IMPERIAL COLLEGE OF SCIENCE,
TECHNOLOGY AND MEDICINE

DEPARTMENT OF ELECTRICAL AND ELECTRONIC
ENGINEERING

FINAL YEAR PROJECT

Interim Report: Augmented Reality for Human Robotic Interaction

Authors:
Aufar LAKSANA
CID: 01093575

Project Supervisor:
Dr. Yiannis Demiris

January 28, 2019

Contents

Chapter 1

Introduction and Requirements

1.1 Introduction

This report was written as part of the Final Year Project for the MEng. Electronic & Information Engineering course. The project is supervised by Dr. Yiannis Demiris at the Imperial College London. The content of the report covers the research and progress of the project so far, between October 2018 until January 2019.

1.2 Motivation

A study on powered wheelchair users showed that there were approximately 3.6 million wheelchair users in the United States alone (Kairy et al. 2014). The study also showed that approximately 30% of the users were operating powered wheelchairs (PWCs) or scooters, and that similar data had been reported in Europe. According to a report examining the recent trends amongst adults aged 65 and older in the United States, the number of elderly adults is projected to more than double from 46 million to over 98 million by 2060; due to increased life expectancy stemming from better healthcare and a reduction in mortality rate at older ages (Mather et al. 2015). As a result of the growing elderly population, it is likely that the number of powered mobility devices will continue to grow.

The study by Kairy et al. (2014) also highlights the problems faced by powered wheelchair users (PWUs). PWUs are often afraid of navigating in crowded areas, or are unable to operate their device safely, due to visual, motor and cognitive disabilities. In order to address these issues, the implementation of smart or intelligent wheelchairs has been proposed. These smart wheelchairs will help the users by providing services such as navigation assistance, allowing the user to carry out daily activities with more ease. An example of navigation assistance is collaborative control, Carlson & Demiris (2012) which utilizes a smart system that recognizes and assists the user when they require help, by manipulating the control signals of the powered wheelchair.

Within the Personal Robotics Lab at the Imperial College London, a lot of work has been done on enhancing the powered wheelchair user experience. One approach, Zolotas et al. (2018) involves the use of an augmented reality (AR) headset to help the user understand their wheelchair's behaviour. The AR headset renders helpful indicators, such as the trajectory of the wheelchair and highlighting potential obstacle collisions.

This project explores the idea implementing a smart system that would further benefit PWUs, by allowing them to navigate in crowded areas and recognizing locations that are frequently visited, such as at home or the shopping mall, and building a map of the location to allow better navigation assistance. An AR headset can also be utilized to display the internal state of the smart wheelchair, such as highlighting objects that determine the frequently visited location, or alerting the user to people moving towards the wheelchair and rendering a suggested path to avoid collision.

1.3 Project Specification

The aim of this project is to design and build a system that will allow powered wheelchair users to more easily conduct routine tasks, such as navigating around the house, or other frequently visited locations, such as the grocery store.

This project involves several hardware components, all of which are available within the Personal Robotics Lab. The hardware includes the following, as well as the sensors already mounted on the wheelchair:

- Powered Wheelchair
- Microsoft Hololens

The system being developed is divided into two major parts, Robotic Behaviour and Augmented Reality Visualization.

1.3.1 Robotic Behaviour

The goal for this section of the project is to design and develop algorithms that will allow for assistive navigation on the powered wheelchair. The system will utilize sensors mounted on the wheelchair to build up a map of the surroundings. Objects in the surrounding area will be detected and marked as potential collisions, depending on the trajectory of the wheelchair. An extension is the ability to detect moving objects, such as people, calculating the trajectory of the object and deciding if a collision is imminent.

A major hardware component of this project is the Microsoft Hololens, a mixed reality headset that can be worn by the powered wheelchair user. The Hololens posses the ability to track the eye movements of the user with an eye-tracker addon. An interesting concept that can be explored is the ability to control the powered wheelchair using the eye-tracker, removing the need for a joystick, allowing users who lack the motor skills to operate the wheelchair. This feature can also be used to check if the user has noticed an object that may cause a collision. Should the user not see the object, the system will first highlight the object, before taking over from the user

1.3.2 Augmented Reality Visualization

Using the Hololens, the goal of this section is to communicate to the user the internal state of the system controlling the powered wheelchair. Using augmented reality, visualizations will be rendered on the Hololens, allowing the user to understand the trajectory of the wheelchair, what potential collisions may occur. The system will also be able to take over control of the wheelchair, as such, it would be beneficial if a warning was displayed to the user right before the system takes over.

Other visualizations include highlighting moving objects and tracking them as they move across the field of vision of the user. Should the user not notice a moving object that may cause a collision, the Hololens will flash a warning and highlight the object in order to attract the attention of the user, allowing them to make adjustments themselves.

Chapter 2

Literature Review

2.1 Object Detection

In this project, the main purpose of object detection algorithms will be to identify objects of the class 'Human' in the surrounding area of the wheelchair. In order to tackle the problem of users being unwilling to navigate in crowded areas, a system must be implemented to help track and navigate around humans.

One of the modern approaches to detecting humans in an image from a camera is to use Deep Convolutional Neural Networks. Human detection,as stated by Vidanapathirana (2018), is a special case of Object Detection and Object Localization. The system would ideally be able to pinpoint the location of the human object relative to the wheelchair, in order to be able to calculate the trajectory of the person.

However, the use cases of object detection is not only limited to detecting humans. There has been research conducted into the use of object detection for recognizing the scene around a robot, ie. the system is able to identify the robot is now in a hallway (Quattoni & Torralba 2009, Espinace et al. 2010).

2.1.1 Human Detection

The main challenges with detecting humans in an image is the large variations in the appearance. A frontal view of a person may be recognized by the algorithm, but a side view may cause problems due to the algorithm not recognizing key features from a different angle (Dalal & Triggs 2004).

SIFT A solution to this problem is outlined by Lowe (2004), in an approach named the Scale Invariant Feature Transform (SIFT), which transforms image data into co-ordinates which are scale invariant relative to local features. This method allows for a large number of features to be extracted from an image. The feature are also distinct, allowing for a single feature to be correctly matched with a low uncertainty against a database of existing features.

A large quantity of features is required for object detection, due to the often cluttered nature of the image. In order to detect a small object in the background, at least 3 features must be correctly matched for reliable identification.

In order to perform object detection, SIFT features are first extracted from the training set of available images. A new image, in our case, a human torso/body, will be recognized by individually comparing the features in the test image with the features in the training set, in a Nearest Neighbours approach utilizing Euclidean distance. However, a drawback of using SIFT is the high computational cost of comparing the features (Wang et al. 2011).

YOLO You Only Look Once (YOLO) is a state of the art object detection system relying on a single neural network to predict bounding boxes around the objects (Redmon et al. 2015). The main advantage of YOLO is that it is extremely fast, as the name, the algorithm only looks once at an image to predict what objects are present.

