# MACHINE LEARNING POWERED SMART VISUAL-SPATIAL AGNOSIA ASSISTIVE DEVICE

By

Adri Shankar Das
20321035
Ababil Hossain Fahad
20321034
Fardin Rahman Sahill
20321025
Mohammed Shakib
20321029

A Final Year Design Project (FYDP) submitted to the Department of Department of Electrical and Electronic Engineering in partial fulfillment of the requirements for the degree of Bachelor of Science in Electrical and Electronic Engineering (BSEEE)

Department of Electrical and Electronic Engineering
Brac University
October 2024

# MACHINE LEARNING POWERED SMART VISUAL-SPATIAL AGNOSIA ASSISTIVE DEVICE

By

Adri Shankar Das
20321035
Ababil Hossain Fahad
20321034
Fardin Rahman Sahill
20321025
Mohammed Shakib
20321029

A Final Year Design Project (FYDP) submitted to the Department of  Department of Electrical and Electronic Engineering in partial fulfillment of the requirements for the degree of Bachelor of Science in Electrical and Electronic Engineering (BSEEE)

Department of Electrical and Electronic Engineering
Brac University
October 2024

## Declaration

It is hereby declared that

1. The Final Year Design Project (FYDP) submitted is my/our own original work while completing degree at Brac University.
2. The Final Year Design Project (FYDP) does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The Final Year Design Project (FYDP) does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I/We have acknowledged all main sources of help.

**Student's Full Name & Signature:**

| | |
|---|---|
| **Adri Shankar Das** | **Ababil Hossain Fahad** |
| 20321035 | 20321034 |
| **Fardin Rahman Sahill** | **Mohammed Shakib** |
| 20321025 | 20321029 |

## Approval

The Final Year Design Project (FYDP) titled “Machine Learning powered smart Visual-Spatial Agnosia assistive device” submitted by

1. Adri Shankar Das (20321035)
2. Ababil Hossain Fahad (20321034)
3. Fardin Rahman Sahill (20321025)
4. Mohammed Shakib (20321029)

of Summer, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Science in Electrical and Electronic Engineering (BSEEE) on 24-10-224.

**Examining Committee:**

Academic Technical Committee (ATC): (Chair)

_______________________________
Abu Hamed M. Abdur Rahim, PhD
Professor, Department of EEE
BRAC University

Final Year Design Project Coordination Committee: (Chair)

_______________________________
Abu S.M. Mohsin, PhD
Associate Professor, Department of EEE
BRAC University

Department Chair:

_______________________________
Md. Mosaddequr Rahman, PhD
Professor and Chairperson, Department of EEE
BRAC University

## Ethics Statement

We have maintained all the ethical approaches for this project. We have checked the plagiarism by Turnitin and the percentage of plagiarism is only 8%.

## Abstract

Visual-spatial agnosia is a neurological visual impairment disorder that occurs after brain strokes in elderly people. These patients are unable to recognise objects surrounding them and also they cannot pursue the distance of the object. This creates a huge barrier between the real world and the world they perceive. This study proposes a smart eyewear assistive device that will help patients recognize the object as well as guide the patient toward the object by showing the distance of the object with the help of machine learning algorithms. A stereo vision algorithm is used to estimate the distance along with the Yolov11 algorithm that detects the objects. Physical data has been recorded to analyze the models' performance with the Yolov11 algorithm with stereo vision algorithm. This study also proposes future works in this area.

**Keywords:** Visual-spatial agnosia; neurological disorder; Yolov11; Stereo vision algorithm; Object detection

## Acknowledgement

We would like to express my deep appreciation to the ATC panel and professors. Their advice and suggestions were of immense importance which helped us in this work. We also value their helpful feedback and ideas throughout the whole process.

# Table of Contents

## List of Tables

## List of Figures

# Chapter 1 Introduction

## 1.1 Introduction

This report presents a machine learning-based smart assistive device to enhance navigation for visually impaired patients with spatial agnosia. We have embedded the electronic component into smart eyeglasses, enabling-by their very nature-adjustments to conform with specific users' unique needs. The proposed device is made up of several integrated subsystems, including microcontrollers, a stereo vision algorithm, and YOLOv11, which interact in concert to provide object detection and notification to users along their path. This document tries to present the development process in a step-by-step manner.

### 1.1.1 Problem Statement

Acute strokes come with severe problems. Although it is cured, it comes with some after-effects or disorders where visual problems are a great concern. A study shows that 1033 patients were observed who had no pre-existing eye problems, 48% of them suffered from visual problems after surviving strokes [1]. One of the visual problems is visual agnosia. Visual agnosia refers to problems in recognizing objects when it's only related to seeing them. These difficulties aren't because of issues with basic vision or other thinking abilities like language and memory or any eye-related problem [2]. It is a neurological disorder that affects a person's ability to perceive and understand spatial relationships and the organization of objects in their visual field. Visual-spatial agnosia is the problem when the patient shows difficulties in perceiving the distance of objects and also cannot identify them [3]. The condition usually results from brain damage, often to the parietal lobe, which plays an important role in processing spatial information.

1.1.2 Background Study

A case study reveals the behavior of B.L., a 51-year-old man, who survived a brain stroke, greatly suffered from visual-spatial agnosia although he had no eye problem at all [4]. Though it is a very concerning problem, no existing medical or dedicated solutions for this problem are found [5]. As a result, although they cure themselves of the stroke, they lead a miserable life.

Although there are several therapies to help the patients, they exist but they cannot help the patient's daily life. We conducted a survey with 15 doctors and 8 medical students to find more information about it as there is no extensive research paper found to gather more information. First, we gathered information about visual spatial agnosia. Then we started to conduct the survey. 86.7% of the doctors do not know about any technological assistive device for this problem. Furthermore, We described our project to the doctors and medical students and among them 80% doctors and 100% medical students said it can assist the patients. Therefore, we propose a smart glass/helmet that can assist people with real-time image processing and provide them proper sense of direction and objects via audio and visual devices. Thus the device reduces suffering with the assistance.

