resolution camera to generate a local perspective view.

Figure 2.33: A hyper omni vision omnidirectional camera. The camera combines a perspective camera and a hyperbolic mirror. The prototype was developed at Osaka University, Japan. This image is courtesy of Yagi et al.

The problem of low resolution is common [22] [104] among several omnidirectional cameras, especially for a single camera system. The image formation on a catadioptric camera depends on the surface of the mirror. In this case, the number of pixels on the perspective camera remained the same on the omnidirectional mirror's azimuth resolu-

tion. Besides that, the amount of reflected light changed the direction on the curvature mirror (see Figure 2.34). Fernandes et al. [105] asserted that the angle of reflection on the curved mirror was twice when compared to the plane mirror.

Nagahara et al. [106] improved Hyper Omni Vision's resolution by using the method of sub-pixel displaced and multi-focused images. The catadioptric camera based on a hyperbolic mirror was used in several applications. Yamazawa et al. [107] mounted the Hyper Omni Vision on a robot. He developed a strategy to assess the movement of the robot in discovering unfamiliar obstacles. The same authors exploited the prototype for motion estimation in real-time.

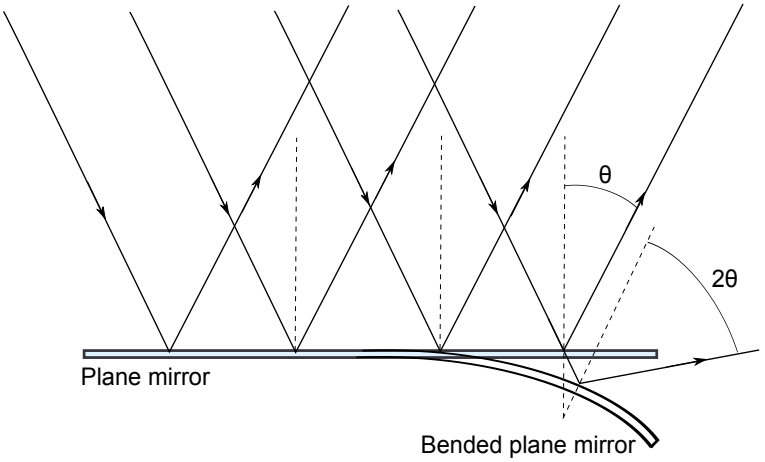


Figure 2.34: The reflected ray light on the plane mirror has changed the direction if the plane mirrors bent downward.

In 2009, Barzack et al. [108] used the catadioptric system that consisted of a perspective camera combined with a hyperbolic mirror for face detection and tracking using the Viola-Jones method. The face tracking method used the region of interest (ROI). The panoramic was extended horizontally to avoid losing and duplication detection. The authors claimed that the hyperboloidal mirror has SVP (or single projection centre). The image resolution was uniformly distributed along the mirror periphery.

2.6.2.4/ ELLIPSE MIRROR

Ellipes mirror: An ellipse is described as the analytical sets of points for which the total of its distances to two foci $(\mathbf{F_1}, \mathbf{F_2})$ is a fixed number (see Figure 2.35). An ellipse

derived from a circle, its structure, and its form depend on the value of eccentricity **e**.

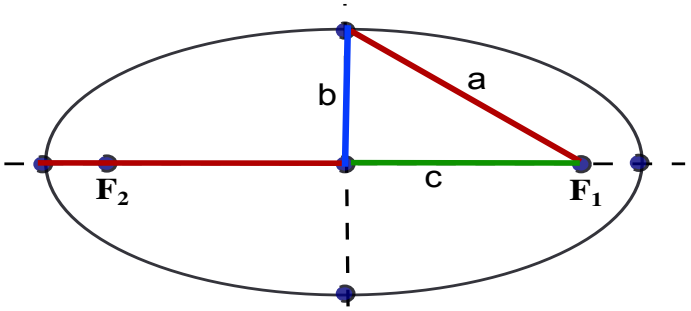


Figure 2.35: The diagram of an ellipse. Where **a** is a semi-major axis and **b** is a semi-minor axis. **c** is a linear eccentricity. The eccentricity is a non-negative value that characterises a shape or a form. The values of eccentricity are tabulated in Table 2.1.

$$\mathbf{e} = \sqrt{1 - \left(\frac{a}{b}\right)^2}, \tag{2.25}$$

Referring to Figure 2.36 the eccentricity of ellipse is defined in Equation 2.25. The form of ellipses varied from the value of **e=0**, where **a** and **b** are equal. If the ellipse is rotated along the symmetry axis, it is called an ellipsoid. The use of such a mirror is rare if compared to other types of mirror. Even if the ellipse meets SVP constraint characteristics, ineffective FoV is obtained due to the reflective part of the ellipse located at the interior part of the mirror. Thus, the camera is placed to face the reflective part of the ellipse.

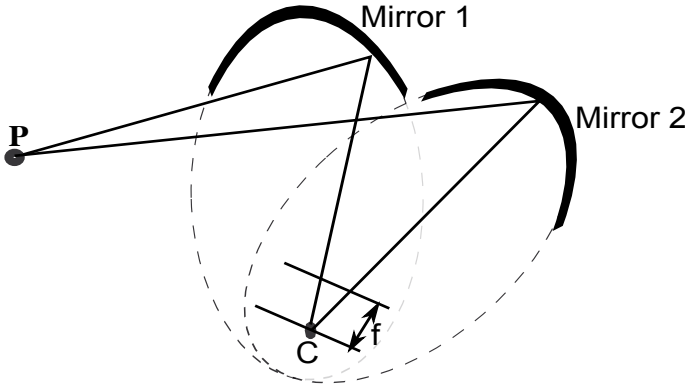


Figure 2.36: Point **P** in space observed by two ellipsoid with different orientations using a single camera (C).

Nene et al. [85] proposed a camera system that achieved a wide FoV using a single

perspective camera. They examined the four types of mirrors that met the SVP constraints. The types of mirror are planar, ellipsoid, parabolic, and hyperbolic, to observe what the cameras had seen. In conclusion, to realise the SVP constraint, the combination between the camera system and the types of the mirror must be correct. This point is discussed literally in several publications, and the summary is provided in Table 2.2.

Coupling between types of camera and mirror	
Type of mirror	*Type of camera*
Plannar	Perspective
Parabolic	Orthography
Hyperbolic	Perspective
Ellepse	Perspective

Table 2.2: The coupling between camera and mirror type that satisfies SVP constraint. The majority is perspective camera, except for parabolic mirror that used an orthographic camera.

2.7/ NON SINGLE VIEWPOINT

The Non-Single viewpoint (NSVP) is the primary set of all types of a mirror for a dioptric and catadioptric camera. The SVP mirror is a subset of NSVP (see Figure 2.37). A camera system has a unique viewpoint if the camera centre is placed on the mirror's focal point. The catadioptric system with a conic section mirror should satisfy the condition of the SVP constraint. Several conditions is considered and addressed. For instance, the pinhole camera aligned to the symmetry line of the parabolic mirror. Otherwise, it would not satisfy the SVP constraint and produce a bad image with undesired distortion. The solution of the equation in [80] represents all types and classes of the mirror in light of SVP constraint. However, specific solutions only covered the theoretical aspect for all available mirrors, such as planar, conical, spherical, ellipse, hyperbolic, and parabolic.

The SVP determined if the reflected ray of light intersects at one point. For example, for the hyperbolic mirror, the SVP constraint is satisfied if the camera centre is placed on the hyperbolic mirror's foci. Jeng et al. [109] proposed a method to identify a set of analytic equations for the unwrapping images obtained from a hyperbolic mirror. To verify the determined equations' accuracy, a fraction of the framework parameters were acclimated to fit the SVP imperative. At the same time, unwrapped images that used

the subsequent streamlined camera model had no distinction from those acquired by a technique dependent on the SVP model. According to Jeng et al., the image unwrapping problem was noted for hypercatadioptric camera in NSVP condition. The hypercatadioptric camera is the combination of a hyperbolic mirror with a perspective lens. Another term is called paracatadioptric, which is the combination between a parabolic mirror and an orthographic lens.

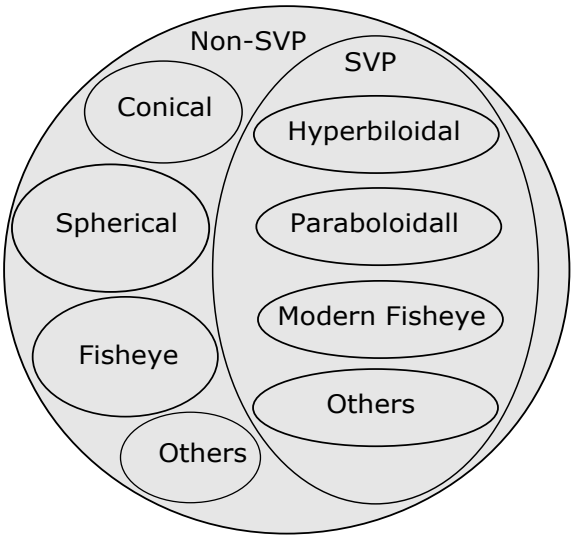


Figure 2.37: The classification of mirrors for the omnidirectional camera. All mirrors are considered to have NSVP. The SVP is classified under NSVP if not fixed with the condition of the SVP constraint. For example, a hyperbolic mirror is out of SVP classification if the camera centre is not aligned with the focus point. The description of each mirror is explained in the following.

The conical, spherical, and most fisheye cameras do not have SVP but instead have a locus of projection centre or diacaustic. The diacaustic is the interaction of a ray of light with a refractive surface (dioptric or fisheye camera). The catacaustic is the interaction of a ray of light with a reflective surface (catadioptric camera).

However, Baretto et al., [110] [111] demonstrated that SVP could approximate the small locus of dioptric systems. Thus these systems also underlie a standard central projection model. Courbon et al. [112] assessed the use of a unified model of the sphere for dioptric cameras in particular. It resolved that the unified model may be applied for fisheye cameras with an expansion to cover the radial distortion of the lens. However, it was mentioned in [113] that the modern generation of fisheye cameras is central. Thus,

the modern fisheye cameras can satisfy the SVP constraints.

2.7.1/ SPHERE SHAPE MIRROR

Sphere shape mirror: The spherical mirror has the SVP constraint if the camera centre is placed on the sphere's centre. Walter et al. [114] used the catadioptric system to combine a perspective camera and a spherical mirror. The system was able to correct some distortions on the scene. The images processed into two main unwrapped images; panoramic image and Birds Eyes View. Both the unwrapped images used for two navigation modalities; Topological Navigation and Visual Path Navigation.

The Birds Eyes Views is defined as how a bird looks and observes the ground from the sky. A bird's ability to see the outline of a particular area or to map has attracted researchers to replicate that unique ability. Sato et al. [115], and Liu et al. [116] used several fisheye cameras to obtain Birds Eyes Views.

Southwell et al. [117] used a catadioptric system with a perspective camera and two level of spherical mirrors. The two spherical mirrors differed in size and geometry (see Figure 2.38). The genius technique preserved two views at once from the double spherical mirror or double mirror lobes. The binocular view from the two lobes recovered the depth of the scene. Fiala et al [118] [119] extended the study of this camera system.

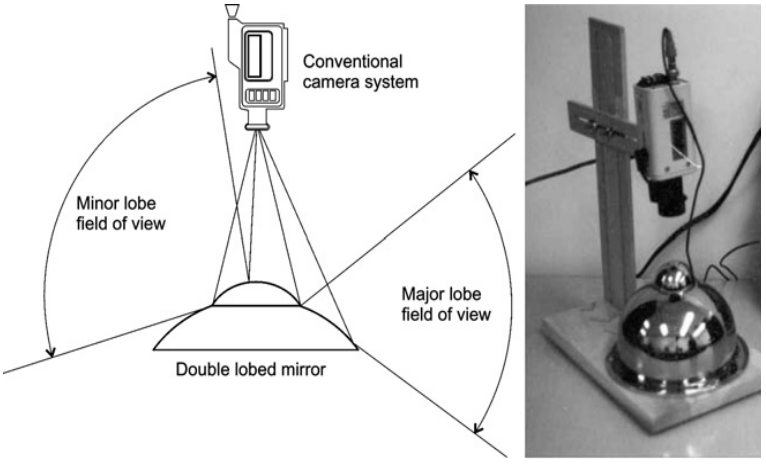


Figure 2.38: A catadioptric camera with a combination of perspective camera and double-lobe mirror. The two lobes have different size and geometry. The result is two different views that reflected a binocular view, a minor lobe FoV from the small lobe and a major lobe FoV from the big lobe. The system captured two images at once for depth estimation and 3D reconstruction. The image is courtesy of Southwell et al.

They aimed to obtain 3D reconstruction using horizontal and vertical lines extracted from

the double lobes mirror. The horizontal line segment was derived using Panoramic Hough Transform. The vertical line segment was detected as straight radial lines. The matching between the two lobes provided 3D points.

Another robotic application using a parabolic catadioptric camera. It referred to the work by Hong et al. [120], and Zhang et al. [121]. Both studies used the same 360° catadioptric camera. It based on a parabolic mirror and perspective camera. The robot's navigation relied on the system, which obtained the mapping of the surrounding area to fulfil the homing task algorithm. Zhang et al. used features matching the current landmark as the current target with the next target view, which was adequate to extract rotation and translation for 3D reconstruction.

2.7.2/ CONE SHAPE MIRROR

Cone shape mirror: A conic mirror is classified as a NSVP, but Baker et al. [122] explained that a conic mirror has a SVP constraint if the camera centre is placed on the apex (summit) of the cone. Since SVP is located on the fixed point on the cone's tip, the conic mirror is unsuitable for developing wide FoV. The slight movement of the pinhole camera caused the SVP to fall on the locus of the cone.

Now, the cone is no longer on its SVP. Li et al. [123] [124] [125], had proven that the cone has SVP. However, SVP depends on the parameters of the compound optical characteristics. Thus, the researchers claimed that the other more extended camera should be used at the pinhole camera to address optical phenomena issues.

In 2003, Li et al. [126] used a prior set-up for robot navigation. The researchers claimed that the current trend of omnidirectional focuses on reducing the cost instead of enhancing the images' quality. They used the concept of a virtual camera, where the camera looks at the world point beyond the mirror. The omnidirectional consisted of a cone mirror and two perspective cameras aligned with the cone's symmetry. The result is a high-resolution images to capture 3D depth values in real-time. Thus, more explication of each set-up element, especially for the cone mirror, is explained in [127].

Li et al. 2006 presented the extensive imaging hypothesis about the cone reflected in its SVP set-up. They demonstrated that the SVP cone mirror catadioptric framework was not exclusively practical yet. The configuration had points of interest for specific applications. The hypothesis gave detailed clarification on the design set-up of the SVP

cone, the benefits and the faults of such frameworks, and how to correct and rectify in developing a functional imaging system.

The Conic Projection Image Sensor (COPIS) project is referred to the earliest work on the conic mirror. The project was led by Yagi et al. [128] [129] [130], and the yield was used in the navigation of a mobile robot. The system demonstrated wide FoV due to the use of the fisheye lens and a conic mirror. The application was meant to map the surrounding area using azimuth information extracted from radial lines. The COPIS project's enhancement is the combination with Multiple Visual Sensing Sensor (MISS) by Yagi et al. [102]. The system combined COPIS with two perspective cameras to obtain a binocular view.

The COPIS is used to understand the global map. In contrast, the binocular is used to understand the details of the fascinating local object. The combination of the two systems offered advantages to capture a scene with high resolution and observe the surrounding area. Southwell et al. [131] produced a catadioptric system for internal pipeline inspection. They used a conic mirror coupled with a perspective camera, which recorded images with 360° that displayed the pipeline's inner surface. They also examined the surface image using an image processing technique.

Pegard et al. [132] built a catadioptric camera using a conic mirror and a perspective camera. The goal was for robot navigation and robot positioning with the help of the target landmark. The vertical lines, such as edges of the door, cupboard, and building, were extracted using image processing tools - Sobel edges detection. The radial distortion was corrected using the available method.

2.8/ Omnidirectional Projection Models

Section 2.3 presented the projection of the pinhole camera model. The pinhole camera model described the relationship between mathematical model of points in the 3D world and the projection of points on the pinhole camera. The same projection model was used to explain the perspective camera. The projection model was expanded to the case of the omnidirectional camera. As depicted in Section 2.3, the presence of strong distortion is considered to yield the accurate and realistic results.

2.8.1/ PROJECTION MODEL CAMERA CATADIOPTRIC

As discussed in the previous subsection, S Baker et al. [133], tested four possibilities to obtain panoramic images using several types of mirrors. The researchers coupled mirror-camera combination to address SVP constraint, as summarised in Table 2.2.

Amongst the four sets of camera-mirrors combination, only two were realised practically and offered further interest for robotics application. The combinations were:

1. A hyperbolic mirror and perspective camera called hyper catadioptric camera.

2. A parabolic mirror and orthographic camera called paracatadioptric camera.

Thus, the two camera-mirror combinations were studied in detail. The camera-mirror configuration proposed a projection model for an omnidirectional camera, which led to the image's formation.

The projection of images on hypercatadioptric and paracatadioptric cameras had weighed in radial and tangential distortions that affected image formation. The reflection and the refraction laws were adhered to in defining the geometry of the camera-mirror combination. The formulation in mathematics appeared essential to determine the mirror and camera parameters, such as the focal point and the polar surface. Therefore, the selected papers [134] [135] explained the configuration in detail. Some of the mathematical equations were taken from the papers to support the explanation given in this section.

2.8.1.1/ HYPERCATADIOPTRIC CAMERA

As mentioned previously, the hypercatadioptric camera is a camera system with the combination of a perspective camera and a concave hyperbolic mirror. The hyperbolic mirror is the family of conic-section with two focal points. As described previously, the camera system has addressed SVP constraint if the centre of the perspective camera with focal distance, $\mathbf{f_p}$, is placed precisely at one of the focal points (foci) (see Figure 2.39).,

Where $\mathbf{P} = (\mathbf{X}, \mathbf{Y}, \mathbf{Z})^{\mathbf{T}}$ is a point in space which is projected to the surface of the mirror $\mathbf{P_M} = (\mathbf{X_M}, \mathbf{Y_M}, \mathbf{Z_M})^{\mathbf{T}}$ through point O, which is a mirror's effective viewpoint. $\boldsymbol{d}$ is a distance between the camera center to the mirror focal point. The characterisation of the

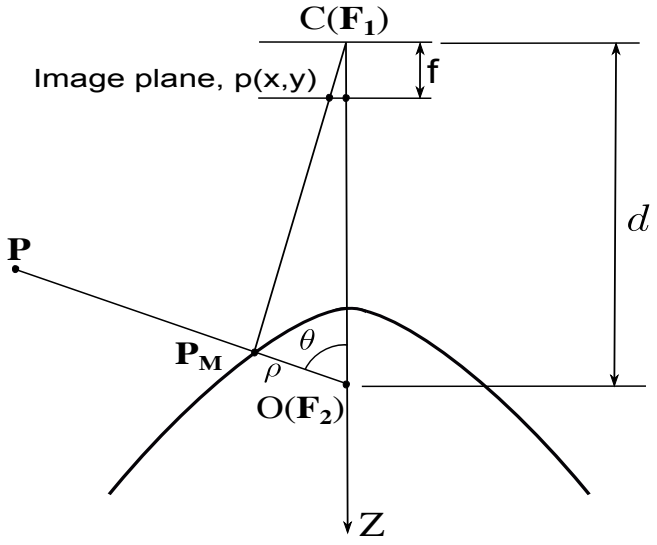


Figure 2.39: The configuration between a perspective camera and a concave hyperbolic mirror. The camera centre, $\mathbf{C}$, placed at the foci $\mathbf{F_1}$ towards the reflecting surface of the hyperbolic mirror.

mirror give by the polar equation that represents ρ,:

$$\rho = \frac{\mathbf{p}}{\mathbf{1 + e\cos\theta}}, \tag{2.26}$$

Where

- **e** is the eccentricity of conic-section for a hyperbolic mirror (**e** >1). See Table 2.1.
- ρ is the parameters to characterise the hyperbolic mirror, which is the distance between the focal point(foci) to the directrix of the conic-section. See figure 2.30.
- $\mathbf{\cos\theta}$ defined as,

$$\mathbf{\cos\theta} = \frac{\mathbf{Z}}{\mathbf{||P||}}, \tag{2.27}$$

$$\mathbf{\cos\theta} = \frac{\mathbf{Z}}{\mathbf{\sqrt{X^2 + Y^2 + Z^2}}}, \tag{2.28}$$

The 3D point on space, $\mathbf{P}$, projected on the hyperbolic surface is determined as follows:

$$\mathbf{P_M} = \frac{\rho\,(\mathbf{X, Y, Z})^{\mathbf{T}}}{\mathbf{\sqrt{X^2 + Y^2 + Z^2}}}, \tag{2.29}$$

The next step is to project the 3D point of the hyperbolic mirror $\mathbf{P_M(X_M, Y_M, Z_M)}$ onto the image plane on the perspective camera. The camera matrix that consisted of

the camera's intrinsic parameters is used to estimate the projection of 3D image to 2D image. Where the image plane coordinate $\mathbf{p(x, y)}$ is given below:

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{f} & \mathbf{0} & \upsilon_{\mathbf{o}} \\ \mathbf{0} & \mathbf{f} & \nu_{\mathbf{o}} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \boldsymbol{d} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{X_M} \\ \mathbf{Y_M} \\ \mathbf{Z_M} \\ \mathbf{1} \end{bmatrix}, \tag{2.30}$$

In order to confirm that the system has addressed the SVP constraint, the camera centre is placed on the second focal point of the hyperbolic mirror. The distance between the two foci had been forced so that $\mathbf{d} = \frac{\mathbf{2ep}}{\mathbf{1-e^2}}$, where $\mathbf{e}$ is the eccentricity of a hyperbolic mirror in conic-section and $\mathbf{p}$ is the parameter of the hyperbolic mirror in light of directrix. Based on all the information, Equation 2.30 is resolved as follows:

$$\mathbf{p} = (\mathbf{x}, \mathbf{y}) = \left(\frac{\frac{\mathbf{1-e^2}}{\mathbf{1+e^2}} \mathbf{f X_M}}{\frac{\mathbf{2e}}{\mathbf{1+e^2}} \sqrt{X_M^2 + Y_M^2 + Z_M^2} + Z_M} + \mu_o, \frac{\frac{\mathbf{1-e^2}}{\mathbf{1+e^2}} f Y_M}{\frac{\mathbf{2e}}{\mathbf{1+e^2}} \sqrt{X_M^2 + Y_M^2 + Z_M^2}} + \nu_o \right), \tag{2.31}$$

where μ_o and ν_o are the principle points of camera frame.

2.8.1.2/ PARACATADIOPTRIC CAMERA

A paracatadioptric camera is a combination of an orthographic camera and parabolic mirror. Such combination is discussed in the previous chapter, and it has two ambiguous configuration parameters. The two unclear configurations are focal point, f, and distance between the two foci, d, where both parameters are towards infinity ($\mathbf{f} \rightarrow \infty, \mathbf{d} \rightarrow \infty$) and the value of eccentricity, $\mathbf{e} = \mathbf{1}$. By considering the new configuration, the Equation 2.31 is extended as follows:

$$p = (x, y) = \left(\frac{p X_M}{\sqrt{X_M^2 + Y_M^2 + Z_M^2}} + \mu_o, \frac{p Y_M}{\sqrt{X_M^2 + Y_M^2 + Z_M^2}} + \nu_o \right), \tag{2.32}$$

2.8.2/ PROJECTION MODEL FISHEYE

The projection of an omnidirectional camera is crucial due to the presence of distortion. Plenty of publications have explained the characteristics of a fisheye camera. The discussion helped the researcher to address the problem before using this camera. For

example, T. Ho et al. [136], explained the optical function of a fisheye lens.

The implementation of this camera is weighed in heavy distortion. The dioptric camera or fisheye camera suffered from strong barrel distortion. Generally, the fisheye camera is classified as NSVP camera. Scaramuzza et al. [137] claimed that all modern fisheye cameras are central and can satisfy SVP. In some robotic applications, this camera is used after the image is projected on a unified spherical model. This unified spherical projection model is a suitable approximation and estimation method for robotic application. Ying et al. [125] proposed a unified model for central catadioptric camera and fisheye camera. The model presented by Ying had remodelled the fisheye camera to a central catadioptric camera with SVP. Courbon et al. [138], proposed a unified model for the catadioptric camera, which can be used for the fisheye camera. They suggested a method by including the distortion parameters. The technique involved projecting the image onto a unit sphere and adhering to the perspective projection on an image plane. Omar et al. [139] developed a projection model to control the translation degree of freedom in the visual sensing application by computing the invariant to rotational motion.

The fisheye camera projection model is equivalent to the pinhole camera model's projection in terms of the distance between a point in space called $\mathbf{p(x, y)}$ and the centre point of the lens. The correlation is depicted in the following for the fisheye lens and pinhole projection model. $\mathbf{x}$ and $\mathbf{y}$ are the coordinate points on the fisheye lens.:

$$\mathbf{r} = \sqrt{(\mathbf{x^2} + \mathbf{y^2})}, \tag{2.33}$$

$$\mathbf{r'} = \sqrt{(\mathbf{x'^2} + \mathbf{y'^2})}, \tag{2.34}$$

And x' and y' are the coordinate of pinhole camera image frame.

The model is not limited to the pinhole model. It can be modelled in another way, perhaps less complicated than the pinhole model. For example, Fitzgibbon et al. [195] proposed a method to estimate the value of r by mixing both the fisheye camera model and a pinhole camera model.

$$\mathbf{r'} = \mathbf{k_1}\frac{\mathbf{r}}{\mathbf{1 - k_2 r^2}}, \tag{2.35}$$

Another way to express the relation on the fisheye lens is:

$$\mathbf{r} = \mathbf{f}\boldsymbol{\theta}, \tag{2.36}$$

Where $\mathbf{r}$ is the angular distance. $\mathbf{f}$ is the radius of the circle, which is equivalent to the focal length of fisheye lens and the angle $\boldsymbol{\theta}$ is the incident angle.

2.8.3/ PROJECTION OF GENERAL MODEL

A non-central projection occurred when all light is not projected at one point (see Figure 2.40). There could be more than one focus point for the lens. The non-central projection may occur when more than one camera is used at a time to capture images. This orientation is called a multi-view system [140]. For this case, each camera does not have the same or uniform focus point. This camera network is difficult to get the evenly focused points. This situation caused the difficulty in creating geometry for image creation.

Several camera systems or polydioptric camera system is described in detail at the beginning of this chapter. It is a solution to the physical problem of the catadioptric camera which affected the quality of the resulting image. The installation of a catadioptric camera on the robot system, created another issue to a space constraint for installation. Besides, the camera system offers a hybrid camera system compounded more than a camera, regardless either central or non-central [141].

The approach in creating this camera system is complicated if compared to the other camera model. Thus, the calibration system for generic models applied for non-central camera systems [142]. For a generic model, each pixel represents a ray of light expressed in the line Plücker [143]. It contained six coordinates to display the direction of the light-emitting in the space.

2.9/ MULTIPLE VIEW GEOMETRY

Along with the vast development of computer and information technology, the 3D system also adhered to the same rhythm. The need for 3D object visualization is accessible in many applications, such as in animation and graphic. In the area of architecture, it used to save an old building in the form of 3D points for future reconstruction. The education and culture identification, as well as in virtual reality, have their interest in the development of 3D images.

The 3D offered the accurate documentation for restoring the old building if the build-

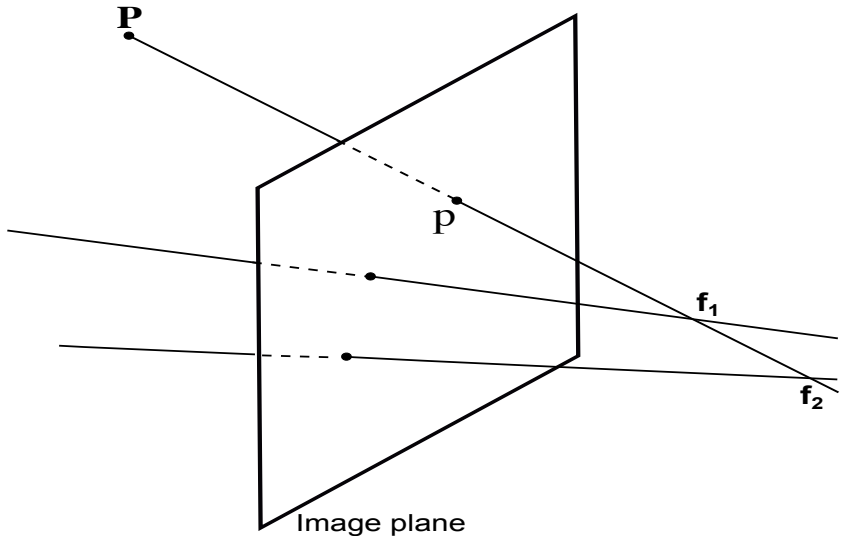


Figure 2.40: The projection of non-central with two focus points ($\mathbf{f_1}$ and $\mathbf{f_2}$) in the system.

ing has been demolished. The accuracy of 3D points is required in a lot of applications to exploit the related details and high quality of display output. The 3D reconstruction technique is divided into two categories, which are the passive and the active techniques. The active technique used a structured light. Dipanda et al. [144] used a camera and a projector to produce a structured light from a 2D image. A combination between laser beam and video camera is another active technique for 3D reconstruction. The passive technique used several images from different angles of view of the same object. Prakoonwit et al. [145] used several images from a perspective camera to reconstruct 3D points from a unique object.

The passive technique is also known as photogrammetry or a structure from motion. This economical method is preferred and more popular. The active technique used expensive instruments in the research rather than using several images from a camera for the same objective. This constraint channelled some studies into devising more straightforward and more cost-effective methods. Zhang et al. [146] used a digital camera to capture a set of images of human to reconstruct a human model based on the captured images. Earlier studies focused only on the calibrated camera. However, in the subsequent development, the technique has been implemented for the uncalibrated camera. The corresponding images can determine the camera parameters.

Sonka et al. [147], mentioned that the proposed 3D vision is to obtain 3D information from a 2D image, which is severe in geometry and radiometry. It discussed the problem of a single image that lacks information to reconstruct the 3D and the question

of radiometric related to the complex process of image intensity. The 3D vision discussed on how human looks at the 3D object through his eyes in a general view. The perception is essential, mainly to estimate the distance between the 3D object and human's eye. The algorithm of 3D reconstruction from the 2D image can be developed by understanding on how human eyes works.

This section discussed the geometry based on several views. These views captured from the numbers of cameras that look at a single object or the same camera but looking at the same object from the different location. Multi-view geometry is a vital tool to outline the 3D reconstruction algorithm from several cameras image formation. Be reminded that, if using two different cameras, the intrinsic or internal parameters of the two cameras should differ. The multi-view geometry performed the relation of the camera's position and the camera's intrinsic parameters to the unique target point in space. It describes the geometric relationship between the multiple views of scenes. Besides, this section related to the general principle of parameters estimation between the camera and point in space. The related parameters and camera properties was identified from the real-world images using state-of-the-art algorithms.

It is easy to understand with the two views called epipolar geometry. The explanation presented in this section is based on the perspective camera. It is also accurate and can be used for other types of lens or camera under the SVP family, as explained previously. The process is explained in detail in [148] and [149].

The 3D point is obtained by using all the information retrieved from the images, as mentioned previously. The number of the image should be at least two until N (N is the number of images). The image can be either static or dynamic. A static image is a non-moving object configured as the two views or n-view stereo configuration. Meanwhile, the dynamic is the moving image at different places and configuring as the structure from motion. The static and dynamic configuration work with the same principle is called epipolar geometry, as described in detail in [148].

2.9.1/ EPIPOLAR GEOMETRY

In 3D reconstruction, the epipolar geometry or the intrinsic projection geometry is the technique to determine the similarity between two images. If the camera looks at an object on the scene, it assumed that there is a geometry connecting among the observed

3D points from the 2D projection. The epipolar geometry depended to the value of intrinsic and extrinsic parameters. The camera needs to recalibrate in order to estimate the new parameters.

If $\mathbf{X}$ is a 3D point and the point is observed from two different angles of view, where the point on the first image is x and and the point 2D on the second image is $\mathbf{x}'$. Figure 2.41 shows the illustration of epipolar geometry. The points $\mathbf{x}$, $\mathbf{x}'$, $\mathbf{X}$, $\mathbf{C}$ and $\mathbf{C}'$ are coplanar, and all the points are located on the epipolar plane. All planes connected on the baseline, which connected the two centres of cameras. This plane called as the epipolar plane. If the only known point is $\mathbf{x}$ in the first image, how can it estimated the location of point $\mathbf{x}'$ on the second image?

The epipolar plane can be determined by the baseline and the reflected ray light that pointed to the centre of the camera, $\mathbf{C}$ through the image plane on the first camera, noted as $\mathbf{x}$. The amputation between the epipolar plane and the second image plane forms a line $\mathbf{l}'$, where the point $\mathbf{x}'$ located. The line $\mathbf{l}'$ is the epipolar line associated with the point $\mathbf{x}$ on the first image. The epipolar geometry is used to simplify the matching for 3D reconstruction and narrow the area in seeking the corresponding points [150]. The epipolar geometry can be applied to estimate the location of the camera and image rectification [151].

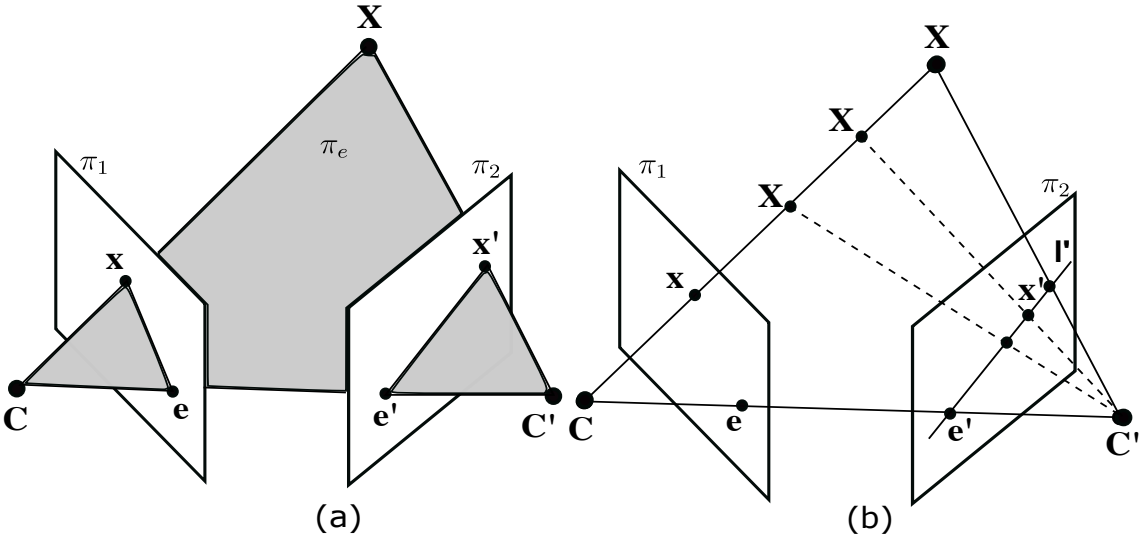


Figure 2.41: Epipolar geometry. Points $\mathbf{C}$, $\mathbf{x}$, $\mathbf{X}$, $\mathbf{x}'$ and $\mathbf{C}'$ are coplanar and all the points located on the epipolar plane.

2.9.1.1/ FUNDAMENTAL MATRIX

The geometrical information from uncalibrated cameras or the multiple viewpoints is vital data if one wishes to recover epipolar geometry. The information is contained in the single key-point called a fundamental matrix. A fundamental matrix is epipolar geometry in the form of a mathematical approach with $\mathbf{3 \times 3}$ array.

It is beneficial to identify the correlation between the views for 3D reconstruction. It started with the 3D equation derived from the two views of images. The projection of 3D point $\mathbf{P(XYZ)}$ on two planes images $\boldsymbol{\pi_1}$ and $\boldsymbol{\pi_2}$, in which the two points are noted as $\mathbf{p_1}$ and $\mathbf{p_2}$, can be associated using projection matrix. The relation of the second camera that refers to the first camera is displayed in the following equation 2.37 and equation 2.38 .

$$\mathbf{p_1 = M_1[P\ 1]^T \rightarrow p_1 = K_1[I|0][P\ 1]^T,} \tag{2.37}$$

$$\mathbf{p_2 = M_2[P\ 1]^T \rightarrow p_2 = K_2[R_{12}|t_{12}][P\ 1]^T,} \tag{2.38}$$

The two equations above show the geometrical relation between the first and second cameras. Here, $\mathbf{R_{12}}$, $\mathbf{t_{12}}$ are the rotation and the translation, respectively, between the first and second cameras. The reference camera is the first camera. Both $\mathbf{K_1}$ and $\mathbf{K_2}$ are intrinsic parameters matrix of the two cameras, which contain the internal parameters of the camera, such as principle point, focal point, distortion, etc. (see Chapter 2, Section 2.3.1). In the case of using the same camera with a different position, the value of intrinsic parameters is equal, where,

$$\mathbf{K_1 = K_2,} \tag{2.39}$$

The relative posed between the two cameras are defined by the epipoles on the images or cameras. It explained the projection of an image from one camera to the other camera. The epipoles are easy to describe as the intersection of the baseline with the image planes. Figure 2.41 shows the illustration of epipolar geometry. The epipoles marked as $\mathbf{e}$ or $\mathbf{e_{12}}$ and $\mathbf{e'}$ or $\mathbf{e_{21}}$ are linked to the centre of the cameras. It connect the first camera's respective projection centre to the image planes on the second cameras and, inversely, from the second camera's projection centre to the image plane on the first camera. The relation between the projection centre and the correspondent image planes

is given by:

$$\mathbf{e_{12}} = \mathbf{K_1 R^{T_{12}} t_{12}}, \tag{2.40}$$

$$\mathbf{e_{21}} = \mathbf{K_2 t_{12}}, \tag{2.41}$$

Where $\mathbf{e_{12}}$ is the epipole on the first image plane and $\mathbf{e_{21}}$ is the epipole on the second image plane.

Equations 2.37 and 2.38 are used to project the 3D points. The equations are simplified and expressed in another manner. Let's say the point $\mathbf{p_1 = (x_1, y_1, 1)}$ and $\mathbf{p_2 = (x_2, y_2, 1)}$ are homogeneous coordinates and $\mathbf{P}$ is the point in space, so:

$$\mathbf{K_1^{-1} p_1 = P}, \tag{2.42}$$

$$\mathbf{K_2^{-1} p_2 = R_{12} P + t_2}, \tag{2.43}$$

Substituting Equation 2.42 to Equation 2.43,a single equation is generated to represent both projection centre and image planes.

$$\mathbf{K_2^{-1} p_2 = R_{12} K_1^{-1} p_1 + t_{12}}, \tag{2.44}$$

Equation 2.44 s simplified by using the anti-symmetric matrix on $\mathbf{t_{12}}$. At this step, the relation of Equation2.44 depends only on the cameras' position. This step completed and simplified the dependency of the mathematics relation of two correspondent points. The anti-symmetric matrix or skew-symmetric matrix of $\mathbf{t_{12}}$ is:

$$\mathbf{t_{[\times]}} = \begin{bmatrix} \mathbf{0} & \mathbf{-t_z} & \mathbf{t_y} \\ \mathbf{t_z} & \mathbf{0} & \mathbf{-t_x} \\ \mathbf{-t_y} & \mathbf{t_x} & \mathbf{0} \end{bmatrix}, \tag{2.45}$$

for all the vector components $\mathbf{(x, y, z)}$ of $\mathbf{t}$. $\mathbf{(t = (t_x, t_y, t_z))}$, then the equation becomes:

$$\mathbf{t_{12[\times]} K_2^{-1} p_2 = t_{12[\times]} R_{12} K_1^{-1} p_1}, \tag{2.46}$$

The equation is simplified by multiplying it with $\mathbf{P_2^T K_2^{-T}}$, and this new mathematical relation is called **epipolar constraint**.

For more detail, Equation 2.46 elaborated as follow:

$$\mathbf{p}_2^{\mathrm{T}}\mathbf{K}_2^{-\mathrm{T}}\mathbf{t}_{12[\times]}\mathbf{R}_{12}\mathbf{K}_1^{-1}\mathbf{p}_1 = \mathbf{0}, \tag{2.47}$$

the equation is simplified as:

$$\mathbf{p}_2^{\mathrm{T}}\mathbf{F}_{12}\mathbf{p}_1 = \mathbf{0}, \tag{2.48}$$

Where,

$\mathbf{F}_{12} = \mathbf{K}_2^{-\mathrm{T}}\mathbf{t}_{12[\times]}\mathbf{R}_{12}\mathbf{K}_1^{-1}$ is a 3×3 matrix, and it brings together several properties for matching between the point from two different views.

The Properties of fundamental matrix

The fundamental matrix has a rank-2 matrix. It has several useful properties to describe the epipolar geometry in algebraic representation.

1. **Correspondences condition**. The fundamental matrix satisfies all pairs of points between two images. In detail, point $\mathbf{p_2}$ on the second image lies on the epipolar line $\mathbf{l'} = \mathbf{Fp_2}$ on the first image. The two points are coplanar and lie on the same epipolar plane. It helps to find the corresponding points on the other images. For the opposite case, any point on the second image $\mathbf{p_2}$ that corresponds to that point lies on the epipolar line $\mathbf{l} = \mathbf{F}^{\mathrm{T}}\mathbf{p_2}$ on the first image.

2. **Transpose**. If a fundamental matrix $\mathbf{F}$ is for the correspondence point $(\mathbf{p_1}, \mathbf{p_2})$, on the opposite case, the fundamental matrix of $(\mathbf{p_2}, \mathbf{p_1})$ is $\mathbf{F}^{\mathrm{T}}$.

3. **Epipolar line**. The epipolar line on the second image can be calculated using the fundamental matrix on the first image. $\mathbf{l'} = \mathbf{Fp_1}$. The epipolar lines on the first image are $\mathbf{l} = \mathbf{F}^{\mathrm{T}}\mathbf{p_2}$ and $\mathbf{l} = \mathbf{p_2F}$.