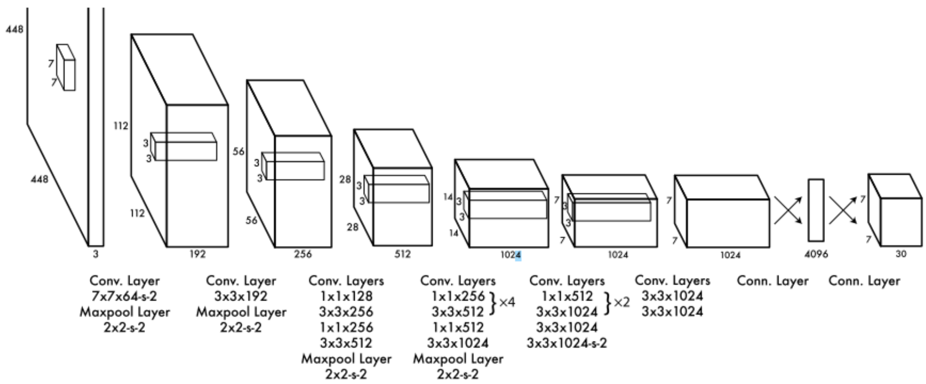

Figure 2.1: The YOLO architecture (Redmon et al. 2015)

The Convolutional Neural Network (CNN) in the YOLO system takes in an image and resizes it to 448x488 pixels. The image is then passed through the CNN layers and is output as a 7x7x30 tensor, containing the co-ordinates of the bounding boxes and the probability distributions over all the classes the network is trained on.

The YOLO algorithm has been used to recognize human actions by Shinde et al. (2018), using the LIRIS Human Activities dataset, which contain human actions such as shaking hands and entering rooms. It was found in the study that only a few frames of a streamed video is required for the YOLO algorithm to recognize the human actions.

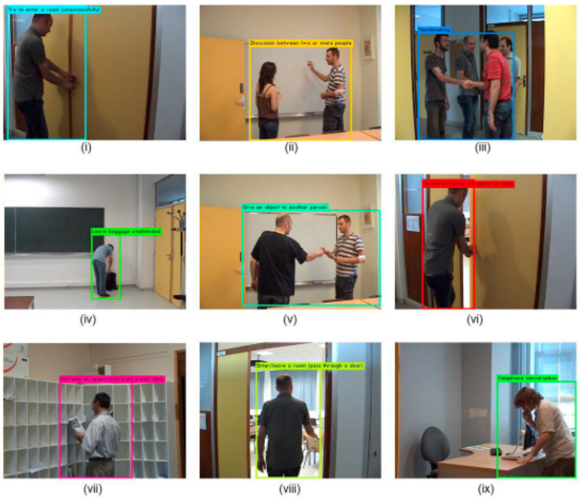

Figure 2.2: The YOLO algorithm recognizing human actions (Shinde et al. 2018)

An extension of this work for the purposes of this project would be to detect humans walking towards the camera mounted on the wheelchair, in order to help with navigating through crowded areas.

2.2 Simultaneous Localization And Mapping (SLAM)

The term mapping refers to a system that will create a map of the surrounding areas, by detecting objects such as walls and other obstacles. In order to help users navigate, the system must analyse the surroundings for potential dangers. As such, it is important to build up a thorough and complete map.

A fundamental method for robot navigation is the Simultaneous Localization And Mapping (SLAM) method. The process allows the system to predict the trajectory of the robot and the location of all objects on-line, without the need of an *a priori* knowledge of the robots location (Bailey & Durrant-Whyte 2006). The method estimates the pose of the robot relative to landmarks which are detected. The popularity of SLAM increased with the emergence of indoor applications of robotic devices. For this project, it is expected that the user will be mostly navigating around the house or indoors, which rules out the use of GPS to bound the localization errors (Cadena et al. 2016).

2.2.1 Existing Work

A review of SLAM techniques can be found in Cadena et al. (2016), which also outlines the standard formulation of the SLAM problem as that of a Maximum a posteriori (MAP) estimation. The formulation relies on Bayes theorem, and using the prior knowledge of the robots pose to maximize the likelihood to estimate the current position of the robot. The variables required to estimate the position are the robot poses, the position of landmarks and the calibration parameters of the sensors.

In order to build an accurate map of the surroundings, the calibration of the sensors providing the measurements is a crucial step. The choice of sensors also matter, as the type of data returned by the sensor may affect the computational complexity of the SLAM algorithm. As such, it is common to have a module in the system that deals with the extraction of relevant features from the sensor data.

A fairly common assumption in SLAM approaches is that the world is static and remains unchanged as the robot moves. This becomes an issue with the goal of this project, which hopes to achieve the ability to detect human objects walking around the wheelchair. This issue will be addressed in a later section.

2.2.2 Visual SLAM

Visual SLAM (vSLAM) is an implementation of SLAM that relies on visual inputs only. As stated in Taketomi et al. (2017), vSLAM is suitable for AR due to the low computational algorithms that can be implemented on the limited resources of an AR headset. The technique of vSlam is mainly composed of three modules:

Initialization In the initialization stage, camera pose estimation is conducted, to transform objects in a 2D image from the camera into a 3D co-ordinate system that the robot understands. This process determines the position and orientation of the camera relative to the object. A part of the environment is reconstructed as part of the initial map using the global co-ordinate system of the robot.

Tracking Here, the reconstructed map is used to estimate the pose of the camera with respect to the map. Feature mapping or feature tracking is conducted on the images in order to get a 2D-3D correspondence between the image and the map. The camera pose can then be calculated from the correspondences by solving the Perspective-n-Point problem (Nistér 2004). This allows the system to identify where on the map the robot currently is.

Mapping When the robot passes through an environment that has previously not been mapped, the 3D structure of the surroundings is calculated from the camera images. The structures are then added to the existing map of the environment.

2.3 Scene Recognition

Wheelchair users spend a substantial part of their time at home. An active area of research called assistive domotics (Rosslin & Tai-hoon 2010), involves developing assistive technology and home automation systems that allow users to interact with common household objects from a seated position in the wheelchair.

The ability for a smart wheelchair to recognize a room in the house would greatly improve the user experience. A map of the surroundings that was previously built of that room can now be loaded when the wheelchair enters the room again, greatly reducing the amount of computational overhead in rebuilding the entire map. The stored maps may contain useful information such as the positions of doorways or other objects of interest that the users can interact with.

Quattoni & Torralba (2009) proposed a scene recognition model that is specifically built for recognizing indoor scenes, by using image prototypes to learn a distance function to map indoor scenes to their correct labels. Scenes containing similar objects tend to have the same labels, and it was found that some objects are more important than others in determining a scenes identity. For instance, a library will contain many bookshelves, whereas a kitchen is unlikely to. A similar method is proposed by Espinace et al. (2010), which relies on object detection to correctly classify scenes. The intuition is that objects can be detected in real time, and using contextual relations, it is possible to associate objects with scenes.

2.4 Head Mounted Display and Control

2.4.1 Microsoft Hololens

The Microsoft Hololens is an untethered holographic computer, allowing for the display of 3D holograms pinned to real world objects. The Hololens is equipped with an array of sensors, making it an ideal choice of hardware for this project.