**Responses from the Doctors**

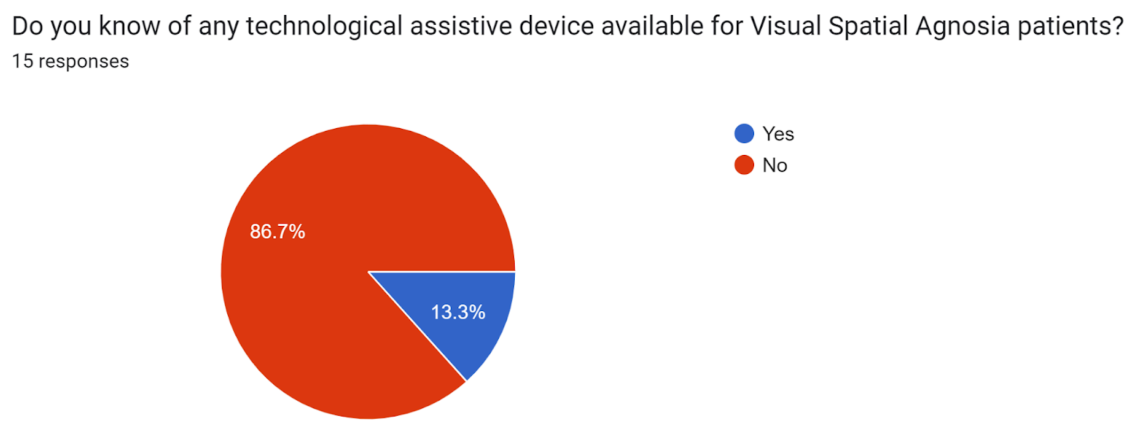

*Figure. 1: Responses from the Doctors*

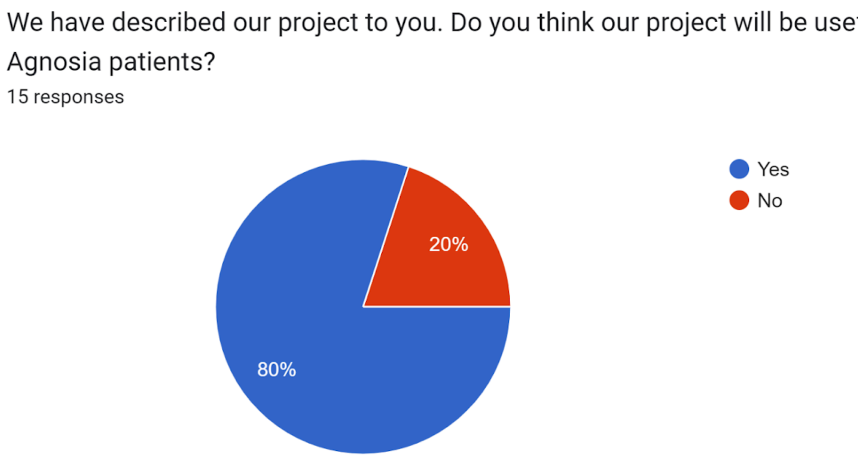


*Figure 2: Responses from the Doctors*

## Responses from the Medical Students

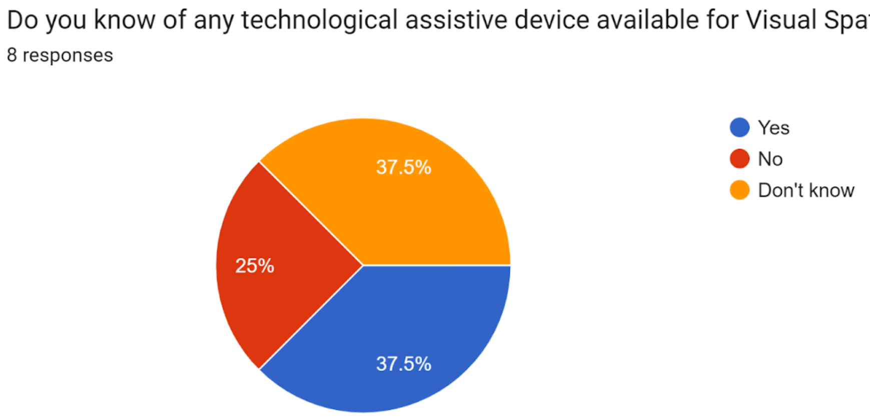


*Figure. 3: Responses from the Students*

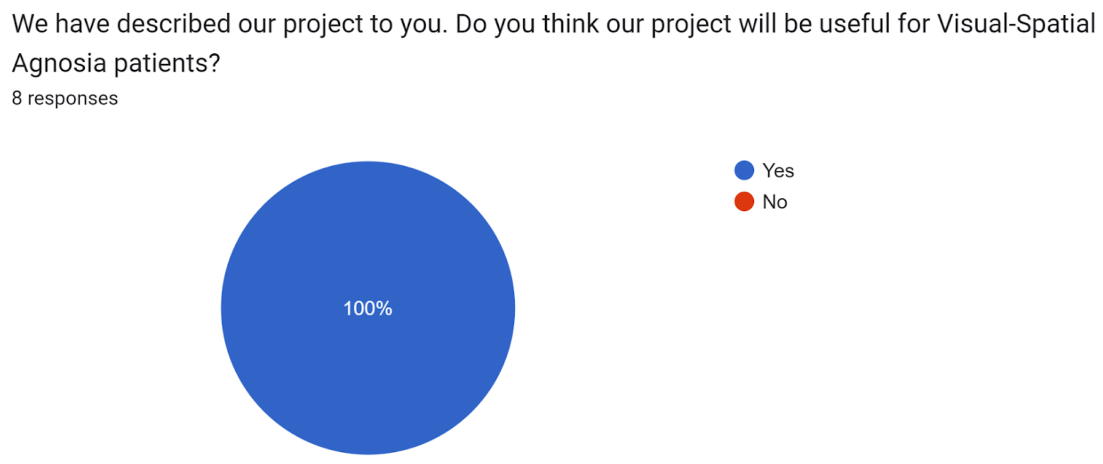


*Figure. 4: Responses from the Students*

1.1.3 Literature Gap

This technology is not practiced much in Bangladesh, nor has there been enough research on it. A study by Rushton shows that people with disabilities in South Asian countries like Nepal, India, and Bangladesh have very limited access to assistive technology. The number of assistive products available for these people is still small, and funding for this area is not good enough. Added to that, there is a big gap in the current research. This problem is mainly due to the lack of awareness about these issues and the few studies and development for the assistive device [6].

Research on vision impairments will not only assist eye doctors but also in the development of better devices that will help people in the community who have such topographical disorientation [7]. Therefore, enough resources should be availed to conduct research into these areas for advancement and improvement of society. The conclusion to this is that more research must be done in order to understand the daily challenges, concerns, and needs of a visually impaired individual. Special elaboration of solutions should be focused on providing them with safe and independent travelling.

## 1.2 Objectives, Requirements, Specification, and Constraints

### 1.2.1. Objectives

The project we aim to develop incorporates the following objectives:

- ➢ Design and Construct a smart Eyeglass for patient of Visual Spatial Agnosia.
- ➢ Monitoring measurements and distances & detection of real-time objects.
- ➢ Show the output through audio and visual devices.
- ➢ Implement the system such that it meets the users' requirements.
- ➢ Social awareness creation regarding person with spatial processing difficulties and the hardships they go through.

Functional and Nonfunctional Requirements:

The requirements are dictated based on the user needs, and these come in two forms: functional and non-functional. Functional requirements are meant to realize the goals of the project at hand, while the non-functional requirements aim at how the user would interact with the device and its general outlook.

Functional Requirements:

| **Requirement** | **Explanation** |
|---|---|
| A Frame | To create and develop a smart glass, the first thing we need is a frame for the glass. |
| Object Recognition and Identification | The device must have a system that can find objects. It should be able to recognize any item that comes in front of the user. |

| | |
|---|---|
| Real-time Feedback | This will instantly display to the user a view of the space information for conceptualization purposes. |
| Notification system | Whenever an object is detected by the device, there should be a alert system that will provide the distance and object name |
| Battery Monitoring and Power Management | As we use a suitable powerbank for power source, it'll display the battery charge on the screen and also user can use power efficiently to ensure it functions for extended periods. |
| Connection Features | The device should be capable of connecting with other devices, such as smartphones, bluetooth, earbuds, monitors, headphones, and many more, for updates, added features, or to function with navigation apps. |

*Table 1: Functional Requirements*

Non-functional Requirements

| **Requirement** | **Explanation** |
|---|---|
| Portability | This device should be lightweight, small, and wearable or carried in such a way that it makes it comfortable for everyday wear. |
| Usability | The device must be easy to use and learn. It should have a user-friendly platform that requires little training to |

| | |
|---|---|
| | perform its functions. The commands and feedback that it provides must be simple and clear. |
| Maintenance and Support | The device should be easily scale up the integral machine learning model or microcomputer, and it should be more user-friendly by allowing easy access for technical support in case of repairs. |

*Table 2: Non Functional Requirements*

### 1.2.2 Specifications

Technical Specification

| **Sub-System** | **Sub-system Specifications** | **Components** | **Component Specifications** |
|---|---|---|---|
| Eye Goggle | No. of Inputs - 2<br>No. of Outputs -2<br>Type of Data - Analog (Input) and Digital (Output) | 2 Pi Cameras | Voltage: 3.3V<br>Resolution: 3280 x 2464 pixel static images RGB |
| | | 2 Speakers | Receive Sensitivity: 108dB<br>Frequency Response: 20-20kHz |
| | | Micro-Display | Resolution: 640x400 RGB<br>Useable Display Area: 0.23" |

| | | | |
|---|---|---|---|
| | | | diagonal<br>Power Consumption: 0.12~195mW<br>Frame Rate: 60Hz<br>Operating Temperature: -20℃~70℃<br>Storage Temperature: -30℃~80℃ |
| | | Power Supply | Voltage: 5V, 5A |
| Processing Unit | No. of Input - 2<br>No. of Output - 2<br><br>Analog (Input) and Digital (Output) | Microcontroller | 5V/5A DC power via USB-C, with Power Delivery support |
| | | Cooling fan | Operating Voltage: 5V |

