POLITECNICO DI TORINO

Master's Degree
in Mechatronics Engineering

Master's Degree Thesis

Analysis and Implementation of a Stereoscopic Visual System for a Teleoperated Avatar Robot

Politecnico
di Torino

TOHOKU
UNIVERSITY

Supervisor
Prof. Marcello Chiaberge
Prof. Yasuhisa Hirata

Candidate
Federico Villa

Academic Year 2023-2024

Contents

Abstract

Thanks to technological, sanitary, and social advancements, the number of older people has greatly increased over the years. Many states are now facing the problem of generational change as the number of new births is lower than it was a few decades ago. As a result, the overall population is older than before.

Almost all elderly people suffer from mobility and physical problems, and physical activity is essential for improving mobility, muscle strength, balance, and coordination, reducing the risk of falls and fractures. One way to address the lack of exercise is to use a treadmill for gait training. However, walking on a treadmill can be monotonous, and users cannot accurately assess their walking ability in a real-world environment.

In past research, a solution was created to make physical activity on the treadmill more realistic and immersive. The virtual walking system consists of a treadmill and a robot. The treadmill is automatic, so the pace changes depending on the speed of the user. This speed is then sent to the robot, which moves around the surrounding environment. A joystick is used to control the robot's direction. The user wears an HMD (Head-Mounted Display) to view the images captured by the robot's cameras.

The visual system is one of the most significant components of this setup. Compared with previous solutions, several hardware and software components have been improved. Notably, a new stereo visual system with two 200° cameras has been implemented, compared to a single 360° camera in previous research. This provides the possibility of having two different images of the same scenario, mimicking the human visual system.

These two videos are processed by an external graphics card, allowing for reduced latency in both sending and actual playback. For video playback, a new camera calibration methodology was analyzed. Specifically, a re-production surface was designed to account for radial-type distortions. This surface automatically produces a clean, distortion-free image without relying on computer vision algorithms to modify the image. The analyzed surface is applied in a Unity application and visualized by the user through a new HMD (Meta Quest Pro).

To make the visual experience even more immersive, a rotation system for the robot's cameras was designed to follow the rotation of the HMD. This means every time the user turns their head, the robot's visual system will follow this movement.

In conclusion, due to these developments in both hardware and software, the system is more reliable and immersive. In particular, the reduction in video latency allows users to feel more confident when using the system. The new system is targeted at elderly people who have physical problems and need to perform light physical activity, as well as those who, being forced to stay home, want to engage in exercise.

Abstract

Grazie ai progressi tecnologici, sanitari e sociali, il numero di persone anziane è aumentato notevolmente negli ultimi anni. Molti Stati si trovano ora ad affrontare il problema del ricambio generazionale, poiché il numero di nuove nascite è inferiore rispetto a qualche decennio fa.

Quasi tutti gli anziani soffrono di problemi fisici e di mobilità, e l'attività fisica è fondamentale per migliorare la mobilità, la forza muscolare e l'equilibrio. Un modo per affrontare la mancanza di esercizio fisico è l'uso di un tapis roulant per allenare la deambulazione. Tuttavia, camminare su un tapis roulant può risultare monotono, e gli utenti non riescono a valutare accuratamente la propria capacità di camminare in un ambiente reale.

In ricerche precedenti, è stata sviluppata una soluzione per rendere l'attività fisica sul tapis roulant più realistica e coinvolgente. Il sistema di camminata virtuale è composto da un tapis roulant automatico, la cui velocità si adatta a quella dell'utente. Questa velocità viene poi trasmessa a un robot che si muove all'interno di un ambiente circostante. Per controllare la direzione del robot, l'utente utilizza un joystick. Inoltre, indossando un HMD (Head-Mounted Display), l'utente può visualizzare le immagini catturate dalle telecamere del robot, creando così un'esperienza di allenamento più immersiva e interattiva.

Il sistema visivo rappresenta uno dei componenti più significativi di questa configurazione. Rispetto alle soluzioni precedenti, diversi aspetti hardware e software sono stati notevol-

mente migliorati. In particolare, è stato implementato un nuovo sistema visivo stereo, composto da due telecamere con un campo visivo di 200°, sostituendo la singola telecamera a 360° utilizzata in ricerche passate. Questo approccio consente di ottenere due immagini distinte dello stesso scenario, emulando così il sistema visivo umano. I due video vengono elaborati da una GPU esterna, riducendo la latenza sia nella trasmissione che nella riproduzione. Per garantire una riproduzione ottimale, è stata sviluppata una nuova metodologia di calibrazione della telecamera, progettando una superficie di riproduzione che tiene conto delle distorsioni radiali. Questa superficie produce automaticamente un'immagine chiara e priva di distorsioni, evitando la necessità di complessi algoritmi di computer vision per modificare l'immagine. La superficie analizzata è integrata in un'applicazione Unity e visualizzata dall'utente attraverso un HMD di nuova generazione (Meta Quest Pro).

Per rendere l'esperienza visiva ancora più coinvolgente, è stato progettato un sistema di rotazione delle telecamere del robot, che segue i movimenti dell'HMD. In questo modo, ogni volta che l'utente ruota la testa, il sistema visivo del robot si allinea a questo movimento, creando un'interazione fluida e immersiva.In particolare, la riduzione della latenza video consente agli utenti di sentirsi più sicuri nell'utilizzo del sistema. Il nuovo sistema è rivolto alle persone anziane che hanno problemi fisici e devono svolgere un'attività fisica leggera, nonché a coloro che, costretti a rimanere a casa, desiderano fare esercizio fisico.

Chapter 1

Introduction

1.1 Social Background

The importance of movement, particularly walking, for both mental and physical health cannot be overstated. Walking regularly can significantly improve cardiovascular health by reducing the risk of heart disease and stroke. This simple activity helps maintain a healthy weight, thereby enhancing overall physical fitness and preventing obesity. Regular physical activity, such as walking, stimulates the release of endorphins, which are natural mood lifters, providing a sense of happiness and relaxation. [1]

Furthermore, incorporating walking into your daily routine can improve sleep quality, boost energy levels, and enhance cognitive function. By making walking a regular part of your life, you invest in a healthier body and a more balanced mind, reaping the benefits of this accessible and effective form of exercise.

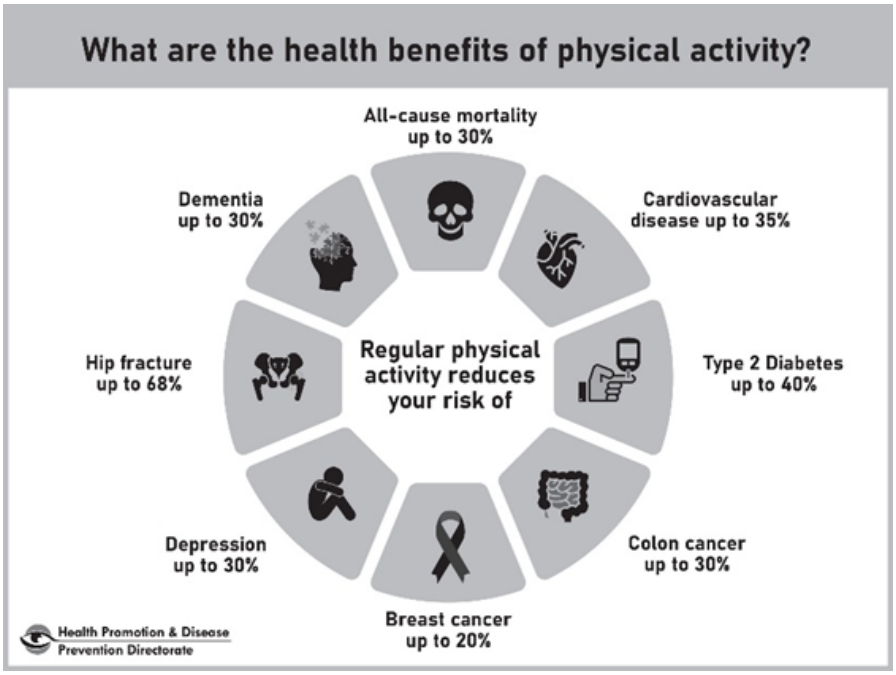

Figure 1.1: Health benefits of physical activity

Today, we live in a world where physical activity is more accessible than ever, and it has evolved to meet the demands of our modern lifestyles. Indoor exercise options have greatly expanded, offering a wide range of activities to suit every preference and need. From home workout videos and online fitness classes to state-of-the-art gyms equipped with the latest exercise machines, there are countless ways to stay active indoors. This shift has been further accelerated by technological advancements, such as fitness apps and wearable devices, which provide personalized workout plans and track our progress. [2]

Moreover, the recent global pandemic has highlighted the importance of indoor physical activity. With outdoor activities restricted and social distancing measures in place, people have turned to home workouts as a safe and effective way to stay fit. The rise of virtual fitness communities and live-streamed exercise classes has made it possible to maintain a healthy lifestyle from the comfort of our homes. As the global population

ages, the need to make indoor physical activity accessible to everybody becomes crucial. Regular exercise is essential for managing chronic conditions such as heart disease and arthritis, which are more common among older adults. However, many seniors and individuals with health issues face challenges with outdoor exercise due to mobility limitations and other factors. Indoor physical activity at home offers a controlled and safe environment, making it an ideal solution for those with limited mobility or health issues.

Region	Number of persons aged 60 years or older in 2017 (millions)	Number of persons aged 60 years or older in 2050 (millions)	Percentage change between 2017 and 2050	Distribution of older persons in 2017 (percentage)	Distribution of older persons in 2050 (percentage)
World	962.3	2080.5	116.2	100.0	100.0
Africa	68.7	225.8	228.5	7.1	10.9
Asia	549.2	1273.2	131.8	57.1	61.2
Europe	183.0	247.2	35.1	19.0	11.9
Northern America	78.4	122.8	56.7	8.1	5.9
Latin America and the Caribbean	76.0	198.2	160.7	7.9	9.5
Oceania	6.9	13.3	92.6	0.7	0.6

Table 1.1: Data on older persons (60+ years) by region for 2017 and 2050.

1.2 Walking Training

Walking exercise is one of the most popular activities suggested for elderly people. It provides a safe way to stay active without putting stress on the body. This is particularly important for those with conditions like arthritis, where high-impact activities could cause pain or damage.
Generally, the walk can be carried out in two different environments:

- **Indoor Walking**: it's the best choice when living in a polluted city and can be done in any weather condition. Moreover, the treadmill displays relevant training data such as velocity, burnt calories, and heart rate.

- **Outdoor Walking**: it has more benefits for humor and mental health, in particular exercise in natural environments leads to greater improvements in psychological well-being.[3]

One common tool for indoor walking is the treadmill. It consists of a moving platform with a wide conveyor belt driven by an electric motor. The treadmill was invented by William Staub, an American mechanical engineer, in the 1960s [4]. Over the years, this device has seen a significant increase in use, and today, treadmills remain the best-selling exercise equipment by a large margin.
The success is due to some key factors:

- Most treadmills come with adjustable settings for speed and incline, as well as pre-programmed workouts.

- Users can do other things while on the treadmill, such as watch television or read, which for many can help keep the exercise interesting.

- The treadmill provides a consistent surface that is easier to navigate than sidewalks, curbs, or trails, reducing the risk of tripping.