Figure 2.3: The Microsoft Hololens hardware (Zeller 2018)

Hardware Specifications The Microsoft Hololens contains the following sensors:

- 1 Intertial Measurement Unit (IMU)
- 4 Environment understanding cameras
- 1 Depth Camera
- 1 2MP Photo/HD video camera
- Mixed reality capture
- 4 Microphones

- 1 Ambient Light Sensor

A full technical specifications for the hardware is available from Microsoft (2015).

Development Mixed Reality applications are developed using the Universal Windows Platform. A computer running Windows 10 will be required to develop applications, since programs such as Unity and Microsoft Visual Studio will need to be installed.

Gaze The concept of Gaze is that of a form of targeting in mixed reality applications (Microsoft 2018*a*). It lets the system to know where the user is looking in the world, and from there, allows the system to discern their intent. Users tend to look at the object or location that they will interact with.

Using the Hololens, a mixed reality application can determine whether a user is currently not looking at an object, allowing the application to give a visual/audio cue to the user to look at the object. This can be used in this project to alert the user to a potential collision with a person that is approaching the wheelchair if the user has not spotted the person already.

No Eye Tracker The Hololens can keep track of its location and rotation relative to the environment, but is unable to capture eye gaze data. This is because there are no internally facing cameras. The user can interact with virtual objects by using a virtual cursor, controlled by changing their physical position and head rotations. Applications can be developed to estimate where the user is looking while wearing the device.

Relying on head gaze to estimate the point that a user is looking at is limited in its accuracy, since the users eyes can move away from the centre of the Hololens, so the user is actually looking at something else. In the research by van der Meulen et al. (2017), an extra eye-tracker from Pupil Labs was attached to the Hololens. It was found that the addition of the eye-tracker greatly increased the accuracy in gaze location calculations. For virtual targets, the eye-tracker had a minimal effect on the estimation, and can be estimated by head rotation alone.

2.4.2 Augmented Reality Visualization

The Microsoft Hololens is able to blend real world and virtual content into environments where digital and physical objects can co-exist and interact. The term 'Mixed Reality' was first introduced by Milgram & Kishino (1994), and refers to the blending of the physical and virtual worlds.

The Hololens allows the developer to create 'Holograms', which are objects of light and sound that are displayed by the headset. Users are able to interact with the holograms through voice, gaze and gestures. Enhanced environment apps are applications that facilitate the placement of digital information on the user's current environment (Microsoft 2018*b*). An example of an enhanced environment application is placing markers in augmented reality on objects that the user can interact with in both the physical and digital worlds.

2.4.3 Gaze/Eye-Tracking Control

The idea of gaze based control has already been explored in the Personal Robotics Lab (Chacón-Quesada & Demiris 2018), by also using the Microsoft Hololens as an AR head-mounted display. It was shown that the user could control the smart wheelchair using the AR user interface, allowing them to navigate through doors and approach people.

Figure 2.4: The AR User Interface demonstrated by Chacón-Quesada & Demiris (2018)

The implemented system is aware of both humans and other objects in the surroundings, and allowing the user to select various options to interact with the objects, as seen in Figure 2.4.

Montenegro-Couto et al. (2018) uses a concept whereby they use an eye-tracker to control the wheelchair motors via gaze rather than a joystick. The options to control the wheelchair are displayed on a screen, and the eye-tracker calculates the 2D co-ordinates of where the person is looking at on the screen. Various options are listed such as the ability to manoeuvre the wheelchair, as well as interact with nearby objects.

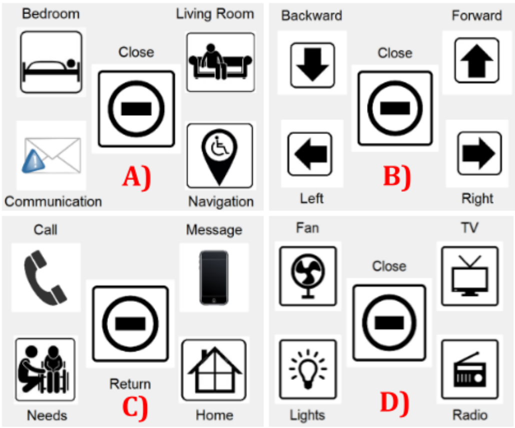

Figure 2.5: The screen displayed to users in Montenegro-Couto et al. (2018)

2.5 Competing Products

One of the main issues with gaze/eye based control is that of the 'Midas touch' problem, where every gaze does not equal a goal. The user may look at an object for a split second, but may not actually want to move towards that object. To counteract this problem, Wästlund et al. (2010) developed a system which displays on-screen buttons that control the wheelchairs movements. The system also stopped the wheelchair when the device approaches an object.

A similar idea was proposed by Arai & Mardiyanto (2011), whereby an eye-tracker would detect the position of the pupil and translate it onto an invisible control panel similar to that of Wästlund et al. (2010). To avoid the Midas touch problem, a sustained gaze at one of the controls was required before the wheelchair would move.

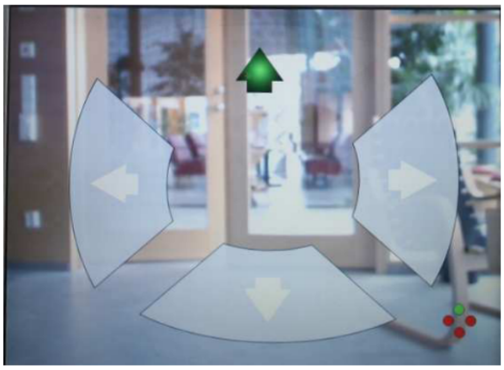

Figure 2.6: The on-screen buttons displayed to users (Wästlund et al. 2010)

Raymond et al. (2018) proposes a system which utilizes a depth camera in conjunction with an eye tracker. The system identifies between eye movements aimed at the floor as a gaze target and other non-navigational eye movements. This approach is different as it does not rely on an artificial user interface to control the wheelchair as proposed by Arai & Mardiyanto (2011), Wästlund et al. (2010).

Chapter 3

Implementation Plan

This section of the report goes over the overall project goals, as well as the steps that will need to be taken in order to achieve the goals within the given time frame.

3.1 Overall Project Goals

As mentioned in Section 1.3, the goal of this project is to develop an intelligent powered wheelchair capable of recognizing objects and humans in its vicinity. By building a map of its surroundings, and through object detection, useful prompts can be displayed to the user. The ability to detect people will greatly decrease the possibility of a collision when navigating through more crowded spaces, and can be achieved by displaying alerts on the head-mounted AR device or through the system taking over control of the wheelchair to avoid the object.

Furthermore, the Microsoft Hololens forms a key part of the overall system design. By making use of the sensors available on the headset, such as the depth camera, as well as the Gaze method to estimate where the user is looking, another goal for this project is to implement a wheelchair control system that solely relies on head pose, rotation and gaze, removing the need for a joystick and allowing users who are paralysed to more easily control their powered wheelchair.

3.2 Hardware Analysis

Before any algorithms can be developed, it is important to understand the hardware available. For instance, camera resolution will affect the image quality, which may become an issue for object detection. As such, a proper analysis of any cameras, sensors and other pieces of hardware should be conducted.

