Employment of Artificial Intelligence Mechanisms for Mixed Reality Systems with Motion Controllers to Support Accessibility and Usability

Carlos Filipe Paiva da Cunha Macieira Antunes

Thesis to obtain the Master of Science Degree in

Information and Enterprise Systems

Supervisors: Prof. Leonel Caseiro Morgado

Prof. Carlos Eduardo Dias Coutinho

Examination Committee

Chairperson:

Supervisor:

Members of the Committee:

July 2022

" You will face many defeats in life, but never let yourself be defeated."

Hindi Proverb

Acknowledgments

There are so many people to whom I wish to express my gratitude, because of their dedication and commitment that there are not enough words to express all the help provided during the development of this dissertation.

To my brother, my niece and my nephew, I would like to wish them all the success in their lives, so they can find true happiness in whichever path they wish to follow.

Concerning my father, I do not have enough compliments to describe is unconditional and constant help during all the problems I have been through, especially during my disease, without him, I would not be here presenting this dissertation and also for all the guidance he provided that came from his amazing soul. He keeps helping me all the time, which is something amazing, his strong knowledge on electronics engineering helped me in guiding this dissertation concerning several areas like connecting web systems with drones.

My greetings to the professors Carlos Coutinho and Leonel Morgado for giving me the opportunity to develop this project and also for their orientations in order to adjust to the necessary requirements and objectives.

My gratitude to the professor Miguel Mira da Silva for clarifying my doubts during the outgoing of this master, which was an amazing help in order to achieve the wanted results.

My special thanks for my aunt that provided an incredible help concerning the treatment of health care information.

To Xerox, for providing unpaid leaves which allowed me to invest more time and commitment to the dissertation and also for presenting CareAR, an application that adds virtual elements on chat facilitating daily supporting activities on printer troubleshooting issues.

To my, colleagues and teammates, I would also like to show my appreciation for all the moments and experiences we had during our team projects, improving my skills in teamwork and leadership.

Also, would like to show my gratitude to the several doctors that attended me during my disease, reducing the level of tinnitus making possible for me to return to the usual daily activies, which I could not for a long period of time.

Thanks IST, ISCTE and UAb for entering my live and providing all this knowledge and experience.

Resumo

Atualmente, as tecnologias de virtualização estão a mesclar o mundo físico com o digital em várias áreas, como a medicina e a educação. Os elementos virtuais podem ter mecanismos de inteligência artificial mais responsivos na interação com o utilizador.

No quotidiano, as aplicações de realidade mista possuem alguns problemas, como aumentar corretamente durante o movimento, colisão dos elementos virtuais com objetos do mundo real, inteligência artificial responsiva limitada e não são fáceis de serem usadas por pessoas com deficiências visuais, auditivas, físicas, fala, cognitivas, motoras, linguagem, aprendizagem e neurológicas. Existe necessidade de criar novas funcionalidades nas aplicações virtuais para responder a carências humanas e sociais tendo por base a fundamentação epistemológica.

Esta dissertação demonstra protótipos de realidade mista com novas funcionalidades quando comparado com as tradicionais defendendo um conceito de interação virtual permitindo responder a várias necessidades como médicas, adaptação ao mundo real, acessibilidade a pessoas com determinadas incapacidades e melhorar processos de aprendizagem.

Ao movimentar a câmera do telefone, do computador e de um drone, o software irá processar as informações fornecidas pelo ambiente e utilizará algoritmos de inteligência artificial para posicionar os elementos virtuais no ecrã, calcular as distâncias e interagir com o utilizador de forma dinâmica respondendo aos movimentos realizados.

De forma a garantir a validade dos métodos, foram efetuados testes, que permitem analisar o comportamento dos utilizadores e registrar o que acontece nos protótipos para fins demonstrativos e obter conclusões sobre formas de melhorar as funcionalidades dos sistemas.

Palavras-Chave: Realidade Mista, Inteligência Artificial, Engenharia de Software, Controladores de Movimento, Drones e Visão de Computador.

Abstract

Currently, virtualization technologies are merging the physical world with the digital in various areas, like medicine and education. Virtual elements can have artificial intelligence mechanisms that are more responsive concerning user interaction.

Nowadays, mixed reality applications have some problems, such as correctly zooming during movement, collision of virtual elements with real world objects, limited responsive artificial intelligence and not easy to be used by people with visual, hearing, physical, speech, cognitive, motor, language, learning and neurological.

There is a need to create new functionalities in virtual applications to respond to human and social needs based on epistemological foundations.

This dissertation demonstrates mixed reality prototypes with new features when compared to traditional ones, defending a concept of virtual interaction allowing to respond to various needs such as medical, adaptation to the real world, accessibility to people with certain disabilities and improving learning processes.

By moving the camera of the phone, the computer and a drone, the software will process the information provided by the environment and will use artificial intelligence algorithms to position the virtual elements on the screen, calculate the distances and interact with the user dynamically, responding to the movements performed.

In order to guarantee the validity of the methods, tests were carried out, which allow analyzing the behavior of users and recording what happens in the prototypes for demonstrative purposes and obtaining conclusions on ways to improve the functionalities of the systems.

Keywords: Mixed Reality, Artificial Intelligence, Software Engineering, Motion Controllers, Drones and Computer Vision.

Table of Contents

List of Figures

List of Tables

Abbreviations

AI **A**rtificial **I**ntelligence

AIML **A**rtificial **I**ntelligence **M**ark Up **L**anguage

AE **A**fter **E**ffects

API **A**pplication **P**rogramming **I**nterface

AR **A**ugmented **R**eality

CGI **C**omputer **G**enerated **I**mages

DSR **D**esign **S**cience **R**esearch

DSRM **D**esign **S**cience **R**esearch **M**odel

FPS **F**irst **P**erson **S**hooter

FR **F**unctional **R**equirements

GUI **G**raphical **U**ser **I**nterface

IR **I**nfra**r**ed

HTML **H**yper**t**ext **M**arkup **L**anguage

HTTP **H**yper **T**ext **T**ransfer **P**rotocol

HTTPS **H**yper **T**ext **T**ransfer **P**rotocol **S**ecure

HUD **H**eads **U**p **D**isplay

IDE **I**nterface **D**evelopment **E**nvironment

JSON **J**ava**S**cript **O**bject **N**otation

MAS **M**ulti **A**gent **S**ystems

MR **M**ixed **R**eality

ML **M**achine **L**earning

NLP **N**atural **L**anguage **P**rocessing

NoSQL **N**ot **o**nly **S**tructured **Q**uery **L**anguage

PPG Photoplethysmography

REST **R**epresentational **S**tate **T**ransfer

ROI **R**egion **o**f **I**nterest

SLAM **S**imultaneous **L**ocalization **A**nd **M**apping

SOA **S**ervice-**O**riented **A**rchitecture

SOAP Simple **O**bject **A**ccess **P**rotocol

TPS **T**hird **P**erson **S**hooter

UAV **U**nmanned **A**erial **V**ehicle

UDDI **U**niversal **D**escription, **D**iscovery, and **I**ntegration

UI User **I**nterface

VR **V**irtual **R**eality

VFX Visual Effects

WSDL **W**eb **S**ervices **D**escription **L**anguage

XML E**x**tensible **M**arkup **L**anguage

Chapter 1

1

1. Introduction

This chapter introduces the research preformed being divided in six main sections. First, will mention the context and motivation of the dissertation, second will show a detailed definition of the problem. The objectives appear as third and the research questions at fourth. The final part will show the research methodologies used and an outline section where an explanation of the document structure is showed mentioning an overview of the upcoming chapters.

1.1 Context and Motivation

New technologies have been emerging and appearing in society like Augmented Reality (AR), Virtual Reality (VR), Mixed Reality (MR), Internet of Things (IoT) changing the behavior of the user (Hoyer et al., 2020); these new technologies work in real time and may use specific artificial intelligence (AI) mechanisms, which improves the usability that might vary depending on the context (Kim, 2019). In order to make the learners more interested in the subjects they are studying, several approaches may be considered like the creation of multimedia contents (Jonassen, 2000; Jenkins, 2007; Carmigniani et al., 2010). There is a necessity to improve the behavior of artificial intelligent agents for mixed reality applications to make the applications more efficient and to provide more functionalities, retrieving important information from the surrounding environment like positioning, distances and the Photoplethysmography (PPG) signal allowing to implement several mechanisms, like collisions between virtual elements and the real world, motion controllers and vital signs of the user. With the objective to improve the usability and accessibility of the softwares for patients with disabilities like visual, auditory, physical, speech, cognitive, language, learning and neurological, this study focuses on the movement of 3 body parts: hand motion (Guo et al., 2011), eye gaze (Antunes, 2017) and general body (including arms, legs, hands, feet and head).

The usage of motion controllers can facilitate the interaction for different type of persons, which can help people with different type of disabilities, making the applications in some cases easier to use.

Virtual Reality applications have the ability of providing the sensation of immersion, usually the VR glasses combine a process of blending two images, one for each eye allowing for the user to see a 3D virtual world (Yalon-Chamovitz & Weiss, 2008). On a perspective of accessibility, can help for some cognitive, auditory and physical disabilities.

Drones are used nowadays for several purposes like getting geographical images, disaster response, operations of rescue, healthcare and agriculture.

The mixed reality applications can have more tasks to achieve objectives in several areas like, physiotherapy, medical (Le et al., 2010), unmanned aerial vehicles (UAVs), formation among others and should also provide the possibility to be used by people with different type of disabilities.

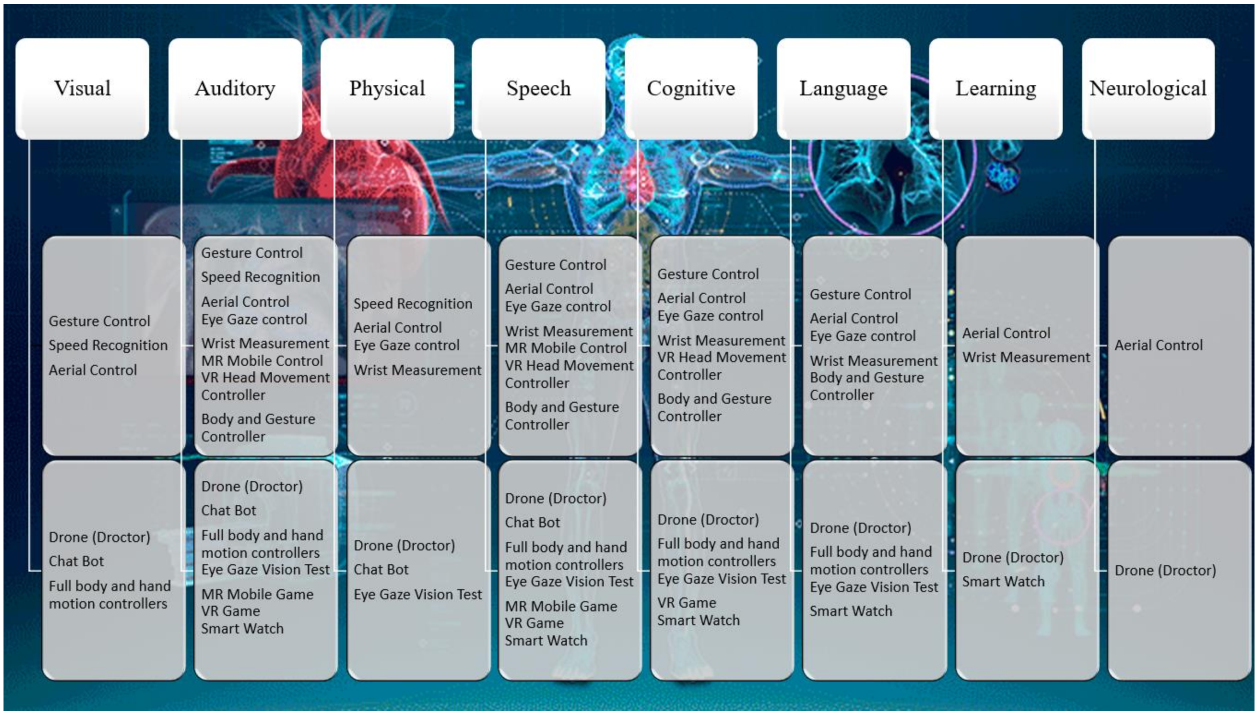


Figure 1. Common disabilities and technologies used to reduce the usability gap, plus the applications developed that can be used for the mentioned disorders depending on the degree of incapacity.

1.2 Problem Definition

Nowadays mixed reality applications have some problems, like correctly augmenting during motion (Babu et al., 2018), collision of the virtual elements with real world objects (Breen et al., 1995), limited responsive AI (Acemoglu, & Restrepo, 2020; Ghafghazi et al., 2021; Kim & Song, 2022) and are not easy to be used by people with visual, auditory, physical (Jia et al. 2007), speech, cognitive, motor (Chin et al., 2018) language, learning and neurological disabilities (Trewin, 2018).

The employment of artificial intelligence in virtual elements can be enhanced and provide more useful functionalities than the current ones; the interaction with the user in software engineering terms can be improved. Many gaps can be found in MR systems like the collision of virtual elements with the real-world environment objects (Breen et al., 1995) or the missing of responsive AI in some cases, or a correct and more functional treatment of the information provided by the environment.

Concerning the problem of accessibility and usability for users of different ages and disabilities, several applications were developed taking in consideration the degrees of handicap.

The developed “Droctor” system, a web hand gesture drone controller (Sarkar et al, 2016) is able to be used by people with hearing and low visual capacity. The objective of this system is to allow a doctor to control by hand motion via web, a drone that captures the vital signs data of the patient without contact and without the need of the user to perform any actions.

Concerning visual, auditory, speech, language and some motor and cognitive incapacities, a motion controller of the movement (Yashoda et al., 2018) of several body parts like hands, fingers, arms, legs, head, body and feet was implemented in several applications.

For people with visual incapacities a chat bot for speech recognition was tested (Weeratunga et al., 2015).

Users with movement and earing problems can use the Aerial Control and the Eye Gaze control.

The iris movement controller allows to resolve the problem of usability for people with certain levels of auditory, physical, speech, cognitive, language, learning and neurological incapacity.

This work aims to a scientific approach based on the combination of technologies to solve issues that occur every day, looking for an automation of the reply from the information system by improving the AI response mechanisms to the inputs from the user.

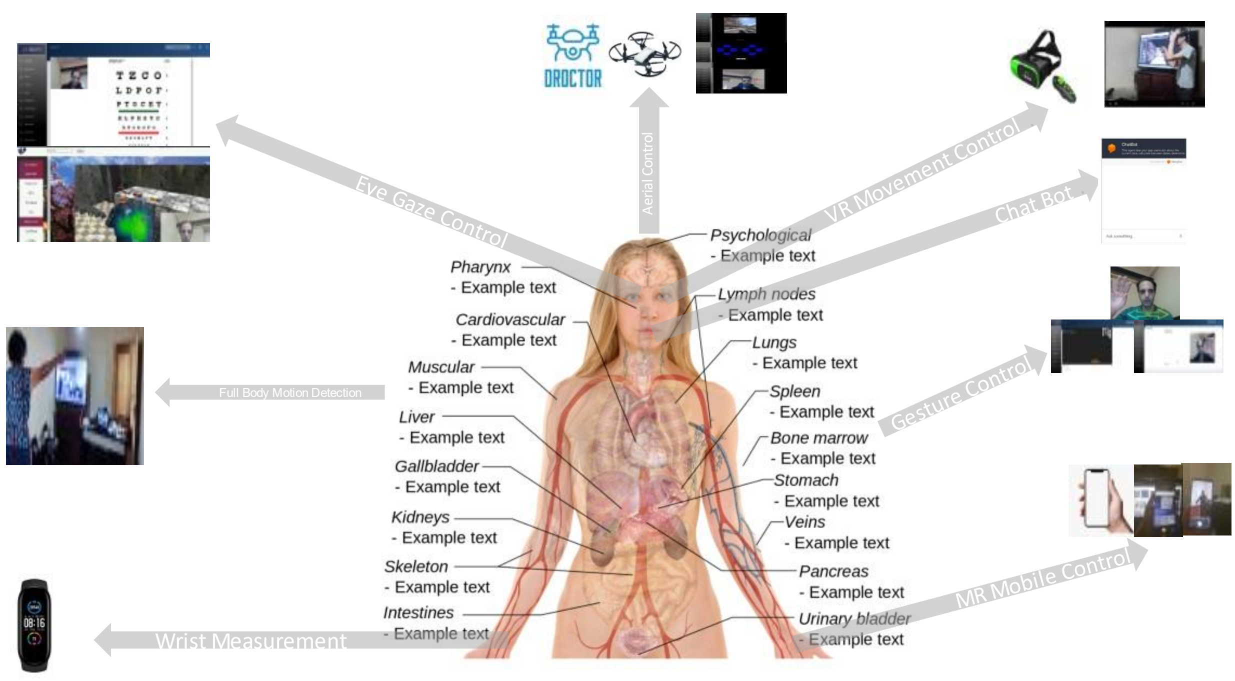


Figure 2. Developed Systems for different types of disabilities, some with motion control for interaction with some of the human body parts.

The objective is to find an innovative concept as a solution for this situation considering the engineering aspects such as using inputs, AR elements and AI mechanisms in order to produce an output.

Zhao et al. (2018) looked for a solution of controlling drones movement using motion detectors, increasing the interactivity of the system and allowing the use for people with visual and certain motor problems, MR systems can interact with motion controllers in order to facilitate the manipulation of AI agents.

The detection of occlusions is not an easy process, because scenarios of low lightning do not provide enough information, making difficult to determine where the virtual objects should be placed in the final screen (Walton & Steed, 2017).

Several strategies were followed like a Mixed Reality (MR) application for mobile phones that allow interaction with AI augmented agents that respond to the movements of the phone and have collisions with obstacles of the real world. A Virtual Reality (VR) application was also developed.

To the best of my knowledge, currently there is not a system like a MR First Person Shooter (FPS) with shooting enemies. Therefore, a Systematic Literature Review was conducted in order to identify the technical limitations of this problem (Wang, 2021).

Motion controllers are needed for people with certain types of disabilities in order to perform different type of tasks, like eye gaze for people with motor disabilities (Sunny et al., 2021) or hand controllers for users that have visual problems.

1.3 Objectives

The research in cause aims to find solutions to the problems related with traditional MR applications, developing responsive AI mechanisms and also use of motion controllers (Wozniak et al., 2016) to allow the accessibility for people with different disabilities.

Virtual systems have been developing over the last years, but still faces many issues. The development of theoretical artefacts will help to provide more functionalities to achieve pre-determined objectives and also to let the usage for different types of people.

Concerning the employment of artificial intelligence mechanisms for virtual elements, the research tries to implement new functionalities to the traditional processes. The architectures of multi-agent systems reveal high importance in order for the system to provide accurate detection and increase the process of recommendations (Morais et al., 2012)

The drone has to be programmed in order to access the inputs like gestures captured by the camera and treat the information obtained accurately.

Unmanned aerial vehicles (UAVs) can be controlled by hand motion, making the process easier for people with visual disabilities.

The research objectives to add a new methodology to the process of interaction with the systems using motion controllers. The study focuses on the application of AI for virtual elements of mixed reality systems, where these elements can be controllers of an AI agent or have mechanisms that respond to the inputs performed by the user. The main objectives of the dissertation are:

- Increase the efficiency and functionality of Mixed Reality applications by adding AI mechanisms making easier to provide functionalities and interactivity with the users. (Objective 1).
- Improve the usability and accessibility of Mixed Reality systems by using motion controllers allowing people with different type of disabilities to perform tasks. (Objective 2).

1.4 Research Questions

These questions allow to determine the problem and fundament this study, facilitating the achievement of the proposed objectives.

- How can the implementation of AI mechanisms in mixed reality systems provide more efficiency, control and functionality? (Research Question 1).
- How to avoid collisions between virtual elements and real-world obstacles? (Research Question 2).
- Which type of motion controllers can facilitate the usability and accessibility of mixed reality applications for people with disabilities like visual, auditory, physical, speech, cognitive, language, learning and neurological? (Research Question 3).

The first two questions are related to the development of a methodology using algorithms to perform certain tasks increasing the functionality of the softwares, being more focused on the objective 1. The third research question is more related with the second objective aiming to provide a solution for the problem of accessibility for people with visual, auditory, physical, speech, cognitive, language, learning and neurological. The answers to the questions, allow to satisfy the proposed objectives, due to being focused on the possible solutions for the research problem. The study focuses on the methodologies to develop MR systems increasing the efficiency and usability when compared with traditional processes.

1.5 Research Methodologies

This section mentions the different types of research performed for the dissertation, facilitating the conduct and organization of the work performed.

1.5.1 Exploratory Research

The Exploratory Research can be very useful when answering questions like how and why, helping to understand the nature of the problem aiming to generate hypothesis and getting insights and ideas. The methodology in cause tries to answer the knowledge gaps and analyses the scope. Some of the advantages are determining the reliability of methodological characteristics and obtaining primary knowledge on the object of research (Zukauskas et al., 2018). The questionnaires to the final users and observation will follow a qualitative research process and the performance analysis, the results of the algorithms, the softwares/prototypes and the analysis of code, papers and sites will follow a quantitative methodology.

1.5.2 Research Paradigm and Methodology

The research aims to provide answers to the investigation questions in order to achieve the proposed objectives, using a continuum epistemological perspective, analyzing the nature, origin and scope of knowledge to reach a possible response to the problem. A double approach (positivist, interpretative) requires the combination of the quantitative and qualitative overviews. The research will provide a possible solution by implementing artefacts to reach the proposed objectives and answer the research questions.

The developed solutions were tested and the performance and distances to execute the tasks were measured taking in consideration the existing processes.

The data collection follows a quantitative perspective and the surveys an open-ended format. By using exploratory research, it is important to understand the nature of the problem, so a study of the challenges was done, taking in consideration the existing methodologies used for different types of disabilities as well as an investigation of other researches, using motion controllers and mixed reality systems and the approaches used to resolve problems like collisions between virtual elements and accessibility for certain handicaps.

The study focuses initially on the limitations and possible technologies, like implementing motion controllers and AI mechanisms; on a second phase, prototypes were developed aiming to resolve the problems like implementing collisions in MR systems and development of motion controllers, allowing to move a drone and also offering the possibility to be used by people with different type of disabilities. The third stage is concerning implementation of the developed systems and in the fourth stage, a questionnaire to the users was performed, as well as tests to improve usability based on performance and measurement metrics to qualify the developed artifacts. After the results, will be possible to obtain conclusions mentioning if the presented solution is reliable. For the first objective and the first and second research question a mobile FPS prototype was developed to defend the concept of improving AI mechanics in order to provide more functionalities to achieve objectives and for the second objective and the third research question several prototypes were developed and tested like a hand gesture drone controller allowing the usage for different type of users.

The survey to the testers followed a qualitative approach and the artefacts to defend the concept used a quantitative approach.

The qualitative research might be more flexible or adjustable and may use categorization and patterns, to perform this research methods like interviews, focus groups and observation can be used. Basis in exploration and understanding of the nature of the problem and hypothesis can be determined. Intends to create knowledge and fill research gaps, providing an interpretation of the problem. This type of approach can be adjusted to what is learned following an iteration process aiming to explore the problem.

Quantitative research may have a more mathematical overview and manipulation of observations to describe and explain the phenomena that the observations reflect employing empirical methods. May analyze data in a more numerical perspective in order to achieve conclusions or to justify certain assumptions (Sukamolson, 2007).

Table 1. Quantitative and qualitative approaches. Adapted from (Mack et al., 2005).

	Quantitative	*Qualitative*
General framework	• Seek to confirm hypotheses about phenomena • Instruments use more rigid style of eliciting and categorizing responses to questions • Use highly structured methods such as questionnaires, surveys, and structured observation	• Seek to explore phenomena • Instruments use more flexible, iterative style of eliciting and categorizing responses to questions • Use semi-structured methods such as in-depth interviews, focus groups, and participant observation
Analytical objectives	• To quantify variation • To predict causal relationships • To describe characteristics of a population	• To describe variation • To describe and explain relationships • To describe individual experiences • To describe group norms
Question format	Closed-ended	Open-ended
Data format	• Numerical (obtained by assigning numerical values to responses)	• Textual (obtained from audiotapes, videotapes, and field notes)
Flexibility in study design	• Study design is stable from beginning to end • Participant responses do not influence or determine how and which questions researchers ask next • Study design is subject to statistical assumptions and conditions	• Some aspects of the study are flexible (for example, the addition, exclusion, or wording of particular interview questions) • Participant responses affect how and which questions researchers ask next • Study design is iterative, that is, data collection and research questions are adjusted according to what is learned

1.6 Document Structure

This work is organized as follows. The present Chapter 1 is where the context, motivation and definition of the problem is done mentioning also the objectives, challenges and research questions which this dissertation intends to answer followed by the research methodologies used, that in this case was the exploratory research due to focusing on the problem showing important insights.

The second chapter mentions the theoretical background which provides concepts and principles and also to prove the nature of the problem, as well as related work and investigations showing how other authors approached the problem.

Chapter three show the development done for mixed reality applications focusing on the AI mechanisms demonstrating a process and prototypes to add functionalities and also a possible solution to the problem of collision of the virtual elements with real world objects.

The forth chapter will be more focused on motion controllers for MR systems aiming to increase the usability for different type of disabilities taken in consideration certain aspects like eye gaze, hand motion and presenting several artefacts, being one of them a hand gesture drone controller.

On the fifth chapter shows the tests and results obtained taking in consideration qualitative and quantitative aspects, by using a questionnaire and measuring distances and performance in order to justify and defend the proof of concept in cause.

The final chapter will show the conclusions obtained from all the developed process, looking for an efficient solution for the previously described issues taking in consideration the work performed and the results obtained.

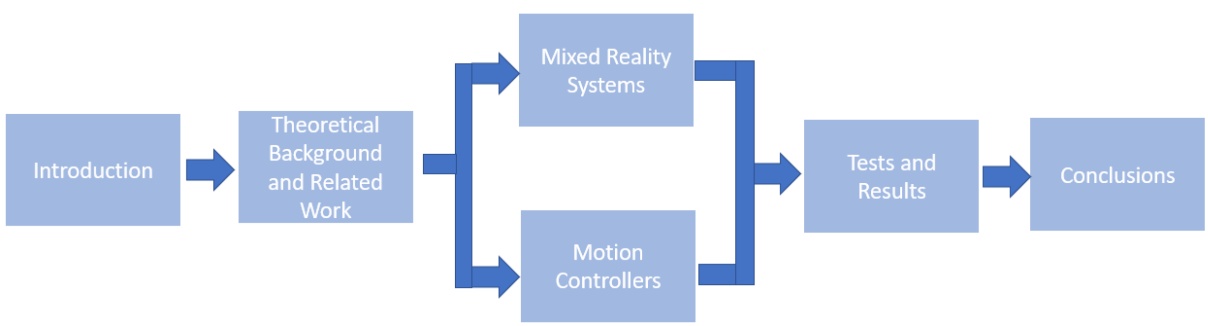


Figure 3. Dissertation structure.

Chapter 2

Theoretical Background and Related Work

2

Taking in consideration the objectives, the proposed questions, the necessary technologies and the research performed by other authors, this section will provide more insights concerning the topics related with the investigation in cause, demonstrating the nature of the problems, how other researchers approached for a solution and which limitations and challenges exist.

The chapter will mention the difficulties in obtaining ubiquitous AR and also to develop systems that respect the usability standards making easier the accessibility concerning several types of disabilities. Initially will describe AI agents and how AI systems are based, the section 2.2 describes challenges and methodologies concerning mixed reality applications, followed by theoretical concepts of Computer Vision for Artificial Intelligence.

Chapter 2.4 describes fundaments related to usability and accessibility mentioning needs and how systems can be adapted to users with different type of disabilities followed by a section related to motion controllers considering other projects and how users can interact with the systems via eye gaze and hand motion and finally a chapter about web development mentioning characteristics and functionalities.

2.1 Multi-Agent Systems

An AI system is composed by an agent and an environment. The agents act in this environment, and there may be other agents.

Concerning simple reflex agents, generally the function is based on a set of rules, actions and conditions. Historical data is ignored and act only based on current perception. The action/condition allows to define a state, for example if the condition is true, perform a certain action.

Model-based reflex agents use models to assess the environment, in these models it is important to find a rule, whose condition satisfies the current situation. There are models that use the history of perception and internal memory to make decisions about an internal "model" of the environment. Internal memory allows these agents to store part of their navigation history and then use this semi-subjective history to help them understand things about their current environment, even when everything they need to know cannot be directly observed.

Goal-Based Agents make decisions depending sometimes on the distance to the target having to choose between multiple possibilities, usually selecting the one that reaches a desired target state. All of their actions have the objective of reducing the distance from the goal to the value of 0. Decisions can be represented explicitly and can be modified, making the process more flexible. This implementation generally requires research and planning and the behavior can be easily altered.

Utility-based agents do not act only on a non-objective basis, but in a better way to achieve it. The utility of the agent differs from its counterparts, helping to choose the best alternatives, when there are several available. Aims to maximize the utility function, this function maps a state with utility measurements. The process is flexible and adapts to changing environments, the agent is able to make rational decisions, even when the objectives are inadequate.

Learning Agent is related to Machine Learning (ML), this type of agent manages to learn from past experiences and performs actions without these being explicitly defined in the Code.

The learning agent consists of two elements: the learning element and the performance element. The performance element offers the possibility for the agent to select the actions based on the perceptions received. The learning element assesses the behavior of the performance element and proposes improvements. Sometimes, it also suggests actions and mechanisms that allows to improve the behavior of this agent.