4. **Epipoles**. For any point $\boldsymbol{x}$ the epipolar line $\mathbf{l'} = \mathbf{Fx}$ contains the epipoles $\mathbf{e'}$, the point from the second image on the first image plane. Thus, $\mathbf{e'}$ satisfies $\mathbf{e'}^{\mathrm{T}}(\mathbf{Fx}) = (\mathbf{e'}^{\mathrm{T}}\mathbf{F})\mathbf{x} = \mathbf{0}$ for all $\boldsymbol{x}$. It follows that $\boldsymbol{e'}^{T}\boldsymbol{F} = \mathbf{0}$. $\boldsymbol{e'}$ is the left null-vector of F. Similarly, when $\mathbf{Fe} = \mathbf{0}$, $\mathbf{e}$ is the right null vector of $\mathbf{F}$.

5. **The determinant of $\mathbf{F}$ is always zero**. $\mathbf{det(F)} = \mathbf{0}$. This follow the fact that for all $\boldsymbol{n}\times\boldsymbol{n}$ matrices, the determinant obeys the multificative property : $\mathbf{det(A\dot{B}=det(A)\dot{d}et(B)}$. In $\mathbf{F} = [\mathbf{e'}]_{\times}\mathbf{p'p}$, the matrix $[\mathbf{e'}]_{\times}$ is a skew matrix where the determinant of a skew matrix is equal to zero. Hence, $\mathbf{det(F)} = \mathbf{0}$.

6. If $\mathbf{F}$ is the fundamental matrix for a given ordered pair of cameras or ordered pair position of single camera. The fundamental matrix become $\mathbf{F^T}$ if the ordered of cameras or position of camera reversed.

Step to calculate and estimate the fundamental matrix

The fundamental matrix is 3×3 matrices, which represented the geometry between two images. Languet Higgin introduced it (1981) [152], a British researcher who then prolonged his research on the essential matrix. He noted the difficulty to find the difference between the two images. It is the visual and the relation of mathematics between the two images that explain the images' precise orientation more accurately.

Later in 1992, the research was extended by Richard Hartley [153] and Oliver Faugeras [154] who proposed a solution for the fundamental matrix. The essential matrix is related to the 3D coordinates of the two images. The fundamental matrix is related to the image coordinates at those 3D points. In 1997, Hartley justified the fundamental matrix performance and proved the theory by experimenting with real images [155].

The 8-point algorithm, which was initially explored by Languet Higgin, is a method that is seldom used to calculate the fundamental matrix from the eight or more corresponding points. This algorithm has moderate benefits and easy for implementation. If $\mathbf{p_1} = [\mathbf{x_1}\ \mathbf{y_1}\ \mathbf{1}]^\mathbf{T}$ and $\mathbf{p_2} = [\mathbf{x_2}\ \mathbf{y_2}\ \mathbf{1}]^\mathbf{T}$, each point in the linear equation in $\mathbf{F}$ is unknown. As mention in the properties, each $(\mathbf{p_2}, \mathbf{p_1})$ correspondences gives an equation $\mathbf{p_2 F p_1}$ to estimate the elements of $\mathbf{F}$. This equation is written as :

$$\begin{bmatrix} \mathbf{x_2}\ \mathbf{y_2}\ \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{f_{11}} & \mathbf{f_{12}} & \mathbf{f_{13}} \\ \mathbf{f_{21}} & \mathbf{f_{22}} & \mathbf{f_{23}} \\ \mathbf{f_{31}} & \mathbf{f_{32}} & \mathbf{F_{33}} \end{bmatrix} \begin{bmatrix} \mathbf{x_1} \\ \mathbf{y_1} \\ \mathbf{1} \end{bmatrix} = \mathbf{0}, \tag{2.49}$$

Equation 2.49, which is called epipolar constraint is expanded and the new equation developed.

$$\begin{bmatrix} \mathbf{x_2}\ \mathbf{y_2}\ \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{f_{11}x_1 + f_{12}y_1 + f_{13}} \\ \mathbf{f_{21}x_1 + f_{22}y_1 + f_{23}} \\ \mathbf{f_{31}x_1 + f_{32}y_1 + F_{33}} \end{bmatrix} = \mathbf{0}, \tag{2.50}$$

Equation 2.50 is solved,

$$\mathbf{x_2(f_{11}x_1 + f_{12}y_1 + f_{13}) + y_2(f_{21}x_1 + f_{22}y_1 + f_{23}) + f_{31}x_1 + f_{32}y_1 + f_{33} = 0}, \tag{2.51}$$

Simplify Equation 2.51

$$\mathbf{x_2 f_{11} x_1 + x_2 f_{12} y_1 + x_2 f_{13} + y_2 f_{21} x_1 + y_2 f_{22} y_1 + f_{23} y_2 + f_{31} x_1 + f_{32} y_1 + f_{33} = 0}, \tag{2.52}$$

Rearrange Equation 2.52

$$\mathbf{x_2x_1f_{11} + x_2y_1f_{12} + x_2f_{13} + y_2x_1f_{21} + y_2y_1f_{22} + y_2f_{23} + x_1f_{31} + y_1f_{32} + f_{33} = 0,} \tag{2.53}$$

The solution of the equations is in the form of a linear equation $\mathbf{Af = 0}$.

Where,

$\mathbf{A}$ is the $\mathbf{1 \times 9}$ matrix given by $\mathbf{A = [x_2X_1\ X_2y_1\ X_2\ y_2X_1\ y_2y_1\ y_2X_1\ y_1\ 1]}$ and $\mathbf{f = [f_{11}\ f_{12}\ f_{13}\ f_{21}\ f_{22}\ f_{23}\ f_{31}\ f_{32}\ f_{33}]^T}$

$$\begin{bmatrix} \mathbf{x_2x_1\ x_2y_1\ x_2\ y_2x_1\ y_2y_1\ y_2x_1\ y_1\ 1} \end{bmatrix} \begin{bmatrix} \mathbf{f_{11}} \\ \mathbf{f_{12}} \\ \mathbf{f_{13}} \\ \mathbf{f_{21}} \\ \mathbf{f_{22}} \\ \mathbf{f_{23}} \\ \mathbf{f_{31}} \\ \mathbf{f_{32}} \\ \mathbf{F_{33}} \end{bmatrix} = \mathbf{0,} \tag{2.54}$$

Given N correspondences $(\mathbf{X_{1i}X_{2i}})$, where $\mathbf{i = 1, 2, 3..., N}$. The equation $\mathbf{Af = 0}$ is developed until the $\mathbf{i^{th}}$ correspondences and stacking together to yield $\mathbf{Af = 0}$ where $\mathbf{A}$ is a $\boldsymbol{N \times 9}$ matrix. Eight correspondence points are needed to estimate $\mathbf{F}$. The condition of the solution by performing the $\mathbf{SVD}$ of $\mathbf{A}$ to obtain $\mathbf{UDV^T}$. Then, zeroing out all the smallest singular value in $\mathbf{D}$, and forming the product $\mathbf{UD'V_T}$ where $\mathbf{D'}$ is the modified form of $\mathbf{D}$. The value of $\mathbf{f}$ taken from the last column of $\mathbf{V}$.

$$\begin{bmatrix} \mathbf{x_2^1x_1^1} & \mathbf{x_2^1y_1^1} & \mathbf{x_2^1} & \mathbf{y_2^1x_1^1} & \mathbf{y_2^1y_1^1} & \mathbf{y_2^1x_1^1} & \mathbf{y_1^1} & \mathbf{1} \\ \mathbf{x_2^2x_1^2} & \mathbf{x_2^2y_1^2} & \mathbf{x_2^2} & \mathbf{y_2^2x_1^2} & \mathbf{y_2^2y_1^2} & \mathbf{y_2^2x_1^2} & \mathbf{y_1^2} & \mathbf{1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{x_2^nx_1^n} & \mathbf{x_2^ny_1^n} & \mathbf{x_2^n} & \mathbf{y_2^nx_1^n} & \mathbf{y_2^ny_1^n} & \mathbf{y_2^nx_1^n} & \mathbf{y_1^n} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{f_{11}} \\ \mathbf{f_{12}} \\ \mathbf{f_{13}} \\ \mathbf{f_{21}} \\ \mathbf{f_{22}} \\ \mathbf{f_{23}} \\ \mathbf{f_{31}} \\ \mathbf{f_{32}} \\ \mathbf{F_{33}} \end{bmatrix} = \mathbf{0,} \tag{2.55}$$

The fundamental matrix's calculation has been widely studied to solve the problem in multi-view geometry. Several papers have probed into this topic and discussed the best solution. New approaches have been introduced by mixing between several algorithms,

for example, the robust estimation, where Zhang et al. used least median square [150] to discard false matching in the set. Random sample consensus (RANSAC) is another related approach that exploited the capability to interpret or smoothing the data containing significant gross error [156]. This point is described by Richard Hartley and Andrew Zisserman in [148].

Olivier Faugeras in [157] provided an exceptional explanation of the technique in computer vision. The report is based on the author and his team's actual work. Mainly, its study the machine visual perception for robotic application. The book is related to the mathematical principle in-line to their work, rather than describing the technique to provide the result. Michealsen et al. [158] use Good Sample Concesus (GOODSAC) to search the best model parameters when the measurement is contaminated with outliers. This strategy applied to the essential matrix with more precision than RANSAC. Another approach was introduced by Jin Yu et al., [159]. The paper explained the optimization process by minimizing the outliers.

The proposed framework developed a new method to remove outlier, which controlled between reliability and speed. One of the properties of the fundamental matrix is the determinant. Using the existing algorithm and forcing the constraint of $\mathbf{det(F) = 0}$. It reduced the number of points to seven. The information of this approach is explained in [160] and [161]. This reviewer approach's asserted that the implementation of this technique is more complicated than that with the eight-point algorithm.

2.9.1.2/ ESSENTIAL MATRIX

The essential matrix stated the relation between two cameras or two positions of cameras looking at the same point in space. Similar to the matrix fundamental, the connection mathematics considered both rotation matrix and translation vector or previously known as rigid transformation, for both calibrated cameras. The difference between the fundamental matrix and the essential matrix is the relation of rotation and translation derived from the camera's retina plan in the essential matrix.

The essential matrix is used for calibrated camera, while the fundamental matrix for the uncalibrated camera. For both cases, the goal is the same - to estimate the rotation matrix and translation vector between the two posses. These aspects are required to figure precisely better yield for the actual application. Suppose the essential matrix is

considered as the relation of mathematics between the two cameras. All the correspondence points obtained are not expressed by the image plan but expressed from the camera plane or the retinal plane. However, suppose the camera matrix is known. In that case, the relation between points at the image plane and the retinal plane is expressed as:

$$\mathbf{p_r} = \mathbf{K^{-1}p_i}, \tag{2.56}$$

where,

$\mathbf{p_i}$ = point on image plane

$\mathbf{p_r}$ = point on retinal plane

$\mathbf{K^{-1}}$= camera matrix

Since there is a similarity in the solution between fundamental and essential matrices, the final solution must also be the same. Taking back Equation 2.48 of the fundamental matrix, the relation between two correspondence points $\mathbf{p_1}$ and $\mathbf{p_2}$ can be established, where $\mathbf{p_1}$ is the point at the first camera and $\mathbf{p_2}$ is the point at the second camera.

The epipolar constraint and the essential matrix writes as,

$$\mathbf{p_2^T E_{12} p_1} = \mathbf{0}, \tag{2.57}$$

where, $\mathbf{E_{12}}$ is essential matrix that established the geometry relationship between two calibrated cameras. When comparing the elements in Equations 2.47 and 2.48, the straightforward relationship between the fundamental matrix and the essential matrix is as follows:

$$\mathbf{F_{12}} = \mathbf{K_2^{-T} E_{12} K_1^{-1}}, \tag{2.58}$$

where,

$\mathbf{E_{12}}$=essential matrix

$\mathbf{K_2}$= instrinsic matrix of camera 2

$\mathbf{K_1}$= intrinsic matrix of camera 1

$\mathbf{F_{12}}$=fundamantal matrix

From the the equation, the essential matrix deduces as

$$\mathbf{E_{12}} = \mathbf{t_{12[\times]}R_{12}}, \tag{2.59}$$

Step to compute the essential matrix

The calculation of essential matrix can be obtained by adhering to the epipolar constraint $\mathbf{P_2^T E_{12} P_1} = \mathbf{0}$. The points $\mathbf{P_2}$ and $\mathbf{P_2}$ are the correspondence points related to the object being looked upon from the two cameras. These points are noted as $\mathbf{P_2(x_2\ y_2\ z_2)}$ and $\mathbf{P_1(x_1\ y_1\ z_1)}$, and consist of six correspondence points to contribute in the calculation. The solution is called a linear solution, and the six points are fed to the essential matrix. Based on the essential matrix calculation, the two cameras' position relationship was obtained with the unknown scaling factor. The linear solution is written as follows:

$$\mathbf{AE} = \mathbf{0}, \tag{2.60}$$

The linear equation is derived from Equation 2.57, where $\mathbf{P_2(x_2\ y_2\ z_2)}$ and $\mathbf{P_1(x_1\ y_1\ z_1)}$ are the correspondence points and,

$$\mathbf{E} = \begin{bmatrix} \mathbf{e_{11}} & \mathbf{e_{12}} & \mathbf{e_{13}} \\ \mathbf{e_{21}} & \mathbf{e_{22}} & \mathbf{e_{23}} \\ \mathbf{e_{31}} & \mathbf{e_{32}} & \mathbf{e_{33}} \end{bmatrix}, \tag{2.61}$$

According to the epipolar constraint in Equation 2.57, the linear solution adheres as following steps.

$$\begin{bmatrix} \mathbf{x_2^n\ y_2^n\ z_2^n} \end{bmatrix} \begin{bmatrix} \mathbf{e_{11}} & \mathbf{e_{12}} & \mathbf{e_{13}} \\ \mathbf{e_{21}} & \mathbf{e_{22}} & \mathbf{e_{23}} \\ \mathbf{e_{31}} & \mathbf{e_{32}} & \mathbf{e_{33}} \end{bmatrix} \begin{bmatrix} \mathbf{x_1^n} \\ \mathbf{y_1^n} \\ \mathbf{z_1^n} \end{bmatrix} = \mathbf{0}, \tag{2.62}$$

where $\mathbf{n}$ is the number of estimation points.

$$\begin{bmatrix} \mathbf{x_2^n\ y_2^n\ z_2^n} \end{bmatrix} \begin{bmatrix} \mathbf{e_{11}x_1^n + e_{12}y_1^n + e_{13}z_1^n} \\ \mathbf{e_{21}x_1^n + e_{22}y_1^n + e_{23}z_1^n} \\ \mathbf{e_{31}x_1^n + e_{32}y_1^n + e_{33}z_1^n} \end{bmatrix} = \mathbf{0}, \tag{2.63}$$

Solve the matrix multiplication

$$\mathbf{x_2^n x_1^n e_{11} + x_2^n y_1^n e_{12} + x_2^n z_1^n e_{13} + y_2^n x_1^n e_{21} + y_2^n y_1^n e_{22} + y_2^n z_1^n e_{23} + z_2^n x_1^n e_{31} + z_2^n y_1^n e_{32} + z_2^n z_1^n e_{33} = 0}, \tag{2.64}$$

By simplifying the Equation 2.64 into the matrix form,

$$\begin{bmatrix} x_2^1x_1^1 & x_2^1y_1^1 & x_2^1z_1^1 & y_2^1x_1^1 & y_2^1y_1^1 & y_2^1z_1^1 & z_2^1x_1^1 & z_2^1y_1^1 & z_2^1z_1^1 \\ x_2^2x_1^2 & x_2^2y_1^2 & x_2^2z_1^2 & y_2^2x_1^2 & y_2^2y_1^2 & y_2^2z_1^2 & z_2^2x_1^2 & z_2^2y_1^2 & z_2^2z_1^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_2^nx_1^n & x_2^ny_1^n & x_2^nz_1^n & y_2^nx_1^n & y_2^ny_1^n & y_2^nz_1^n & z_2^nx_1^n & z_2^ny_1^n & z_2^nz_1^n \end{bmatrix} \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{31} \\ e_{32} \\ e_{33} \end{bmatrix} = \mathbf{AE} = \mathbf{0}, \tag{2.65}$$

The position of cameras were extracted from the extrinsic parameter and the essential matrix. To do this, the method prescribed in [148] has adhered. The orientation of the camera is free to rotate at the axis. Due to that, its also changed the position. However, the direction of the camera need to ensure so that the object is located in front of both cameras.

Accordingly, in geometry, it is also possible for both cameras to see the object on the opposite direction. This finding intended to discover the best possible configuration so that both cameras can reconstruct the precisely points. Using the camera matrices $\mathbf{M_1[I0]}$ and $\mathbf{M_2[Rt]}$, the position relative of both cameras is as specified by rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$. The first step is the extraction. Let's consider the SVD of $\mathbf{E_{12}}$ which is as follows:

$$\mathbf{E_{12}} = \mathbf{U}\ \mathbf{diag(1\ 1\ 0)}\mathbf{V^T}, \tag{2.66}$$

or

$$\mathbf{E_{12}} = \mathbf{U\Sigma V^T}, \tag{2.67}$$

with the diagonal matrix,

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{2.68}$$

with $\mathbf{U}$ and $\mathbf{V}$ choose the $\mathbf{det(U) > 0}$ and $\mathbf{det(U) > 0}$.

In the other hand, $\mathbf{t_{12}}$ extracted, which is correspond to the third column of the orthogonal matrix. So,

$$\mathbf{t_{12}} = \pm\mathbf{U(0\ 0\ 1)^T}, \tag{2.69}$$

or the translation vector is either $+\mathbf{U_3}$ or $-\mathbf{U_3}$. The sign positive or negative shows the orientation rotation of the camera.

Similar to translation that uses the orthonormal matrix, the two possible rotations are

$$\mathbf{R}_{12}^{1} = \mathbf{U}\mathbf{W}\mathbf{V}^{\mathbf{T}}, \tag{2.70}$$

and

$$\mathbf{R}_{12}^{1} = \mathbf{U}\mathbf{W}^{\mathbf{T}}\mathbf{V}^{\mathbf{T}}, \tag{2.71}$$

Where the orthonormal matrix is given by

$$\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{-1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}, \tag{2.72}$$

Equation 2.69, 2.70 and 2.71, are equations of rotation and translation between two cameras. They are four possible orientation between camera 1 and 2.

Let's consider camera one as the reference camera. Among all the possible orientation, only one is the correct solution to use as the best configuration for 3D reconstruction. The four possible arrangements are listed in the following:

$$\mathbf{M}_{2}^{1} = [\mathbf{R}_{12}^{1}| + \mathbf{t}_{12}], \tag{2.73}$$

$$\mathbf{M}_{2}^{2} = [\mathbf{R}_{12}^{2}| + \mathbf{t}_{12}], \tag{2.74}$$

$$\mathbf{M}_{2}^{3} = [\mathbf{R}_{12}^{1}| - \mathbf{t}_{12}], \tag{2.75}$$

$$\mathbf{M}_{2}^{4} = [\mathbf{R}_{12}^{2}| - \mathbf{t}_{12}], \tag{2.76}$$

The four configurations are displayed in Figure 2.42. The best solution is the solution is based on the following:
(a) where the point $\mathbf{X}$ reconstructed in front of both cameras.

Prior studies on camera geometry, it only focused on the perspective camera. As mentioned earlier, the geometry of perspective camera follows the geometry of the pinhole camera model. In the case of a hybrid camera, such as catadioptric and dioptric cameras, it should define the solution to obtain the fundamental and essential matrices, apart from seeking the solution for geometry problem.

Several hybrid cameras are discussed in Chapter 2.6. The discussion included the physical construction of the camera and the perception to see the object. The omnidirectional camera is an NSVP camera. However, the camera system has an SVP camera

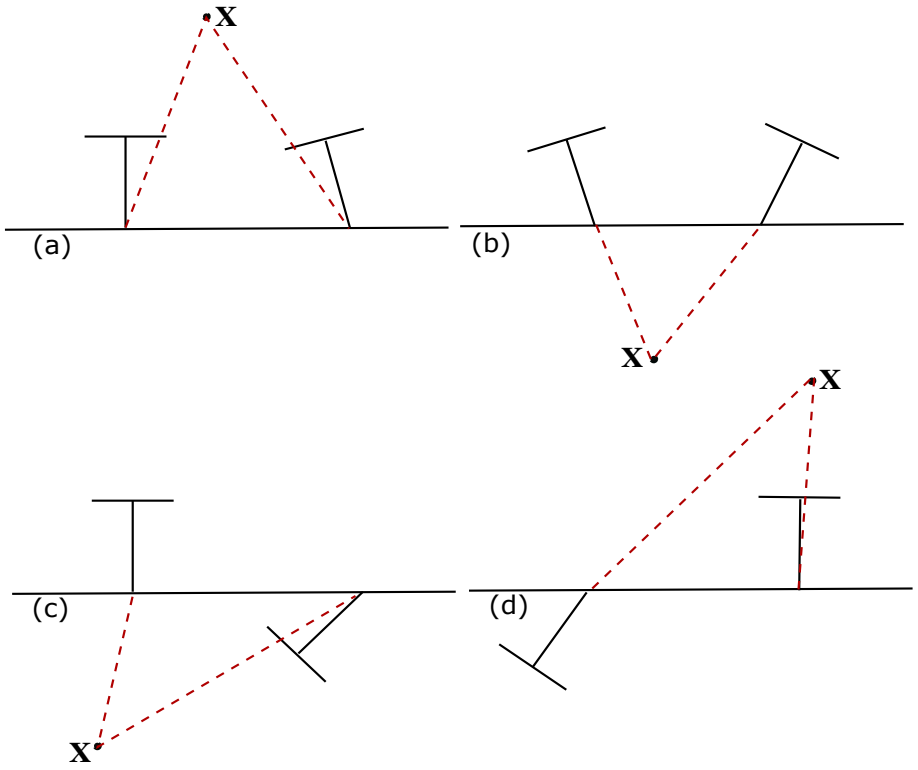


Figure 2.42: The four possible solutions for reconstructing point $\mathbf{X}$. Solution (a) is the best solution for both of the cameras facing the point $\mathbf{X}$. The solution (b) is false because the point $\mathbf{X}$ is reconstructed behind both cameras. Same as the solution (c) and (d), the point $\mathbf{X}$ reconstructed behind one of the cameras.

system if the camera center placed correctly on the mirror's focal point. Otherwise, it does not comply with the SVP constraint and results in undesired distortion and lousy image. For an omnidirectional camera in light of SVP, the spherical model projection may be incorporated. The perspective camera linearly fits the pinhole camera model. The projection model and the epipolar constraint are related to the pinhole model. As for the case of the omnidirectional camera, it is not linear anymore.

The parameters are lifted and adjusted so that they can be used for omnidirectional cameras. Besides, the epipolar line is not as straight as that for the perspective camera. The epipolar line is deformed and follows the form of lenses and mirror of the omnidirectional camera. This point means; the perspective view can adapt to the conic characteristic of the mirror. This idea was highlighted by Geyer et al. [162], which worked on the fundamental matrix of catadioptric by considering the geometry of the mirror.

In 2005, Claus et al. [163] study the ratianal function model for a catadioptric camera. It focussed for radial distortion on the catadioptric mirror. The algorithm was proposed to adapt the camera and lens or mirror to obtain the generalised fundamental

matrix and epipolar constraint for the camera and lens. Micusik et al., [57] discussed the SVP constraint and the type of camera extended to the omnidirectional camera. Following the investigation, the target was to generalise the geometry problem. The results were used for all kinds of cameras concerning the SVP constraint. In 2002, Sturm et al., [164] devised an idea to mix the configuration between omnidirectional and perspective cameras.

The goal of this idea is to obtain the generalised fundamental matrix between both cameras. Later, the idea was continued by Puigi et al. [165] who performed the deep analysis on the two views concerning combining images obtained from omnidirectional and conventional cameras. The research had focused on epipolar geometry and homography. Bastanlar et al. [166] assessed the combination between omnidirectional and perspective cameras and used the findings to perform a structure from motion. The paper [167] by Baretto et al. explained the combination of two omnidirectional cameras, which are catadioptric and dioptric (fisheye) cameras. The mixture is represented a new representation of the geometry system by lifting the perspective plane to 5D projective planes (Veronese map).

2.10/ CONCLUSION

The literature acknowledges the ability to increase the field of view of a camera. Se'veral wide field of view cameras system from state-of-the-art has been discussed. This chapter investigated the detail of the wide field of view component. This investigation includes the technique to optimize the field of view and becomes an important point depicted for the next chapter. The optical distortion is a critical issue for the mirror-based and lens of a camera system. This issue distortion exists at all wide field of view camera system.

Moreover, the single viewpoint constraint is essential in the development of an omnidirectional camera. This constraint is essential in developing a mirror-based and lens-based camera system. The literature highlight discussed the importance of SVP in the development of an image. This chapter also discussed the mirror from the family of a conic section that confirmed an SVP constraint.

The fundamental of geometry is presented in the mathematical approach. The pin-

hole camera is studied and become a fundamental model in the image's projection. The notion of intrinsic and extrinsic parameters are verified from the pinhole camera. It extended for the omnidirectional camera for more complex projection geometry. The representation of both parameters embeds together in a matrix. Then the multiview geometry is discussed in emphasizing the estimation of the fundamental and essential matrix.

This chapter generates the basic idea in developing the outcomes of the research. The research finding is guided to achieve the goal to develop the desire vision system. The research from the state-of-the-art highlighted the current issue. This research gap is vital to solve the issue and develop the novelty in this research.

3

SENSOR DESIGN

3.1/ INTRODUCTION

The wide FoV vision system has become popular in robotic applications and as a human-aided apparatus. The typical application of this vision system is for video surveillance and for further in robotic application. The vision system is connected to the computer network to monitor several places simultaneously. For the past decade, the camera is used as part of the security system parallel to the advanced technology in computer vision. The camera system has been integrated with other algorithms, such as face recognition (a biometric-based application that uniquely identifies a person based on his facial pattern). Earlier in the robotic application, the camera is used to detect paths and objects. It also used for deep application such as the 3D reconstruction and the depth estimation.

As time passes, the robotics industry has become more intelligent by combining machine vision and computer vision. The early application, the vision technique reflected the use of a laser scanner with a bar code. Then, the incorporation of QR code a few years later, along with the smartphone's massive evolution. The next technological expansion initiated autonomous robots capable of performing tasks individually, such as independent manoeuvring and monitoring. In doing so, both the connection and integration between the robot and the vision system are essential. Nonetheless, the robot can perform the assigned task more efficiently if it has all information regarding the surrounding area [168].

The conventional camera system has limited FoV. The FoV referred to an observable open area that the robot can see through the camera (optical sensor). The standard

conventional camera used a standard lens with a focal length ranging from 30 to 40 mm. Let's say $\mathbf{f}$ is the focal length, $\mathbf{l}$ is the long side of the camera or sensor frame, and $\boldsymbol{\theta}$ is the angle of FoV. Figure 3.1 shows the FoV of a conventional camera. The correlations among angle $\boldsymbol{\theta}$, focal length and the camera frame are expressed in Equation 3.1.

$$\boldsymbol{\theta} = \mathbf{2}\tan^{-1}\left(\frac{\mathbf{l}}{\mathbf{2f}}\right) \tag{3.1}$$

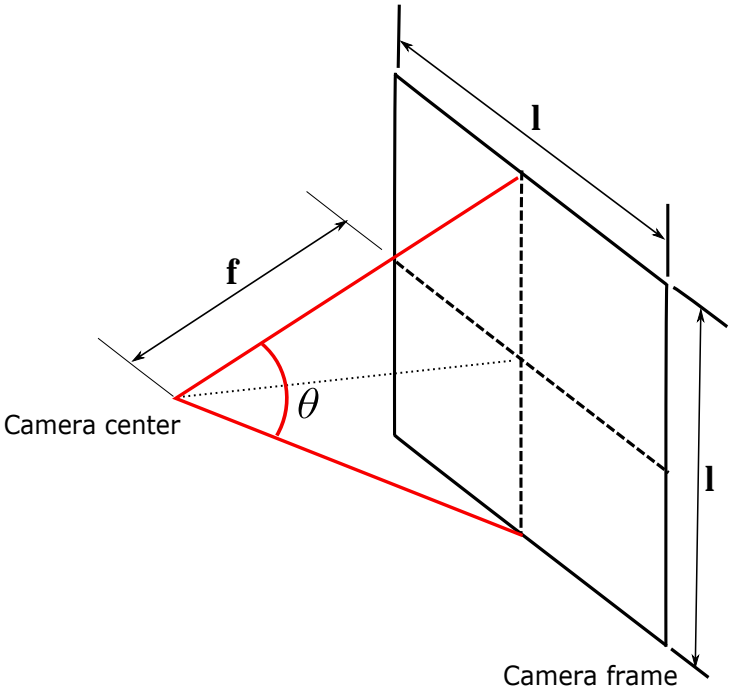


Figure 3.1: The FoV of conventional camera.

Referring to Equation 3.1, the focal length,$\mathbf{f}$ is the only variable that controls the FoV of the camera without changing the physical dimension of the camera. The FoV can be adjusted by increasing or decreasing the focal length. Decreasing $\mathbf{f}$ offered a wider FoV, but there is a limit in decreasing $\mathbf{f}$. The further decrease of focal length and the camera frame can cause distortion issue.

A full FoV offered plenty of advantages, particularly in terms of time response for robotic application. For example, in a mission to hunt a specific target, the 360° FoV vision system reduces response time and saves power consumption. This chapter described the development, design, and calibration of 360° FoV vertical and horizontal vision systems together with a high-resolution mirror to grab more detail information from the scene.

3.2/ PROPOSED SYSTEM : OMNI-VISION

The research and development of camera system as an important feature in robotics application have been carried out extensively. As such, this study took an attempt to increase the FoV. The goal is to increase the efficiency of application in various utilisation such as for monitoring, manoeuvring, and the development of point 3D in space. Chapter 2 listed several techniques used to increase FoV. The methods revealed promising outcomes in FoV expansion. Apart from the advantages, several drawbacks were noted, such as object deformation, low resolution, and rigidity issues.

A complete field of view vision system is always an exciting topic in the robotic application. Cross-knowledge between a biologist and an engineer cum researcher offered assuring results, as mentioned in the previous paragraph. The amount of research work aimed to enhance the efficacy of communication between human and robot or between robots and the surrounding area. The connection should be continuous, tangible, and relevant. In this frame, human perception is always the best and should serve as a reference. Chapter 2 presents several methods to improve the FoV. A proper vision system can operate in real-time within dynamic and cluttered scenes, as well as varied lighting to capture information from the robotics surrounding.

As such, the vision system proposed in this study is composed of several cameras with varied vision capabilities. However, the solution should suit to the robotic application. Hence, the proposed vision sensor is designed to achieve 360° FoV with minimum sensors, apart from considering the rigidity of the system. The proposed vision system is vital for robotic application due to its limited mounting space. Moreover, the designed of the vision sensor was ensured to incur a minimum cost. The design on Omni-Vision aimed to observe 360° (prey-like vision) scene. Simultaneously, observing the detail of the scene in front of the camera system (predator-like vision) using a high-resolution perspective camera. The 360° scene captured by two omnidirectional fisheye cameras placed back to back, while the high-resolution scene captured by the ZED perspective camera. The detail of the cameras used in this camera system discussed and explained in the next sub-chapter.

3.3/ HYBRID CAMERA DESIGN

Several methods are discussed in Chapter 2 to acquire the panoramic image. Chapter 2, presents all omnidirectional camera models, along with their advantages and disadvantages. It included the distortion which appeared on the image.

The omnidirectional camera offered a wide FoV that suited for the new robotic applications. This camera addressed some issues in computer vision application, such as object detection and pattern recognition. The proposed Omni-Vision system aimed to retrieve 360° FoV and captured the information for object detection and 3D reconstructions. Figure 3.2 shows the proposed set-up of Omni-Vision. In contrast, Figure 3.4 portrays the view from the Omni-Vision.

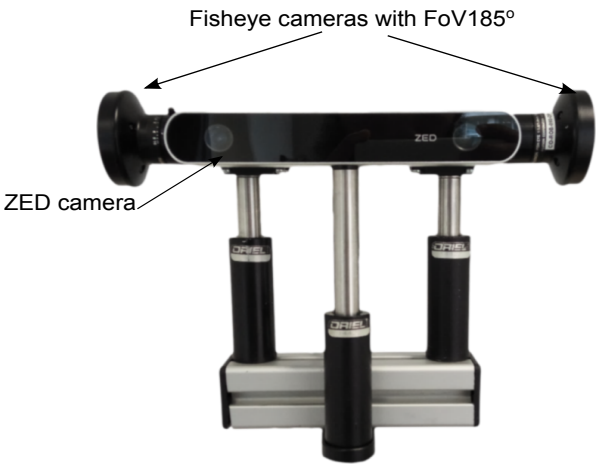


Figure 3.2: The Omni-Vision camera rig. The fisheye cameras are placed back-to-back to acquire 360° FoV. The suggested position takes advantage of the fisheye camera specification that offers more than 180° FoV. A high-resolution ZED stereo vision camera is placed in front of the camera rig. The baseline is parallel with the baseline of fisheye cameras.

3.3.1/ FISHEYE CAMERA

The proposed Omni-Vision sensor used two IDS CCD cameras. It mounted with a Fujinon fisheye lens. The fisheye cameras are placed back-to-back to view a full scene on the left and right of the body. In other words, it covered the whole scene at once. Referring to figure 2.15, the FoV of the fisheye lens is 185°. Essentially, the combination of two fisheye lens placed back-to-back offered 360° FoV both vertically and horizontally.

Why fisheye camera and not catadioptric camera? Fisheye cameras offered FoV as extensive as 185° despite radial distortion, especially at the edge of the image. However, if compared with catadioptric cameras, fisheye cameras do not have a fix round black blind-spot at the center of the image, which is the image of the camera.

The fisheye cameras' designed to minimize the size, rigidity, and robustness. It made these cameras superior to catadioptric cameras for a full field of view. The suitability of fisheye camera adapted to unified spherical model which is aligned with the catadioptric cameras, made it conceivable and exceptionally profitable to exploit the larger FoV in robotic applications.

iDS Camera: The iDS camera is a digital camera with a CCD (Couple Charge Device). The CCD camera is functioned to convert falling photons to the sensor surface into electrons, then accumulated into the capacitor and transformed into electronic signal forms. Compared to another type of digital camera, CMOS (Complementary Metal Oxide Semiconductor), the CCD camera offered more benefits, such as images with low noise, more pixels, and better quality. The CCD cameras produced a high-quality images due to a high number of pixels and excellent light sensitivity. The camera size is compact. It minimized the baseline distance between the cameras in a back-to-back position. Figure 3.3 exhibits the iDS camera.

Figure 3.3: TThe CCD digital camera from Imaging Development System (iDS)

Fujinon Fisheye Lens: The perspective camera mentioned above has limited FoV. Referring to Chapter 2, the Fujinon fisheye camera is mounted on the iDS camera to increase the field of view. The Fujinon fisheye lens, with FoV exceeding 180°, has been widely used for video surveillance. The wide field of view is exploted, hence, an algorithm is vital in order to fuse the images from two fisheye cameras to obtain a panoramic image

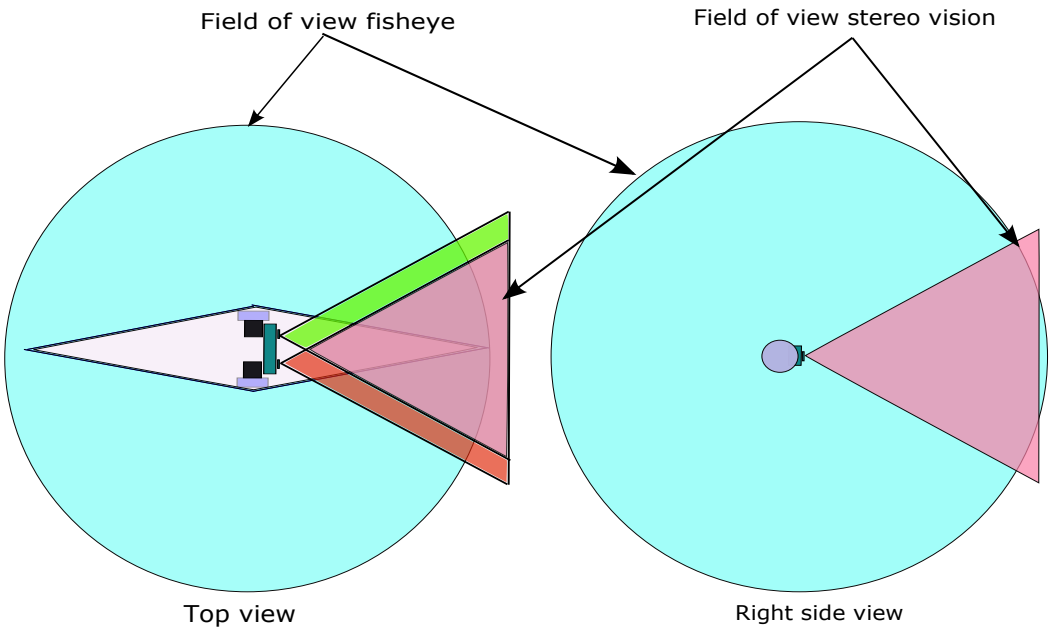


Figure 3.4: The Omni-Vision FoV. Blue represents the coverage area of fisheye cameras. The top and right-side views show that the fisheye cameras can cover full scene. Green and red represent views from stereo vision camera. The violet depicts the area of binocular view of stereo vision camera.

of 360°. Figure 3.5 and Table 3.1 below show the fisheye lens from Fujinon Corporation.

Figure 3.5: Fisheye lens from Fujinon Corporation.

Despite the strong distortion, fisheye camera dont have the circular blind-spot in the middle of the image, as the catadioptric camera. The fisheye camera designed more compact and structurally durable. Thus, the camera preferred over bulky catadioptric cameras for all robotic applications, especially in underwater systems. The adaptation of a standard unified spherical model to calibrate central catadioptric cameras with fisheye cameras is enabled the exploitation of a wider FoV for robotic applications. The omnidirectional camera is more affordable and is a popular photographic tool to capture photos with a wide FoV [176].

Fujinon Fisheye: FE185C046H-1:1.4/1,4mm		
Focal Length (mm)		1.4
Iris Range		F1.4 - F16
Operation	Focus Iris	Fixed Manual
Angle of View (HxV)	1/2” 1/3” 1/4”	185°x185° θ(4.6 185°x144°47’ 144°47′x108°35’
Focusing Range (From Front Of The Lens) (m)		∞ - 0.1
Back Focal Distance (in air) (mm)		9.7
Exit Pupil Position (From Image Plane) (mm)		-61
Filter Thread (mm)		-
Mount		C
Mass (g)		140

Table 3.1: Fujinon fisheye specification.

3.3.2/ ZED CAMERA

The Omni-Vision is a vision system capable of observing the surrounding area up to 360° vertically and horizontally. It has high-resolution windows to capture more information from the front view. The process to determine the object distance from the camera is a critical feature in the robotic application. Stereo vision is a method that resembles the function of human perception, which can be implemented in robots using two cameras mounted on the robot. The concept of stereo vision is briefly explained in Chapter 2.

The classic stereo vision uses two different cameras, pointing to the same object. The new generation of stereo vision camera is composed of two cameras synchronised and built-in within the same casing to achieve a 3D vision. In this present study, the ZED camera used to capture the scene’s details in front of the Omni-Vision camera system. It displayed the ability to perceive long-range depth and 3D tracking capability. The ZED camera is a passive camera that differs from active cameras, such as Kinect. It is built with a high-resolution cameras, the ZED camera utilized to perform triangulation to estimate 3D reconstruction.

Among the best features offered by this camera is its ability in depth perception. The ZED camera can measure the depth over 20 meters at a frame rate of 100 fps. This camera achieved the wide field of view, which is 110° and 70° for horizontal and vertical views, respectively. These features are sufficient to be adapted in this research work.

The ZED camera is equipped with a Software Development Kit (SDK) that adds

features to the application. The developer provides the SDK, including libraries, documentation, a sample of code, and guides. The SDK is offered the solution to integrate the camera with the application. Its used in designing for specific programming language. Figure 2.23 in Chapter 2, portrays the ZED camera with output images (images from left and right cameras, depth, and 3D reconstruction).

3.3.3/ Baseline and Blind-spot

The baseline in the context of this study's is referred to the distance between two fisheye cameras placed back-to-back of each other. This section highlighted the presence of blind-spots in the Omni-Vision cameras due to the baseline. The role of fisheye camera used to observe the surrounding area with 360° FoV. The scope and limitations are explained to address any uprising question about Omni-Vision camera system development.

A blind-spot was predicted in the Omni-Vision system, while developing that vision system using several cameras. However, this study identified the cause of this particular blind spot, and its mitigation measures are prescribed.

A blind-spot area refers to an area that the camera cannot see. In the proposed hybrid camera, the area of the blind-spot was determined by the distance between the opposing cameras. Figure 3.6 illustrates that the area of the blind-pot. It occurred from baseline to the intersection FoV between left and right fisheye cameras and the intersection's periphery. A disc illustrated the blind-spot (refer to Figure 3.6(b)) in the form of a circle or a disc, where the diameter of the disc is equal to the diameter of the unit sphere. Figure 3.7 portrays the position of fisheye camera. The area of blind-spot varied proportional to the baseline distance. This point signifies a correlation between the baseline and the area of the blind spot.

The best position of the cameras without a blind spot is when the baseline distance is zero (see 3.7(a)). However, it is impossible to achieve this due to the connection cables at the back of both cameras. Therefore, on the Omni-Vision camera system, the minimum attainable baseline distance measured 12.4 cm.

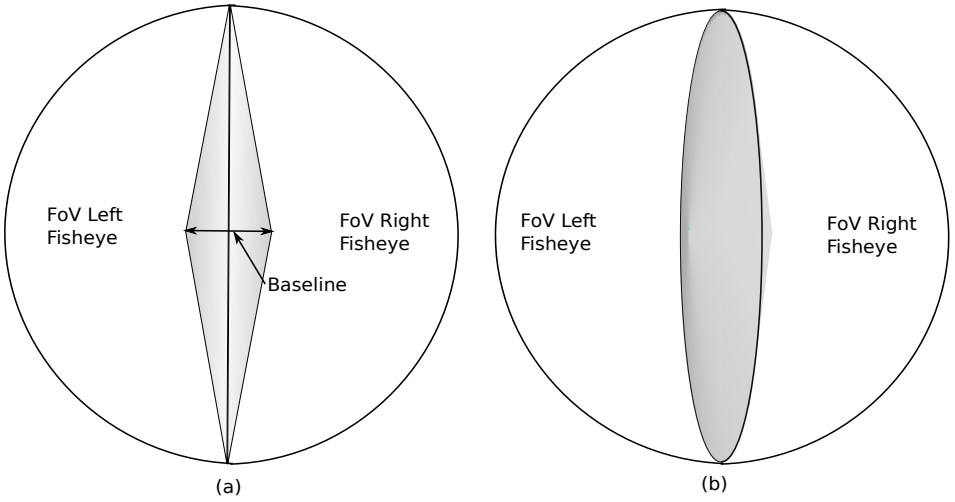


Figure 3.6: The area of blind spot between two fisheye cameras. (a) The blind spot area from in front of the camera rig (see Figure 3.6). (b) The perspective view of the blind spot area.