3.2.1 Powered Wheelchair

The Personal Robotics Lab have built the Assistive Robotic Transport for Adults (ARTA), a smart powered wheelchair, which has been used for several other research projects (Chacón-Quesada & Demiris 2018, Zolotas et al. 2018). From the reports, it can be seen that ARTA already has some built-in obstacle avoidance and shared control methodologies.In addition, it is also equipped with laser range finders, which can return depth information. Chacón-Quesada & Demiris (2018) also utilizes the Hololens camera to capture images of the surroundings, but the main bulk of the image processing is done on a remote computer, which communicates with the head mounted display (HMD).

Further investigation into the sensors and other hardware on the wheelchair will need to be conducted, by asking members of the Personal Robotics Lab who have worked with the wheelchair. Official documentation for the wheelchair is sparse, and it would be less time consuming to ask people directly. Additionally, more sensors can be mounted on ARTA, should the need arise.

3.2.2 Microsoft Hololens

The Microsoft Hololens is a key component of the project, as it will be used to both obtain images and display AR visual cues by rendering holograms. The official documentation (Microsoft 2018*b*) is available online, along with many tutorials on websites such as Youtube and Stackoverflow.

Pupil Labs Eye-Tracker

The Personal Robotics Lab also has access to an add-on for the Hololens, an eye-tracker developed by Pupil Labs. The cameras on the add on run at 120hz and low latency eye tracking software is available from the company, allowing for applications that require fast interaction (Pupil Labs n.d.).

Development

Developing applications for the Hololens requires a computer running Windows 10, and installations of Microsoft Visual Studio, Unity and several other programs (Microsoft 2018*b*).

3.2.3 Hardware Access and Additional Hardware

As ARTA and the Hololens can not leave the Personal Robotics Lab, time slots for development time will need to be arranged with the other members of the lab. Furthermore, after analysing the sensors available on the wheelchair, more sensors may need to be requested. At the time of writing, a Xbox Kinect Camera is being considered as another camera that may be mounted on ARTA.

3.3 Augmented Reality Wheelchair Control

Utilizing the Microsoft Hololens and the Pupil Labs eye-tracker add-on, the goal of this part of the project is for the user to have the ability to stare at a location in front of the wheelchair, and the system will recognize the intent to move towards a goal. This intent is then converted to wheelchair control signals and the user will be able to navigate the powered wheelchair without the use of a joystick.

3.3.1 Cross Platform Interface

The wheelchair control relies on the Robotic Operating System (ROS), which runs on a standard Linux distribution such as Ubuntu 16.04. On the other hand, the Hololens runs on the Universal Windows Platform. As such, an interface between the two operating systems is needed. Members of the Personal Robotics Lab have already developed these interfaces for their previous projects.

3.3.2 Eye-Gaze Estimation

To allow the user to fixate their gaze on a specific point and navigate the wheelchair towards that point, an eye-gaze estimation system will need to be implemented. Using the eye-tracker and depth camera on the Hololens, and approach similar to the one used by Raymond et al. (2018) will be developed. It has been noted that head-gaze can be used to determine a goal destination, but as shown through the research of van der Meulen et al. (2017), there is a significant increase in the accuracy when used in conjunction with an eye-tracker, and it will also greatly reduce the Midas touch problem.

3.4 Human Detection System

Previous research by Chacón-Quesada & Demiris (2018) utilizes the Detectron object detection method (Girshick et al. 2018) and conducts the processing on a remote computer. For this project, the goal of the human detection would be to detect the human in the image, and using depth cameras or laser range finders, determine if the person is moving closer towards the wheelchair and if a collision is imminent.

A personal interest of mine is to work more with Deep Learning and Neural Networks in order to gain a better, practical understanding of the concept. As such, for this project, I would like to work with the YOLO object detection method to be able to recognize humans and track them across frames (Shinde et al. 2018, Redmon et al. 2015).

3.4.1 Testing Pre-Trained Models

There are already pre-trained YOLO networks capable of recognizing people. However, these need to be tested with the Hololens camera to see if the images produced by the Hololens will allow the networks to recognize people.

3.4.2 Estimating Moving Human Trajectories

In order for the wheelchair to be aware of potential collisions with humans, the people in the vicinity must be detected and considered as a landmark by the

Visual SLAM system. A system utilizing vSlam and a Moving Object Detector similar to the work done by Wang et al. (2011) will be employed to locate the positions of people in the surroundings, and estimate their trajectories.

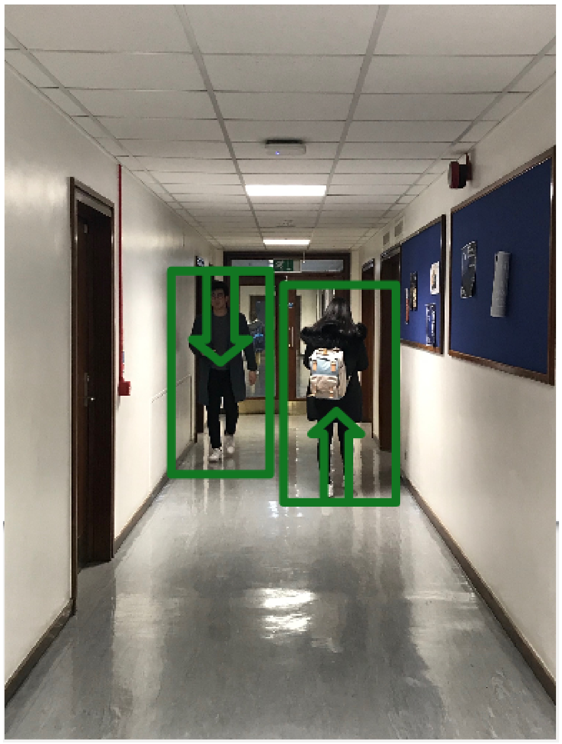

Figure 3.1: Bounding Boxes and Arrows indicating trajectory

3.5 Obstacle Avoidance System

The goal of the obstacle avoidance system is for the wheelchair to be intelligent in detecting obstacles in the surroundings. The system will encompass the Human Detection System mentioned in Section 3.4 as well as using range finders and mapping techniques to prevent the user from colliding with walls and other obstacles.

3.5.1 Utilizing Existing Work

Previous work by Zolotas et al. (2018) uses various software and hardware components atop the Robotic Operating System, which estimate the trajectory of the wheelchair and render holograms on the HMD showing the current path, as well as a suggested path should the current trajectory result in a collision.

An extension of this would be for the system to realize when to take over from the user should they not respond to the visual warnings. The system will then manoeuvre the wheelchair into a trajectory it deems safe that will avoid the oncoming object.

3.5.2 Hologram Warnings

Utilizing the display on the Microsoft Hololens, once a human or object is detected, it will be tracked as it moves across the users field of view. However, users may not always spot the oncoming object, due to being distracted or looking elsewhere.

The Hololens and the Pupil Labs eye-tracker allow for a system to be developed which will know when the user has not spotted the oncoming object. When the system realizes this, a warning hologram will be rendered on the HMD, highlighting the oncoming object and alerting the user with a sound from the direction the object is coming from.