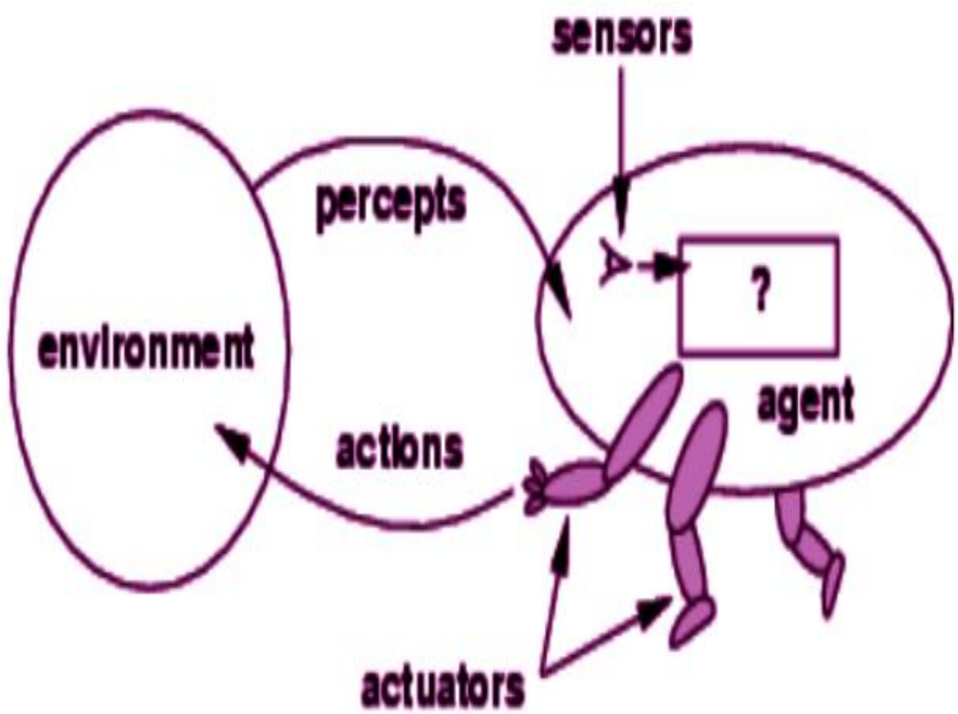


Figure 4. An Intelligent Agent (Russell & Norvig 2003).

Mulit-Agent System (MAS) corresponds to an area of research and study of artificial Intelligence (AI) in modern times. A Multi-Agent System consists of various decision-making agents that interact in a shared environment to achieve common or conflicting goals. It can be said that it is focused on behavior analysis, definition of algorithms and placement of objectives.

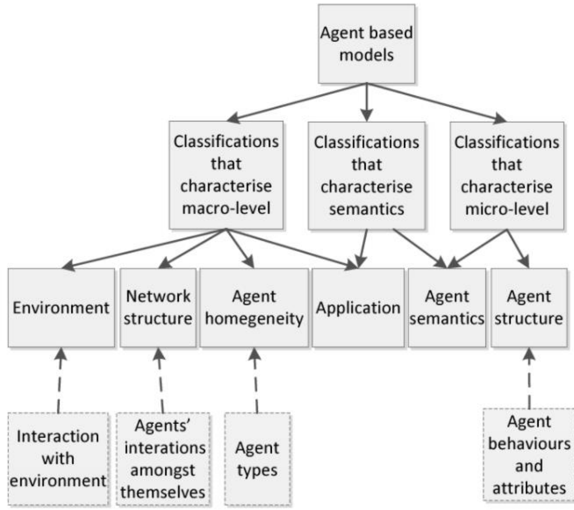


Figure 5. Classification of Agent Based Models. Adjusted Pudane (2017).

In Multi-Agent System (MAS), various agents can interact with each other; these agents can be homogeneous or heterogeneous, and should be analyzed their organization, interaction, relationship with the environment, spheres of influence and the communication process between agents.

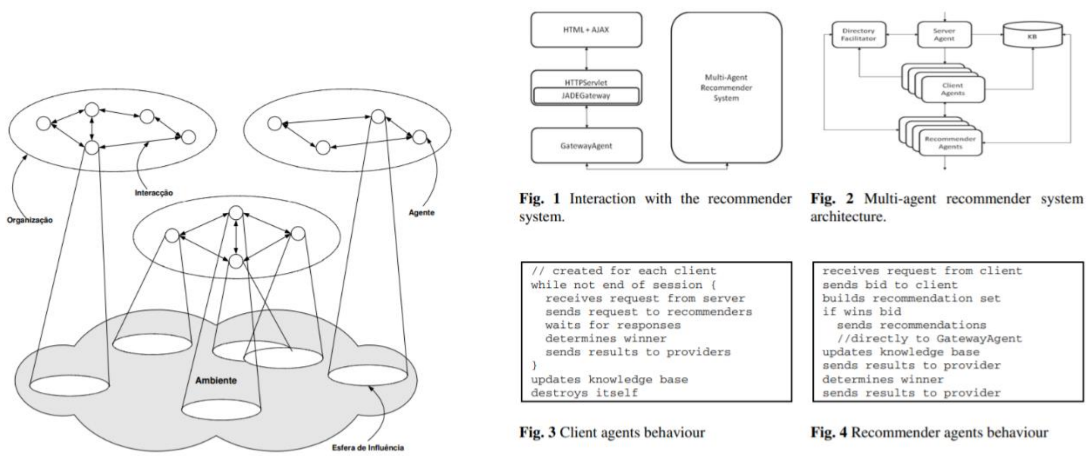


Figure 6. Theorization of architectures and behaviors in recommenders (Morais et al., 2012).

Over time, distributed artificial intelligence has been evolving and has been applied to abstraction and divide and conqueror concepts for problem solving. Distributed Artificial Intelligence focuses on: Granularity of two agents; heterogeneity of two agents; methods for the distribution of control (among the agents) and possibilities of communication between the agents.

In general terms, communications are governed by protocols, which allow to define standards in the process of sending data. The Knowledge Sharing Effort (KSE) is developed in the KQML (Knowledge and

Query Manipulation Language) based on the exchange of messages for knowledge and the KIF (Knowledge Interchange Format).

Learning can be interactive or individual and must have various aspects such as the property of the objective, or the method used. In the coordination of agents, different types of dependencies can be verified, such as independence, unilateral, mutual and reciprocal dependency.

There are several types of architectures, this study will describe some of the more commonly used.

Deliberative Architectures: the decisions of the agents are based on logical reasoning.

Reactive Architectures: are usually performed in real time and the agent acts from interactions with the environment, the behaviors are sometimes defined having in mind the functionalities and it is possible to have various behaviors at the same time.

Hybrid Architectures: combines reactive architecture with deliberative architecture. As actions have in account the environment and the functionalities of the agent.

Architecture by layers: based on the existence of subsystems. It can be considered that on the horizontal layers the software is linked to the input sensors and the output actions and in the vertical the input sensors and the actions have at least one layer between them, from the input to the output this will go through several filters.

Architecture BDI (Belief-Desire-Intention): The belief can be seen as the current state of the agent based on what he considers to be true at a given moment (like the knowledge of the environment).

The wishes are the objectives, which can arise later; the wish can be obtained from the use of a reasoning on the existing sub-desires.

The intentions can be seen as actions and tasks that the agent chooses to perform in order to accomplish his objective.

Generic Architecture of a Social Agent: agents usually have models of other agents. In this type of architecture, the agents can learn, in order to improve their decisions in the future. According with Foo (2007), the behavior of the agents can follow the mathematical model of game theory and search for the Nash equilibrium.

The agents can be evaluated over several perspectives like the Performance Measure, Environment, Actuators and Sensors (PEAS).

2.2 Ubiquitous Augmented Reality and Virtual Reality

Ubiquitous AR has been the most common approach for dynamic objects. The outliers are identified by cues; in example, by re-projecting and removing the matches that are not accordant to the expected motion.

These approaches lead to erroneous estimation of motion in scenes with very few inliers.

PARSAC (Tan et al., 2013) has adapted Random Sample Consensus RANSAC (Fischler & Bolles,1981) to handle dynamic scenes. RANSAC allows to efficiently remove outliers even when the inlier ratio is rather low, so that the camera pose can be reliably estimated on challenging situations (Wang et al., 2022).

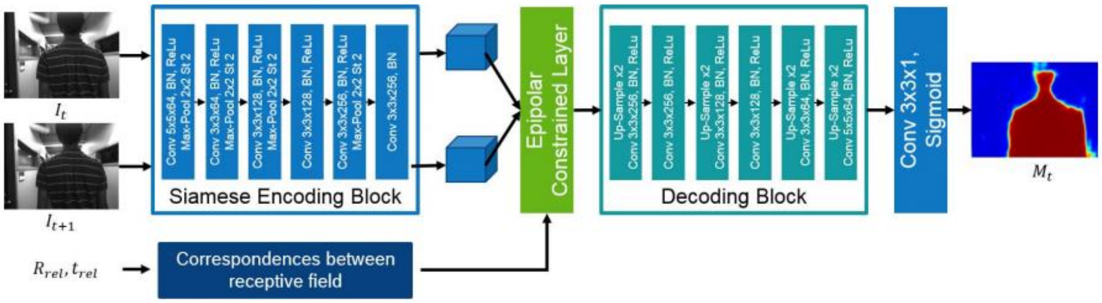


Figure 7. Example of a Deep Neural Network (Babu et al., 2018).

Babu et al., (2018) have developed a Multi-motion MC-VIO algorithm that is able to track landmarks related to primary and secondary motions. Ubiquitous Augmented Reality (AR) should provide exact values anytime, anywhere and should be able to localize and provide virtual elements on static and also moving coordinates of the real world.

Babu et al., (2018) considered that:

1. The camera pose is vital for the approach in a dynamic environment, considering a moving vehicle the objects outside the vehicle are based on the inertial coordinate frame and the objects inside the vehicle in a local coordinate frame. In a dynamic environment, the application of AR is more complex.

2. The camera tracking of the moving objects is required to generate AR anywhere, anytime. On a moving vehicle, for example it will be necessary to view virtual objects inside the vehicle and outside.

The (Babu et al., 2018) approach performs multi-motion tracking in larger dynamic environments.

Random sample consensus (RANSAC) is an iterative algorithm to calculate the in liners removing the outliners consequently. Process: 1. Randomly selecting a subset of the data set 2. Fitting a model to the selected subset. 3. Determining the number of outliers. 4. Repeating steps 1-3 for a prescribed number of iterations.

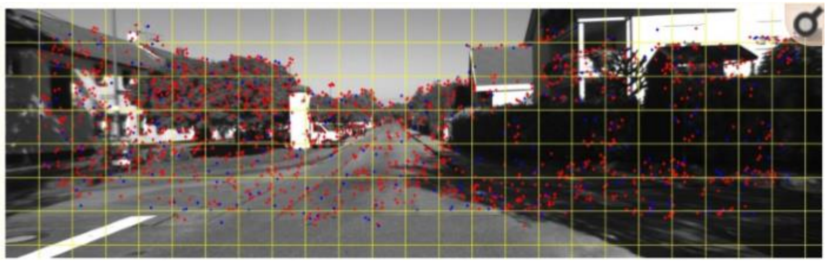


Figure 8. Features detection process before and after bucketing: the red points stand for features before bucketing, the blue points stand for features after bucketing, and the yellow lines depict individual buckets. Source: (Liu et al., 2017).

For some cases, to create AR in an efficient manner, it is necessary to calibrate the camera in order to correct the lens distortion.

One of the commonly used processes to detect collisions for augmented elements is the use of point clouds, for example, Kinect retrieves 4d depth information of the images and programs that use Point Cloud Library (PCL) are able to create cloud points according with overhangs of the real images.

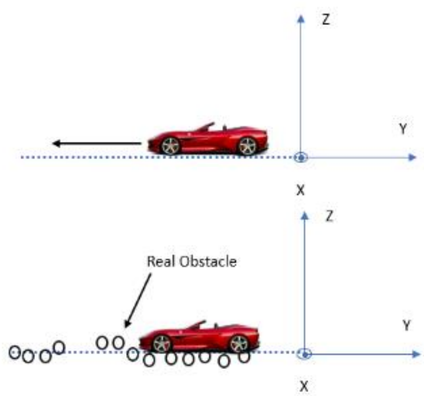


Figure 9. Demonstration of collisions with point clouds.

The moment the augmented element detects a specific point cloud, it will be triggered via code to do another function like stopping. In the image below, a project developed in OSG and PCL that creates virtual elements in real world images using two 3D coordinate systems that interact with each other providing information in order to facilitate the process of placing AR elements.

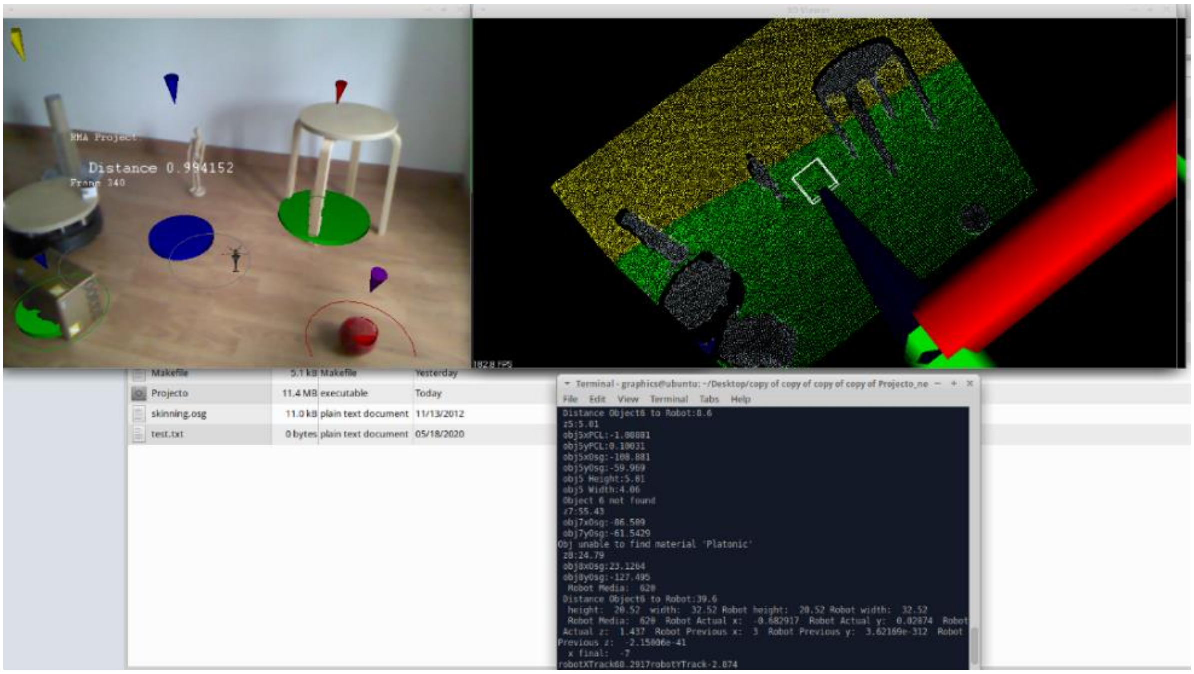


Figure 10. Previous project developed in Point Cloud Library (PCL) and Open Scene Graph (OSG).

Virtual Reality (VR) allows to provide a visual experiment, using specialized hardware like the VR glasses, HUDs (Head mounted displays), data gloves or special installations like Cave Automatic Virtual Environment. VR development requires the use of specialized libraries and add-ons in order to treat the images accordingly so it can be seen on the display. Factors like the player position and orientation should be taken in consideration. A 3D map can be created using engines like Unity or Unreal Engine among other possibilities.

According with Zheng et al. (1998), there are three factors that any VR application should have, that is response to the user actions, the 3D graphics should be in real time, and there should be a sense of immersion. The virtual reality over a high-speed network can be a research tool because big datasets, advanced graphics and real time processing can be used.

The process of data visualization can be described as a sequence of fundamental processing steps (Haber, 1990). Simulation: the data sensing and measurement are used as an input. Data selection and filtering: The data obtained has to be analyzed and selected. Visualization mapping: The data has to be transformed to primitives as well as their properties like color and size. Rendering: The primitives are rendered as images that will appear on the screen.

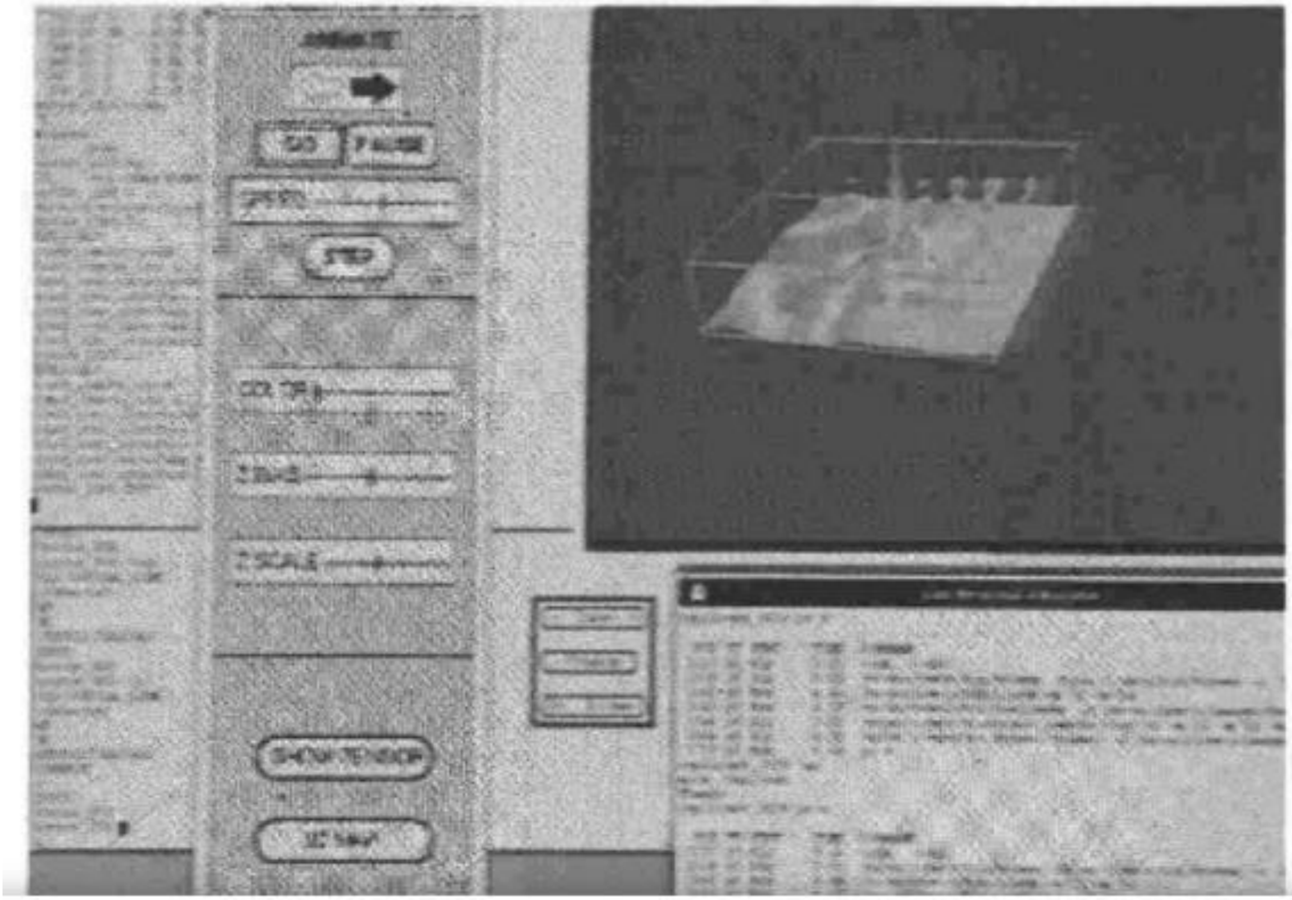


Figure 11. Supercomputer Visualization (Haber, 1990).

Virtual Reality facilitates the interface between man and machine; computer generated images (CGI) and 3D dimension images can be generated and several hardwares may be required in order to create a sense of immersion like the VR glasses.

Krevelen (2010) work is based on automatic calibration, human tracking, use of displays and a global positioning system. Mixed Reality (MR) is based on the insertion of virtual elements into the real world captured via a sensor like a webcam. The real environment can give many information like the positioning, depth and Red, Green, Blue (RGB) values of each pixel which can help the process of inserting computer-generated images into the images captured by a sensor.

Engines like Unity and Unreal Engine allow to add these virtual elements to a real-world environment.

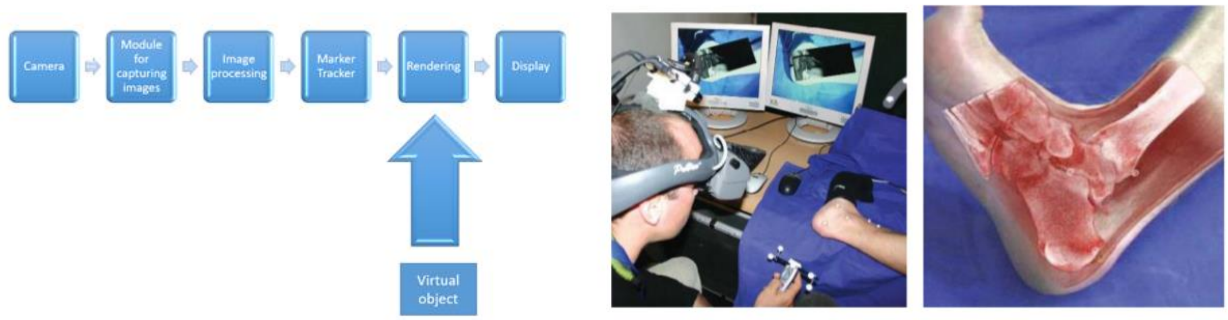


Figure 12. Virtual Object Rendering and Medical Scanner (Krevelen, 2010).

Considering the study of Krevelen (2010) based on automatic calibration and human tracking, various monitors can be used and also a global positioning system, taking into account various UIs and positioning of two pixels. There is a possibility of obtaining different types of mixed reality.

To facilitate the process of explaining the pinhole camera model in the figure below. The point Oc is in the world coordinate system (Xw,Yw, Zw). OcXcYcZc is the camera coordinate system, o′uv is the pixel coordinate system (PCS), and oxy is the retinal coordinate system (RCS) (Wang & Lynch, 2014).

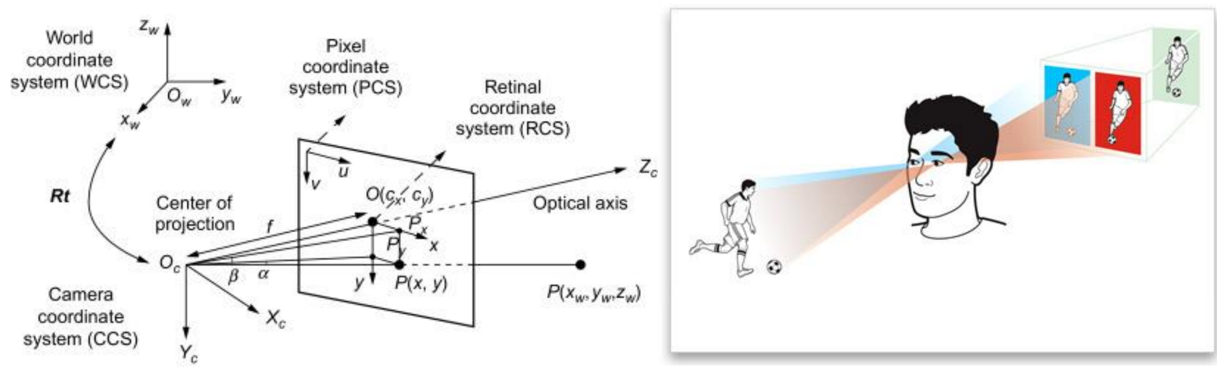


Figure 13. Adjusted from (Huang et al., 2014).

In order to prospectively project a 3D homogenous point Q = (X, Y, Z, 1) to the plane of the camera, with n as the z coordinate of the image plane; the equation provided by resolving the matrix system below will give the homogeneous coordinates of the pixel, q=(x’, y’, w’) (Huang et al., 2014).

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} n & 0 & 0 & 0 \\ 0 & n & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \qquad \lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & o_x & 0 \\ 0 & f & o_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}.$$

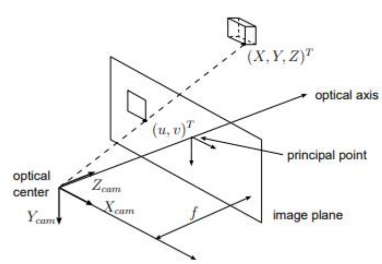


Figure 14. Adapted from the work of Morvan (2009).

In the figure above the left matrix corresponds to the coordinates of the homogeneous 2D point q and the matrix on the right to the homogeneous 3d point D. The matrix in the middle is the transformation matrix that allows to get the coordinates of one point knowing the other. The use of a Multi-motion MC-VIO algorithm may allow to track landmarks related to primary and secondary motions. It has been the most common approach for dynamic objects.

2.3 Artificial Intelligence for Computer Vision and the Photoplethysmographic Signal

The Photoplethysmography (PPG) is an optical technique to detect changes of the blood volume, giving important information about the cardiovascular system; The waveform has two parts, the alternating current (AC) and the direct current (DC). The AC component corresponds to the variations of the blood volume that are synchronized with the heart movements. The DC component is non pulsatile and based on the reflected optical signals and light absorption from the tissues depending on the tissue structure as well as venous and diastolic arterial blood volume (Castaneda et al., 2018).

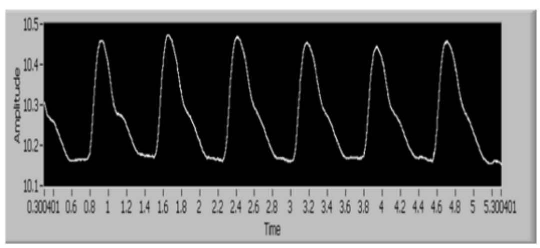


Figure 15. The PPG signal acquired in LabView (Bagha & Saw, 2011).

By using the algorithm methods, such as, the Fast Fourier Transform (FFT) that converts a signal into separated spectral parts providing information about the frequency of the signal and Independent Component Analysis (ICA).

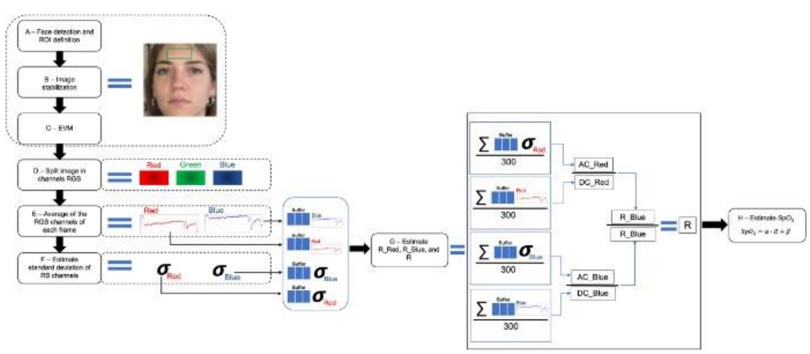


Figure 16. Framework to retrieve an estimation of SpO2 and Framework for respiration detection from the PPG signal (Molinaro et al., 2022).

The formula used to calculate the SpO2 can be based in the AC and DC of the blue and greenchannel (Scully et al., 2011).

$$SpO_2 = A - B\frac{AC_{RED}/DC_{RED}}{AC_{BLUE}/DC_{BLUE}} \tag{1}$$

To determine the cardiac frequency, previous studies have showed that the green channel has stronger plethysmograph signal (Kwon et al., 2012).

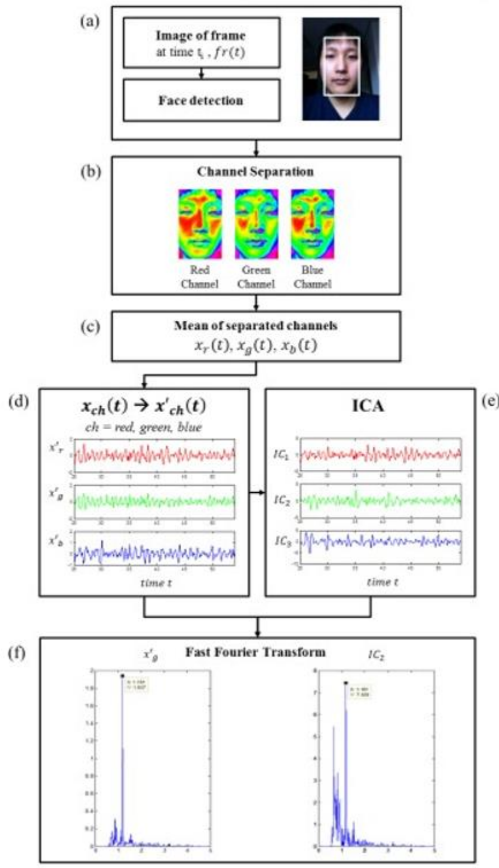


Figure 17. Heart rate extraction (Kwon et al., 2012).

Any channel, red, green and blue can be normalized via a function where μ is the average and σ the standard deviation of Xch(t) respectively.

$$x'_{ch}(t) = \frac{x_{ch}(t) - \mu_{ch}}{\sigma_{ch}} \tag{2}$$

The treatment of the PPG signal and the use of the algorithms described allow to determine vital signs like the cardiac frequency.

Information Systems can use the Photoplethysmogram (PPG) signal to obtain viral signs among other possibilities. In many researches, the beats per minute (BPM) calculation by a camera shows close results to the ones reported by medical devices. The values obtained by a software/prototype can be compared with the ECG values (at 1000Hz), mainly for the green channel. There are many AI for Computer Vision studies that use the Region of Interest (ROI) detection mixed with the Fast Fourier Transforms (FFT) (Kwon et al., 2012). The accuracy of the values reported has to be high and have a low error margin when compared with other equipments. Processes like Joint Approximation Diagonalization of Eigen-matrices (JADE) to separate observed mixed signals into latent source signals and FFT (ECG uses the Fourier Series) allows to calculate the vital signs values based on the spectrum considering duration, frequency and distance, the analysis should take in consideration noise and luminosity.