Figure 1.2: Treadmill

Initially, the treadmill was used exclusively for cardio and fitness activities, but over the years it was also introduced for rehabilitation purposes for patients with movement and coordination problems.

The treadmill allows walking to be simulated in a controlled environment. This makes it especially useful for people who have had lower extremity injuries or surgeries or who suffer from neurological conditions such as stroke or multiple sclerosis.

In rehabilitation, the treadmill is also used to analyze the patient's walking patterns; physical therapists can observe the movement in detail and act in real-time to correct any abnormalities in the patient's gait. This is useful for improving gait quality and accelerating functional recovery.

1.2.1 Virtual Reality Training

In recent years, the growth of Augmented Reality (AR) and Virtual Reality (VR) has been remarkable, driven by technological advances and increased computing power of devices. Initially developed for the video game industry, these technologies have found applications in various fields, such as industry, medicine, education and vocational training.

It is important to emphasize the difference between AR and VR:

Virtual Reality (VR) is a technology that simulates three-dimensional digital environments, allowing users to immerse themselves in virtual experiences as if they were real. Through devices such as VR viewers, headsets and motion controllers, VR isolates users from the physical world and transports them into a virtual space to explore, interact and experience simulated situations.

Augmented Reality (AR) is a technology that enriches the real environment by superimposing digital elements, such as images, text, sound, or three-dimensional objects, onto the physical world. Using devices such as smartphones, tablets or AR glasses, users can view this virtual content in real-time, integrated with what they see in the surrounding world.

AR and VR may seem like two similar concepts to each other, but the substantial difference is that VR transports users into entirely new digital worlds, offering an interactive experience through the use of headsets or specialized glasses. In contrast, AR enhances the real-world surroundings by adding digital elements on top of them, providing additional information or expanding their functionality.[5]

A very interesting application of virtual reality for patients with Parkinson's desease has been studied in the studies of N. Brandín-De la Cruz [6], where there is a use of

both a bodyweight support and Virtual reality environment. The results obtained are very positive from both clinical and patient perspectives, this research has opened new perspectives and new approaches for medical rehabilitation.

Figure 1.3: N. Brandín-De la Cruz case study

1.3 Past Research

At Tohoku University's Smart Robots Design lab, a design of an autonomous treadmill connected to a Robot avatar was created by student Promsutipong Kengkij [7].

This system aims to make indoor walking motion more realistic, thus making physical activity more fun and stimulating. The main goal is to be able to visualize a realistic immersive environment while walking on the treadmill, the system is designed for different types of users, support rails are provided for the elderly or for people who are in rehabilitation or have motor diseases.

The complete system consists of an autonomous treadmill and a robot avatar, in fact the user walking on the treadmill makes the robot move, which can move freely in an outdoor environment. With the treadmill autonomous the speed can vary freely as if it were a real walk, a joystick is used to enable users to move the robot left or right.

To make the activity more realistic a VR system has been installed, specifically the user, wearing a head set can view on screen the images captured by a video camera installed on the robot.

In this way, it is evident how the sense of realism compared to a classic indoor physical activity is greater.

Figure 1.4: Overall system idea

1.4 Research Topic

One of the problems in the previous research was the significant lack of reality in live video streaming. The perception of the depth was very low, which caused a sense of disorientation in the user's view. With the old video system, composed of a 360 FOV camera, there was the sense of not controlling the moving robot and this can cause real damage to the user's safe state.

Another problem was the too-high latency in the streaming. Latency in video refers to the delay between the capture or creation of the video content and its display. High latency can cause disorientation due to the significant delay between real-time events.

This research is about the implementation of a new stereoptical video system composed by two fisheye 200 FOV cameras Sony IMX 219, mounted on a Raspberry Pi 5, that can capture live streaming.

The images obtained from fisheye cameras have distortions that do not allow the image to be displayed optimally, to solve this a new correction system was analyzed.

This new method consists of directly modifying the display on which the images are projected, these modifications are based on a mathematical model that considers the distortion parameters of the cameras.

To reduce the latency, it's implemented a new GPU Intel ARX, connected with the Raspberry with an ethernet cable. GPUs are designed to handle large volumes of data simultaneously, speeding up video encoding and decoding compared to traditional CPUs.

The video streaming is playing in a Unity application, that is implemented on a HDM meta quest pro. The HDM and the camera system servo are connected, in fact, the head angle values, obtained by the Unity sensor, are used to turn in real-time the cameras, in this way the camera system can follow the orientation of the user's head.

Chapter 2

System Configuration

Previous research explored an immersive virtual walking system utilizing an avatar robot. This system allows users to experience the sensation of outdoor training while remaining indoors. Users can control the avatar robot using a treadmill with automatic speed control and a joystick, enabling the avatar robot to mimic their movements. Equipped with a camera on the avatar robot and a VR visor, users can view their surroundings and enhance the sensation of being outdoors [7] .

Figure 2.1: Virtual walking idea

2.1 Treadmill

The key aspect of the virtual walking system is to create an experience that allows users to walk naturally. Traditional treadmills, which operate at a constant speed, fall short of this goal. To address this, previous research developed an automatic speed control treadmill designed to enable a more natural walking experience. This section provides an overview of the mechanical system and explains the control method used for the automatic speed control treadmill.

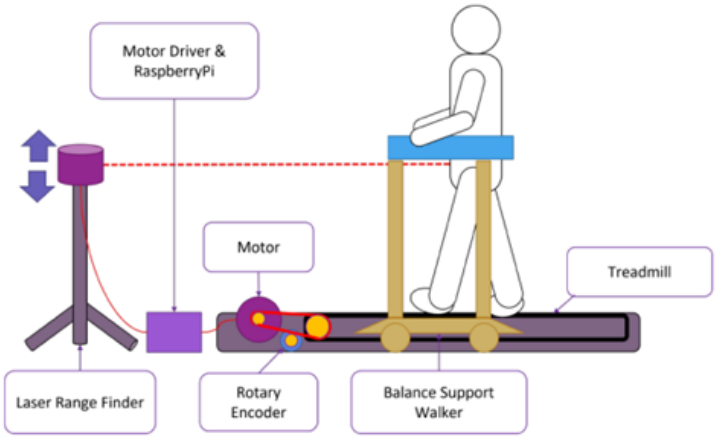

Figure 2.2: Treadmill model system

As shown in figure 4, the main components are:

- The Laser Range Finder (LRF) (Hokuyo UBG-04lx-f01) is used to track the user's position on the treadmill, providing feedback for the control system outlined below. The sensor is positioned on a tripod to allow for height adjustments, as it needs to be at the same level as the user's pelvis. The scan data are filtered to isolate only the distance measurement between the LRF and the user, ensuring the control

system functions accurately.

- Motor Driver and Raspberry Pi: The motor is a DC brushed type with a voltage of 110 VDC, a rated current of 6.8 A, a power output of 1 horsepower (746 W), and a speed of 3100 RPM. It drives the pulley that transfers mechanical power to the treadmill's pulley through a belt. Speed control is achieved by varying the voltage applied to the motor. When an extra load (a person) walks on the treadmill, the motor draws additional current to increase torque and maintain the set speed.

- Rotary Encoder (Omron E6A2-CW3C): This device measures the user's walking speed directly from the belt. Since the rotary encoder is connected to the belt that drives the treadmill, the belt's linear velocity can be interpreted as the treadmill's linear velocity.

- Raspberry Pi 3 Model B+ with Ubuntu 16.04 and ROS Kinetic serves as the computing unit to interface with the laser range finder and rotary encoder. It processes the data and executes high-level control algorithms, sending control signals to the motor driver through GPIO (General Purpose Input/Output) pins.

- Balance Support Walker is added to the system to provide a sense of balance and safety to the user during the treadmill training.

According with the user's velocity the treadmill speed is modified. To implement this, a P controller with position-feedback gain for velocity is used.

$$v(t) = K_p\,(x(t) - x_0) + C$$

- K_p, Proportional gain
- $x(t)$, User's position
- x_0, Reference position (located at the end of the treadmill)
- C, Bias constant

The treadmill belt is analyzed in three areas, the first one, the red one, defined as the "stopping zone," is the one where the treadmill speed is null until $x(t) > x_0$, as a consequence, also the robot velocity becomes null. Next, there is the yellow zone, where the treadmill speed follows the proportional law, and finally, there is the green zone where the position $x(t) > x_{\text{MAX}}$. In this case, the $v(t)$ is saturated to a maximum value v_{MAX}.

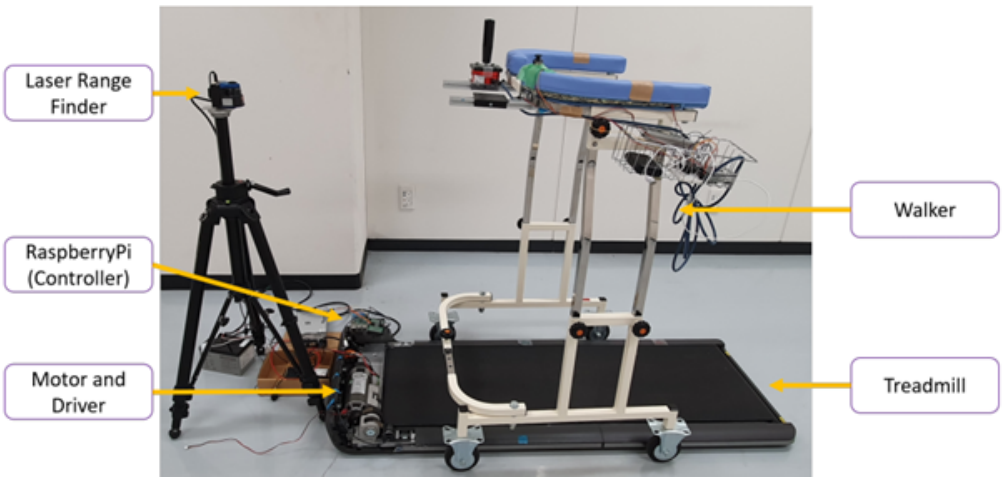

Figure 2.3: Treadmill system use

2.2 Avatar Robot

The focus of this system is the relationship between the treadmill and the robot, particularly the latter, that is an avatar robot that can impersonate the user's figure in the external environment.

When the user walks on the treadmill, he can teleoperate the robot, adjusting both its speed and the direction of its trajectory. An avatar robot avatar offers a way to be present and interact remotely when physical presence is difficult or impossible.

In recent decades, the use of robot avatars has grown in many areas, for example, in the rehabilitation and health sector, robots have been introduced to improve the recovery process of patients by making it more efficient and functional. Another area of interest is the social sector, especially in recent years, many businesses are using these solutions within restaurants or stores to better automate the work process. As an example, OhmniLabs Ohmni and Beam-Pro by Suitable Technologies are two avatar robots used to navigate and interact in different environments, these robots are equipped with an iPad mount and can navigate around environments while providing high-quality video conferencing capabilities. They feature autonomous driving, obstacle avoidance, and a self-balancing system.

Additionally, these systems have found applications in education and collaborative workspaces, enabling individuals to attend classes or meetings remotely with a greater sense of presence.

Advanced developments in sensors and AI are now allowing avatar robots to perform more complex tasks, such as manipulating objects or responding to user commands with greater accuracy. As the technology evolves, avatar robots are becoming indispensable tools for bridging physical distances and enhancing human-robot interaction in various sectors.