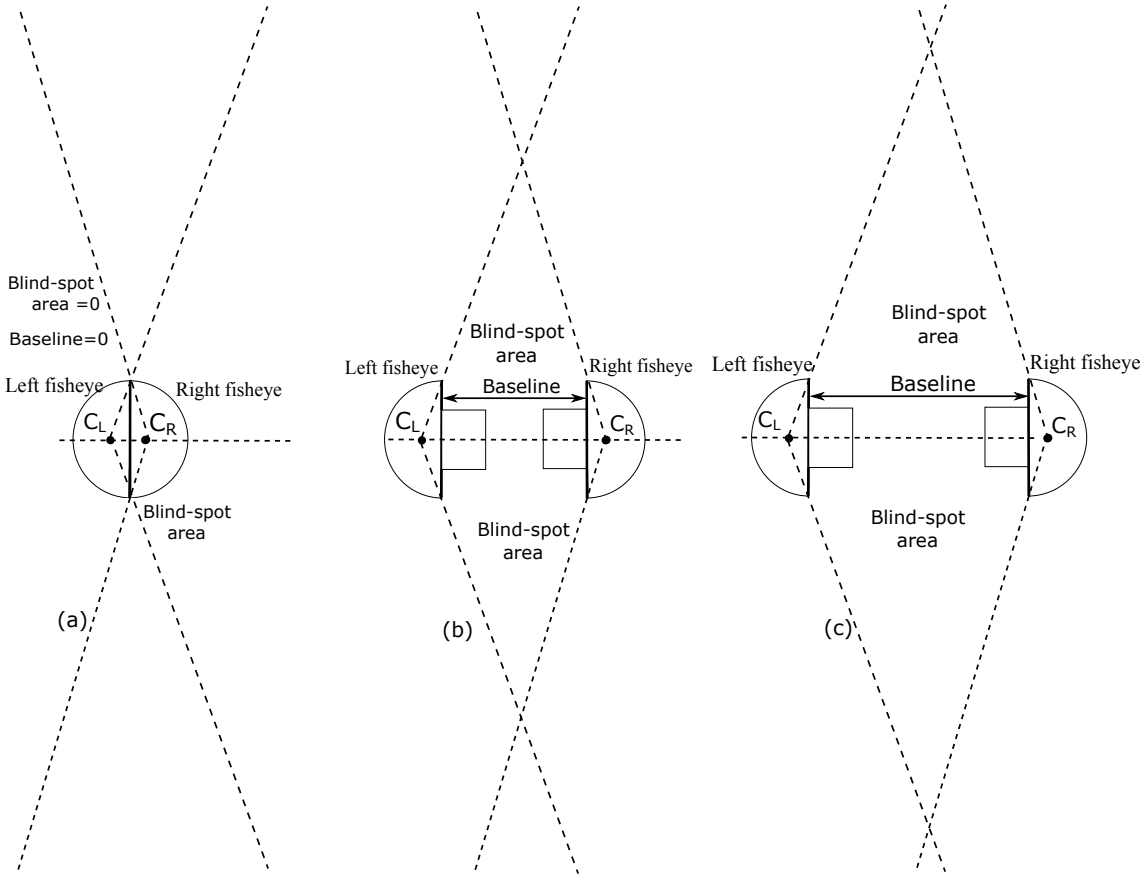


Figure 3.7: The increasing of blind-spot area according to the change of baseline distance.

The goal of calculating disc diameter is to support the existence of blind-spot on the Omni-Vision camera system, illustrated in the form of a disc. Figure 3.8 portrays the cross-section area of the disc. The volume of the disc can be calculated by:

$$V = 2\pi(OM)(ON),$$

where $ON = OM \tan 87.5°$= radius of the disc.

The baseline is not defined by the calculation, but was determined by the minimum distance fisheye camera placed back-to-back. The baseline should be minimized in order to minimize the size of blind-spot.

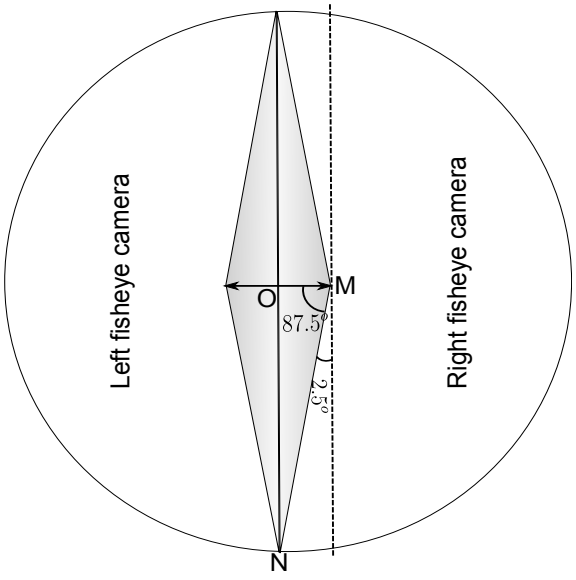


Figure 3.8: The illustration shows the calculation of blind-spot disc diameter.

3.4/ CONCLUSION

Formulation of cameras with 360° view, both vertically and horizontally, is described in this chapter. The system was coupled with a high-resolution stereo view window for detailed viewing. The Omni-Vision camera is a hybrid camera with two view structures; surround and binocular views.

The hybrid camera was adapted from a sensory study involving two groups of animals; prey and predator. Two types of cameras were selected, namely omnidirectional camera and stereo vision camera. The fisheye camera selected due to its size strength, compactness and suitability to mount on the robot. The fisheye camera has a FoV that exceeds 180° in all directions. More importantly, it is generated the blind-spots in com-

parison to catadioptric camera systems.

The ZED camera refers to a high-resolution camera with a binocular view. This camera is used to capture detailed information on the scene. The baseline identified the presence of a blind spot area in the proposed camera design. Nonetheless, after analysing the results, a technique is prescribed to reduce the blind spot area. The proposed camera is called Omni-Vision, and it is equipped with the mentioned camera types. The detail development and implementation of the proposed camera is described in the following chapter.

4

Multi-camera Calibration and Fusion

4.1/ Introduction

An autonomous robot can perform specific tasks in an uncertain environment without any intervention from human to control it. In other words, autonomous robots possessed the ability to recognise and adapt to their surroundings. Implementing a real-time image is a tricky issue to solve. The issue included the camera calibration. An abnormal or a distorted image is produced from the incorrectly or uncalibrated camera. Figure 4.1 displays an example of an abnormal or a distorted image.

(a) Distorted image (b) Real image

Figure 4.1: An example of image captured using uncalibrated camera.

The goal of calibration process is to ensure that the captured image reflects the real situation. In the 3D reconstruction, the calibrated camera gave the exact coordinate of its position. In general, the calibrated camera gave the correct information to the processor

and result in reducing error. The camera's calibration process, is a common problem in computer vision [177]. It involved several factors, including the technique and the environment. A slight change in the physical parameters of a camera can produce an output image that differs from the real situation.

The main problem in the camera calibration process is the process to estimate the internal parameters of the camera. To acquire and obtain the right image, several factors should be considered, including light intensity, the colour of objects around the target object, and the quality of the camera used. Hence, the calibration process needs the optimum condition to estimate the related parameters, either internal or external in producing the optimum image and locating the position of the camera simultaneously. The result with this optimum set-up will optimize the quality of 360° image for surveillance and robotics application.

4.2/ CAMERA CALIBRATION

This section presents the calibration techniques used in this study. The calibration technique differed to the type of camera. In this study, the Omni-Vision camera system used two different camera, the perspective and the omnidirectional camera. Both camera have different characteristic and the geometry. So that they need the different strategy and technique of calibration.

The result of this study facilitated to other researchers to implement this calibration technique in robotic applications and further develop it to the higher level of application. The vision of the robot is highly dependent on the quality of the calibration process. Failure to perform a good calibration process will paralyse the robot from seeing clearly. The most serious is a calculation error in estimating the position of the image or the position of the camera. This low performance will cause the camera to fail to transform the real world into the image. Therefore, the robot is not able to perform its task exactly as expected.

The technique of calibration differed from one type of camera to another. For example, the omnidirectional camera has calibration techniques and methods that differ from other camera types. In fact, numerous methods have been proposed for camera calibration, auto-calibration, and implementation in structure from motion [56] [57] [67] [178] [68] [179] [180].

4.2.1/ INTRINSIC CALIBRATION

3D reconstruction is an exciting science domain in geometry computing. Before performing 3D reconstruction, camera calibration is essential to determine the parameters from the camera, which are intrinsic and extrinsic parameters.

Chapter 2 has generally described the functions of intrinsic and extrinsic parameters in creating an image. Nevertheless, image calibration is a significant problem upon dealing with computer vision [181]. Its primary purpose is to capture an image using the camera and estimate the image's transformation from the real 3D object to the image plane. The essence of the perspective error included in one of the vision error. The perspective errors occurred when an image captured by the camera is not perpendicular to the real object.

Figure 4.2(a) exhibits the object taken by a camera in a perpendicular position that produces a correct perspective. On the other hand, Figure 4.2(b) displays the object taken by a camera with a specific angle that produces a vast perspective error. This point shows that despite using the same object, the resulting perspective may differ.

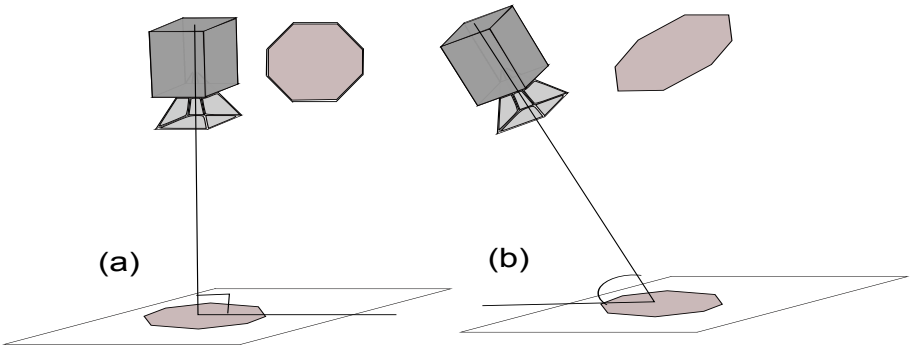


Figure 4.2: Views from different camera angles.

One reason for this to happen is the existence of various camera models, as discussed in Chapter 2. This section presents camera calibration for several models, such as pinhole and spherical cameras.

Generally, the calibration process is required when the camera captures a previously known object with some variances in the produced image. The image is corrected via the comparison of images, and the calibration is essential if there is geometry different between the images. In order to simplify the calibration process, a calibration tool kit called calibration grid (chessboard) is required, as illustrated in Figure 4.3. The num-

ber of images taken during calibration ascertains the accuracy and the precision of the calibration decision.

4.2.1.1/ PERSPECTIVE

Numerous studies have tapped into several calibration techniques. For instance, Xu et al. [182] examined several calibration methods and showed the results obtained after using the latest calibration method. The differences contained therein were classified into several calibration methods, which could be used to determine the parameters of some camera models.

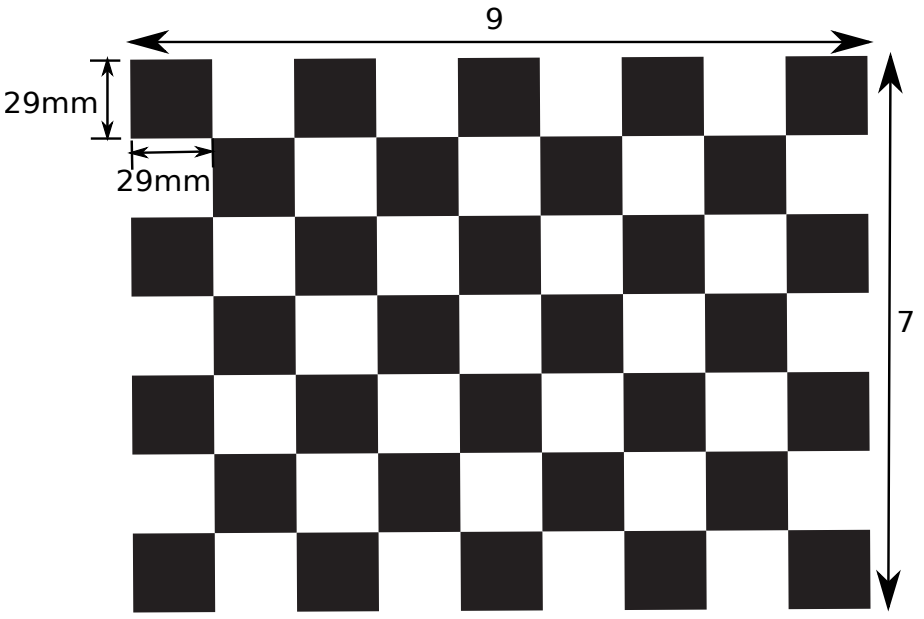


Figure 4.3: The pattern (chessboard) used for camera calibration. The size of the chessboard is $\mathbf{7 \times 9}$ with 63 squares. The size of each individual square is $\mathbf{29mm \times 29mm}$.

On the other hand, Salvi et al. [183] employed linear techniques using computational matrix transformation for camera calibration. The technique used specific matrix sizes to transform 3D objects into 2D objects.

Another technique used to calibrate the camera is the non-linear optimisation technique. A calibration technique becomes non-linear when the lens in the camera model is imperfect. In this case, the camera parameters are obtained from the trial repeatedly to achieve optimal outcomes. The functions performed can identify the best

distance and image projections by using the same approaches multiple times. The advantage of this technique is that when an object is calibrated repeatedly, sufficient data are gathered to increase other objects' accuracy.

Xu et al. [182] proposed a calibration method called the two-step technique. This technique uses linear parameters to compute several parameters. In the second step, the rest of the parameters are repeated iteratively. This technique is enabled to obtain a fast calibration configuration by breaking down the number of iterations performed. This two-step technique uses the advantages of the methods mentioned above.

Earlier researchers have widely carried out studies on calibration techniques. According to Lenz and Tsai (1988), [184], the two calibration techniques is referred to use a scale factor and image centre. The perpendicular scale factor is used to capture images from the camera, then transformed into a matrix. A calibration using an imaging centre is another technique. It referred to take a straight line against the fall of light to identify the centre of the object taken.

Meanwhile, Zhang (1999) [185] depicted that the two calibration techniques are photogrammetric calibration and self-calibration. The photogrammetric calibration technique is applied to test the object to be calibrated. On the other hand, self-calibration is performed by processing the camera's data to get the optimal object without executing any calibration test.

A the moment, the perspective camera referred to the pinhole camera. But, in a modern camera, some perspective cameras are also a wide-angle camera even though the level is not as large as omnidirectional. Therefore, the intrinsic matrix and the distortion coefficient of are critical parameters that must be determined. Both of these parameters were discussed in Chapter 2. Calibration geometric was initiated 15 years ago when several new calibration approaches were proposed. Tsai et al., [186] proposed a calibration method that reduces the parameters that need to be calculated or estimated. This method takes into account the internal and external parameters of the camera and lens. Neglecting the parameters provided by the manufacturer will definitely affect the camera calibration process badly. The proposed method produces accurate and complete results.

Heikkila et al., [187] proposed a four-step camera calibration. This procedure is extended from the two-step method developed by Xu et al. The additional steps to this method are to compensate for distortion caused by circular features. The fourth step is

to correct the distorted image coordinates. This method may be performed in both linear and non-linear settings.

An easy-to-understand calibration method should always be an option. The method prescribed by Zhang [185] is the most popular method to calibrate a perspective camera. This method can be performed using the MATLAB toolbox developed by J. Y. Bouguet [188]. Zhang's method is translated step-by-step in the MATLAB toolbox to extract the intrinsic parameters of the camera.

The point 2D is denoted as $p = (u, v)^T$, while point 3D reflects $P = (X, Y, Z)^T$ with Z being a vector null. Let's say H is the matrix homography from at least three images.

$$\mathbf{H} = \begin{bmatrix} \mathbf{h_{11}} & \mathbf{h_{12}} & \mathbf{h_{13}} \\ \mathbf{h_{21}} & \mathbf{h_{22}} & \mathbf{h_{23}} \\ \mathbf{h_{31}} & \mathbf{h_{32}} & \mathbf{h_{33}} \end{bmatrix}, \tag{4.1}$$

At each image, a homography between the checkerboard (point $\mathbf{P}$ or object) and the image point $\mathbf{p}$ is calculated. The output of the calculation is used to construct matrix $\mathbf{V}$.

$$\mathbf{V} = \left[\mathbf{h_{i1}h_{j1}}, \mathbf{h_{i1}h_{j2}} + \mathbf{h_{i2}h_{j1}}, \mathbf{h_{i2}h_{j2}}, \mathbf{h_{i3}h_{j1}} + \mathbf{h_{i1}h_{j3}}, \mathbf{h_{i3}h_{j2}} + \mathbf{h_{i2}h_{j3}} + \mathbf{h_{i3}h_{j3}}\right], \tag{4.2}$$

The singular value decomposition (SVD) produced the following matrix $\mathbf{D}$:

$$\mathbf{D} = \begin{bmatrix} \mathbf{D_{11}} & \mathbf{D_{12}} & \mathbf{D_{13}}.... & \mathbf{D_{1k}} \\ \mathbf{D_{21}} & \mathbf{D_{22}} & \mathbf{D_{23}}.... & \mathbf{D_{2k}} \\ \mathbf{D_{31}} & \mathbf{D_{32}} & \mathbf{D_{33}}.... & \mathbf{D_{3k}} \\ \mathbf{D_{41}} & \mathbf{D_{42}} & \mathbf{D_{43}}.... & \mathbf{D_{4k}} \\ \mathbf{D_{51}} & \mathbf{D_{52}} & \mathbf{D_{53}}.... & \mathbf{D_{5k}} \\ \mathbf{D_{61}} & \mathbf{D_{62}} & \mathbf{D_{63}}.... & \mathbf{D_{6k}} \end{bmatrix}, \tag{4.3}$$

The coefficient $\mathbf{k}$ denotes the length of the column of a matrix $\mathbf{D}$. Upon taking the last column of a matrix $\mathbf{D}$, all elements are arranged to produce matrix $\mathbf{B}$. The matrix $\mathbf{B}$ (6D vector) is a symmetry matrix used to estimate the intrinsic parameters of a camera.

The elements in matrix $\mathbf{B} = [\mathbf{b_{11}}\ \mathbf{b_{12}}\ \mathbf{b_{13}}\ \mathbf{b_{21}}\ \mathbf{b_{22}}\ \mathbf{b_{23}}\ \mathbf{b_{31}}\ \mathbf{b_{32}}\ \mathbf{b_{33}}]$ are the input of $\mathbf{b}$, which are necessary to extract intrinsic parameters.

$$\mathbf{B} = \begin{bmatrix} \mathbf{D_{1k}} & \mathbf{D_{2k}} & \mathbf{D_{4k}} \\ \mathbf{D_{2k}} & \mathbf{D_{3k}} & \mathbf{D_{5k}} \\ \mathbf{D_{4k}} & \mathbf{D_{5k}} & \mathbf{D_{6k}} \end{bmatrix} = \begin{bmatrix} \mathbf{B_{11}} & \mathbf{B_{12}} & \mathbf{B_{13}} \\ \mathbf{B_{21}} & \mathbf{B_{22}} & \mathbf{B_{23}} \\ \mathbf{B_{31}} & \mathbf{B_{32}} & \mathbf{B_{33}} \end{bmatrix} = \mathbf{K^{-T}K^{-1}}, \tag{4.4}$$

Finally, the intrinsic parameters are calculated individually from the following equa-

tions.

$$\upsilon_0 = \frac{\mathbf{B_{12}B_{13} - B_{11}B_{23}}}{\mathbf{B_{11}B_{22} - B_{12}^2}}, \tag{4.5}$$

$$\lambda = \frac{\mathbf{B_{33} - [B_{13}^2 + \upsilon(B_{12}B13 - B11B_{23})]}}{\mathbf{B_{11}}}, \tag{4.6}$$

$$\alpha = \sqrt{\frac{\lambda}{\mathbf{B_{11}}}}, \tag{4.7}$$

$$\beta = \sqrt{\frac{\lambda \mathbf{B_{11}}}{\mathbf{B_{11}B_{22} - B_{12}^2}}}, \tag{4.8}$$

$$\gamma = \frac{-\mathbf{B_{12}}\alpha^2\beta}{\lambda}, \tag{4.9}$$

$$\mu_0 = \frac{\lambda\upsilon_0}{\beta} - \frac{\mathbf{B_{13}}\alpha^2}{\lambda}, \tag{4.10}$$

Next, the intrinsic values are arranged in matrix $\mathbf{K}$, as follows:

$$\mathbf{K} = \begin{bmatrix} \alpha & \gamma & \mu_0 \\ \mathbf{0} & \beta & \upsilon_0 \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}, \tag{4.11}$$

4.2.1.2/ OMNIDIRECTIONAL-FISHEYE

As mentioned in the previous chapter, the omnidirectional camera is the combination of a perspective camera and a lens or a reflected mirror. The system is called dioptric or fisheye. Meanwhile, the system is called catadioptric if the perspective camera is combined with the reflected mirror. Both systems share some similarities in the camera calibration process. The combination of a perspective camera with a lens or a reflected mirror changes the camera's geometry.

The calibration process needs to estimate the new internal parameter of a global system. Such as the coordinate of principle point in format pixel dimension, the focal length in terms of pixel dimension $(\mathbf{fx, fy})$, and the skew parameter $(\mathbf{s})$. The other relevant parameter is the eccentricity (ξ).

The eccentricity is a coefficient that defined the value of the radial distortion of the camera lens. Suppose the FoV of a camera exceeds 180°. In such cases, the correct value of eccentricity can be identified based on the exact location of such objects in space. This point means; if an object is located at 180° from the camera's FoV, the object should lie on the 180° line of the camera in reality. Otherwise, the eccentricity should be

refined to obtain a more accurate result. The value of eccentricity determines the level of distortion. This value reflected the envelopment of the 180° hemisphere image. The positive side of the lens and camera combination is the increase of FoV from 70° to more than 180° FoV vertically and horizontally.

The optical lens may produce a systematic error at the curve. To minimise the systematic effect or recover new intrinsic parameters, the camera with lens must be re-calibrated, and the output observation or measurement should be in-line with the results obtained [189]. A unified spherical model representation is commonly used for the omnidirectional camera. The eccentricity coefficient obtained during the calibration process is referred to the value of the radial distortion of the lens. The calibration of the omnidirectional camera uses the same procedure as the perspective camera. The chessboard displayed in Figure 4.3 can serve as a reference object during the calibration process. The image of the chessboard should be taken from several different angles. It took from the different positions and angles to minimize the projection error. Besides, the chessboard should be adequately placed on the planar. The camera should neither be too far nor too close to the pattern. This procedure is another solution to reduce errors.

The two methods for calibration of the omnidirectional camera were developed by Scaramuzza [190], and Mei [191]. The two approaches are popular in the robotic community due to their easy-to-understand and straightforward calibration procedures. Besides, the toolbox provided eases the use of this method.

Mei [191] asserted that the calibration process is simplified by making some assumptions based on the theory of Barreto [192]. Among the assumptions is that the value of eccentricity should start with 1 (one). Hence, it is assumed that the principal point must be in the middle of the camera lens, primarily to obtain a balanced image. Besides, the distortion coefficient should be small and close to zero. Applying these assumptions eases the calculation and estimation of the focal length. If the value of the eccentricity coefficient is 1 (one), then the focal length value is assumed to be linear. This assumption should be based on at least three image points owned by non-radial lines. These steps must be followed to estimate the intrinsic parameters using a known size grid.

Earlier, Scarmuzza [190] had already carried out a series of experiments to calibrate an omnidirectional camera. The method proposed by Scaramuzza differed from that proposed by Mei. Scaramuzza omitted the projection model developed by Baretto.

He modelled the 3D point projection of the image based on Taylor's series of compositions. As opposed to the method used previously, the points were not projected on the unified spherical model but projected directly on the reflection mirror of a catadioptric camera or a fisheye camera lens.

Puigi et al. l [193] reviewed the existing omnidirectional camera calibration methods within the computer vision community. Despite the variances noted in the calibration methods, each calibration method yielded accurate reconstruction results.

The generic camera model from the omnidirectional camera was assessed for better comprehension [25]. The auto-calibration of the fisheye camera is described in [194], where Micusik et al. generalised the method of simultaneous linear estimation and omnidirectional camera model introduced by Fitzgibbon [195]. In 2005, Thirthala et al. [196] used a 1D radial camera to calibrate a fisheye camera. Both central and non-central cameras can be reduced to a 1D radial camera by assuming that the centre of radial distortion is identified. A simple fisheye distortion calibration is explained in [197]. A simple calibration process is proposed to avoid costly minimisation and optimisation algorithm to determine the accuracy.

4.2.1.3/ EXPERIMENT AND RESULT

PERSPECTIVE: ZED CAMERA. The ZED camera is referred to a compact camera developed by Stereolabs. It is a factory-calibrated camera, wherein the intrinsic parameters are provided in the technical documentation. Table 4.1 tabulates the intrinsic parameters provided by the manufacturer, where $(\mathbf{fx}, \mathbf{fy})$ is the focal distances and (μ_o, υ_o) is the principal point of the camera lens.

Instrinsic parameters		
	Left perspective camera	Right perspective camera
$(\mathbf{f_x}, \mathbf{f_y})$	1401.79, 1401.79	1398.1, 1398.1
(μ_o, υ_o)	1158.08, 646.31	1106.48, 653.948

Table 4.1: Factory calibration results.

The ZED camera's intrinsic parameters can also be manually estimated using a toolbox developed by Bouguet [188]. A $29mm \times 29mm$ chessboard is used as a reference image during the calibration process. Table 4.2 presents the estimated calibration values

using the toolbox. Upon comparing the two tables, the parameter values in both tables are almost identical. The resulting error between the two values can be reduced by increasing the number of image samples during the calibration process.

Instrinsic parameters		
	Left perspective camera	Right perspective camera
$(\mathbf{f_x}, \mathbf{f_y})$	1396.00, 1403.00	1391.00, 1401.00
(μ_o, υ_o)	1084.00, 605.00	1051.00, 597

Table 4.2: The calibration result estimated by Bouguet toolbox.

OMNIDIRECTIONAL: FISHEYE CAMERA The omnidirectional fisheye camera is a hybrid camera. This camera is the combination of a perspective camera and the fisheye lens to increase the FoV until 180°. However, the lens attached in front of the camera produces a barrel distortion that changes the camera geometry and the internal parameters.

The calibration method used Mei's calibration toolbox. The calibration toolbox is developed to estimate the value of eccentricity (ξ), which is crucial for enfolding the fisheye image onto the unit sphere. Similar to the perspective camera, a chessboard is used as the reference object. The distance between the camera and the pattern should be neither too far nor too close. Another critical point is the angle between the camera and the chessboard. This procedure is integral to reduce re-projection error. Besides, the number of sample images is sufficient for the calibration process to reduce error. Table 4.3 tabulates the intrinsic parameters of left and right fisheye estimated using the toolbox.

Intrinsic Parameters		
	Left fisheye camera	Right fisheye camera
$(\mathbf{f_x},\mathbf{f_y})$	657.00, 657.00	667.00, 666.00
(μ, υ)	610.00, 481.00	609.00, 481.00
ξ	1.4392	1.4709

Table 4.3: Intrinsic parameters of the left and right fisheye camera.

4.2.2/ EXTRINSIC CALIBRATION

As depicted in Section 3.3 (see Chapter 3), the Omni-Vision camera rig is composed of three cameras - two fisheye cameras and a high-resolution ZED camera. Images from the cameras were fused to obtain a full FoV horizontally and vertically. To

do so, a spherical model representation was used for all the cameras. This proposed camera rig is technically valid as an omnidirectional camera with a single effective viewpoint if the spherical approximation is employed.

The spherical representation is retained the properties of the projection geometry. The perspective camera's intrinsic parameters (ZED) and the fisheye camera were calculated and estimated using the toolbox developed by Bouget [188]. At the same time, the intrinsic parameter of the fisheye camera estimated by Mei calibration toolbox [191]. Next, the rotation and translation between the cameras were estimated using epipolar geometry that yielded the relative positions between all the respective cameras.

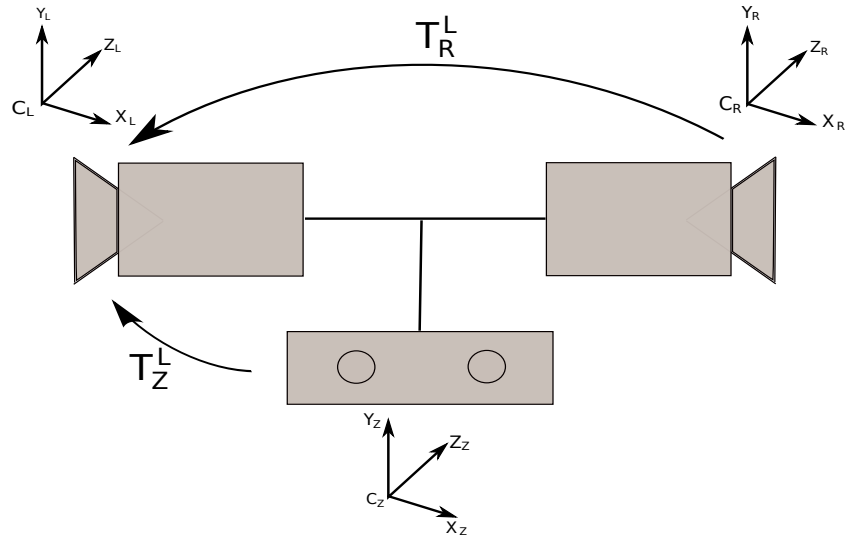


Figure 4.4: The transformation matrix obtained from the relationships between all the cameras. The two fisheye cameras with more than 180° FoV were placed back to back. At the anterior of the Omni-VIsion placed a high-resolution perspective camera called ZED..

The goal of the calibration process is to obtain the internal parameters of cameras. It used to project the image to the unit sphere, which denoted as $\mathbf{X_s^{Lf}}$, $\mathbf{X_s^{Rf}}$ and $\mathbf{X_s^z}$. From the extrinsic calibration, two transformation matrices were obtained that displayed the correlations between the cameras.

Let the left fisheye camera assigned as a reference, so the transformation matrices between the ZED camera and right fisheye camera are $\mathbf{T_R^L}$ and $\mathbf{T_z^R}$. The transformation matrix consists of information regarding rotation ($\mathbf{R}$) and the translation ($\mathbf{t}$). Figure 4.4 il-

lustrates the correlations among all the cameras. Once the images is projected to the unit sphere (the proposed camera rig), an overlap area was noted between all the cameras. The image from the ZED camera overlapped with the left and right fisheye cameras. The overlapping between the fisheye cameras occurred mainly because the FoV of both fisheye cameras exceeded 180°; displaying the relationship between those extracted from the overlapping area along the periphery of both fisheye.

The relationship was estimated using a straight forward approach of point matching. The IPOA was used to estimate the rotation between the two fisheye cameras and between the fisheye sphere and the projection of the ZED camera. The assumption was that it is a rigid transformation without the translation. The results of the extrinsic calibration are explained in detail in the next chapter.

4.2.2.1/ Unified Spherical Model

The unified spherical camera model was proposed by Barreto [111]. The image formation on the dioptric camera was affected by radial distortion. These models of projection were developed for cameras with radial distortion. This model is primarily an image produced by a camera that uses a lens or reflected mirror to expand the FoV.

In order to use this model, images containing radial distortion were projected onto the unified spherical model in the shape of a hemisphere. The radial distortion value should expand uniformly on the image. In contrast, the midpoint of the image is the minimum distortion point. The distortion value grows more extensively as it approaches the side of the image. The best-used example is an image produced by a dioptric camera or also known as the fisheye camera.

Courbon et al., [198] assessed the unified spherical model using a central catadioptric camera and fisheye camera. Figure 4.5 illustrates radial distortion distribution from a fisheye lens.

Hence, the point on the scene is not linear with the point on the dioptric image. The image formed in the dioptric camera with radial distortion adhered to the three-step procedure. Following a pinhole model, a world point $\mathbf{X}$ originates the projective ray denoted as $\mathbf{x} = \mathbf{PX}$, where $\mathbf{P}$ is a conventional $\mathbf{3 \times 4}$ projection matrix. This ray is transformed into a 2D projective point $\mathbf{x'} = \mathbf{Kx}$. This corresponds to the second step, where $\hbar()$ is a linear transformation, $\mathbf{K}$ (intrinsic parameters matrix) for dioptric cameras.

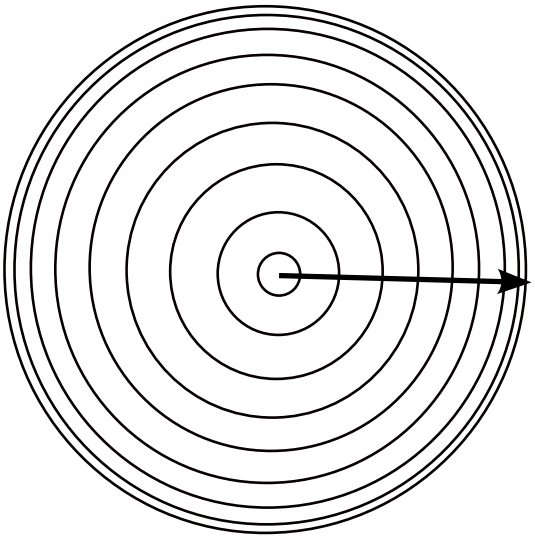


Figure 4.5: The radial distributes uniformly on the image taken from a fisheye camera. The direction of flash shows the increasing radial distortion value from a minimum at the midpoint of the image to and maximum near the edge of the image.

The non-linear transformation, δ equations (equation 4.12 dan 4.13) are related to image point $\mathbf{x''}$ with $\mathbf{x'}$ that models the radial distortion of the lens, where ξ is the amount of radial distortion. Figure 4.6 displays the step-by-step unified spherical model developed by Barreto.

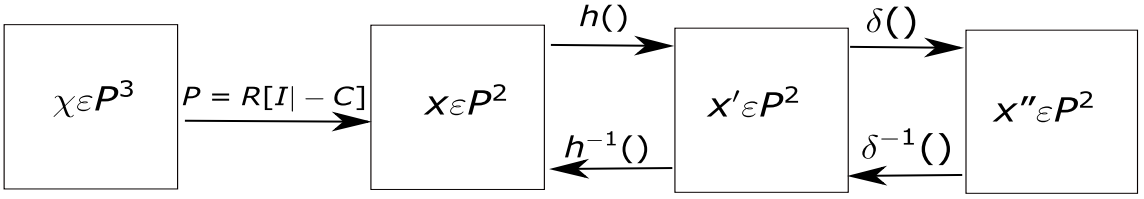


Figure 4.6: The step of unified spherical model developed by Barreto. The figure is the courtesy of Barreto [111].

$$\delta(\mathbf{x'}) = \left(2\mathbf{x'}, 2\mathbf{y'}, \mathbf{z'} + \sqrt{\mathbf{z'^2} - 4\xi\left(\mathbf{z'^2} + \mathbf{y^2}\right)}\right)^{\mathbf{T}}, \tag{4.12}$$

where $\xi < 0$

$$\delta^{-1}(\mathbf{x''}) = \left(\mathbf{x''z''}, \mathbf{y''z''}, \mathbf{z''^2} + \xi\left(\mathbf{x''^2} + \mathbf{y''^2}\right)\right)^{\mathbf{T}}, \tag{4.13}$$

$$\mathbf{x'} = \hbar\left(\mathbf{x}\right) = \mathbf{Kx}, \tag{4.14}$$

$$\mathbf{x}' = \hbar^{-1}(\mathbf{x}') = \mathbf{K}^{-1}\mathbf{x}', \tag{4.15}$$

Since $\mathbf{x}'$ is a homogeneous vector that represents a point in oriented projective plane, $\lambda\mathbf{x}'$ represents the same point whenever λ is a positive scalar.
Assuming $\lambda = \frac{1}{\sqrt{\mathbf{x}''^2+\mathbf{y}''^2}}$ in equation 4.13 , the following is gained:

$$\begin{aligned} \mathbf{x}' &= \frac{\mathbf{x}''\mathbf{z}''}{\sqrt{\mathbf{x}''^2 + \mathbf{y}''^2}} \\ \mathbf{y}' &= \frac{\mathbf{y}''\mathbf{z}''}{\sqrt{\mathbf{x}''^2 + \mathbf{y}''^2}}, \\ \mathbf{z}' - \xi &= \frac{\mathbf{z}''^2}{\sqrt{\mathbf{x}''^2 + \mathbf{y}''^2}} \end{aligned} \tag{4.16}$$

Using equation 4.16, a ray of light intersects on the paraboloid with a vertex $\mathbf{O}''$. The coordinate system of this paraboloid is the origin of that paraboloid with a vertex $\mathbf{O}'$. The origin coordinate paraboloid is given in the following equation.

$$\mathbf{x}'^2 + \mathbf{y}'^2 - (\mathbf{z}' - \xi) = \mathbf{0}, \tag{4.17}$$

Figure 4.7 portrays the position of all points. This figure explained the steps to project the world point $\mathbf{x}'$ to the image plane.

The model was extended by Mei [191] in which a calibration toolbox was developed for an omnidirectional camera. The model was enhanced by introducing another type of distortion called tangential distortion, incorporated with radial distortion. Figure 4.8 presents the Mei's projection model from the pinhole camera to a unified spherical model. The eccentricity, ξ parameter that reflected the amount of radial distortion. It depended on the mirror's shape for a catadioptric camera or the lens for a fisheye camera.

Upon comparing the value of ξ, since Barreto used a paraboloid, the value of eccentricity ξ obtained by Barreto is below 0. On the other hand, Mei used a unified spherical model, and the value of eccentricity exceeded one. Mei's projection model [191] was applied as a reference, and it outlined the procedures to place the image on the unified model.

Let us consider a 3D point $\mathbf{P} = (\mathbf{X}, \mathbf{Y}, \mathbf{Z})^{\mathbf{T}}$ on the camera frame, and it is projected

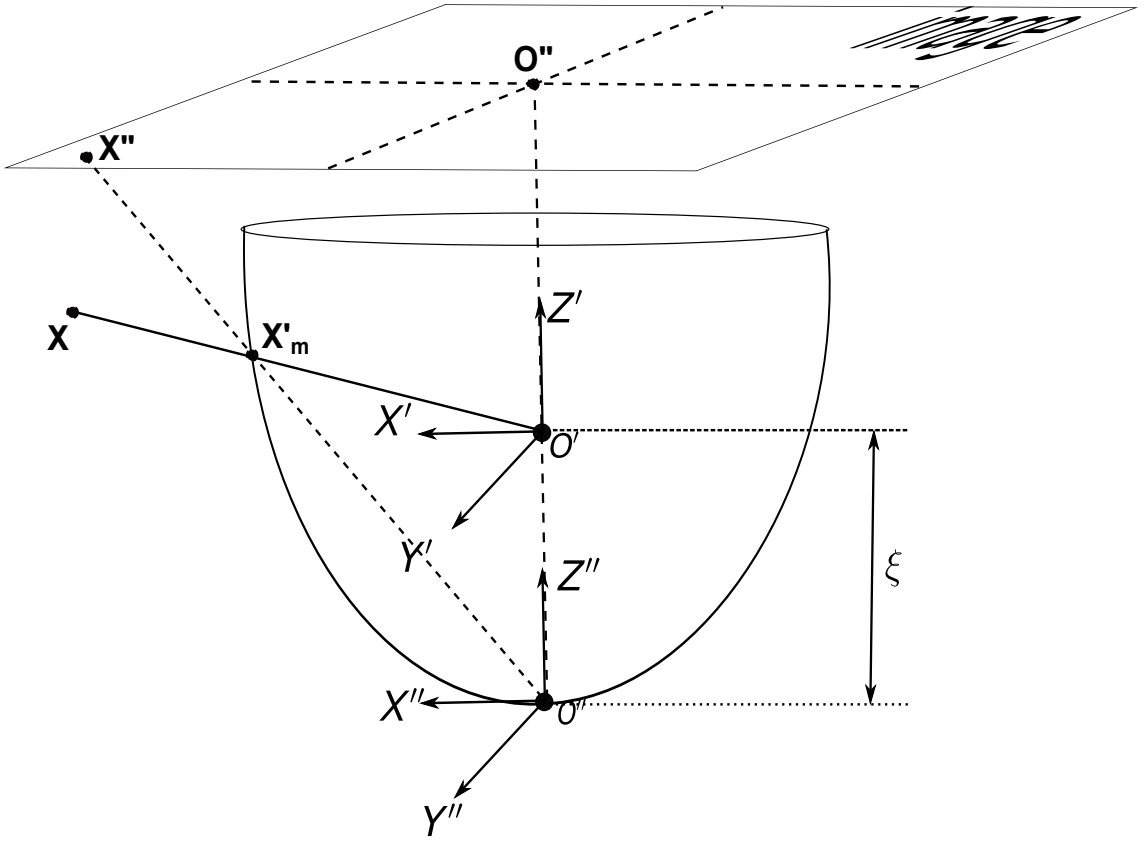


Figure 4.7: The paraboloid model for fisheye camera. The model explains the projection of image in space to the image plane. The paraboloid assumes radial distortion from the fisheye lens. The figure is courtesy of Barreto [111].

to the unit sphere with C_m as a center of the sphere.

$$\mathbf{P_s} = \frac{\mathbf{P}}{||\mathbf{P}||} = (\mathbf{X_s}, \mathbf{Y_s}, \mathbf{Z_s})^{\mathbf{T}}, \tag{4.18}$$

Next, point $\mathbf{P_s}$ is mapped to a new reference frame with a new centre $\mathbf{C_p}$.

$$(\mathbf{P_s})_{\mathbf{F_m}} \rightarrow (\mathbf{P_s})_{\mathbf{F_p}} = (\mathbf{X_s}, \mathbf{Y_s}, \mathbf{Z_s})^{\mathbf{T}}, \tag{4.19}$$

After that, the point is projected onto a normalised plane.

$$\mathbf{m_u} = (\frac{\mathbf{X_s}}{\mathbf{Z_s} + \xi}, \frac{\mathbf{Y_s}}{\mathbf{Z_s} + \xi}, \mathbf{1})^{\mathbf{T}} = \hbar \mathbf{X_s}. \tag{4.20}$$

The models of distortion (both tangential and radial) are added in the projection model.