Some OpenCV algorithms allow to obtain the ROI and later divide the capture into 3 Independent Components (ICA) (Kwon et al., 2012). The 3 divisions can be normalized before applying FFT and present the values of the 3 ICA with normalized Red, Green, Blue (RGB) values. Studies mention that the green channel has a stronger plethysmographic signal (PPG).

In AI for Computer Vision, the sense of sight is usually studied in more depth.

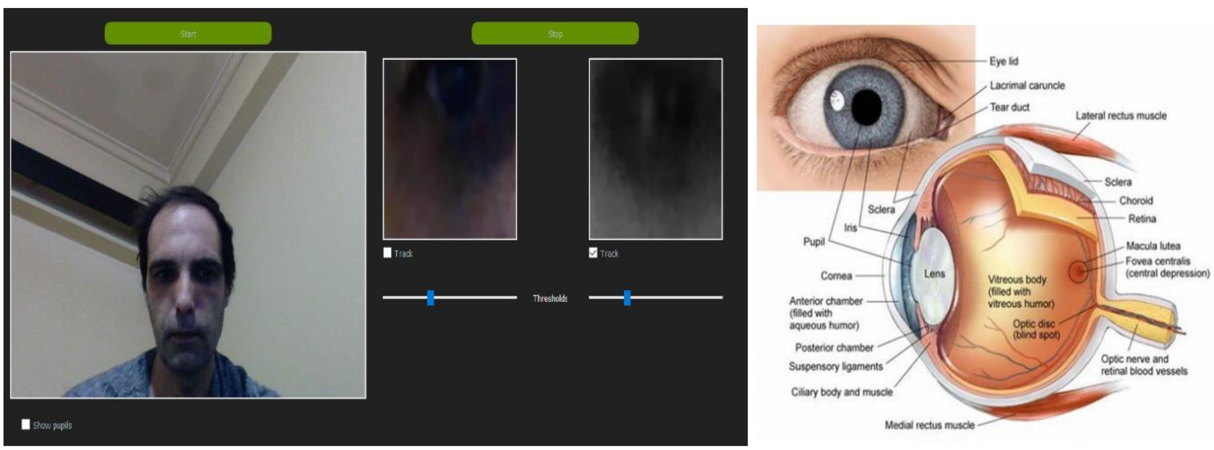


Figure 18. Human Eye, (Ahirwal, 2020).

In the human retina there is the fovea which deals with cones with high density, which are colored (red, blue or green) sensitive cells that require bright light conditions and the periphery that works with high density of rods, which are color-insensitive cells, thus being favorable for low light conditions.

Fixing objects at night can make them invisible: the lack of rods in the fovea limits our ability to perceive what we are focusing on in low-light conditions

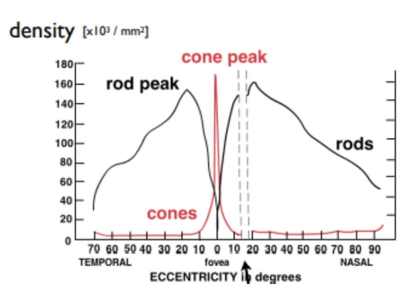


Figure 19. Eccentricity in degrees (Osterberg,1935).

In the Eccentricity plot, the zone that does not have degrees is due to an optic nerve, but the interpolation of the eyes allows to solve the problem (Osterberg, 1935).

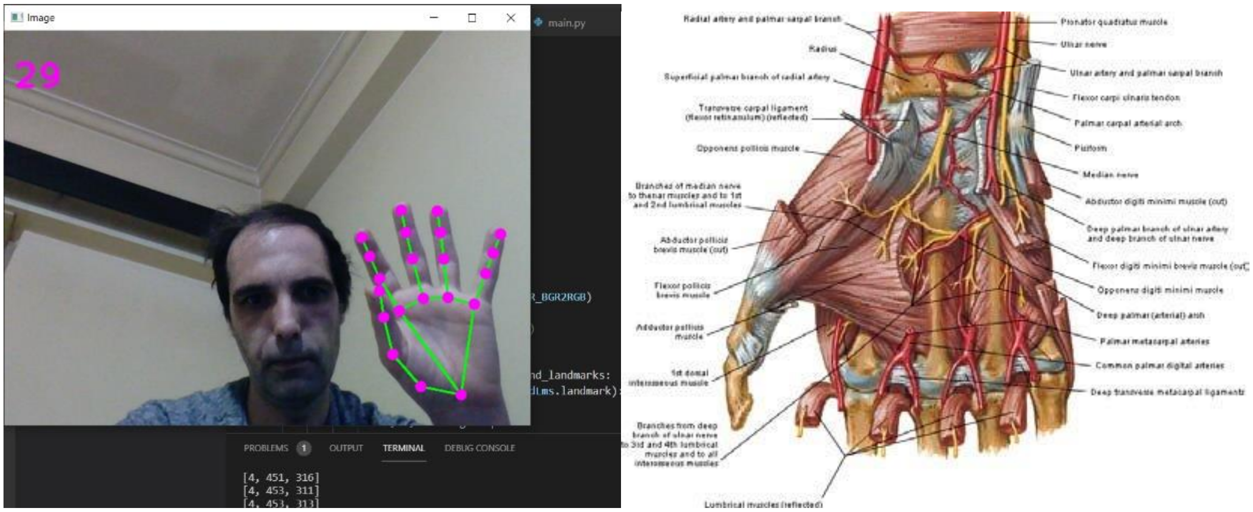


Figure 20. Human hand detection and Human hand (Zecca et al., 2006).

Gestures can be used as a form of communication with AI agents as an alternative to verbal communication, allowing people with speech disabilities to interact with the system, certain gestures can be identified after classification and make for example drones move according with the mean of the gesture.

2.4 Usability and Content Accessibility

The Web Content Accessibility Guidelines (WCAG) 2.0 aims to make web content generally accessible with a focus on people with disabilities. There are several types of disabilities that must be taken into account, such as, visual auditory, physical, speech, cognitive, language, learning and neurological. These measures take into account various aspects such as security, compatibility, navigation, video presentation, text presentation, adaptability, assistance with Input (such as correctors), possible accessibility and usability improvements can be done with the keyboard or the Liquid Crystal Display (LCD) by improving visibility like adjusting color combinations that facilitate the process of focusing and making easier to read the showed image. Information Systems (IS) should be predictable, and must follow patterns and time constraints such as defining frames that allow the user to perform his task via the application in a more efficient and effective manner.

The WCAG is able to address a number of questions, but can not respond to all possible types of deficiencies, because disabilities vary according with different aspects like grade, for certain levels the system may not be used with efficiency or in cases where patients have combinations of different disabilities. The age of the users has also to be taken in consideration, some Hz (Hertz) frequencies might not be heard after certain ages; guidelines can be used in order to make the web content easier to use for older people, because normally the abilities change or decrease due to aging. The usability has to be managed and developed to improve the obtention of results for users in general.

In information systems usability, systems must be: easy to use, effective and pleasant to use, minimize errors, increase satisfaction, decrease frustration, make tasks more productive and hide unnecessary technical details from the user.

The design of interfaces requires identifying needs and establishing requirements (where users are defined under correspondent metrics, needs are ascertained and profiles are drawn). Whenever possible there should be alternative designs taking into account conceptual and physical models. In this process, there must be the development of demo versions and tests, such as prototypes and Mockups to be able to present a solution closer to the need of the customer. What was developed, should be evaluated in several aspects like the number of errors, or the number of requirements fulfilled. The system should focus on the user and identify the more accurate usability criteria.

Usability objectives are:

Effectiveness or effective use: how good is the application at producing the expected results? Efficiency or efficient use: should be fast, have automatic error and exception handling. Safety or safe to use: avoid unwanted situations (prevent errors). Usefulness or good use: number of functions correctly offered. Should be easy to learn and memorize.

The World Wide Web Consortium (W3C) in accessibility aims to be developed for patients with disabilities and to allow these systems to be operated by users with difficulties to: understand, comprehend, navigate, interact and contribute via web.

Online accessibility tries to cover all disabilities that affect access to the web, such as: auditory, cognitive, neurological, physical, speech and visual.

The International Organization for Standardization (ISO) 9241-11:2018 is more focused on the Ergonomics, improving the human-machine interaction; showing usability definitions and concepts distinct from those of accessibility, such as the possibility for a specific user to manipulate specific tools in order to meet specific goals in terms of effectiveness, efficiency, and satisfaction, for a specific context of use, while the ISO/TR 16982:2002 is more focused on the User-Centred Design and can be defined as the design of products (and services) that users can use both for a specific purpose or to perform other requested operations and tasks with very little effort and great efficiency.

Table 2. WCAG compliance levels. Adapted from (Noh et al., 2015).

Compliance Level	**A**	**AA**	**AAA**
Importance	Importance 1	Importance 2	Importance 3
Concept	Must	Should	May
Definition	Must comply	Should comply	May comply
Significance	Fulfillment of basic requirements	Removal of grave defects	Removal of slight defects
	Guarantee of web accessibility	Increase in web accessibility	Improvement of web accessibility
	Access to web contents is impossible in case of noncompliance	Causes difficulty in accessing web contents in case of noncompliance	Causes inconvenience in accessing web contents in case of noncompliance

WCAG A (minimum) represents the most simple considerations, based in a criterion that has to be met and possess the most basic notions and the main standards according to a set of rules. WCAG AA deals with higher and most common barriers for certain deficiencies within certain grades; must be complied with and holds more rigid rules, which greatly facilitate accessibility to the website. WCAG AAA (maximum) represents the highest level of accessibility, and is possible to perform, but may have high requirements and might be difficult to apply to the entire site, as an example the definition of the front end code easing the processes of reuse, where all pages use the same master page or just maintaining certain elements such as the head or the left and right menus and the footer.

What happens should be evaluated as the number of errors, or the number of requirements met. Must focus on the ease of use and identify processes to improve accessibility. The objectives of usability are: Effectiveness or effective use (in example: How much is the system capable of producing considering the expected result?) Efficiency: High speed of processing, with automatic handling of errors and exceptions. Safe to use: Avoid undesirable situations (prevent errors). Provide utility: the number of functionalities must be correctly offered and aiming to resolve the problem in an optimal way. Should be easy to learn how to use and to memorize. Other aspects of W3C are considered in accessibility to ensure that the systems are developed for people with disabilities and allowing users with difficulties to: perceive, understand, navigate, interact and contribute using the web. Online accessibility also aims to be implemented in mobile phones, smart watches, Smart TVs and other devices with different screen sizes (the use of bootstrap may help in this process) and also for different input modes.

Other cases to be considered are the improvement of the access to people with mutated abilities due to aging, people with temporary impairments such as temporary weakness, or people with situational limitations, like being in environments with strong sunlight, or situations without certain types of hardware like sound columns making the hearing of audio not possible, and people who use slow Internet connections or that have very limited bandwidth.

Dix et al. (2004) shows various forms of assessment, such as assessment through eye tracking which can be done by a web camera or other sensors allowing eye gaze control to interact with the application. On the work of (Dix et al., 2004) there are also references to hand writing and virtual reality among others as a form of Human-Machine interaction.

Figure 21. Eye tracking equipment. Retrieved from (Dix et al., 2004).

People with disabilities can use tools and sites if these are well defined. Currently, there are many accessibility barriers that make it difficult or impossible for some people to use them. The images must have alternative text, for people who cannot see, a reader can be used, which reads the information on a page aloud, including the alternative text of the visual image.

Other objectives are to facilitate the use of search engines and help access for situations with low bandwidth.

Some people cannot use the applications well, such as older users with limited motor control. The accessibility of a system should not depend on the mouse. All functionalities must be possible to carry out. For persons with certain disabilities there is he possibility to use assistive technologies like mimic, voice input, voice recognition, transcriptions for audio or transcriptions of examples. The combination of colors must have the usability standards, as well as the placement of elements must respect pre-defined and standard rules in order to facilitate use. In the study of (Vanderheiden & Henry, 2003) an analysis of various aspects, such as, the difficulties as factors that influence the design of interfaces is performed.

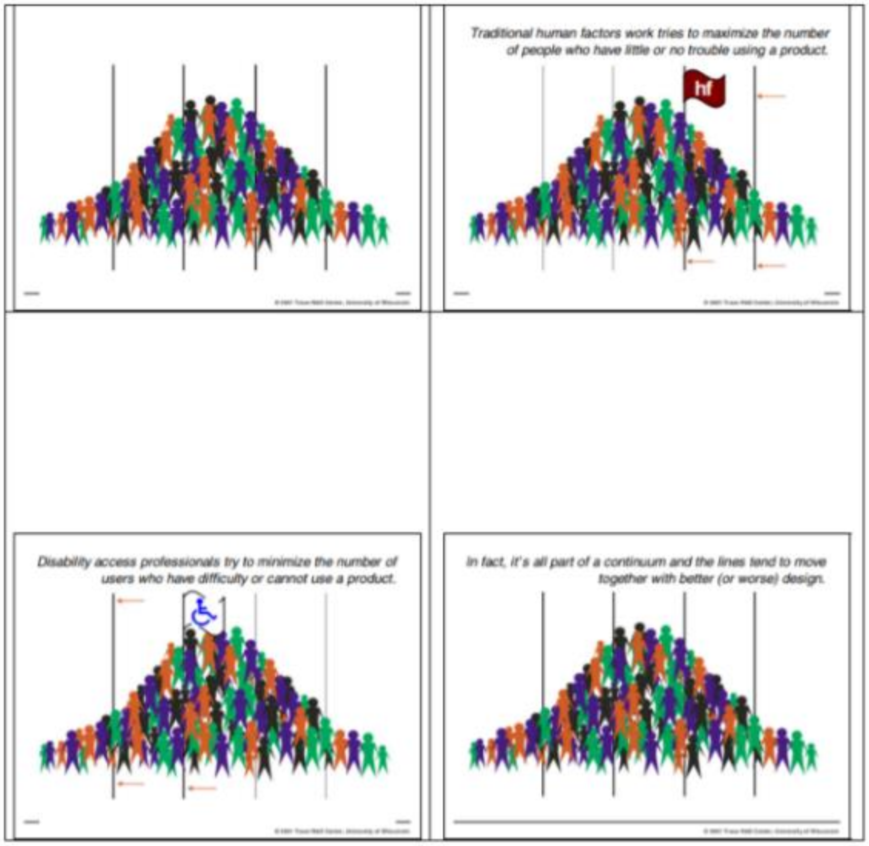


Figure 22. Analysis of accessibility and usability. Retrieved from (Vanderheiden & Henry, 2003).

In the work of (Nielsen, 1994) we can find heuristics and evaluation methods that allow to define strong points and weaknesses regarding the usability of the applications.

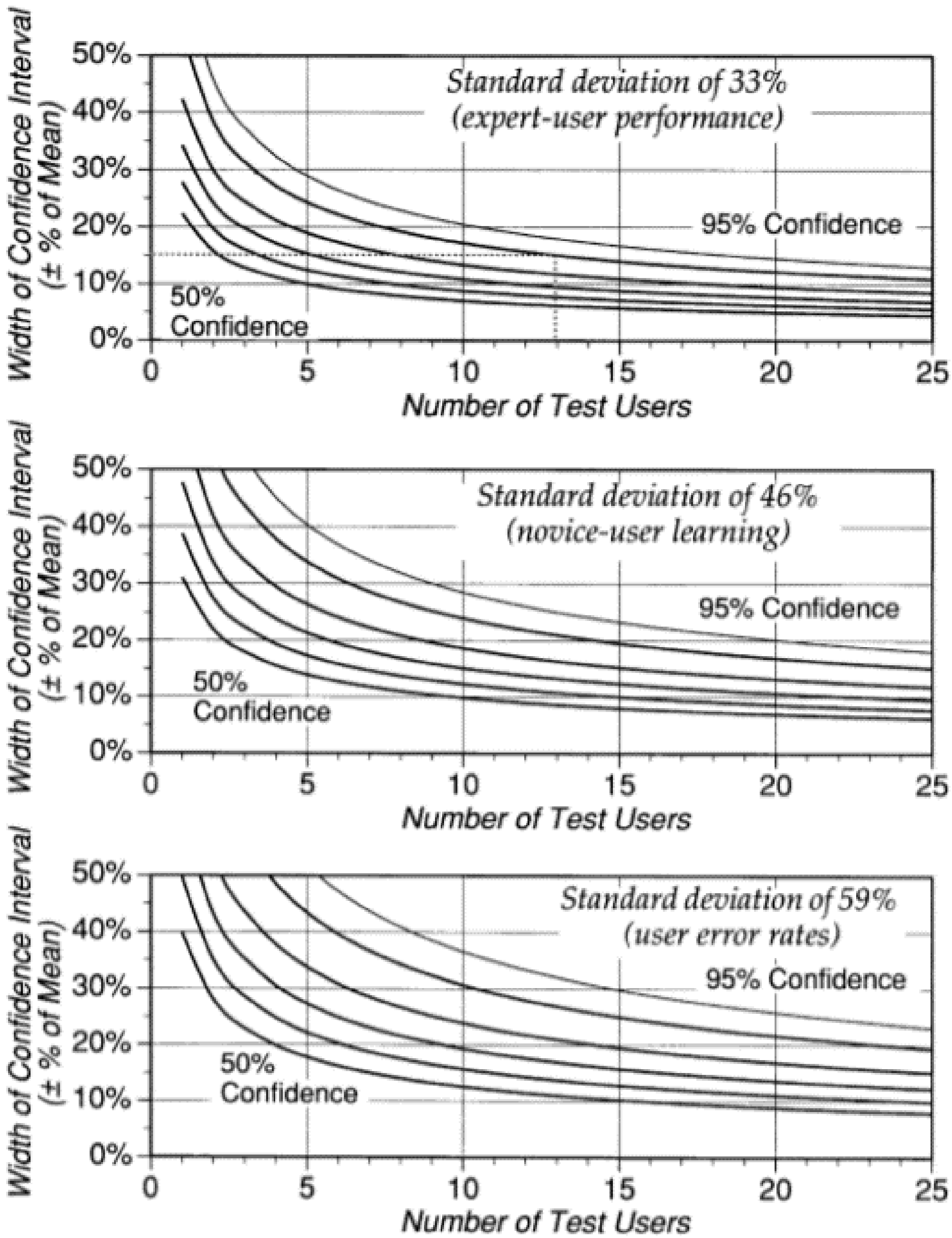


Figure 23. Test with Users. Retrieved from (Nielsen, 1994).

New technologies have appeared in this area that facilitate accessibility or use for people with special needs. Tests with users can give important information in the definition of strengths and weaknesses, such as discovering errors and finding points to improve. In the usability of information systems, the Nielsen heuristics are commonly used as a standard practice, allowing to validate and improve the interfaces, being commonly used in User Experience (UX) and can be associated with the WCAG. Garrett (2020) shows references to various aspects of UX such as colors, the placement of elements must follow a set of rules in order to facilitate accessibility.

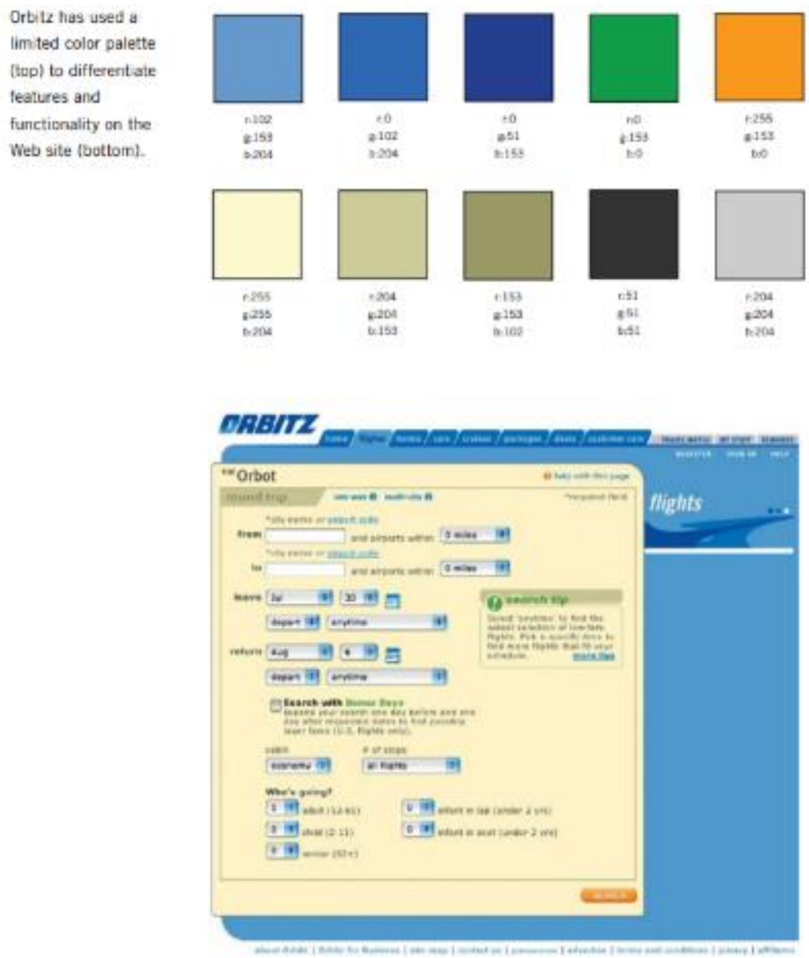


Figure 24. Reference to the choice and combination of colors. Retrieved from (Garrett, 2020).

2.5 Motion Controllers and the Equations of Distance and Movement

Nowadays motion can control applications, there are several known mechanisms which allows the interaction of the user with a system based on his movements. Eye gaze can improve the dialogues between humans and computers and the process of selection can be faster than using a mouse or a keyboard (Sibert & Jacob, 2000).

Hand gestures can be captured by the webcam using JavaScript specific libraries allowing also to interact with the system. There are many ways of interacting with a computer being the most common the keyboard and the mouse, but now there are advanced controllers that use camera and image treatment mechanisms to determine positioning and movement of parts of the body allowing to execute preprogrammed commands (Sziládi et al., 2016).

Figure 25. Leap Motion Controller (LMC) showing a virtual representation of the hand and distinguishing the fingers.

The equation of the distance between two points in a 3D space is given by the formula:

$$AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{3}$$

The Euclidian distance formula is:

$$\sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}. \tag{4}$$

The equation of movement of a UAV can be defined by the formula:

$$\tau = H(\mathbf{q})\ddot{\mathbf{q}} + C(\mathbf{q}, \dot{\mathbf{q}}) + G(\mathbf{q}) \tag{5}$$

Where H is the inertia matrix, C the centripetal matrix and G the gravitational matrix (Barrientos et al., 2002).

These formulas allow to perform measurements and can be used for evaluation of performance, the shorter distance taken to perform a certain task may indicate less effort.

2.6 Web Development

In web development, it is common to use databases and APIs, the browser has to be able to open the page and process the information. HTML can combine with CSS to make the page layout and esthetical aspects combined JavaScript or Angular for functionalities.

Nowadays web applications commonly use Web services and APIs to communicate with parts of the application or other applications. On the HTML code, there are several inputs or tags that allow to display forms. The application can be divided in tiers, and the data layer makes the interface with the database. According with the study of (Haris, 2019) it is pointed out that PHP is the programming language more used; the research also mentions the several advantages of the MVC architecture like security. The process of comparing frameworks is not easy and several criterias should be taken in account like debugging, how it connects with the database among many other factors. Many technologies can be used in web development like .NET, Java Spring MVC, on the frontend it can be used Angular or JavaScript. The choice of technology is an important factor. Concerning web development, the servers can be configured according with the type of application that is going to be implemented. PHP appeared in 1995 by Rasmus Lerdorf, a technology that has been used in large scale, especially on the backend side.

Occurances of patterns:	Sub-system 1	Sub-system 2	Sub-system 3	Sub-system 4	Sub-system 5	Sub-system 6	Sub-system 7
Files with code duplication (over 30 lines)	3	1	-	7	2	3	8
Data injection	86	4	1	-	-	-	-
Thread and synchronization issues	-	-	6	-	-	-	-
Cross site scripting	-	-	3	-	-	-	-
Usage of resource after it was closed	-	2	1	-	-	-	-
Code that is likely to cause NullPointerException	5	2	2	3	2	-	20
Calling methods of immutable objects without using the result	10	-	-	2	1	-	-
Not closing resources	-	7	2	2	2	4	190

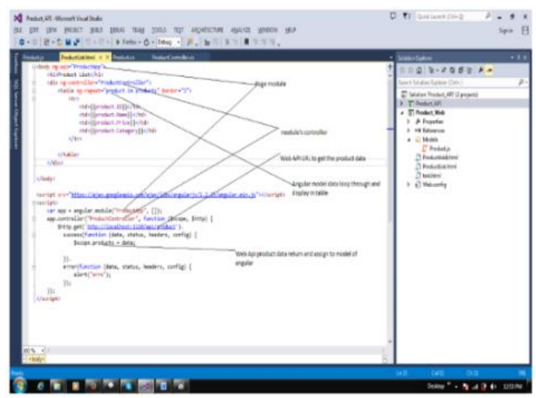


Figure 26. Application of MVC with frontend, backend and data layer with C# and code metrics (Haralambiev, 2011).

According with (Haralambiev, 2011) some metrics that can be used for code performance are: Files with code duplication (over 30 lines), data injection, thread and synchronization, cross site scripting, usage of resource after it was closed, code that is likely to cause NullPointerException, calling methods of immutable objects without using the result and not closing resources.

On the controller can be defined several operations like, associating the http command to a certain function, such as, the Create Read Update and Delete (CRUD) functions, or the details of the created entities. Microsoft has the Entity Framework and the ADO.NET, which helps with the development process.

Node.JS can be used in several platforms being a runtime environment that allows to use JavaScript.

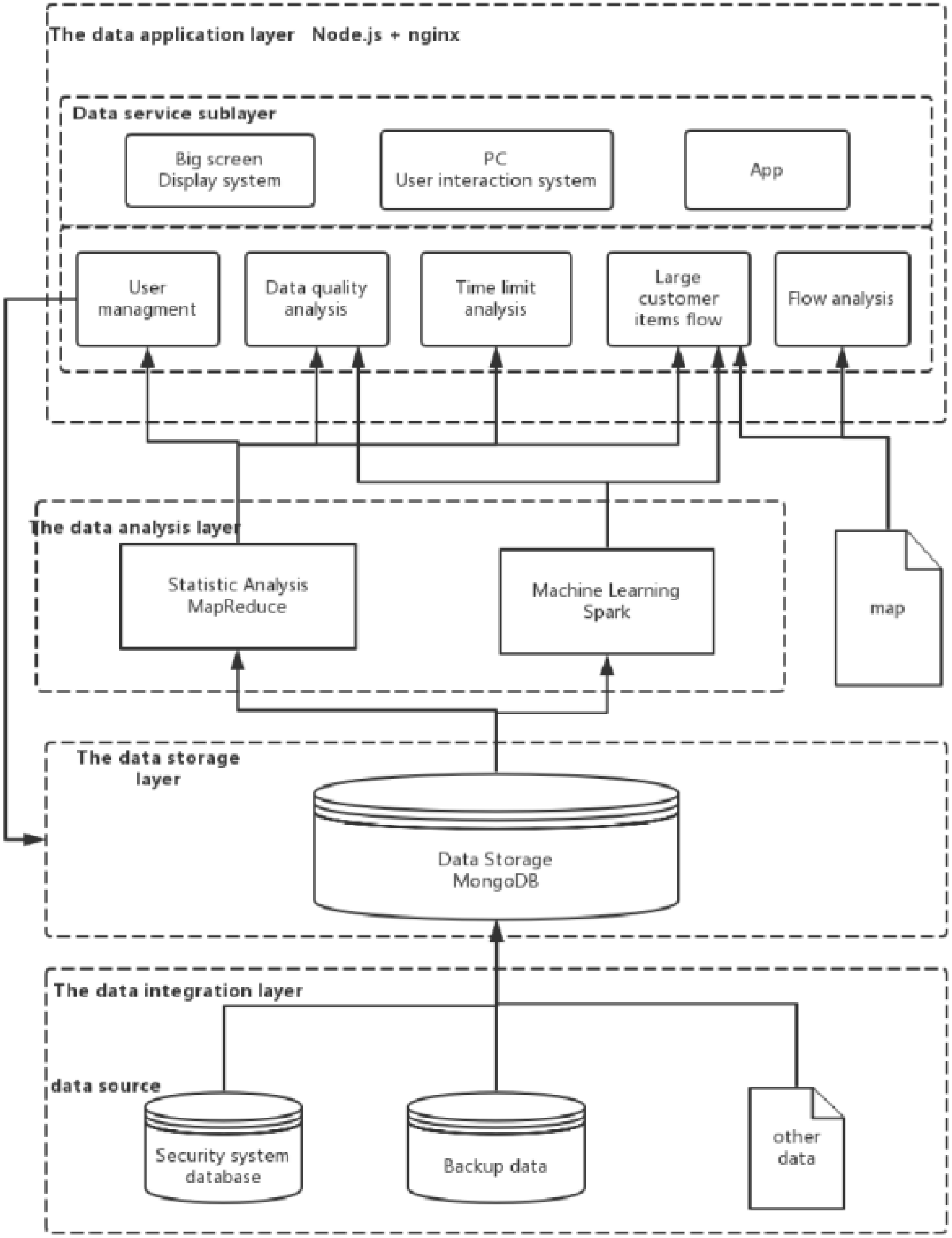


Figure 27. Example of an architecture in Node.JS. Adapted from (Liang et al., 2017).

Node.JS has an API for Input and Output (IO) operations. Chitra and Satapathy (2017) compared the IIS with the Node.js in terms of performance and concluded that Node.js might be more appropriate for IO systems and not for applications that require high CPU usage.