*Table 3: Technical specifications*

1.2.3 Technical and non-technical consideration and constraint in design process

| | |
|---|---|
| **Insufficient Fund** | As we are students we do not have sufficient fund to use high accuracy sensors and micro-computer |
| **Low Processing Power** | As we are building a prototype and we do not have high-specification hardware as a result our processing would not be as fast as we expected. |

| | |
|---|---|
| **Latency** | As we do not have fast processing hardware we will not be able to reduce the latency up to real-time. |
| **Low resources and literature availability** | The problem has no particular neurological solution. Also, there are not enough resources or research papers online. As a result, we have to rely on small resources and the survey that we conducted. |

*Table 4: Technical and non-technical consideration and constraint*

### 1.2.4 Applicable compliance, standards, and codes

| Code Number | Definition |
|---|---|
| IEEE SA - P1451.99 | This standard defines a method for data sharing, interoperability, and security of messages over a network, where sensors, actuators, and other devices can interoperate, regardless of underlying communication technology. |
| ISO/IEC 21989:2002 | Information technology — Telecommunications and information exchange between systems — Private Integrated Services Network — Specification, functional model and information flows — Short message service. |
| IEEE SA - P3109 | This standard defines a binary arithmetic and data format for machine learning-optimized domains. It also specifies the default handling of exceptions occurring in this arithmetic. |
| IEEE SA - P3123 | The standard defines specific terminology utilized in artificial intelligence and machine learning (AI/ML). The standard provides clear definitions for relevant terms in AI/ML. Furthermore, the standard defines requirements for data formats. |
| IEEE 11073 (X73) | A foundation for medical systems and devices that are interoperable is established by this family of standards. It facilitates smooth data interchange and communication across various medical devices, enhancing clinical processes and patient care. |
| IEEE 1471 | This standard provides rules and guidance on how to describe the architecture of technical or software-intensive systems, which is vital for medical software and devices used in biomedical engineering. |

| | |
|---|---|
| IEEE 1073 | It addresses medical device’s and information system’s communication within a healthcare and protective environment. The standard focuses on the interconnection of medical devices with patient’s monitoring, data storage, and information collection or manipulation systems. |
| IEEE 1599 | This standard gives guidance on the modeling and simulation for medical imaging systems, it helps engineers and researchers design and evaluate medical or biomedical imaging technologies. |

*Table 5: Applicable compliance, standards, and codes*

### 1.3 Systematic overview/summary of the proposed project

The proposed device will consist of sensors and display attached to different positions of the walking distance to detect the object, as well as bounding area of the object. The sensors (Stereo vision Camera) shows at the display, the distance of the object so that the patient can see and he/she will get a voice command too. The object will be detected by the Stereo vision Camera and the notification will be given via voice command by a blutooth. Here we use rasberrry pi 5 as micro-computer and use YOLOv11 algorithm for object detection and high accuracy. Other than this, we power up the pi5 using a suitable powerbank so that the device can be used for extended periods of time and does not require constant charging.

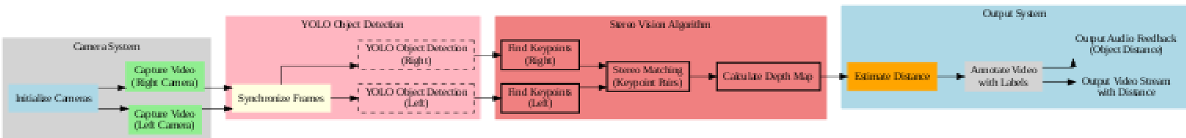


*Figure. 5: Overview of the proposed system*

### 1.4 Conclusion

The chapter was based on finding an engineering problem, which, with the acquired knowledge and skills, we handled. The next thing that did was some extended background research to find out what existed and just how effective in solving our problem. We did our research and found other available solutions and their shortcomings, found literature gaps, and areas that needed improvement. We furthered the relevance of our project with the prevailing and future scenario. We outlined in the end the objectives, functional and non-functional requirements, specifications, limitation, relevant compliance standards, and codes with regard to our project. This initial stage proved to be quite crucial as it enhanced our knowledge regarding the subject matter.

# Chapter 2 Project Design Approach

## 2.1 Introduction

Ideas, procedures, materials, and deliverables are planned out during the project design phase, which is the first stage of the project lifecycle. Because it provides a more comprehensive picture, a project design comes before a project plan, which contains more specific information. Following the determination of the problem, we attempted to ascertain the number of methods in which we may satisfy the problem's criteria. Here, we looked for ideas that would work best for our project.

## 2.2 Identify multiple design approach

**Design approach 1: Smart V-Glass Assistant**

Yolov11 algorithm to detect objects and stereo vision algorithm to measure distance.

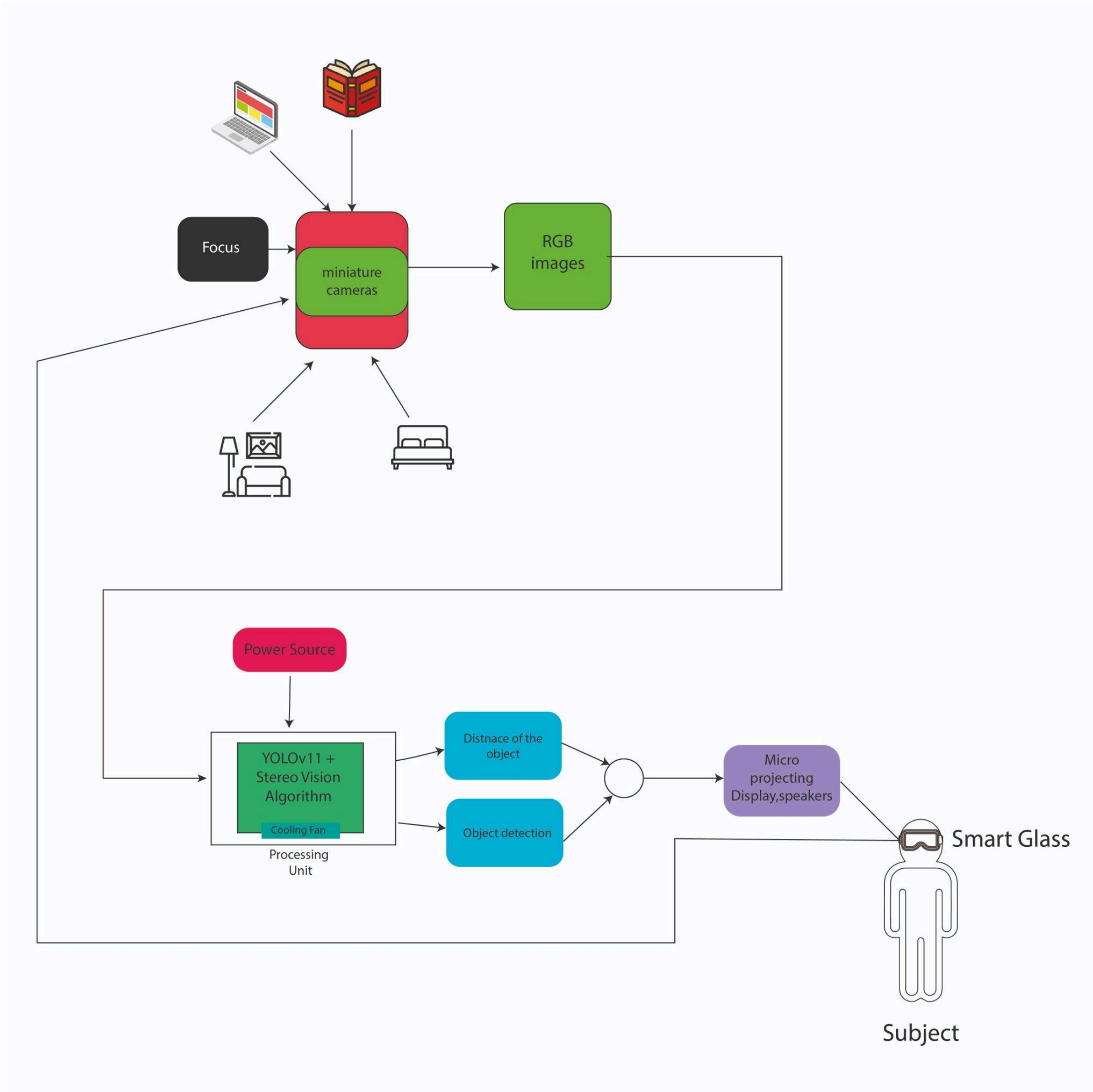


*Figure. 6: Design approach 1*

In this design approach we intend to use an eye glass that will have two input devices: 2 miniature rgb cameras; and a small projection display and a earbud set for the output. The processing unit of the system will be connected through wire externally. First, when the user focuses to an object, both of the cameras will capture images to measure the distance and to detect the object. There will be two rgb pi cameras to act as stereo vision cameras. The cameras will replicate the human eyes. right side camera will act as rignt eye and left side camera will act as the left eye. The miniature cameras will give a RGB image that can be processed to detect objects.