2.2.1 Fetch Freight 100

The avatar robot used in this research is a Fetch Freight100 OEM Base by Fetch Company. The Fetch Freight100 OEM Base is widely used in agriculture, warehousing, and industry due to its extensibility and versatility, allowing it to support various applications as a safe autonomous mobile platform [8].

Key features include a differential drive with 8 wheels supporting up to 100 kg. The base features a modular top plate with 73 threaded mounting points, allowing for easy integration of hardware and sensors. A 2D laser scanner with a 25m range enables mapping, localization, obstacle avoidance, and object detection. Additionally, the robot is equipped with two 3D depth cameras for superior vision, allowing it to avoid both ground-based and overhanging dynamic obstacles.

For connectivity, the robot includes two power ports, a direct communication bus, Ethernet, and USB 3.0 ports on the top of the base for accessory connections. A side panel provides access to Ethernet, USB 3.0, and DisplayPort. The robot also features Wi-Fi connectivity, making it easy to integrate into various environments and applications.

Figure 2.4: Freight100 OEM Base

2.2.2 Differential Drive

This robot is based on differential drive, that is a vehicle where two drive wheels have different angular velocity, respectively right velocity W_r and left W_l, both wheels are mounted on a common axis, and the direction of motion can be either forward or backward. The fetch has two motor each of these connected with one wheel, the different angular velocity allows the robot to rotate around a point (Centre of Rotation) that lies along its common left and right wheel axis. Each hub motor is held in a drop suspension configuration, allowing the robot to keep traction when crossing obstacles, without allowing the robot to sway side-to-side when manipulating on flat ground. Four casters provide stability during movement.

Model equations:

- $\omega = \dfrac{r(w_R - w_L)}{d}$
- $v = \dfrac{r(w_R + w_L)}{2}$

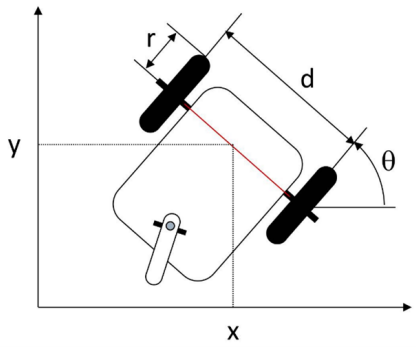

Figure 2.5: Differential drive model

The physical inputs are the angular velocity of the two wheels, w_R and w_L. d is the distance between the wheels and r is the wheel's radius.

2.3 Overall system connection

The overall system can be studied in two subsystems. The one related to the treadmill and the one related to the robot avatar, in figure 2.6 the treadmill includes the components with a green box, while those with orange borders are related to the fetch.

- **Treadmill**: The user controls the linear speed of the robot by walking on the treadmill. Also, through a joystick, the user can vary the robot's rotation. These two data are managed by the Unity application that will connect the treadmill system with the robot system. In addition to this data, the rotation speed of the headset itself is also managed.

- **Avatar robot**: The Robot Controller, by receiving information from the Headset, can control the robot's direction and speed and the rotation of the robot's visual system. In addition, the robot controller also has sensor data from the two Lidar reporting information for object detection. Finally, the visual system will send information regarding the external environment to the 'application on Unity, where the user can interact.

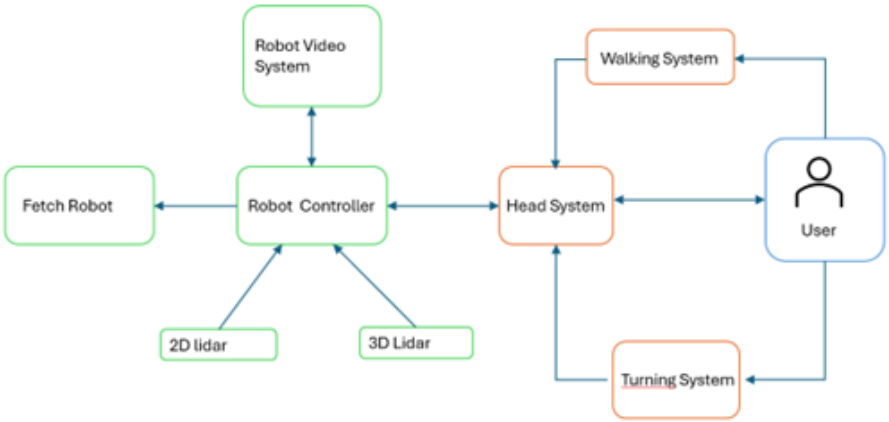

Figure 2.6: Overall system connection

Chapter 3

Stereoscopic System

Stereopsis is the visual ability of humans to perceive the depth of space through binocular mechanisms. The human organism is equipped with a visual system consisting of two frontal eyes that receive slightly different images of the same object.

Such differential perspective is known as parallax binocular since the eyes are separated horizontally by about 6/6.5 cm in an adult. The horizontal retinal disparity generated is used to communicate the relative depth of objects.

Stereoscopic or binocular vision is defined as a person's ability to be able to integrate images coming from two different eyes. Our brain then combines these two images into one, allowing us to perceive depth and see the world in 3D [9].

The main features of binocular vision:

- **Binocular disparity**, which is the difference in the position of the images on the right and left of the same object. The disparity values of objects arranged in a scene allow us to infer their size and depth. In the process of the creation of stereoscopic images, the disparity can be defined differently than it would appear, offering an important artistic cue.
- **Fusion**, the brain's ability to combine the two slightly different images captured by our eyes to create a unique perception of depth and three-dimensionality. The activity of the muscular apparatus serves to visually align the axes to project the images onto corresponding retinal areas, allowing the mind to create a single, unified representation.
- **Stereopsis**, this process determines the spatial position of each point of the image concerning the point of fixation. Derived through analysis of the horizontal retinal disparity subsisting between the two images [10].

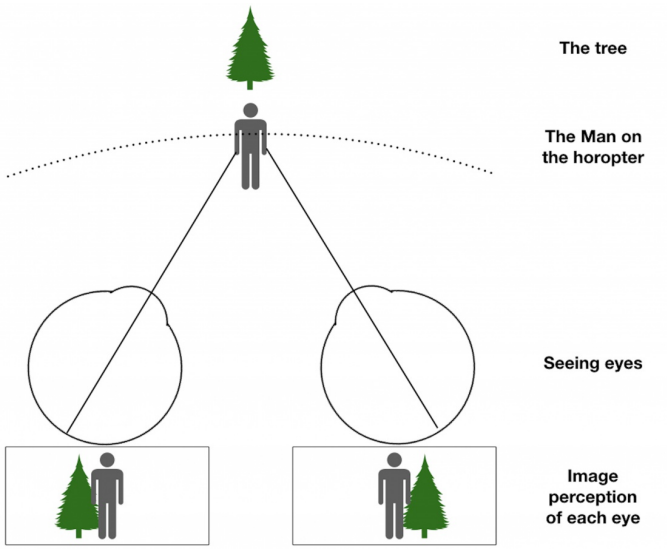

Figure 3.1: Binocular Vision

3.1 Pinhole Camera Model

The pinhole camera model describes the mathematical relationship between the coordinates of a point in three-dimensional space and its projection into the image plane of an ideal pinhole camera, where the camera aperture is considered a point.
In this type of model, distortions aren't considered, due to not considering the nature of the lens and the finite size of the aperture hole. Consequently, it is one of the simpler methods to map a 3D point in the camera plane [11].
In the geometrical model, the main features are:

- Camera coordinate system, where the origin is defined as the center of projection.
- Focal length f, the distance between the center of projection and the image plane.
- Image coordinate system, where the 3D point is projected.
- Pixel coordinates refer to the position of a point in an image, measured in pixels. In this coordinate system, the origin (0, 0) is usually located at the bottom-right corner of the image.

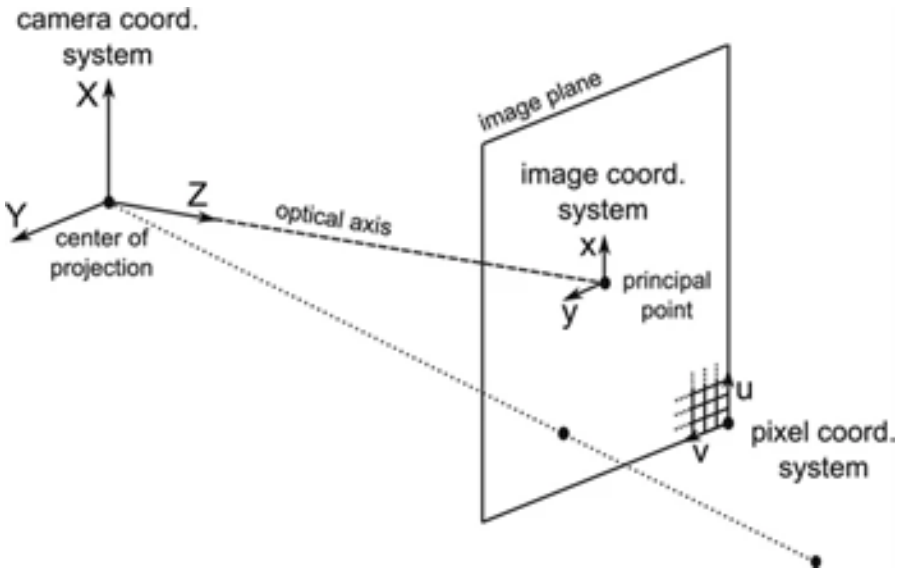

Figure 3.2: Pinhole Camera Model

Considering a point with coordinate (X, Y, Z) with respect to the camera reference system. Applying similar triangle analysis, it is possible to determine the coordinates in the reference system of the image plane:

- $x = f\frac{X}{Z}$
- $y = f\frac{Y}{Z}$

The coordinates in the pixel coordinate system are obtained by applying a translation corresponding to the shift of origin and scaling corresponding to the conversion from metric coordinates to pixel ones:

$$u = k_u(x + x_0) = k_u f\frac{X}{Z} + k_u x_0$$

$$v = k_v(y + y_0) = k_v f\frac{Y}{Z} + k_v y_0$$

The factors k_u and k_v represent the density of pixels in the image sensor along the horizontal (per column) and vertical (per row) directions.

The expression in matrix form is

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \begin{bmatrix} k_u f & 0 & k_u x_0 & 0 \\ 0 & k_v f & k_v y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Figure 3.3: New stereoscopic video system

3.2 Distortion Effects

The captured image can have distortion effects due to several aspects, in particular, the focal length plays an important role. Depending on its distance, a barrel or pincushion effect can characterize the image. In addition, wide-angle lenses include more pieces of glass that are curved so the portions of the image that are at the edges of the frame can become skewed, and the image will reflect this curvature.
It's possible to describe two main categories of distortions, Radial and Tangential.

Radial Distortions

Radial distortions cause straight lines to appear curved. Radial distortion becomes larger the farther points are from the center of the image.
They can be divided in two categories:

- **Bearing Distortion** (Barrel Distortion): Straight lines appear curved outward, creating a bearing-like effect. This type of distortion is common in wide-angle lenses.
- **Pincushion Distortion** (Pincushion Distortion): Straight lines appear curved inward as if the image were compressed toward the center. This effect is most common in telephoto lenses.