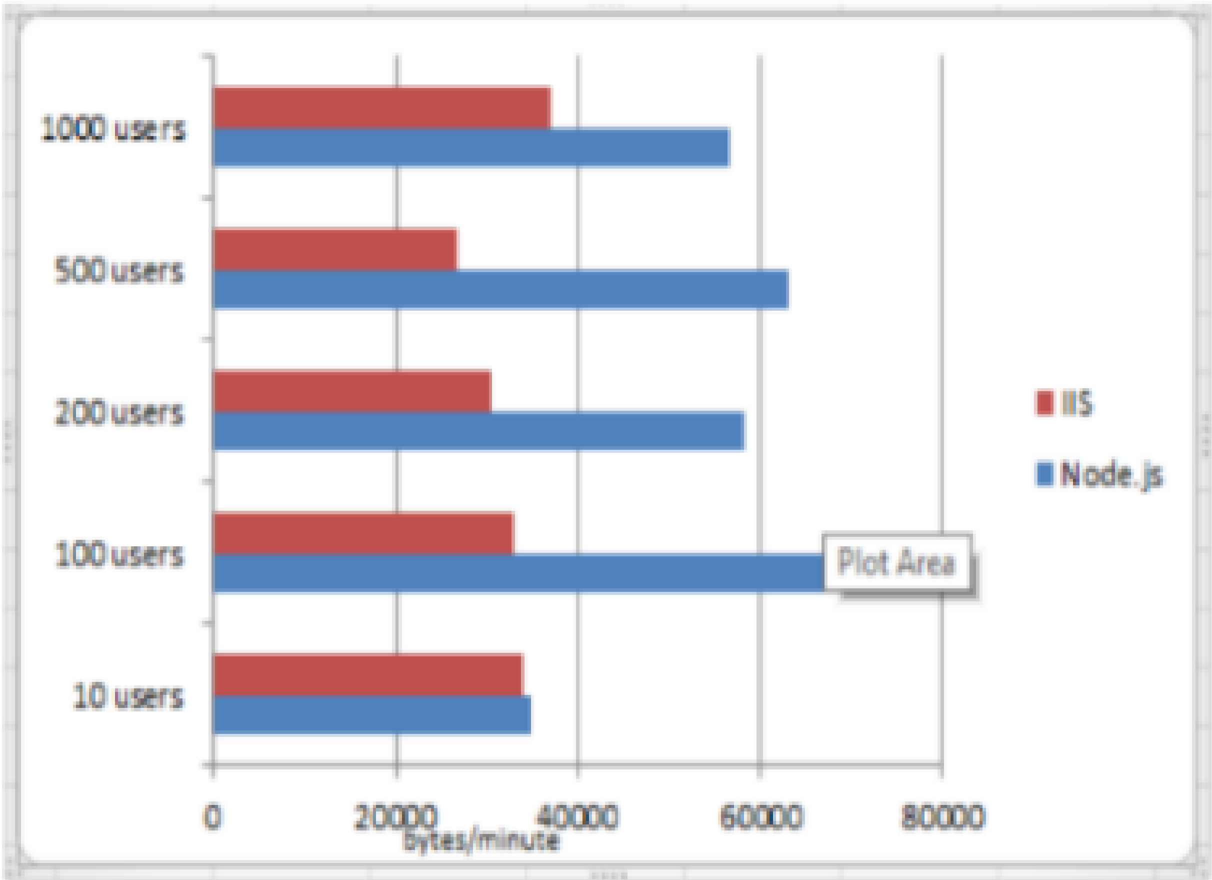


Figure 28. Speed comparison between Node.js and IIS (Chitra & Satapathy, 2017).

When a HTTP request arrives for a resource in the server, the HTTP.sys will ask the Windows Activation Service information from the configuration store (applicationHost.config) and details will be provided like the application pool and site configuration, the World Wide Web Publishing Service will use this information to configure the HTTP.sys. The Windows Activation Service will start a worker process for the pool in cause and returns a response to the HTTP.sys.

Chapter 3

Mixed Reality and Virtual Systems Overview

3

This section shows the work performed to implement artificial intelligence mechanisms on the systems and improve the virtual elements when interacting with the user and the real world and also how to refine the display of information for projects based on AI for Computer Vision (CV).

The process of employing AI on MR applications can be analyzed in many perspectives, on the proposed solution, the collision between virtual elements with real world objects was implemented via known processes like Vuforia. The behavior of the AI agents is responsive to the user input which can be motion using triggers like distance proximity and sensors like the camera that treat the information of the surroundings. Several prototypes were developed, such as, a drone gesture controller that collects the vital signs, a mobile MR application with flexible AI, responsive motion controllers for usability interaction (hand motion, eye gaze and body movement) and a Web MR application for health care and training.

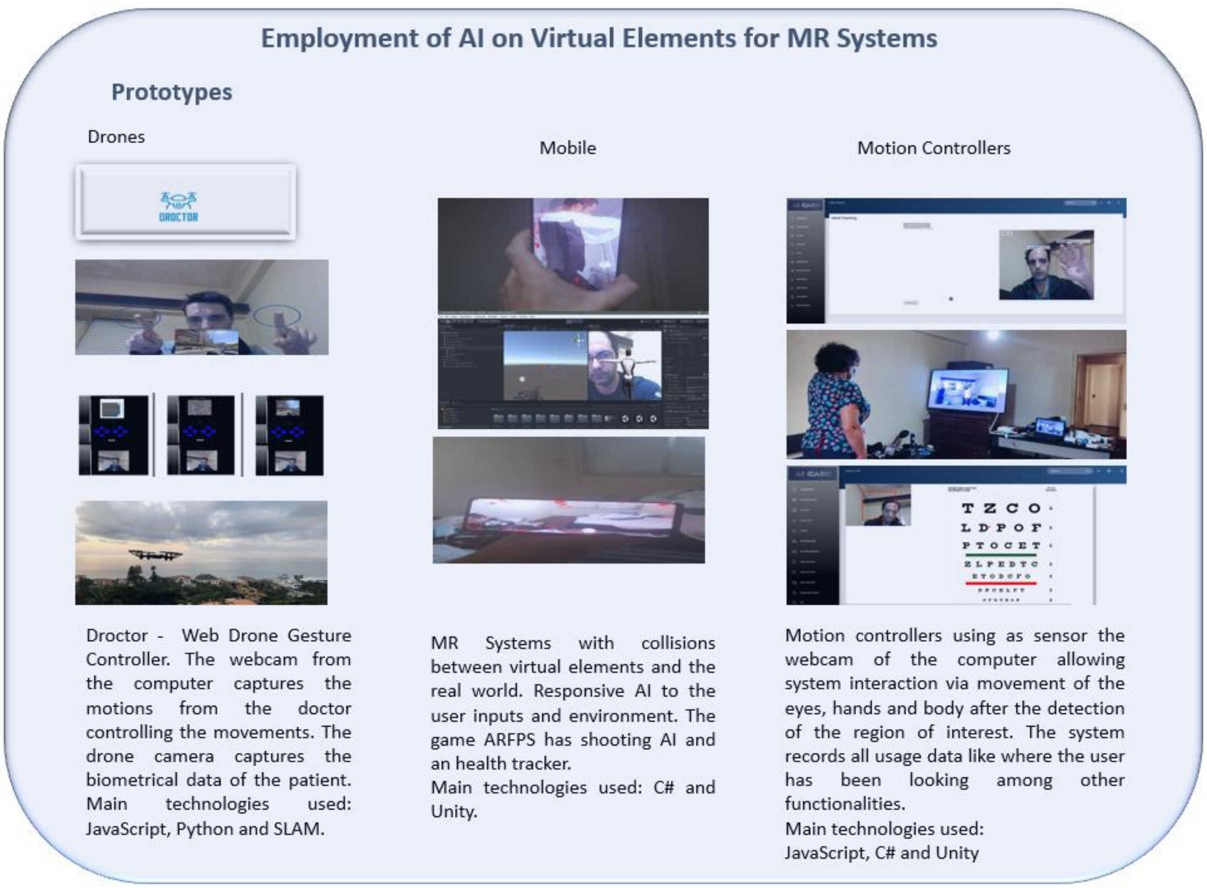


Figure 29. Prototypes of MR system where the augmented elements have AI mechanisms implemented.

With the objective to improve the performance and functionality of the Information Systems (IS), the dissertation aims to a strategy of developing the AI of the virtual elements on Mixed Reality (MR) systems in order to make the applications more useful and efficient for many areas like healthcare, gaming, teaching, robotics, Unmanned Aerial Vehicles (UAV), physiotherapy, among others.

In order to implement MR in mobile devices, the type of operating system of the user has an important factor. Some of the prototypes were developed for Android only, for which the IDE and libraries of Android Studio were used.

Vuforia allows projects in Unity to have 3D virtual elements without markers and facilitates the process of making collisions between the virtual elements and objects of the real world. Some of the projects used free assets from the Unity and Unreal Engine stores.

Concerning the usability and accessibility of MR applications for patients with limitations like visual auditory, physical, speech, cognitive, language, learning and neurological, this study focuses on 3 types of motion controllers: hand controller, eye gaze and body motion (head, hands, arms, feet and legs) allowing the applications to be used by people with certain disabilities for determined handicap degrees.

Figure 30. Developing applications for different types of disabilities.

For this study, several applications were developed for several incapacities, the “Droctor” web drone controller is able to be used in almost all cases because a doctor controls with gesture via web, a drone that captures some of the viral signs of the patient without contact and without the need of the user to perform any actions.

Concerning the visual disabilities, a hand motion controller was implemented in several applications as well as a chat bot for speech recognition.

Users with movement and earing problems can use the Aerial Control and the Eye Gaze responsive system. Several strategies were followed like a Mixed Reality (MR) application for mobile phones that allow interaction with AI augmented agents that respond to the movements of the phone and have collisions with obstacles of the real world. Some of the artefacts developed require the installation of specific libraries for the development activities and in some cases the presence of specific hardware.

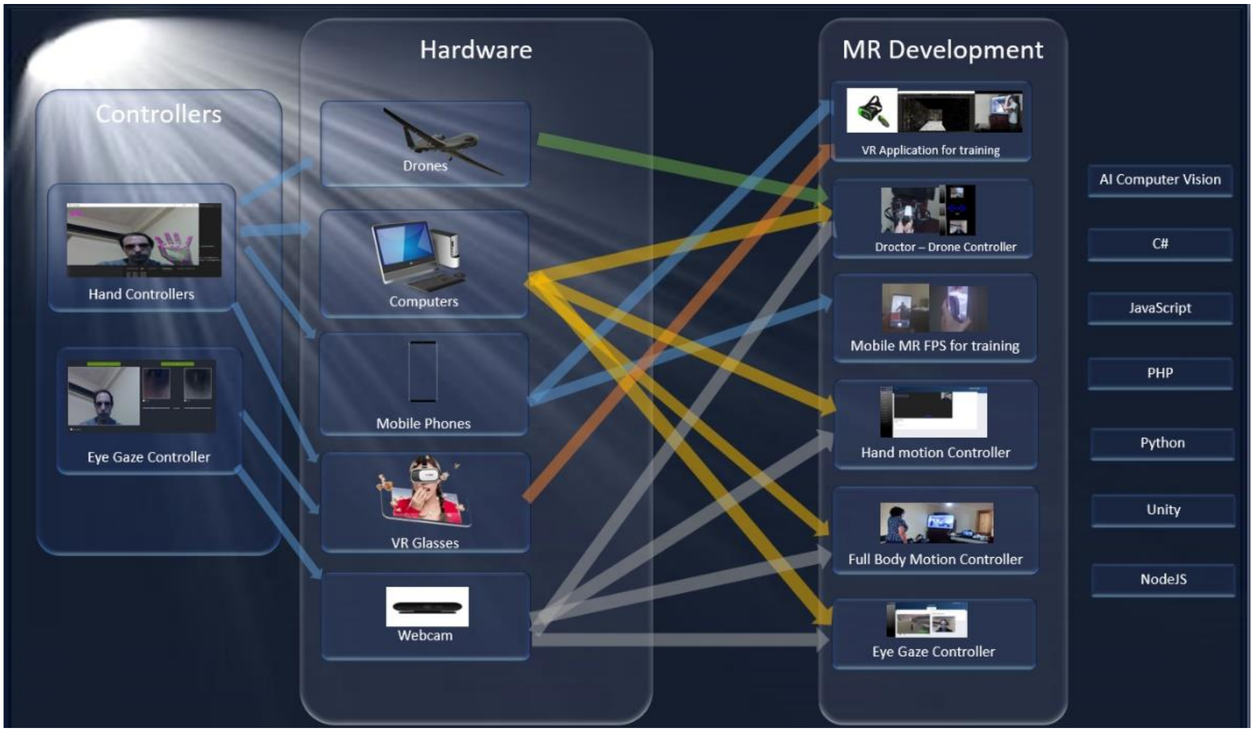


Figure 31. Schema mentioning the hardware required for the usage of the prototypes.

The concept is based on using real world data obtained by the camera and treat the information in order to provide new functionalities.

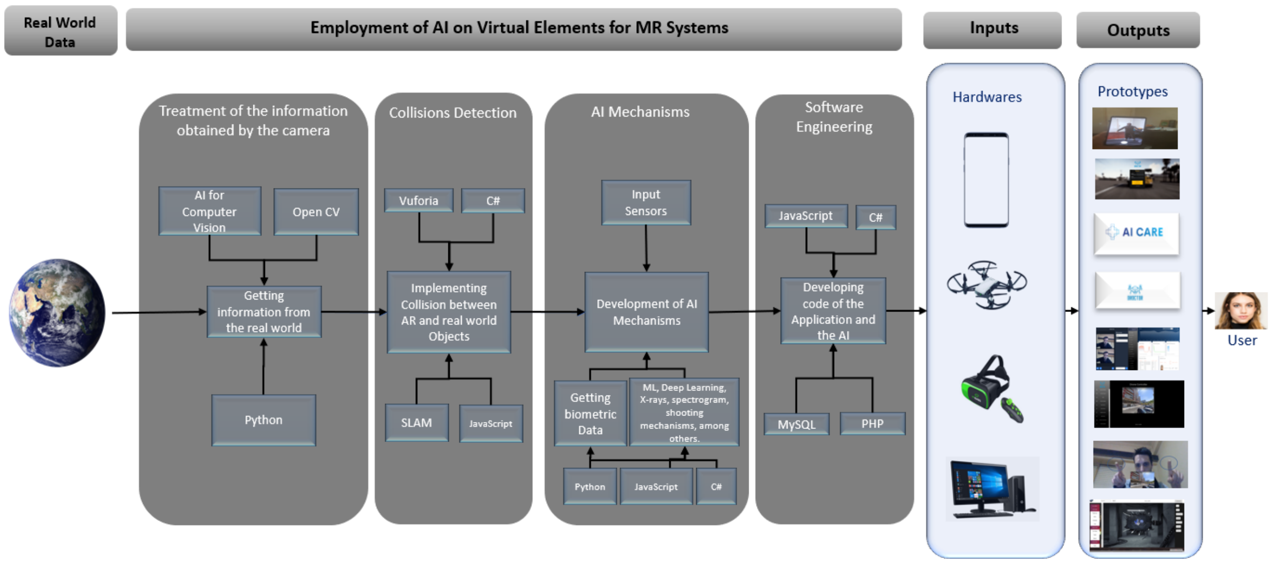


Figure 32. Placing AI mechanisms on the Information Systems.

Using the camera as a sensor of an AI system can increase the number of functionalities and usability of mixed reality systems, in order to develop these type of systems specific libraries like the OpenCV or the WebGazer.js may be required. With the objective of making the applications more robust, the collision detection with physical objects assume an important factor especially when implementing a drone controller or a mobile augmented reality (AR) first person shooter (FPS).

For this project some of the information was stored in a MySQL database, the web camera, phone camera and drone camera were used as sensors and several programming languages were required like C#, python, JavaScript among others.

AI for computer vision techniques were implemented in order to achieve certain functionalities like detection of the region of interest (ROI) or signal processing.

3.1 Employment of Artificial Intelligence Mechanisms on Mixed Reality Applications for Mobile Phones

Concerning the problem of avoiding collisions between the virtual elements generated and the objects of the real world was used an AI mechanism called simultaneous localization and mapping (SLAM) which is possible to obtain via the Vuforia engine for Unity.

The implementation of MR in mobile devices has many aspects to be considered namely the type of operating system. The prototype used to defend the concept is Android compatible, so the IDE and libraries of Android Studio were installed for development purposes.

Vuforia allows projects in Unity to have 3D virtual elements without markers and assets from the unity store were imported to develop a system where the computer-generated images are correctly placed in the LCD and with collisions with the physical objects, as a possible way to achieve the first objective of this research.

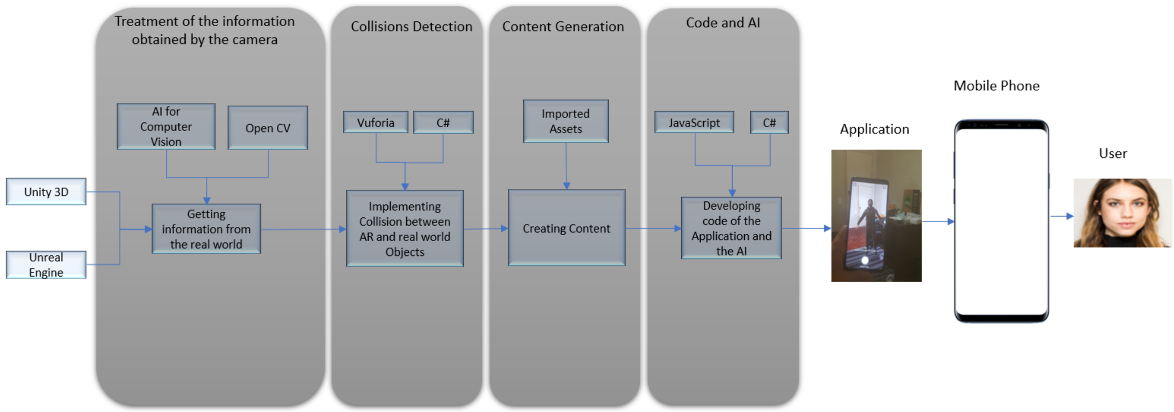


Figure 33. Use of Unity engine for MR.

Concerning Unreal Engine, the MR can be implemented via a combination of visual scripting with C++ and also allows the importation of libraries to help the development process.

With the objective to improve the performance of the usage of mobile phone applications, a strategy of making collisions between the virtual elements and the real world with a responsive AI to the several scenarios was developed. The image obtained by the phone camera appears in a 3D plane with associated C# scripts, on the Unity engine.

During development it is possible to test the image appearing on the panel by using the web camera and when the final solution is compiled and the .APK file created the panel will render the images obtained by the phone camera.

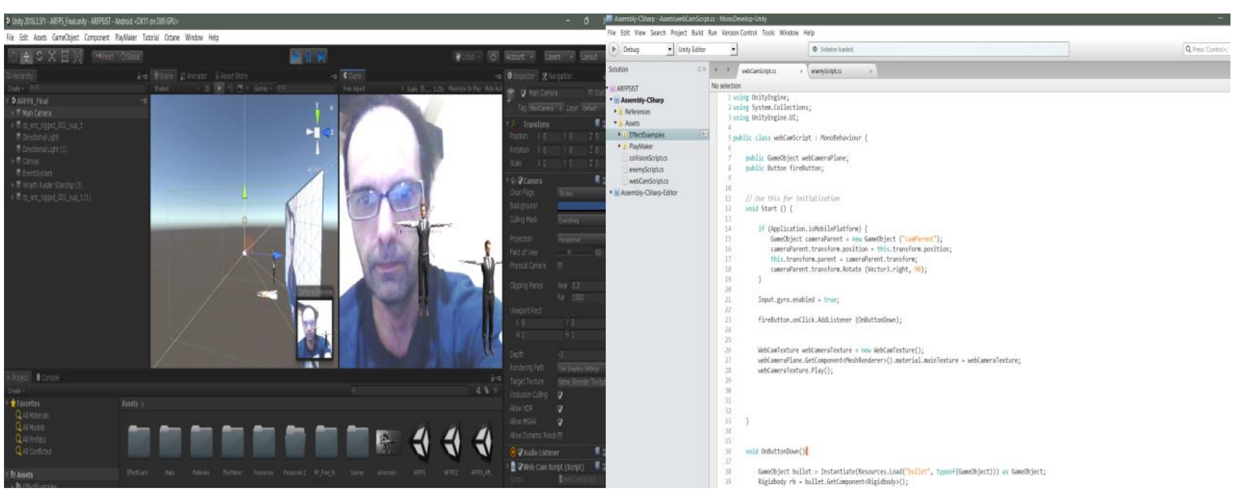


Figure 34. Source code to place a background texture with what the phone camera will capture.

For the objective of developing a demonstrative prototype, a First Person Shooter (FPS) in MR was implemented for the phone with shooting AI agents that followed a methodology based on using a box collider to delimiter the AI Area, where the AI agents can move without going too far from the phone camera view, as demonstrated on the picture below, the camera of the unity engine points to a plane, this plane used a C# script written in MonoDevelop that allows to set the parameters to calibrate with the real-world environment. In order to get these results were used 3D programming techniques, like transforms, quaternions and textures.

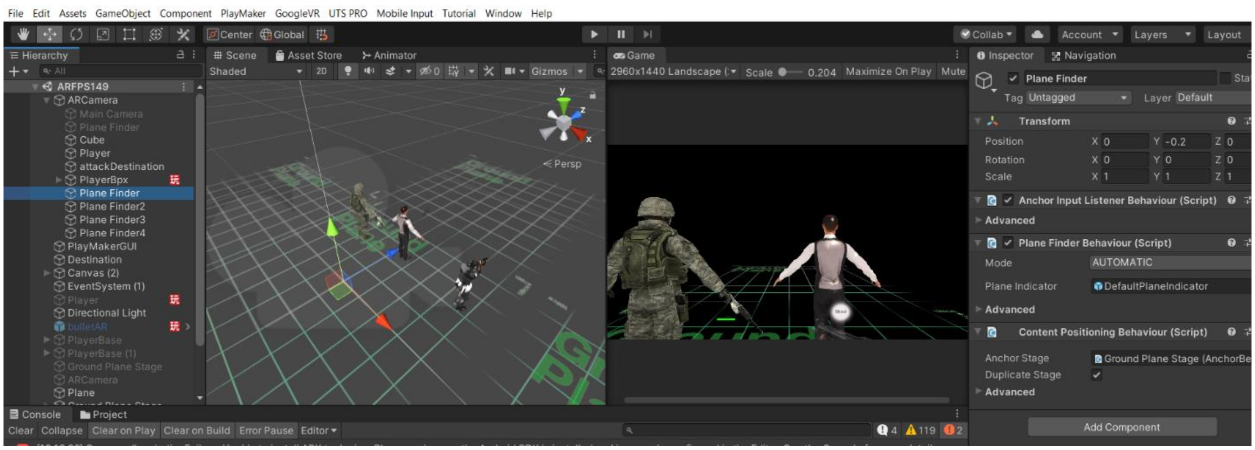


Figure 35. Use of Unity to setup the camera and placement of AI agents.

For the HUD (Heads Up Display) the unity canvas is the parent node, having associated images that interact with the user, like the weapon and the fire button. Concerning the movement coordination this was scripted via C# in Unity with the help of the class Gyroscope. The AI agents have several scripts associated like the collision script that allows to play animations once they collide with another object and other techniques like the unity Coroutines which allow to wait for a certain period of time facilitating the process of correctly placing virtual objects in the real-world environment captured by the camera.

There are two known possible processes among several of implementing MR systems, one via Unity with C# and another using Unreal Engine and C++. To implement responses of the AI agents like the enemies of a FPS game, a scripting process was followed combined with animations that respond adequately to the inputs of the user, some variables were taken in consideration like time and distance, making the interactions in a sequenced manner.

The bullets use a ray cast function taking on consideration a certain direction and speed in the 3D world coordinates, when the bullets collide against certain type of object, a trigger mechanism is activated playing the correspondent animation depending of the type of object that they collide with.

3.2 Collision Detection Between the Real-World Objects and the Augmented Elements

Vuforia engine was used in order to obtain the AR Camera associated to the Unity Engine which allows the use of simultaneous localization and mapping (SLAM) and create occlusions with augmented and real image elements.

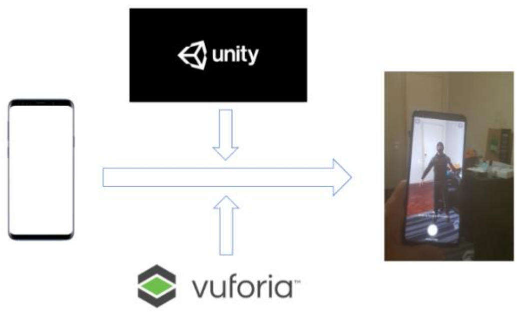


Figure 36. Use of Vuforia to correctly place AI agents in the scenario captured by the phone camera.

With the use of Vuforia AR Camera and ground plane detection it is possible to place correctly the computer-generated image in the LCD taking in consideration the 3D physics of the environment captured by the phone camera.

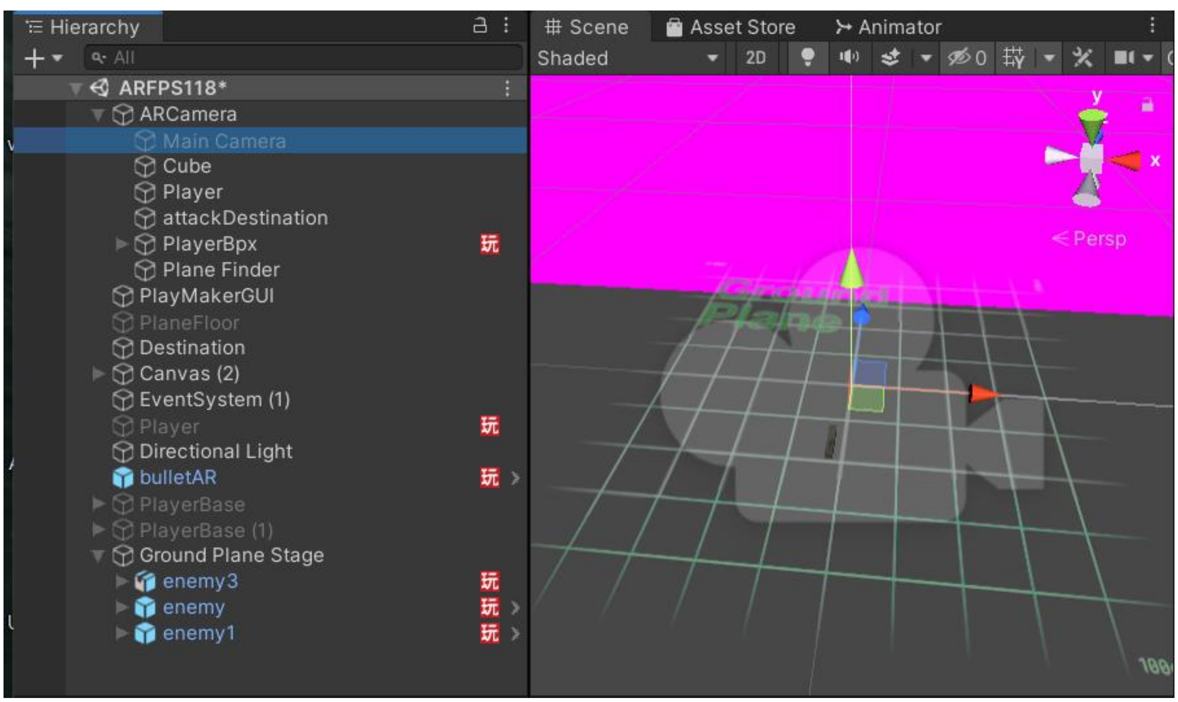


Figure 37. Use of Unity to setup Vuforia AR camera.

Via the Unity IDE it is possible to configure the camera settings and the lightning and also to implement AI mechanisms like the behavior of the virtual elements among other possible functionalities. The programming language used to develop the prototype was C# and some libraries and classes were imported in order to implement all the features of a mixed reality first person shooter for mobile phones compatible with some of the operating systems of the android family.

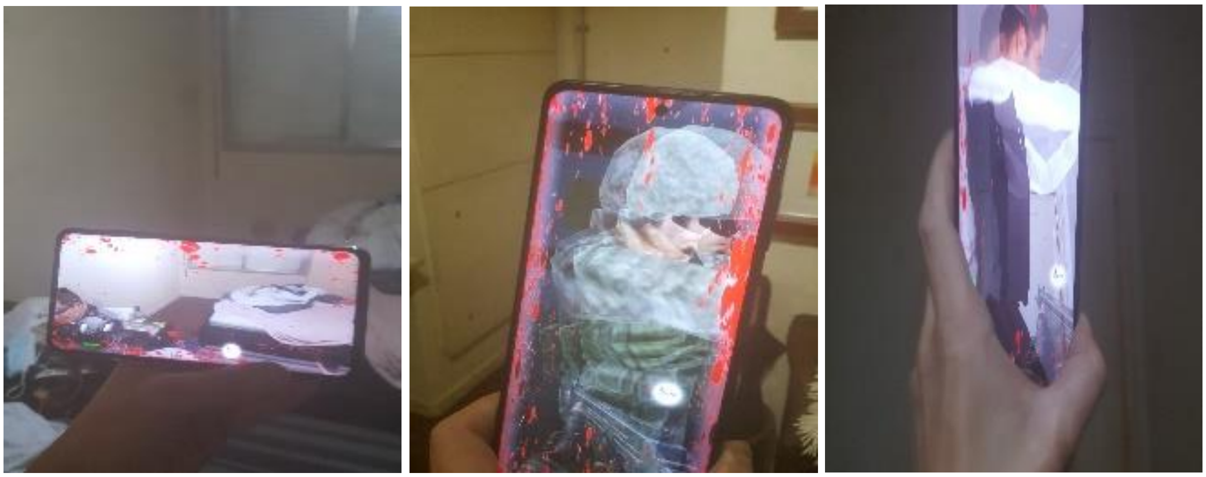


Figure 38. Preview of the prototype ARFPS.