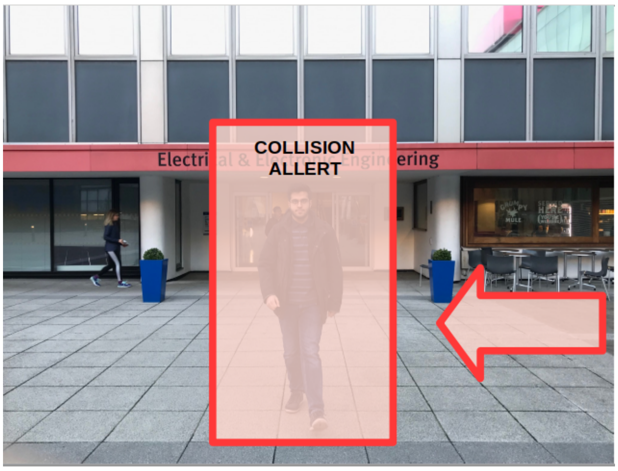

Figure 3.2: A mockup of the hologram rendered on the HMD

Figure 3.2 shows the potential holograms that will be rendered on the Microsoft Hololens screen. The position of the pink arrow will vary, and will be rendered at the spot where the user is currently looking, in order to attract their attention to the collision.

3.6 Project Timeline

The project timeline is listed in the Appendix Figures 1 and 2. The Gantt chart displays the project outline, and what tasks are dependant upon each other.

3.7 Fallback

Should time constraints become an issue, the part of the project that can be reduced is the Obstacle Avoidance System outlined in Section 3.5. It may become too burdensome to implement a system that will take over control from the user and work reliably. As such, just displaying the warning to the user in the form of a hologram and forcing the wheelchair to stop if they do not pay attention to the warning will be a suitable reduction.

This will leave more time for testing of the Human Detection System (Section 3.4) and the Augmented Reality Wheelchair control (Section 3.3), which are considered as core components of this project, and will be demoed during the presentation.

3.8 Extensions

The following section outlines the possible extensions to the project, should time allow. The technical implementation details have not been fully researched, but a basis has been formed from the literature review.

3.8.1 Scene Recognition

As mentioned in Section 2.3, Scene Recognition can be of used to aid powered wheelchair users conduct routine tasks. For instance, by recognizing the device has returned to an area marked as the user's home, it may be a good idea to turn off the human detection system, since it is unlikely that the user will encounter many people. This would reduce in the amount of computation going on in the background. Another use may be to load stored maps of an area that has been previously explored, such as a map of the house, and only unexplored or changed areas would be updated of the map.

As explored by Quattoni & Torralba (2009), as well as Espinace et al. (2010), object detection algorithms can be used to recognize scenes. By training the YOLO object detector to recognize objects other than humans, we may be able to map a relationship between objects and scenes.

This would involve training another convolutional neural network to recognize objects in an image from the Hololens, and associating that object with a label. An example of this would be recognizing pots and pans and labelling the scene as a kitchen. As such, an indoor scene dataset would be required to train the network, which would have to be gathered manually and labelled. An alternative would be to make use of indoor scene datasets previously used by Quattoni & Torralba (2009).

3.9 Overall System Architecture

The high level overview of the project is displayed in the Appendix, Figure 3. An onboard computer running Linux will do most of the processing, such as the gaze-estimation and human detection. The Hololens will provide the Linux system with gaze and eye-tracker data, as well as the images for the YOLO human detector.

The individual units will communicate with the wheelchair control unit running on the Robotic Operating System. Stop signals are sent to the control unit when the human detector estimates that a collision is about to happen. The Eye-Gaze Estimation unit will send driving instructions to reach a point the user is looking at.

Chapter 4

Evaluation Plan

4.1 Human Detection Test

To be able to accurately test the recognition rate of the Human Detection System, we must gather some test data that the YOLO network was not trained on.

4.1.1 Gathering Test Data

Using the Hololens camera, video recordings of people walking in the corridors and walkways around the Imperial College London campus will be gathered. A few sample test locations are shown in Figure 4.1. A mixture of indoor and outdoor scenes will be captured, since this will best reflect the possible locations a wheelchair user may visit. These videos will be kept as a test set for the human detection system.

The reasoning behind using pre-recorded videos to test the system this is so that a small test-bench can be developed, so that after the network is trained, a quick test run can be initiated to see if the recognition accuracy has improved. This will save a large amount of time, since the alternative involves transferring the network onto the onboard Linux computer and communicating with the Microsoft Hololens and live images. Furthermore, it will be difficult to recreate test conditions using live images, and it would involve multiple people who would serve as actors.

(a) Queens Lawn Walkway

(b) Senior Common Room Walkway

(c) EEE Entrance (Indoors)

(d) Sherfield Walkway

Figure 4.1: Some sample test locations where video can be captured

4.1.2 Testing with Pre-Recorded Data

A test-bench will be developed which allows the developer to visually see the results of the object detection. The YOLO method draws bounding boxes around recognized objects, as shown in Figure 4.2.

To create the test-bench, after the video is recorded,the number of people walking in each video will have to manually be counted, before labelling each file with that number. This is so that it is possible to test the number of recognitions the algorithm finds in each video.

Figure 4.2: HMD Yolo Bounding Boxes

4.2 Eye-Gaze Control Test

To test the Eye-Gaze control, the ARTA wheelchair is required. The user will have to sit in the wheelchair and fully operate the device as if they were a disabled person.

4.2.1 Initial Test: Movement

The initial test will be conducted in a corridor within the Electrical and Electronic Engineering Department. The goal of the test will be for the user to operate the wheelchair by fixing their gaze on a marked point at the end of the corridor. The wheelchair will then move along the corridor until it has reached the marked point, before stopping (Figure 4.3).

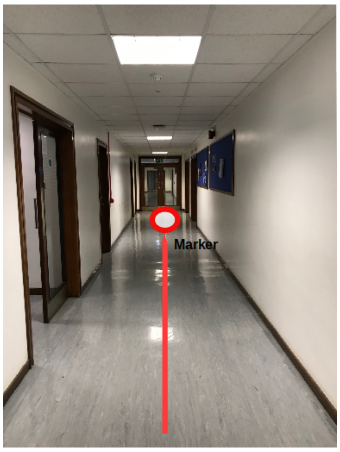

Figure 4.3: The marker and the ideal path for the wheelchair to take

4.2.2 Advanced Tests

To fully test the navigational capabilities of the eye-gaze system, more complex tests such as turning around corners or entering doors must be conducted. A list of potential indoor tests are listed below:

- Navigate to the end of a corridor and turn left/right.
- Navigate to the end of a corridor through a door.
- Navigate down a corridor and turn left/right into a door.
- Navigate towards a marker, lose sight of point, continue to navigate to that point.
- Navigate down corridor with people around without colliding with anyone.

4.3 Obstacle Avoidance Test

The obstacle avoidance tests can be divided into two sections, display and control. The display tests rely solely on the Hololens, rendering holograms and playing sound when a person or object will collide with the wheelchair. The control tests will require the use of the wheelchair, and can be tested in the full system test.