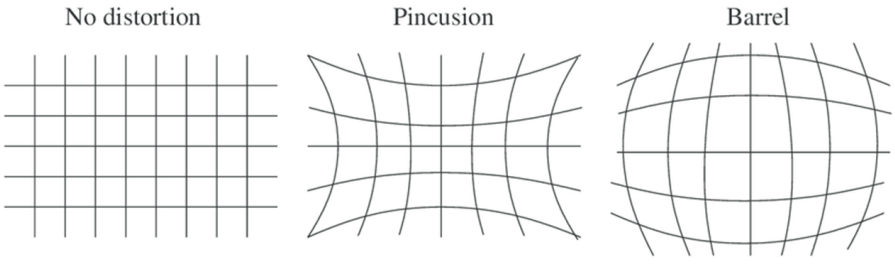

Figure 3.4: Radial Distortions

Tangential Distortions

Tangential distortions are caused by lens misalignment relative to the sensor plane. It's sometimes also called de-centering distortion, because the primary cause is due to the lens assembly not being centered over and parallel to the image plane.
This type of distortion can cause images to a pear tilted or distorted laterally.

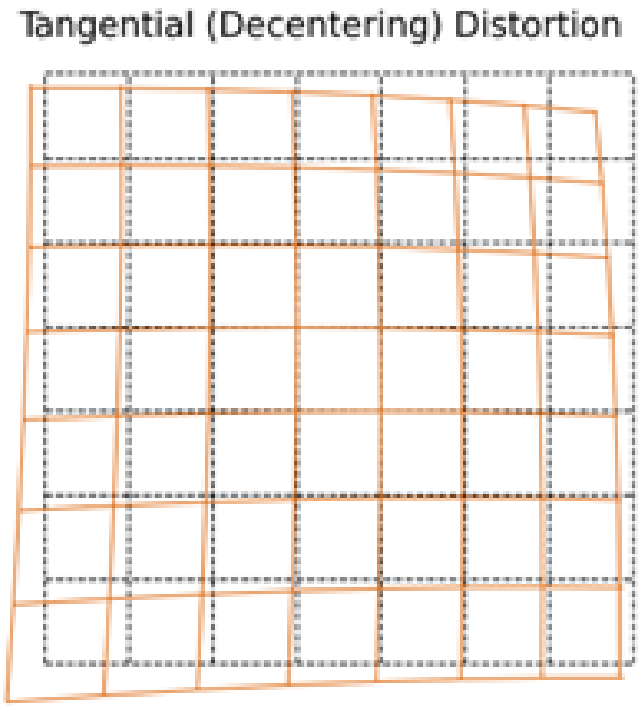

Figure 3.5: Tangential Distortions

3.3 Fisheye Lens

In the new video system are used two cameras IMX219 Camera, with 800 megapixels and 200 FOV [12]. They are wide-angle photographic lenses, the images produced by a fisheye are circular, although in some cases when 180° coverage is not needed, part of the edge, or even the entire circular edge, is excluded.

The special feature of fisheye lenses is their ability to produce curved lines, unlike wide-angle lenses. While in wide-angle lenses the curved lines are a defect to be corrected, usually in fisheye lenses they are a characteristic feature.

With fisheye lenses, only one horizontal and one vertical line, which form a cross in the center of the image, remain straight and perfectly perpendicular. All other lines appear symmetrically curved.

A fisheye lens allows the photographer to add more artistic expression and drama to his photographs [13].

Figure 3.6: Sony IMX 219 FOV 200°

Figure 3.7: Fisheye image example

Chapter 4

Image Correction

The novelty brought in this research is the method of image display inside the dome. To make the experience more immersive, the inner surface of a hollow hemisphere was chosen as the dome; the original fisheye image will be directly projected onto it.

This innovative approach ensures that the viewer is fully enveloped in the visual content, creating a more engaging and realistic simulation environment. Unlike traditional flat or cylindrical screens, the hemispherical dome eliminates the limitations of peripheral vision, allowing for a more expansive field of view.

This chapter will explain the projection of a fisheye image in a hemispherical dome and the effect this projection has on the original image distortions. To address these distortions, tests were conducted to study the characteristics of the camera used for image capture. These tests were essential to gather data on how the lens and sensor affect the projection, enabling the creation of a mathematical model to correct the distortions. The resulting model ensures that the projected images are accurately represented on the curved surface of the dome, preserving spatial consistency and enhancing the overall visual experience.

4.1 Fisheye Equidistant Projection

Fisheye images cannot be modeled with the pinhole model because of fundamental differences in projection principles. The pinhole model assumes a perspective projection that creates images with straight lines kept straight, whereas fisheye lenses capture a much wider field of view, up to 180 degrees or more, by intentionally distorting straight lines into curves. This nonlinear distortion allows fisheye lenses to represent a wide angle of view in a single frame, a feature impossible to replicate with the pinhole pattern.

The projection of pinhole model images is represented by the following relationship:

$$r(\theta) = f \tan \theta$$

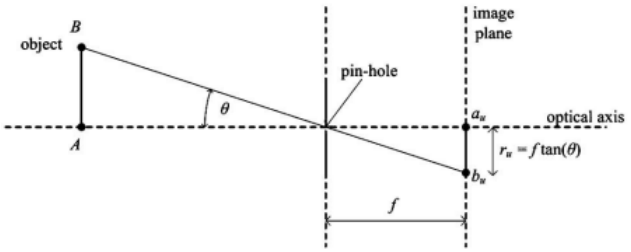

Figure 4.1: Pinhole Projection

Where r is the fisheye radial distance of a projected point from the center, f is the focal distance, θ is the angle between the generic vector position in the scene and the focal axis vector.

There are different types of fisheye image projection:

- $r(\theta) = 2f \tan \theta$ stereographic projection
- $r(\theta) = f\theta$ equidistant projection

- $r(\theta) = 2f \tan \theta/2$ equisolid projection
- $r(\theta) = f \sin \theta$ orthogonal projection

The equidistant projection is one of the most widely used. This type of projection maintains a linear relationship between the angle of view field and the radial distance from the center of the image. Therefore, each uniform increase in the field angle corresponds to a uniform increase in the radial distance [14].

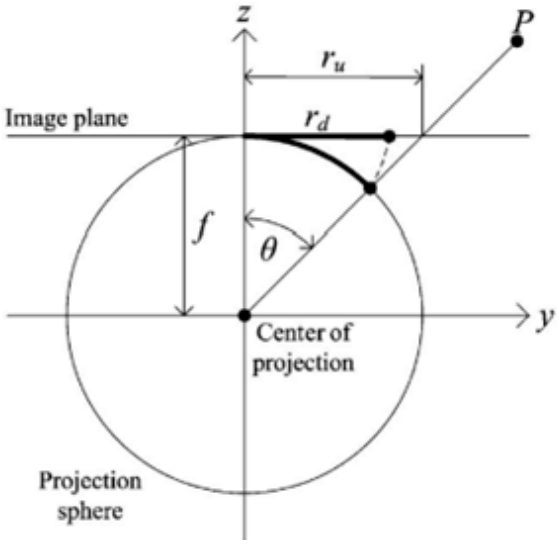

Figure 4.2: Equidistant Projection

The images obtained from the IMX camera can be approximated as a fisheye image with FOV 180, these have considerable radial distortions, particularly barrel distortions, distortion that curves the straight lines toward the center of the image.

By projecting a hemispherical fisheye image onto a hemispherical surface the distortions, are corrected due to the natural correspondence between the two curved surfaces. When this image is mapped onto a hemispherical surface, every point in the image is realigned with the curvature of the surface, visually eliminating the distortions. This process

restores more natural proportions and angles, making the image representation truer to reality. In addition, a hemispheric dome envelops the user, increasing the sense of immersion and taking full advantage of the wide field of view of the fisheye projection. This approach makes the VR experience smoother and more convincing.

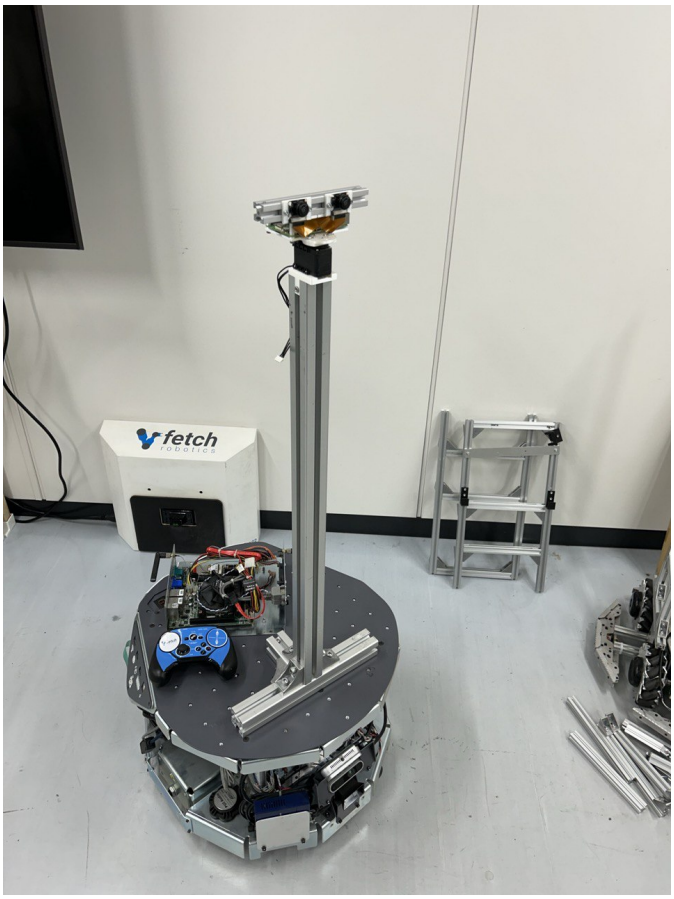

Figure 4.3: Fetch Robot with the new video sytem

4.2 Fisheye Lens Correction

In a perfectly circular fisheye projection, the radial distance r from the center of the image is directly proportional to the latitude (the angle between the 3D vector and the optical axis of the camera). This implies that each linear increment in the distance from the center of the image corresponds to a linear increment in the angle of the 3D scene [15].

Along any radial line in the fisheye, two points separated by the same distance in pixels correspond to the same angular distance in the 3D scene.

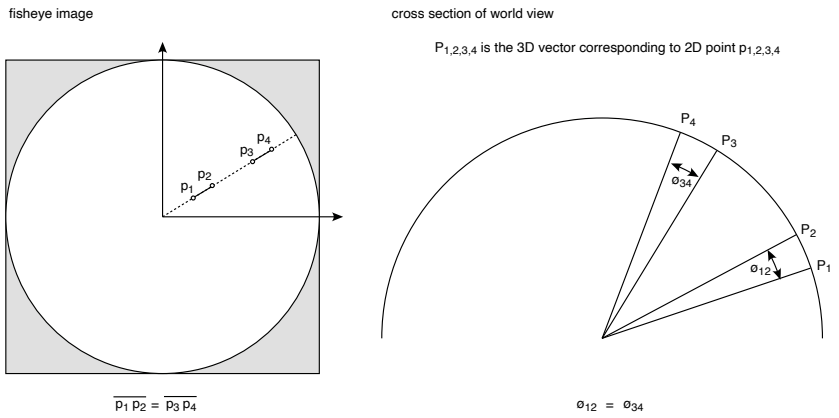

Figure 4.4: Ideal Fisheye Properties

In practice however, many fisheye lenses do not show linear proportionality between the distance r on the sensor or fisheye circular image and the field angle (or "field angle") of the 3D vector corresponding to that distance [16].