Concerning the shooting system, a ray cast function was used taking in consideration direction, speed, position and distance and also spawner game object mechanisms in order for the bullets to appear after the verification of certain conditions like when the user presses the shooting button or when the AI agent or enemy is at a certain distance of the main player.

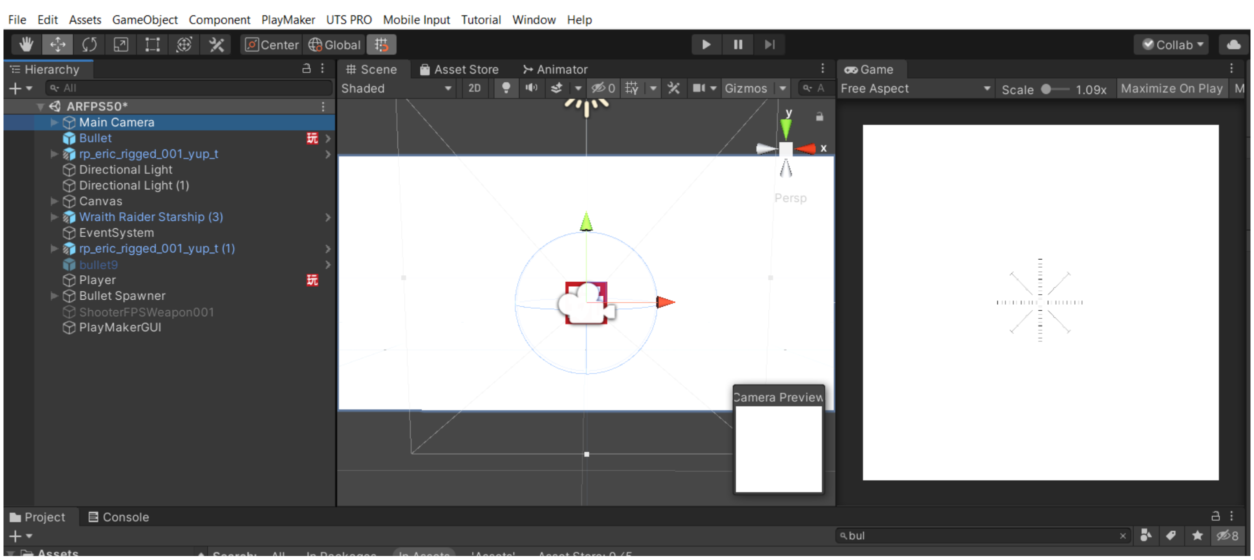


Figure 39. Placing the camera in the Unity Game Engine

The animations follow a set of conditions and sequences in order to be activated, the picture below shows a schematic of the necessary order and trigger mechanisms associated to the animation transitions process during gameplay.

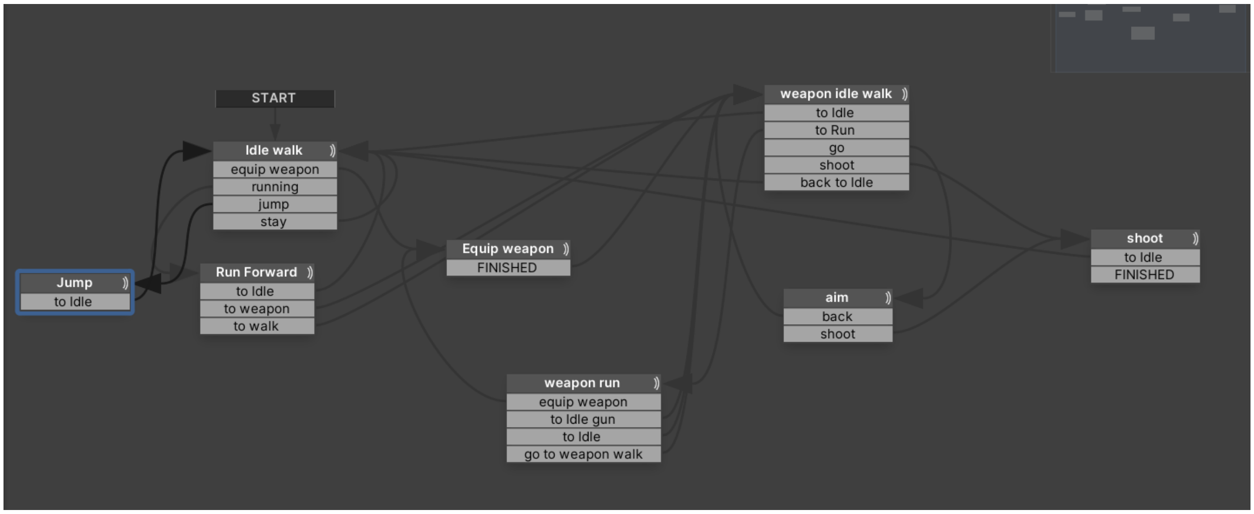


Figure 40. Making animation transitions in Unity

The animations follow an AI mechanism in order to provide more functionalities to MR systems as demonstrated in this prototype, it is possible to notice that the virtual elements can have more functionalities and interaction with the users. In the future will be possible that applications may have more uses from the real-world data captured by a sensor, the infrared (IR) cameras for example can provide night vision.

The process of treating information has been improving and the AI strategies nowadays are able to treat huger amounts of data making predictions with a high rate of accuracy.

The presented developed prototype illustrates a process of responding to the second research question by demonstrating a set of processes to achieve the proposed objective, for this demonstration was used the C# programming language and the Unity game engine among other technologies.

3.3 Employment of Artificial Intelligence Mechanisms on Mixed Reality Systems for Healthcare

The disease diagnosis process can be very complicated, many factors should be considered like the vital signs data on a primary care stage. With the objective to improve online healthcare, AI mechanisms can be combined with real world data displaying the information via virtual elements correctly paced in the images captured by a camera.

In order to measure the heartbeats during gameplay an application was developed that measures the heartbeats in real time with the use of Open CV and the Python programming language that is based in AI for computer vision mechanisms.

The virtual elements that appear on screen have AI mechanisms allowing to display the cardiac frequency of the user. The results from the webcam detection were compared and tested with the ones provided by a clock sensor and medical devices like an oximeter and a pulse blood pressure monitor.

Figure 41. Heartbeat detection and comparison with a clock and medical devices.

Via AI processes, it is possible to obtain an approach of the values of the vital signs of a patient by using a camera as a sensor, the results can be displayed via a MR system where the virtual elements have specific AI Mechanisms being displayed correctly in the screen after ROI detection.

The prototype developed allows to track the face of the user first and afterwards will report the heartbeat in real time. The picture below is the heartbeat measurement on a web application with the use of Open CV and the python programming language.

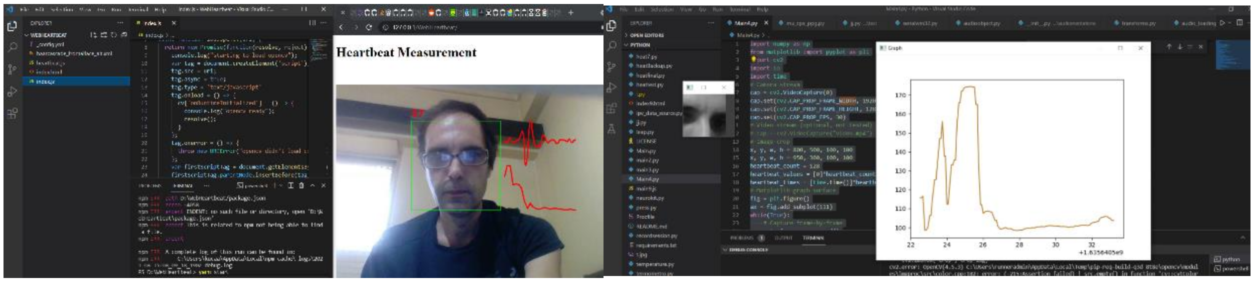


Figure 42. Heartbeat detection and comparison.

The breath rate can also be obtained via AI for Computer vision processes treating the Photoplethysmogram (PPG) signal obtained by the infrared (IR) sensor, which on this prototype is the web camera.

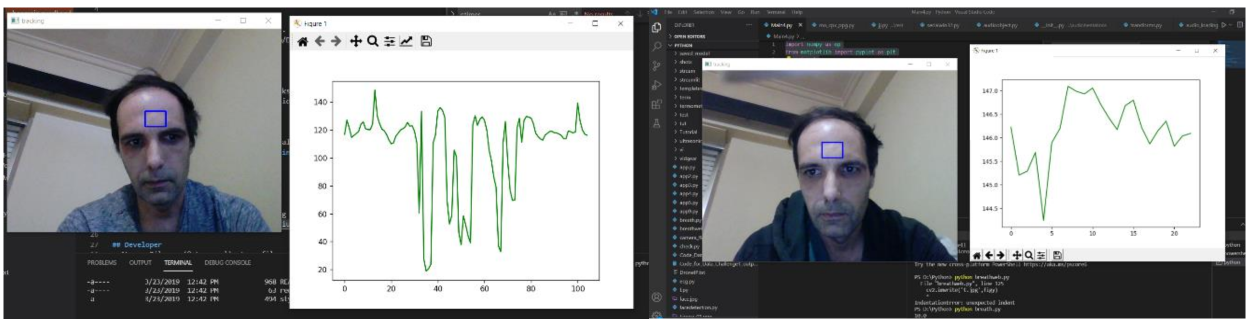


Figure 43. Breath rate detection and comparison.

This value can be obtained by counting the number of pikes, maximum in the heart rate, every time the heart rate gets a maximum means you have inhaled, showing the number of breads per minute.

Below is a snapshot of the execution of the program prototype. It is possible to see, there are low frequency waveforms mixed with high frequency sawtooth (small spikes).

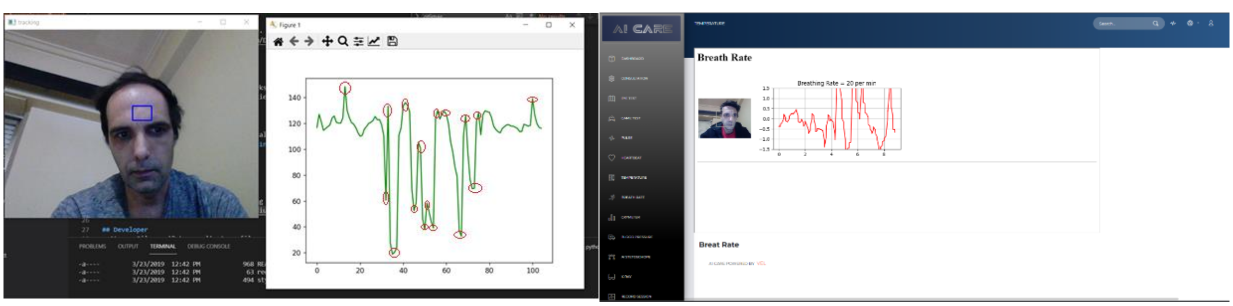


Figure 44. Breath rate obtained via the pikes of the reading of the cardiac frequency.

It can be seen that each small peak corresponds to a pulse beat (heartbeat). And the great peak-valley transition corresponds to the respiratory cycles.

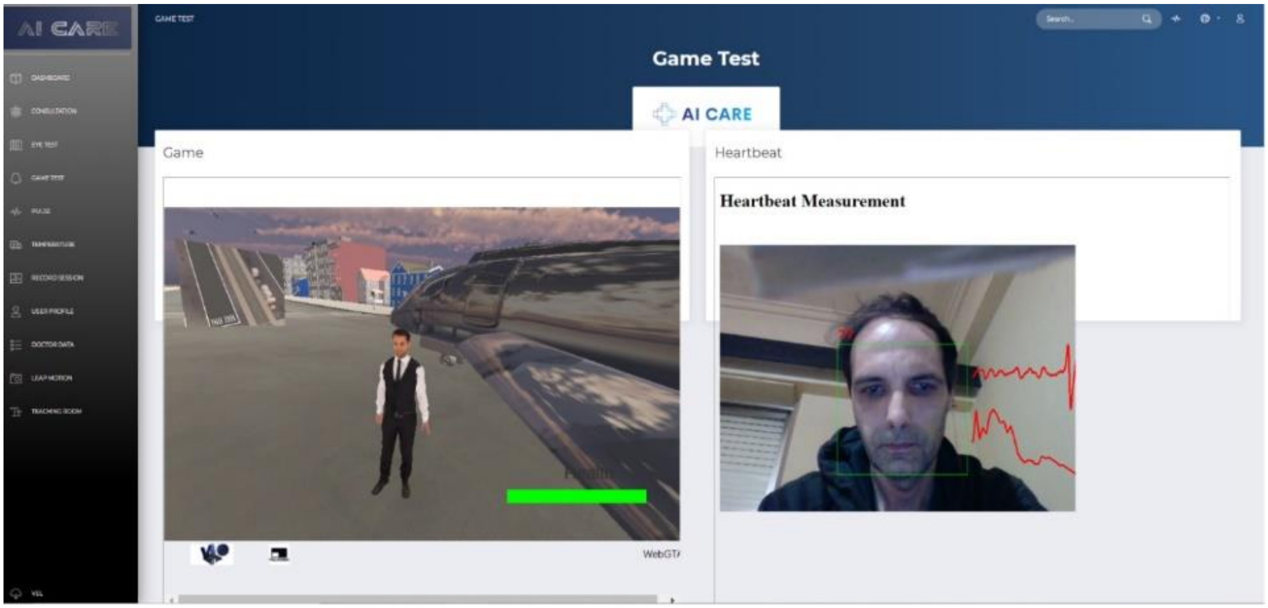


Figure 45. Web application to obtain the cardiac frequency data during gameplay.

Mixed reality systems can have AI mechanisms on the virtual elements allowing to perform functionalities for different areas like healthcare.

The developed prototype intends to provide a possible answer to the first research question of the investigation in cause showing an example that can be used for healthcare activities.

3.4 Virtual Reality Implementation for Training and Usability

Concerning the hardware required for virtual reality testing, this research has focused on glasses that allow to insert the mobile phone, the mobile applications for these devices usually are divided in two screens, one for each eye.

Figure 46. Virtual Klack 3D VR and VR Glasses Smartphone WPS, 3D, Bluetooth with remote.

Table 3. Details of the VR glasses used

Virtual Klack 3D VR	VR Glasses Smartphone WPS, 3D, Bluetooth with remote
Supported systems: Android and iOS.	3D Virtual Reality Glasses for Smartphone;
Maximum allowable phone size: 16.5 x 8.5 cm	Compatible with Smartphones 4.7" - 6.5";
Ideal for phones ranging from 11 to 15 cm.	Field of View: 95º, Adjustable Focus and Distance;
The visor is made of ABS plastic and polycarbonate.	Smartphone Requires Gyroscope Sensor;
	Velcro Headband with Adjustable Measure;
	Bluetooth controller, 195x115x135mm.

Virtual Reality applications have the ability of providing the sensation of immersion, usually the VR glasses combine a process of blending two images, one for each eye allowing for the user to see a 3D virtual world. On a perspective of accessibility can help for some cognitive, auditory and physical disabilities.

The VR devices might have a remote or not, without a remote a virtual laser bean can work as a pointer generating interaction with parts of the application that are programmed to perform certain functionalities, using collisions triggers. The Google VR SDK was used, which allows to provide the VR format for unity. As demonstrated in the figure below in order to generate the effect of the immersion and for the application to be correctly seen when using this type of VR glasses described, the application has to be divided in two screens the first for the left eye and the second for the right eye.

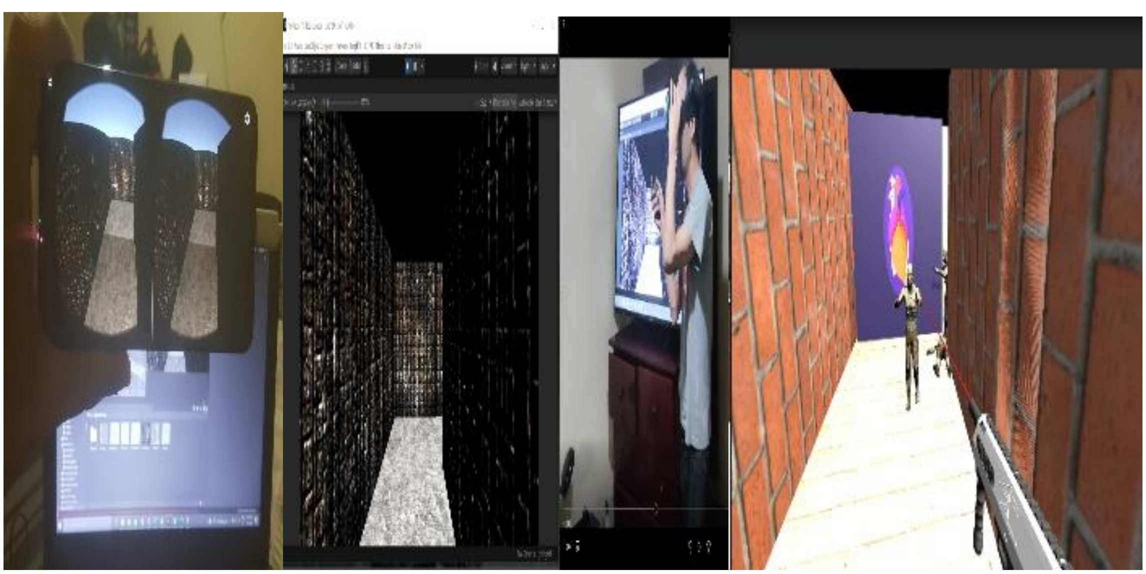


Figure 47. VR App on the mobile phone before using glasses; baking the AI area, usage of the VR glasses, game for training that uses a laser bean based on head movement and when the mission is accomplished shows an educational video on how to proceed in case of heart attacks.

In terms of usability, VR technologies have showed relevant results for training activities, some VR applications simulate an operating room, showing doctors how to proceed concerning certain types of surgeries. For some functionalities under some contexts VR can help to achieve the objectives in a more interactive and efficient way.

The VR application developed has the objective of providing training where videos explaining how to proceed for certain clinical emergencies are played after successfully completing the game, aiming for an innovative way of providing formation.

With the objective of making the applications more interactive and accessible to people with auditory and visible disabilities and to improve the performance and functionality of the Information Systems (IS), the dissertation shows a strategy of using another sensor instead of a camera, which in this case is the leap motion controller that allows to detect and place on the screen the hands of the user.

There has been an increased use of VR technologies for training activities, the implementation of AI mechanisms allow to provide more functionalities to the applications.

Figure 48. VR Leap Motion Controller. Features highly accurate V4 hand and finger tracking. Offers a 135° field of view and up to 80 cm range. Tracks objects and captures high-speed infrared footage Interacts directly with digital content, VR & AR apps.

A system that uses this type of sensor allows the user to interact with the application via alternate processes than the traditional computer and keyboard allowing the possibility of being used for different type of users and ages.

The VR application developed for this sensor allows the simulation of a surgery displaying the details of certain devices.

The purpose is based in simulating a real-world scenario in order to make the formation more appealing and interactive taking in consideration the risks of performing certain surgery operations.

Figure 49. Developed Healthcare Simulator with leap motion, for training purposes.

The objective is to facilitate the training of the doctors to perform a surgery, in this case the leap motion sensor allows to display the hands on the screen, allowing a different type of interaction making the application more appealing because some surgery operations require the use of the hands, which the movements can not be trained by using a mouse or typing a keyboard.

3.5 Artificial Intelligence for Computer Vision in Healthcare using Mixed Reality Systems

In order to improve the clinical process, the proposed solution of technologies aims to facilitate the diagnosis process, provide geographic information and in case of emergency, making the work of the physician simpler avoiding the current high costs on machinery required.

The Droctor application developed is a prototype of a drone web hand motion controller to obtain vital signs of a patient, uses AI mechanisms on the virtual elements and provides the location of the drone in Google maps.

On the dashboard it is possible to check all the patients associated to a certain doctor, but the graphs and the health values displayed are only of the chosen user that is going to be diagnosed. The vital signs data can be added manually and preference is given for having values for each month.

On the web application appears eight buttons that allow to move the drone, the images captured by the drone camera appear on real time and display the vital signs values as virtual elements after ROI detection which in the case of the cardiac frequency for this application is the face of the user, allowing the doctor to know where the patient is, get the vital signs without the risk of contagion at lower cost than traditional processes and without the need of relocating from one place to another.

The drone was tested more in indoor environments but can also be used outdoor. MR systems can have more functionalities by implementing AI mechanisms.

For the drone in cause, DJI offers libraries that can be imported in the code allowing to implement controller mechanisms.

The UAV used has a short flight time of around 13 minutes, which is enough time to perform the process of obtaining vital signs for one user, but other drones with higher capacities like cameras with more pixels, batteries that can last longer may show more efficient results.

Figure 50. Drone 720p Ryze by DJI Tello. Flight time: up to 13 min

Image transmission distance: 100m. HD 720p broadcast. Intel processor. Box Contents Accessories: Battery, Propellers, Propeller Guard, Compartment. Compatible with remote controls.

The drone application was done using a combination of several technologies like JavaScript, PHP, python among others, some operations required the importation of specific libraries like OpenCV and the DJI in order to let the implementation of functionalities such as getting the vital signals or implementing a drone controller.

The geo location used the Google Maps API that requires an account in Google Cloud and by IP is able to get the latitude and longitude coordinates which are presented in a map with a blue pointer, the type of map can be changed according to the possibilities supplied by Google Maps.

In order to start the drone, it is required that the drone has battery and that is connected to the computer which for the DJI model used was only possible via wi-fi. By pressing the button connect, the application will use the drone IP and after pressing the button take off, the drone starts flying.

A hand gesture controller was also implemented, providing two types of control being the other web buttons, on the same screen of the web MR application is possible to have the map, the controller and the images obtained by the drone camera reporting the vital signals if a user face is correctly detected.

The project aims to defend the concept of AI mechanisms for healthcare by providing an innovative approach when compared with traditional processes that have higher costs, are more laborious and take longer time to perform the activities avoiding also physical contact.

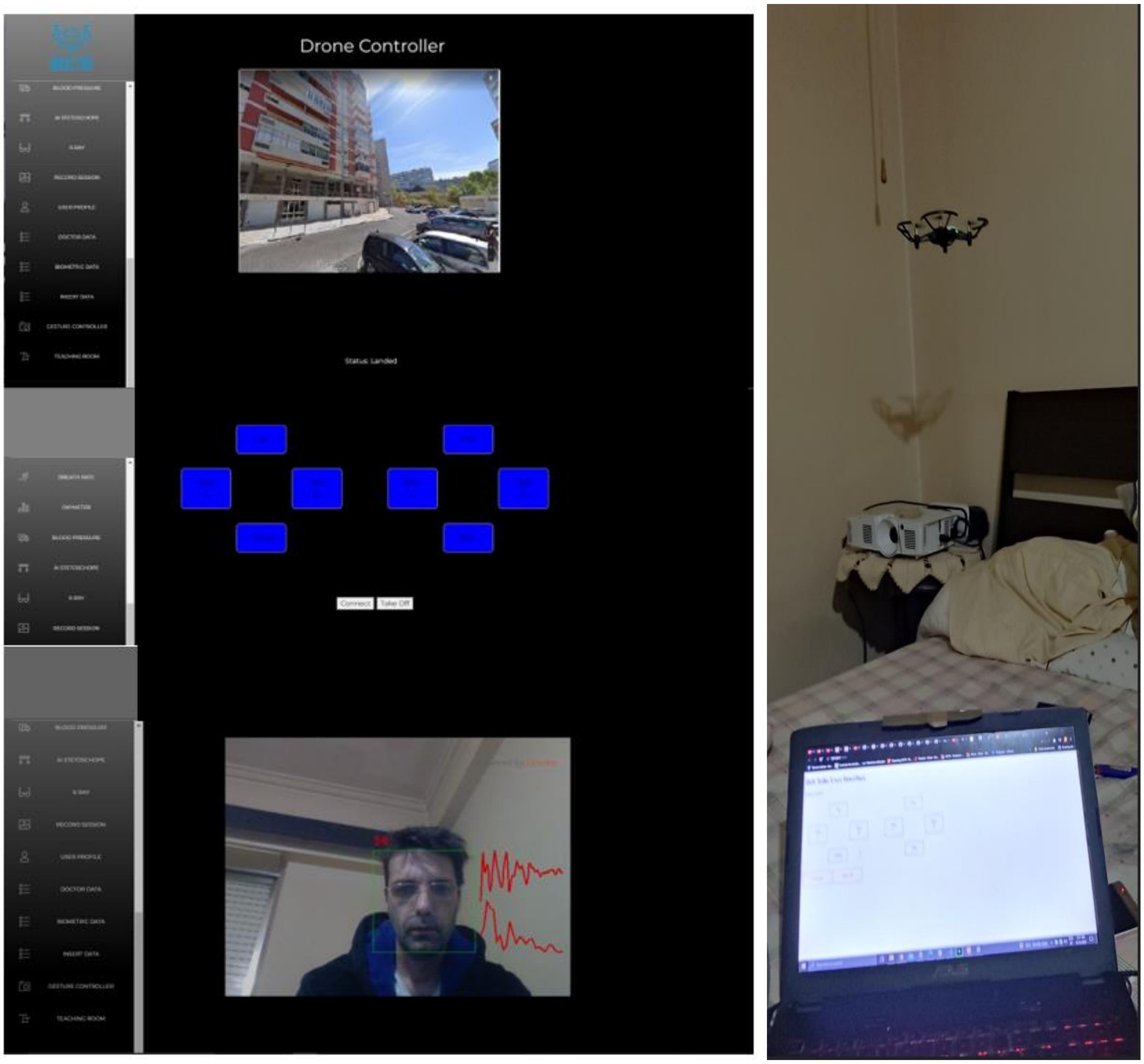


Figure 51. Developed prototype “Droctor”. Controlling a drone to obtain the vital signs with geo location.

After getting the vital signs the application also offers the possibility of viewing the medical records for each patient showing the data for different times which can complement the current clinical record of a user helping the doctor in certain activities like diagnosis or determining the severity health status.

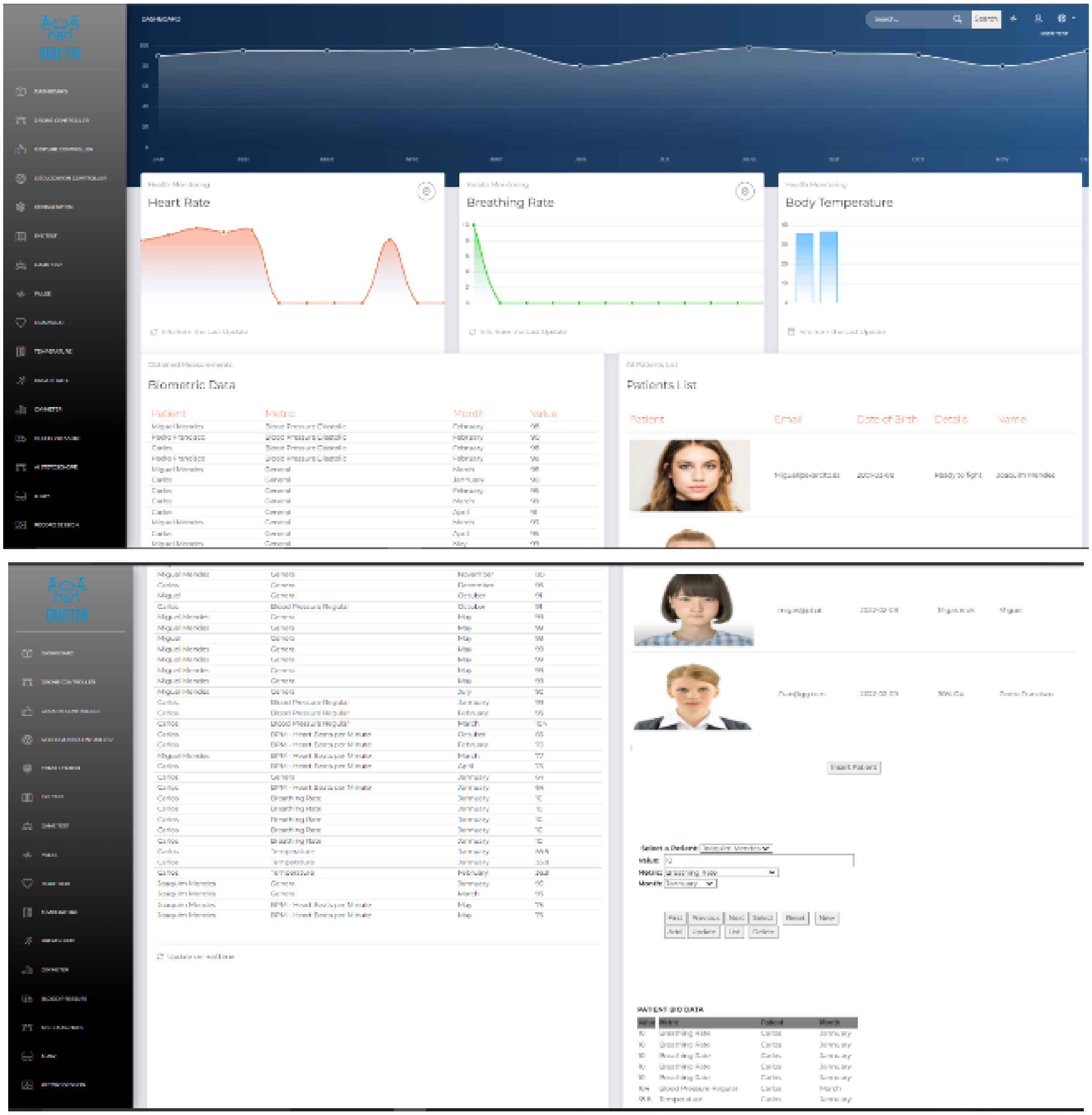


Figure 52. Dashboard of the developed prototype, the "Droctor" application and the possibility to edit the data of all the patients.