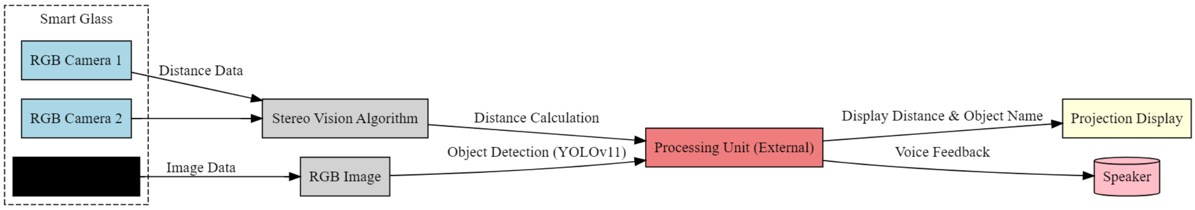


*Figure. 7: Overview of Design approach 1*

The monochrome data and rgb image will go to the processing unit. Here, the processing unit will run YOLOv11 to catagorize and detect the object. The both camera's data will be processed and the distance (in centimeters) of the object will be found. These two parameters will be then showed via the same glass and these two text will appear on the object of the glass. When we will show the image on the display, that time we will show only the right side cameras image. Also, a speaker will loudly read these two parameters. Thus, the subject will be able to recognize the object and will have an approximation about the distance of the object. As a result, the subject or the patient will be able to move freely as needed and lead a preferable life.

**Design approach 2: Smart V-Watch navigation assistant:**

Using the Mask R-CNN algorithm to detect the object and stereovision algorithm to measure the distance of the objects from the user.

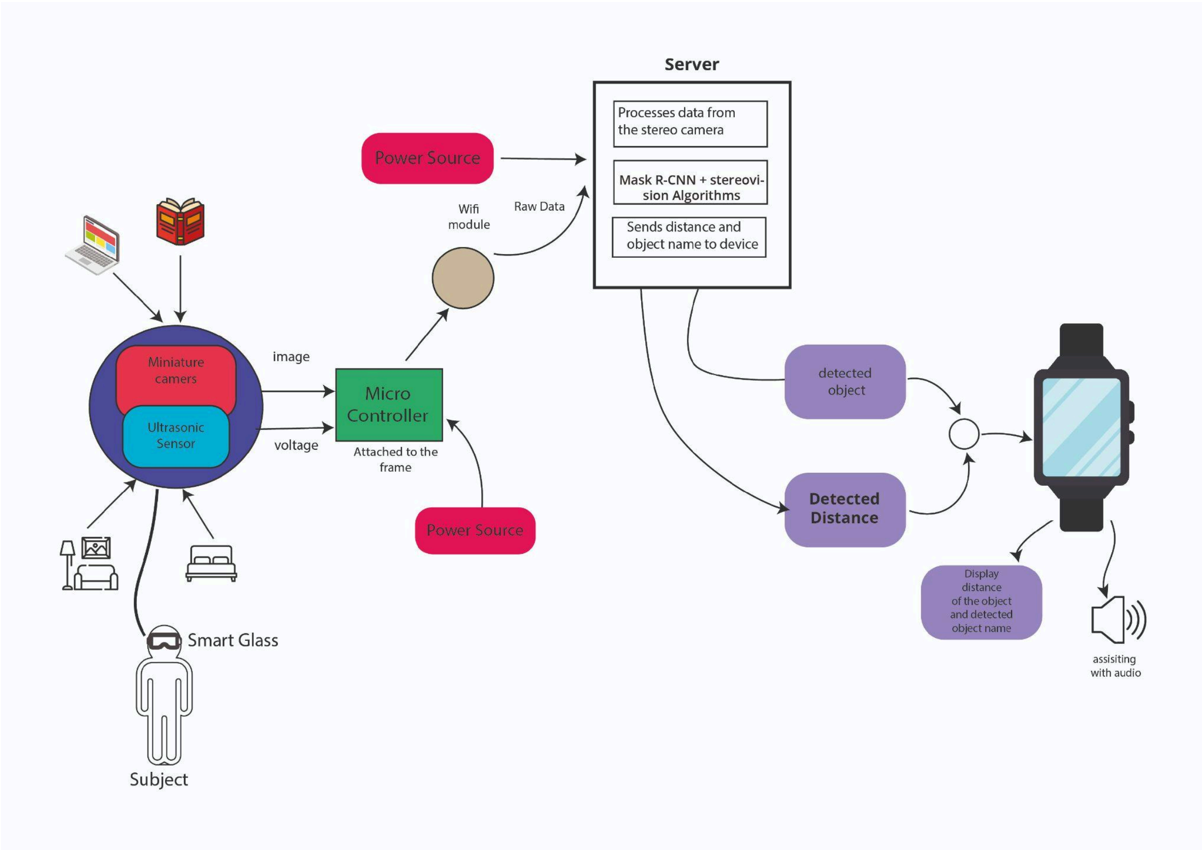


*Figure. 8: Design approach 2*

In this design approach, a  wearable glass and an app (smart watch/mobile) will operate as an input and output device and a server to process data. The wearable glass will include the input sensors and a microcontroller. The processing unit will be on cloud or a server computer. First, when the device is turned on, and when the user will aim the glass to a object, the miniature camera will capture the image of the object and the ultrasonic sensor will generate ultrasonic wave from the sender and then receive it and thus give a voltage output. The captured image and voltage will be sent to a server using a microcontrollers' wifi module. The surver will run the Mask R-CNN to detect the object. This surver will also calculate the distance of the focused object from the voltage data that the ultrasonic sensor gave.

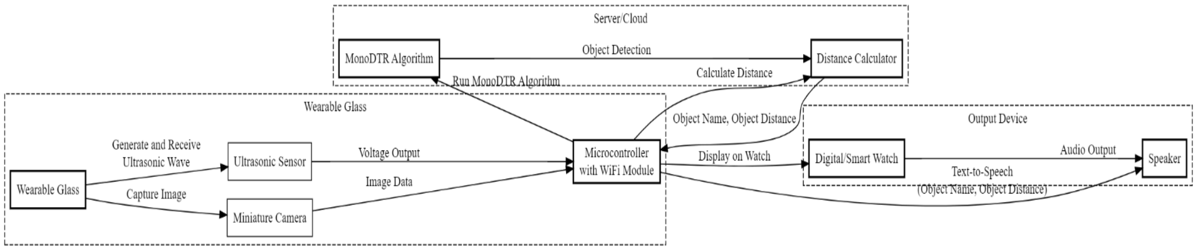


*Figure. 9: Overview of Design approach 2*

These two parameters, 1. Object name, 2. Object distance (in centimetre) will be then sent to the microcontroller to show the output. The microcontroller will send the output to a digital watch that can show the text of the detected object name and distance of the object. The digital or smart watch will also play a audio reading these two information so that the user or the subject can easily listen to it and do not need to repeatedly see the watch. Thus, the subject or the patient will be able to move freely without any interruption and have a relaxing fast moving life.

**2.3 Analysis of multiple design approach**

The two design methods were compared in this section using a variety of criteria, including cost, efficiency, usability, maintainability, and manufacturing feasibility. The evaluation of each of these strategies using these standards is displayed in the following table.