4.3.1 Collision Warnings

To test the collision warnings, we can play pre-recorded videos of people moving towards the camera. The system will have to detect when the person is getting too close and if a collision is possible.

If a collision is possible, the Hololens hologram renders would go through a process such as the one shown in Figure 4.4, whereby the object is tracked as it moves closer towards the camera.

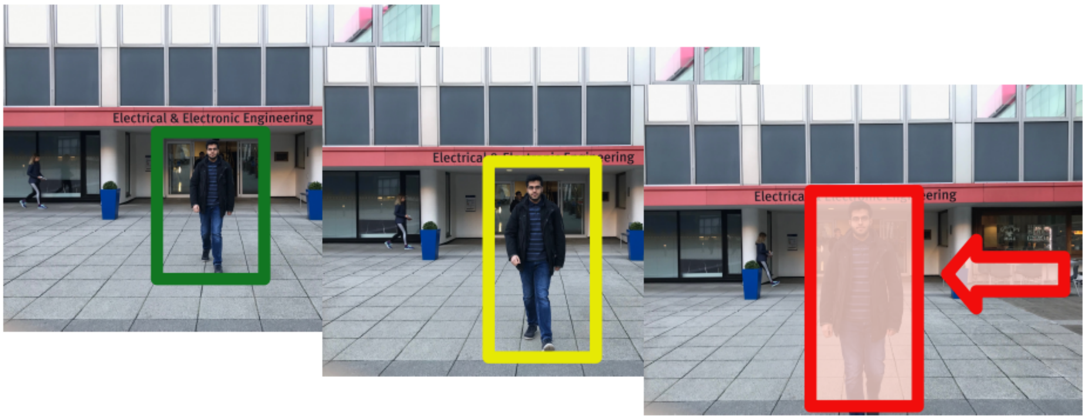

Figure 4.4: The obstacle avoidance render with collision warning

4.4 Full System Test

The full system test is a combination of all the tests in this section. It tests the smart wheelchairs ability to be controlled by eye-gaze, as well as its human and object detection capabilities to stop the wheelchair from colliding with a landmark.

An obstacle course of sorts will be setup featuring all the elements from previous tests. The goal of the course will be to navigate through the corridor and turn into a doorway, all whilst avoiding moving humans and maneuvering around obstacles in the way, such as cardboard boxes. The wheelchair would should be able to stop before it hits an object, as well as display warnings to the user right before.

Chapter 5

Ethical, Legal & Safety Plan

5.1 Safety

5.1.1 Physical Safety

For safety reasons, the user operating the wheelchair should be physically capable of taking control of the joystick. This is especially important when conducting some of the advanced tests (Section 4.2.2). These tests involve the use of people walking in vicinity of the powered wheelchair. The user should have access to a kill switch to stop the wheelchair instantaneously, should a collision be imminent and the system fail to stop the wheelchair autonomously.

Furthermore, the speed of the wheelchair should be limited to a slow pace, since the wheelchair will be operated indoors and around other people. It is important to make sure that any human participants in the test are aware of the hazards of the wheelchair so they can take appropriate action, such as being ready to step out of the way should the wheelchair not stop.

The corridors used for testing the wheelchair must be clear of all non-test obstacles, and must be cordoned off so that unknowing passer-bys do not enter the testing area and be in danger of collision.

5.2 Legal

5.2.1 Recording People for Test Data

In order to test the Human Detection System, video recordings of people walking along walkways will need to be taken. To avoid any breaches of GDPR, the following precautions must be taken:

- Before recording, inform people in the area that we will be recording. People who do not wish to be recorded will be given time to leave.
- Ensure there are alternative routes to the paths being recorded, so people can avoid being recorded.
- Should no one want to be recorded, enlist the help of willing students as actors, who will recreate the scene of people walking in and out of the shot.
- Avoid recording in areas where minors frequent. The Queens Tower Lawn often has schoolchildren on their lunch break, so efforts will be made to avoid that location.

If people do not consent to recording after a recording has been completed, effort will be made to blur out the faces of that individual.

Bibliography

Arai, K. & Mardiyanto, R. (2011), 'A prototype of electric wheelchair controlled by eye-only for paralyzed user', *Journal of Robotics and Mechatronics* **23**(1).

Bailey, T. & Durrant-Whyte, H. (2006), 'Simultaneous localisation and mapping (SLAM): Part I - The essential algorithms.', *IEEE Robotics and Automation Magazine* **13**(2), 99–108.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I. & Leonard, J. J. (2016), 'Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age', *IEEE Transactions on Robotics* **32**(6), 1309–1332.

Carlson, T. & Demiris, Y. (2012), 'Collaborative control for a robotic wheelchair: Evaluation of performance, attention, and workload', *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **42**(3), 876–888.

Chacón-Quesada, R. & Demiris, Y. (2018), 'Augmented Reality Control of Smart Wheelchair Using Eye-Gaze-Enabled Selection of Affordances', pp. 1–4.
URL: *http://www.imperial.ac.uk/personal-robotics/robots/*

Dalal, N. & Triggs, W. (2004), 'Histograms of Oriented Gradients for Human Detection', *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR05* **1**(3), 886–893.

Espinace, P., Kollar, T., Soto, A. & Roy, N. (2010), 'Indoor scene recognition through object detection', *Proceedings - IEEE International Conference on Robotics and Automation* pp. 1406–1413.

Girshick, R., Radosavovic, I., Gkioxari, G., Dollár, P. & He, K. (2018), 'Detectron'.
URL: *https://github.com/facebookresearch/detectron*

Kairy, D., Rushton, P. W., Archambault, P., Pituch, E., Torkia, C., El Fathi, A., Stone, P., Routhier, F., Forget, R., Demers, L., Pineau, J. & Gourdeau, R. (2014), 'Exploring powered wheelchair users and their caregivers' perspectives on potential intelligent power wheelchair use: A qualitative study', *International Journal of Environmental Research and Public Health* **11**(2), 2244–2261.

Lowe, D. G. (2004), 'Distinctive Image Features from Scale-Invariant Keypoints', *International Journal of Computer Vision* pp. 1–28.

Mather, M., Jacobsen, L. A. & Pollard, K. M. (2015), 'Aging in the United States', *Population Bulletin* **70**(2).

Microsoft (2015), 'Microsoft HoloLens HoloLens Device Specifications'.

Microsoft (2018*a*), 'Gaze - Mixed Reality — Microsoft Docs'.
URL: *https://docs.microsoft.com/en-us/windows/mixed-reality/gaze*

Microsoft (2018*b*), 'Windows Mixed Reality Documentation'.

Milgram, P. & Kishino, F. (1994), 'A TAXONOMY OF MIXED REALITY VISUAL DISPLAYS', *IEICE Transactions on Information Systems* **E77**(12), 1–15.