The access is encrypted and the user management can be done by doctors and administrators, allowing to correct values and also to change the vital signs values obtained via an automatic process.

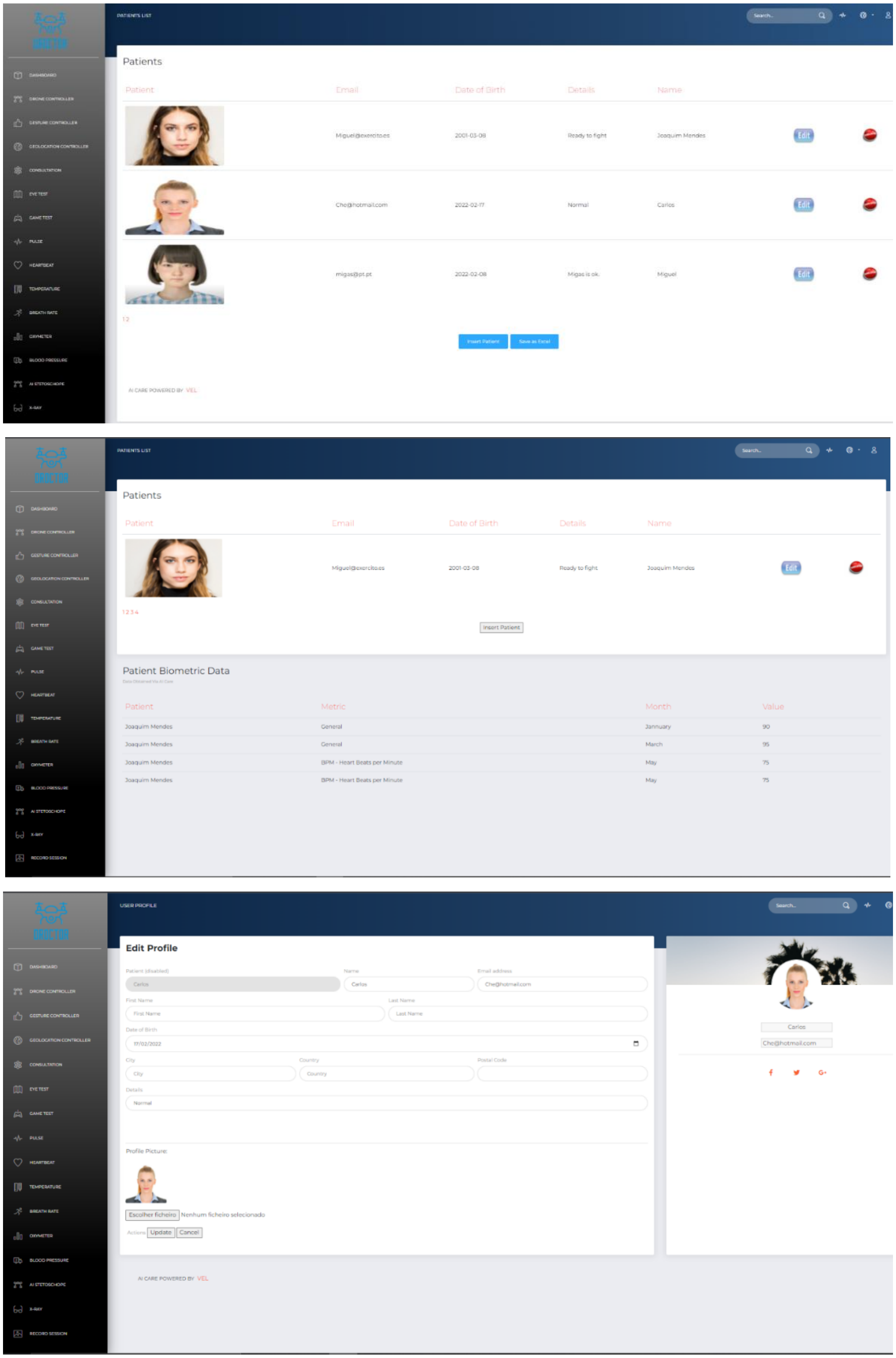


Figure 53. List of users. List of vital signs data for each user and form to edit general information of the patient. Form to change health values of just one person and possibility to record a session that was developed mainly in Python with Flask combined with JavaScript.

The system has several contents to facilitate the primary care diagnosis like recording sessions, playing a serious game to get the important rest state of the patient. The visual test can be done with an eye gaze controller, helping the healthcare professional to know where the user is looking during a virtual consultation. A gesture control was also implemented to facilitate the usability of people with visual disability or without a mouse or keyboard among other possibilities. Another feature is the multi video chat room allowing senior doctors to teach skills and how to use the system. For the pulse measurement, an API receiver mechanism from the developed mobile application in java to get the cardiac frequency via infra-red (IR) camera was implemented.

The video chat constructed mainly via NodeJS allows to interact with the patients with mixed reality (MR) and obtain the vital signs data on real time, helping the doctor to decide if an emergency procedure should be triggered. The health state values are obtained on real time, and the video chat gives the possibility to the doctor to provide instructions to the patient.

The drone can facilitate the process of reaching several areas, providing important information like the location coordinates and the vital signs of a user, the research would allow an innovative approach of performing general clinic activities when compared with traditional methods and can be very helpful for situations of contagious diseases or cases that require constant monitoring.

The AI mechanisms of the virtual elements can be used to control drones and also for reporting vital signs of a user after ROI detection.

3.6 Virtual Chat

With the objective to improve the performance and functionality of the Information Systems (IS), the research shows a strategy of implementing MR chats and use of chat bots which can be very useful for people with visual disabilities and providing clinical information in a more interactive process.

Due to the nature of the project, the communication can be done in several ways.

A video chat was developed in C# and Unity in order to make possible the communication between two users in a virtual system.

Implementing chats in a MR application may make easier the process of achieving certain objectives, such as, the communication between a health professional and a patient to perform certain activities like getting the vital signs or provide instructions on real time to perform certain tasks.

Figure 54. Virtual chat.

Virtual chats, may allow to make conversations while showing the surrounding environments, which can provide more information.

Conversations with machines can be viewed from different perspectives. In the case of accessibility, there may be a need to create a speech recognition system, generally based on natural language processing.

In AIML (Artificial Intelligence Mark Up Language) it is common to define a default value, if the user enters values that are not in the knowledge base, the AI agent responds a message asking if the user can repeat the data input process.

Speech recognition systems have been evolving, although it is still noted that accents, hoarseness and other factors, such as the capacity and quality of the microphone, can make the recognition process difficult.

In the photo below a possible architecture of using AI Agents for speech recognition with the Amazon APIs.

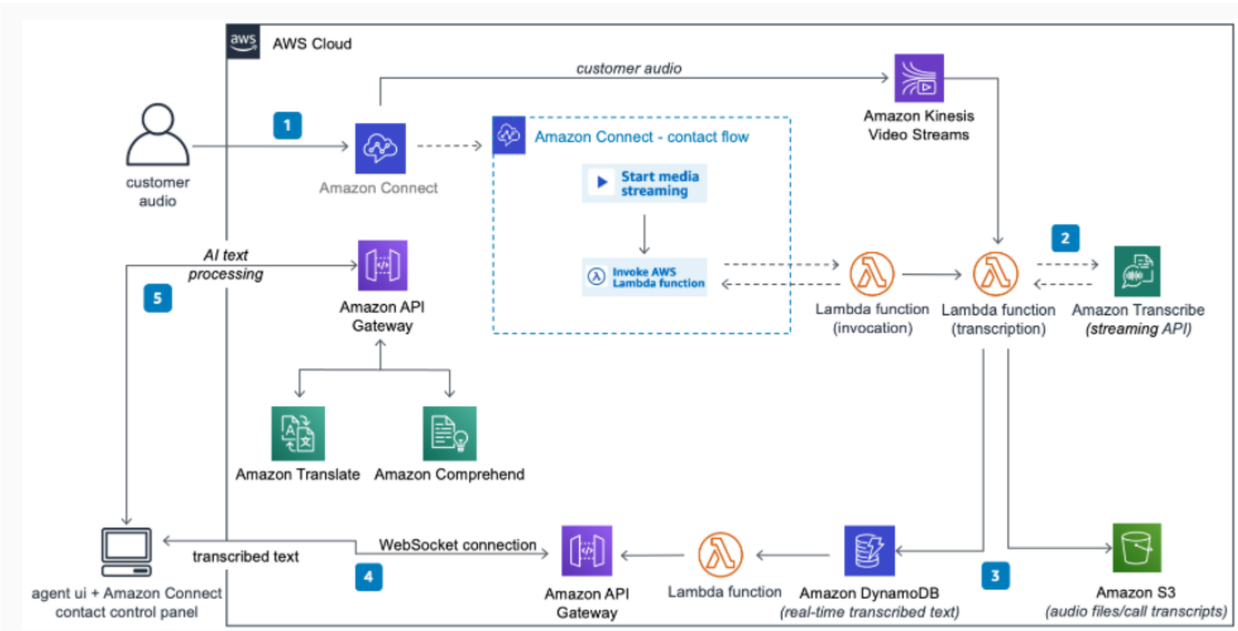


Figure 55. Using AI agents for speech recognition with Amazon Web Services.

AI agents, in some cases, respond to the values obtained by sensors. DialogFlow, among several possibilities, allows the usage of keyboard and microphone as input and the screen and speakers as output. For speech recognition, a chat bot was developed with the DiailogFlow from Google, although it has some limitations.

The answers have to respond correctly to what the user inserted and requires long databases and storage capacities, making difficult the implementation process of a full conversation. Many companies have been performing research on chat bots and recommender systems, the use of natural language processing (NLP) has improved this process but there still are many research challenges.

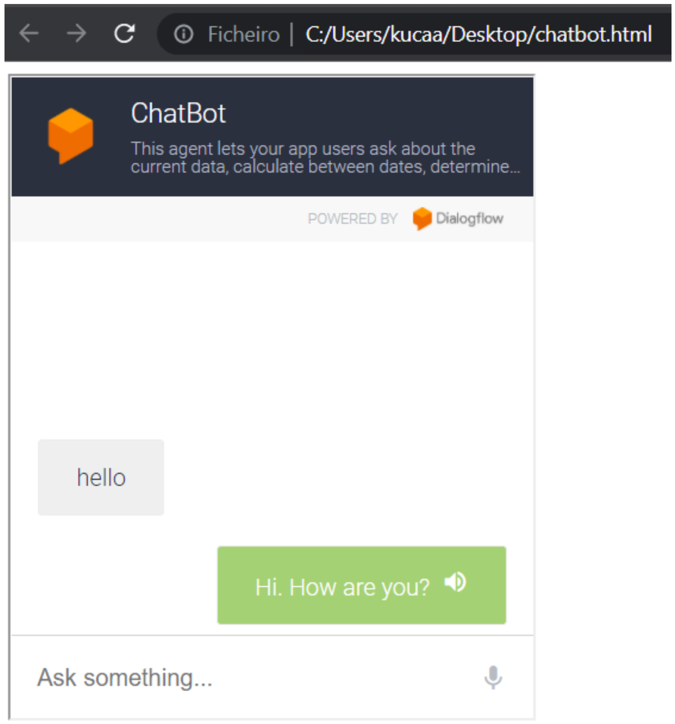


Figure 56. Implementation of a chatbot.

Currently there are many chatbots in node Js and Python Flask, in the case of using python there may be a need to create a database to store the input and output values as well as the possibility of using specific libraries and even using technologies of natural language processing.

Communication agents can be used in online stores. Currently, there are advanced systems of AI recommenders that verify user data and show the most suitable products for the profile.

LISP is based on the processing of lists and in AIML there is the use of tags with input and output parameters that allow responding to user requests similar to the implementation process in DialogFlow.

Voice recognition can be used for control purposes also.

Chapter 4

Mixed and Virtual Reality Systems with Motion Controllers

4

This chapter will describe the implementations performed to implement motion controller systems due to improve usability, it has four main sub sections being the first focused on hand motion tracking and controlling the second focuses on the use of eye gaze presenting a prototype of performing a visual test, the third a full body motion controller that can be used for physiotherapeutic activities and the fourth on commanding a drone based on motion.

Motion Controllers can help the accessibility to Information Systems for people with several disabilities like visual auditory, physical, speech, cognitive, language, learning and neurological. The use of these type of technologies has been increasing over the last years but have some limitations with many research challenges where this section is going to focus, presenting prototypes and processes to use for certain activities and to achieve objectives, such as, getting the vital signs of a patient using a hand gesture drone controller.

4.1 Employment of Artificial Intelligence Mechanisms for Gesture Control

With the objective to perform a more effective training for users of different ages and disabilities and to make the learning process easier, an online motion controller was implemented via web using several programming languages like JavaScript.

Open CV is able to detect the players movement. Via python code it is possible to detect the position of the hands during web camera video streaming. This technology detects the hands movement while the user is interacting with the online platform, allowing to control virtual objects that interact with the system.

Another possibility of getting hand movement instead of a camera is by the use of specific sensors like leap motion that has a library that allows to use the Unity game engine to develop applications which can be used for different areas like training and healthcare.

With JavaScript it is possible to develop applications that interact with the movement of the hands in real time, facilitating the use for people that have visual limitations or that do not have a keyboard or a mouse.

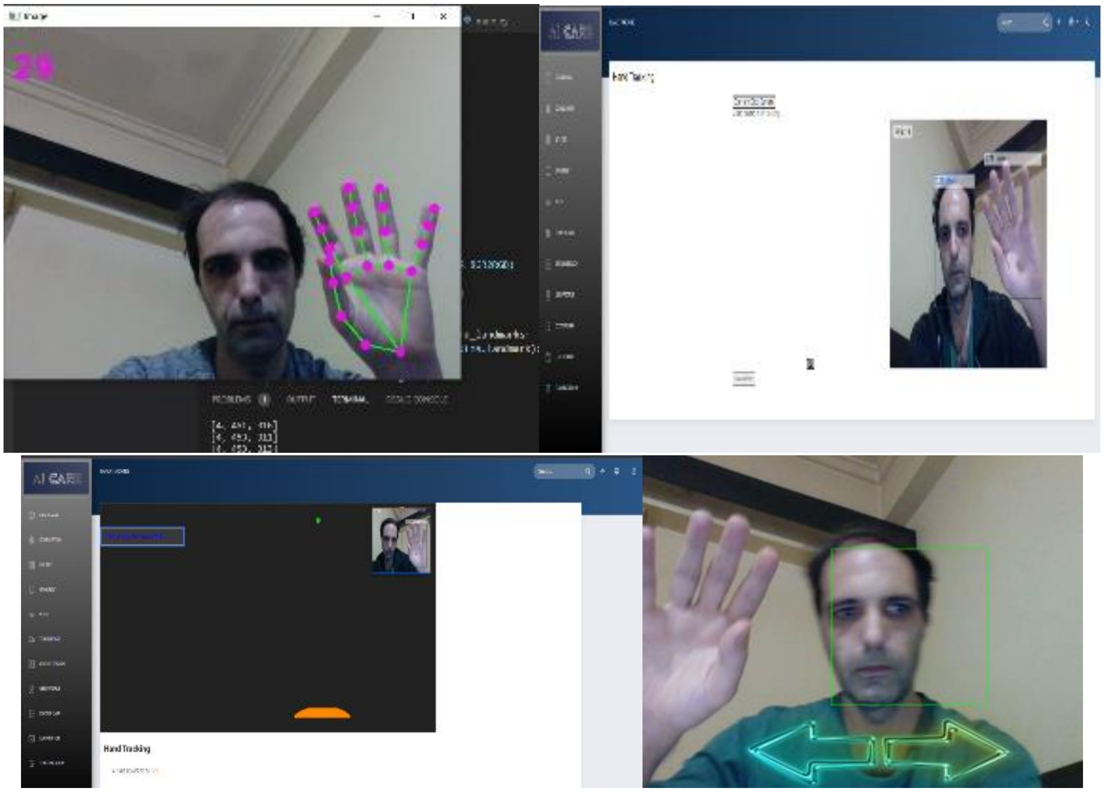


Figure 57. Hand detection and comparison. Using gesture control.

The developed prototype used the handtrack.js to detect hands from the video captured by a web camera and to control virtual elements, in order to be able to interact, the system first has to detect a hand, after detection the hand movements will control the virtual objects allowing to perform tasks like playing games.

This type of application can be very useful for physiotherapeutic activities concerning the hands and also makes the process of interaction easier as it does not require to use other hardwares like keyboard or mouse.

The use of hand controllers has been increasing but still faces many challenges, the process of hand recognition may fail, scenarios with low light may difficult the use and the quality of the camera and level of lightning may also affect the normal functioning of these type of systems.

4.2 Employment of Artificial Intelligence Mechanisms for Eye Gaze

The eye movement can interact with a web application, after detection of the iris via the web cam. On the prototype was used JavaScript libraries in order to make possible the realization of tests and to facilitate the interaction.

Eye gaze controllers can be very helpful for people with motor disabilities, controlling the application with the movement of the eyes and has many research challenges. Via Open CV using python it is possible to detect the movement of the iris.

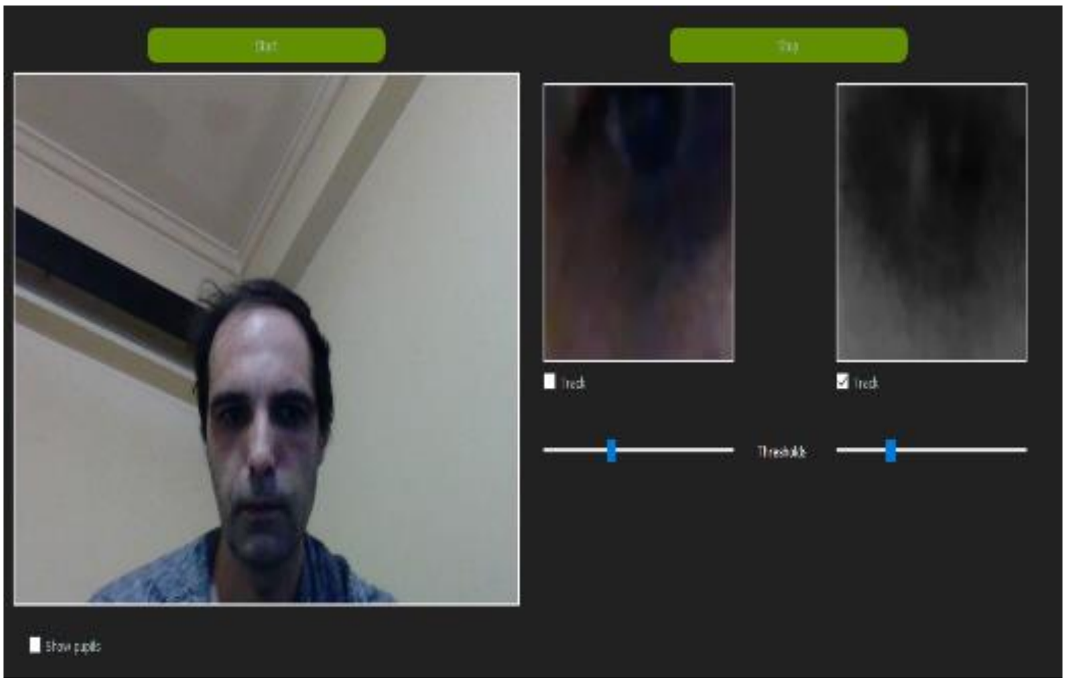


Figure 58. Use of Python to detect the eyes.

The detection process faces many AI challenges, it can be improved and for some cases may require a calibration process where the user looks to certain points of the screen in order to configure the use.

The presented prototype shows a possible solution for performing the visual test online where an orange dot works as a cursor mentioning where the patient is looking which facilitates the evaluation of a visual test providing more insights.

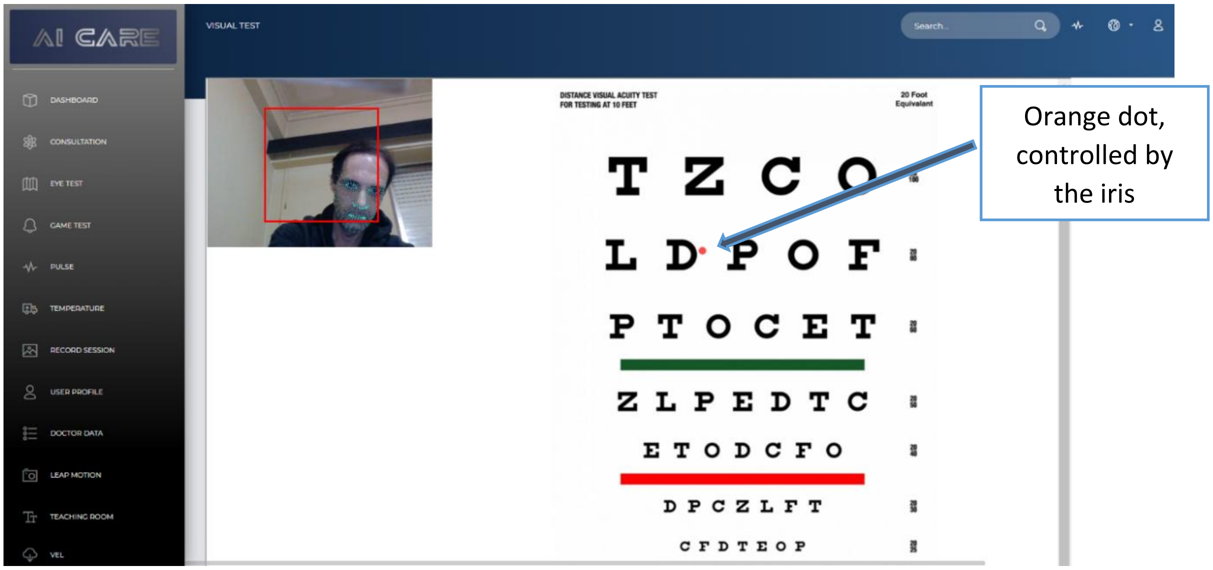


Figure 59. Eyes detection and interaction with the web virtual application.

Via JavaScript and using the webgazer.js it is possible to detect the eyes automatically without the need of calibration on real time and will show on the screen a point that corresponds to where the detected user is looking, the detection is done via web camera.

During the gameplay, the systems records where the user was looking and via AI presents a prediction of the user emotional states like happy, surprise and neutral, which can be used for evaluation purposes as well as the logs obtained for the eye movement during usage of the system.

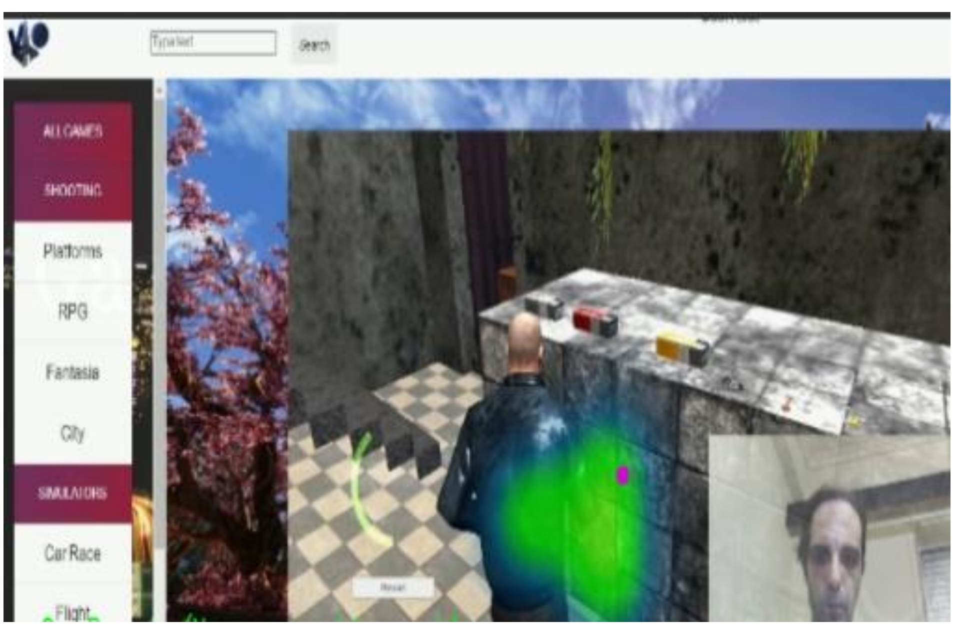


Figure 60. Displaying the area where the user is looking in a recording.

#Gaze cordinates [0,0] : left top corner, [1,1] bottom right corner				
#[RespondentNr]	[StimulusNr]	[timestam	[Gaze x %]	[Gaze y %]
0	0	376	0.948	0.92
0	0	503	0.93	0.913
0	0	640	0.913	0.885
0	0	678	0.883	0.842
0	0	794	0.899	0.784
0	0	948	0.669	0.627
0	0	1081	0.699	0.619
0	0	1213	0.718	0.563
0	0	1347	0.767	0.547
0	0	1479	0.837	0.54

```
Time, Pupil Center X , Pupil Center Y , Glint Center X , Glint Center Y , Pupil Diameter , Blink , DbBlink , HeadGesture , GazeX , GazeY
67,921,630,0,0,53,,,,0,0
181,923,630,0,0,52,,,,0,0
250,926,631,0,0,53,,,,0,0
320,935,636,0,0,50,,,,0,0
385,948,650,0,0,52,,,,0,0
450,961,656,0,0,53,,,,0,0
504,0,0,0,0,0,,,,0,0
599,0,0,0,0,0,,,,0,0
708,0,0,0,0,0,,,,0,0
797,0,0,0,0,0,,,,0,0
912,0,0,0,0,0,,,,0,0
1089,0,0,0,0,0,,,,0,0
1244,0,0,0,0,0,,,,0,0
1399,0,0,0,0,0,,,,0,0
1531,817,669,0,0,46,,1,,0,0
1601,0,0,0,0,0,,,,0,0
1749,0,0,0,0,0,,,,0,0
```

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Recording	Computer	Sensor	Project na	Export dat	Participant name	Variable1	Recording	Recording	Recording	Recording	Recording	Recording	Timeline n	Recording	Recording	Recording	Recording	Recording
2	0	1.74E+11		Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
3	21731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
4	25731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
5	29731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
6	33731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
7	37731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
8	41731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
9	45731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
10	49731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
11	53731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
12	57731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
13	61731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
14	65731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
15	69731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
16	73731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
17	77731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
18	81731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
19	85731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
20	89388	1.74E+11	Mouse	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
21	89731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10

Entire Recording								
Duration of inter	**Participan**	**Variable1**	**1**	**Average**	**Median**	**Count**	**Total Time**	**Total Recording Duration**
Recording1	Participant		15.66	15.66	15.66	1	15.66	15.66
Average			15.66	15.66	15.66	1.00	15.66	15.66
Count			1					
Variance								
Standard Deviati								

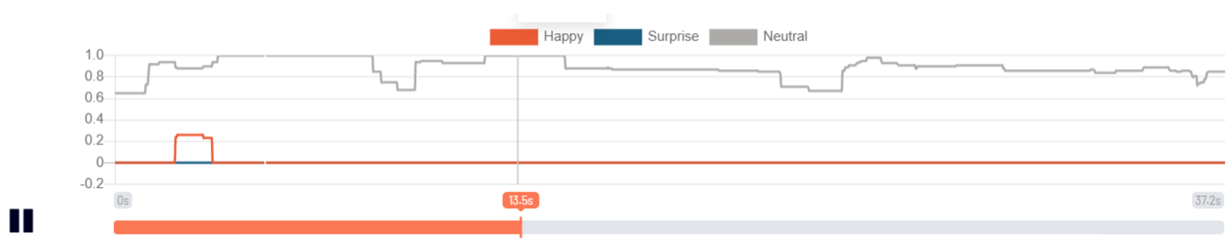


Figure 61. Eye gaze statistics for different time frames.

During tests was recorded sessions, these recordings allow access several type of informations like the predicted state of emotion of the user, the time taken, the coordinates in the screen of where the user was looking among other information in order to test if the proposed prototype is valid.

Eye gaze controller still have many research challenges, the detection process is not easy as it has two region of interest and requires complex calculations, webgazer.js showed high results and the implementation process requires less labor.

4.3 Employment of Artificial Mechanisms for Full Body Motion Controllers

Learning activities can be more interactive and accessible for users of different ages and types of disabilities. Based on the context of training and to facilitate the performance of physiotherapeutic tasks a controller that detects, hands, arms, legs, head and feet was used to answer the third research question of this investigation.

Based on the creation of AI Agents, like the creation of a Computer-generated Imagery (CGI) girl, the sensor which in this case is the webcam requires to capture the all body in order to associate geometries and bones. A virtual girl replicates the user movements on a projected Liquid Crystal Display (LCD) that aims to help in physiotherapeutic terms by providing movements and exercises that can help some motor disabilities. The prototype was developed in C# using the Game Engine Unity providing a Mixed Reality (MR) experience. On a perspective of virtual storytelling, this project uses a process of providing interactivity to the user for story creation.