Now, it contains three radial and two tangential distortion parameters.

$$\mathbf{x_c = x_1 + k_1 r^2 + k_2 r^4 + k_5 r^6 + 2k_3 xy + k_4(r^2 + 2x^2),} \tag{4.21}$$

$$\mathbf{y_c = y_1 + k_1 r^2 + k_2 r^4 + k_5 r^6 + 2k_4 xy + k_3(r^2 + 2y^2),} \tag{4.22}$$

where:

$$\mathbf{r = \sqrt{x^2 + y^2},} \tag{4.23}$$

and the sum of distortion is

$$\mathbf{m_d = m_u + D(m_u, V),} \tag{4.24}$$

where $\mathbf{V}$ is the coefficient of distortion.

$$\mathbf{V = (k_1, k_2, k_3, k_4, k_5, k_6),} \tag{4.25}$$

Finally, point $\mathbf{m_d}$ is projected to the image plane using $\mathbf{K}$, which is a generalised camera projection matrix. The values of $\mathbf{f}$ and $\boldsymbol{\eta}$ should also be generalised to the whole system (camera + lens).

$$\mathbf{p = Km_d = \begin{bmatrix} f_1\eta & f_1\eta\alpha & u_0 \\ 0 & f_2\eta & v_0 \\ 0 & 0 & 1 \end{bmatrix} m_d,} \tag{4.26}$$

where the $\mathbf{[f_1, f_2]^T}$ is the focal length, $\mathbf{(u_0, v_0)}$ is the principal point, and α is the skew factor. Finally, by using the projection model, the point on the normalised camera plane can be lifted to the unit sphere by using the following equation:

$$\mathbf{\hbar^{-1}(m_u) = \begin{bmatrix} \frac{\xi + \sqrt{1+(1-\xi^2)(x^2+y^2)}}{x^2+y^2+1} x \\ \frac{\xi + \sqrt{1+(1-\xi^2)(x^2+y^2)}}{x^2+y^2+1} y \\ , \frac{\xi + \sqrt{1+(1-\xi^2)(x^2+y^2)}}{x^2+y^2+1} - \xi \end{bmatrix}} \tag{4.27}$$

4.2.2.2/ EXPERIMENT AND RESULT

PROJECTION 2D FISHEYE ONTO THE UNIT-SPHERE: The next step after the calibration process is to use the parameters' values and project the 2D image from the fisheye

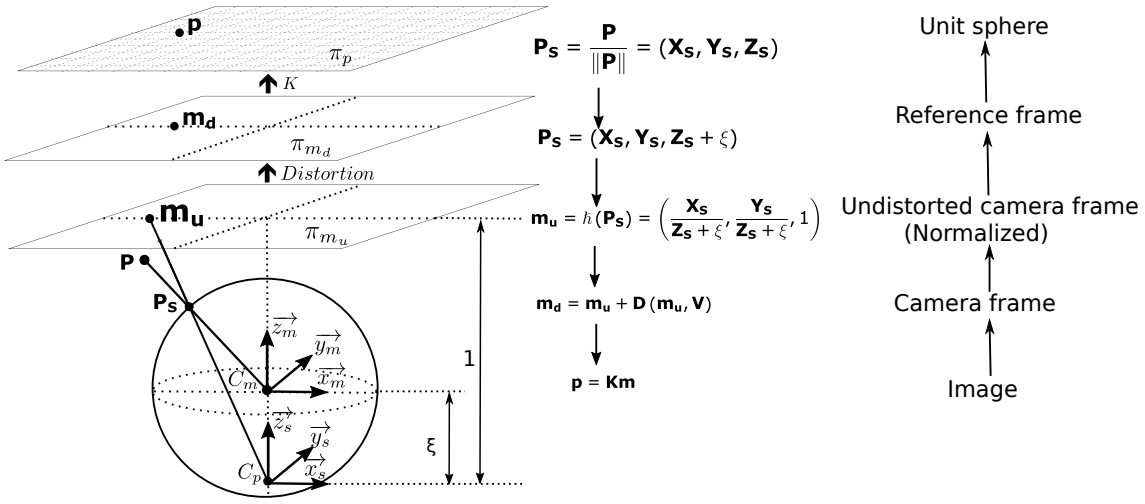


Figure 4.8: Mei's projection model from pinhole camera to unified spherical model. This figure is courtesy of Christopher Mei [191].

camera onto the unit sphere. As mentioned frequently in the previous section, the fisheye lens used in the Omni-Vision system has more than 180° FoV. As depicted in Table 3.1 regarding Fujinon fisheye lens specification, the FoV of the fisheye lens is 185°. Figure 3.4 shows the FoV of the Omni-Vision camera rig.

The coverage area of the individual fisheye lens is 185° vertically and horizontally. The angle θ refers to the angle covered by each fisheye camera (left and right). In contrast, angle α is the angle where both cameras have a mutual covering area (overlapping area). The angle α was expected to be 2.5° and fall beyond the fisheye image's perimeter. The mutual covering area is an important area. This area is exploited to fuse both cameras. The fusion process is explained in the next section. The rough idea is to develop and use the algorithm to enfold the 2D image captured from the fisheye camera using intrinsic parameters.

Algorithm 1: Projecting 2D fisheye image onto the unit sphere.

Result: Hemisphere of fisheye image.

1 Initialization : Instrinsic parameters from left and right fisheye camera and 2D image captured from left and right fisheye camera;

2 Project the 2D image onto unit sphere using mapping function Equation 4.27 from Mei projection model.;

Algorithm 1 displays the steps to project an image onto the unit sphere. Figure 4.9

portrays the results of the projection. The enfolding details were neglected. The purpose of this task is to perform an algorithm and deduce the possibility of the projection onto the unit sphere for the left and right fisheye cameras.

ESTIMATE THE EXTRINSIC PARAMETERS: In addition to intrinsic parameters, camera calibration must also include external calibration to determine the camera's position in the world coordinates. Depending on the position of the cameras used, the algorithm developed to yield desired outcomes. The proposed hybrid camera called Omni-Vision (see Figures 3.2 and 3.4), is a combination of three cameras. The combination includes left and right fisheye cameras that offer vertical and horizontal 360° (sphere) view, along with the combination of the globe and ZED camera. Since the ZED camera is positioned between the two fisheye cameras, the rigid transformation between the two fisheye cameras and the ZED camera must be taken into account.

In other words, the extrinsic calibration estimation is a direct reference to the proposed hybrid camera. Since this experiment estimated the rigid transformation between the left and right fisheye cameras, the left fisheye camera served as the coordinate reference. Figure 4.10 illustrates the view captured from the proposed Omni-Vision camera rig by referring to the cameras' position.

ESTIMATE THE EXTRINSIC PARAMETERS: Between the Two Fisheye Cameras

As mentioned earlier, the left and right fisheye shared an overlapping area. The overlapping area appeared along the periphery of the fisheye camera. The rigid transformation contained information about the rotation and translation of the camera. The rigid transformation extracted using the mutual points between the two fisheye cameras. The one-to-one matching between the two images (left and right fisheye images) refers to the technique employed to estimate rigid transformation. In order to minimise the development of error during the calculation, the number of clicking points should increase using the manually click method. Figure 4.10 illustrates the view captured from the proposed Omni-Vision camera rig by referring to the cameras' position.

Next, the manually selected features were projected on the unit sphere using the intrinsic parameters calibration results tabulated in Table 4.3. The 3D point projected onto the unit sphere was applied to estimate the rigid transformation between the left and right fisheye cameras. Figure 4.12 displays all the projected points onto the unit sphere.

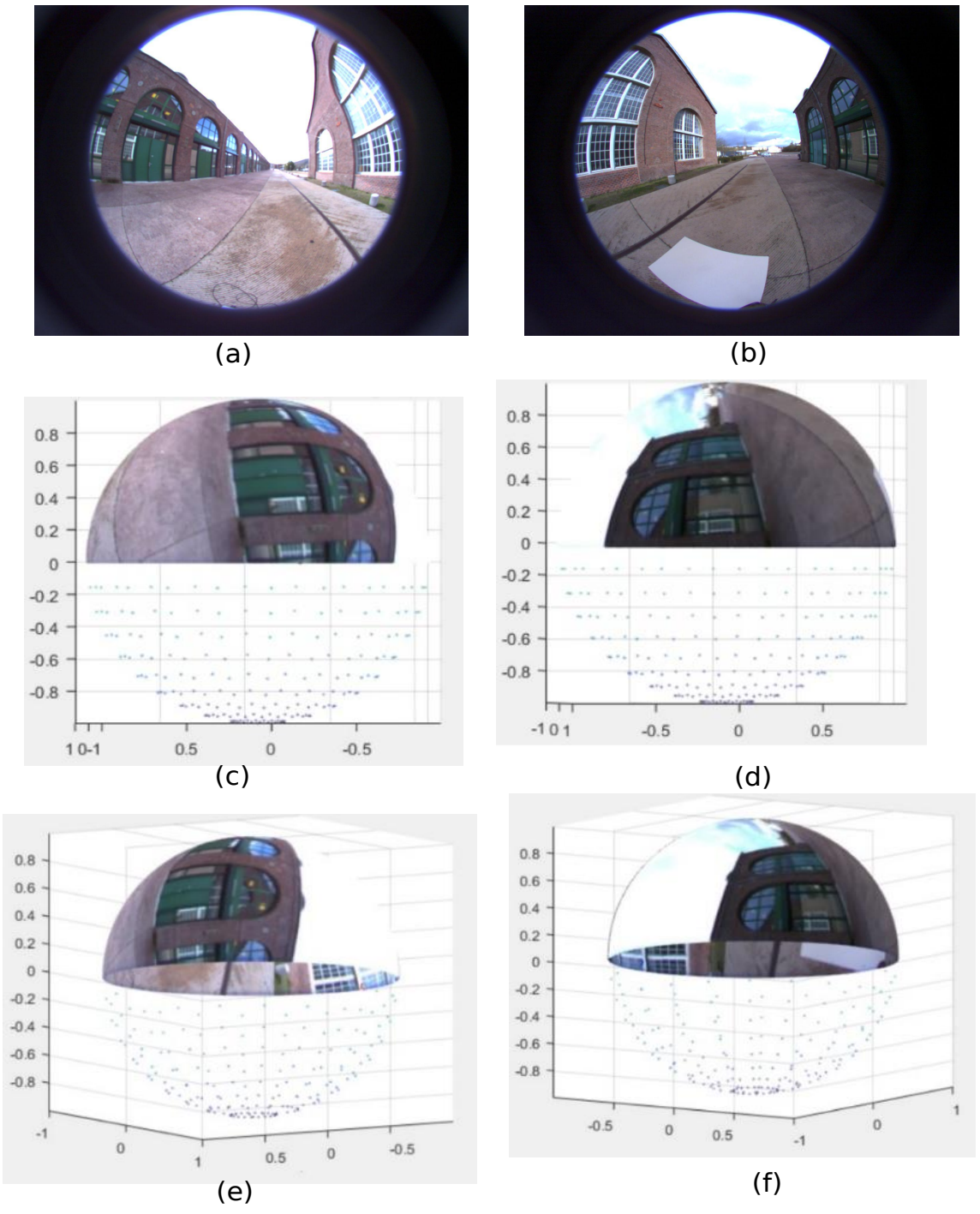


Figure 4.9: (a) and (b) are the 2D images taken from the left and right fisheye cameras. (c) and (d) are the hemisphere of left and right images after they were projected onto the unit sphere. (e) and (f) are the orthography views of left and right hemispheres to show the complete form of hemisphere.

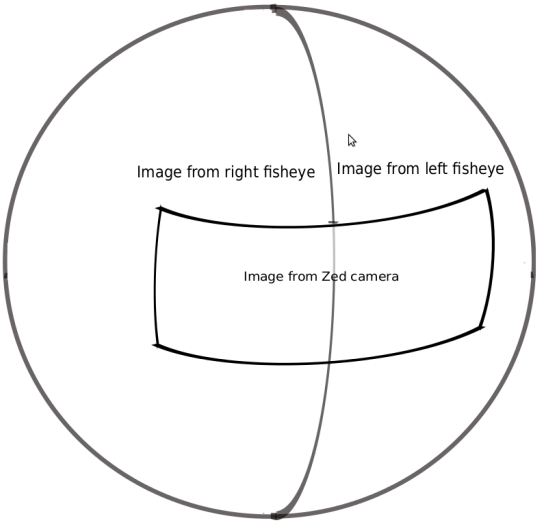


Figure 4.10: The view from Omni-Vision camera rig.

The values of estimated rigid transformation are listed in the following:

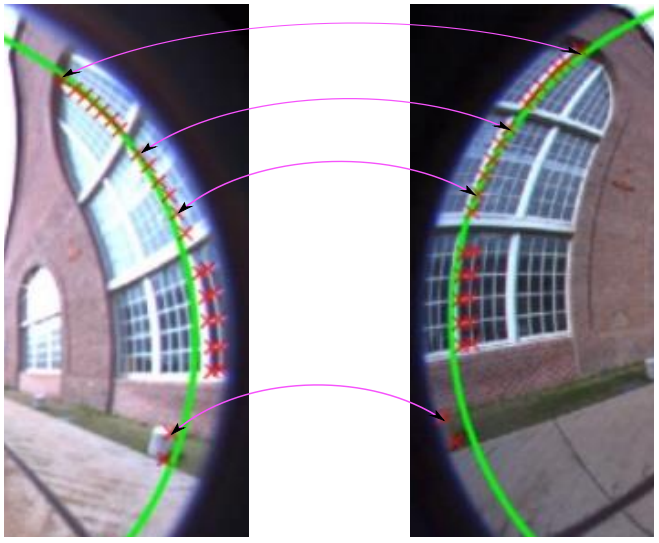
$$\mathbf{T} = \begin{bmatrix} \mathbf{-1.0000} & \mathbf{-0.0017} & \mathbf{0.0073} & \mathbf{-0.0039} \\ \mathbf{-0.0019} & \mathbf{0.9997} & \mathbf{-0.0250} & \mathbf{0.0085} \\ \mathbf{-0.0073} & \mathbf{-0.0250} & \mathbf{-0.9997} & \mathbf{0.1143} \\ \mathbf{0.0000} & \mathbf{0.0000} & \mathbf{0.0000} & \mathbf{1.0000} \end{bmatrix}, \tag{4.28}$$

The rigid transformation was obtained using SVD, and the result was close to the pure rotation.

The initial result of the rigid transformation ($\mathbf{T}$) in Equation 4.28 appeared to be almost excellent. However, as noted in the last column that contained the vector translation, there is the existence of translation parameters in projecting the features points onto the unit sphere. Based on the Omni-Vision camera configuration and the desired output, the selected features from left and right fisheye cameras were projected onto the unit sphere at the exact centre point (see Figure 4.14). According to the Omni-Vision setup, the images from both fisheye cameras used Mei's projection model to project the image to the unit-sphere. The points on the left fisheye image projected to the unit sphere using the projection model. In the same manner, the same method applied to the right fisheye image. Based on this method, if both projection images combined, the spheres will share the same centre point. Thus, base on this theory, the translation, which is the distance from both unit sphere centre point, is equal to zero. Theoretically, there is no translation, and the value should be zero in the last column in matrix $\mathbf{T}$. Another method was incor-

2D image from left fisheye
(a)

2D image from right fisheye
(b)

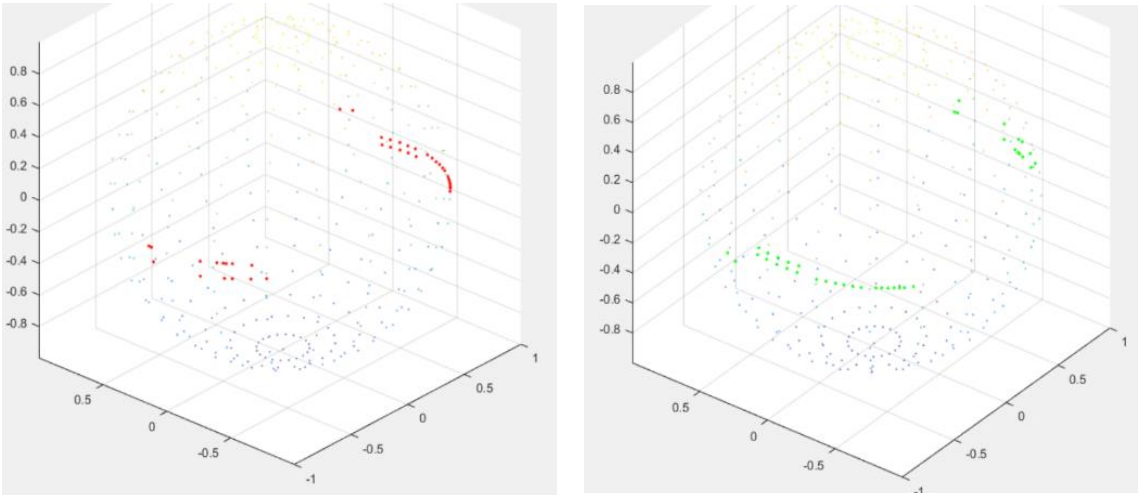


(c)

Figure 4.11: The 2D images from left fisheye camera (a) and right fisheye camera (b). An example of one-to-one manually clicking (c).

porated to estimate the rigid transformation with pure rotation by forcing the translation vector to zero. This iterative strategy integrated the IPOA method to determine pure rotation.

The rotation matrix has three degrees of freedom. The magnitude and the direction of $\mathbf{R}$ are defined by the arbitrary and the angle of axes $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$. (see Equations 2.13, 2.14 and 2.15). The sum of rotation, $\mathbf{R}$ is described in Equation 2.16. Where, $\theta_{\mathbf{x}}$, $\theta_{\mathbf{y}}$ and $\theta_{\mathbf{z}}$ are Euler angles that affected the rotation matrix $\mathbf{R}$. The Euclidean distance between reference 3D points (left fisheye) and 3D points from right fisheye was minimised using a



(a) (b)

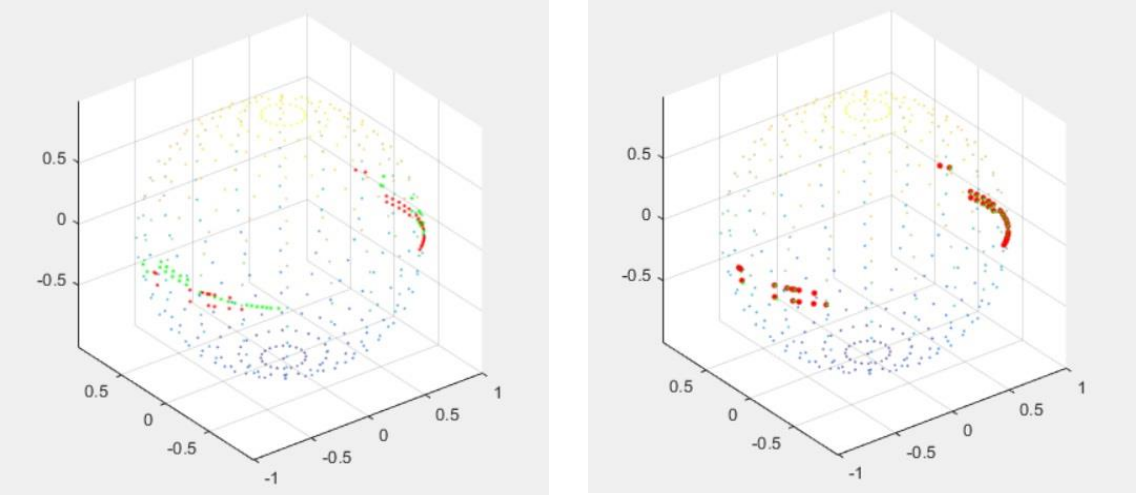
Figure 4.12: The features points form both left and right fisheye then projected onto the unit sphere to estimate the rigid transformation. Figure (a) shows all the projected point from the left fisheye, and Figure (b) shows all the projected point from the right fisheye.

cost function.

The new rigid transformation matrix is as follows:

$$\mathbf{T} = \begin{bmatrix} \mathbf{-1.0000} & \mathbf{-0.0048} & \mathbf{0.0085} & \mathbf{0.0000} \\ \mathbf{-0.0045} & \mathbf{0.9994} & \mathbf{-0.0335} & \mathbf{0.0000} \\ \mathbf{-0.0087} & \mathbf{-0.0334} & \mathbf{-0.9994} & \mathbf{0.0000} \\ \mathbf{0.0000} & \mathbf{0.0000} & \mathbf{0.0000} & \mathbf{1.0000} \end{bmatrix}, \tag{4.29}$$

The new rigid transformation matrix displayed in Equation 4.29 was applied to project the features points onto the unit sphere. Figure 4.13 illustrates the new estimation rigid transformation matrix, which revealed exceptional alignment between the two-unit spheres. The projection points on the left fisheye (red colour) and right fisheye (green colour) projected on the same unit sphere. After applying the transformation matrix, the left fisheye points rotate 180° and then aligned with the points on the left fisheye (as manually clicked one to one). This result served as the basic structure for the proposed Omni-Vision system.



(a) (b)

Figure 4.13: The initial points project on the same unit sphere. (a) is points on the sphere using the initial rigid transformation (Equation 4.28). (b) is after using the new rigid transformation matrix (Equation 4.29), where the points from the left and right fisheye cameras are aligned on the sphere.

4.3/ MULTI-CAMERA FUSION

Fusion is the process of merging images from multiple cameras into a single image. The inclusion of images must first refer to a foundation on which all images will refer to that same basis. The unified spherical model is a reference that should be adhered by each image. Each image captured by the camera becomes non-homogeneous due to the camera's internal and external parameters. The camera type also contributes to non-homogeneity due to the presence of different degrees of distortion. An example is to combine images from perspective and fisheye cameras. The method applied in this thesis is to project all the images onto a unified spherical model before combination.Finally, all the image combined to produce a unique image consist of the three images from the three cameras. In this technique, distortion is amalgamated into the image to produce a perfect spherical image.

The main objective of the proposed camera setup is to produce a 360° FoV unit sphere with high quality to handle visualization. The proposed setup used a fisheye camera that does not generate any blind spots in front of the camera, and obtained a clean image. It can capture an image with useful information from the scene..

The discussion in Section 4.2 is extended in Section 4.3. The method is explained

in detail, and understanding the process is essential for further research. In this section, the actual image obtained from fisheye cameras and ZED camera were combined to produce a single image of the Omni-Vision hybrid camera. The following describes pure rotation without translation.

4.3.1/ PURE ROTATION REGISTRATION

A common way to fuse the image from Omni-Vision is to calibrate the camera set up so that the relative position between the cameras are known. Let the features from the left and right fisheye cameras (projected onto unit sphere) reflect $\mathbf{P^L}$ and $\mathbf{P^R}$, respectively. The transformation between the two fisheye cameras is signified as $\mathbf{T} \in \mathbf{R^{4\times4}}$, hence:

$$\mathbf{P^L = TP^R}, \tag{4.30}$$

The transformation matrix, $\mathbf{T}$, was estimated in accordance to that depicted in [199]. As the Extrinsic Calibration method used in [200] or Singular Value Decomposition (SVD) [56] failed to correct or recover the transformation matrix $\mathbf{T}$. This problem is due to the pure rotation properties. Therefore, a pure rotation matrix was employed to address this issue by enforcing the transformation matrix with zero translation [201], as indicated in the following:

$$\min_{\mathbf{R}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{\Psi}(||\mathbf{P_i^L - RP_i^R}||), \quad \mathbf{s.t.} \quad \mathbf{RR^T = 1, det(R) = 1}, \tag{4.31}$$

Where $\mathbf{R}$ is the desired pure rotation matrix, $\mathbf{\Psi(\cdot)}$ function is the Huber-Loss function for robust estimation. By solving the above equation, a pure rotation matrix that minimised the registration errors between the two fisheye cameras' fusions was obtained. In this case, the IPOA was adopted to solve the problem.

At the initial stage, the positions of the left (red colour) and right (blue colour) hemispheres appeared to overlap each other. The centre of both hemispheres also overlapped, and indicated no translation between the left and right images. The reference image referred to the image from the left camera. The right fisheye image rotated at the centre of rotation (the centre of the hemisphere) if the transformation matrix was applied on it, thus producing a unit sphere. Figure 4.14 illustrates the position of the left and right hemispheres. The left-hand-side image presents the overlapping left and right hemi-

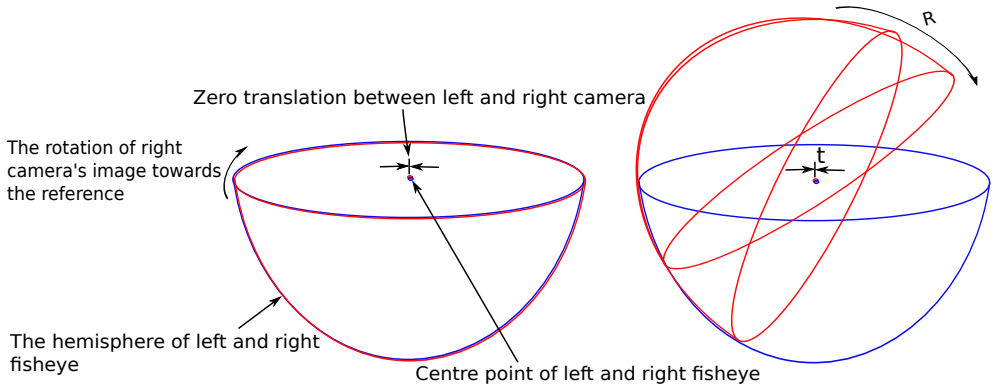


Figure 4.14: Pure rotation and transformation matrix that contains rotation with zero translation.

spheres at the same position and orientation. After applying the transformation matrix, the right fisheye hemisphere rotated 180° to produce a unit sphere.

4.3.2/ FUSION TWO FISHEYE CAMERAS

The fusion of two fisheye images was performed by using internal parameters and a rigid transformation matrix containing rotational and translation information. The internal parameters were used to produce the exact hemisphere shape of each camera. Next, the relative relationship between the left and right fisheye cameras was estimated to merge where the points (left and right fisheye cameras) in the overlapping areas were concatenated to generate a complete sphere. The two fisheye images were successfully combined in [202]. Figure 4.15 portrays the initial result of the fusion between the two fisheye images.

4.3.2.1/ EXPERIMENT AND RESULT

The internal parameters of the left and right fisheyes were applied to the respective camera to generate hemispheres. Figure 4.16 illustrates the left and right fisheye hemispheres. Since the left and right cameras used are different, the two cameras' internal parameters will also be different (see Table 4.3). The rigid transformation matrices displayed in Equation 4.29 were applied on the right fisheye's hemisphere to produce a complete sphere.

Figure 4.15: The first attempt to fuse two fisheye images onto the unit sphere using the transformation matrix $\mathbf{T}$ with pure rotation and zero translation.

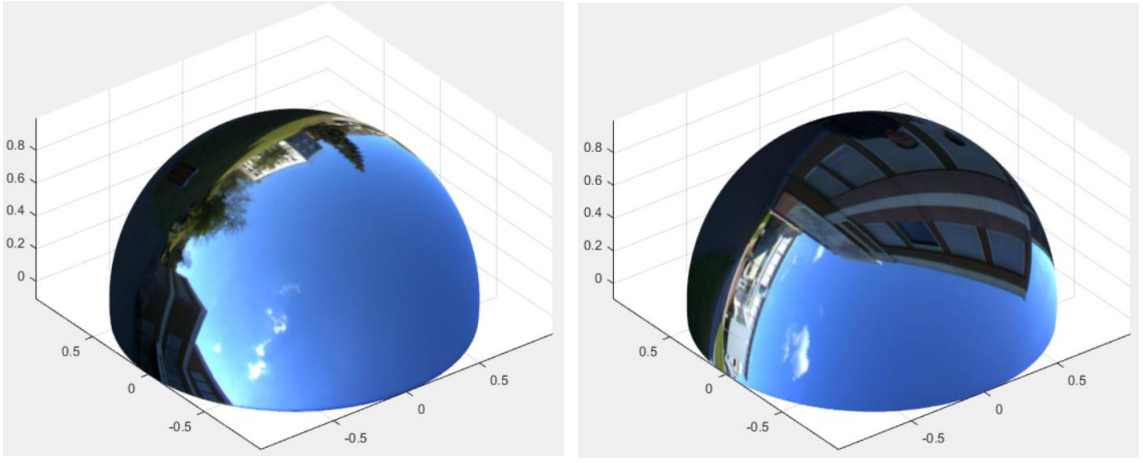


Figure 4.16: The 2D images from left and right fisheyes projected onto the unit sphere to produce hemispheres for both cameras.

Figure 4.17 portrays the full spherical result after combining the left and right hemispheres. The vertical lines on the spheres signify the differences in intensity between the left and right fisheye images. Figure (a) displays the image seen from the front of the Omni-Vision camera, while Figure (b) is rearview.

ESTIMATION THE THREE DIMENSIONAL REGISTRATION ERROR

The computation of registration error during mapping on the unit sphere was performed to verify the registration method's efficiency. The proposed projection error analy-

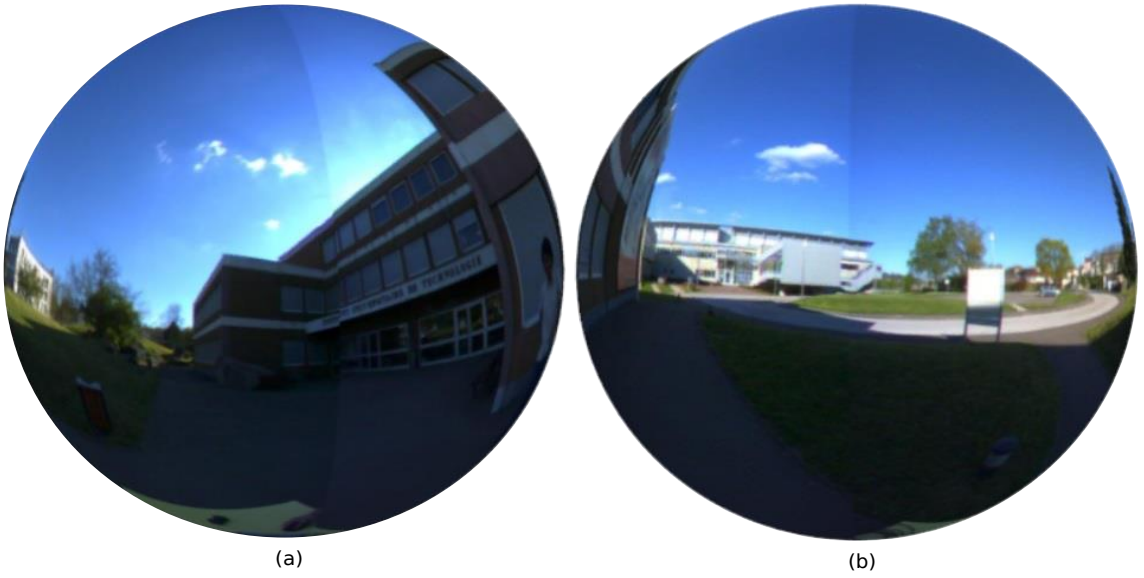


Figure 4.17: The full sphere for front view (a) and back view (b) using the Omni-Vision camera rig.

sis aims to investigate and measure the quality of the projected image on the overlapping area. The overlapping area is the mutual area looking from both fisheye cameras. This overlapping area locates along the periphery of left and right fisheye cameras. The Root Means Square Error (RMSE) is frequently applied to calculate errors. Both rigid 3D transformation matrix and parameter ξ (obtained from calibration) were employed to determine the residual errors, which reflect the difference between the points on the left and right hemispheres. Figure 4.18 exhibits the two correspondence points each from left and right fisheye that did not overlap at the same coordinate, where (u_1, v_1) and (u_2, v_2) are the 2D coordinates on left and right fisheye images.
Three methods have been compared:

1. IPOA: The method proposed in this study - Pure rotation was estimated based on matching features and IPOA.

2. SVD : The transformation (rotation and translation) matrices were estimated using matching features with SVD [200].

3. CNOC : The method used in Calibration Non Overlapping Cameras [199].

The image sequences were taken from several different settings, including inside and outside a building with bright and cloudy weather conditions. The feature points were

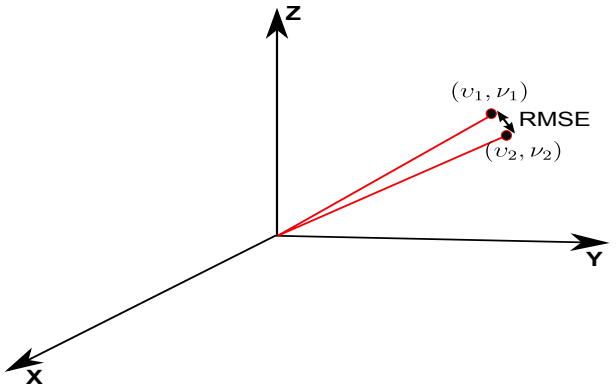


Figure 4.18: The calculation of registration error.

selected in the over-lapping area. The same data set was used for all three methods. Figures 4.19, 4.20 and 4.21 indicate that the proposed method yielded the lowest registration errors. The experiment has proven that the technique of IPOA could reduce 3D registration errors.

Figure 4.19, 4.20 and 4.21 shows that the proposed method gives low registration

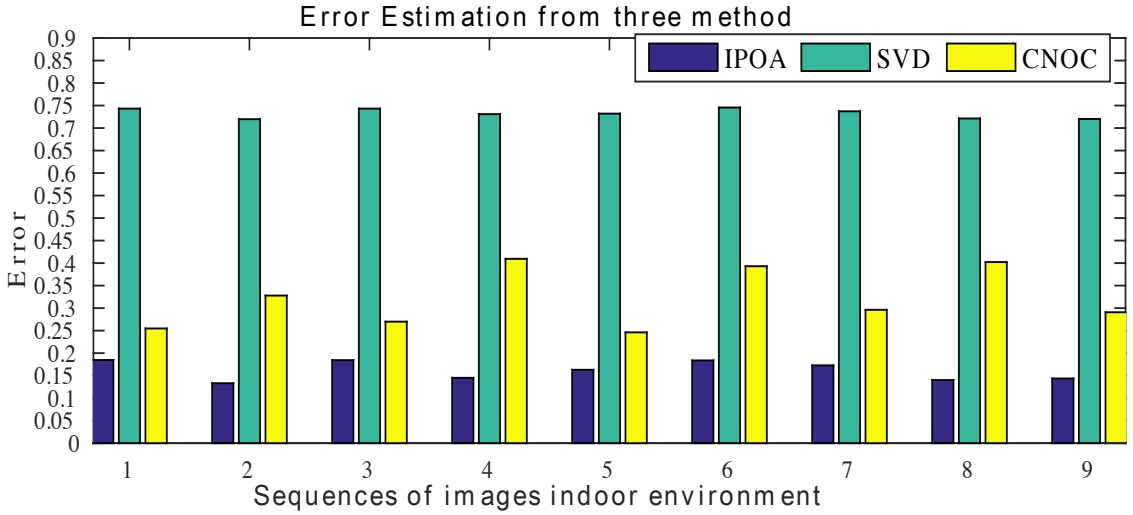


Figure 4.19: Sample images captured from inside of the building.

on the overlapping area. In detail, the precise error analysis by calculating the average error on each environment. The average error is 16.0 pixels in indoor, 18.77 pixels in outdoor clear sky and 17.72 in outdoor cloudy. The overall average is 17.49 pixels. The difference between the average of each environment and the overall average is around **±1** pixel. This difference can be concluded that registration errors in the overlapping area are not significant in the environment's different conditions.

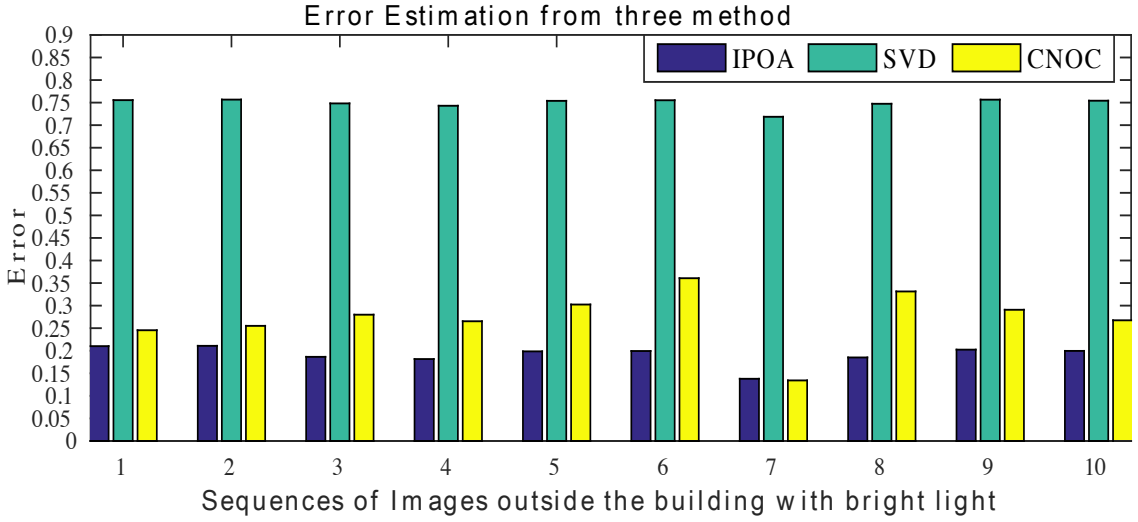


Figure 4.20: Sample images were taken at outside building with the bright condition.

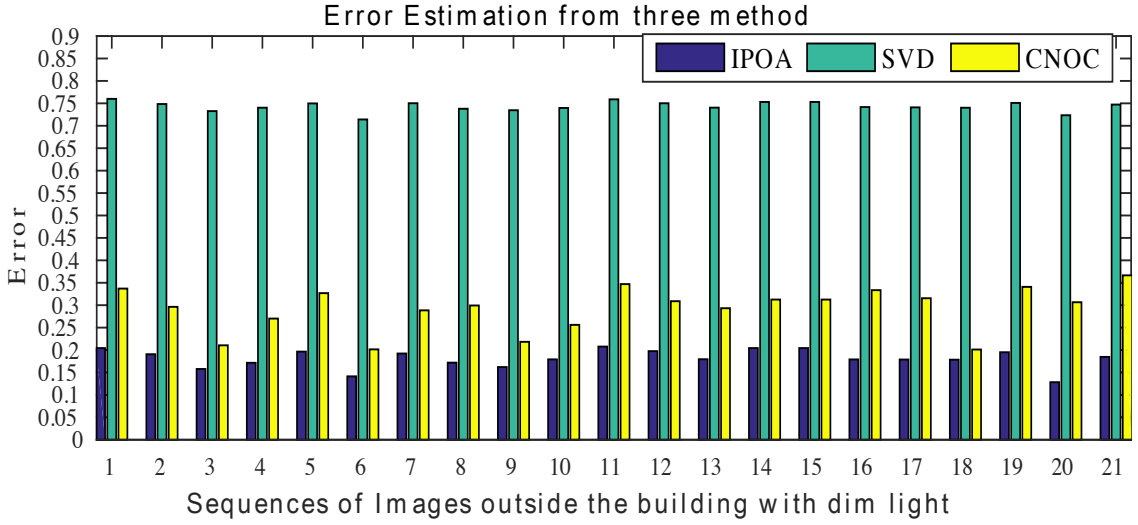


Figure 4.21: Sample images were taken at outside building with the cloudy condition.

4.4/ CAMERA CALIBRATION USING ZERO-DEGREE OVERLAPPING CONSTRAINT

One of the research objectives is to propose a new method for a multi-camera setup. As mentioned in Section 3.3, the Omni-Vision sensor is composed of three cameras: two fisheye camera and one high-resolution ZED camera. The two fisheye cameras with more than 180° FoV were placed opposite each other. The ZED camera was positioned to view the anterior of the camera rig with a high-resolution view. Since the FoV of the fisheye

camera exceeds 180°, the proposed setup has an overlapping area between the two fisheye cameras along the periphery and the overlapping between the two fisheye cameras and ZED camera. This study made use of the advantages offered by the overlapping FoV of all cameras. It is the primary aim of this study to project the three images onto the unit sphere. This point was enabled using a new fisheye camera calibration method. It used the constraint of overlapping zero-degree lines of the two fisheye cameras utilising a Unified Camera Model.