Montenegro-Couto, E. H., A. Hernandez-Ossa, K., L. C. Bissoli, A., Sime, M. & F. Bastos-Filho, T. (2018), 'Towards an Assistive Interface To Command Robotic Wheelchairs and Interact With Environment Through Eye Gaze', *Anais do V Congresso Brasileiro de Eletromiografia e Cinesiologia e X Simpósio de Engenharia Biomédica* (January).
URL: *https://www.even3.com.br/anais/cobecseb/78867*

Nistér, D. (2004), 'A Minimal Solution to the Generalised 3-Point Pose Problem', *Journal of Mathematical Imaging and Vision* **27**(1), 560–567.

Pupil Labs (n.d.), 'Pupil Labs - VR AR'.
URL: *https://pupil-labs.com/vr-ar/*

Quattoni, A. & Torralba, A. (2009), 'Recognizing Indoor Scenes', *International Surgery* **56**(3), 182–186.

Raymond, L.-a., Piccini, M., Subramanian, M., Pavel, O. & Zito, G. (2018), 'Natural Gaze Data-Driven Wheelchair'.

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2015), 'You Only Look Once: Unified, Real-Time Object Detection'.
URL: *http://arxiv.org/abs/1506.02640*

Rosslin, J. R. & Tai-hoon, K. (2010), 'Applications, Systems and Methods in Smart Home Technology : A Review', *International Journal of Advanced Science and Technology* **15**(January 2010), 37–48.

Shinde, S., Kothari, A. & Gupta, V. (2018), 'YOLO based Human Action Recognition and Localization', *Procedia Computer Science* **133**(2018), 831–838.
URL: *https://doi.org/10.1016/j.procs.2018.07.112*

Taketomi, T., Uchiyama, H. & Ikeda, S. (2017), 'Visual SLAM algorithms: a survey from 2010 to 2016', *IPSJ Transactions on Computer Vision and Applications* **9**(1), 16.
URL: *http://ipsjcva.springeropen.com/articles/10.1186/s41074-017-0027-2*

van der Meulen, H., Kun, A. L. & Shaer, O. (2017), 'What Are We Missing? Adding Eye- Tracking to the HoloLens to Improve Gaze Estimation Accuracy', *Proceedings of the Interactive Surfaces and Spaces - ISS '17* pp. 396–400.
URL: *http://dl.acm.org/citation.cfm?doid=3132272.3132278*

Vidanapathirana, M. (2018), 'Real-time Human Detection in Computer Vision — Part 2'.
URL: *https://medium.com/@madhawavidanapathirana/real-time-human-detection-in-computer-vision-part-2-c7eda27115c6*

Wang, Y. T., Feng, Y. C. & Hung, D. Y. (2011), 'Detection and tracking of moving objects in SLAM using vision sensors', *Conference Record - IEEE Instrumentation and Measurement Technology Conference* pp. 1078–1082.

Wästlund, E., Sponseller, K. & Pettersson, O. (2010), 'What you see is where you go', *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10* (January), 133.
URL: *http://portal.acm.org/citation.cfm?doid=1743666.1743699*

Zeller, M. (2018), 'HoloLens hardware details - Mixed Reality — Microsoft Docs'.
URL: *https://docs.microsoft.com/en-us/windows/mixed-reality/hololens-hardware-details*

Zolotas, M., Elsdon, J. & Demiris, Y. (2018), 'Head-Mounted Augmented Reality for Explainable Robotic Wheelchair Assistance'.

Appendices

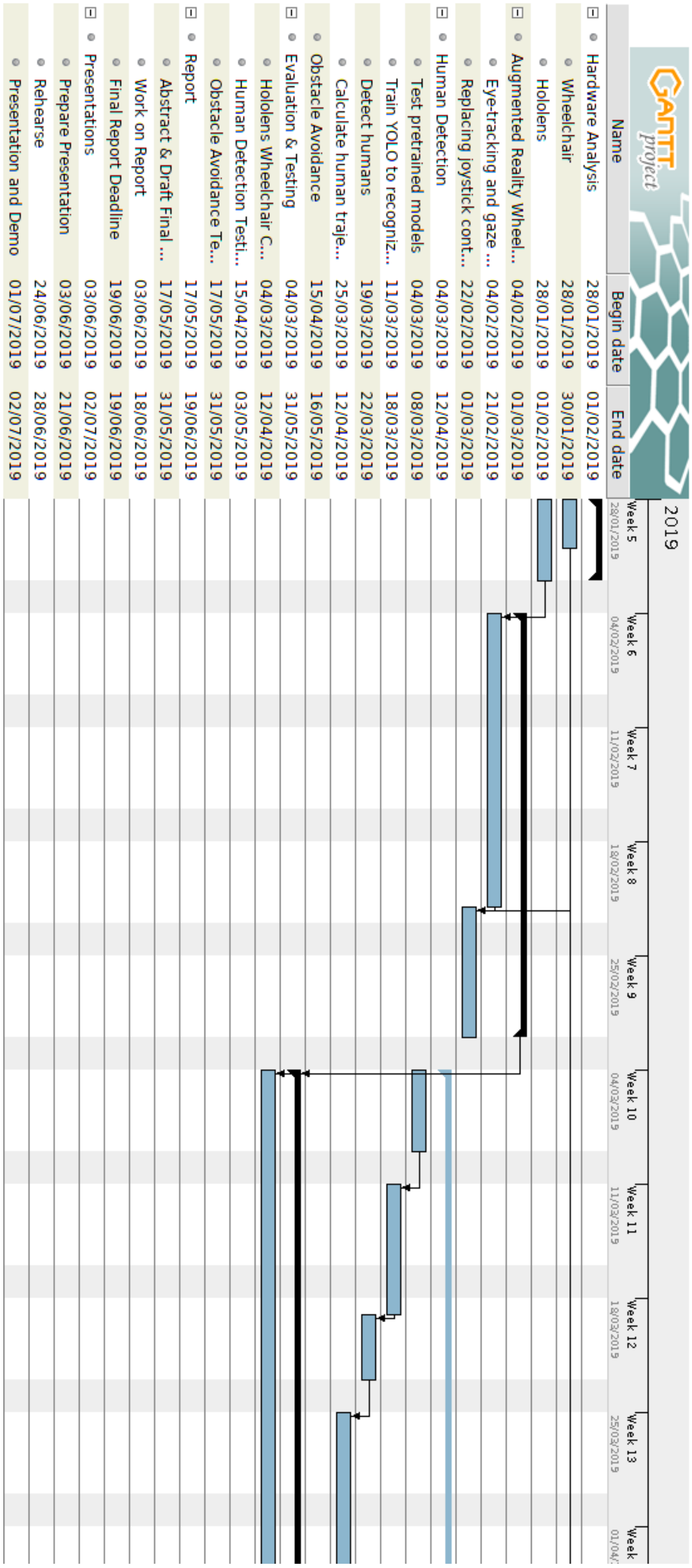

Figure 1: Project Timeline (Part 1)