Based on this data table we will try to analyze which particular design approach is more beneficial for us and for future work which will give us a short idea to continue our research.

| Multiple Design Approach | Design Approach 1 | Design Approach 2 |
|---|---|---|
| Efficiency been fulfilling requirements | Object detection and distance measurement system | Object detection and distance measurement |
| Cost | Low | High |
| Usability | The users don't have to use their body part the system is in action | The users have to use their body part while the system is in action |
| Weight | Light | Heavy |
| Size | Compact | Compact |
| Manufacturability | Easy to mass produce | a very time consuming to mass produce |
| Maintainability | It is easily maintain and takes less time for troubleshoot | It is easily maintained and takes less time for troubleshoot |
| Upgradability | Easy to improvement new features | Difficult to import new features to the system because of the hardware complexibility |

*Table 6: Analysis of multiple design approach*

Here, a weighted matrix was used to perform a quantitative comparison. In order to determine which of these eight aspects were most crucial to users, we looked at research papers, articles, and polls that examined comparable goods. We then collectively gave each feature a weight and evaluated each design strategy according to how well it satisfied the requirements. In order to identify the optimal strategy, the ratings were finally totalled and compared.

| Feature | Approach-1 (10) | Approach-2 (10) |
|---|---|---|
| Accuracy | 8 | 6 |
| Dataset | 8 | 7 |
| Speed | 8 | 4 |
| Weight | 9 | 6 |
| Budget | 8 | 8 |
| Number of components | 7 | 8 |
| Power Required | 8 | 5 |
| Using time limit | 8 | 7 |
| Average | 8 | 6.38 |

*Table 7: Weighted average*

**2.4 Conclusion**

We listed the many strategies that may be used to accomplish the intended outcome in this section. Based on our study, two of these designs were chosen. We had to ensure during the selection process that the system's goals, criteria, and demands were fulfilled as well as that the specified limits were taken into consideration. To figure out the specifics of each concept and examine its potential, requirements, and limits, more study was done. To identify the best design strategy, both quantitative and qualitative studies were conducted in the third segment.

Based on user feedback, eight criteria were selected for the qualitative approach. Finally, we collectively assessed each of the 2 designs' performance on these specified criteria based on the study conducted in the preceding sections. This comparison was later represented in the weight matrix, which eventually assisted us in identifying the best course of action for our system.

# Chapter 3 Use of Modern Engineering and IT Tool

## 3.1 Introduction

Choosing the right tools for the creation and validation of the finished prototype was our next challenge after deciding on the best design strategy. Software and hardware tools were the two categories into which the tools were separated based on their intended purpose. Tasks like 3D modeling, code compilation, and functional verification required software tools for simulation, interface, and visualization. We initially made a list of every resource that was available before comparing them to identify the best ones in order to select the right program. We considered the advantages and disadvantages of utilizing each of the many software solutions during this procedure.

## 3.2 Select appropriate Engineering and IT tools

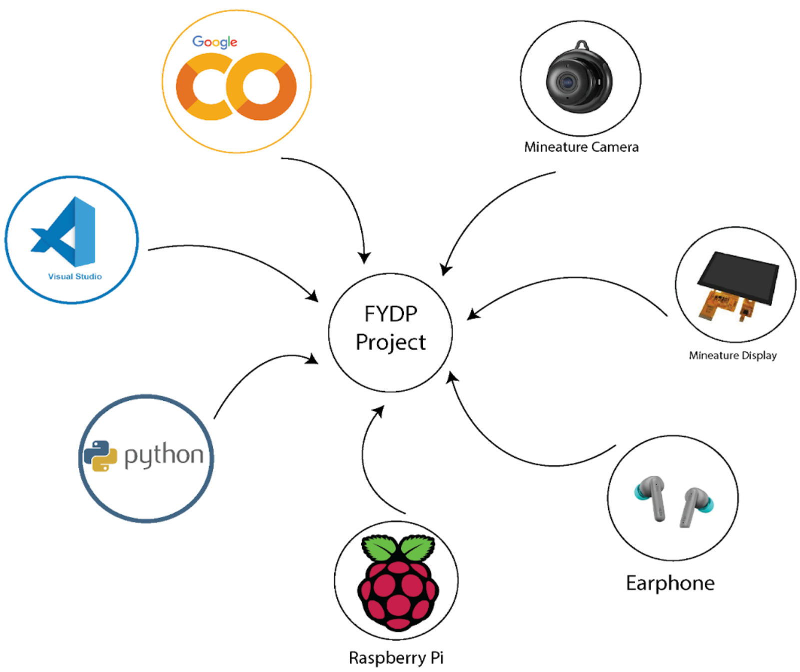


*Figure. 10: Overview of used Engineerig tools*

**Selection of Modern Engineering/IT Tools**

**1. Google Colab** is basically a web-based Python development environment that is used for training machine learning models. Free GPUs and TPUs are available in the environment to perform and accelerate the training processes of deep learning models.

**Benefits**

Powerful computers: Google Colab grants free access to powerful GPUs and TPUs that help speed up AI/ML tasks.

Pre-installed libraries: Most AI and machine learning libraries like TensorFlow, PyTorch, and OpenCV are already installed, saving your time.

**2. Raspberry Pi 5** is a low-cost, small, single-board, credit-card-sized computer that can be used to run applications, process data, and interface with other hardware components, such as sensors and cameras. In such a model, this might also serve as the central processing unit of the assistive device.

**Benefits**

Affordable: Projects with the Raspberry Pi 5 are relatively cheap to make and use.

Compact & light: Compact in size, thus applicable in wearable or portable devices.

Wide range of connectivity options: Littered with connectors-expose USB, HDMI, GPIO pins, and more-to which one could connect all sorts of other components-sensors, displays, and the like.

**3. Bluetooth,** this enables devices to talk with each other wirelessly. In the case of this project, it will be able to connect extra parts such as an external speaker, headsets, and even possibly a smartphone.

**Benefits**

Wireless connectivity: It enables the device to communicate with exterior parts without using cables for better comfort on the part of the users.

Low Power Consumption: Very suitable for battery-run devices in that it consumes less power compared to other wireless alternatives, like Wi-Fi.

**4. Miniature Display** would provide the user with small visual displays of feedback or messages relating to found objects, device status, or an alert of some type.

**Benefits**

Small, light; perfect for usage in portable devices since it negligibly adds weight or size.

Display Variations: The display has been designed to show various types of information based on the setting of the device.

**5. Python** programming language is mainly used in the development of machine learning algorithms, data processing, and controlling of devices.

**Benefits**

Python has plenty of libraries on AI, data processing, computer vision, and even more, including work with hardware. Among many, the key ones are TensorFlow, OpenCV, NumPy.

Easy to learn and use: Python's syntax is readable, which allows for rapid development and prototyping.

**6. Fusion 360** is a CAD, or Computer-Aided Design software. It is used in designing the mechanical parts of the assistive device, such as enclosure, mounts, and special mechanical parts.

**Benefits**

3D modeling allows the precise designing and customizing of the device components.

You can easily 3D print with the help of Fusion 360. Designs made in Fusion 360 can be saved for instant 3D printing.

**7. YOLOv11** is used in this real-time object detection for the project, helping find items surrounding the user to help in navigation and creating space awareness.

**Benefits**

Real-time Detection: The system detects an object that the user is looking at and provides information immediately.

One-stage detection: YOLO works effectively because, in one step, it detects and names the objects.

Flexibility: It can be trained to search for many different objects depending on user needs.

**8. Stereo vision algorithms** use two cameras to act like human eyes. This allows the device to see how far away things are. It helps the device know where objects are in space.

**Benefits**

Depth information gives valid measurements of distance to objects that could be used in navigation.

3D perception helps the device create a proper understanding of how objects look and where they remain; it helps in giving assistance based on space.