This leads to a non-uniform distortion effect, where image compression becomes more evident toward the edges. This behavior is common on less expensive lenses.

Lens manufacturers often provide data describing the curve that relates the distance r with the field angle (or "field angle") for a specific lens. This data is essential for the correction of distortions and accurate representation of the scene. For high-quality lenses, manufacturers can provide specific curves for each lens, considering variations that may occur during the manufacturing process [17].

Figure 4.5: Real Fisheye

Figure 4.6: Ideal Fisheye

4.2.1 Calibration Test

A typical expression that describe the relation between the distance r with the field angle θ :

$$\theta(r) = a_0 + a_1 r + a_2 r^2 + \ldots + a_n r^n \tag{4.1}$$

However, Sony does not provide an expression formula for the IMX 219 sensor. Therefore, a calibration test is performed to determine its characteristics and behavior in practical use.

The procedure involves several steps :

1. **Determining the Zero Parallax Position :** Initially, the user aligns a close object and a distant one, adjusting the camera's position on the slider rail until both objects remain aligned despite rotations of the camera/lens. For the IMX 219, the zero-parallax point appears closer to the front of the lens than usual.

   The term "zero parallax" refers to the axis or point in a stereoscopic setup, such as a stereo camera or VR system, where objects appear aligned without parallax, meaning they do not appear to shift relative to a monocular view.

2. **Choosing an Object in the Scene :** Next, an object in the scene is selected and aligned at mid-height and width on the lens. This alignment can be facilitated using tools like a GIMP with crosshairs or utilizing a camera's built-in crosshair or alignment grid.

3. **Rotating the Camera/Lens:** The camera/lens is rotated through the lens's field of view in increments of 5 degrees. At each step (angle), a photograph is captured. For the IMX 219 lens, this rotation spans 180 degrees.

4. **Pixel Distance Measurement :** For each photograph (angle), the distance in pixels from the center horizontally to the chosen object is measured. These distances are then plotted against the corresponding angle.

5. **Curve Fitting :** A function is fitted to the plotted data points. This function helps describe the relationship between the angle of rotation and the pixel distance from the center, aiding in understanding the lens's distortion characteristics.

Three IMX 219 cameras were analyzed in the calibration process. It is important to emphasize how for all the chambers the zero-parallax axis is located almost exactly between the front of the lens and the sensor.

In the images below the setup for the experiment is represented.

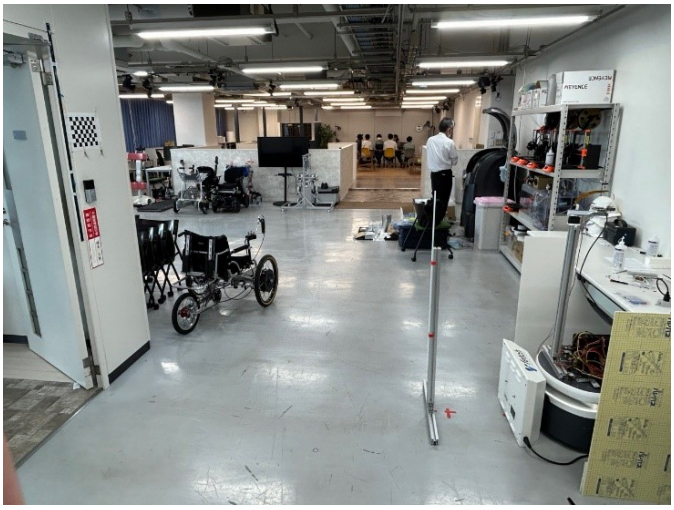

Figure 4.7: Experiment Setup

Figure 4.8: Experiment Setup

4.2.2 Experiment Results

The test was performed on 3 different cameras, so it is possible to both assess the difference between ideal and real behavior but also whether there is distinction between the same cameras.

The graph shows on the y-axis the image distance along the normalized radius, and on the x-axis angle the field angle. For convenience, the horizontal axis was chosen as the reference axis point for calculating the distance of the image centers of the various steps. The curve obtained reflects the a priori assumptions; in fact, the compression effect begins to occur at the edges of the image. As a result, peripheral details are squeezed into a smaller portion of the image. Another interesting aspect is the concavity of the curve, this is an indication that the compression effect, in fact, the points farthest from the center are compressed, but less drastically than in a curve with a convex curve.

Considering the graphs obtained, it can be established that the behavior of the tested cameras is the same for all of them.

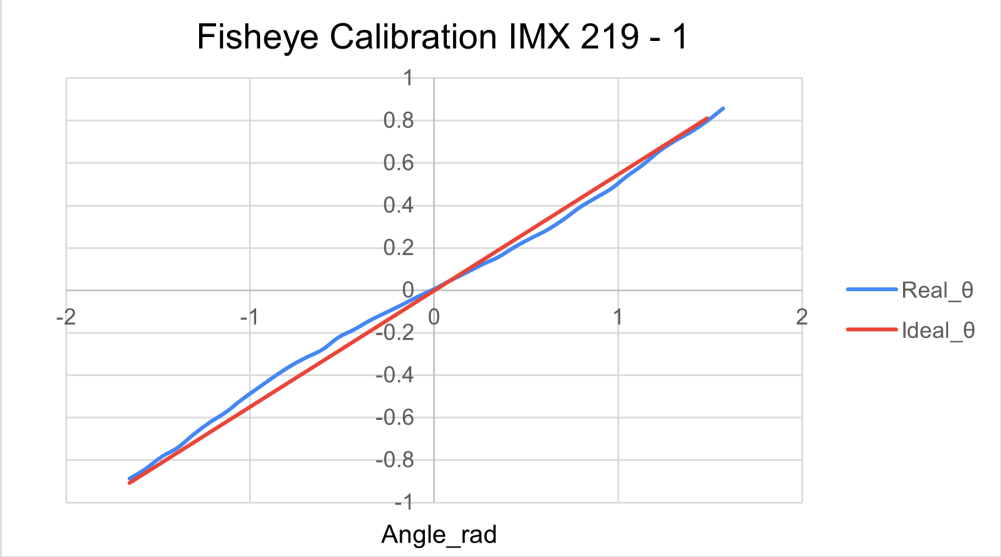

Figure 4.9: Experiment Setup Camera 1

From the graph, it is possible to compare the difference between the linear equation representing ideal fisheye and the nonlinear equation of the second degree, which on the contrary identifies the real behavior of the camera.

Ideal Behavior	$\frac{r}{R} = 0.5474 \cdot \theta$
Real Behavior	$\frac{r}{R} = -0.0029 \cdot \theta^2 + 0.517 \cdot \theta + 0.01145$

Table 4.1: Comparison between the ideal and real behaviors of Camera 1.

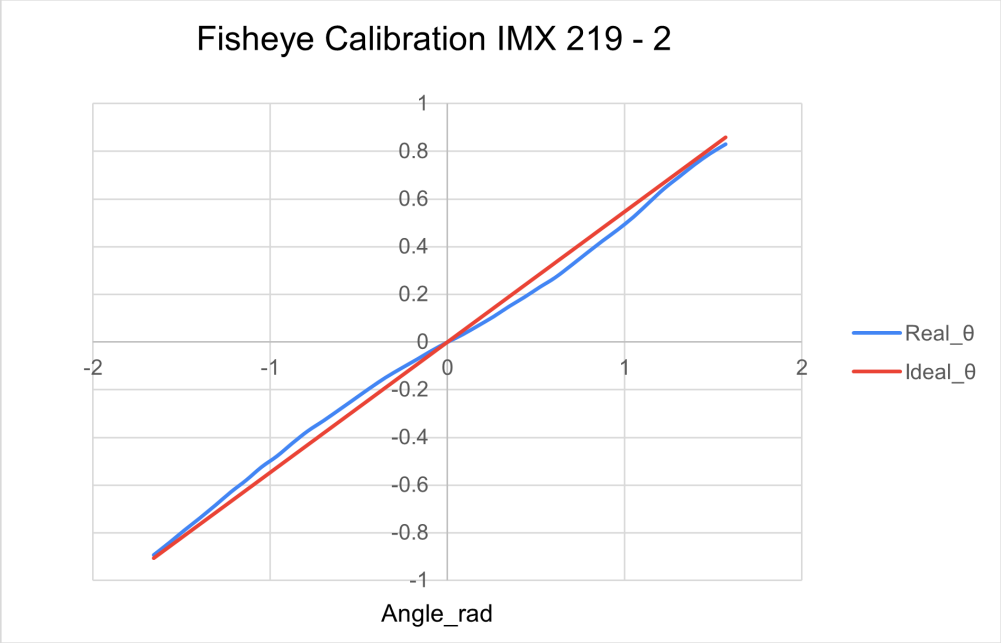

Figure 4.10: Experiment Setup Camera 2

Ideal Behavior	$\frac{r}{R} = 0.5474 \cdot \theta$
Real Behavior	$\frac{r}{R} = -0.0016 \cdot \theta^2 + 0.5138 \cdot \theta + 0.00015$

Table 4.2: Comparison between the ideal and real behaviors of Camera 2.

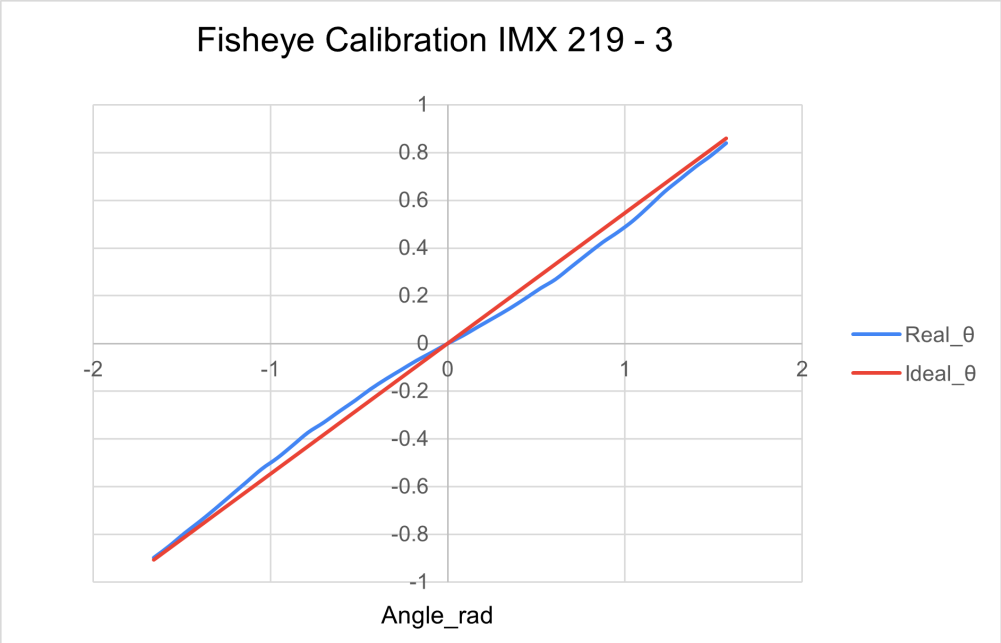

Figure 4.11: Experiment Setup for Camera 3

Ideal Behavior	$\frac{r}{R} = 0.5474 \cdot \theta$
Real Behavior	$\frac{r}{R} = -0.0037 \cdot \theta^2 + 0.515 \cdot \theta$

Table 4.3: Comparison between the ideal and real behaviors of Camera 3.