Nevertheless, in order to yield outstanding outcomes, several preliminary findings were determined and verified first. The internal parameters of the camera were the most critical points to consider. It assumed that both cameras were positioned correctly and had similar magnification value. This point indicates that both left and right cameras' centre points fall on the same line and direction. The objects displayed in the overlapping area had the same size on both cameras (see Figure 4.22). In order to ease the experimentation, the parameters in both cameras were assumed to have no significant variance. A unique parameter was incorporated for the formation of a hemisphere, namely eccentricity, ξ. As previously explained, the value of eccentricity influences the shape of the hemisphere. The mistaken eccentricity may cause an inaccurate position on the 180° lines of the real-world camera and the image plane. This condition could result in a significant difference between an object's position in the real world and its position in the unit sphere. Based on Figure 4.23 below, the correct value of eccentricity should locate the black dot on the zero degrees plane of a hemisphere or a unit sphere. In order to obtain the correct value of eccentricity, it should be re-estimated using a qualitative experiment. Figure 4.25 and 4.28 illustrate the experimental set-up to re-estimate the value of ξ. The experiment was conducted based on the following **Assumptions**:

- If ξ is estimated correctly, the 180° line of the fisheye camera should ideally lay on the zero degrees plane of the Unit Sphere. The image should be projected on the 180° line, as illustrated in Figure 4.23.
- The projected image should be not stretched out on the unit sphere. Figure 4.24(a) displays the correctly estimated ξ. The size of the chessboard is not stretched at the edge of the image. Figure 4.24(b) uses a high value of ξ, where the chessboard stretched especially at the edge of the image.
- The correct calibration (registration) of the multi-fisheye camera setup is derived

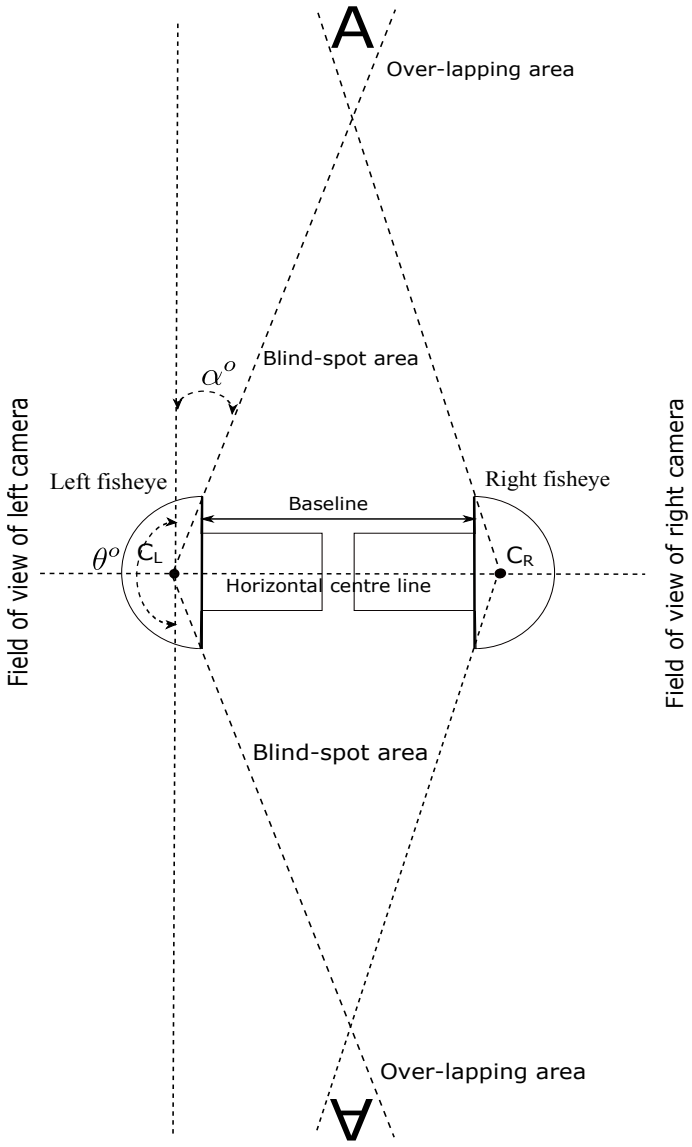


Figure 4.22: The image formation from the two fisheye cameras overlaps at the boundaries of the two fisheye cameras. The assumption that the exact size of the image 'A' captured on both camera's image plane. The size of the blind-spot area varies by the baseline distance. Minimizing the baseline will also minimize the blind-spot area. C_L is the left camera centre, and C_R is the right camera centre.

from the correct overlapping area.

4.4.1/ Zero-crossing Plane Distance Minimization

Let $\{\mathbf{x}_i\}_{\mathbf{i=1}}^{\mathbf{n}}$ and $\{\mathbf{P}_i\}_{\mathbf{i=1}}^{\mathbf{n}}$ be set of points located on the 180° line of the fisheye image and their projections onto a unit sphere, receptively. Figure 4.28 shows the unit sphere enfolded using initial and estimated ξ. Simultaneously, the calibration problem can be

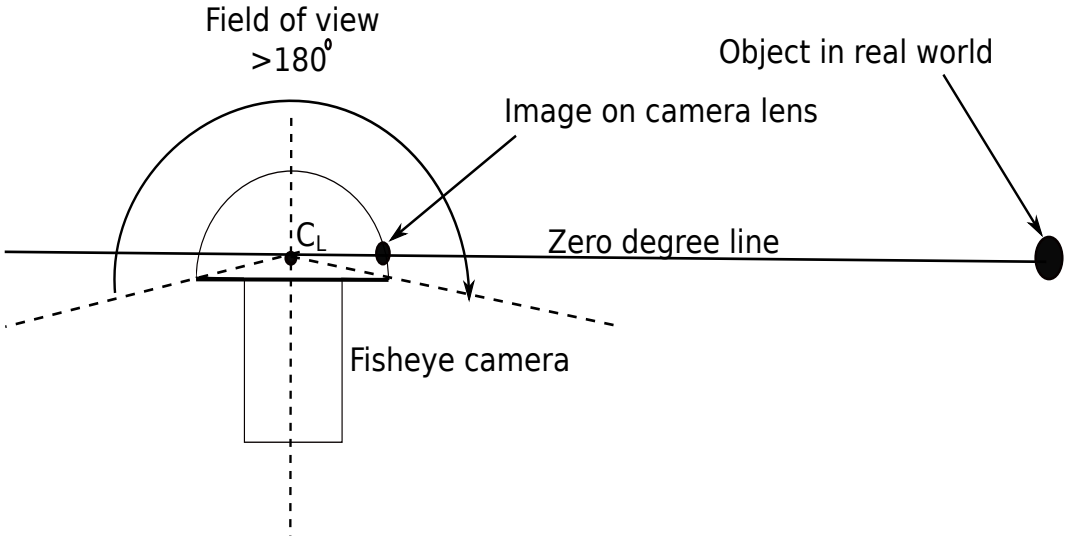


Figure 4.23: The example of experiment set-up to re-estimate the value ξ.

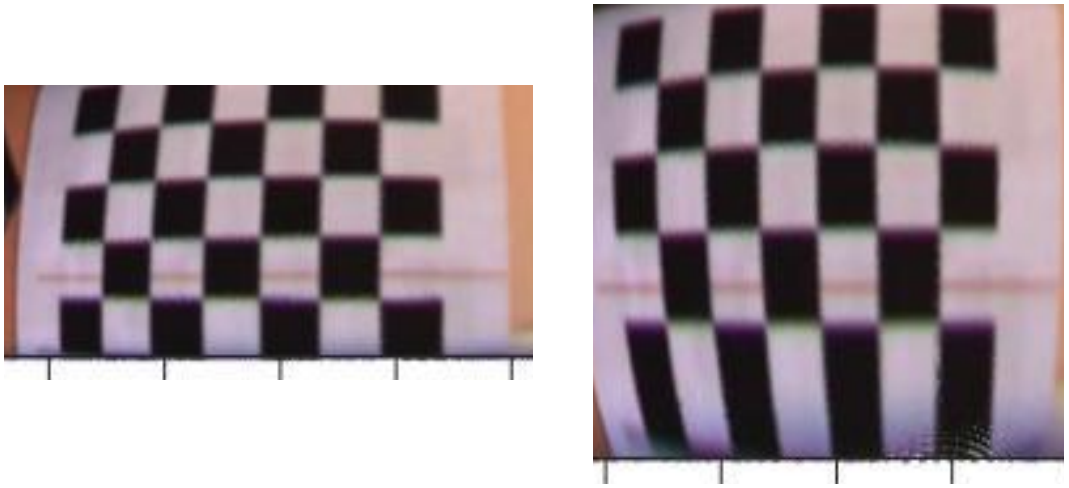


Figure 4.24: The comparative results using the correct estimation and fault values of ξ. (b) the image stretched at the edge of the fisheye image.

simplified as a minimization problem such that the distance between the 180° line and the zero-crossing plane of the unified sphere is minimized. In other words, the distance between $\mathbf{P}$ and the Zero-crossing plane of the unit sphere is required to be minimized. The formula is denoted as follows:

$$\min_{\xi} \sum_{i=1}^{n} \psi(\|\mathbf{f}(\mathbf{x_i}, \xi)\|_2), \tag{4.32}$$

where $\mathbf{f}(\mathbf{x}, \xi)$ Equation 4.27 is the mapping function from the fisheye image ($\mathbf{x}_i$) to the unified camera model ($\mathbf{P_i}$), operator $\|\cdot\|$ stands for the *l2*-norm, $\psi(\cdot)$ function is the

adopted Huber-Loss function for robust estimation purpose. Since the mapping function $\mathbf{f}(\mathbf{x}, \xi)$ in Equation 4.32 is not linear, the Interior Point Optimization Algorithm scheme, a global non-linear optimization method, was employed.
The overall structure to re-estimate the eccentricity may be summarized in **Algorithm 2**.

Algorithm 2: Re-estimate the value of eccentricity, ξ

Result: To estimate the correct value of ξ

1 Initialization : Start with the initial value eccentricity from the calibration process;
2 Do iteration to estimate ξ using cost function by minimizing the z-component;
3 if ξ previous = ξ current;
4 Exit iteration;
5 Otherwise continue the iteration goto step 2.

4.4.1.1/ EXPERIMENT AND RESULT

The camera rig baseline (from the left fisheye lens to the right fisheye lens) measured. Two parallel lines with equal distance from each other and the centre line are drawn. The camera platform is faced and aligned in front of the pattern (chessboard). The centre line touches the edges of both fisheye images (see Figures 4.25 and Figure 4.28). The middle part of the parallel line (green and red) corresponds to the edge of each camera's fisheye lens. The green line (left fisheye) and the red line (right fisheye) is a 180° fisheye lens line, forced to the zero planes when fisheye the image is projected onto the unit sphere. It is done by expanding the cost function for estimate ξ, which minimizes the pixel of z-component on the selected row using IPOA.

The edge of both fisheye images touched the centre line (black) on the chessboard pattern. The line that appeared on the chessboard pattern corresponds to the 180° line of the left and right fisheye lens. The visible track of line around the edge of the image signified that the unit sphere's zero-crossing line should fall on the circular line. Figures 4.26 and 4.27 display the construction of the 180° circular line. The circular line was defined by assuming that all points fell on the pattern corresponding to the 180° line of the fisheye cameras.

Figure 4.29 displays four images captured from the same fisheye camera. All the

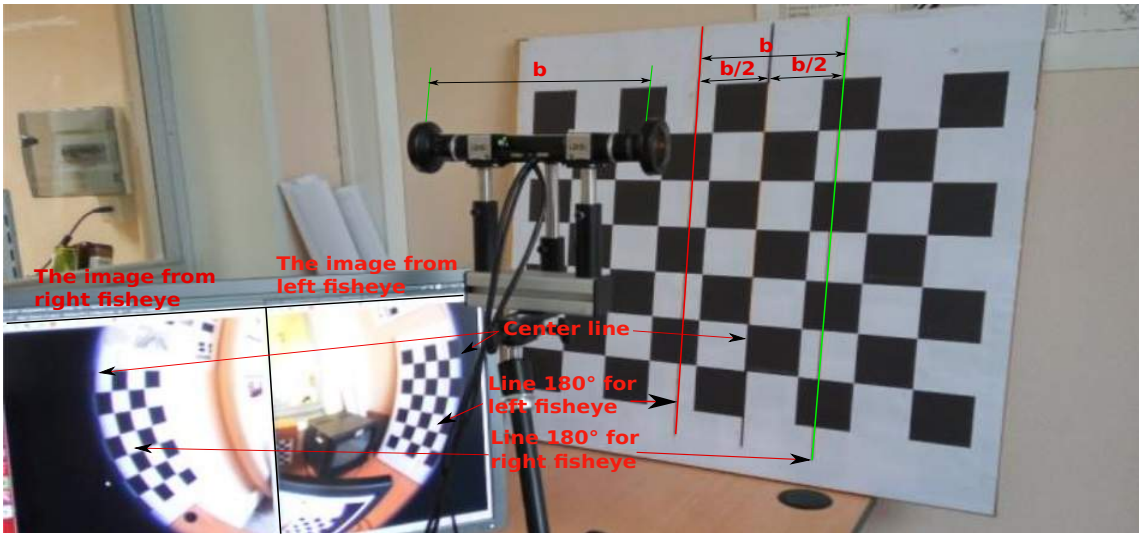


Figure 4.25: Experimental setup to calibrate the value of ξ. The baseline of the camera rig (from the left fisheye lens to the right fisheye lens) was measured. Two parallel lines with the same distance to each other and a centre line displayed a pattern. The rig was facing in front of the pattern. The centre line touched the edges of both fisheye camera images.

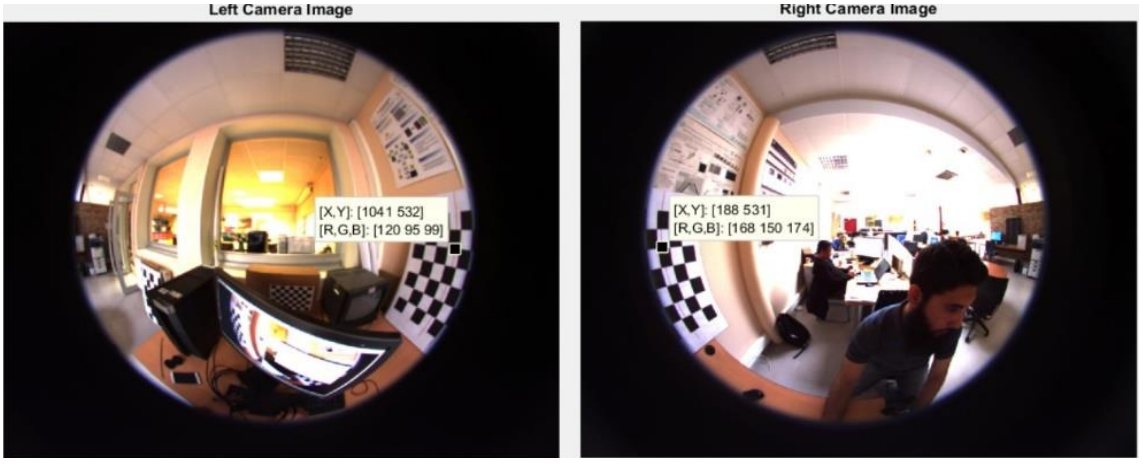


Figure 4.26: The 2D images captured from left and right fisheye cameras. The circular 180° line should fall and overlap on the pattern.

images are projected to the unit sphere using different values of ξ. In determining the value of ξ, Barretto and Mei proposed different ranges of value. In Barreto's projection model, the value of ξ is below 1 (one). While, in Mei projection model, the value of ξ must exceed 1 (one). Since this studies used Mei's projection model, the estimated value of ξ is **1.5632**. This estimated value serves as the initial value of ξ, which is used to recalculate in order to obtain the optimum value of ξ. As explained previously, the correct value of ξ

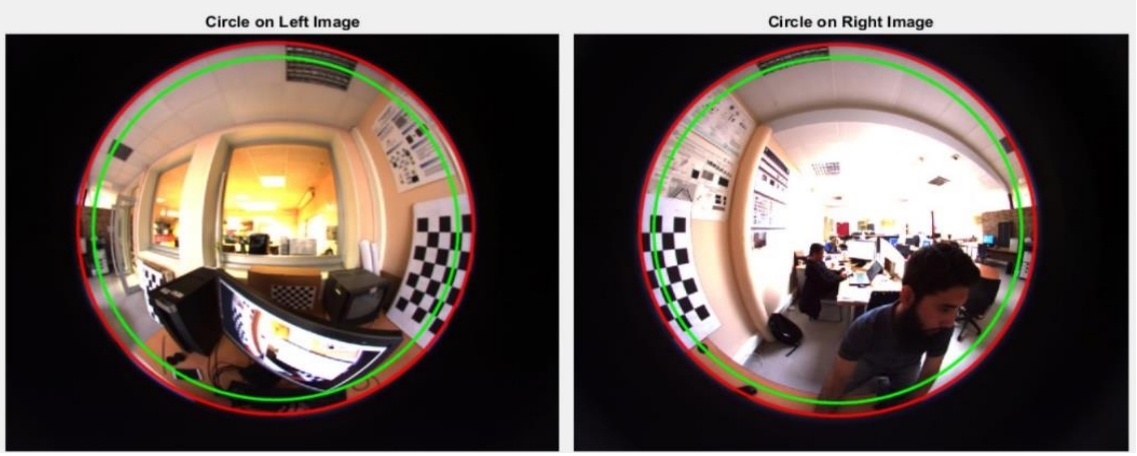


Figure 4.27: The circular line that correctly overlaps on the line on both left and right 2D images.

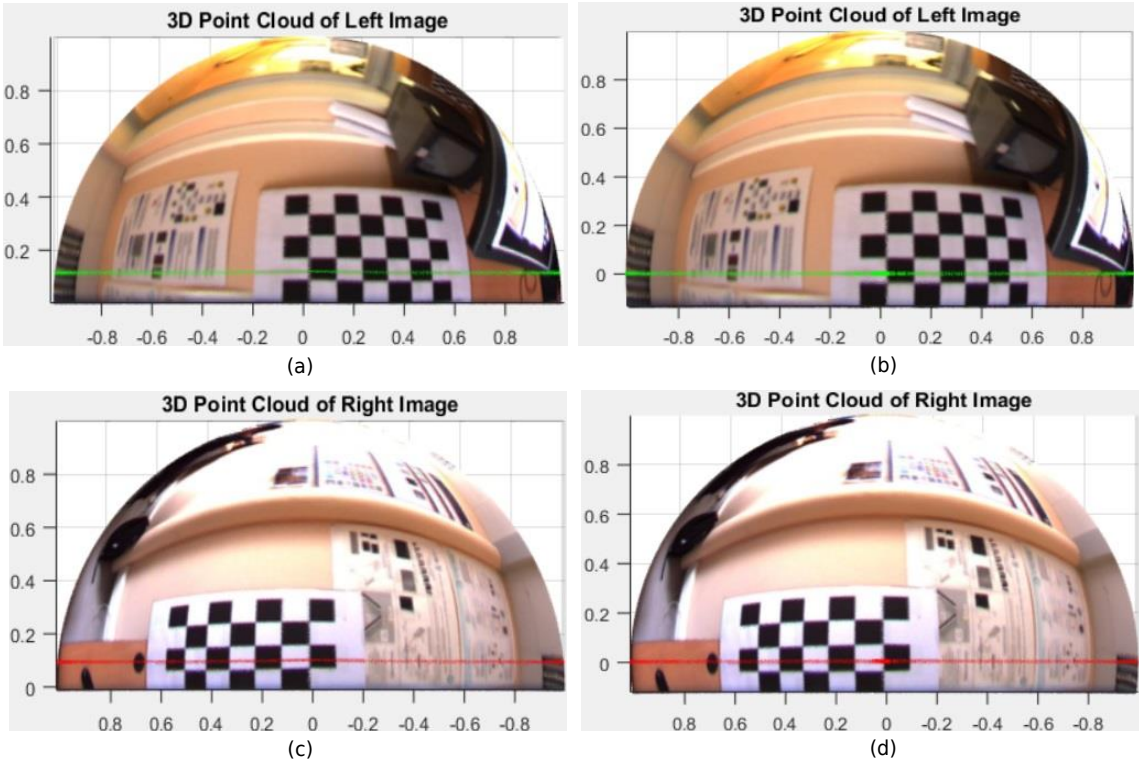


Figure 4.28: The left image (a) and the right image (c) were projected with the initial estimate of ξ. The 180° lines should ideally fall on the zero planes (see Figure 4.27). After the iterative estimation of ξ for both cameras, the 180° line fall on the zero planes as portrayed in the image (c) and image (d).

confirmed based on the point at 180° from the camera that should fall on the zero-degree line. It was decided after performing a detailed observation during the enfolding to the

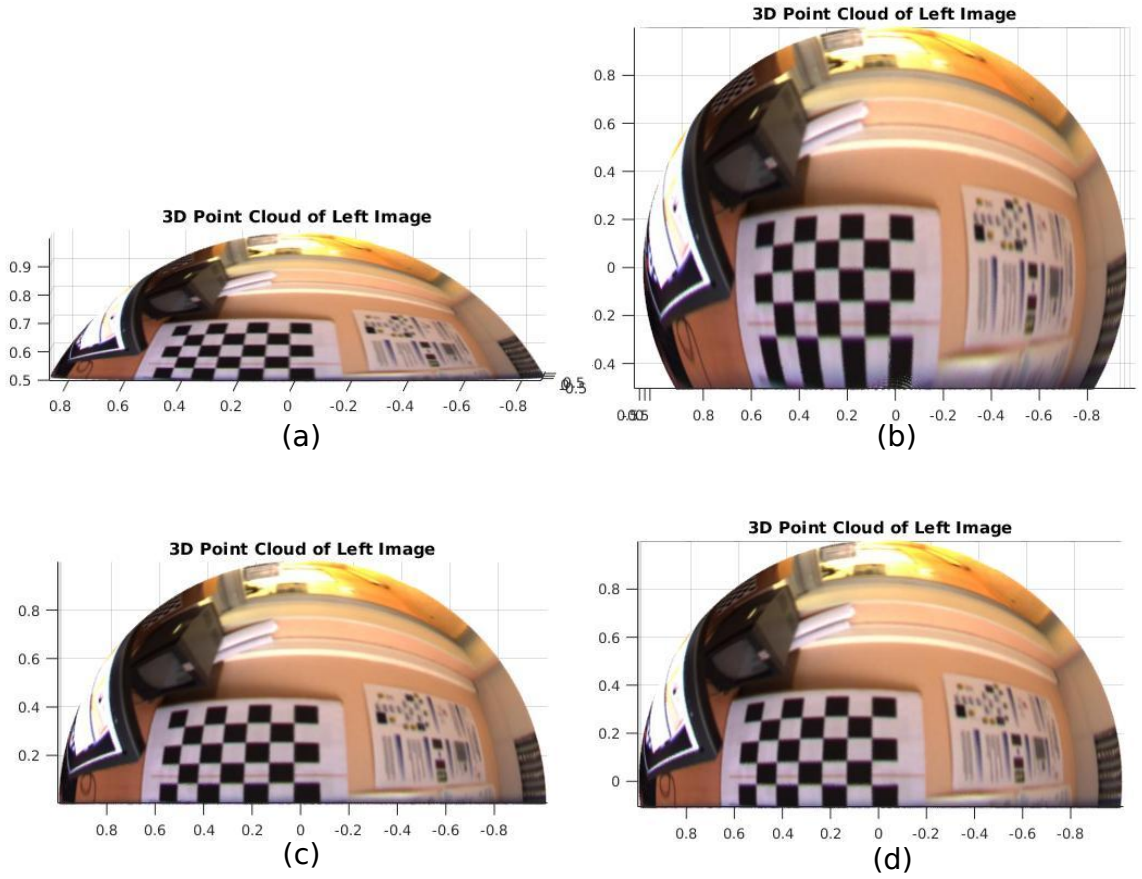


Figure 4.29: The fisheye image after projecting onto unit sphere.

unit sphere and the fusion between the left and right fisheye cameras (see Figure 4.28).

The projection errors should be hindered but always detected refer to the compression of the object at the edge of the fisheye image and the stretching of the image at the edge of the fisheye image. These errors generate double images at the mutual object in the overlapping area after the fusion process.

The initial value of eccentricity estimated using Mei's Calibration Toolbox. As discussed previously, the eccentricity value defined the level of distortion. The value of eccentricity influences the formation of the hemisphere. This statement refers to Figure 4.29. Image (a) is the formation of hemisphere using eccentricity $\xi = \mathbf{0.7}$. As observed, part of the image was compressed along the periphery of the image. To look at the effect of coefficient eccentricity on forming a hemisphere, its values changed to $\xi = \mathbf{1.75}$. As a result, a stretched image was observed along the periphery of the image (b). The initial value of eccentricity $\xi = \mathbf{1.4392}$ was retrieved from the calibration and produced image

(c). However, the 180° line did not fall correctly on the unit sphere's zero planes. The optimum value of eccentricity ξ was re-estimated from the initial value of ξ using the cost function by minimising the z-component. After the iteration, the new value of ξ was estimated as **1.5632**. It produced image (d), which resulted in virtually correct projection of a fisheye image onto the unit sphere and the 180° line lies correctly on the zero lines of a unit sphere.

4.5/ FUSION OF PERSPECTIVE CAMERA ONTO UNIT SPHERE

The Omni-Vision camera rig is a combination of three cameras to produce 360° vertical and horizontal images with one window for stereo viewing. All the cameras are referred to a single reference camera to combine them. In the set-up of Omni-Vision camera system, the left fisheye camera is assigned as the reference camera. It means; all camera coordinates should be referred to as the left fisheye camera.

The next challenge was to incorporate the perspective camera (ZED) over the unit sphere. Figure 4.30 illustrates the image seen from each of the cameras. The estimation of rotation between the ZED camera and fisheye cameras was similar that the estimation of rotation carried out between right and left fisheye cameras. Based on Figure 4.4 (camera position on Omni-vision) and Figure 4.10 (desired image composition), the images captured from the ZED camera were shared between left and right fisheye cameras. The matching point was manually selected in a 2D image to estimate the ZED camera's rigid transformation towards the left fisheye camera. The fusion of fisheye cameras with ZED camera using a unified model representation yielded desired outcomes based on the proposed multi-camera setup.

Let $\mathbf{x^z}$ and $\mathbf{x^{Lf}}$ be the image features that correspondence between ZED camera and left fisheye cameras, respectively. Let $\mathbf{P^Z}$ and $\mathbf{P^{Lf}}$ be the features that correspondence (mapped from $\mathbf{x^z}$ and $\mathbf{x^{Lf}}$) to unified sphere. The fusion of ZED camera with fisheye cameras can be framed as a minimisation problem from the features that correspondence to the unified sphere, which is defined as:

$$\underset{\theta_{x,y,z}}{\operatorname{argmin}} \sum_{i=1}^{n} \Psi \left(\left\| \mathbf{P}_S^{Lf} - \mathbf{P}_S^{Z}(\theta_{x,y,z}) \right\|_2 \right), \tag{4.33}$$

where $\Psi(\cdot)$ is the Loss function for the purpose of robust estimation, while

$$\mathbf{P}(\theta_{x,y,z}) = \mathbf{R}(\theta_{x,y,z}) \begin{bmatrix} x_s & . & . & . & . & x_s^n \\ y_s & . & . & . & . & y_s^n \\ z_s & . & . & . & . & z_s^n \end{bmatrix}, \tag{4.34}$$

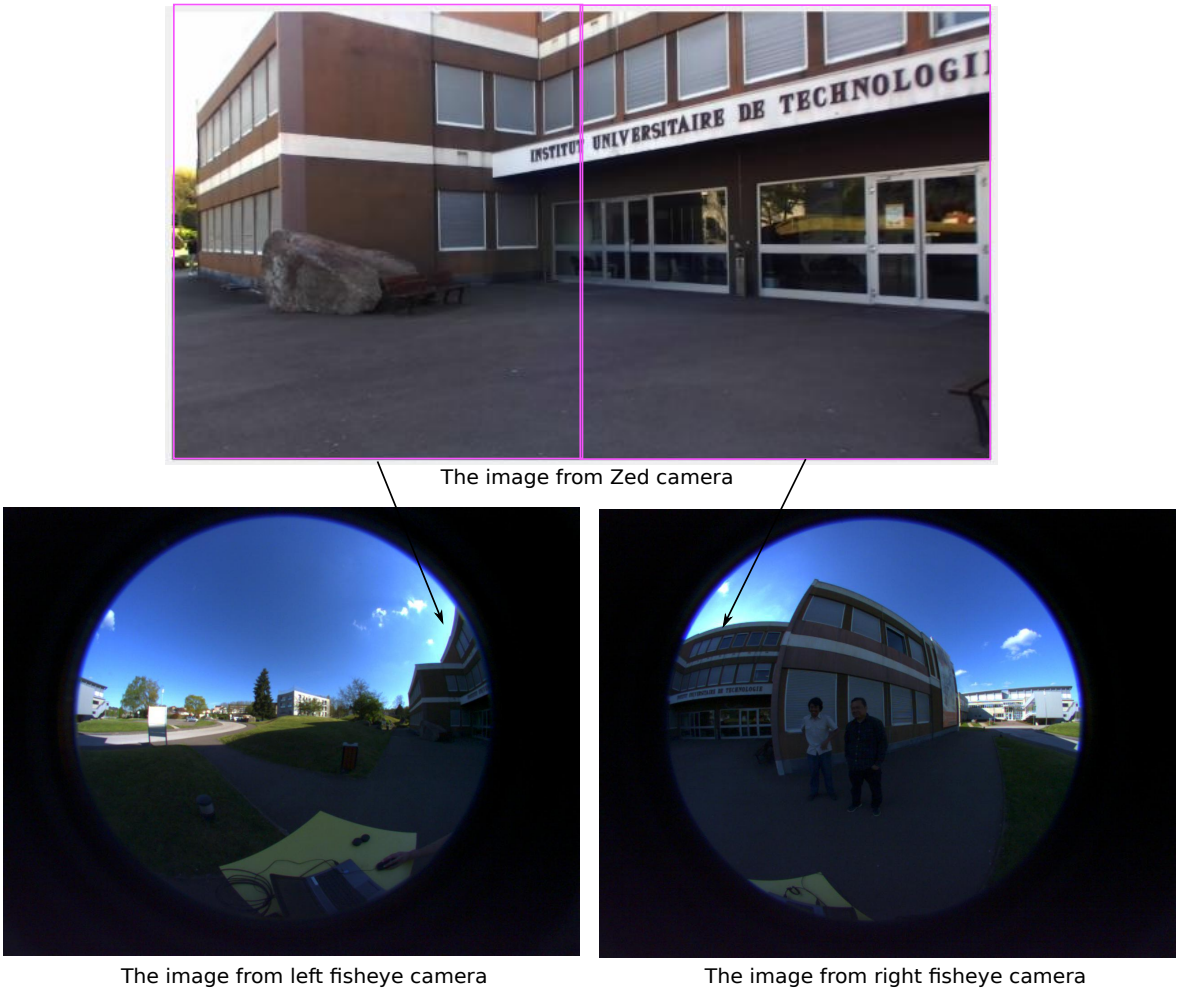


Figure 4.30: The images from the three cameras.

The equation 4.34 stands for the ZED camera sphere registration points to the left fisheye camera (the reference). The $\mathbf{R}(\theta_{x,y,z})$ is the desired pure rotation matrix with estimated rotation angles $\theta_{x,y,z}$. In order to solve this problem, the IPOA was applied; similar to solving Equation 4.31.

4.5.1/ EXPERIMENT AND RESULT

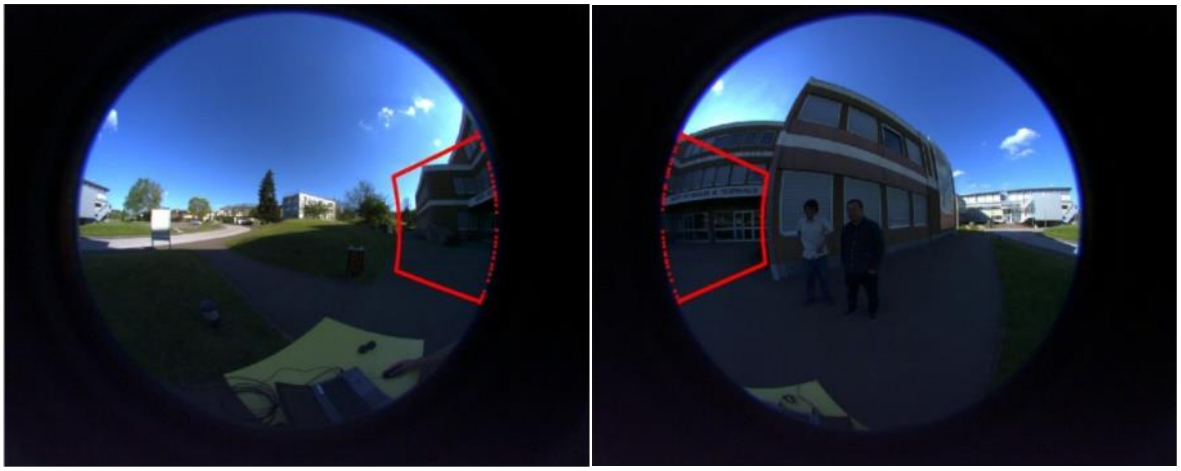
The rigid transformation between the ZED camera and the left fisheye camera was estimated as given in the following:

$$\mathbf{T} = \begin{bmatrix} -0.0143 & -0.0290 & -0.9995 & 0.0000 \\ -0.0062 & 0.9996 & -0.0289 & 0.0000 \\ 0.9999 & 0.0058 & -0.0145 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix}, \tag{4.35}$$

The outcomes from the transformation matrix signified that the matrix coincided with the camera's position against the reference camera. The matrix transformation estimate was accurate and verified the ZED camera's position in the Omni-Vision camera rig.

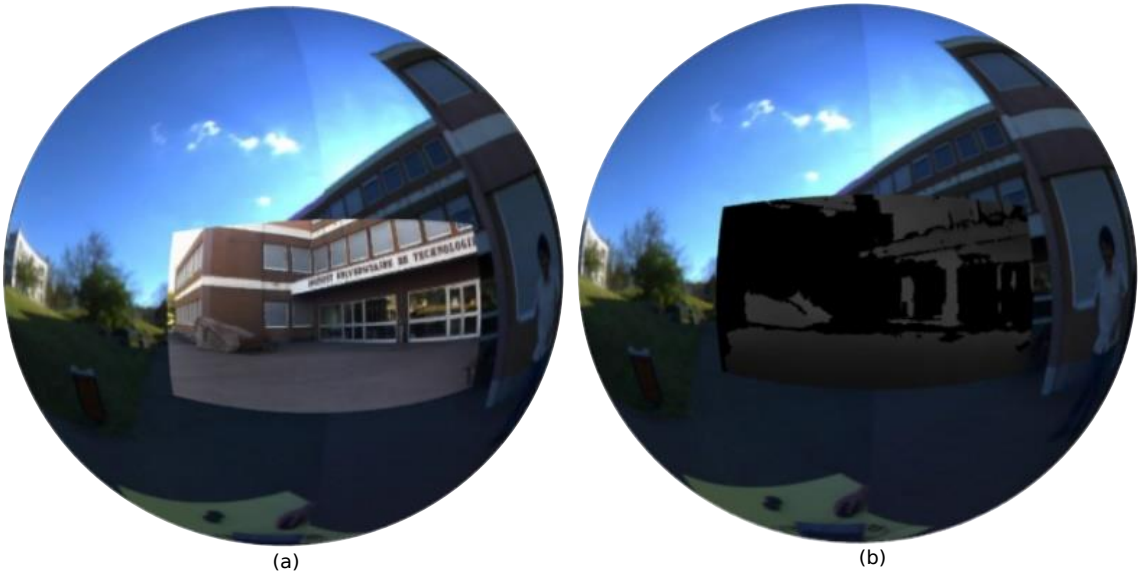
The next step is to estimate the area covered by the ZED image over the unit sphere. It was performed by projecting the perimeter border of the ZED image over the spherical unit. Since the ZED image shared between left and right fisheye cameras, the area covered should be identical. The pixels or overlapping areas between the unit sphere and ZED images were identified using the k-Nearest Neighbour (k-NN) algorithm. The overlapping area on the unit sphere projected from the fisheye camera. It was then projected back into the image plane using Mei's projection model (see Figure 4.8).

The initial result retrieved from the three cameras' fusion is illustrated in Figure



(a) (b)

Figure 4.31: The boundary perimeter of the ZED camera projected onto the unit sphere.



(a) (b)

Figure 4.32: The ZED image fused onto the unit sphere. Image (a) is the high-resolution ZED image onto the unit sphere, and image (b) is the ZED depth image onto the unit sphere.

4.32. The algorithm displayed the successful yield that fused the three images from a different camera and different geometry. The outcome was exceptional. It is the early step to fuse the N number of the image(s) from the N number of the camera(s) from difference geometries.

The presence of high distortion in the fisheye image had made a combination with ZED image a difficult task (distortion close to zero). Figure 4.32 shows the initial result of the proposed multi-image fusion. The result was only an estimate, and yielding a more accurate result meant addressing the distortion of the fisheye cameras. The appropriate distortion coefficient was incorporated into the ZED image to generate a robust combination. Based on the result analysis, an error was found between objects on the bounded ZED image and fisheye images. Figure 4.33 displays the error projection due to improper handling of distortion. Hence, the next task was to develop a new algorithm to yield optimal outcomes.

4.5.2/ Optimization of fusion ZED image onto the unit sphere

The optimization process is vital to achieving an excellent final result. For this context, a non-linear optimization solution is preferred in this application. The Interior Point

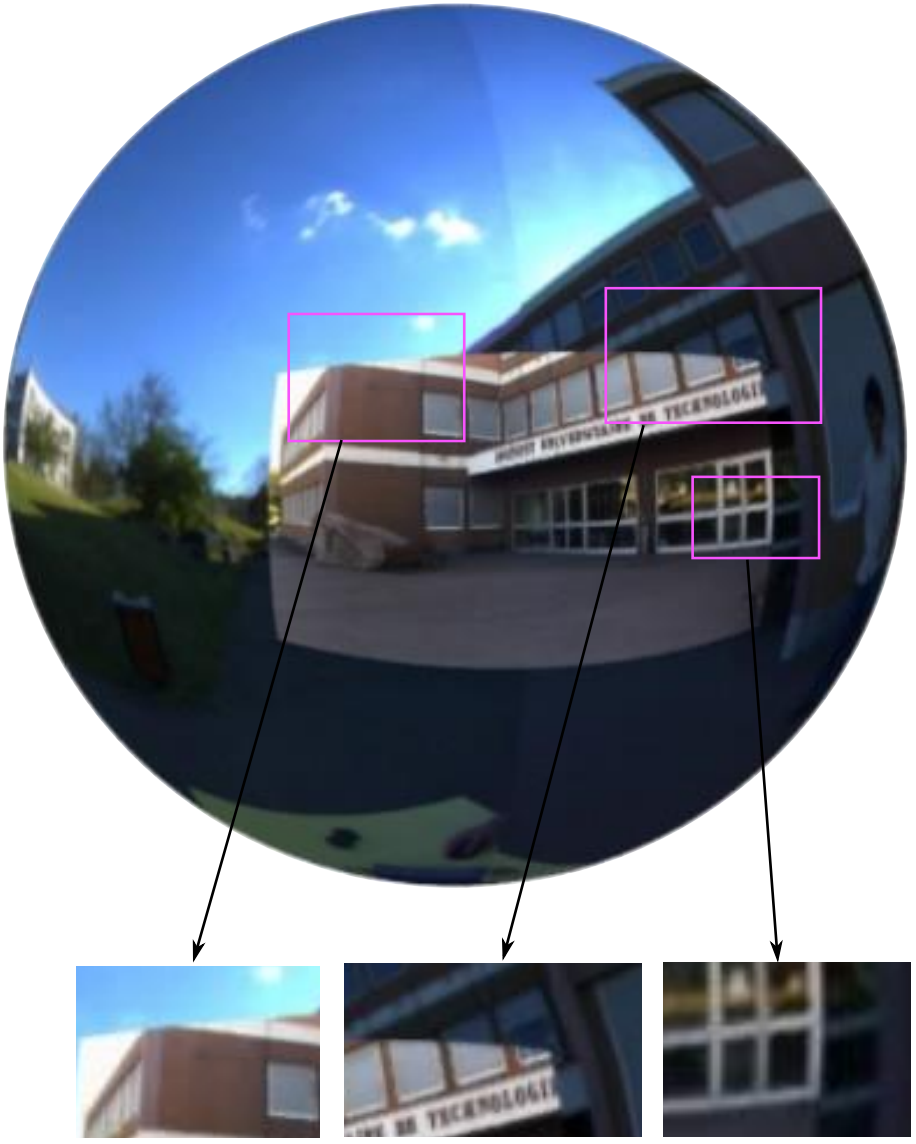


Figure 4.33: The fusion error noted between the boundary of the ZED image and the fisheye image.

Optimization (IPO) is a non-linear optimization algorithm used to solve a non-linear optimization problem. The parameters IPO estimated on the development of the unit sphere of the fisheye image. It is also used to remap (with distortion) the ZED features (perspective image) onto the unit sphere. The IPO parameters consist of thirteen elements.

- Paramerters 1 to 9 are corresponds to the parameters H of Affine Projective Distortion.
- Parameter 10 is correspond to the coefficient of distortion (δ).
- Parameters 11 to 13 are corresponds to rotation angles in X, Y and Z axis.

The transformation matrix of ZED parametrized by parameter number 11 to 13 with zero translation. The perspective image from the ZED camera is converted to camera frame using IPO parameter and K matrix from calibration.

Figure 4.10 presents the images captured from all the cameras with respect to their assigned position. A new method is proposed in this study to fuse the perspective image onto the unit sphere. The massive difference between perspective and spherical images refers to distortion that could deform the object in the scene [203]. The direct matching point from the perspective was deficient in matching the features on the unit sphere. It was due to the characteristics of the unit sphere that consisted of several levels of distortion. The affine projective distortion parameters were added to the perspective image plane before projected to the unit sphere's projection. In this case, the distortion proposed by Mei was neglected to perform mapping onto the unit sphere.

$$\begin{aligned} \mathbf{P} &= \mathbf{K}^{-1} \cdot \mathbf{H} \cdot \mathbf{I}, \\ \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{1} \end{bmatrix} &= \mathbf{K}^{-1} \cdot \mathbf{H} \cdot \mathbf{I}, \end{aligned} \tag{4.36}$$

where, $\mathbf{H}$ is affine projective distortion parameters. $\mathbf{K}$ is a camera matrix. $\mathbf{I}$ is an image frame.

$\mathbf{H}$ = Affine projective distortion parameters.

$\mathbf{K}$ = Camera matrix.

$\mathbf{I}$ = Image frame.