### 3.3 Use of modern engineering and IT tools

- Python: we have chosen Python as our primary programming language to develop and execute our algorithms. Python's popularity in the fields of data science and machine learning stems from its ease of use, extensive libraries, and strong community support. This makes it an ideal choice for our project, allowing us to leverage its powerful features and readily available resources.
- Google Colab for Simulations: To facilitate collaboration among our four-person team, we utilized Google Colab. This free, cloud-based platform enables us to write, execute, and share Python code seamlessly. Google Colab's real-time collaboration features ensure that team members can contribute and make adjustments simultaneously, streamlining our workflow and enhancing productivity.
- Raspberry Pi 5: We implemented a machine learning algorithm to achieve our desired outcomes, utilizing a Raspberry Pi microcontroller for its user-friendly interface and significant capabilities. The Raspberry Pi's versatility and ease of integration with various sensors and modules make it a suitable choice for our project, enabling us to handle the computational demands efficiently.
- Earphones and miniature display: we incorporated earphones, a miniature display, and a miniature camera. These tools were selected to create a compact and functional system, allowing us to interact with and test our machine learning models in a real-world environment. The earphones provide audio output, the miniature display offers visual feedback, and the camera captures input data, all of which are essential for the comprehensive testing and validation of our system.
- YOLOv11 algorithm: Most of the projects which dealt with real-time object detection have been using YOLOv11. The major reason for choosing the YOLOv11 algorithm is that, generally, the YOLO model has a great reputation in object detection tasks.

### 3.4 Conclusion

In this chapter, we learned how new engineering and IT tools can be used in order to design, develop, and test a solution. To be able to perform this, we needed to acquire all information that we needed about software and hardware which can give us what we want. First of all, we divided the jobs into two categories: software and hardware. For each, we tried to specify the tools we could use. As in other projects, we chose the best options by making comparisons between options. However, the varied tools this time necessitated doing the comparisons in different ways. We produced the comparisons ourselves through our research and based on what we decided was relevant. Comparisons were not simply different between the hardware and software groups, they were different for each subsystem as well. In the comparative study, we have also taken into consideration the shortcomings of the respective tools and tried to investigate how those can be set right. Later, with the chosen tools, we have developed our solution and assessed it. Also, for executing our tasks, we need to learn some techniques and hence we had to consult various resources, including manual, tutorials, documents, and online communities, for help.

# Chapter 4 Optimization of Multiple Design and Finding the Optimal Solution

## 4.1 Introduction

In this particular segment we are going to identify multiple solutions for our design approach and choose the optimal solution based on our observation and analysis. In order to find the optimum solution, we decided to Implement each design approach and then conduct a comparative analysis. Previously in Chapter 2 we have created a Table based on different features and their weight of both of the design approaches so in this segment we can analyze the table in detail.

## 4.2 Optimization of multiple design approach

**Design Approaches Analysing**

Approach-1: Utilizes YOLOv11 and a stereo vision algorithm.

Approach-2: Utilizes Mask R-CNN and a stereo vision algorithm for distance measurement and object detection with bounding boxes.

**Feature-by-Feature Analysis**

**1. Accuracy**

Approach-1: Score - 8

YOLOv11 is known for its high accuracy in object detection and segmentation. The score of 8 reflects that this approach performs well in identifying and delineating objects, which is crucial for users with visual-spatial agnosia.

Approach-2: Score - 6

Mask R-CNN, while fast and efficient, may not match the segmentation accuracy of YOLOv11. A score of 6 indicates that this approach is less precise, which could result in occasional misidentification or less detailed object recognition.

**2. Dataset**

Approach-1: Score - 8

The dataset used with YOLOv11 is likely comprehensive, contributing to the high accuracy and reliability. This indicates thorough training with diverse and well-annotated data.

Approach-2: Score - 7

The dataset for Mask R-CNN is also robust but might not be as exhaustive or detailed as that used with YOLOv11. This slight difference can affect overall performance, particularly in challenging scenarios.

**3. Speed**

Approach-1: Score - 8

YOLOv11, though accurate, is computationally intensive. However, achieving a speed score of 8 suggests optimized implementation, ensuring the device remains responsive and suitable for real-time use.

Approach-2: Score - 4

Despite Mask R-CNN reputation for speed, the score of 4 indicates potential bottlenecks in the system, perhaps due to integration issues with the stereo vision algorithm or hardware limitations.

**4. Weight**

Approach-1: Score - 9

A higher score for weight suggests that the hardware and overall design of this approach are lightweight, which is beneficial for portability and ease of use.

Approach-2: Score - 6

Approach-2 is comparatively heavier. This could be due to the hardware requirements of Mask R-CNN or less efficient design choices, impacting user comfort and device portability.

**5. Budget**

Approach-1: Score - 8

Both approaches have a score of 8 for budget, indicating that the cost-effectiveness of the designs is well-balanced. They are likely within the acceptable range for a final year project, providing good value for the features offered.

Approach-2: Score - 8

Similar to Approach-1, this score suggests that the design is cost-effective and within budget constraints, offering a reasonable balance of performance and cost.

**6. Number of Components**

Approach-1: Score - 7

A score of 7 indicates a slightly more complex design with potentially more components, which could mean more maintenance but might offer better modularity and potential for upgrades.

Approach-2: Score - 8

With a higher score, Approach-2 is simpler in terms of components, potentially reducing the complexity and improving reliability and ease of assembly and maintenance.

**7. Power Required**

Approach-1: Score - 8

This approach is efficient in power consumption, with a score of 8 suggesting good optimization, which is crucial for portable devices to ensure longer usage times between charges.

Approach-2: Score - 5

A lower score indicates higher power consumption, which could be due to the processing demands of Mask R-CNN. This affects battery life and the practicality of the device for extended use.

**8. Using Time Limit**

Approach-1: Score - 8

With a score of 8, this approach likely has a longer operational time, making it more practical for daily use by individuals with visual-spatial agnosia.

Approach-2: Score - 7

This score indicates a slightly shorter usage time, which might be a result of higher power consumption and less efficient power management. Based on the provided scores and we can say:

**Comparative Decision**

Approach-1 (YOLOv11 and stereo vision) has an overall higher average score of 8, indicating superior performance in accuracy, speed, weight, power efficiency, and overall user experience. It is well-suited for users needing precise and reliable assistance, though it might be slightly more complex and component-heavy.

Approach-2 (Mask R-CNN and stereo vision), with an average score of 6.38, performs adequately but falls short in critical areas like accuracy, speed, and power efficiency. It is simpler and might be easier to maintain but may not provide the same level of reliability and responsiveness as Approach 1.

Considering these factors, Approach-1 appears to be the more robust and user-friendly choice for a smart assistant device tailored for visual-spatial agnosia, providing better overall performance and a more satisfactory user experience.

### 4.3 Identify optimal design approach

After doing the comparative analysis to find our optimum solution we can say that our design approach is logically the optimum solution here. We will try to understand the first design approach which is smart V glass assistant in detail.

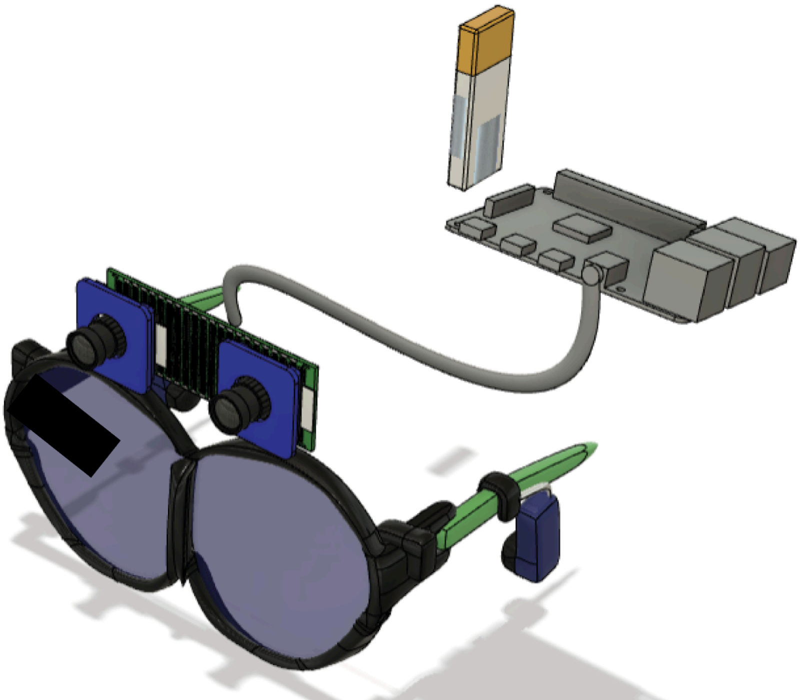

*Fig. 11: 3D design of our 1st design approach*

This device helps the patients supporting them to recognize objects and also the spatial relationship between the user and the object. Fig. 1(Chapter 2 first figure) is the design of the device. The eye frame has two RGB cameras as input devices, one mini-display attached to one lens, and two speakers near the two ears as output devices. There is a processing unit where the microcontroller and the battery units are situated. When the user focuses on an object, the two cameras take multiple photos and send them to the processing unit. The processing unit processes the data using YOLOv11 and stereo vision algorithms and sends it to the display and speakers. The user can watch the output as augmented reality pointers, such as distance markers, and highlighted objects with names. With the help of the speaker set, the user also gets auditory guidance about the objects and their distance.