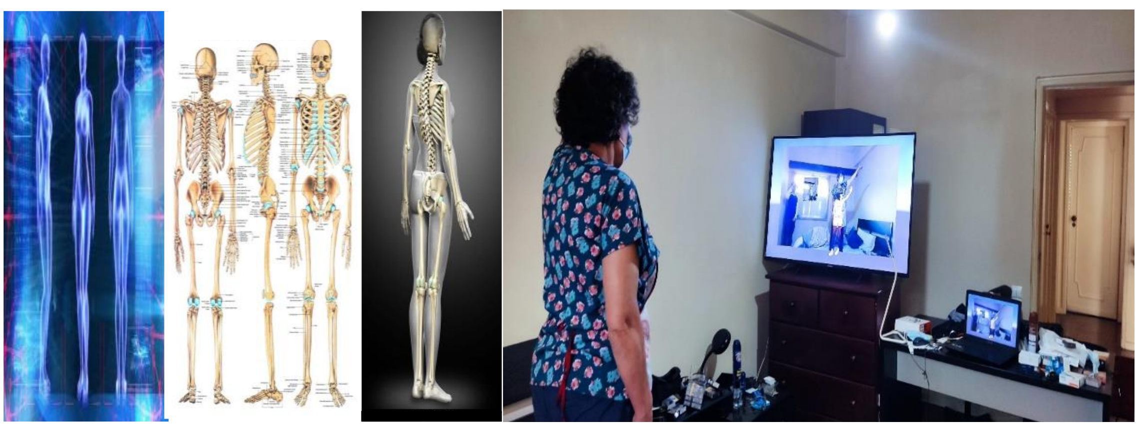


Figure 62. Full body motion and detector based on the bones structure.

The technology used can help people with visual problems to interact with the system with all parts of the body having the head, arms, body and legs more relevance concerning the detection process.

Via C# the developed prototype is able to associate the animations to the values detected by the camera.


Figure 63. Performing test with full body motion controller.

During tests it was noticed a high level of satisfaction from the users, mainly due to being a new technology. The MR system in cause allows to place virtual elements in the images captured by the camera, the objective was possible to achieve via the Unity game engine.

In order for the application to work the camera has to detect all part of the bodies, so users can not be close to the sensor and should be at a distance superior to 80 cms to perform better results.

There are other sensors like the Wii 2 that allow to interact with many parts of the body.

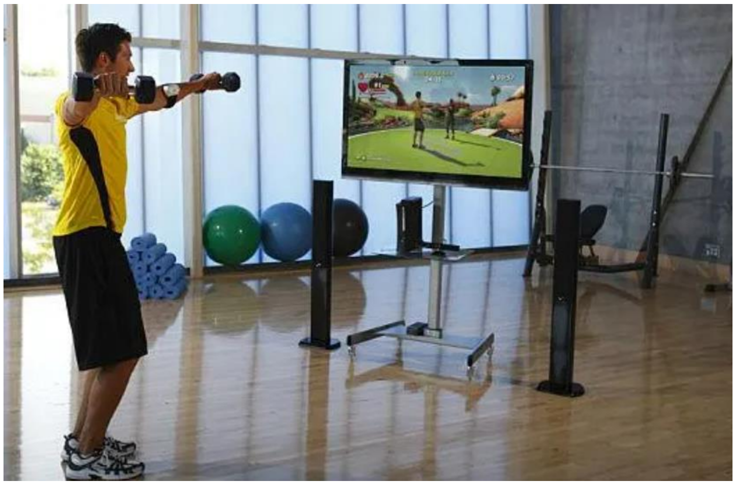


Figure 64. Wii 2 for full body tracking.

4.4 Employment of AI Mechanisms on Virtual Elements for Drones with Gesture Controller

In order to implement MR on the images obtained by the camera, the open CV library was used and most of the code was written in Python and JavaScript. For the prototype was developed a web and gesture controller.

A DJI Tello drone was used to get physical data, although the flight time is not high, it is enough for testing one user at a time and the quality of the images obtained by the camera was enough to perform the AI mechanisms, although other drones have better capacities and compatibilities with other technologies.

Figure 65. Drone 720p Ryze by DJI Tello and the mobile phone application controller. Flight time: up to 13 min. Image transmission distance: 100m. HD 720p broadcast. Intel processor. Box Contents Accessories: Battery, Propellers, Propeller Guard, Compartment Compatible with remote controls and virtual reality devices. Maximum speed of 8 m/s.

Concerning objective 2, a hand gesture drone controller was developed, by controlling the drone with gestures it is possible to be used by people with some motor and visibility limitations. For the drone controllers was imported the DJI libraries in Python.

For testing purposes, the objective was to get the vital signs with a drone, which has some research challenges like collisions with walls.

The implementation of AI mechanisms allows an improvement on MR applications, providing more functionalities and answering problems that are not possible to achieve via traditional processes.

The connection of the used drone with the computer was done via wi-fi and to get the vital signals, the patient does not have to perform any actions and the doctor has to point the drone camera to the face of the user.

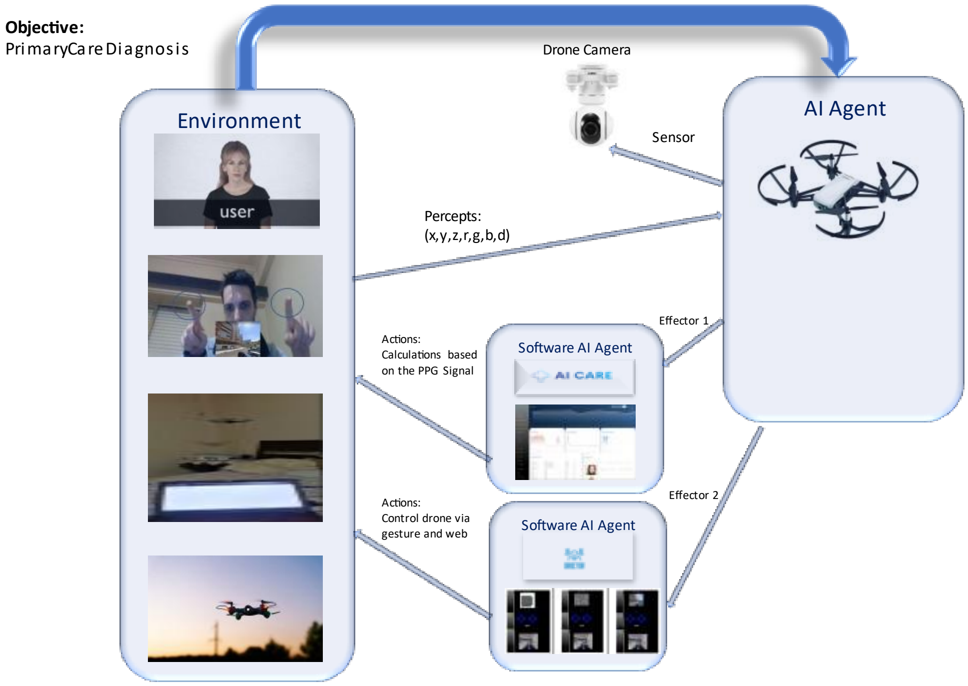


Figure 66. Perspective from the AI Drone agent ("Droctor").

The AI agent uses the camera as a sensor and via OpenCV it is possible to detect the ROI, which requires the drone camera to capture the face of the end user, so the drone can not be very close or very far from the patient.

The web application that controls the drone was developed using the programming languages PHP, JavaScript, Python, Flask among others, the database was developed with MySQL, storing data such as the vital signs of the patients for different time frames.

When the drone is active, the coordinates of latitude and longitude will be obtained from the IP of the UAV and will be displayed in a map on the application via the Google Maps API allowing to know where the vehicle is currently placed.

In order to get the vital signs of a user, the camera of the drone has to identify a previous defined region of Interest such as a face, after correctly tracking the area, the process of obtaining vital signs will start.

There are two type of controllers one via hand gestures and the other one by using web buttons, which allow to give instructions to the UAV via Wi-Fi.

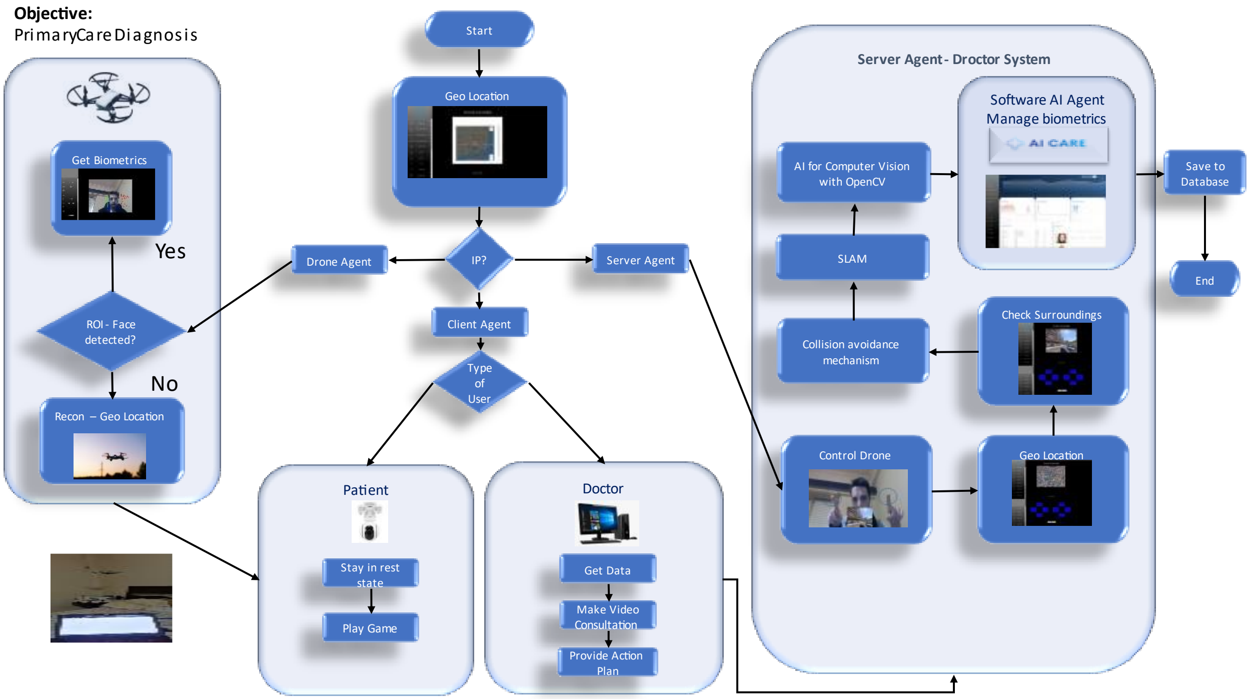


Figure 67. Process of obtaining vital signs data via drone.

The presented AI system has several functionalities, like getting the vital signs, check the surroundings, obtain the location of the UAV among others.

The "Droctor" web based application shows a list of users and the respective bio data, which can be modified by an user with administration access in order to make changes, because the readings may fail or more information may be required to perform a diagnosis.

The interaction between user and drone is made via an online system that allows to control the drone via gestures or buttons facilitating the accessibility for different types of people with different ages.

Mixed reality systems have been improving over the last years allowing the implementation of new functionalities but still have limitations with research challenges where this dissertation tries to focus. The use of virtualization technologies can also be applied to drones displaying the information obtained by the camera mixed with virtual elements with AI mechanisms.

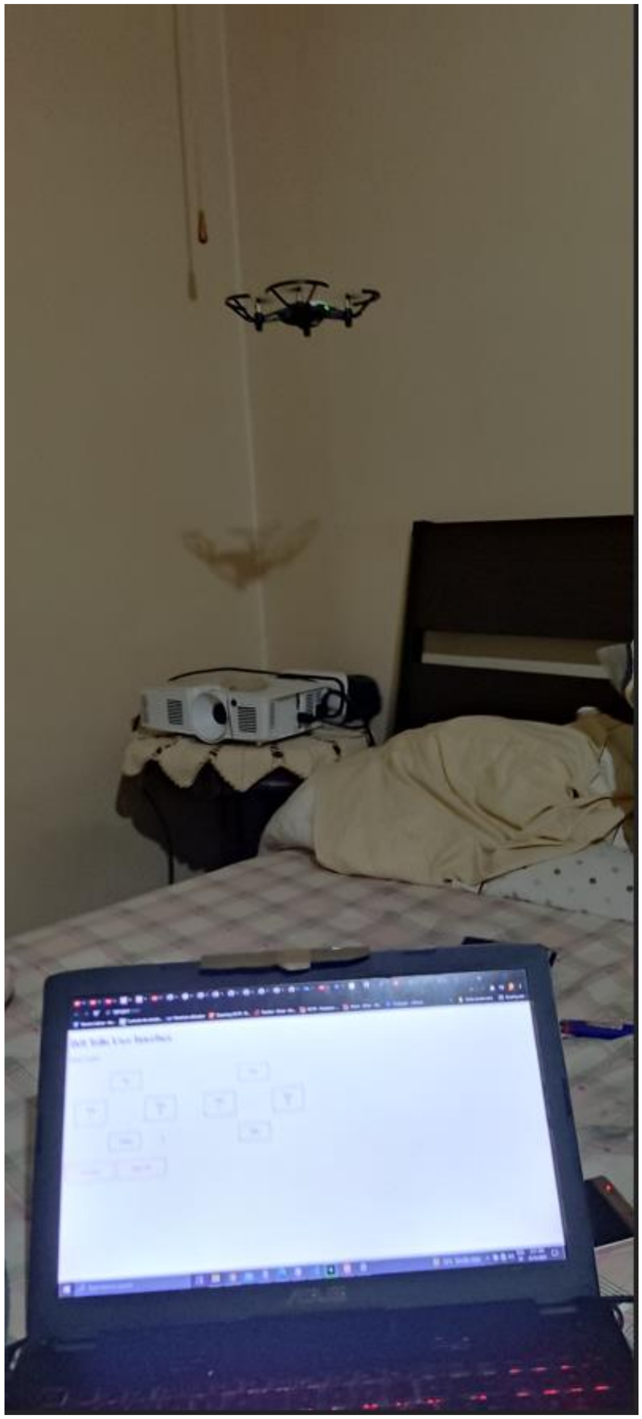


Figure 68. Performance of tests with the drone web controller.

The tests are important to determine and evaluate the system developed, for this project the distance of the user to the cameras is important, if very close or very far the ROI may not be identified and the information will not be processed.

The initial tests were done with web buttons to control the drone and after was tested the hand gesture controller.

4.4.1 Collisions Avoidance

In order to avoid collisions of the drone with physical objects in the real-world environment several strategies were implemented like setting a maximum speed limit and defining a distance from which the unmanned aerial vehicle (UAV) will not move towards the defined direction, DJI Tello has a library that allows to program the control of the drone via Python, and the OpenCV Library was also used to treat images.

Simultaneous localization and mapping (SLAM) can be seen as a mechanism for creating and updating maps based on the information obtained by the sensor, which in the proposed solution is the camera of the drone that is accessible from the web application by Internet Protocol (IP) via Wi-Fi helping also to locate the controlled device.

4.4.2 Geo Location

In order to make the process to intervene in case of emergency much easier, the web application using the google maps API shows where the drone is located on the map, allowing also to facilitate the navigation when the drone is not physically visible to the pilot. The drone camera will show the place where the patient is on the web application, providing important information about the environment.

Via JavaScript is possible to locate based on Internet Protocol (IP) and calling APIs that will use the variables of latitude and longitude to show on the map where the UAV is placed.

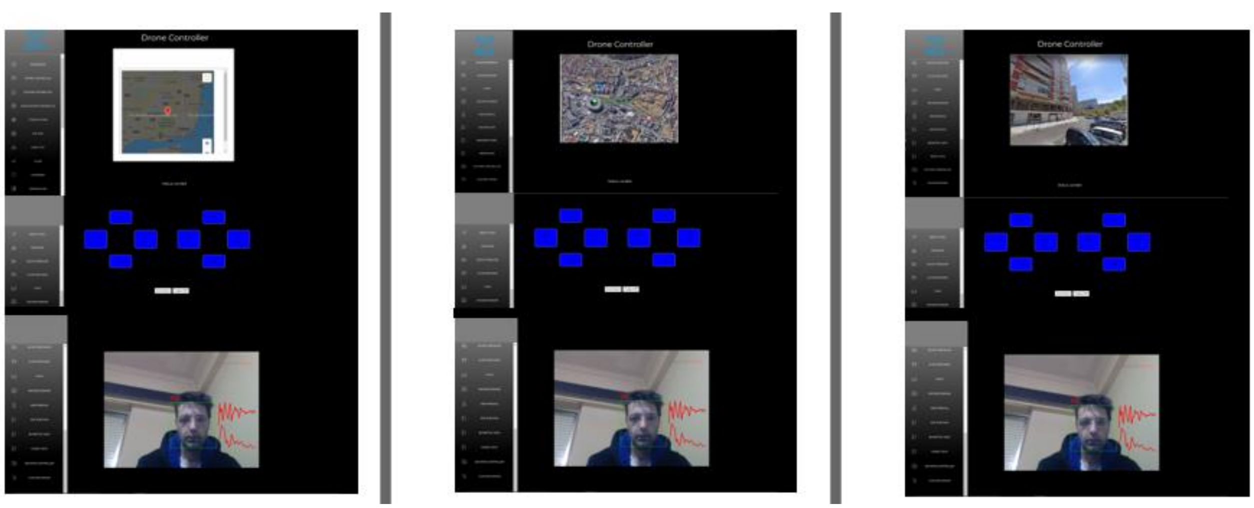


Figure 69. Locating and controlling a drone with a map with geo location.

4.4.3 Gesture Controller

Due to accessibility purposes, and in order to make the application accessible for people with certain disabilities like vision problems, the drone can be controlled by gestures, the web cam of the computer will look for the Region of Interest (ROI) which in the solution provided are the two fingers and provide interaction via motion.

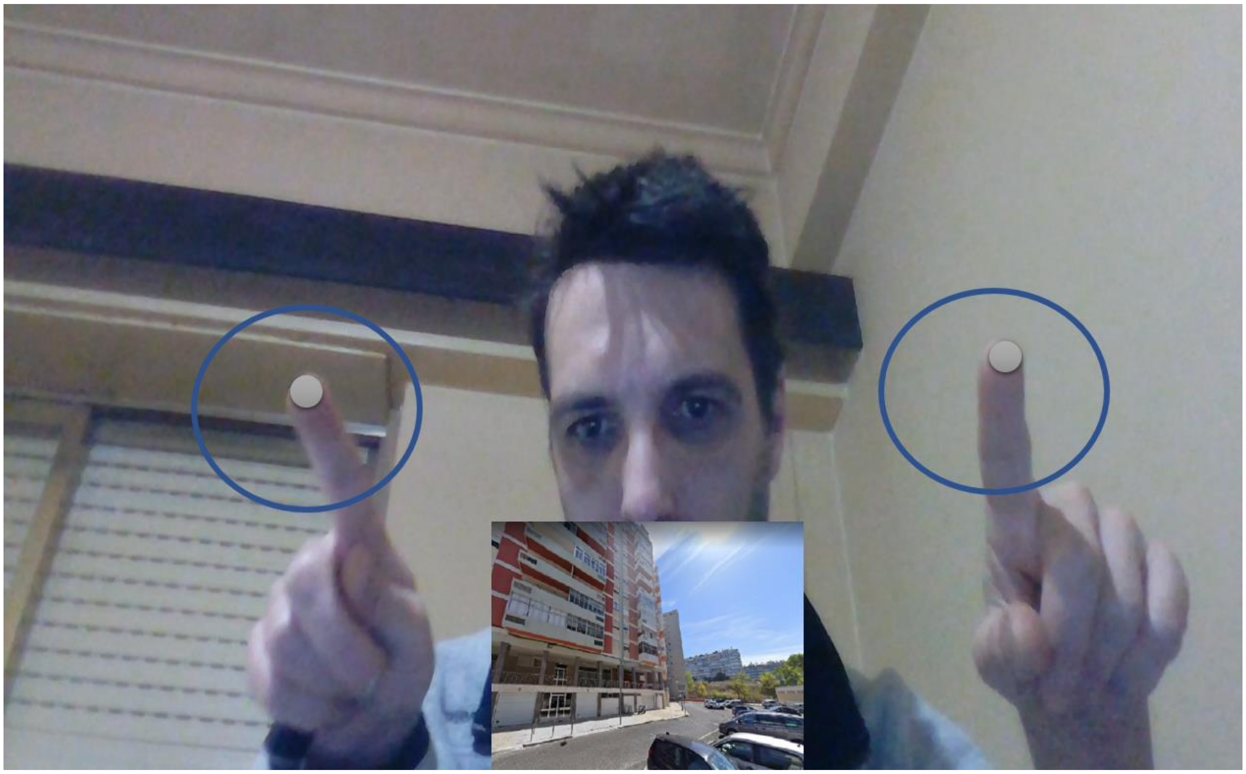


Figure 70. Controlling the drone with the fingers.

4.4.4 Drone Doctor

There are two cameras being used in this multi-agent system, one is the web cam of the computer the other, the camera of the UAV, the first camera helps the doctor to control the flying device and also uses AI for computer vision mechanisms that allow the pilot to see himself, while performing his tasks, the second camera is in the drone, that will treat the Photoplethysmogram (PPG) signal obtained allowing to know the BPM and other important medical information via the face detected by the done camera. The level of light influences the accuracy of the readings, but after tests even in low light environment the levels obtained are almost the same as the one perceived by medical instruments like infra-red (IR) thermometers or oximeters.

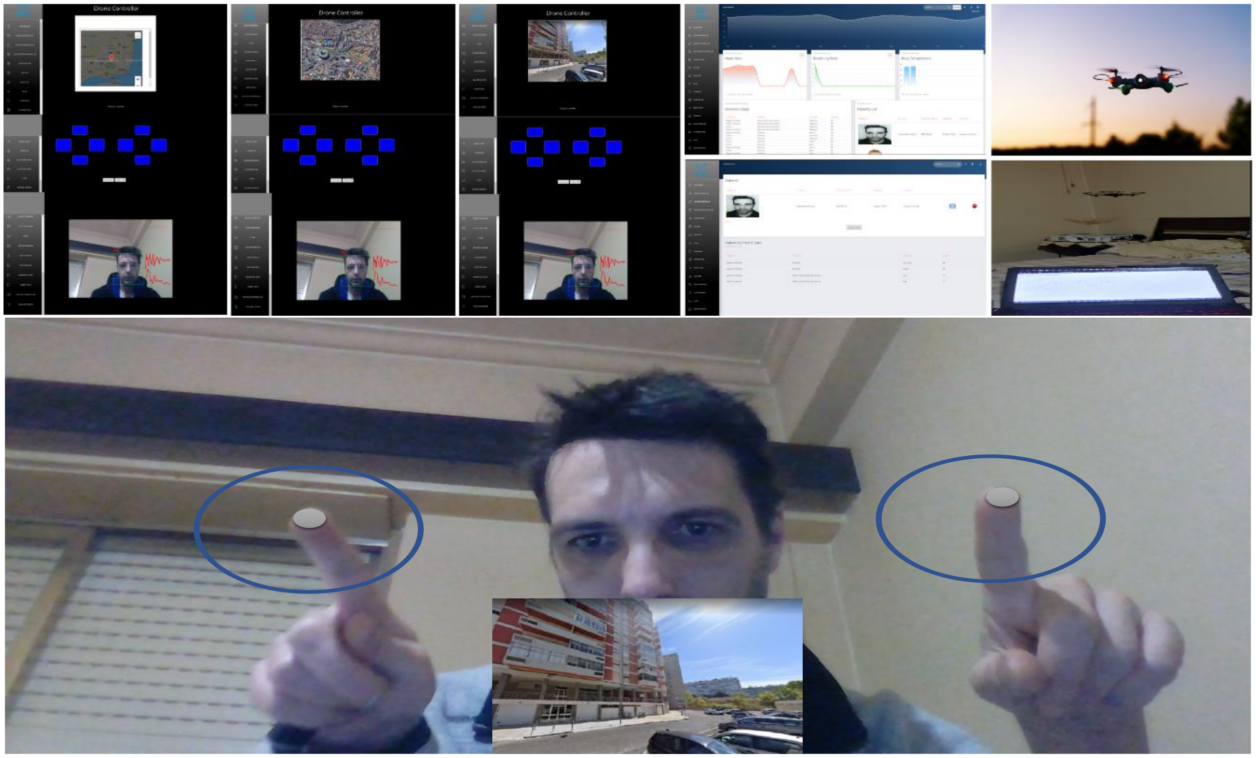


Figure 71. Drone controller via keyboard and gesture movements with geo localization to obtain the vital signs data of the patients.

In order to avoid collisions of the drone with real world objects, the Simultaneous localization and mapping (SLAM) was used, updating a map of the environment and tracing the AI agent location. The main technologies used on this project were: Python, Flask, PHP, MySQL, the Google Maps API, among others.

A great part of the project was done using the programming languages Python and JavaScript, the information obtained by the drone camera is treated via AI Computer Vision and placed on a virtual web application on order to use the information obtained by the environment and getting vital signs data via the Photoplethysmogram (PPG) signal. The gesture control is based on the Region of Interest (ROI) detection and the virtual elements on the image obtained by the computer web cam make easier to control an unmanned aerial vehicle (UAV).

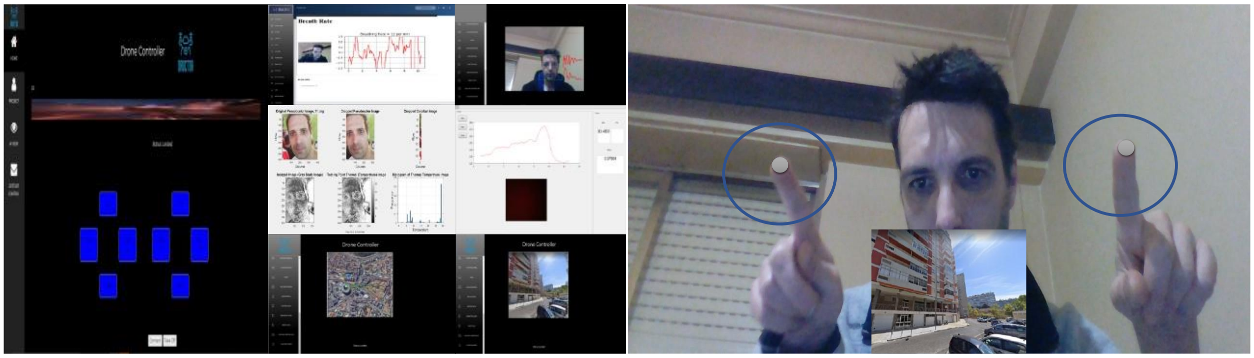


Figure 72. Controlling the drone, obtaining the vital signs data and using MR to report the information obtained by AI Computer Vision.

AI Mechanisms may allow a better performance and add new functionalities to the current existing applications making some tasks easier like the diagnosis on the primary care of a patient, which can be done online and without physical contact avoiding the contagion of the Covid-19 disease.

Considering the objective 1 of this research, the use of AI mechanisms can be applied to virtual elements allowing to control a drone, concerning the second objective the application of motion controllers allow people with different type of disabilities to perform tasks.

This project aims to answer the research questions by providing details of the AI mechanisms used, how the problem of collisions was treated and an implementation of different types of motion controllers to provide accessibility for different disabilities.

Chapter 5

Tests, Results and Discussions

5

This chapter presents the results and tests of the applications in order to verify if the objectives are achieved and to validate the research questions. Concerning the first and second objective of this research a questionnaire process was used.

Concerning the second and the third research question the time and distance to perform the tasks was measured in order to evaluate the artifacts developed. The tests follow a combination of two types of approaches, the qualitative and the quantitative.

5.1 Questionaries to the Users

First was asked to sign a right grant form before proceeding with the questions and the use of the applications. The second test was based on using the prototypes, being the sessions recorded and the third step to fill an evaluation questionnaire.

Concerning which of the motion controllers is the most important, in general all participants found them important but none considered the hand motion controller the least important.

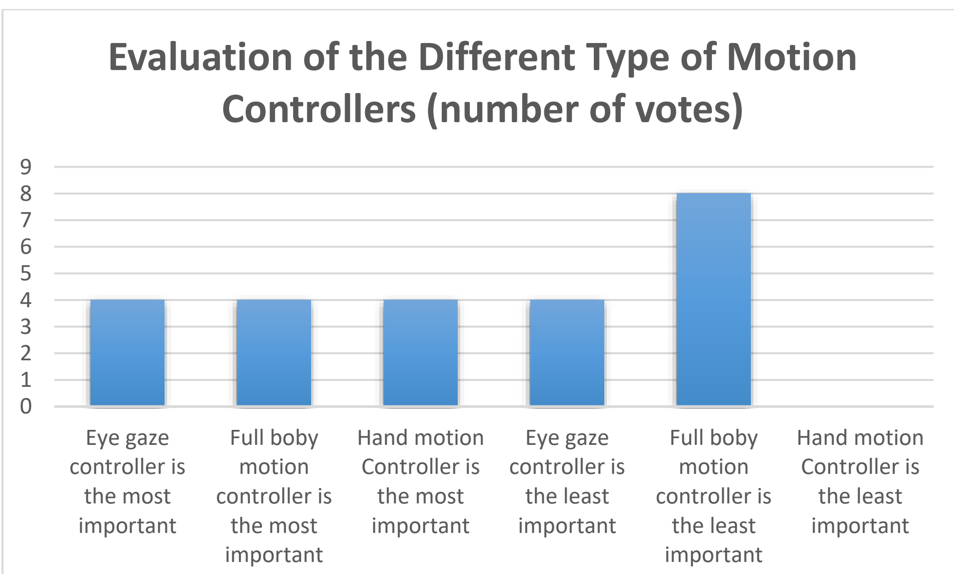
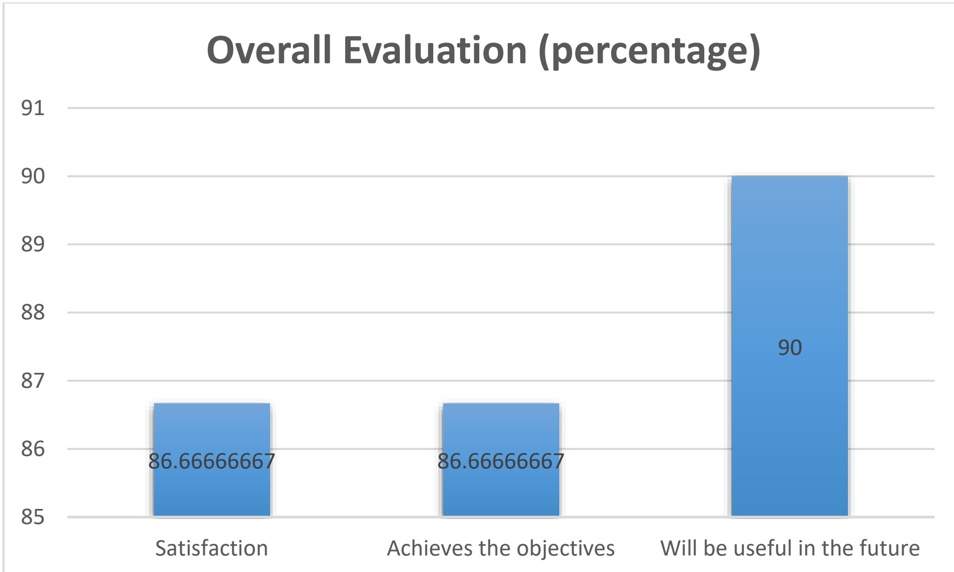


Figure 73. Results of the questionaries concerning the importance of each type of motion controller.