Replace all the matrices,

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{f_x} & \delta & \upsilon_o \\ \mathbf{0} & \mathbf{f_y} & \nu_o \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{h_{11}} & \mathbf{h_{12}} & \mathbf{h_{13}} \\ \mathbf{h_{21}} & \mathbf{h_{22}} & \mathbf{h_{23}} \\ \mathbf{h_{31}} & \mathbf{h_{32}} & \mathbf{h_{33}} \end{bmatrix} \cdot \begin{bmatrix} \upsilon \\ \nu \\ \mathbf{1} \end{bmatrix}, \tag{4.37}$$

By extracting Equation 4.37,

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{f_x h_{11}} + \delta \mathbf{h_{21}} + \upsilon_o \mathbf{h_{31}} & \mathbf{f_x h_{12}} + \delta \mathbf{h_{22}} + \upsilon_o \mathbf{h_{32}} & \mathbf{f_x h_{13}} + \delta \mathbf{h_{23}} + \upsilon_o \mathbf{h_{33}} \\ \mathbf{f_y h_{21}} + \nu_o \mathbf{h_{31}} & \mathbf{f_y h_{22}} + \nu_o \mathbf{h_{32}} & \mathbf{f_y h_{23}} + \nu_o \mathbf{h_{33}} \\ \mathbf{h_{31}} & \mathbf{h_{32}} & \mathbf{h_{33}} \end{bmatrix} \cdot \begin{bmatrix} \upsilon \\ \nu \\ \mathbf{1} \end{bmatrix}, \tag{4.38}$$

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{1} \end{bmatrix} = \begin{bmatrix} \upsilon(\mathbf{f_x h_{11}} + \delta \mathbf{h_{21}} + \upsilon_o \mathbf{h_{31}}) + \nu(\mathbf{f_x h_{12}} + \delta \mathbf{h_{22}} + \upsilon_o \mathbf{h_{32}}) + \mathbf{f_x h_{13}} + \delta \mathbf{h_{23}} + \upsilon_o \mathbf{h_{33}} \\ \upsilon(\mathbf{f_y h_{21}} + \nu_o \mathbf{h_{31}}) + \nu(\mathbf{f_y h_{22}} + \nu_o \mathbf{h_{32}}) + \mathbf{f_y h_{23}} + \nu_o \mathbf{h_{33}} \\ \upsilon \mathbf{h_{31}} + \nu \mathbf{h_{32}} + \mathbf{h_{33}} \end{bmatrix}, \tag{4.39}$$

the following was observed:

$$f(1) = \upsilon(\mathbf{f_x h_{11}} + \delta \mathbf{h_{21}} + \upsilon_o \mathbf{h_{31}}) + \nu(\mathbf{f_x h_{12}} + \delta \mathbf{h_{22}} + \upsilon_o \mathbf{h_{32}}) + \mathbf{f_x h_{13}} + \delta \mathbf{h_{23}} + \upsilon_o \mathbf{h_{33}}, \quad (4.40)$$

$$f(2) = \upsilon(\mathbf{f_y h_{21}} + \nu_o \mathbf{h_{31}}) + \nu(\mathbf{f_y h_{22}} + \nu_o \mathbf{h_{32}}) + \mathbf{f_y h_{23}} + \nu_o \mathbf{h_{33}}, \quad (4.41)$$

$$f(3) = \upsilon \mathbf{h_{31}} + \nu \mathbf{h_{32}} + \mathbf{h_{33}}, \quad (4.42)$$

solving Equation 4.39 and normalising it with the 3rd row,

$$f(K, H, \delta, I) = \begin{bmatrix} \frac{f(1)}{f(3)} \\ \frac{f(2)}{f(3)} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{1} \end{bmatrix}, \quad (4.43)$$

The values of $\mathbf{x}$, $\mathbf{y}$ and ξ ($\xi = \mathbf{0}$ for the perspective camera) were replaced into inverse mapping function (Equation 4.27) for projecting onto the unit sphere.

Figure 4.34 illustrates that the ZED camera image overlapped onto the unit sphere. It also recovered the scale between ZED and fisheye cameras using parameter δ. Upon comparing the previous result with the result after fusion. The result was enhanced after adding the affine projective distortion to the ZED image. The result displayed in Figure 4.35(b) shows that the images from ZED and fisheye cameras were aligned, especially after addressing the distortion.

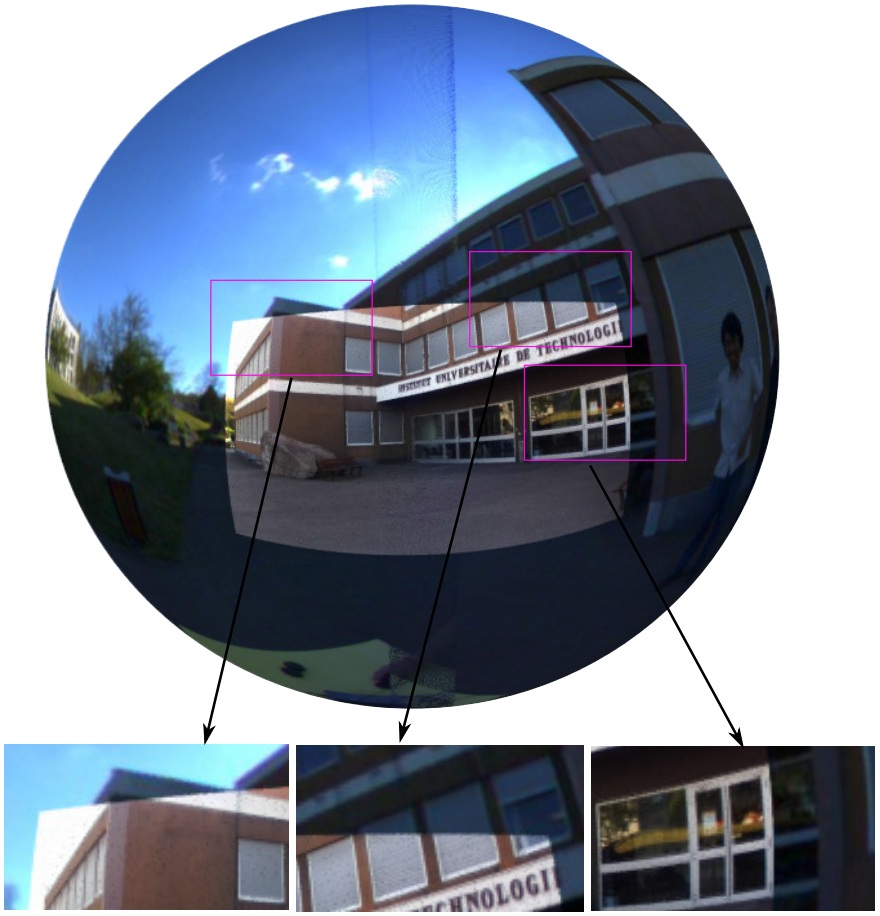


Figure 4.34: The perimeter boundary of the ZED camera was aligned with the fisheye image.

Figure 4.35: The ZED image was successfully aligned onto the unit sphere (left and right fisheye image). The borders of ZED and fisheye images were zoomed to view the details of the fusion. Photo (a) is the fusion using the previous method, and photo (b) results from the proposed method.

4.6/ CONCLUSION

Several vital points that function as the core of this thesis are highlighted in this chapter. Calibration is a fundamental process that must be performed before using a camera. The flow of the calibration process was similar for both the perspective and omnidirectional cameras. The only significant variance was the value of the radial distortion parameters that had been considered to obtain optimal and precise results. Intrinsic calibration of the perspective camera was performed using Bouget's toolbox, while Mei's toolbox was employed for the omnidirectional cameras.

The distortion of a fisheye camera was determined by using parameter eccentricity, ξ. The value of ξ was re-estimated and optimised using a cost function to minimise the z-component of pixels on the selected line via IPOA. The left and right fisheye cameras were fused using a rigid transformation matrix comprised of a rotation matrix and translation vector. The rigid 3D transformation matrix was estimated using the overlapping features by assuming that it was pure rotation. The IPOA was applied to estimate the rotation between the sets of projected points. The registration error was calculated to evaluate the performance of the proposed method.

The same procedure was employed to fuse the ZED camera onto the unit sphere. The initial result of the ZED projection used the coefficient distortion from Mei's projection model. The fusion result was optimised by adding the affine projective distortion to the ZED image before the unit sphere's projection. The final result is the optimum fusion image from three cameras. The image's position is referred to as the position of the camera on the Omni-Vision camera system. The combination of three cameras enabled the Omni-Vision to capture a complete 360° image, with a stereovision window for detail information of the scene.

5

System implementation - 3D Reconstruction

5.1/ Introduction

Along with the advancement of technology in the field of Information Technology and computers, 3D technology has vastly enhanced. For instance, computer graphics technology has undergone rapid progress in its display of 3D models. In the past, 3D identification and reconstruction processes were widely developed and installed for robotic applications [204] [205].

To date, 3D object display is diversified for use in other fields, including animation and graphics applications [206], architecture [207], education and cultural identification [208], as well as virtual reality [209] [210]. An example of preserving culture is referred to the remodelling of historic buildings using 3D techniques. The 3D reconstruction is a simple technique used to document the reconstruction and restoration of historic buildings in case of demolition. This method extracts vital information about the historical materials that can be preserved by computer technology. Besides, multiple other applications require accuracy in terms of geometry, details, and high display quality. The cutting-edge computer technology makes visualising 3D more manageable. Compared to the past, the cost of producing 3D was costly due to the use of an expensive and hi-end computers then. To date, 3D visualisation is performed rapidly at a reasonable cost. In fact, interactive 3D has been implemented for game consoles and continuous scientific observations, such as the movement of Earth [211], as well as to extract geothermal [212] and medical data [213] [214].

Studies in the field of 3D reconstruction and linear modelling of 2D images are often categorised into the scope of 3D vision in the field of digital image processing or computer vision. The 3D vision discussed on how humans perceive 3D objects within the sensory system. The two categories of 3D reconstruction techniques are passive and active techniques.

AN ACTIVE TECHNIQUE: The active technique or object scanning demands control of structured light. Some researchers used a camera and a projector or a viewer to produce structured light [176]. Another way to do so incorporates a laser beam and a video camera.

A PASIVE TECHNIQUE: The passive technique is performed by using two or more images of an object in various positions with the camera [215] [216]. This technique is known as the adoption of photogrammetry or structure from motion.

The object scanning technique, at times, requires expensive equipment and specialised skills for operation. As a cost-effective and straightforward approach, reconstructed objects based on images are popular amongst researchers [217]. Initially, studies within this domain focused only on calibrated cameras. Later the advancement of 3D reconstruction enables integration with the non-calibrated camera. The camera parameters can be determined from varied imageries in one order. All parameters required for the reconstruction process may be extracted from the image pair.

As discussed in the previous chapter, Omni-Vision hybrid cameras use three cameras to generate vertical and horizontal 360° panoramic images. The algorithm described in the previous chapter had successfully incorporated images from various types of cameras. Therefore, this camera's capabilities to take pictures are described in this chapter, besides analysing the view by achieving 3D reconstruction. The use of these cameras is related to observation (safety) and robotics application. The use of Omni-Vision as a camera sensor offers various advantages.

Since the proposed camera has 360° viewing space, it can speed up the time to act. For example, suppose a robot is assigned to detect an object that is placed behind the robot. In that case, the Omni-Vision camera equips the robot with sufficient information - the object's coordinates in this instance. This Omni-Vision camera saved the energy consumption as the robot not need move in search of the object. The following section describes the implementation of an Omni-Vision camera rig for 3D reconstruction.

5.2/ EPIPOLAR GEOMETRY OMNIDIRECTIONAL CAMERA

In 3D reconstruction, epipolar geometry is a technique that identifies the match between two images. Epipolar geometry for the perspective camera is briefly explained in Chapter 2 (see Section 2.9.1). The issue of epipolar camera over images with high distortion, which reflects the image on the sphere, was assessed in this present study [56]. The epipolar geometry for an omnidirectional camera has been vastly studied. It is originally used for the catadioptric camera as a model. Nonetheless, this present study has extended it to a dioptric or fisheye camera, which is the type of omnidirectional camera employed in the proposed Omni-Vision system. The knowledge of epipolar geometry for a perspective camera is applied to omnidirectional cameras. Figure 2.41, illustrates the epipolar geometry for perspective camera. The epipolar geometry served as a reference to describe the epipolar geometry for a calibrated omnidirectional camera (internal parameters are known) in detail.

As the projection model referred to projection to the unit sphere, epipolar geometry and all projection geometries adhere to the same projection model. The essential matrix that connects image one to image two does not change in terms of structure, which has a rotation matrix of $\mathbf{3 \times 3}$ size and vector translation. The epipolar line, as displayed in Figure 2.41, is likely the intersection point where $\mathbf{x'}$ should fall. For omnidirectional cameras, this line is represented by epipolar circles $\mathbf{C_1}$ and $\mathbf{C_2}$. Point $\mathbf{P}$ projected on both spheres is the intersection point between the epipolar plane and two spheres (see Figure 5.1).

Figure 5.2, shows the epipolar constraint and the two positions of fisheye camera are considered as point $\boldsymbol{P}$ in the space. The coordinate $(\boldsymbol{v_1}, \boldsymbol{v_1})$ and $(\boldsymbol{v_2}, \boldsymbol{v_2})$ are the corresponding image coordinates on the two fisheye images. Point $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$ are the projection point $\mathbf{P}$ onto the unit sphere image at the two varied fisheye positions [111]. The projection to unit sphere used the inverse mapping function in Equation 4.27. After the projection, it was found that points $\mathbf{P}$, $\mathbf{p_1}$, $\mathbf{p_2}$, $\mathbf{O_1}$ and $\mathbf{O_2}$ are coplanar, thus:

$$\begin{aligned} \overline{\mathbf{O_1O_2}} \times \overline{\mathbf{O_2P_1}} \cdot \overline{\mathbf{O_2P_2}} &= \mathbf{0}, \\ \mathbf{O_1^2} \times \mathbf{P_1^2} \cdot \mathbf{P_2} &= \mathbf{0}, \end{aligned} \tag{5.1}$$

where, $\mathbf{O_1^2}$ and $\mathbf{P_1^2}$ are the coordinates of $\mathbf{O_1}$, and $\mathbf{P_1}$ in coordinate system $\mathbf{X_2, Y_2, Z_2}$. The transformation between system $\mathbf{X_1, Y_1, Z_1}$ and $\mathbf{X_2, Y_2, Z_2}$ can be described by rotation $\mathbf{R}$

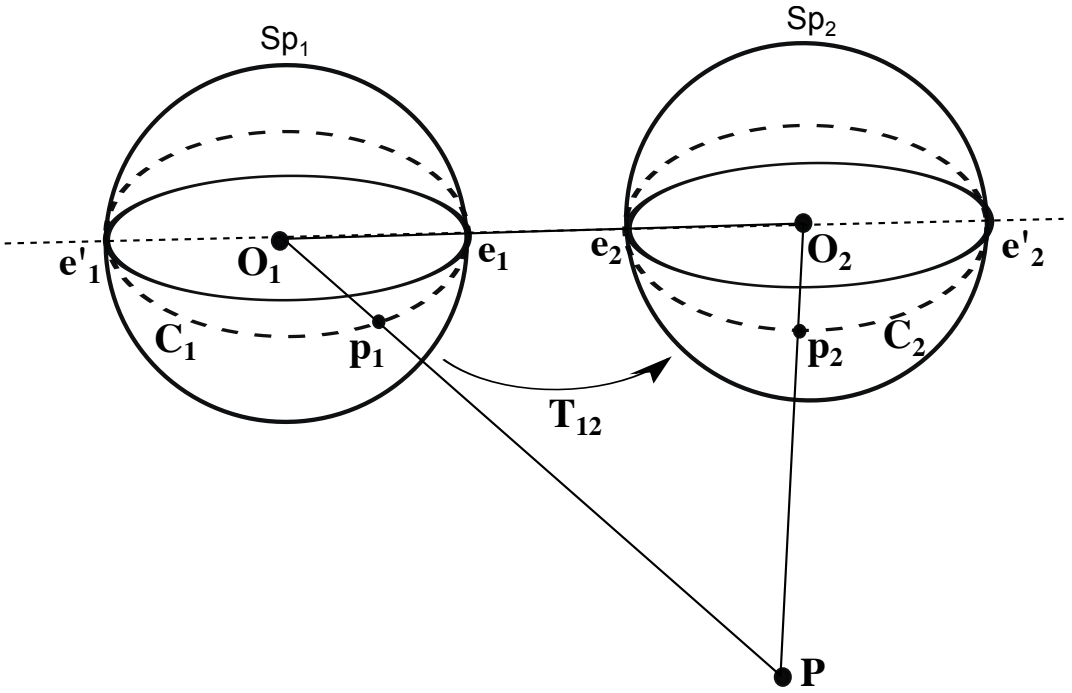


Figure 5.1: The epipolar geometry for omnidirectional camera.

and translation $\mathbf{t}$.

The transformation equations are as follows::

$$\begin{aligned} \mathbf{O_1^2} &= \mathbf{R \cdot O_1 + t = t,} \\ \mathbf{P_1^2} &= \mathbf{P_1 \cdot O_1 + t,} \end{aligned} \tag{5.2}$$

By substituting 5.2 in 5.1, the following is obtained,

$$\mathbf{P_2^T E P_1 = 0,} \tag{5.3}$$

where, $\mathbf{E = [t]_\times R}$ is the essential matrix that consists of rotation and translation.

Referring to Section 2.9.1.2, the method used to compute the essential matrix is also valid for omnidirectional images. Suppose an object is detected on both images/spheres. In that case, the relationship between the two images/spheres can be translated to the essential matrix, $\mathbf{E_{12}}$ (i.e., the essential matrix between image one and image two). In order to estimate the essential matrix, points that correspond to the pairs on the fisheye images are stacked into the linear system. Hence, the overall epipolar constraint is as follows:

$$\mathbf{Uf = 0,} \tag{5.4}$$

where: $\mathbf{U = [u_1, u_2, ..., u_n]^T}$

and $\mathbf{u_i}$ and $\mathbf{f}$ are vectors constructed by stacking column of matrices $\mathbf{P_i}$ and $\mathbf{E}$ respectively.

$$\mathbf{P_i = P_i P_i'^T,} \tag{5.5}$$

$$\mathbf{E} = \begin{bmatrix} \mathbf{f_1} & \mathbf{f_4} & \mathbf{f_7} \\ \mathbf{f_2} & \mathbf{f_5} & \mathbf{f_8} \\ \mathbf{f_3} & \mathbf{f_6} & \mathbf{f_9} \end{bmatrix}, \tag{5.6}$$

The essential matrix can be estimated with linear least square by solving Equations 5.4 and 5.5, where $\mathbf{P_i^{'}}$ is the projected point which corresponds to $\mathbf{P_2}$ of Figure 5.2, $\mathbf{U}$ is $\mathbf{n \times 9}$ matrix and $\mathbf{f}$ is $\mathbf{9 \times 1}$ vector with 9 elements of $\mathbf{E}$.

The initial estimate of the essential matrix was used for the robust estimation of the essential matrix. A modified iteratively re-weighted least square method for Omni-Vision camera is proposed, as initially explained in [75] [218]. This assigned minimal weights to outliers and noisy correspondences. Weight assignment was performed by identifying the residual $\mathbf{r_i}$ for each point.

$$\mathbf{r_i = f_1 x_i' x_i + f_4 x_i' y_i + f_7 x_i' z_i + f_2 x_i y_i' + f_5 y_i y' i + f_8 y_i' z_i + f_3 x_i z_i' + f_6 y_i z_i' + f_9 z_i z_i',} \tag{5.7}$$

$$\mathit{err} \rightarrow \min_{\mathbf{f}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \left(\mathbf{w_{Si} f^T u_i}\right)^2, \tag{5.8}$$

$$\mathbf{w_{Si}} = \frac{\mathbf{1}}{\mathbf{\nabla r_i}}, \tag{5.9}$$

where: $\mathbf{r_i} = (\mathbf{r_{xi}^2 + r_{yi}^2 + r_{zi}^2 + r_{xi'}^2 + r_{yi'}^2 + r_{zi'}^2})^{\frac{1}{2}}$,
and, w_{Si} is the weight (Sampson's weighting) g) assigned to each set of corresponding point and $\mathbf{\nabla r_i}$ is the gradient, while $\mathbf{r_{xi}}$ and so on are the partial derivatives derived from Equation 5.7 ($\mathbf{r_{xi} = f_1 x_i' + f_2 y_i' + f_3 z_i'}$).
After computing all the weights $\mathbf{U}$ matrix is updated as follows:

$$\mathbf{U = WU,} \tag{5.10}$$

where, $\mathbf{W}$ is diagonal matrix of the weights computed using Equation 5.9.

The essential matrix was estimated at each step and was forced to attain rank 2 in each iteration. The procrustean approach was adopted, and SVD was employed for this purpose. The proposed hybrid camera system is a combination of three cameras that use multi-view geometry without any modification. It is in line with the discussion on epipolar geometry for the omnidirectional camera. The conventional method has been discussed by researchers in the field of computer analysis and computer vision, such as image processing. This method's advantages have attracted the attention of many researchers for image processing from omnidirectional cameras for robotic navigation, pattern recognition, and surveillance.

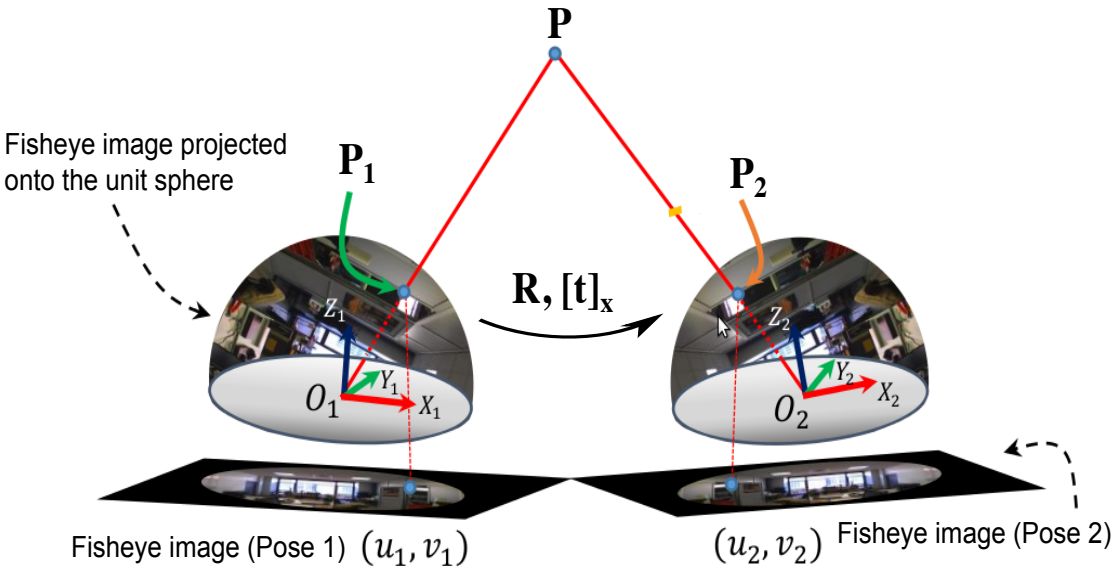


Figure 5.2: Details of epipolar geometry omnidirectional camera.

5.3/ FEATURES MATCHING

The image description is a vital task related to computer vision. The detection of a specific object in an image by matching the feature points refers to the process of finding a feature point match between feature points on a reference image and feature points on a target image. Feature points are general information or unique characteristics of an object in an image.

The main problem in the process of image matching and classification is identifying uniformity of information related to size, intensity, rotation, scale, noise, and viewpoint. Other problems refer to the scene's sensitivity to detect several similar objects (e.g., homogenous structure and texture), complicated lens distortion, and non-affine distortion on the scene. Thus, matching the features of the two images becomes a difficult task. As a result, plenty of outliers are generated. Nevertheless, the proposed Omni-Vision camera rig addresses all these issues.

Plenty of image descriptors have been proposed to identify the correct similar features on various images. The features extracted from an image are divided into two categories; global and local types. The global feature or context depicts the image as a whole to speculate the complete object target. For example, images with some similar significant subjects. The global type can be used on specific images, although it is con-

sidered less successful than the local type.

In contrast to the global descriptor is the local descriptor. The local descriptors describe the image patches or the key-point of an object in the image. The SIFT (Scale Invariant Features Transform) [219] and SURF (Speed Up Robust Features) [220] [221] are some successful local descriptors applied in many applications. In order to produce SURF with high performance, it may be combined with Maximally Stable Extremal Regions (MSER) [222]. This combination increases the speed and accuracy of the local features, where MSER is a region detector. Integrating local and global descriptors has generated exceptional outcomes [223] [224]. The approach is robust to local appearance ambiguity and enables non-rigid transformations without displaying convincing evaluations and results.

The features descriptor technique is still being updated with a new technique. The semantic context depicted in [225] refers to an augmented version of a local feature. It increases the discriminant power of the feature without affecting the variability of the image shape. Descriptor works on an image that contains multiple similar regions while depicting slight non-affine distortions.

5.4/ TRIANGULATION

The development of robot portrayed its significance in robotics and automation domains, in parallel with the advancing world of technology and intense global competition. As the demand for robots rises, the quality of robots is expected to be enhanced. Robots are required to be more innovative and have the ability to detect its surrounding accurately. More examples are:

- Measure the distance of an object.
- Identify an object and manipulate the object in a specific location with different positions and orientations accurately [226].

Therefore, computer vision developed a computer system to help robots view the world. With this development, robots can recognise objects, determine the distance of objects. Model objects in 3D by relying on information in the form of images of objects observed from two or more points of view [227]. The ability to measure the distance to an object

becomes one of the most essential and fundamental things for a robot in order to recognise the environment.

Many studies have proposed a method for robots to measure distances, such as ultrasonic sensors, infra-red sensors, and lasers. In this study, the triangulation technique was implemented to estimate the distance between the camera and the target object. It used a camera to recognise objects, as well as to measure the distance of objects and the depth of the scene [228] [229] [230].

The scene from two images can be reconstructed using parameters $\mathbf{P_1}$, $\mathbf{P_2}$, $\mathbf{R}$ and $\mathbf{t}$ as described in [68]. Referring to Section 5.2 and Section 5.3, as weel as Figure 5.1 and Figure 5.2, the coplanar consists of points $\mathbf{O_1}$, $\mathbf{P_1}$, $\mathbf{P}$, $\mathbf{P_2}$ and $\mathbf{0_2}$. The line passing through the centre of the first camera $\mathbf{O_1}$ and the point projected at unit sphere $\mathbf{P_1}$ is defined as:

$$\mathbf{aP_1}, \tag{5.11}$$

By assuming the first image as the reference image, the line passing through the centre of the second camera $\mathbf{0_2}$ and point projected at unit sphere $\boldsymbol{P_2}$ can also be defined as:

$$\mathbf{bRP_2 + t}, \tag{5.12}$$

where: $\mathbf{a}$, $\mathbf{b} \in \mathbb{R}$ (Real number)

$\mathbf{R}$ is the rotation matrix, and $\mathbf{t}$ is the translation vector between the two images. It assumed that the 3D point generated from each image did not overlap each other. Still, each point was located near each other in minimal residues. The triangulation method was employed in this context to minimise the 3D line from the first and second images. This query's solution should be solved as a least-square problem - a classic solver to minimise the distance between two points.

$$\min_{\mathbf{a,b}} \|\mathbf{aP_1 - bRP_2 - t}\|, \tag{5.13}$$

$$\begin{bmatrix} \mathbf{a}^* \\ \mathbf{b}^* \end{bmatrix} = \left(\mathbf{A^T A}\right)^{-1} \mathbf{A^T t}, \tag{5.14}$$

where;

$$\mathbf{A} = \begin{bmatrix} \mathbf{P_1}, & \mathbf{-RP_2} \end{bmatrix}, \tag{5.15}$$

Next, the 3D point of P was determined by calculating the middle point of the minimal distance between lines Equation 5.11 and 5.12. The equation is given as follow:

$$\mathbf{P_k} = \frac{\mathbf{a^*P_1 + b^*RP_2 + t}}{\mathbf{2}}, \tag{5.16}$$

where, $\mathbf{k = 1, 2, 3, 4...}$ are the number of 3D points in space.

The two images estimated the 3D point in space. The 3D point estimated by considering the line intersection between the line 3D from the first image, the line 3D from the second image, and the rigid transformation between the images. Nonetheless, the precision was only the estimation that relied on the quality and the similarity between the two images. Referring to Figure 2.42, four possible solutions can be applied to estimate the position of a 3D point. In this context, the solution for camera configuration is that the 3D points should be located in front of both cameras.

Theoretically, the correct solution should be the optimal solution that offers a positive depth of a scene. However, the solution of the method appeared to be challenging due to the different geometry properties. A more straightforward solution was undertaken by solving the uncertainty and the quality of the result obtained. Hence, the distance between the norm of the reconstructed point using each solution and the projected point on the unit sphere $\mathbf{P_1}$ on the first image was determined. After that, the four distances were compared, and the minimum distance was selected as the best solution.

The method iterated for all the points projected on the first image. The following figure displays the results retrieved based on proposed method. Synthetic data used to assess the performance of the prescribed method. The results obtained with the real image are discussed in the next section.

5.5/ 3D RECONSTRUCTION FROM N VIEWS.

The proposed Omni-Vision camera system has an advantage due to the panoramic image that it offered. The camera system provided the immense information 360° horizontally and vertically from the scene, hence ready to exploit the image for various applications, such as surveillance, mapping, and 3D reconstruction. Due to its complete view system, the Omni-Vision camera system generated a visual memory useful for robotic application. However, this notion is only valid for areas without occlusion.

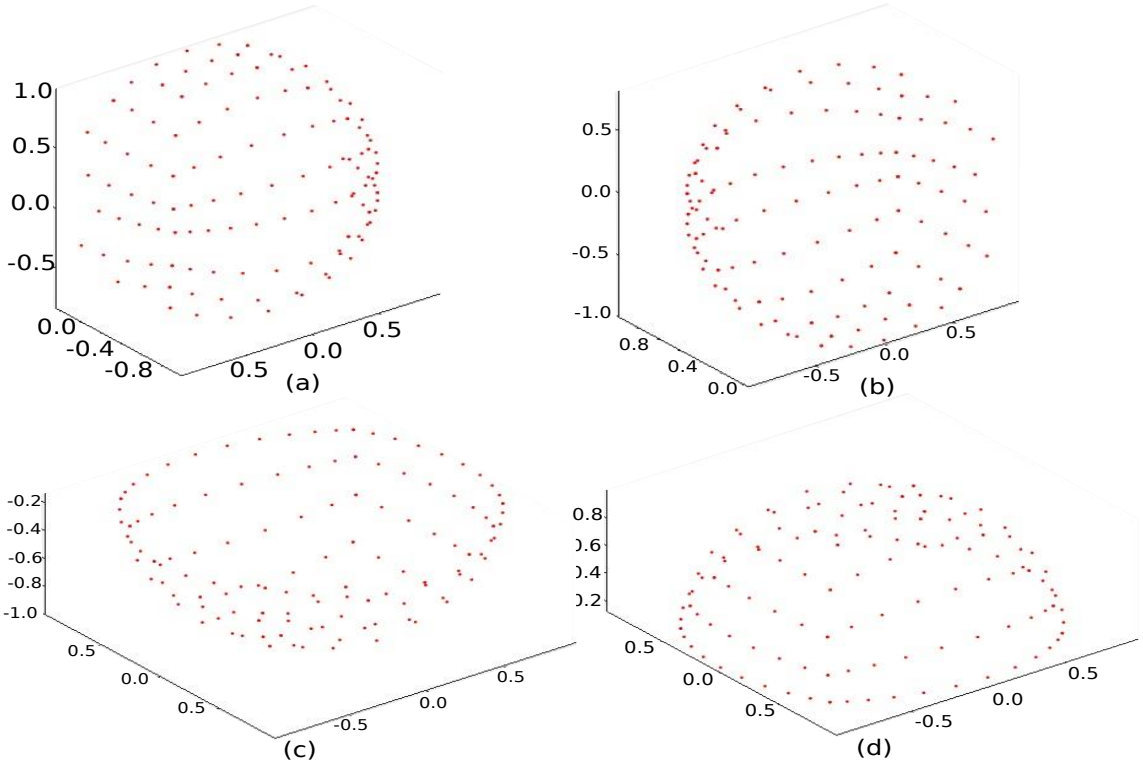


Figure 5.3: Synthetic data implemented in each of the four possible solutions. Each solution gave different results that affected the position of the 3D points. Only one solution gave realistic outcomes based on the orientation of cameras installed in the proposed camera rig, Omni-Vision.

The visual ability of the Omni-Vision camera is illustrated in Figure 5.4. At positions 1, 2, and 3, the right fisheye views the object. At positions 4 and 6, the object is viewed by the left fisheye. At position 5, both left and right fisheye cameras see the object because it falls in the overlapping area.

When compared to other hybrid cameras equipped with a perspective camera at the place of the fisheye camera, while retaining the same movement and orientation as Omni-Vision, the limited FoV would only allow the right camera to view the object, particularly at positions 1 and 2. The system would lose the image at other positions.

The Omni-Vision refers to a vision system that consists of an N-number of cameras. Its design and development are explained in detail in Chapter 3. The N-number of cameras produces an N-number of images with varied geometries. This study proposes an algorithm to fuse the N-number of cameras and the rich information extracted from the scene for multiple uses. Next, the camera system's performance was evaluated by measuring its ability and accuracy to perform some tasks, such as surveillance and 3D

reconstruction.

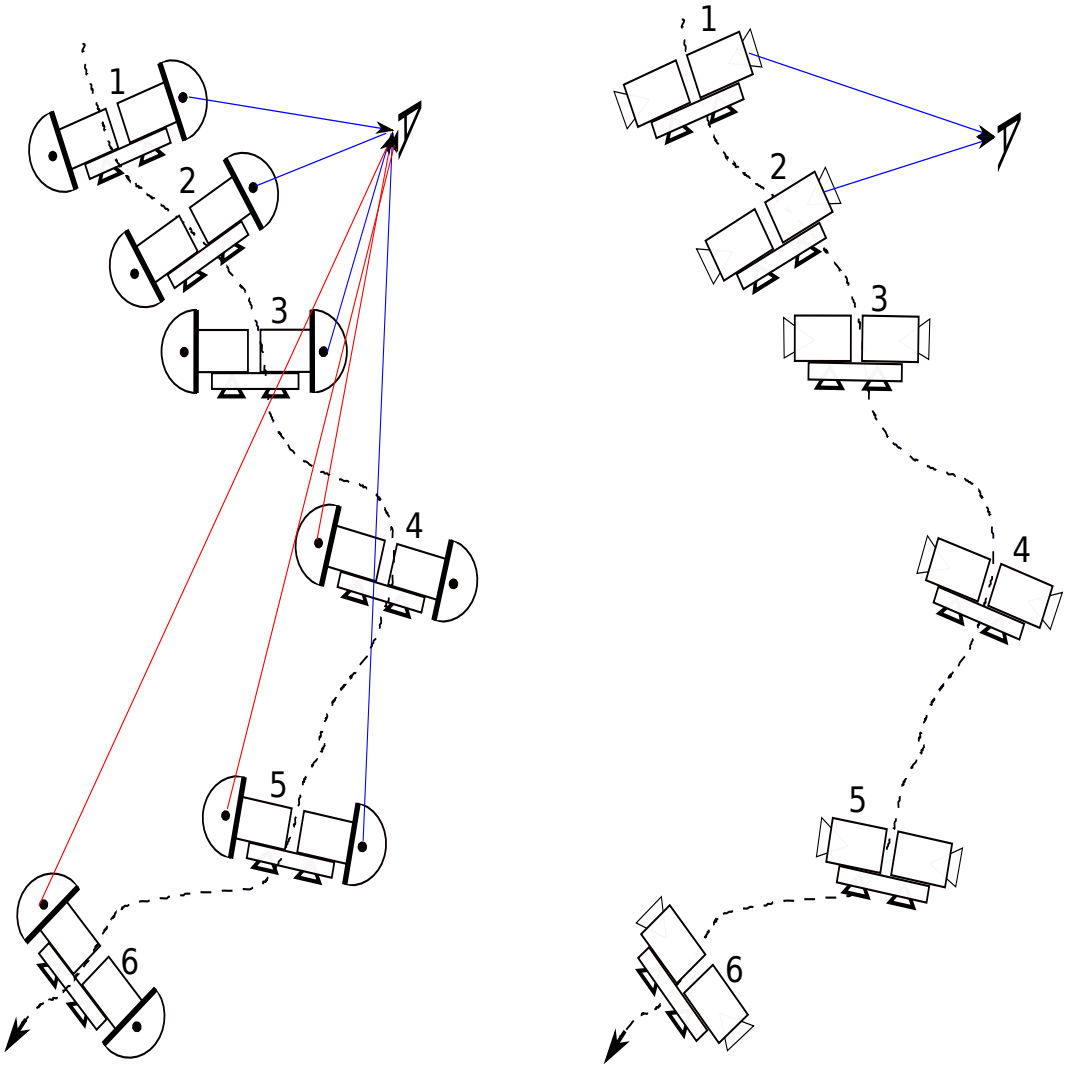


Figure 5.4: View of two camera systems. On the left is the proposed camera system, Omni-Vision. On the right is a hybrid camera that uses a perspective camera on both sides.

However, in the case of motion estimation or structure from motion, the algorithm may not function as expected due to a single effective viewpoint constraint. Kim et al., [231], applied the spherical approximation approach for a system that consisted of several cameras. They concluded that the approximation outperformed the generalised camera model.

5.5.1/ 3D RECONSTRUCTION FROM FISHEYE CAMERA

The Omni-Vision camera consists of the wealthiest information of the surrounding area. The surrounding area is captured by the two fisheye cameras with 180° FoV. The wide FoV enabled the full view of 3D reconstruction. The image captured using the fisheye camera had low resolution due to the presence of high distortion.

A high-resolution perspective image is usually employed to reconstruct a 3D scene [232] [233] [234] [235]. Image with high resolution captured from perspective camera enables zooming without losing the original quality of the image. However, the characteristics of a fisheye camera differ from those of a perspective camera. The use of a fisheye camera should be extended to a higher level for broader applications. This flow is in line with the study objective, which is to study the capability and the ability of the fisheye camera or the proposed Omni-Vision camera system.

The 3D reconstruction of a scene was assessed using the same camera but with projection from two different locations. Initially, proposed by Longuet-Higgins et al. [152], the conventional perspective camera was used to extract rigid transformation $(\mathbf{T})$ and to conclude the correlation between rotation $(\mathbf{R})$ and translation $(\mathbf{t})$. The pipeline of extraction is discussed in the previous chapter.

The 3D points constructed based on several steps. Figure 5.5, shows the process flow. The input of the block diagram is the image from n-view cameras. The feature and matching points were detected using the existing method described in Section 5.3.

The features point and matching use the existing method in Section 5.3. The method that can maximise the correct features with affordable accuracy was implemented. The estimation of the essential matrix had been necessary to determine the correspondences between the two images. The essential matrix was calculated using the omnidirectional camera's epipolar constraint, as explained in Section 5.2. It contains information about rotation and translation. Besides matching the combination of rotation and translation to the four possible solutions (see Figures 2.42 and Figure 5.3(d).

In order to develop the epipolar constraint of the omnidirectional camera, the solution with 3D synthetic data and different structure of the point cloud is proposed. For the epipolar constraint estimation, the 3D pyramid and wall cube pattern synthetic data (see Figure 5.6(a)) were applied. At the initial position (see Figure 5.6(b)), the pyramid pattern synthetic data were projected onto the unit sphere (red). Next, at another position, the

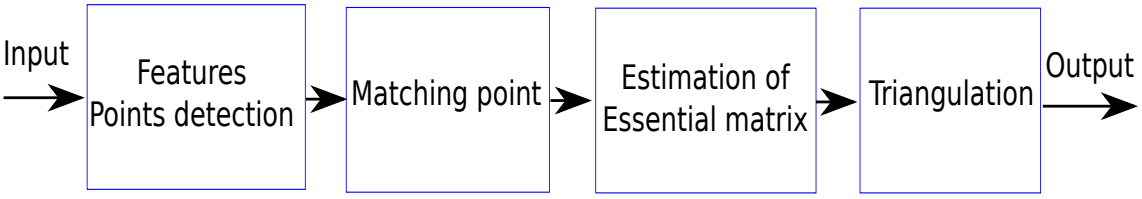


Figure 5.5: The block diagram of 3D reconstruction.

synthetic data were projected again onto the unit sphere (blue) with known rotation and translation relative to the initial position. After that, the essential matrix was calculated using the technique defined in Section 5.4. Equation 5.17 is the result of the rigid transformation matrix without the addition of noise. Finally, based on Figure 5.6(c), the 3D point was reconstructed using rotation matrix and translation vector.

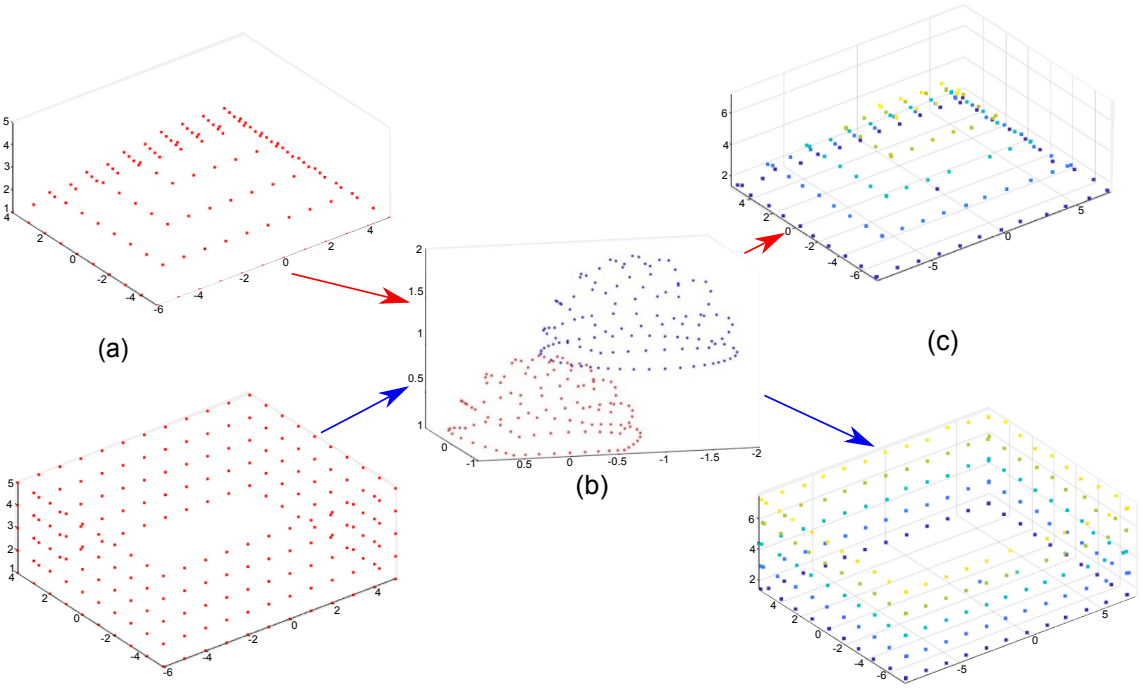


Figure 5.6: Synthetic data for 3D reconstruction. Two synthetic shapes of data (pyramid and wall cube) were used to test the method's accuracy.

$$\mathbf{T}_{\text{without noise}} = \begin{bmatrix} \mathbf{0.9254} & \mathbf{0.0180} & \mathbf{-0.3785} & \mathbf{0.0000} \\ \mathbf{0.1632} & \mathbf{0.8826} & \mathbf{0.4410} & \mathbf{-0.7071} \\ \mathbf{0.3420} & \mathbf{-0.4698} & \mathbf{0.8138} & \mathbf{0.7071} \\ \mathbf{0.0000} & \mathbf{0.0000} & \mathbf{0.0000} & \mathbf{1.0000} \end{bmatrix}, \tag{5.17}$$

Another scenario that can create is when relevant data are mixed with noisy or destructive data. Figure 5.7 displays the input pyramid pattern synthetic data mixed with synthetic noise. For this scenario, Gaussian noise was generated and incorporated into the pyramid pattern synthetic data. It assumed that the maximum distance between the pixels of a fisheye image after projecting onto the unit sphere is the standard deviation of the Gaussian noise.