**MaskRCNN Algorithm**

The Mask RCNN or Mask Region-Convolutional Neural Network is an object detection algorithm that extracts the feature maps of the input image (where CNN works as a backbone) & it extends the faster RCNN model which adds a branch to predict the segmentation mask in each ROI [8]. In order to detect any small, large, or medium objects Mask RCNN has the overall best performance. The main purpose of RCNN is to take an input picture and produce bounding boxes, where each bounding box contains an item and its category. The goal of Mask RCNN is to address the instance segmentation issue in computer vision and machine learning. It might be able to identify different items in a picture or video.

It receives an image containing one or more objects and returns an image with masks, bounding boxes, and classes. It mainly works based on 2 stages (which are linked in the background structure which are: (i) generating proposals about the regions of the image that we give as our input, (ii) Predicting the class of the object with bounding box and creating masks at pixel level of the object given in the input image based on the generated proposals in 1st stage [9].

**YOLOv11 Algorithm**

A cutting-edge object identification model, the YOLOv11 algorithm improves on the effectiveness and precision of its predecessors, including YOLOv8 and YOLOv10. YOLOv11, which is geared for real-time computer vision workloads, has improved training pipelines, efficient feature extraction, and an improved architecture. Both cloud-based and edge deployments can benefit from these enhancements as they increase mean Average Precision (mAP) scores while consuming less computing resources [10].

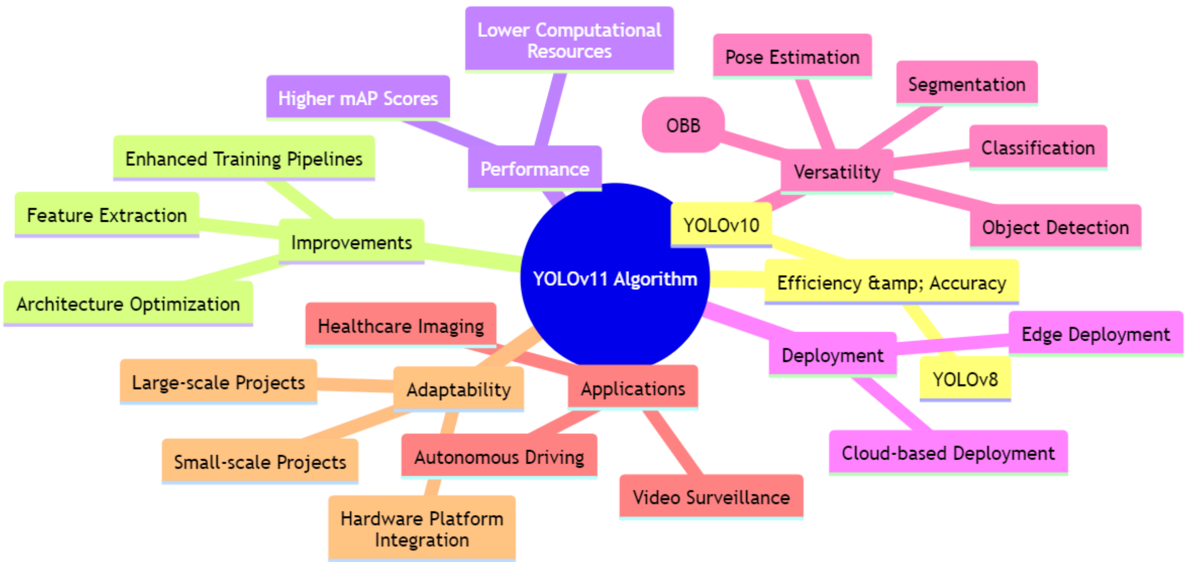


*Figure. 12: YOLOv11 algorithm*

A cutting-edge object identification model, the YOLOv11 algorithm improves on the effectiveness and precision of its predecessors, including YOLOv8 and YOLOv10. YOLOv11, which is geared for real-time computer vision workloads, has improved training pipelines, efficient feature extraction, and an improved architecture. These improvements use less computing resources and result in higher mean Average Precision (mAP) ratings, making it appropriate for both cuple based and edged development.

YOLOv11 is highly versatile, supporting various tasks like object detection, segmentation, classification, pose estimation, and oriented object detection (OBB). Its speed and precision make it ideal for demanding applications such as autonomous driving, healthcare imaging, and video surveillance. Additionally, the model is adaptable for large-scale or small-scale projects, ensuring smooth integration across different hardware platforms and environments.

**Stereo vision algorithm**

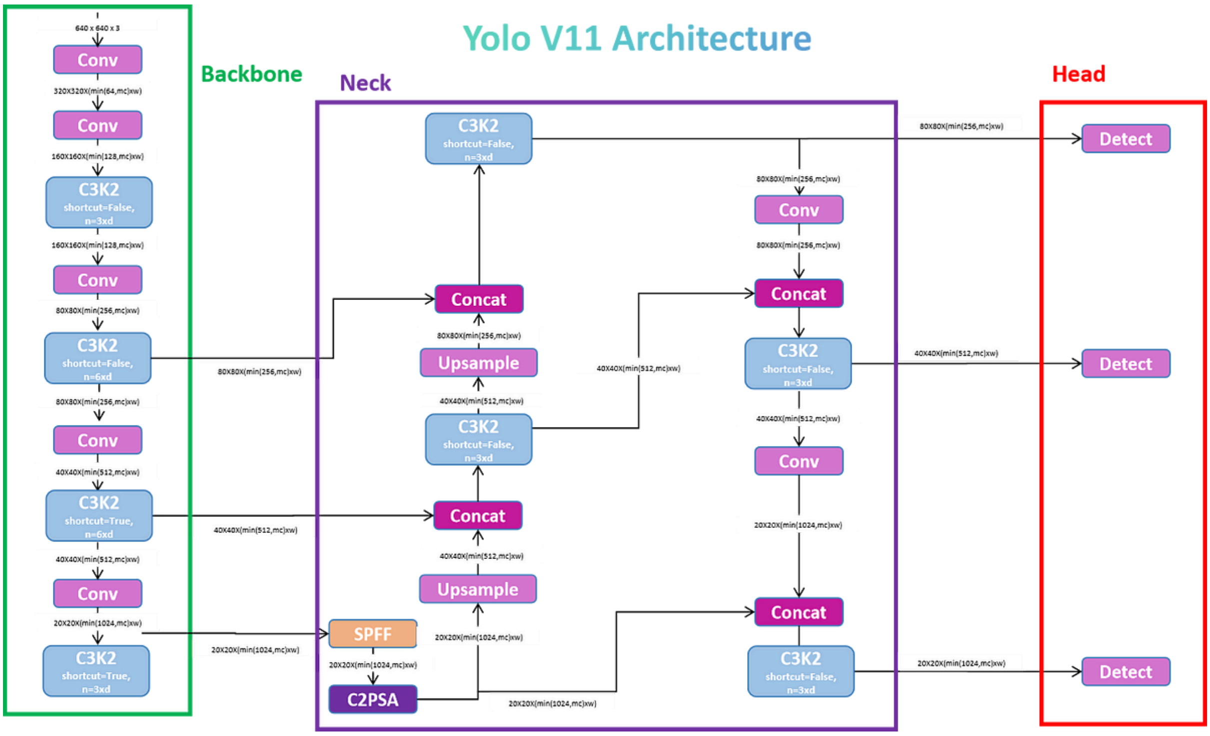


*Figure 13: Yolov11 Architecture*

Fig. 2 explaining the Geometric idea of stereo vision algorithm and the distance measurement operation of stereo vision system. Here *f* is the focal length of the camera, *d1* is the distance between the camera and the object, $d_3$ is object width, and *x* is camera's internal length [11].