4.3 Creation of a Dome and UV Map Algorithm

The creation of the hemispheric dome is done on Blender software, this choice is because the software is very convenient for creating gaming applications and virtual environments, in fact the interconnection between Blender and Unity is optimal [18].

A UV map is a two-dimensional representation of the coordinates of a three-dimensional surface, which allows a texture (an image) to be applied accurately to a 3D model. The UV map "unwinds" the surface of the model in a 2D plane, where each vertex of the 3D model is mapped to a position on the texture. This process ensures that the texture aligns correctly with the model, allowing visual details such as colors, patterns, and materials to be added accurately.

The main idea is to change the UV map of a hemisphere, which can only be used if the fisheye image is ideal. This modification is based by following the mathematical model obtained from the calibration of the camera in fact from those parameters it is possible to analyze how the UV map of the dome should be so that if the real fisheye image is projected onto it the effects of distortions are corrected, so it is as if projecting the image onto the modified dome would have the same effect if the fisheye was ideal in the dome without modification.

The creation and editing operations were carried out using the Blender application. After obtaining a hemisphere to modify the UV map by algorithm, Blender Python API must be used [19].

The algorithm of UV mapping creation and modification is characterized by the following steps:

1. Creation of the default UV mapping characterized by vertices positioned as if they were on concentric rays. This UV map is a perfect equidistant one.

2. For each vertex, calculate the distance from the center (radius), store in a list the different significant radii that characterize the UV map.

3. Store the coordinates of each vertex in a dictionary, where each key is the reference radius of the corresponding vertices.

4. For each reference radius, calculate the corresponding theta angle by inverse formula (Real Behavior Table 4.1). Once the angle is obtained, one can calculate the ideal distance that ray would have in an image without distortion. (Ideal Behavior Table 4.1). For each vertex, calculate the difference between its radius and the ideal radius and shift its position by a magnitude equal to the difference of the two radii.

The following images it's possible to notice the difference between a UV map before the correction and the one after the correction.
The new map turns out to be consistent with the formulas obtained from the calibration in fact, the area where it is most modified is the area near the margin.

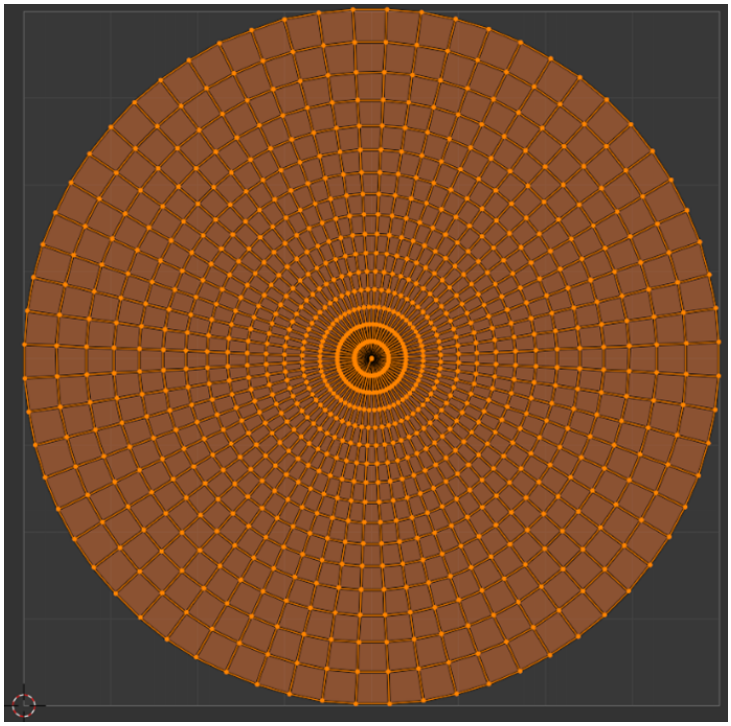

Figure 4.12: UV map before Correction

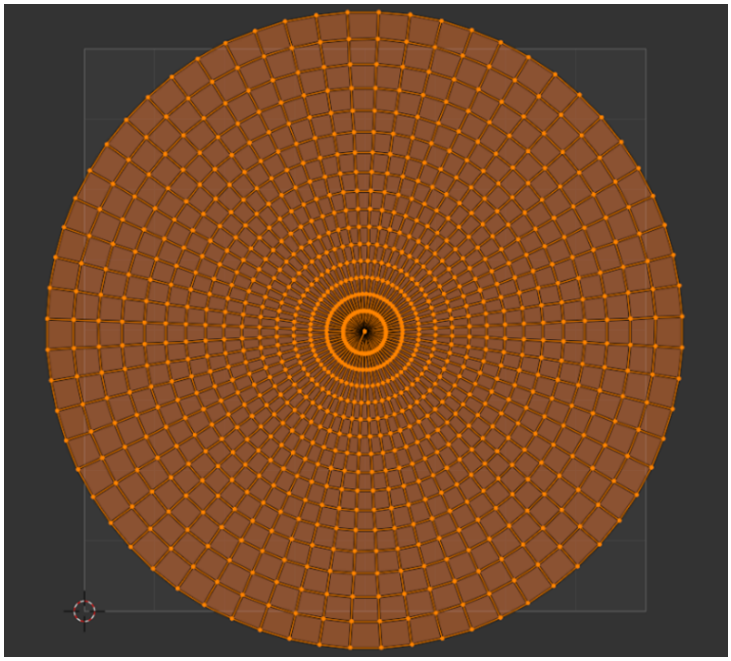

Figure 4.13: UV map after Correction

Chapter 5

Unity Application

One of the key parts of this robotic system is the virtual experience that the user can live using the HMD. It was decided to program the graphical user interface (GUI) on Unity, which offers various tools and packages to facilitate the creation of virtual environments. Unity is a game and application development platform, highly versatile and powerful. Originally designed for game development, Unity has become an essential tool for many other industries, including robotics. In recent years, Unity has also found application in the field of robotics because of its powerful simulation and visualization capabilities. It enables the creation of realistic simulated environments where robots can be tested before being implemented in the real world, proving particularly useful for designing, testing, and optimizing robot control algorithms. Unity is used to develop interactive user interfaces that allow human operators to control and monitor robots on different platforms, including PCs, mobile devices, and VR viewers. In addition, Unity can display sensory data in real-time, and this can be very useful if you want to connect robots with information that Unity can compute [20].

This software can be integrated with the Robot Operating System (ROS), a flexible and open-source platform for robot software development, creating powerful simulation and control tools for robots.

For this application, Unity is mainly used for 2 different purposes:

- **Streaming video**
- **ROS connection for head movement and steering robot**

5.1 Streaming Environment

In Unity, a scene has been created in which two mathematically modeled surfaces are placed for visual distortion correction. Each of these surfaces is used for the projection of images captured by the robot cameras. Specifically, one of the surfaces is dedicated to the video from the right camera, while the other is dedicated to the left camera. To make the experience more immersive, the two surfaces do not remain fixed, but follow the rotation of the headset. In this way, when the user rotates the head, the surfaces follow this movement, always allowing high perception of the visual environment. For safety reasons, no image is transmitted at the bottom of the projection surfaces, as this could lead to a state of confusion and a feeling of instability when walking on the treadmill. To avoid this, it was decided that in the lower part the user can view the real outdoor environment.

To make the application as usable as possible, a download system for projection surfaces was integrated. To achieve this, AssetBundle was used. This solution allows projection surfaces to be dynamically loaded and managed, optimizing resources and improving the efficiency of the application. AssetBundle allows downloading only the data needed at the time of need, reducing the initial loading time and improving the user experience [21].

After modeling the hemispherical domes on Blender and using the algorithm for correcting the UV map, it is possible to include this surface in the Unity scene. In fact, on Blender the mesh is saved as file.fbx, so that the changes are retained. This file is then inserted into the scene.

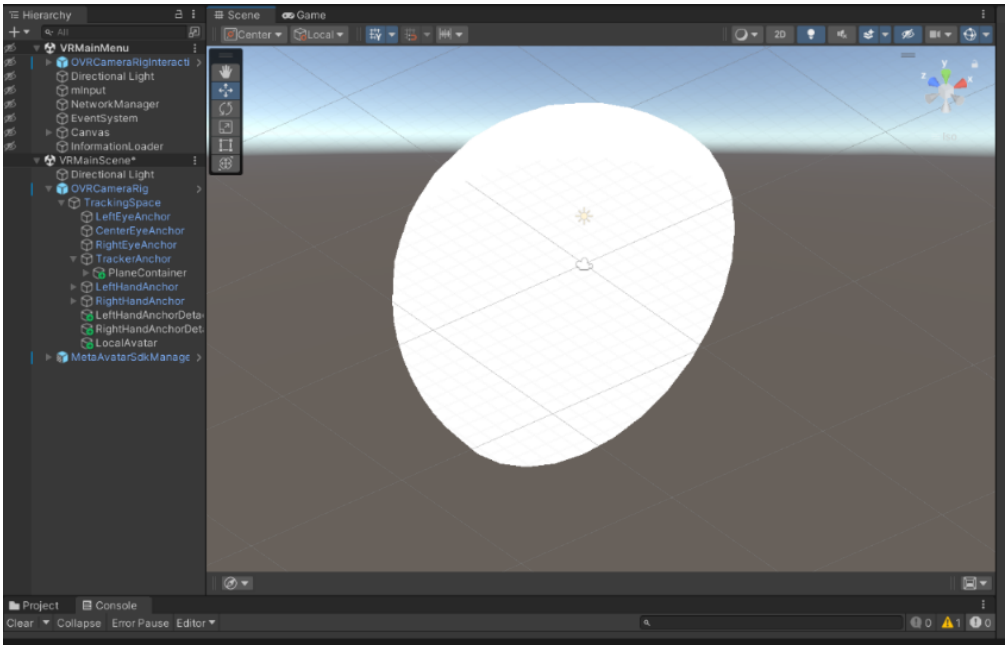

Figure 5.1: Unity Dome

5.2 Data Connection

In the Unity application, all data related to the user's head movement, the robot's rotation, and its linear speed are managed. This approach leverages Unity's ability to unify all control parameters (such as speed and rotation) within a single virtual environment, simplifying the coordination between the VR headset, joystick, and robot movements. By centralizing these controls, there's no need to handle separate data streams, making the overall system more cohesive and easier to manage.

The data managed are:

- **Robot Rotation Angle**: The user can control the robot rotation by means of a joystick, particularly by moving the left joystick thumbstick. Controller tracking is a feature that allows controllers to be used to interact with the virtual environment, perform gestures, and manipulate objects within a scene.

  Joystick_angle = OVRInput.Get(OVRInput.Axis2D.PrimaryThumbstick)[0]

- **Head Rotation**: Unity provides access to data regarding the rotation of the VR, and specifically for this application, the rotation around the y-axis is the one to be considered.

  By doing so, head tracking can be obtained, and this data is then used to rotate the robot's visual system. In this way, the communication between the user's head movements and the robot's visual system can be performed. The servo used for the robot's visual system is a Dynamixel.