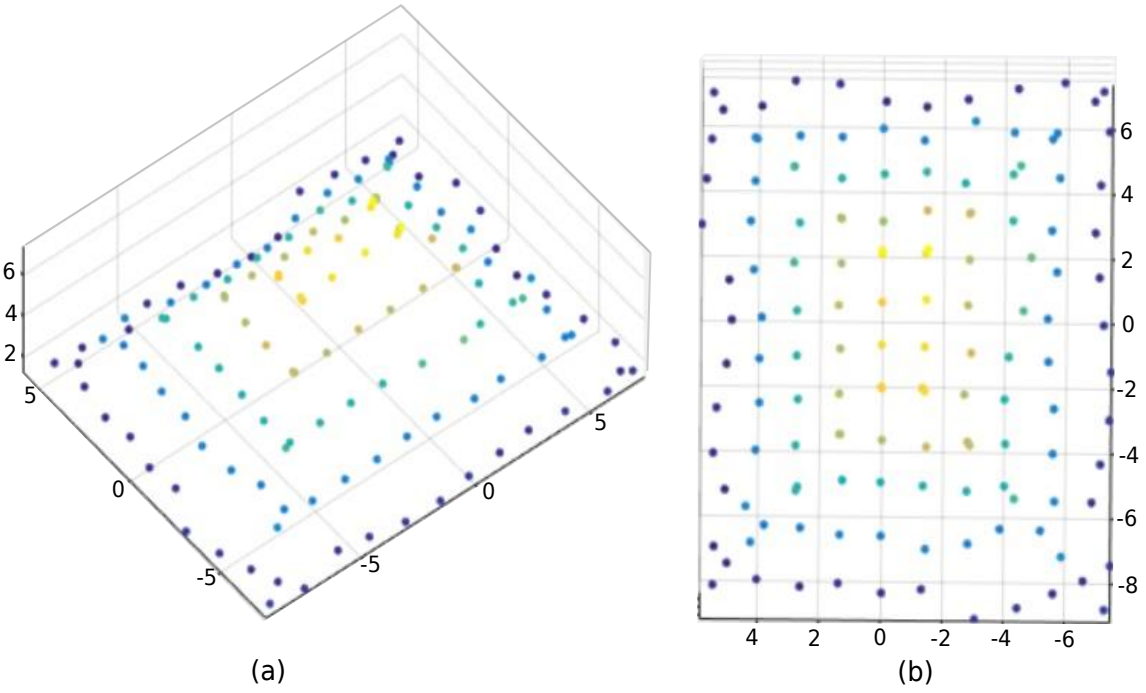


Figure 5.7: The pyramid shape synthetic data add with the Gaussian noise for 3D reconstruction.

$$\mathbf{T}_{\textbf{with noise}} = \begin{bmatrix} \mathbf{0.9255} & \mathbf{0.0091} & \mathbf{-0.3787} & \mathbf{-0.0042} \\ \mathbf{0.1683} & \mathbf{0.8857} & \mathbf{0.4327} & \mathbf{-0.7263} \\ \mathbf{0.3394} & \mathbf{-0.4642} & \mathbf{0.8181} & \mathbf{0.6873} \\ \mathbf{0.0000} & \mathbf{0.0000} & \mathbf{0.0000} & \mathbf{1.0000} \end{bmatrix}, \tag{5.18}$$

Equation 5.18 shows the result of rigid transformation using synthetic data with noise.

This test proved that the algorithm could be applied to the presence of unexpected noise. Based on both experiments (see equation 5.17 and 5.18). There is no significant difference between pure synthetic data and synthetic data with noise. The experiment revealed that Gaussian noise did not affect the performance of the algorithm.

5.6/ EXPERIMENTS AND RESULTS

5.6.1/ ERROR ANALYSIS

The quantification of the error made by creating a synthetic data pyramid and wall cube (see Figure 5.6). It compared the reconstructed points with the synthetic input. Gaussian noise added in the synthetic data with characterized as multiples of standard deviation, equal to the largest pixel to pixel distance over the spherical projection.

The error plot is shown in Figure 5.8. The average Euclidean distance of 50 attempts is plotted for each value of standard deviation (0.005) with 50 multiples of 0.05 for pyramid and wall cube reconstruction. This error pattern is also consistent with other synthetic 3D patterns.

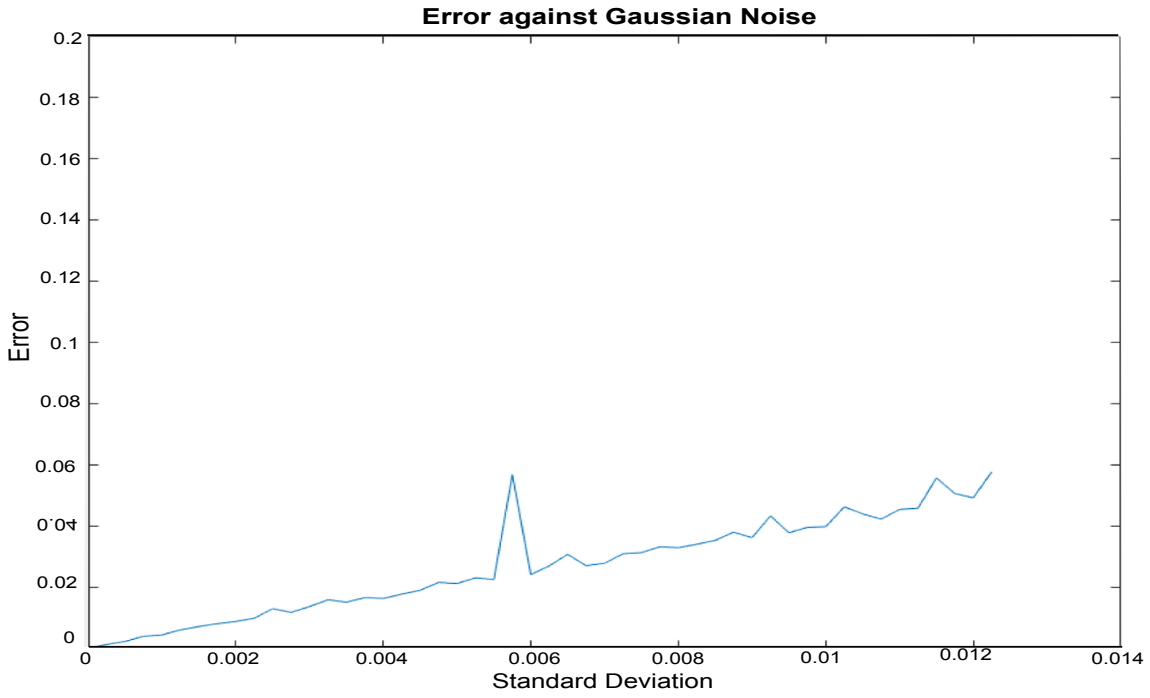


Figure 5.8: Euclidean distance plot between synthetic ground truth and reconstruction for multiple standard deviations.

5.6.2/ 3D RECONSTRUCTION USING OMNIDIRECTIONAL CAMERA

The final result portrayed as a block diagram in Figure 5.5. It referred to the reconstruction of 3D points using the least means square triangulation algorithm (see descrip-

tion in Section 5.4). The feature descriptors and extraction are depicted in Section 5.3. Several techniques were considered to extract features from the scene for adaptation to work with the image derived from the fisheye camera. As depicted earlier, the number of detection and match points on the image depended on several factors, such as size, intensity, and rotation. Therefore, the threshold value was set to reduce the number of outliers.

The mean threshold value provides the best quality of feature points with no or fewer outliers. The result from the stern value of the threshold can be used to estimate the essential matrix. The soft value of the threshold offers more extraction points. Still, the points are usually mixed with outlier, thus giving false result during 3D reconstruction.

In order to verify the capability of the algorithm, the real image captured from the fisheye camera was analysed. The experiment used one camera to capture the same image from two positions, which reflected the camera's movement perpendicular to the object. The camera and the object's positions are displayed in Figure 5.9.

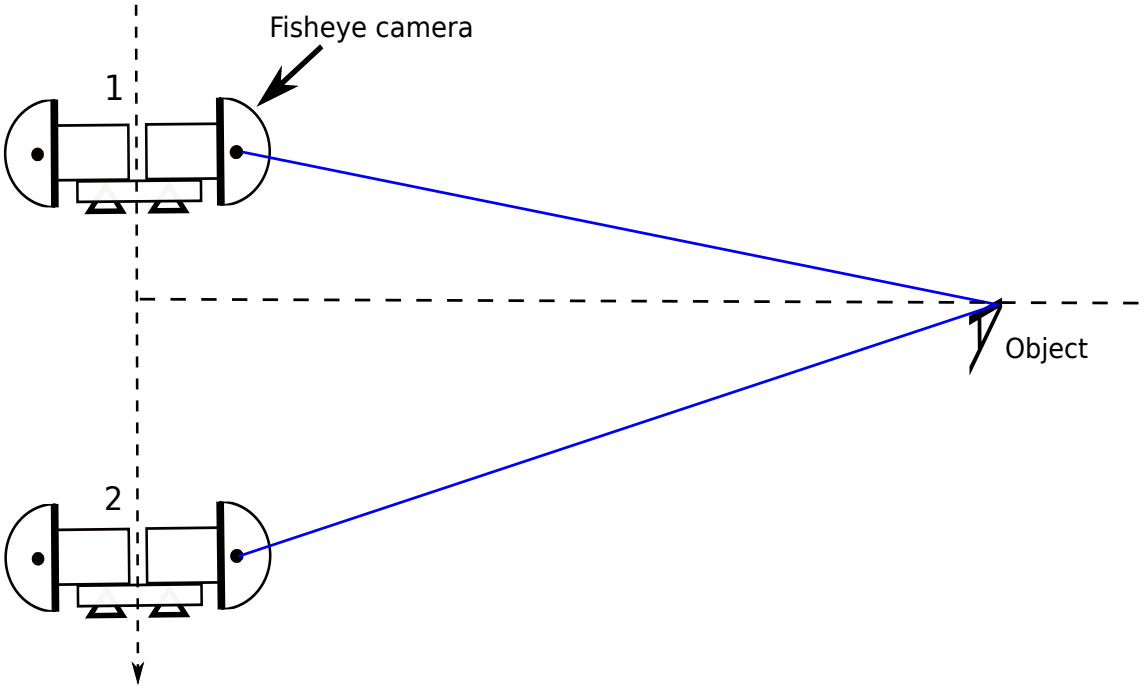
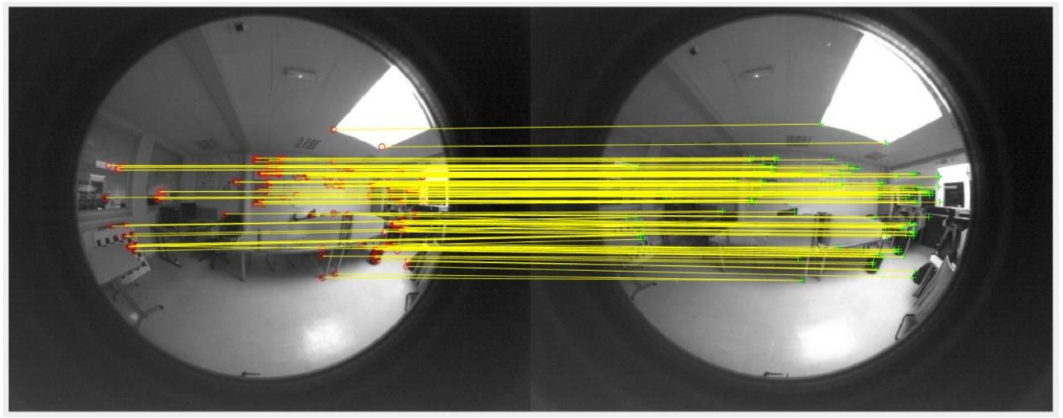


Figure 5.9: Positions of camera and the object.

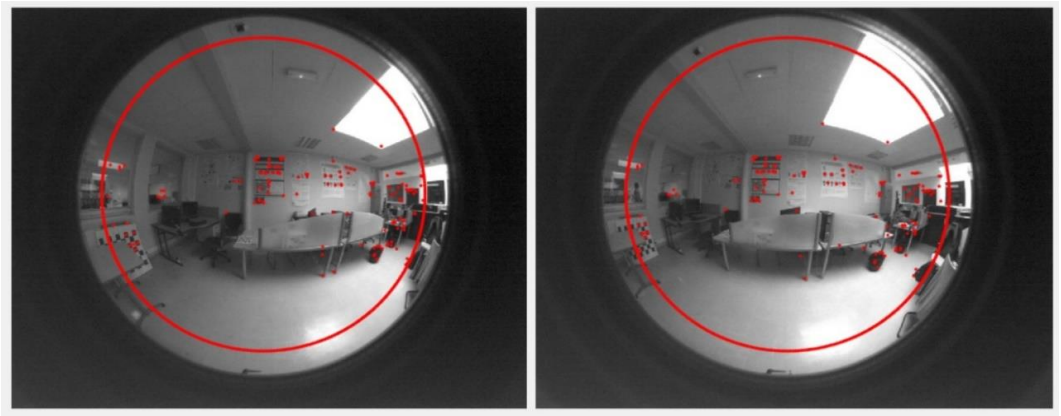
The distance between the camera position and the object was about 2 meters. After capturing the image from the fixed position, point detection was performed to obtain

the point corresponding to the two images. The matching performed between the corresponding points within the two images. Next, the triangulation algorithm was applied to the matched points to generate 3D points.

Several types of features descriptor discussed in Section 5.3 were used to detect



(a): Automatic point detection using features descriptor



(b) The correspondences point between two images from different posses..The circule indicates the 180-degree line of the fisheye camera.

Figure 5.10: Detection of features using automatic features detection.

the features from two images automatically. The threshold was defined by quantifying the best correspondence points without or less outlier. Figure 5.10 displays the matching correspondence points between two images captured at different poses using the automatic features detector. The red circle refers to the 180° line of the fisheye camera (see Section 4.4.1).

As for 3D reconstruction, the maximum number of quality features points required to visualise the object in a 3D environment had been determined. The point selection was performed manually to obtain the maximum points without the outlier.

Figure 5.11 illustrates the result of the process. The original images (see Figure5.11(a)) took from two different positions. In comparison, the distance between the

two positions limited to hinder extreme intensity that could affect the matching algorithm's performance. Figure 5.11(b) exhibits the connection between the corresponding points from the images. Figure 5.11(c) the anaglyph result, which refers to a stereoscopic vision where the two images were superimposed and printed in order to identify the corresponding points of the two images on a single image. As an analogy of human perception, the visual cortex (processes visual information) of the brain fuses anaglyph into the perception of a 3D scene or composition similarity to the study goal.

This study's final goal is to develop the pipeline for 3D reconstruction using the linear least square triangulation. First, the method for the left fisheye image was developed. Next, the same algorithm or method was used for the right fisheye image. Both results were fused to obtain the full view of 3D reconstruction.

Figure 5.12, 5.13 and 5.14 illustrate the estimation of 3D points from front view, top view, and left-side view of the scene, respectively. The real position of the objects could be identified by referring to Figure 5.9. The position of each point estimated in a 3D environment. Points near the edges of the fisheye image were difficult to triangulate due to lens distortion. A small error was noted while clicking the points, whereby a vast difference was generated while calculating the 3D position.

The outcomes were significant as the algorithm attempted to calculate the 3D points for the left of the front and the top views. Essentially, the estimation of 3D points at the centre of the image gave a promising result for 3D reconstruction.

The results concluded that the fisheye image enabled 3D reconstruction using least means square algorithm. However, the precision while detecting features points should be at a high level, especially on the part of the lens dominant with radial distortion. On the positive side, this result revealed the capability of the fisheye camera compared to the perspective camera with limited FoV.

The experiment had proven the significant advantages of using stereo vision from a fisheye camera. With its large FoV exceeding 180°, the proposed hybrid camera's stereo vision successfully generated a large 3D map with slight movement. The combination with another fisheye or omnidirectional camera could produce a complete 3D reconstruction of the environment.

The Omni-Vision camera system had achieved 360° FoV horizontally and vertically. The back-to-back view generated by the two fisheye cameras and the Omni-Vision structure has been briefly explained in the previous chapter. A complete view of 3D re-

construction is successfully performed based on the method prescribed above using the proposed hybrid camera.

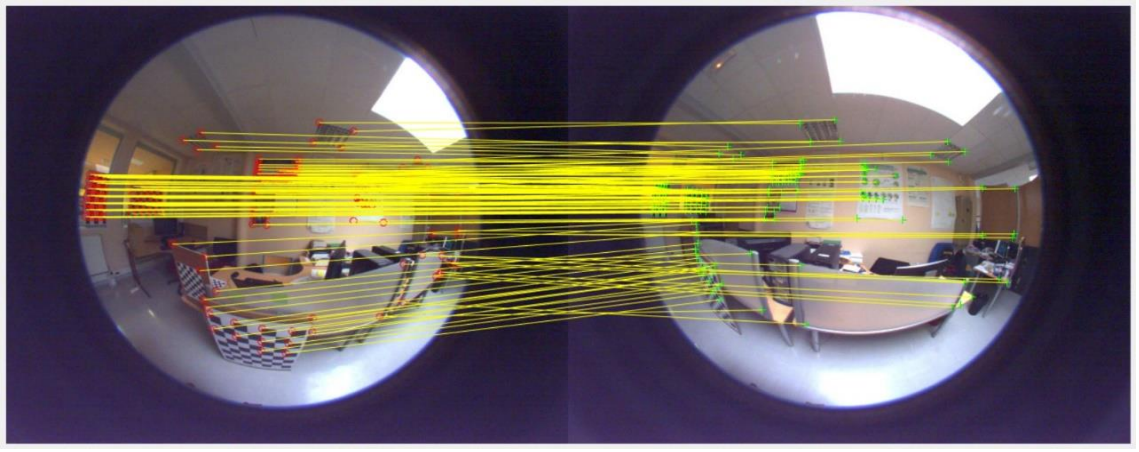
Figure 5.15 displays the perspective view of the 3D result, while Figure 5.10(a) shows the raw image from a fisheye camera. In order to understand the referred 3D points, the objects marked with various colours. Red indicates the ceiling lamp, green denotes the table panel, and blue signifies the chessboard.

5.7/ CONCLUSION

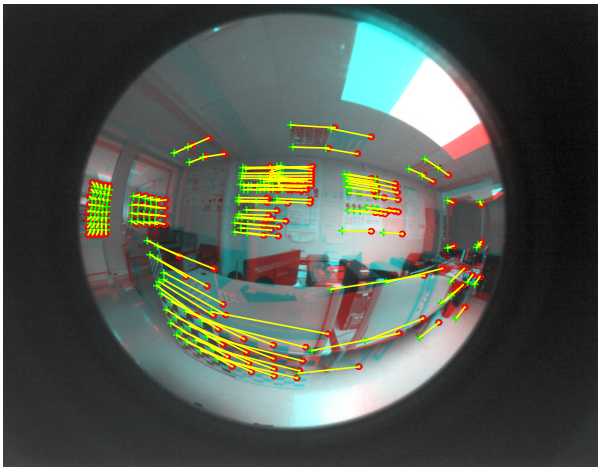
This chapter presents the evaluation that was carried out to assess the system performance. The ability to show 3D reconstructions by the proposed system to perform the task based on the best method has been evaluated. The proposed Omni-Vision camera system had successfully performed 3D reconstruction by exploiting the benefits from the hybrid camera.

The epipolar geometry of the omnidirectional camera was studied as the initial step of the evaluation. Next, suitable features descriptor and matching for omnidirectional camera identified to maximise the detection points. First, a test was conducted using synthetic data. The algorithm appeared stable despite the presence of Gaussian Noise in the synthetic data. The projection of 3D points was calculated using the Least Mean Square Triangulation algorithm, which estimated the promising outcomes of 3D reconstruction. After that, the algorithm fed real data from the fisheye camera. The 3D points estimation results exemplified successful 3D reconstruction from a fisheye camera.

(a): The image taken from fisheye camera



(b) The feature description and points matching between the two images.



(c): Image anaglyph of the wo images.

Figure 5.11: Features detection using manual point clicking.

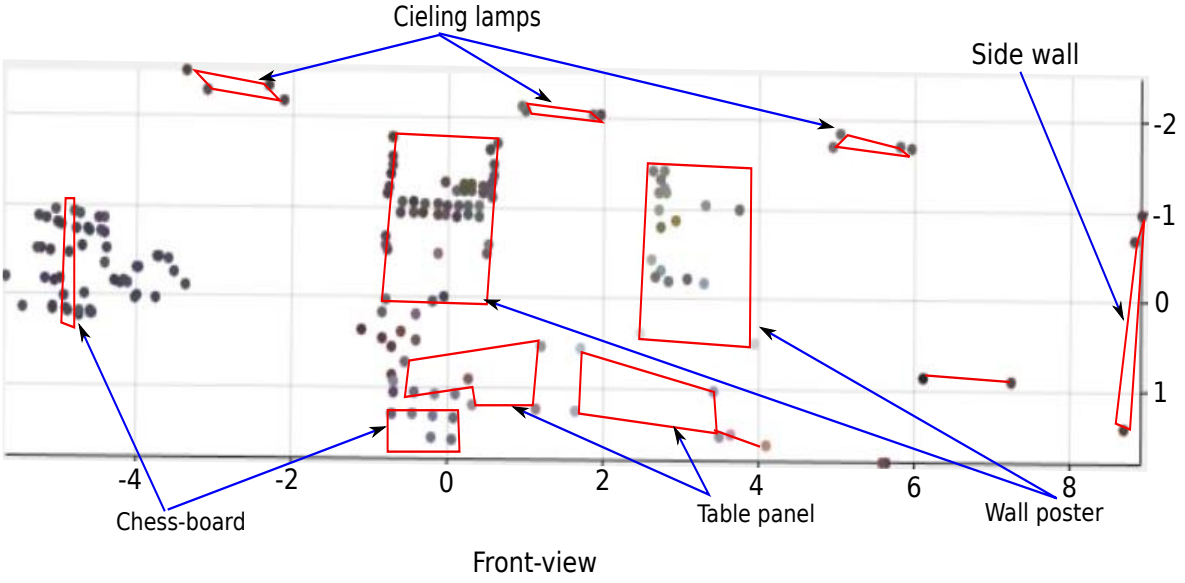


Figure 5.12: Estimation of 3D points looking from front view.

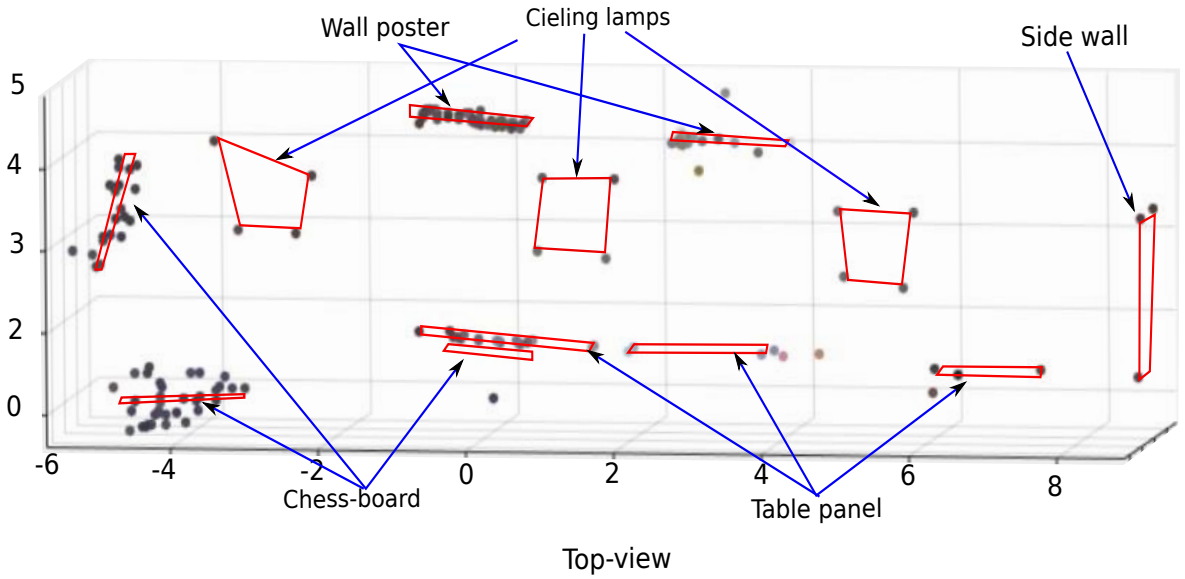


Figure 5.13: Estimation of 3D points looking from top view.

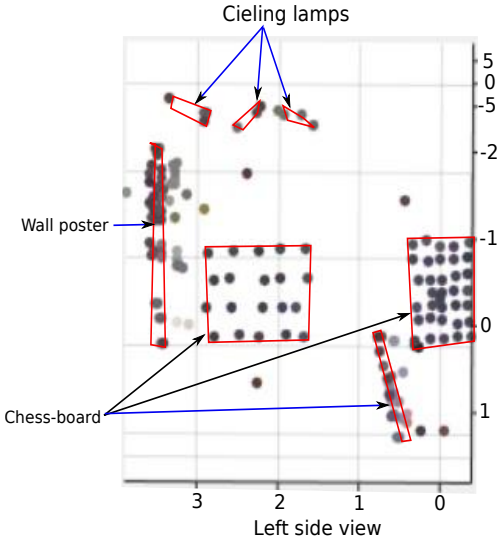


Figure 5.14: Estimation of 3D points looking from side view.

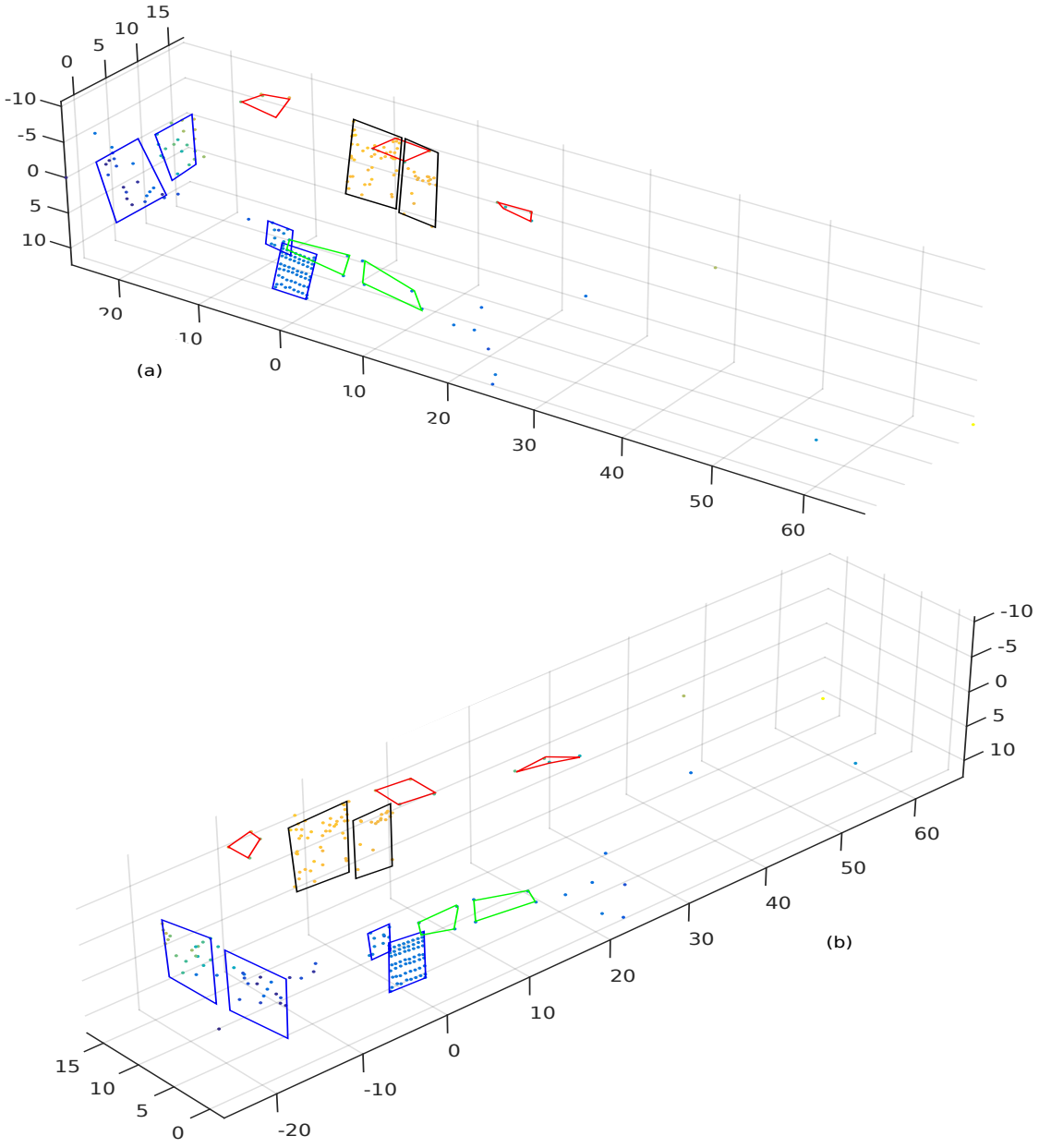


Figure 5.15: Estimation of 3D points in perspective view.

6

Conclusion and Future Works

6.1/ Conclusion

This study has successfully developed a hybrid camera that performs 360° full-view scene vertically and horizontally. The general idea of this research was bio-inspired on how animals view the world in their daily routine. As such, the study combined the visual ability of two groups of animal. The prey category has a wide FoV, while the predator category possesses binocular or stereo vision view. These proposed visions replicated to devise a unique vision system called Omni-Vision. Along with the research work undertaken for this thesis, a full range of literature references and resources contributes to an exceptional starting point regarding the topic at hand.

This thesis was organised as follows: the first chapter introduces the general idea of this research work. The second chapter presents the state-of-the-art, the definition and the history of the optic, along with the camera's evolution and geometry. The third chapter describes the proposed system by considering rigidity and robustness aspects while minimising the number of sensors and cost. The fourth chapter focuses on multicamera calibration and fusion. A new calibration and fusion method is introduced as an enhanced technique. In the fifth chapter, the proposed system executed to evaluate its performance in the real task. Lastly, the sixth chapter concludes the research work and suggests a list of future endeavours.

The general idea of the influence of bio-inspiration on human life is discussed. It describes the two animal categories, prey and predator, which had substantially motivated this research work. The visions of prey and predator animals replicated with a comprehensive FoV camera and binocular/stereo-vision camera. Next, both systems combined

to produce a unique vision system with both capabilities. The expansion of technology reflects the broad application of the vision sensor for extensive exploration beyond human capabilities. The application of vision sensors for autonomous and non-autonomous robots, as well as for video surveillance purposes, suffers from limited FoV. Hence, a novel hybrid camera system with a full view and a window for stereovision view is proposed in this study to address the shortcoming. Additionally, an algorithm is proposed to minimise the number of cameras, rigidity, and ergonomic for robotics application.

The history of optic and camera, including the technical aspects of the camera, is described. The pinhole camera model is the basis of a camera system, which is the reference for any hybrid camera. Internal parameters are essential to describe the unique nature of their optical geometry. Meanwhile, extrinsic parameters define the transformation between the unknown camera reference frame and the known world frame. In short, extrinsic parameters describe the location and the orientation of the camera in the space. The distortion parameters are important dealing with wide FoV camera. Two types of distortion that seldom affect the image's quality are discussed, emphasising radial distortion and the related coefficient. On the positive side, radial distortion increases the FoV of the camera. The omnidirectional camera is the hybrid camera that can increase FoV using several techniques, such as incorporating a reflecting mirror and convex lens and combining several cameras. The three techniques share a common objective - to maximise FoV. However, their advantages are dictated based on their varied unique abilities. The binocular or stereo vision camera captures the depth information of the scene. The stereo vision was obtained using a single camera to capture the same object at different poses or using two cameras to capture the same object at a certain baseline. To date, a compact stereo vision camera can be used to capture the scene's details and for 3D reconstruction. The SVP constraint is a critical aspect when working with a mirror, a lens or using several cameras. There are several forms of mirrors: flat, conic, elliptic, hyperbolic, and parabolic (spherical). Amongst them, flat, hyperbolic and parabolic mirror displayed SVP. The SVP is required to generate pure perspective images from the sensed image. As for the NSVP, such as the fisheye camera, a catadioptric camera with conic and spherical mirrors only approximates the SVP. The pinhole projection model describes the relationship between the 3D world point and the pinhole camera's projection. After that, this projection model is expanded for the omnidirectional camera due to distortion, which deforms the scene. Multiple geometries are a vital tool for 3D reconstruction. The

relationship between several views, such as epipolar geometry, fundamental matrix, and essential matrix, is studied for uncalibrated and calibrated cameras.

The study's goal is to investigate a vision sensor's ability to perform a full-view vision vertically and horizontally, along with high-resolution windows to capture more information about the scene. The criteria emphasised in the proposed system design are cost, rigidity, and ergonomic factors for practical system usage. The fisheye camera selected due to its compact structure and easy-to-manipulate in all robotic applications, including underwater application. Besides, the fisheye camera has no blind spot in the centre of the image than the catadioptric camera. The fisheye camera has a wide FoV that exceeds 180°, thus minimising the proposed hybrid camera's size. Compliance with a unified spherical model for camera calibration further justified selecting this type of camera to provide an exceptionally effective FoV. In order to capture a 360° FoV environment, two fisheye cameras with 185° FoV installed by placing them back to back. A high-resolution ZED camera (stereo vision camera) is placed at the system's anterior to retrieve details of the scene. The blind spot in this combination of three cameras identified and studied, which depends on the fisheyes' cameras baseline.

The three selected cameras were fused to obtain a full view of the scene. Intrinsic calibration was performed for all the cameras to estimate their internal parameters. This process is essential for system execution, such as for 3D reconstruction. The extrinsic calibration determines each camera's position, with the left fisheye camera serving as the reference. At the same time, placing the fisheye cameras back to back with FoV exceeding 180°, an overlapping area produced along the periphery of both fisheye images. This overlapping area exploited to estimate the transformation matrix, which consisted of a rotation matrix and a translation vector. In order to obtain a full 360° view, the images from the three cameras projected onto the unit sphere (with spherical approximation). To do so, Mei's projection model, which is based on a unified spherical model, was employed. The first fusion process is to fuse both left and right fisheye cameras. Pure rotation matrix was applied by enforcing the transformation matrix containing zero translation while minimising the 3D registration errors between the fusion and the two fisheye images. It performed by incorporating the IPOA. The outcomes compared with other methods derived from the literature. Since the FoV of the fisheye camera is more than 180°. The hemisphere formation is assessed by assuming that the 180° line of the fisheye camera should ideally lay on the unit sphere's zero degrees plane. The projected image

should not stretch out on the unit sphere. As the image from planar shape changed to hemisphere, radial distortion generated, and the eccentricity parameter represented. The value of eccentricity affected the formation of the hemisphere. Apart from eccentricity, the baseline (the distance between the two fisheye cameras) also affected the formation of a nearly perfect hemisphere. Hence, a method to re-estimate the value of eccentricity is proposed to fulfil the assumption. This approach incorporated the baseline to address the issue revolving around lengthy baseline distance. The exact process was applied when the perspective ZED camera was fused with the unit sphere. The ZED camera's fusion onto the unit sphere optimised by introducing the affine projective distortion parameter onto the ZED image. The coordinates were fed to the mapping function (Equation 4.27) for re-projection onto the unit sphere. The final result of the fusion revealed that the previous alignment between ZED and fisheye cameras was indeed optimised and yielded promising outcomes for future implementation of the proposed camera system.

The performance evaluated by implementing the vision system for 3D reconstruction. In order to do so, the epipolar geometry of the omnidirectional camera investigated to get the match between the two images. Both epipolar constraint and robust estimation methods (essential matrix calculation) adjusted and realised for images captured using fisheye cameras. The image later projected onto the unit sphere. The features matching process is a critical stage in 3D reconstruction. The optimum essential matrix developed using the points that were estimated by the local features descriptor. Hence, SIFT, SURF, and MSER were the features descriptors used to produce sufficient points for that purpose. Then the points assessed the triangulation process using the Linear Least Square method. The algorithm's performance was assessed using synthetic data, which yielded better outcomes despite the presence of noise. After that, the algorithm pipeline's achievement demonstrated by applying the real images retrieved from fisheye cameras for 3D reconstruction.

6.2/ FUTURE WORKS

A dioptric camera system or the fisheye camera (its commercial term) offers a wide FoV, but has lower resolution as its drawback. This type of camera system is useful for collecting detailed information about the scene. Theoretically, fusing two fisheye cameras

increases the FoV of the camera system. This has been proven with the proposed Omni-Vision camera system that successfully offers a full view FoV. However, due to its low resolution, this camera system is unsuitable for precise and detailed applications. Such camera system is suitable for applications that demand ultra-wide FoV and relatively low accuracy, such as robot navigation, remote inspection, and area surveillance. As such, future work may apply the algorithm from image processing to enhance the quality of the images, for instance, image sharpening and balancing the intensity between left and right images. These can polish the images and enhance feature description. However, the numbers of filter used should be limited because it will reduce the image's critical features.

The intensity level between all of the cameras can detect on the border of the fisheye image on the unit sphere. It becomes critical of one of the cameras facing the light and another camera blotting out the light. To overcome this problem, a method is suggested to integrate both cameras with an autocross exposure algorithm to equalize both images' intensity.

The 3D reconstruction can be enhanced if the number of features points is optimised. Hence, the work depicted in [236] may be extended by enhancing the quality and the quantity of feature description. It was asserted that the method could address ambiguities that occur in images with multiple similar motifs, slightly complicated non-affine distortions, outliers, and detector errors.

In order to yield more reliable matching, a new algorithm is sought for features matching and 3D reconstruction using fisheye image. The new algorithm should solve issues related to radial distortion, especially at the periphery of the fisheye image.

A new system may be developed to increase the FoV of vision system by combining several vision sensors. This demands an effective algorithm to fuse numbers of cameras. Future work may also verify the performance of the algorithm by fusing different types of cameras, such as thermal imaging camera and night vision camera. It would be interesting to add another animal capability to view the world.

Another future recommendation is to produce 360°3D points from three cameras. The preliminary idea is to combine and align all the 3D points constructed by each camera. A method by Khoualed et al. [237] is proposed for the alignment process. The author aligned the 3D data from sixteen different cameras and aligned all the data to produce a 3D view.

Real-time video from a video stream is a challenging task, especially for the proposed Omni-Vision system, which is composed of a combination of several cameras. This work becomes more challenging when dealing with fisheye cameras that have high-level distortion. Hence, future work may apply this Omni-Vision camera system in real-time streaming mode and exploit the advantages offered by this camera vision for scene analysis. Therefore, this study suggests the incorporation of Omni-Vision into the robotic platform to perform several tasks, including navigation, object detection, and surveillance.

I

REFERENCES

BIBLIOGRAPHY

[1] J. Lim, C. Mccarthy, D. Shaw, L. Cole, and N. Barnes, "Insect inspired robots," 01 2006.

[2] R. Moore, *A Bio-Inspired Stereo Vision System for Guidance of Autonomous Aircraft*. 01 2011.

[3] N. Carey and W. Stürzl, "An insect-inspired omnidirectional vision system including uv-sensitivity and polarisation," pp. 312–319, 11 2011.

[4] S. Mafrica, *Bio-Inspired Visual Sensors for Robotic and Automotive Applications*. PhD thesis, 2016.

[5] H. Bakstein and A. Leonardis, "Catadioptric image-based rendering for mobile robot localization," in *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–6, 2007.

[6] L. Silveira, F. Guth, P. Drews-Jr, P. Ballester, M. Machado, F. MORAES, N. Duarte Filho, and S. Botelho, "An open-source bio-inspired solution to underwater slam," vol. 48, 12 2015.

[7] M. Björkman and J.-O. Eklundh, "Vision in the real world: Finding, attending and recognizing objects," *International Journal of Imaging Systems and Technology*, vol. 16, pp. 189 – 208, 01 2006.

[8] D. Kragic and M. Vincze, "Vision for robotics," *Foundations and Trends in Robotics*, vol. 1, pp. 1–78, 01 2009.

[9] Mastcam-Z, "Mastcam-z will be the main eyes of nasa's next mars rover." https://mars.nasa.gov/mars2020/spacecraft/instruments/mastcam-z/.

[10] L. Llorente, S. Barbero, D. Cano, C. Dorronsoro, and S. Marcos, "Myopic versus hyperopic eyes: axial length, corneal shape and optical aberrations," *Journal of Vision*, vol. 4, pp. 5–5, 04 2004.

[11] J. W. Serway, R. A dan Jewett, *Physics for Scien- tists and Engineers with Modern Physics.* Thomson Learning, Inc,, 2008.

[12] R. Möller, “Insect visual homing strategies in a robot with analog processing,” *Biological cybernetics*, vol. 83, pp. 231–43, 10 2000.

[13] K. Weber, S. Venkatesh, and M. Srinivasan, “Insect-inspired robotic homing,” *Adaptive Behavior*, vol. 7, pp. 65–97, 1998.

[14] D. Lyons and F. Fang, “Robust homing with stereovision,” p. 2, 05 2018.

[15] B. Cartwright and T. Collett, “Landmark learning in bees. j comp physiol a,” *Journal of Comparative Physiology*, vol. 151, pp. 521–543, 01 1983.

[16] N. Carey and W. Stürzl, “An insect-inspired omnidirectional vision system including uv-sensitivity and polarisation,” in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 312–319, 2011.

[17] F. Marefat, A. Partovi, and A. Mousavinia, “A hemispherical omni-directional bio inspired optical sensor,” pp. 668–672, 05 2012.

[18] Xiaoming Deng, Fuchao Wu, Yihong Wu, and Chongwei Wan, “Automatic spherical panorama generation with two fisheye images,” in *2008 7th World Congress on Intelligent Control and Automation*, pp. 5955–5959, 2008.

[19] T. Ho, I. D. Schizas, K. R. Rao, and M. Budagavi, “360-degree video stitching for dual-fisheye lens cameras based on rigid moving least squares,” in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 51–55, 2017.

[20] H. M. Dadourian, “Review: Florian cajori, a history of physics,” *Bull. Amer. Math. Soc.*, vol. 35, p. 739, 09 1929.

[21] G. Abetti, “Ronchi v. the nature of light,” *Scientia, Rivista di Scienza*, vol. 65, no. 106, p. 719, 1971.