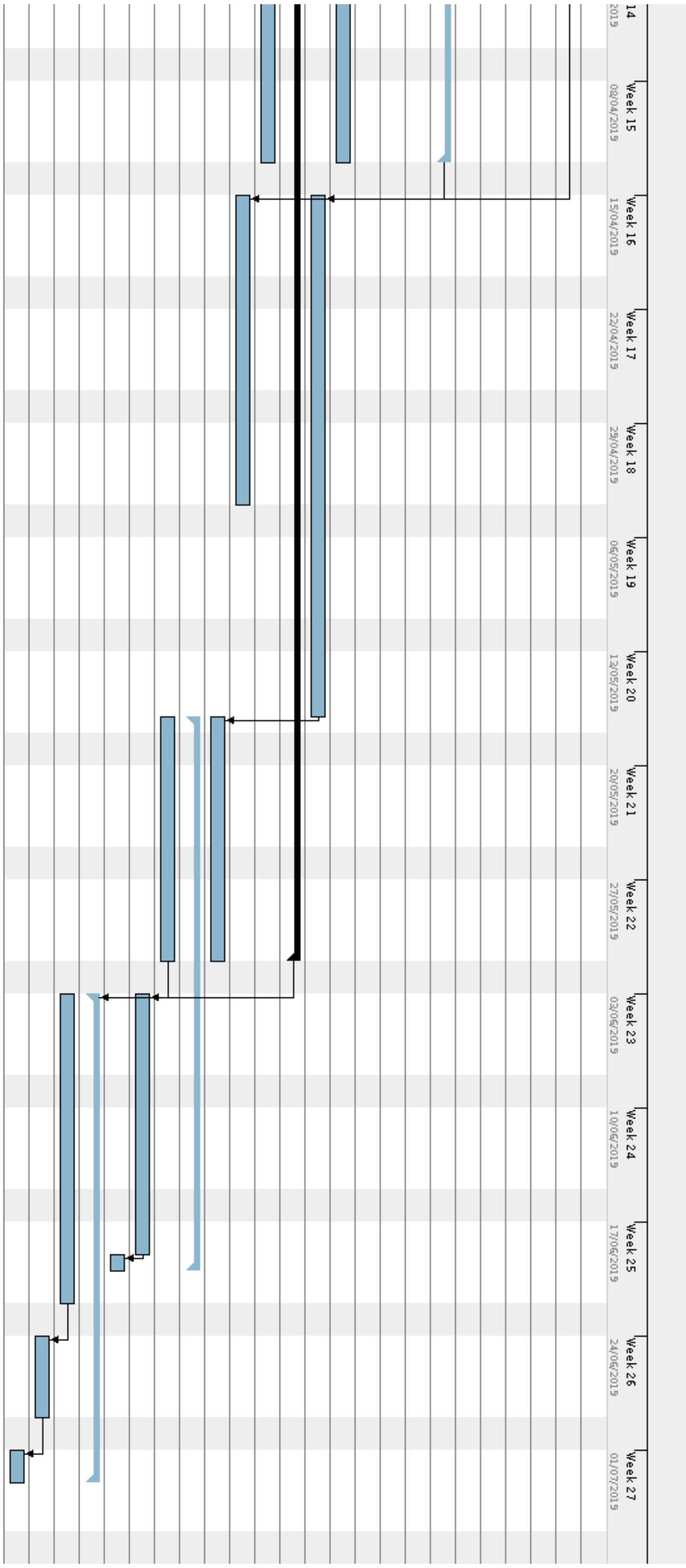

Figure 2: Project Timeline (Part 2)

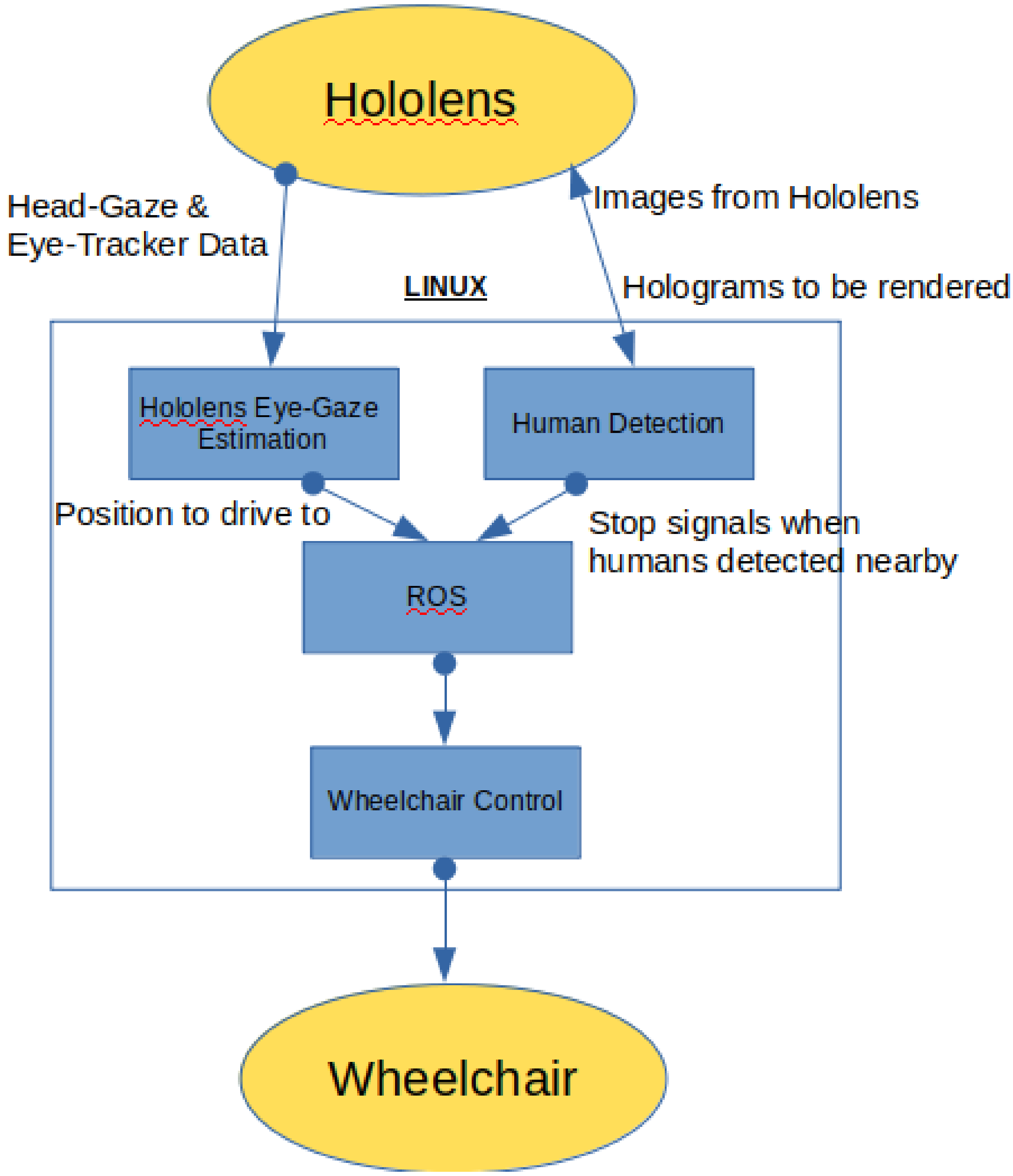

Figure 3: Overall System Architecture