  Head_angle = Camera.main.transform.eulerAngles.y

Figure 5.2: Meta Quest Pro Controller

- **Robot Linear Velocity**: On Unity, in addition to having tracking data taken from VR, data regarding the linear velocity of the robot is handled. This data is sent from the treadmill to the Unity application via a WebSocket protocol.

A `C#` script has been implemented in Unity to send all robot-related data to the Fetch Computer, specifically the **Joystick_angle** and **Robot_Linear_Velocity** to the **cmd_vel** topic, and the **Head_angle** to the **dynamixel** topic in ROS.

Communication between Unity and ROS is achieved through the use of a Rosbridge Server, which enables bidirectional communication between ROS and WebSocket. WebSocket is a communication protocol that provides a low-latency channel for transmitting data between a client (such as a web browser or a Unity application) and a server. Unlike traditional HTTP communications, which require a separate request for each message, WebSocket allows for a persistent connection, enabling continuous data exchange without the need for repetitive requests and responses. This is particularly useful for real-time applications, such as robot control, where latency is a critical factor.

In the implemented system, the client is represented by the Unity application, which sends data via WebSocket to the ROSbridge server, which is responsible for translating and routing the messages to the appropriate ROS topics. The messages sent are of JSON

type, a lightweight and easily readable data exchange format, which is used to transmit the information collected in Unity (such as joystick angle, robot linear velocity, and head angle) to the ROSbridge server.

When Unity sends the data, it is first converted into JSON format by the C# script, and then sent via WebSocket to the ROSbridge server. Once received, the JSON messages are decoded by the server, which publishes them to the respective ROS topics:

1. **cmd_vel topic:** This topic is used to send movement commands to the robot, such as linear velocity vector and angular velocity vector.

2. **dynamixel topic:** This topic is used to send commands to the robot's Dynamixel motors, particularly to control the orientation of the robot's head (and thus the head angle, which could be related to the position of the robot's camera system).

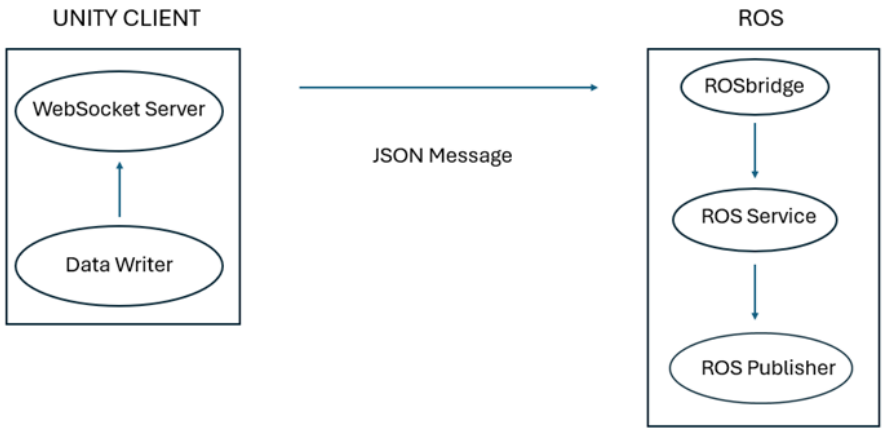

Figure 5.3: Connection between Unity and ROS

Chapter 6

Experimental Results

This chapter reviews the results obtained from the tests and experiments conducted to evaluate the effectiveness of the visual system. To adequately examine both the visual system and the overall use of the system, two experimental modes were designed.
The main focus of the former is on understanding the projected environment and depth perception. The latter is tested through an experiment aimed at identifying different distances between various objects in the same scene. In the second experiment, the focus is on the overall use of the overall system, evaluating aspects such as safety, reliability and user experience.

6.1 Assessment of depth and relative distance

In this experiment, the effectiveness of the new visual system in providing excellent depth perception is evaluated. In order to evaluate this feature, the treadmill subsystem is not connected, and the user, wearing VR, must evaluate the features of a scene viewed by the robot avatar's visual system. The experiment is conducted on 5 students of the SMART ROBOTICS DESIGN LAB, four of them had an experience with a VR device while the others did not, so before performing the test, a test of interaction between the students and the VR was carried out. In the scene, 3 objects were placed in front of the robot avatar at different distances from each other. The ultimate goal of the test is to evaluate the accuracy with which distances between objects and those with the robot avatar are identified. The 5 participants tested the old visual system, characterized by a 360 FOV theta camera, and the new system. The experiment was run twice so that they could have a more accurate analysis of the data. In order to make the distance recognition process easier and more intuitive, a 1m bar was placed on the scene.

In particular, the following information should be evaluated:

1. Distances between robots and objects

2. Order of the sequence of objects from nearest to farthest

The next pages report the values of the experiments performed with an analysis of the final results.

Experiment 1 Ricoh Theta 360 FOV	Ground Truth	Mohammed	Murase	Ricardo	Gustavo	Tadano
Tall box	1.3	0.9	1.5	1.6	1.7	1.6
Small box	1.9	1.5	2.4	2.2	2	2.3
Alcohol	2.6	2.1	3.1	3	3	3
Order	T,S,A	T,S,A	T,S,A	T,S,A	T,S,A	T,S,A

Figure 6.1: Distances values express in meter

Mean error values	Mohammed	Murase	Ricardo	Gustavo	Tadano
Tall box	0.4	0.2	0.3	0.4	0.3
Small box	0.4	0.5	0.3	0.1	0.4
Alcohol	0.5	0.5	0.4	0.4	0.4

Figure 6.2: Mean error values Experiment 1 Ricoh Theta

Mean error values on objects	
Tall box	0.32
Small box	0.34
Alcohol	0.44

Figure 6.3: Mean error values on objects Experiment 1 Ricoh Theta

Mean error values on the participants	Mohammed	Murase	Ricardo	Gustavo	Tadano
	0.4	0.4	0.3	0.3	0.4

Figure 6.4: Mean error values on the participants Experiment 1 Ricoh Theta

Experiment 1 Sony IMX 200 FOV	Ground Truth	Mohammed	Murase	Ricardo	Gustavo	Tadano
Tall box	1.3	1.1	1	1.2	1.3	0.9
Small box	1.9	2	1.8	2	1.8	2.1
Alcohol	2.6	2.3	2.4	2.4	2.8	2.4
Order	T,S,A	T,S,A	T,S,A	T,S,A	T,S,A	T,S,A

Figure 6.5: Distances values express in meter

Mean error values	Mohammed	Murase	Ricardo	Gustavo	Tadano
Tall box	0.2	0.3	0.1	0	0.4
Small box	0.1	0.1	0.1	0.1	0.2
Alcohol	0.3	0.2	0.2	0.2	0.2

Figure 6.6: Mean error values Experiment 1 Sony IMX

Mean error values on objects	
Tall box	0.2
Small box	0.12
Alcohol	0.22

Figure 6.7: Mean error values on objects Experiment 1 Sony IMX

Mean error values on the participants	Mohammed	Murase	Ricardo	Gustavo	Tadano
	0.2	0.2	0.1	0.1	0.3

Figure 6.8: Mean error values on the participants Experiment 1 Sony IMX

Experiment 2 Ricoh Theta 360 FOV	Ground Truth	Mohammed	Murase	Ricardo	Gustavo	Tadano
Tall box	1.9	1.6	1.5	2.4	2.2	2
Small box	2.4	2	1.9	2.7	2.8	3
Alcohol	1.3	1.1	1	1	1.5	1
Order	A,T,S	A,T,S	A,T,S	A,T,S	A,T,S	A,T,S

Figure 6.9: Distances values express in meter

Mean error values	Mohammed	Murase	Ricardo	Gustavo	Tadano
Tall box	0.3	0.4	0.5	0.3	0.1
Small box	0.4	0.5	0.3	0.4	0.6
Alcohol	0.2	0.3	0.3	0.2	0.3

Figure 6.10: Mean error values Experiment 2 Ricoh Theta

Mean error values on objects	
Tall box	0.32
Small box	0.44
Alcohol	0.26

Figure 6.11: Mean error values on objects Experiment 2 Ricoh Theta

Mean error values on the participants	Mohammed	Murase	Ricardo	Gustavo	Tadano
	0.3	0.4	0.4	0.3	0.3

Figure 6.12: Mean error values on the participants Experiment 2 Ricoh Theta

Experiment 2 Sony IMX 200 FOV	Ground Truth	Mohammed	Murase	Ricardo	Gustavo	Tadano
Tall box	1.9	2.1	2	1.9	1.9	1.8
Small box	2.4	2.6	2.7	2.1	2.5	2.5
Alcohol	1.3	1.4	1.3	1.2	1.2	1.3
Order	A,T,S	A,T,S	A,T,S	A,T,S	A,T,S	A,T,S

Figure 6.13: Distances values express in meter

Mean error values	Mohammed	Murase	Ricardo	Gustavo	Tadano
Tall box	0.2	0.1	0	0	0.1
Small box	0.2	0.3	0.3	0.1	0.1
Alcohol	0.1	0	0.1	0.1	0

Figure 6.14: Mean error values Experiment 2 Sony IMX

Mean error values on objects	
Tall box	0.08
Small box	0.2
Alcohol	0.06

Figure 6.15: Mean error values on objects Experiment 2 Sony IMX

Mean error values on the participants	Mohammed	Murase	Ricardo	Gustavo	Tadano
	0.2	0.1	0.1	0.1	0.1

Figure 6.16: Mean error values on the participants Experiment 2 Sony IMX

In each experiment, the distance values given by the five participants are described. From these values, the average error for each object and participant can be calculated in order to analyze the characteristics of each system.

From the data obtained, it can be seen that using both systems there is no problem in detecting the correct sequence of objects, this indicates that neither has an insufficient degree of perception.

Analyzing the average error on objects, for both systems, the error for the most distant objects is greater than that of the nearest object, this is an indication of how the accuracy of the system goes down as the distance from the visual system increases, this phenomena is also common with human visual perception.

All data regarding the errors of the old system are higher in absolute value than those of the new system, consequently depth perception is improved with the latter, it is important to note how in the second experiment the data improve for both systems, this is due to the fact that the participants have developed more familiarity with the system.

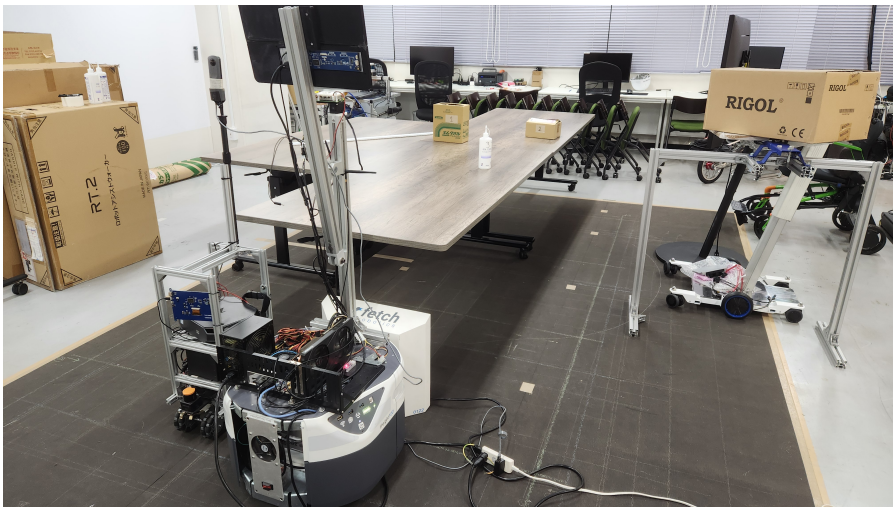

Figure 6.17: Distances values express in meter

6.2 Overview on the safety and sense of use of the system

In the second experiment, the system as a whole is evaluated, as both the treadmill and robot avatar subsystems are now used, and these are connected to each other. At this stage, the "sense of use" of the system is analyzed, which refers to the overall perception of the user during interaction. This includes aspects such as comfort, reliability, and responsiveness of the device. The evaluation is based on how easy the device is to use, how predictably it responds, and how well it manages to provide an immersive feeling to the user.