[22] E. Hecht and A. Zajac, “Optics addison-wesley,” *Reading, Mass*, vol. 19872, pp. 350–351, 1974.

[23] J. L. Coolidge, “The rise and fall of projective geometry,” *The American Mathematical Monthly*, vol. 41, no. 4, pp. 217–228, 1934.

[24] A. Ryberg, B. Lennartson, A.-K. Christiansson, M. Ericsson, and L. Asplund, "Analysis and evaluation of a general camera model," *Computer Vision and Image Understanding*, vol. 115, pp. 1503–1515, 11 2011.

[25] J. Kannala and 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, pp. 1335–40, 09 2006.

[26] D. Brown, "Close-range camera calibration," *Photogramm. Eng.*, vol. 37, 12 2002.

[27] J. Heikkila, "Geometric camera calibration using circular control points," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, pp. 1066 – 1077, 11 2000.

[28] R. Swaminathan and S. Nayar, "Non-metric calibration of wide-angle lenses and polycameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, p. 1172, 10 2000.

[29] R. H. THOULESS, "Apparent size and distance in vision through a magnifying system," *British Journal of Psychology*, vol. 59, no. 2, pp. 111–118, 1968.

[30] J. P. Barreto, R. Swaminathan, and J. Roquette, "Non parametric distortion correction in endoscopic medical images," in *2007 3DTV Conference*, pp. 1–4, 2007.

[31] C. Brauer-Burchardt and K. Voss, "A new algorithm to correct fish-eye- and strong wide-angle-lens-distortion from single images," in *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, vol. 1, pp. 225–228 vol.1, 2001.

[32] F. Devernay and O. Faugeras, "Straight lines have to be straight: Automatic calibration and removal of distortion from scenes of structured enviroments," *Mach. Vision Appl.*, vol. 13, p. 14–24, Aug. 2001.

[33] A. Basu and S. Licardie, "Alternative models for fish-eye lenses," *Pattern Recogn. Lett.*, vol. 16, p. 433–441, Apr. 1995.

[34] K. Miyamoto, "Fish eye lens," *J. Opt. Soc. Am.*, vol. 54, pp. 1060–1061, Aug.

[35] G. Klancar, M. Kristan, and R. Karba, "Wide-angle camera distortions and non-uniform illumination in mobile robot tracking," *Robotics and Autonomous Systems*, vol. 46, pp. 125–133, 02 2004.

[36] R. Hartley and S. B. Kang, "Parameter-free radial distortion correction with center of distortion estimation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, pp. 1309–21, 09 2007.

[37] S. Beauchemin and R. Bajcsy, "Modelling and removing radial and tangential distortions in spherical lenses," pp. 1–21, 01 2000.

[38] Z. Kukelova and T. Pajdla, "A minimal solution to the autocalibration of radial distortion," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–7, June 2007.

[39] R. Strand and E. Hayman, "Correcting radial distortion by circle fitting," in *BMVC*, 2005.

[40] A. E. Conrady, "Decentred Lens-Systems," *Monthly Notices of the Royal Astronomical Society*, vol. 79, pp. 384–390, 03 1919.

[41] S. K. Nayar, "Catadioptric omnidirectional camera," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997. Proceedings., 1997*, pp. 482–488, IEEE, 1997.

[42] C. Hart, "A low-cost omni-directional visual bearing only localization system," 2014.

[43] S. Yoon, W. Han, S. K. Min, and K. S. Roh, "Global localization of mobile robot using omni-directional image correlation," in *Proceedings of the First Pacific Rim Conference on Advances in Image and Video Technology*, PSIVT'06, (Berlin, Heidelberg), p. 433–441, Springer-Verlag, 2006.

[44] M. Liu and R. Siegwart, "Topological mapping and scene recognition with lightweight color descriptors for an omnidirectional camera," *Robotics, IEEE Transactions on*, vol. 30, pp. 310–324, 04 2014.

[45] R. Lukierski, S. Leutenegger, and A. Davison, "Rapid free-space mapping from a single omnidirectional camera," pp. 1–8, 09 2015.

[46] C. Zhang, J. Xu, N. Xi, Y. Jia, and W. Li, "Development of an omni-directional 3d camera for robot navigation," in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), 2012*, pp. 262–267, IEEE, 2012.

[47] K. Watanabe, R. Kawanishi, T. Kaneko, A. Yamashita, and H. Asama, "Obstacle avoidance based on plane estimation from 3d edge points by mobile robot equipped with omni-directional camera," *Advances in Intelligent Systems and Computing*, vol. 193, 01 2013.

[48] J. Delgado-Galvan, A. Navarro-Ramirez, J. Nunez-Varela, C. Puente, and F. Martinez, "Vision-based humanoid robot navigation in a featureless environment," vol. 9116, pp. 169–178, 06 2015.

[49] C. Hart, E. Kreinar, D. Chrzanowski, K. Daltorio, and R. Quinn, "A low-cost robot using omni-directional vision enables insect-like behaviors," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2015, pp. 5871–5878, 06 2015.

[50] O. Cogal and A. Akin, "A new omnidirectional multicamera system for high resolution surveillance," vol. 9120, p. 91200N, 05 2014.

[51] F. Depraz, V. Popovic, B. Ott, P. Wellig, and Y. Leblebici, "Real-time object detection and tracking in omni-directional surveillance using gpu," p. 96530N, 10 2015.

[52] G. Caron, E. Marchand, and E. Mouaddib, "Photometric visual servoing for omnidirectional cameras," *Autonomous Robots*, vol. 35, 10 2013.

[53] M. Liu, "Visual homing from scale with an uncalibrated omnidirectional camera," *IEEE Transactions on Robotics*, 08 2013.

[54] I. Marković, F. Chaumette, and I. Petrović, "Moving object detection, tracking and following using an omnidirectional camera on a mobile robot," in *IEEE International Conference on Robotics and Automation (ICRA), 2014*, pp. 5630–5635, IEEE, 2014.

[55] F. Pasteau, V. K. Narayanan, M. Babel, and F. Chaumette, "A visual servoing approach for autonomous corridor following and doorway passing in a wheelchair," *Robotics and Autonomous Systems*, vol. 75, 09 2014.

[56] P. Chang and M. Hebert, "Omni-directional structure from motion," in *Omnidirectional Vision, 2000. Proceedings. IEEE Workshop on*, pp. 127–133, IEEE, 2000.

[57] B. Micusik and T. Pajdla, "Structure from motion with wide circular field of view cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1135–1149, 2006.

[58] R. Kawanishi, A. Yamashita, and T. Kaneko, "Construction of 3d environment model from an omni-directional image sequence," 2008.

[59] S. Kim and S.-Y. Oh, "Slam in indoor environments using omni-directional vertical and horizontal line features," *Journal of Intelligent and Robotic Systems*, vol. 51, pp. 31–43, 01 2008.

[60] D. Li, D. Weng, H. Zhou, and J. Xie, "Motion interactive system with omni-directional display," in *2013 International Conference on Virtual Reality and Visualization*, pp. 237–240, 2013.

[61] K. Kawauchi and J. Rekimoto, "Quantized reality: Automatic fine-grained spherical images acquisition for space re-construction," in *Proceedings of the 13th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, VRCAI '14, (New York, NY, USA), p. 235–238, Association for Computing Machinery, 2014.

[62] S. Kasahara, S. Nagai, and J. Rekimoto, "Jackin head: Immersive visual telepresence system with omnidirectional wearable camera," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 3, pp. 1222–1234, 2017.

[63] S. Ramalingam, S. Bouaziz, P. Sturm, and M. Brand, "Geolocalization using skylines from omni-images," *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pp. 23–30, 2009.

[64] S. K. Nayar, "Omnidirectional vision," in *Robotics Research* (Y. Shirai and S. Hirose, eds.), (London), pp. 195–202, Springer London, 1998.

[65] J. Neumann, C. Fermüller, and Y. Aloimonos, "Polydioptric camera design and 3d motion estimation," vol. 2, pp. II– 294, 07 2003.

[66] F. Huang, K. Fehrs, G. Hartmann, and R. Klette, "Wide-angle vision for road views," *Opto-Electronics Review*, vol. 21, 04 2013.

[67] J. Gluckman and S. K. Nayar, "Ego-motion and omnidirectional cameras," in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pp. 999–1005, Jan 1998.

[68] C. Ma, L. Shi, H. Huang, and M. Yan, "3d reconstruction from full-view fisheye camera," 2015.

[69] M. Milford, "Vision-based place recognition: How low can you go?," *The International Journal of Robotics Research*, vol. 32, pp. 766–789, 06 2013.

[70] S. Kim, J. Bazin, H. Lee, K. Choi, and S. Park, "Ground vehicle navigation in harsh urban conditions by integrating inertial navigation system, global positioning system, odometer and vision data," *Radar, Sonar & Navigation, IET*, vol. 5, pp. 814 – 823, 11 2011.

[71] J. Nguyen, S. Su, and H. Nguyen, "Spherical vision cameras in a semi-autonomous wheelchair system," *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2010, pp. 4064–7, 08 2010.

[72] K. Daniilidis, R. Klette, and A. Leonardis, *Imaging Beyond the Pinhole Camera*. 04 2004.

[73] Y. Liu, S. Lai, C. Zuo, H. Shi, and M. Zhang, "A master-slave surveillance system to acquire panoramic and multiscale videos," *TheScientificWorldJournal*, vol. 2014, p. 491549, 03 2014.

[74] C.-H. Chen, Y. Yao, D. Page, B. Abidi, A. Koschan, and M. Abidi, "Heterogeneous fusion of omnidirectional and ptz cameras for multiple object tracking," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, pp. 1052 – 1063, 09 2008.

[75] R. I. Hartley and P. Sturm, "Triangulation," *Computer vision and image understanding*, vol. 68, no. 2, pp. 146–157, 1997.

[76] N. Ayache and F. Lustman, “Trinocular stereo vision for robotics,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 1, pp. 73–85, 1991.

[77] K. Saurav, G. Daya, and Y. Sakshi, “Sensor fusion of laser & stereo vision camera for depth estimation and obstacle avoidance,” *International Journal of Computer Applications*, vol. 1, 02 2010.

[78] R. Moore, *A Bio-Inspired Stereo Vision System for Guidance of Autonomous Aircraft*. 01 2011.

[79] Stereolabs, “Zed - depth sensing and camera tracking.” https://www.stereolabs.com/zed/specs/.

[80] S. Baker and S. K. Nayar, “A theory of catadioptric image formation,” in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pp. 35–42, Jan 1998.

[81] V. S Nalwa, “A true omni-directional viewer.,” *Technical report. Bell Laboratories*, 01 1996.

[82] T. Kawanishi, K. Yamazawa, H. Iwasa, H. Takemura, and N. Yokoya, “Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors,” in *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*, vol. 1, pp. 485–489 vol.1, Aug 1998.

[83] H. Hua and N. Ahuja, “A high-resolution panoramic camera,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, pp. I–I, Dec 2001.

[84] H. H. Kar-Han Tan, , and N. Ahuja, “Multiview panoramic cameras using mirror pyramids,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 941–946, July 2004.

[85] S. A. Nene and S. K. Nayar, “Stereo with mirrors,” in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pp. 1087–1094, Jan 1998.

[86] J. Gluckman and S. K. Nayar, “Catadioptric stereo using planar mirrors,” *International Journal of Computer Vision*, vol. 44, pp. 65–79, Aug 2001.

[87] EdmundOptics, “Telecentric lenses,” February 2019.

[88] R. KINGSLAKE, "Optical system design," in *Optical System Design* (R. KINGSLAKE, ed.), pp. 245 – 261, Academic Press, 1983.

[89] M. Watanabe and S. Nayar, "Telecentric optics for computational vision," vol. 1065, 07 1996.

[90] M. Watanabe and S. Nayar, "Telecentric optics for constant-magnification imaging," 02 1996.

[91] M. Watanabe and S. K. Nayar, "Telecentric optics for focus analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 12, pp. 1360–1365, 1997.

[92] M. Born and E. Wolf., *Principles of Optics.* London:Permagon,, 1965.

[93] J. Gluckman, S. K. Nayar, and K. J. Thoresz, "Real-time omnidirectional and panoramic stereo," in *In Proceedings of the 1998 DARPA Image Understanding Workshop*, pp. 299–303, Morgan Kaufmann, 1998.

[94] S. K. Nayar and V. Peri, "Folded catadioptric cameras," in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, vol. 2, pp. 217–223 Vol. 2, June 1999.

[95] S. Cornbleet, *Microwave and optical ray geometry.* 1984.

[96] P. Sturm, "A method for 3d reconstruction of piecewise planar objects from single panoramic images," in *Proceedings IEEE Workshop on Omnidirectional Vision (Cat. No.PR00704)*, pp. 119–126, June 2000.

[97] D. Rees, "Panoramic television viewing system," *U.S. Patent No:3505465 350-25*, 1970.

[98] TSvoboda, S. Pajdla, and V. Aclav Hlav, "Central panoramic cameras: Geometry and design," 01 1998.

[99] V. Grassi Jr and J. Okamoto Jr, "Development of an omnidirectional vision system," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. XXVIII, 03 2006.

[100] Y. Yagi and M. Yachida, "Omnidirectional sensing for human interaction," in *Proceedings of the IEEE Workshop on Omnidirectional Vision 2002. Held in conjunction with ECCV'02*, pp. 121–127, June 2002.

[101] K. Yamazawa, Y. Yagi, and M. Yachida, "Omnidirectional imaging with hyperboloidal projection," in *Proceedings of 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '93)*, vol. 2, pp. 1029–1034 vol.2, July 1993.

[102] Y. Yagi, H. Okumura, and M. Yachida, "Multiple visual sensing system for mobile robot," in *Proceedings of the 1994 IEEE International Conference on Robotics and Automation*, pp. 1679–1684 vol.2, May 1994.

[103] M. Yachida, "Omnidirectional sensing and combined multiple sensing," in *Proceedings 1998 IEEE and ATR Workshop on Computer Vision for Virtual Reality Based Human Communications*, pp. 20–27, Jan 1998.

[104] S. schön Lin, "High resolution catadioptric omni-directional stereo sensor for robot vision," in *In: IEEE International Conference on Robotics and Automation*, pp. 1694–1699, 2003.

[105] J. C. A. Fernandes and J. A. B. C. Neves, "Image quality in catadioptric systems for 360 º field-of-view for mobile robots," in *2006 IEEE International Conference on Industrial Technology*, pp. 994–997, Dec 2006.

[106] H. Nagahara, Y. Yagi, and M. Yachida, "Resolution improving method from multifocal omnidirectional images," in *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, vol. 1, pp. 654–657 vol.1, Oct 2001.

[107] K. Yamazawa, Y. Yagi, and M. Yachida, "Obstacle detection with omnidirectional image sensor hyperomni vision," in *Proceedings of 1995 IEEE International Conference on Robotics and Automation*, vol. 1, pp. 1062–1067 vol.1, May 1995.

[108] A. Barczak, J. Okamoto Jr, and V. Grassi Jr, "Face tracking using a hyperbolic catadioptric omnidirectional system," *Res. Lett. Inf. Math. Sci*, vol. 13, pp. 55–67, 01 2009.

[109] Sheng-WenJeng and W.-H. Tsai, "Analytic image unwarping by a systematic calibration method for omni-directional cameras with hyperbolic-shaped mirrors," *Image Vision Comput.*, vol. 26, pp. 690–701, 2008.

[110] J. P. Barreto and K. Daniilidis, “Unifying image plane liftings for central catadioptric and dioptric cameras,” *OMNIVIS*, vol. 2004, p. 2004, 2004.

[111] J. P. Barreto, “A unifying geometric representation for central projection systems,” *Computer Vision and Image Understanding*, vol. 103, pp. 208–217, 2006.

[112] J. Courbon, Y. Mezouar, and P. Martinet, “Evaluation of the unified model of the sphere for fisheye cameras in robotic applications,” *Advanced Robotics*, vol. 26, pp. 947–967, 2012.

[113] D. Scaramuzza, *Omnidirectional Camera*, pp. 552–560. Boston, MA: Springer US, 2014.

[114] N. Winters, J. Gaspar, G. Lacey, and J. Santos-Victor, “Omni-directional vision for robot navigation,” in *Proceedings IEEE Workshop on Omnidirectional Vision (Cat. No.PR00704)*, pp. 21–28, June 2000.

[115] T. Sato, A. Moro, A. Sugahara, T. Tasaki, A. Yamashita, and H. Asama, “Spatio-temporal bird’s-eye view images using multiple fish-eye cameras,” in *Proceedings of the 2013 IEEE/SICE International Symposium on System Integration*, pp. 753–758, Dec 2013.

[116] Y.-C. Liu, K.-Y. Lin, and Y.-S. Chen, “Bird’s-eye view vision system for vehicle surrounding monitoring,” pp. 207–218, 02 2008.

[117] D. Southwell, A. Basu, M. Fiala, and J. Reyda, “Panoramic stereo,” in *Proceedings of 13th International Conference on Pattern Recognition*, vol. 1, pp. 378–382 vol.1, Aug 1996.

[118] M. Fiala and A. Basu, “Panoramic stereo reconstruction using non-svp optics,” vol. 4, pp. 27– 30 vol.4, 02 2002.

[119] M. Fiala and A. Basu, “Robot navigation using panoramic tracking,” *Pattern Recognition*, vol. 37, pp. 2195–2215, 05 2002.

[120] J. Hong, X. Tan, B. Pinette, R. Weiss, and E. M. Riseman, “Image-based homing,” in *Proceedings. 1991 IEEE International Conference on Robotics and Automation*, pp. 620–625 vol.1, April 1991.

[121] Z. Zhang, R. Weiss, and E. M. Riseman, "Feature matching in 360 degrees waveforms for robot navigation," in *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 742–743, June 1991.

[122] S. Baker and S. K. Nayar, "A theory of single-viewpoint catadioptric image formation," *International Journal of Computer Vision*, vol. 35, 11 1999.

[123] S.-S. Lin and R. Bajcsy, "Single cone mirror omni-directional stereo," *Technical Reports (CIS)*, 01 2001.

[124] S. S. Lin and R. Bajcsy, "The true single view point (svp) configuration for omnidirectional view catadioptric system using cone mirror," 2001.

[125] S.-S. Lin and R. Bajcsy, "Single-view-point omnidirectional catadioptric cone mirror imager," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 840–845, May 2006.

[126] S.-S. Lin and R. Bajcsy, "High resolution catadioptric omni-directional stereo sensor for robot vision," in *2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422)*, vol. 2, pp. 1694–1699 vol.2, Sep. 2003.

[127] S.-S. Lin and R. Bajcsy, "Single-viewpoint, catadioptric cone mirror omnidirectional imaging theory and analysis," *Journal of the Optical Society of America. A, Optics, image science, and vision*, vol. 23, pp. 2997–3015, 01 2007.

[128] Y. Yagi and S. Kawato, "Panorama scene analysis with conic projection," in *EEE International Workshop on Intelligent Robots and Systems, Towards a New Frontier of Applications*, pp. 181–187 vol.1, July 1990.

[129] Y. Yagi, S. Kawato, and S. Tsuji, "Collision avoidance using omnidirectional image sensor (copis)," in *Proceedings. 1991 IEEE International Conference on Robotics and Automation*, pp. 910–915 vol.1, April 1991.

[130] Y. Yagi, S. Kawato, and S. Tsuji, "Real-time omnidirectional image sensor (copis) for vision-guided navigation," *IEEE Transactions on Robotics and Automation*, vol. 10, no. 1, pp. 11–22, 1994.

[131] D. Southwell, B. Vandegriend, and A. Basu, "A conical mirror pipeline inspection system," in *Proceedings of IEEE International Conference on Robotics and Automation*, vol. 4, pp. 3253–3258 vol.4, April 1996.

[132] C. Pegard and E. M. Mouaddib, "A mobile robot using a panoramic view," in *Proceedings of IEEE International Conference on Robotics and Automation*, vol. 1, pp. 89–94 vol.1, April 1996.

[133] S. K Nayar and S. Baker, "Catadioptric image formation," 04 1997.

[134] X. Ying and Z. Hu, "Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model," in *Computer Vision - ECCV 2004* (T. Pajdla and J. Matas, eds.), (Berlin, Heidelberg), pp. 442–455, Springer Berlin Heidelberg, 2004.

[135] C. Demonceaux, *Etude de caméras sphériques : du traitement des images aux applications en robotique*. Habilitation à diriger des recherches, Université de Bourgogne, Nov. 2012.

[136] T.-H. Ho, C. Davis, and S. D. Milner, "Using geometric constraints for fisheye camera calibration," 08 2008.

[137] D. Scaramuzza, "Omnidirectional camera," *Computer Vision: A Reference Guide, Editors: Katsushi Ikeuchi, ISBN: 978-0-387-30771-8 (Print) 978-0-387-31439-6 (Online), Springer, April, 2014*, 01 2014.

[138] J. Courbon, Y. Mezouar, L. Eckt, and P. Martinet, "A generic fisheye camera model for robotic applications," in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1683–1688, Oct 2007.

[139] O. Tahri, Y. Mezouar, F. Chaumette, and P. Corke, "Decoupled image-based visual servoing for cameras obeying the unified projection model," *IEEE Transactions on Robotics*, vol. 26, pp. 684–697, Aug 2010.

[140] R. Pless, "Using many cameras as one," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2, pp. II–587, June 2003.

[141] L. Heng, G. H. Lee, and M. Pollefeys, "Self-calibration and visual slam with a multi-camera system on a micro aerial vehicle," *Auton. Robots*, vol. 39, pp. 259–277, Oct. 2015.

[142] P. Sturm and S. Ramalingam, "A generic concept for camera calibration," vol. 3022, 05 2004.

[143] J. Plucker, "On a new geometry of space," *Philosophical Transactions of the Royal Society of London*, vol. 155, pp. 725–791, 1865.

[144] A. Dipanda and S. Woo, "Towards a real-time 3d shape reconstruction using a structured light system," *Pattern Recognition*, vol. 38, pp. 1632–1650, 10 2005.

[145] S. Prakoonwit and R. Benjamin, "3d surface point and wireframe reconstruction from multiview photographic images," *Image and Vision Computing*, vol. 25, pp. 1509–1518, 09 2007.

[146] D. Zhang, Z. Miao, and M. Wang, "Images based human volumetric model reconstruction and animation," 01 2010.

[147] V. H. Sonka.M and R. Boyle, *3D Vision*. 3rd ed. Toronto: Thomson, 2008.

[148] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. New York, NY, USA: Cambridge University Press, 2 ed., 2003.

[149] A. Heyden, "Tensorial properties of multilinear constraints," *Mathematical Methods in the Applied Sciences*, vol. 23, pp. 169—-202, 2000.

[150] Z. Zhang, Z. Zhang, R. Deriche, R. Deriche, O. Faugeras, O. Faugeras, Q. tuan Luong, Q. tuan Luong, P. Robotique, and P. Robotvis, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence*, vol. 78, pp. 87–119, 1995.

[151] R. Hartley and R. Gupta, "Computing matched-epipolar projections," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 549–555, June 1993.

[152] H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, vol. 293, pp. 133 – 135, September 1981.

[153] R. I. Hartley, “Estimation of relative camera positions for uncalibrated cameras,” in *ECCV*, 1992.

[154] O. D. Faugeras, “What can be seen in three dimensions with an uncalibrated stereo rig,” in *ECCV*, 1992.

[155] R. I. Hartley, “In defense of the eight-point algorithm,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 580–593, June 1997.

[156] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, pp. 381–395, 1981.

[157] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric View-point .* MIT Press, 1993.

[158] E. Michaelsen, W. Hansen, M. Kirchhof, J. Meidow, and U. Stilla, “Estimating the essential matrix: Goodsac versus ransac,” 06 2011.

[159] Jin Yu, A. Eriksson, Tat-Jun Chin, and D. Suter, “An adversarial optimization approach to efficient outlier removal,” in *2011 International Conference on Computer Vision*, pp. 399–406, Nov 2011.

[160] T. S. Huang and A. N. Netravali, “Motion and structure from feature correspondences: a review,” *Proceedings of the IEEE*, vol. 82, no. 2, pp. 252–268, 1994.

[161] R. Hartley, “Euclidean reconstruction from uncalibrated views.,” pp. 237–256, 10 1993.

[162] C. Geyer and K. Daniilidis, “Structure and motion from uncalibrated catadioptric views,” 12 2001.

[163] D. Claus and A. W. Fitzgibbon, “A rational function lens distortion model for general cameras,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, vol. 1, pp. 213–219 vol. 1, 2005.

[164] P. Sturm, “Mixing catadioptric and perspective cameras,” in *Proceedings of the IEEE Workshop on Omnidirectional Vision 2002. Held in conjunction with ECCV’02*, pp. 37–44, 2002.

[165] L. Puig, P. F. Sturm, and J. J. Guerrero, "Hybrid homographies and fundamental matrices mixing uncalibrated omnidirectional and conventional cameras," *Machine Vision and Applications*, vol. 24, pp. 721–738, 2012.

[166] Y. Bastanlar, A. Temizel, Y. Yardimci, and P. Sturm, "Effective structure-from-motion for hybrid camera systems," in *2010 20th International Conference on Pattern Recognition*, pp. 1654–1657, 2010.

[167] J. P. Barreto, "Lifted fundamental matrices for mixtures of central projection systems," 05 2020.

[168] C. Geyer and K. Daniilidis, "A unifying theory for central panoramic systems and practical applications.," pp. 445–461, 01 2000.

[169] U. Sur, "Bio-inspired hemispherical digital cameras of wide-angle field of view," *Current Science*, vol. 107, pp. 18–19, 07 2014.

[170] V. Coffey, "Where life meets light: Bio-inspired photonics," *Opt. Photon. News*, vol. 26, pp. 24–31, Apr 2015.

[171] M. Güzel and M. Ünal, "A survey of insect eye inspired visual sensors," pp. 139–142, 11 2015.

[172] W. Larkin, R. Etienne-Cummings, and J. Van der Spiegel, "Bioinspired imaging: Discovery, emulation, and future prospects [scanning the issue]," *Proceedings of the IEEE*, vol. 102, pp. 1404–1410, 10 2014.

[173] W.-B. Lee, H. Jang, S. Park, Y. Song, and H.-N. Lee, "Compu-eye: a high resolution computational compound eye," *Optics Express*, vol. 24, p. 2013, 02 2016.

[174] W. Maddern and G. Wyeth, "Development of a hemispherical compound eye for egomotion estimation," 01 2008.

[175] S. Li, "Full-view spherical image camera," in *18th International Conference on Pattern Recognition, 2006. ICPR 2006.*, vol. 4, pp. 386–390, IEEE, 2006.

[176] O. Knill and J. Ramirez-Herran, "Space and camera path reconstruction for omnidirectional vision," 09 2007.

[177] Y. gu and M. Veloso, "Multi-model motion tracking under multiple team member actuators," vol. 2006, pp. 449–456, 05 2006.

[178] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A flexible technique for accurate omnidirectional camera calibration and structure from motion," in *Fourth IEEE International Conference on Computer Vision Systems (ICVS'06)*, pp. 45–45, 2006.

[179] Geyer and Daniilidis, "Mirrors in motion: epipolar geometry and motion estimation," in *Proceedings Ninth IEEE International Conference on Computer Vision*, pp. 766–773 vol.2, 2003.

[180] T. Svoboda, T. Pajdla, and V. Hlavac, "Epipolar geometry of panoramic cameras.," pp. 218–231, 06 1998.

[181] L. de la Fraga and O. Schütze, "Direct calibration by fitting of cuboids to a single image using differential evolution," *International Journal of Computer Vision*, vol. 81, pp. 119–127, 02 2009.

[182] Q. Xu, D. Ye, R. Che, and Y. Huang, "Accurate camera calibration with new minimizing function," pp. 779–784, 01 2006.

[183] J. Salvi, "An approach to coded structured light to obtain three dimensional information," 1998.

[184] R. K. Lenz and R. Y. Tsai, "Techniques for calibration of the scale factor and image center for high accuracy 3d machine vision metrology," *Proceedings. 1987 IEEE International Conference on Robotics and Automation*, vol. 4, pp. 68–75, 1987.

[185] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, pp. 1330 – 1334, 12 2000.

[186] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE J. Robotics Autom.*, vol. 3, pp. 323–344, 1987.

[187] J. Heikkilä and O. Silvén, "A four-step camera calibration procedure with implicit image correction," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.

[188] J. Y. Bouguet, "Camera calibrattion toolbox for matlab," 2015.

[189] H. Ingensand, A. Ryf, and T. Schulz, "Performances and experiences in terrestrial laserscanning," 06 2003.

[190] D.Scaramuzza, "Omnidirectional camera and calibration toolbox for matlab.," May 2009.

[191] C. Mei and P. Rives, "Single view point omnidirectional camera calibration from planar grids," in *IEEE International Conference on Robotics and Automation, 2007*, pp. 3945–3950, IEEE, 2007.

[192] J. P. Barreto and H. Araujo, "Issues on the geometry of central catadioptric image formation," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 2, pp. II–II, Dec 2001.

[193] L. Puig, J. Bermudez-Cameo, P. Sturm, and J. Guerrero, "Calibration of omnidirectional cameras in practice. a comparison of methods," *Computer Vision and Image Understanding*, vol. 116, pp. 120–137, 05 2012.

[194] B. Micusik and T. Pajdla, "Estimation of omnidirectional camera model from epipolar geometry," vol. 1, pp. I–485, 07 2003.

[195] A. W. Fitzgibbon, "Simultaneous linear estimation of multiple view geometry and lens distortion," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, pp. I–I, 2001.

[196] S. Thirthala and M. Pollefeys, "Multi-view geometry of 1d radial cameras and its application to omnidirectional camera calibration," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2, pp. 1539–1546 Vol. 2, 2005.

[197] C. Hughes, M. Glavin, and E. Jones, "Simple fish-eye calibration method with accuracy evaluation," *Electronic Letters on Computer Vision and Image Analysis*, vol. 10, pp. 54–62, 12 2011.

[198] J. Courbon, Y. Mezouar, and P. Martinet, "Evaluation of the unified model of the sphere for fisheye cameras in robotic applications," *Advanced Robotics*, vol. 26, pp. 947–967, 07 2012.

[199] P. Lébraly, E. Royer, O. Ait-Aider, and M. Dhome, "Calibration of non-overlapping cameras - application to vision-based robotics," in *Proc. BMVC*, pp. 10.1–12, 2010. doi:10.5244/C.24.10.

[200] M. I. Lourakis and R. Deriche, *Camera self-calibration using the singular value decomposition of the fundamental matrix: From point correspondences to 3D measurements*. PhD thesis, INRIA, 1999.

[201] A. A. Othmani, C. Jiang, N. Lomenie, J.-M. Favreau, A. Piboule, and L. F. L. Y. Voon, "A novel computer-aided tree species identification method based on burst wind segmentation of 3d bark textures," *Machine Vision and Applications*, vol. 27, no. 5, pp. 751–766, 2016.

[202] A. Z. Jamaluddin, O. Mazhar, O. Morel, R. Seulin, and D. Fofi, "Design and calibration of an omni-rgb+ d camera," in *13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), 2016*, pp. 386–387, IEEE, 2016.

[203] Q. Zhang, Y. Li, R. S. Blum, and P. Xiang, "Matching of images with projective distortion using transform invariant low-rank textures," *J. Vis. Comun. Image Represent.*, vol. 38, p. 602–613, July 2016.

[204] M. Y. Kim and H. Cho, "An active trinocular vision system of sensing indoor navigation environment for mobile robots," *Sensors and Actuators A: Physical*, vol. 125, no. 2, pp. 192 – 209, 2006.

[205] J. Gomes-Mota and M. I. Ribeiro, "Mobile robot localisation on reconstructed 3d models," *Robotics and Autonomous Systems*, vol. 31, no. 1, pp. 17 – 30, 2000.

[206] W.-S. Lee and N. Thalmann, "Fast head modeling for animation," *Image and Vision Computing*, vol. 18, pp. 355–364, 03 2000.

[207] S. Horna, D. Meneveaux, G. Damiand, and Y. Bertrand, "Consistency constraints and 3d building reconstruction," *Computer-Aided Design*, vol. 41, no. 1, pp. 13 – 27, 2009.

[208] A. Koutsoudis, F. Arnaoutoglou, and C. Chamzas, "On 3d reconstruction of the old city of xanthi. a minimum budget approach to virtual touring based on photogrammetry," *Journal of Cultural Heritage*, vol. 8, no. 1, pp. 26 – 31, 2007.

[209] F. Bruno, S. Bruno, G. De Sensi, M.-L. Luchi, S. Mancuso, and M. Muzzupappa, "From 3d reconstruction to virtual reality: A complete methodology for digital archaeological exhibition," *Journal of Cultural Heritage*, vol. 11, no. 1, pp. 42 – 49, 2010.

[210] A. Dipanda and S. Woo, "Towards a real-time 3d shape reconstruction using a structured light system," *Pattern Recognition*, vol. 38, no. 10, pp. 1632 – 1650, 2005.

[211] P. Young and G. Wadge, "Flowfront: Simulation of a lava flow," *Computers & Geosciences*, vol. 16, no. 8, pp. 1171 – 1191, 1990.

[212] M. J. O'Sullivan, K. Pruess, and M. J. Lippmann, "State of the art of geothermal reservoir simulation," *Geothermics*, vol. 30, no. 4, pp. 395 – 429, 2001.

[213] N. Ezquerra, L. de Braal, E. Garcia, C. Cooke, and E. Krawczynska, "Interactive, knowledge-guided visualization of 3d medical imagery," *Future Generation Computer Systems*, vol. 15, no. 1, pp. 59 – 73, 1999.

[214] S. E. Mahmoudi, A. Akhondi-Asl, R. Rahmani, S. Faghih-Roohi, V. Taimouri, A. Sabouri, and H. Soltanian-Zadeh, "Web-based interactive 2d/3d medical image processing and visualization software," *Computer Methods and Programs in Biomedicine*, vol. 98, no. 2, pp. 172 – 182, 2010.

[215] S. Prakoonwit and R. Benjamin, "3d surface point and wireframe reconstruction from multiview photographic images," *Image and Vision Computing*, vol. 25, pp. 1509–1518, 09 2007.

[216] J.-S. Park, "Interactive 3d reconstruction from multiple images: A primitive-based approach," *Pattern Recognition Letters*, vol. 26, no. 16, pp. 2558 – 2571, 2005.

[217] D. Zhang and Z. Miao, "Photorealistic 3 d volumetric model reconstruction by voxel coloring," 2010.

[218] P. Torr and D. Murray, "Murray, d.w.: The development and comparison of robust methods for estimating the fundamental matrix. international journal of computer vision 24, 271-300," *International Journal of Computer Vision*, vol. 24, pp. 271–300, 09 1997.

[219] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–, 11 2004.

[220] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer Vision – ECCV 2006* (A. Leonardis, H. Bischof, and A. Pinz, eds.), (Berlin, Heidelberg), pp. 404–417, Springer Berlin Heidelberg, 2006.

[221] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346 – 359, 2008. Similarity Matching in Computer Vision and Multimedia.

[222] R. Kimmel, C. Zhang, A. Bronstein, and M. Bronstein, "Are mser features really interesting?," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, pp. 2316–20, 06 2011.

[223] G. Carneiro and A. Jepson, "Pruning local feature correspondences using shape context," vol. 3, pp. 16 – 19 Vol.3, 09 2004.

[224] E. Mortensen, H. Deng, and L. Shapiro, "A sift descriptor with global context," vol. 1, pp. 184–190, 01 2005.

[225] S. Khoualed, T. Chateau, and U. Castellani, "Semantic-context-based augmented descriptor for image feature matching," vol. 7725, 11 2012.

[226] H. Alzarok, S. Fletcher, and A. Longstaff, "3d visual tracking of an articulated robot in precision automated tasks," *Sensors*, vol. 17, p. 104, 01 2017.

[227] A. Laurentini, "Surface reconstruction accuracy for active volume intersection," *Pattern Recognition Letters*, vol. 17, no. 12, pp. 1285 – 1292, 1996.

[228] T. Borangiu and A. Dumitrache, *Robot Arms with 3D Vision Capabilities*. 04 2010.

[229] B. Fevery, W. Bart, B. Luc, L. Ramon, and C. Torre-Ferrero, *Industrial Robot Manipulator Guarding Using Artificial Vision*. 03 2010.

[230] P. Corke, *Robotics, Vision and Control - Fundamental Algorithms in MATLAB®*, vol. 73. 01 2011.

[231] J.-S. Kim, M. Hwangbo, and T. Kanade, "Spherical approximation for multiple cameras in motion estimation: its applicability and advantages," *Computer Vision and Image Understanding*, vol. 114, pp. 1068–1083, 10 2010.

[232] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *2011 IEEE Intelligent Vehicles Symposium (IV)*, pp. 963–968, 2011.

[233] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski, "Building rome in a day," *Commun. ACM*, vol. 54, p. 105–112, Oct. 2011.

[234] A. Akbarzadeh, J.-M. Frahm, P. Mordohai, B. Clipp, C. Engels, D. Gallup, P. Merrell, M. Phelps, S. Sinha, B. Talton, *et al.*, "Towards urban 3d reconstruction from video," in *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, pp. 1–8, IEEE, 2006.

[235] A. Henrichsen, *3D reconstruction and camera calibration from 2D images*. PhD thesis, University of Cape Town, 2000.

[236] S. Khoualed, T. Chateau, U. Castellan, and C. Samir, "An augmented representation of activity in video using semantic-context information," in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 4171–4175, 2014.

[237] S. Khoualed, U. Castellani, and A. Bartoli, "Semantic shape context for the registration of multiple partial 3d views," in *Semantic Shape Context for the Registration of Multiple Partial 3D Views*, vol. 14, 01 2009.

II

APPENDICES

The list of publications during the thesis period are:

1. Jamaluddin, A.Z., Mazhar, O., Morel, O., Seulin, R., Fofi, D. (2016). Design and calibration of an omni-RGB+D camera. 2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), 386-387.

2. Jamaluddin A.Z., Jiang C., Morel O., Seulin R., Fofi D. (2017) 3D Reconstruction from Specialized Wide FoV Camera System Using Unified Spherical Model. In: Battiato S., Gallo G., Schettini R., Stanco F. (eds) Image Analysis and Processing - ICIAP 2017. ICIAP 2017. Lecture Notes in Computer Science, vol 10484. Springer, Cham.

3. Mazhar O, Jamaluddin AZ, Jiang C, Fofi D, Seulin R and Morel O (2017) Design and Calibration of a Specialized Polydioptric Camera Rig. Front. ICT 4:19. doi: 10.3389/fict.2017.00019

4. Jamaluddin AZ, Jiang C, Fofi D, Seulin R and Morel O (2017) Calibrage et Fusion Système de Camèra Omni-RGB+D à l'Aide d'un Modèle de Caméra Unifiée pour la Reconstruction 3D. ORASIS 2017 sciencesconf.org:orasis2017:139298

III

List of Figures

List of Figures

IV

List of Tables

List of Tables

Document generated with LaTeX and:
the LaTeX style for PhD Thesis created by S. Galland — http://www.multiagent.fr/ThesisStyle
the tex-upmethodology package suite — http://www.arakhne.org/tex-upmethodology/

Abstract:

Camera calibration for a wide field of view is an essential topic in computer vision because of its numerous applications. The use of cameras is prevalent for a robot to see the world or other specialised uses. However, most robots fail to complete the task because of their narrow FoV. This research aims to produce a new camera system with ultra FoV and resolve limited perception. Besides, the system adds a high -resolution stereo view to capture the details of the scene. The literature allows us to study the type of camera, geometry and various method to increase FoV.

A novel calibration method based on Interior-Point Optimization is proposed to outperform state-of-the-art methods. This proposal conducts the design and modelling of a bimodal camera system using two fisheyes and high-resolution ZED camera. The calibration method leads to develop a hybrid camera system with full FoV, which can analyse scene details with high resolution within a window. The method combines multi-cameras setups composed of different characteristics and geometries. It increases the accuracy of the fusion between cameras while handling different levels of optical distortion.

The method is evaluated in terms of the quality of the sphere shape by comparing its accuracy to those obtained with two state-of-the-art approaches. The obtained quantitative results show that the registration errors recorded by our method are much lesser compared to other methods. These results demonstrate clearly that our approach improves the accuracy of the state-of-the-art methods. The end of the application is to implement the system to perform a 360° 3D reconstruction.

Keywords:

Omnidirectional camera, Multi-camera, Camera calibration, Stereovision, Unified spherical model, 3D reconstruction

Résumé :

L'étalonnage de la caméra pour un large champ de vision (CdV) est un sujet essentiel en vision par ordinateur en raison de ses nombreuses applications. L'utilisation de caméras est courante pour qu'un robot puisse détecter son environnement ou pour d'autres utilisations spécialisées. Cependant, la plupart des robots ne parviennent pas à terminer leur tâche en raison de leur CdV étroit. Cette recherche vise à produire un nouveau système de caméra ultra CdV et à résoudre la perception limitée. En outre, le système ajoute une vue stéréo à haute résolution pour capturer les détails de la scène. La littérature nous permet d'étudier le type de caméra, la géométrie et l'augmentation du CdV.

Une nouvelle méthode d'étalonnage basée sur l'optimisation des points intérieurs est proposée pour surpasser les méthodes de pointe. Cette proposition conduit la conception et la modélisation d'un système de caméra bimodale utilisant deux caméras Fisheye et ZED haute résolution. Notre méthode d'étalonnage conduit à développer un système de caméra hybride avec CdV complet, qui peut analyser les détails de la scène avec une haute résolution dans une fenêtre. La méthode combine des configurations multi-caméras composées de différentes caractéristiques et géométries. Il augmente la précision de la fusion entre les caméras tout en gérant différents niveaux de distorsion optique.

La méthode est évaluée en termes de qualité de la forme de la sphère en comparant sa précision à celles obtenues avec deux approches de pointe. Les résultats quantitatifs obtenus montrent que les erreurs d'enregistrement relevées par notre méthode sont bien moindres par rapport aux autres méthodes. Ces résultats démontrent clairement que notre approche améliore la précision des méthodes de pointe. La fin de l'application est d'implémenter le système de reconstruction 3D, et l'objectif est une reconstruction 3D à 360°.

Mots-clés :

Caméra omnidirectionelle, Multi-caméra, Caméra calibration, Stereovision, Modèle Shérique Unifié, Reconstruction 3D

■ École doctorale SPIM - Université de Bourgogne/UFR ST BP 47870 F - 21078 Dijon cedex

■ tél. +33 (0)3 80 39 59 10 ■ ed-spim@univ-fcomte.fr ■ www.ed-spim.univ-fcomte.fr