In general, all users considered the motion controllers an important feature and that will be useful in the future.



The use of AI mechanisms in MR systems was considered important by all the testers and that may be more used in the future, having more functionalities and collisions between the real-world objects and virtual elements makes the applications more appealing.

Concerning the motion controllers, it was noticed that the hand motion controller was not neglected and very appealing for cases of users with visual disabilities, something that was more highlighted for testers of older ages.

The questionnaire allowed to validate all the research questions of this investigation, allowing to retrieve qualitative conclusions.

5.2 Tests of Collisions

Some of the MR prototypes were tested in order to verify if collisions between the virtual elements and the real-world objects occurred assuring that the augmented elements would not appear inside of real-world objects. The tests were performed in an indoor environment with artificial and natural light.

Table 4. Collision tests

Tester	*Time to Complete (in seconds)*	*Collisions Occurred*	*Errors*
Tester 1	54	12	0
Tester 2	102	14	0
Tester 3	30	4	1
Tester 4	75	8	0
Tester 5	140	9	0
Tester 6	87	13	1
Tester 7	60	10	0
Tester 8	34	6	0
Tester 9	103	10	0
Tester 10	109	8	1
Tester 11	56	6	0
Tester 12	80	3	0

These tests were considered for answering the second research question, in order to verify the validity of the artefact developed. The use of Vuforia engine facilitated the process of implementation and also to provide efficient results.

5.3 Tests and Analysis of the Motion Controllers

The tests were performed in an indoor environment with artificial and natural light, most of the readings were close to the ones detected with medical devices.
Via OpenCV is possible to measure the distance of the face from the camera during readings.

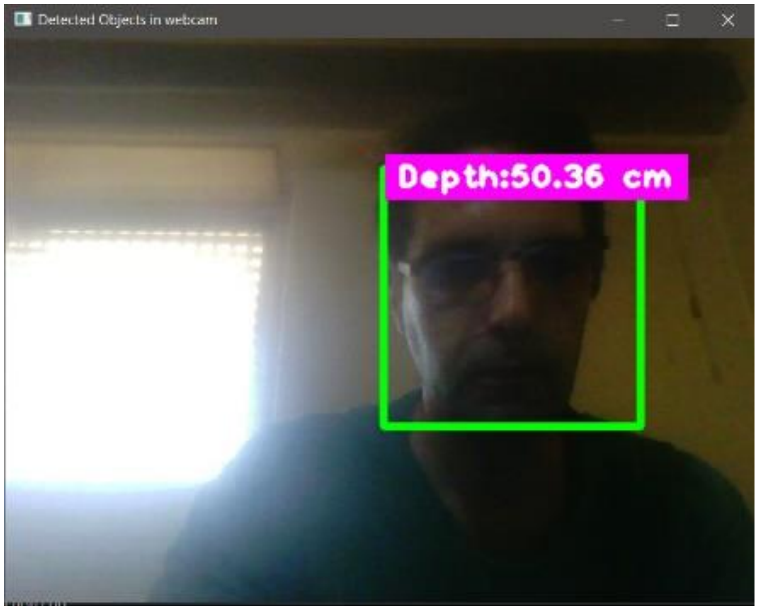


Figure 74. Measuring distance from the ROI to face.

Table 5. Qualitative evaluation results (distance from camera, cms).

	Distance from camera	Standard deviation
tester 1	55.4	5.54
tester 2	53.6	7.34
tester 3	65.4	8.68
tester 4	72.5	5.68
tester 5	67.8	8.68
tester 6	50.5	10.44
tester 7	63.4	1.08
tester 8	55.8	8.08
tester 9	58.7	3.08
tester 10	64	3.058333
tester 11	68.3	7.358333
tester 12	55.9	5.04

Average	60.94166667

Table 6. Quantitative evaluation results (time to completion in seconds).

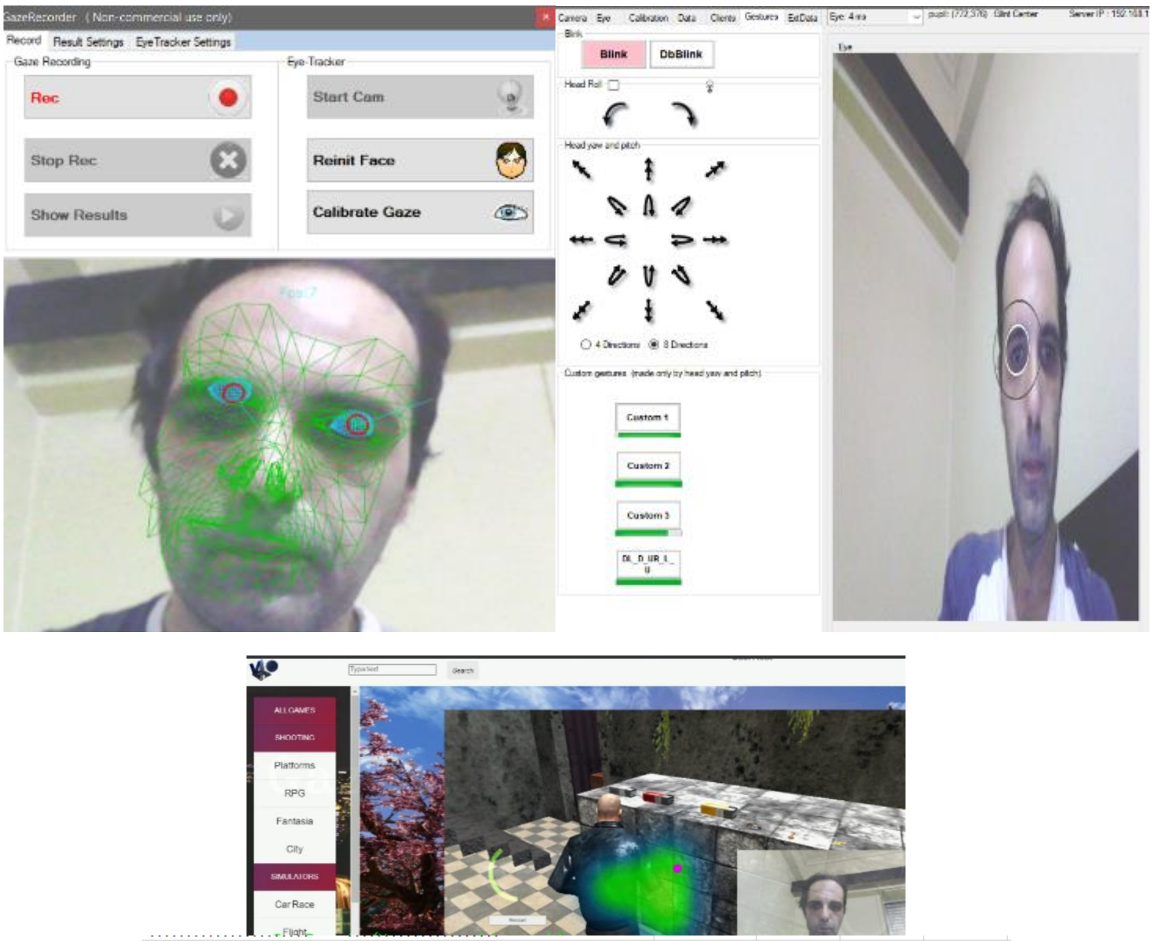
	Time taken	Standard deviation
tester 1	186	60
tester 2	214	88
tester 3	75	51
tester 4	140	14
tester 5	110	16
tester 6	98	28
tester 7	200	74
tester 8	89	37
tester 9	95	31
tester 10	81	45
tester 11	104	22
tester 12	120	6

Average	126

In order to check the accuracy results, tests were done taking in consideration the performance of specific tasks, concerning the performance of hand motion activities it was noticed that after ROI detection, users can easily perform the tasks, but detecting the area of interest was difficult for some cases.

5.4 Tests of Eye Gaze Motion

The tests were performed in indoor environments using in some cases artificial light and other cases natural light. The movement of the Iris was recorded and data was obtained like the coordinates of where the eyes are looking in the screen and predictions of emotional state.



#Gaze cordinates [0,0] : left top corner, [1,1] bottom right corner					
#[RespondentNr]	[StimulusNr]	[timestam	[Gaze x %]	[Gaze y %]	
0	0	376	0.948	0.92	
0	0	503	0.93	0.913	
0	0	640	0.913	0.885	
0	0	678	0.883	0.842	
0	0	794	0.899	0.784	
0	0	948	0.669	0.627	
0	0	1081	0.699	0.619	
0	0	1213	0.718	0.563	
0	0	1347	0.767	0.547	
0	0	1479	0.837	0.54	

Figure 75. Recording sessions of eye gaze.

The data obtained during recordings can also be used to predict emotional states. The use of MR applications with motion controllers has been Increasing.

```
Time, Pupil Center X , Pupil Center Y , Glint Center X , Glint Center Y , Pupil Diameter , Blink , DbBlink , HeadGesture , GazeX , GazeY
67,921,630,0,0,53,,,,0,0
181,923,630,0,0,52,,,,0,0
250,926,631,0,0,53,,,,0,0
320,935,636,0,0,50,,,,0,0
385,948,650,0,0,52,,,,0,0
450,961,656,0,0,53,,,,0,0
504,0,0,0,0,0,,,,0,0
599,0,0,0,0,0,,,,0,0
708,0,0,0,0,0,,,,0,0
797,0,0,0,0,0,,,,0,0
912,0,0,0,0,0,,,,0,0
1089,0,0,0,0,0,,,,0,0
1244,0,0,0,0,0,,,,0,0
1399,0,0,0,0,0,,,,0,0
1531,817,669,0,0,46,,1,,0,0
1601,0,0,0,0,0,,,,0,0
1749,0,0,0,0,0,,,,0,0
```

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Recording	Computer	Sensor	Project na	Export dat	Participant name	Variable1	Recording	Recording	Recording	Recording	Recording	Recording	Timeline n	Recording	Recording	Recording	Recording	Recording
2	0	1.74E+11		Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
3	21731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
4	25731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
5	29731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
6	33731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
7	37731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
8	41731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
9	45731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
10	49731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
11	53731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
12	57731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
13	61731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
14	65731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
15	69731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
16	73731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
17	77731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
18	81731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
19	85731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
20	89388	1.74E+11	Mouse	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10
21	89731	1.74E+11	Eye Tracke	Project1	5/20/2021	Participant1		Recording	5/20/2021	5/20/2021	10:11:13.5	09:11:13.5	15664	Timeline1	Tobii I-VT	1.162.324	1350	2400	10

Entire Recording

Duration of inter	Participan	Variable1	1	Average	Median	Count	Total Time	Total Recording Duration
Recording1	Participant		15.66	15.66	15.66	1	15.66	15.66
Average			15.66	15.66	15.66	1.00	15.66	15.66
Count			1					
Variance								
Standard Deviati								

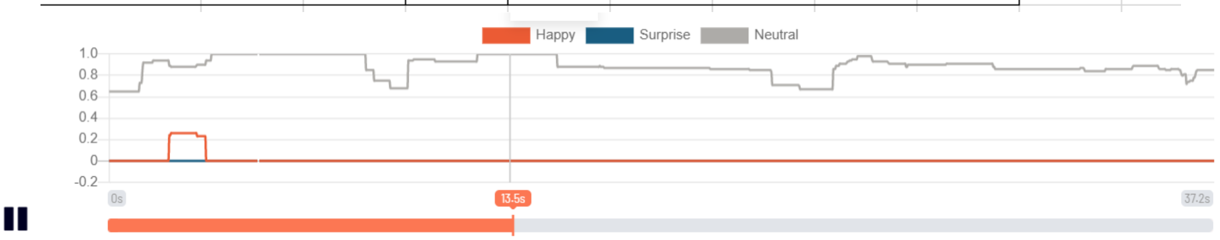


Figure 76. Prediction of emotions

Chapter 6

Conclusions, Limitations and Future Work

6

This dissertation presents a set of processes and prototypes to increase the usability of mixed reality applications. The study took in consideration related work and theoretical concepts, certain aspects were given relevance, such as the collisions of virtual elements with real world objects and the implementation of responsive AI in order to make the applications more useful.

The accessibility problem was analyzed presenting possibilities by using different types of motion controllers to facilitate the use for people with different types of disabilities and ages.

It was noticed difficulties in the process of obtaining the ROI in some cases like low lightning or the user not positioning correctly in front of the camera.

The research aimed to increase the efficiency and functionality of Mixed Reality applications by adding AI mechanisms to the virtual elements making easier to achieve pre-defined objectives.

Over the last years was noticed an increased use of motion controllers, allowing to help the use for several type of users and also reducing the dependency on a mouse and a keyboard.

Some of the processes described have low cost, are not so laborious and faster than traditional methodologies.

6.1 Answer to the Research Questions

The answers to the research questions are:

Research question 1: How can the implementation of AI mechanisms in Virtual elements provide more efficiency, control and functionality?

The use of AI mechanisms in virtual elements allow to perform several activities, like controlling drones, displaying the vital signs of a user, have more functionalities on the applications like shooting mechanisms in a first-person shooter for mobile, among many other possibilities.

The interaction between user and an AI agent has been developing over the last years, providing more functionalities to these agents allowing to make the applications more useful and provide more information that can be used in several areas like healthcare.

Research question 2: How to avoid collisions between virtual elements and real-world obstacles?

Several processes can be used to answer the question like SLAM, a prototype was developed using Vuforia to take in consideration the real-world environment.

The misplacement of virtual elements may impact the overall process of a MR application, it is important to find a solution that provides interaction and responsive AI, allowing the CGI to be placed correctly in an LCD.

Research question 3: Which type of motion controllers can facilitate the usability and accessibility of mixed reality applications for people with disabilities like visual, auditory, physical, speech, cognitive, language, learning and neurological?

Three different types of motion controllers, namely eye gaze, hand motion and full body were given more relevance in this dissertation. The mixed reality applications can be accessible for different types of users, using the iris has a controller for example can help people with motor disabilities and hand controllers can facilitate the use to persons with visual difficulties, this question depends on the level of incapacity, the showed types of motion controllers are only able to be used for certain disabilities and for certain levels.

6.2 Future Investigation

Mixed reality applications are currently being used in different areas like mechanics, training, healthcare among others, but many research challenges are still open, there is an increase need to make these applications accessible to different types of disabilities and also a development of the AI mechanisms that allow to implement important functions like reading the vital signs via a sensor like a camera, or allowing to make predictions facilitating the process of detecting diseases in medical images.

Drones are being used in several areas of investigation and can have more functionalities reducing the human error, it is expected that cars will be able to drive automatically and some functions currently done by humans to be replaced.

The future of the applications has been followed by an increasing development of the correspondent hardware, CPUs have more memory capacity, it is possible to treat higher amounts of data at faster speeds, data science has been growing and answers to problems like predicting when a stock is going to be replaced or identifying a person by pointing a camera to the eyes or face from a huge dataset is being used, advanced algorithms in large amount of data can give information about a person identified via a drone camera, a situation that is currently being done in China, these technologies are being enhanced and having more functionalities.

References

Acemoglu, D., & Restrepo, P. (2020). The wrong kind of AI? Artificial intelligence and the future of labour demand. *Cambridge Journal of Regions, Economy and Society*, *13*(1), 25-35.

Ahirwal, P. (2020). Correlation of Neurochemical Profile to the Retinal Fundus Imaging Features: Pilot Experiment.

Antunes, J. E. (2017). *Gaze-directed gameplay in first person computer games*.

Bagha, S., & Shaw, L. (2011). A real time analysis of PPG signal for measurement of SpO2 and pulse rate. *International journal of computer applications*, *36*(11), 45-50.

Babu, B. P. W., Yan, Z., Ye, M., & Ren, L. (2018, October). On Exploiting Per-Pixel Motion Conflicts to Extract Secondary Motions. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 46-56). IEEE.

Barrientos, A., Peñin, L. F., Balaguer, C., & Aracil, R. Fundamentos de robótica. 2002.

Breen, D. E., Rose, E., & Whitaker, R. T. (1995). Interactive occlusion and collision of real and virtual objects in augmented reality. *Munich, Germany, European Computer Industry Research Center*. In *Computer Graphics Forum* (Vol. 15, No. 3, pp. 11-22). Edinburgh, UK: Blackwell Science Ltd.

Castaneda, D., Esparza, A., Ghamari, M., & Soltanpur, C. (2018). HJI jo b. *Nazeran, and bioelectronics," A review on wearable photoplethysmography sensors and their potential future applications in health care*, *4*(4), 195.

Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. (2010). Augmented reality technologies, systems and applications. Multimedia Tools and Applications, 51(1), 341–377. https://doi.org/10.1007/s11042-010-0660-6.

Chin, C. A., Barreto, A., Cremades, J. G., & Adjouadi, M. (2008). Integrated electromyogram and eye-gaze tracking cursor control system for computer users with motor disabilities.

Chitra, L. P., & Satapathy, R. (2017, February). Performance comparison and evaluation of Node. js and traditional web server (IIS). In *2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET)* (pp. 1-4). IEEE.

Dix, A., Finlay, J., Abowd, G., & Beale, R. (2004) *Human Computer Interaction*. (3rd Edition). ISBN:0130461091. Upper Saddle River, NJ, USA: Prentice-Hall, Inc

Fischler, M.A., & Bolles, R.C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM, 24*, 381-395.

Foo, N. (2007). Social Agents. 4830. 14-14. https://doi.org/10.1007/978-3-540-76928-6_2.

Garrett, J. (2002). *The Elements of User Experience: User-Centered Design for the Web*. ISBN: 978-0735712027. UK: Pearson Education.

Ghafghazi, S., Carnett, A., Neely, L., Das, A., & Rad, P. (2021). AI-Augmented Behavior Analysis for Children With Developmental Disabilities: Building Toward Precision Treatment. *IEEE Systems, Man, and Cybernetics Magazine*, *7*(4), 4-12.

Jia, P., Hu, H. H., Lu, T., & Yuan, K. (2007). Head gesture recognition for hands-free control of an intelligent wheelchair. *Industrial Robot: An International Journal*.

Guo, K., Zhang, S., Zhao, S., & Yang, H. (2021). Design and Manufacture of Data Gloves for Rehabilitation Training and Gesture Recognition Based on Flexible Sensors. *Journal of Healthcare Engineering*, *2021*.

Haber, R. B. (1990). Visualization techniques for engineering mechanics. *Computing Systems in Engineering*, 1(1), 37–50. https://doi.org/10.1016/0956-0521(90)90046-n.

Haralambiev, H., Boychev, S., Lilov, D., & Kraichev, K. (2011, April). Applying source code analysis techniques: A case study for a large mission-critical software system. In *2011 IEEE EUROCON-International Conference on Computer as a Tool* (pp. 1-2). IEEE.

Haris, N. A., & Hasim, N. (2019). PHP frameworks usability in web application development. *Int. J. Recent Technol. Eng*.

Hoyer, W. D., Kroschke, M., Schmitt, B., Kraume, K., & Shankar, V. (2020). Transforming the Customer Experience Through New Technologies. *Journal of Interactive Marketing*, 51, 57–71. https://doi.org/10.1016/j.intmar.2020.04.001.

Huang, P., Meng, Z., Guo, J., & Zhang, F. (2017). Tethered Space Robot: Dynamics, Measurement, and Control. (1st ed.) Academic Press. https://doi.org/10.1016/B978-0-12-812309-6.00001-4.

Jenkins, H. (2007). Afterword: The Future of Fandom. In C. Sandvoss, C. L. Harrington, & J. Gray (Eds.), Fandom: Identities and Communities in a Mediated World (pp. 357-364). New York: New York University Press.

Jonassen, D. H. (2000). *Computers as mindtools for schools: Engaging critical thinking* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Kim, T. (2019). Algorithm Design and Implementation of the User-optimized AI Curation System with AR-based characters.

Kim, T., & Song, H. (2022). Communicating the Limitations of AI: The Effect of Message Framing and Ownership on Trust in Artificial Intelligence. *International Journal of Human–Computer Interaction*, 1-11.

Krevelen, R. & Poelman, R. (2010). A Survey of Augmented Reality Technologies, Applications and Limitations. *International Journal of Virtual Reality* (ISSN 1081-1451). 9. 1. https://doi.org/10.20870/IJVR.2010.9.2.2767.

Kwon, S., Kim, H., & Park, K. (2012). Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference. 2012. 2174-7. https://doi.org/10.1109/EMBC.2012.6346392.

Le, D. N., Van Le, C., Tromp, J. G., & Nguyen, G. N. (Eds.). (2018). Emerging technologies for health and medicine: virtual reality, augmented reality, artificial intelligence, internet of things, robotics, industry 4.0.

Liang, L., Zhu, L., Shang, W., Feng, D., & Xiao, Z. (2017, May). Express supervision system based on NodeJS and MongoDB. In *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)* (pp. 607-612). IEEE.

Liu, Y., Gu, Y., Li, J., & Zhang, X. (2017). Robust stereo visual odometry using improved RANSAC-based methods for mobile robot localization. *Sensors*, *17*(10), 2339.

Mack, N., Woodsong, C., Km, M., Guest, G., & Namey, E. (2005). *Qualitative research methods: a data collectors field guide*. ISBN: 0939704986.

Morais A. J., Oliveira, E., & Jorge A. M. (2012) A Multi-Agent Recommender System. In: Omatu S., De PazSziládi, G., Ujbányi, T., & Katona, J. (2016, October). Cost-effective hand gesture computer control interface. In *2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 000239-000244). IEEE.

Molinaro, N., Schena, E., Silvestri, S., Bonotti, F., Aguzzi, D., Viola, E., ... & Massaroni, C. (2022). Contactless Vital Signs Monitoring from Videos Recorded with Digital Cameras: An Overview. *Frontiers in Physiology*, 160.

Morvan, Y. (2009). Acquisition, compression and rendering of depth and texture for multi-view video*. Nano Letters - NANO LETT*. https://doi.org/10.6100/IR641964.

Nielsen, J. (1993). *Usability Engineering*. ISBN: 0125184050, 9780125184052. Boston, USA: Academic Press.

Noh, K., Jeong, E., You, Y., Moon, S., & Kang, M. (2015). A study on the current status and strategies for improvement of web accessibility compliance of public institutions. *Journal of Open Innovation: Technology, Market, and Complexity*. https://doi.org/10.1186/s40852-015-0001-0.

Pudane, M. (2017). Classification of agent-based models from the perspective of multi-agent systems. 2017 *5th IEEE Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)*, 1-6. https://doi.org/10.1109/AIEEE.2017.8270547.

Russell, S. & Norvig, P. (2003). Artificial Intelligence: A Modern Approach. ISBN: 978-0134610993. UK: Pearson

Sarkar, A., Patel, K. A., Ram, R. G., & Capoor, G. K. (2016, March). Gesture control of drone using a motion controller. In *2016 International Conference on Industrial Informatics and Computer Systems (CIICS)* (pp. 1-5). IEEE.

Scully, C. G., Lee, J., Meyer, J., Gorbach, A. M., Granquist-Fraser, D., Mendelson, Y., & Chon, K. H. (2011). Physiological parameter monitoring from optical recordings with a mobile phone. *IEEE Transactions on Biomedical Engineering*, *59*(2), 303-306.

Sibert, L. E., & Jacob, R. J. (2000, April). Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 281-288).

Sukamolson, S. (2007). Fundamentals of quantitative research. *Language Institute Chulalongkorn University*, *1*(3), 1-20.

Sunny, M. S. H., Zarif, M. I. I., Rulik, I., Sanjuan, J., Rahman, M. H., Ahamed, S. I., & Brahmi, B. (2021). Eye-gaze control of a wheelchair mounted 6DOF assistive robot for activities of daily living. *Journal of NeuroEngineering and Rehabilitation*, *18*(1), 1-12.

Tan, W., Liu, H., Dong, Z., Zhang, G., & Bao, H. (2013, October). Robust monocular SLAM in dynamic environments. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 209-218). IEEE.

Trewin, S. (2018). AI fairness for people with disabilities: Point of view. *arXiv preprint arXiv:1811.10670*.

Vanderheiden, G. C., & Henry, S. L. (2003). Designing flexible, accessible interfaces that are more usable by everyone. In *Proceedings of the 2003 CHI Conference on Human Factors in Computing Systems (CHI 2018)*.

Yalon-Chamovitz, S., & Weiss, P. L. T. (2008). Virtual reality as a leisure activity for young adults with physical and intellectual disabilities. *Research in Developmental Disabilities*, *29*(3), 273-287.

Yashoda, H. G. M. T., Piumal, A. M. S., Polgahapitiya, P. G. S. P., Mubeen, M. M. M., Muthugala, M. A. V. J., & Jayasekara, A. G. B. P. (2018, May). Design and development of a smart wheelchair with multiple control interfaces. In *2018 Moratuwa Engineering Research Conference (MERCon)* (pp. 324-329). IEEE.

Walton, D. R., & Steed, A. (2017, November). Accurate real-time occlusion for mixed reality. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (pp. 1-10).

Wang, M.L. & Lynch, J. (2014). Sensor Technologies for Civil Infrastructures.

Wang, B., Lan, J., & Gao, J. (2022). LiDAR filtering in 3D object detection based on improved RANSAC. *Remote Sensing*, *14*(9), 2110.

Wozniak, P., Vauderwange, O., Mandal, A., Javahiraly, N., & Curticapean, D. (2016, September). Possible applications of the LEAP motion controller for more interactive simulated experiments in augmented or virtual reality. In *Optics Education and Outreach IV* (Vol. 9946, pp. 234-245). SPIE.

Weeratunga, A. M., Jayawardana, S. A. U., Hasindu, P. M. A. K., Prashan, W. P. M., & Thelijjagoda, S. (2015, December). Project Nethra-an intelligent assistant for the visually disabled to interact with internet services. In *2015 IEEE 10th International Conference on Industrial and Information Systems (ICIIS)* (pp. 55-59). IEEE.

Zecca, M., Roccella, S., Cappiello, G., Ito, K., Imanishi, K., Miwa, H., Carrozza, M., Dario, P., Takanishi, A. & (2006). *From the Human Hand to a Humanoid Hand: Biologically-Inspired Approach for the Development of RoboCasa* Hand #1. ROMANSY. 487. ISBN: 978-3-211-36064. Germany: Springer. https://doi.org/10.1007/3-211-38927-X_37.

Zhao, Z., Luo, H., Song, G. H., Chen, Z., Lu, Z. M., & Wu, X. (2018). Web-based interactive drone control using hand gesture. *Review of Scientific Instruments*, *89*(1), 014707.

Zheng, J.M. & Chan, K.W. & Gibson, I. Virtual reality (1998), Virtual reality. https://doi.org/10.1109/45.666641

Zukauskas, P., Vveinhardt, J. & Andriukaitienė, R. (2018). Exploratory Research. https://doi.org/10.5772/intechopen.70631.

Appendices

Appendix A - User Questionnaire

Introduction and objectives

First, we would like to show our gratitude for the participation in these tests.

As mentioned previously this test is under the scope of the Master thesis of Information and Enterprise Systems, in which the objective is to test functional concepts to make the applications useful and easy to use.

This session is divided in 3 parts, the first in answering a questionnaire where some personal data will be asked, this will be used for an analysis and comparison between different uses of the application, the second part is the core of the test where it will be asked to use the application freely, and the third part a questionnaire about the evaluation and how you feel concerning the tested prototypes.

Appendix B - Right Grant Form

Consentment

I, ____, carrier of the Passport nº _____________ authorize the audio the processing of my data for studying purposes.

_________, ___ of April 2022

____

(Signature)

Appendix C - Pre-Test Questionnaire

Pre-Test Questionnaire

Initial questionnaire

Sex: ___ (M or F)

Age: ___

Professional situation: ____________________

District of residence: ___________________

Nationality: _______________

Date of birth: ____/____/_______

Appendix D - Questionnaire after test

Final questionnaire

For the first eight questions, answer each affirmation with the value that more adjusts to your experience. It will be used a scale from 1 to 5 where 1 means “I totally disagree” and 5 means “I totally agree”. For question 9 use 1, 2 or 3, being 1 the one with higher priority and 3 the one with less.

1. The application is easy to use.

   ○ 1 ○ 2 ○ 3 ○ 4 ○ 2

2. The application achieves the objectives.

   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3. I was satisfied with the application.

   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

4. The application will be used in the future.

   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

5. The several functions of the application are well integrated.

   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

6. The application shows a lot of inconsistency.

   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

7. The users will learn quickly to control the application.

   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

8. It was necessary to learn many new things in order to use the application.

   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

9. Which of the 3 motions controllers do you consider to have more priority of use? (Fill the boxes with the values 1,2 and 3, being 1 the one you give more priority and 3 the one with less).

   Full body controller [ ]

   Hand motion controller [ ]

   Eye gaze controller [ ]