In order to evaluate the sense of use of the system, the following factors are defined:

- **Reliability and Safety**
- **Immersion and Realism**
- **Mental Exhaustion**
- **Satisfaction**

The experiment was conducted on 10 students, all of them had no previous experience with the system. Before starting the actual test, a test session was conducted to allow users to gain familiarity and confidence with the device.
The experiment consists of using the system within the SMART ROBOT DESIGN laboratory, thus going to simulate an indoor navigation environment typical of a company office. Users should be able to move freely within the laboratory, choosing their own path to follow. In this way, they will not feel obliged to follow a predefined trajectory, but will be able to explore freely, based on their own sense of using the robot.

Answers
Excellent
Very good
Sufficient
Insufficient
Disastrous

Figure 6.18: Answers Table

The following are the definitions of the parameters and their respective results.

Reliability and Safety

Students were asked to evaluate the entire system in terms of reliability and safety, focusing on the device's ability to function consistently and without malfunction. This criterion is crucial to the research, as the system is primarily intended for use by elderly individuals or those with motor impairments. Consequently, ensuring a strong sense of safety is one of the system's most important features.

Below are the test results, with the reported ratings on the horizontal axis and the number of students on the vertical axis.

Figure 6.19: Answers Table Reliability and Safety

The data obtained indicate how in general the system is considered safe enough, the negative valuations are due to the feeling that the support for the upper body and arms is not stable enough, especially if the end-user is an elderly person with motor problems.

Immersion and Realism

The immersion and realism criterion assesses how well the system succeeds in engaging the user in a unique and immersive experience. The realism of the experience is characterized by the two main activities of use, the first is the sense of reality in walking on the automatic treadmill, the second of the experience of viewing the projection of images captured by the robot's visual system.

By achieving as realistic an experience as possible, the feeling of distance between user and device can be reduced and thus make the overall experience more comfortable.

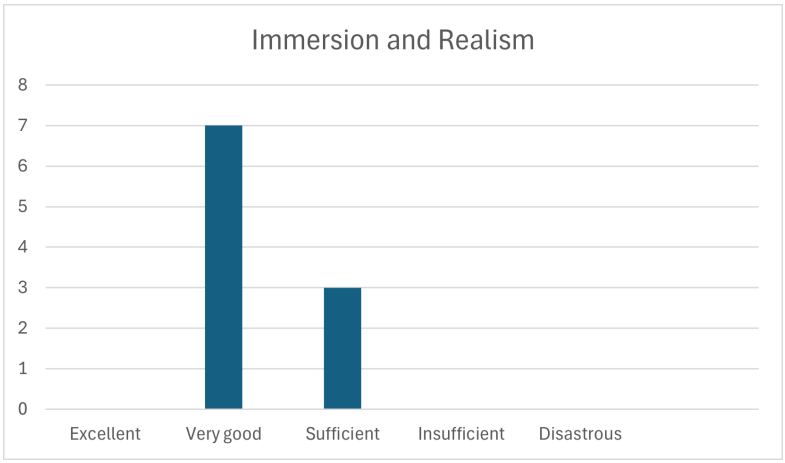

Figure 6.20: Answers Table Immersion and Realism

The data reported on the immersion of the experience are very positive, especially the feedback related to the visual system and the images viewed in VR indicate how the system is very efficient.
One consideration for improvement was regarding the delay between the movement of the 'robot avatar and the viewed images.

Mental Exhaustion

The mental fatigue factor in a system that combines VR and a treadmill, analyzes the level of cognitive effort required by the user to interact with the virtual environment while walking. This aspect is particularly important as the user must simultaneously manage various physical and cognitive stimuli: walking on the treadmill, controlling the robot avatar, and interacting with the virtual environment.

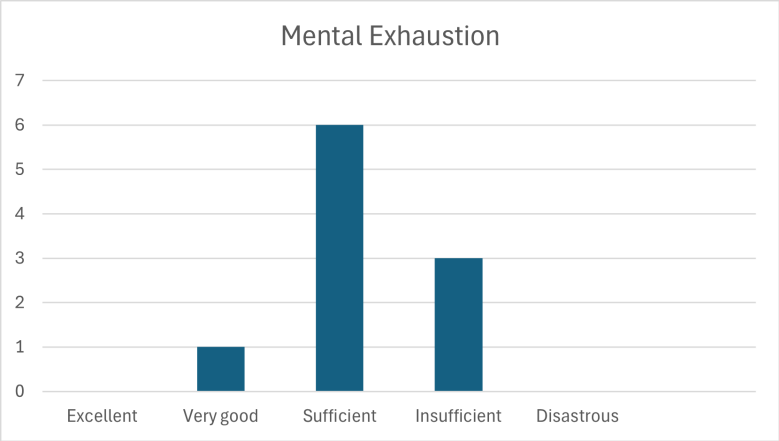

Figure 6.21: Answers Table Mental Exhaustion

Regarding the mental fatigue caused by using the system surely there must be room for improvement, participants stated that however it is a bit difficult to use the whole system, because connecting both walking on the treadmill and viewing the external environment in which to move requires no small amount of mental effort.

Satisfaction

The satisfaction criterion analyzes the degree of appreciation and well-being the user experiences during and after interacting with the device. It is based on the overall perception of the user experience, including aspects such as enjoyment, comfort, and the effectiveness of the device in meeting the user's expectations. A high level of satisfaction implies that the user feels motivated to continue using the system, having positive experience in terms of performance, ease of use, and overall enjoyment.

Figure 6.22: Answers Table Satisfaction

In this last graph it is possible to see that all the participants the system is very fascinating especially because of the stimuli it causes, in particular all of them reported that walking on a treadmill in this way is much more enthusiastic than walking on a traditional treadmill. In fact, during walking one still has to control more devices such as the visual system and the rotation angle of the robot, this on the one hand is very stimulating but on the other hand ,as reported before, it causes more fatigue.

Chapter 7

Conclusions

This thesis describes the development process of a stereoscopic vision system for a robotic avatar, a central component of a comprehensive robotic system that includes the automatic treadmill and the avatar itself. This vision system represents a significant step forward in creating a more immersive and realistic interaction between the user and the virtual or remote environment in which the robot operates.

The implementation of the stereoscopic vision system has significantly improved depth perception in images, enhancing the realism and overall engagement of the experience. Initially, the images were captured in a fisheye format with a 200° FOV, which introduced strong geometric distortions. To address these distortions, the images were projected onto a hemispherical dome. This solution not only enveloped the user in an immersive visual environment but also corrected many of the typical distortions of the fisheye format, thereby improving the overall visual quality.

However, to fully eliminate unwanted compression effects caused by the non-ideal characteristics of the cameras, a mathematical model was developed through an accurate experimental process. This model identified and quantified the intrinsic characteristics of the cameras used, providing the foundation for further optimizations.

Using the collected data, an algorithm was created to adjust the UV map and fix the

remaining distortions in the hemispherical dome. This adjustment made the images more accurate and free of unwanted compression effects, improving the overall visual quality of the system.

An additional improvement to the sense of realism was achieved by implementing a system that allows the visual component to follow the user's head movements. This functionality enables the user to gain a complete view of the environment surrounding the robot, significantly reducing feelings of disorientation while using the treadmill.

The goal of developing this vision system was to make the robot's behavior and abilities more like those of humans. Stereoscopic vision was chosen to better understand how humans perceive depth and distance. Instead of viewing a single image to deduce environmental features, the user perceives two distinct images, one for the left eye and one for the right. This setup mimics human binocular vision, letting the user explore the environment as if they were seeing through the robot's eyes.

Regarding the overall system's safety, this aspect plays a crucial role as the system is primarily designed for elderly users and individuals with motor problems. Safety is not just a technical requirement but also a fundamental factor for the trust of the users, many of them might initially hesitate to use an advanced technological device. In this context, the support device for the arms is a key component of the system. Currently, this device moves freely in space, which could lead to instability or difficulties for the user.

A possible way to improve safety is by adding guides to the wheels of the support device, so it can only move along the treadmill's axis. This change would reduce the risk of unwanted movement and make the system more stable, offering better support for users with motor difficulties.

Bibliography

[1] Marie H Murphy, Elaine M Murtagh, and Janne Boone-Heinonen. Walking – the first steps in cardiovascular disease prevention. *Journal of Cardiovascular Health*, 2011.

[2] Sunghoon Kim, Weisheng Chiu, Heetae Cho, and John Chee Keng Wang. Increasing exercise participation during the covid-19 pandemic: the buffering role of nostalgia. *Journal of Sports Science*, 2022.

[3] Fabien Beaumont, Guillaume Polidori, Fabien D Legrand, and Philippe Jeandet. Effects of outdoor walking on positive and negative affect: Nature contact makes a big difference. *Nature and Health Journal*, 2022.

[4] Wikipedia. Wikipedia treadmill. Accessed: 2024-11-19.

[5] Askar Garad, Ahmed Al-Ansi, Abdullah M Al-Ansi, and Mohammed Jaboob. Analyzing augmented reality (ar) and virtual reality (vr) recent development in education. *Educational Technology Review*, 2020.

[6] N Brandín-De la Cruz. Immersive virtual reality and antigravity treadmill training for gait rehabilitation in parkinson's disease: a pilot and feasibility study. *Rehabilitation and Movement Science*, 2021.

[7] P Kengkij. Remote controlling avatar robot using autonomous systems. *Robotics and Automation Systems*, 2020.

[8] Zebra Technologies. Zebra robotics automation. Accessed: 2024-11-19.

[9] Patterson R and Martin WL. Human stereopsis. *Vision Science Journal*, 2018.

[10] L Benhaim-Sitbon. Binocular fusion disorders impair basic visual processing. *Journal of Visual Perception*, 2020.

[11] K Ikeuchi. Computer vision. *International Journal of Computer Vision*, 2019.

[12] Sony. Sony imx219pq datasheet. Accessed: 2024-11-19.

[13] M L Bauer. Fish-eye lens designs and their relative performance. *Optical Engineering Journal*, 2017.

[14] Martin Glavin. Accuracy of fish-eye lens models. *Optics Journal*, 2016.

[15] P Bourke. Classification of fisheye lenses. *Lens Technology Journal*, 2015.

[16] P Bourke. Nonlinear lens distortion. *Applied Optics Journal*, 2017.

[17] P Bourke. Fisheye lens correction. Accessed: 2024-11-19.

[18] Wikipedia. Blender (programma). Accessed: 2024-11-19.

[19] blender.org. Blender api documentation. Accessed: 2024-11-19.

[20] Wikipedia. Unity (motore grafico). Accessed: 2024-11-19.

[21] unity3d.com. Unity asset bundles documentation. Accessed: 2024-11-19.

List of Figures