So the total distance of the object with focal length is,

$$Distance = f + d_1 \quad (1)$$

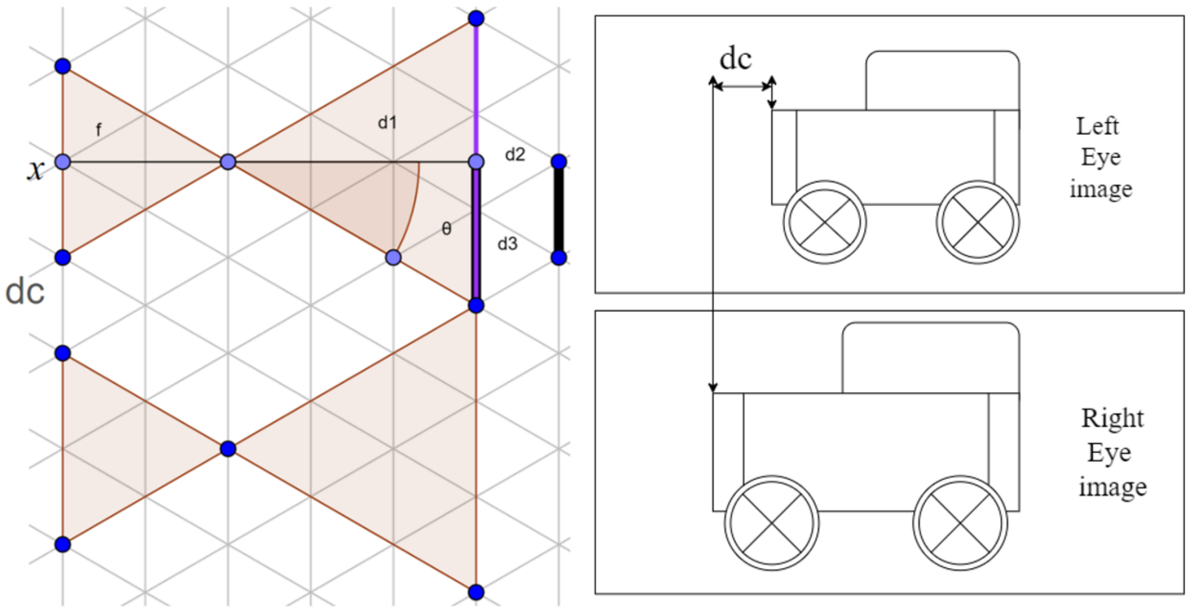


*Fig. 14: Geometric idea of stereo vision algorithm and Distance measurement operation*

Now in order to find *d1*, from the conventional geometric literature, it can be written as:

$$\frac{x}{f} = \frac{d_2}{d_1} \qquad (2)$$

$$d_1 = \frac{fd_2}{x} \qquad (3)$$

However, to apply the formula in (3), the conversion of units between $d_2$(in c.m.) and x(in pixels) is required, where $d_2$ is the distance of the total length of the camera that can capture capacity [11]. Because of this reason, it is necessary to take a different approach which is implemented by using the Pythagorean Triangle and trigonometric formulas,

$$Distance = f + d_2 = f + \frac{d_2}{2tan\theta} \qquad (4)$$

$$\frac{d_3}{d_2} = \frac{p}{N} \qquad (5)$$

$$Distance = f + \frac{Nd_3}{2p\ tan\boldsymbol{\theta}} \qquad (6)$$

P is pixel count, and N is total number of pixels.

Now. suppose in the $d_2$ capacity range, there is an object with a width $d_3$ of and the angle it produces with the focus axis is $\boldsymbol{\theta}$.

Now according to the equation, the pixels which are denoted as $p$. Now the question arises how the distance of the object can be measured in terms of pixels. The answer is that here are two cameras to compare the distance with pixel count [11]. Here are another camera and let the 2 cameras be placed in a way which is across. Now according to Fig. 2 if two similar points of the object are selected in the picture taken from the two cameras and then count the pixels where it can analyze the difference between those two point's pixel count, which definitely will be similar to the difference between the distance of two cameras.

Therefore, from this idea if the distance of the object from the right camera is a1 and similarly from the right camera is $a_2$ where, $p_1$and $p_2$ are the pixel count of both cameras respectively then the corresponding equations are:

$$a_1 = f + \frac{Nd_c}{2p_1\ tan\boldsymbol{\theta}} \qquad (7) \qquad\qquad a_2 = f + \frac{Nd_c}{2p_2\ tan\boldsymbol{\theta}} \qquad (8)$$

In this scenario, here are two equations and two unknowns. So, if they are solved, the value of the unknown variables will be found. The challenge of this whole algorithm is to match the distance of the two cameras and the pixel count of dc .

**YOLOv11 with stereovision Algorithm**

To achieve the distance with the object detection, the calculation of the bounding box's coordinates is needed. Using the YOLOv11, the coordinates of the bounding boxes are found and then these coordinates are then fed to the stereo vision algorithms. From the mentioned equations unknowns a1 & a2 is needed to be calculated. The stereo vision algorithm calculates the distance of the object using the mentioned equations and then a cost matrix is calculated to reduce the error. The main errors to be reduced are the objects between two images, which minimizes the loss in the vertical distances, horizontal distances, and area differences. Finally, using the already calibrated parameters, which are tan and P the distance of the objects are calculated and shown beside the identification of the object.

**Dataset description**

For the development of the smart eyewear device, the COCO dataset is utilized. The dataset has day-to-day data on a large scale. The dataset has a wide range of object detection, captioning, and segmentation features well-known in the field of computer vision for their variety and intricacy. The COCO dataset consists of over 330,000 images, of which more than 200,000 are labeled, and contains 80 object categories with 1.5 million object instances. This dataset is designed to reflect real-world scenes that contain multiple objects in many images and complex backgrounds that enable users to test their projects precisely [12].

**Key features of the COCO dataset**

Large Scale: The dataset contains a large number of labeled images and object instances, ensuring a thorough and exhaustive training and testing dataset for machine learning models.

Complex scenes: The images often contain multiple objects in the dataset to test the real-world scenarios with the ML models. The presence of different levels of obstruction and complex backdrops imitates real-life situations and tests the effectiveness of detection and segmentation algorithms.

Diverse categories: The dataset has 80 object categories that encompass a diverse array of typical things encountered in everyday life, hence bolstering the capacity of models trained on it to generalize.

Annotations: The COCO dataset provides comprehensive annotations, such as bounding boxes, key points for human poses, picture descriptions, and object segmentation masks, which greatly assist in performing diverse computer vision tasks.

This study specifically employed the object recognition and segmentation annotations obtained from the COCO dataset. The annotations play a crucial role in the training of our MaskRCNN and stereo vision models, enabling them to reliably distinguish items and distances in the project.

**4.4 Conclusion**

In conclusion, integrating YOLOv11 with a stereovision algorithm offers significant advantages for object detection and distance measurement. YOLOv11's enhanced architecture ensures precise and real-time detection across a variety of tasks, while stereovision adds depth perception, enabling accurate distance estimation between the camera and objects. This fusion improves not only the spatial understanding of the environment but also enhances performance in critical applications like autonomous vehicles, robotics, and surveillance systems. With the combined strengths of YOLOv11's speed and accuracy and stereovision's depth capabilities, this approach outperforms traditional methods by providing more reliable and context-aware results in complex scenarios.

# Chapter 5 Completion of Final Design and Validation

## 5.1 Introduction

Now that we have verified our design, the next task for us would be to implement the design. So after gathering all the knowledge we have implemented the entire embedded system. A most of the components were not available in our country's market so we have to order them from abroad as well and after gathering all of the components and lots of trial and error we were successfully able to validate our desired outcome.

## 5.2 Completion of final design

The final design of the smart eyeglass integrates the YOLOv11 algorithm with stereovision technology to enable real-time object detection and precise distance measurement. The system uses two synchronized cameras to simulate binocular vision, providing depth perception similar to human eyes. The left and right camera images are processed to detect objects and extract depth information. YOLOv11 identifies objects with high precision, while the stereovision system calculates the disparity between the two views to estimate the distance to detected objects.

### 5.2.1 Final Hardware

**The eye glass section:**

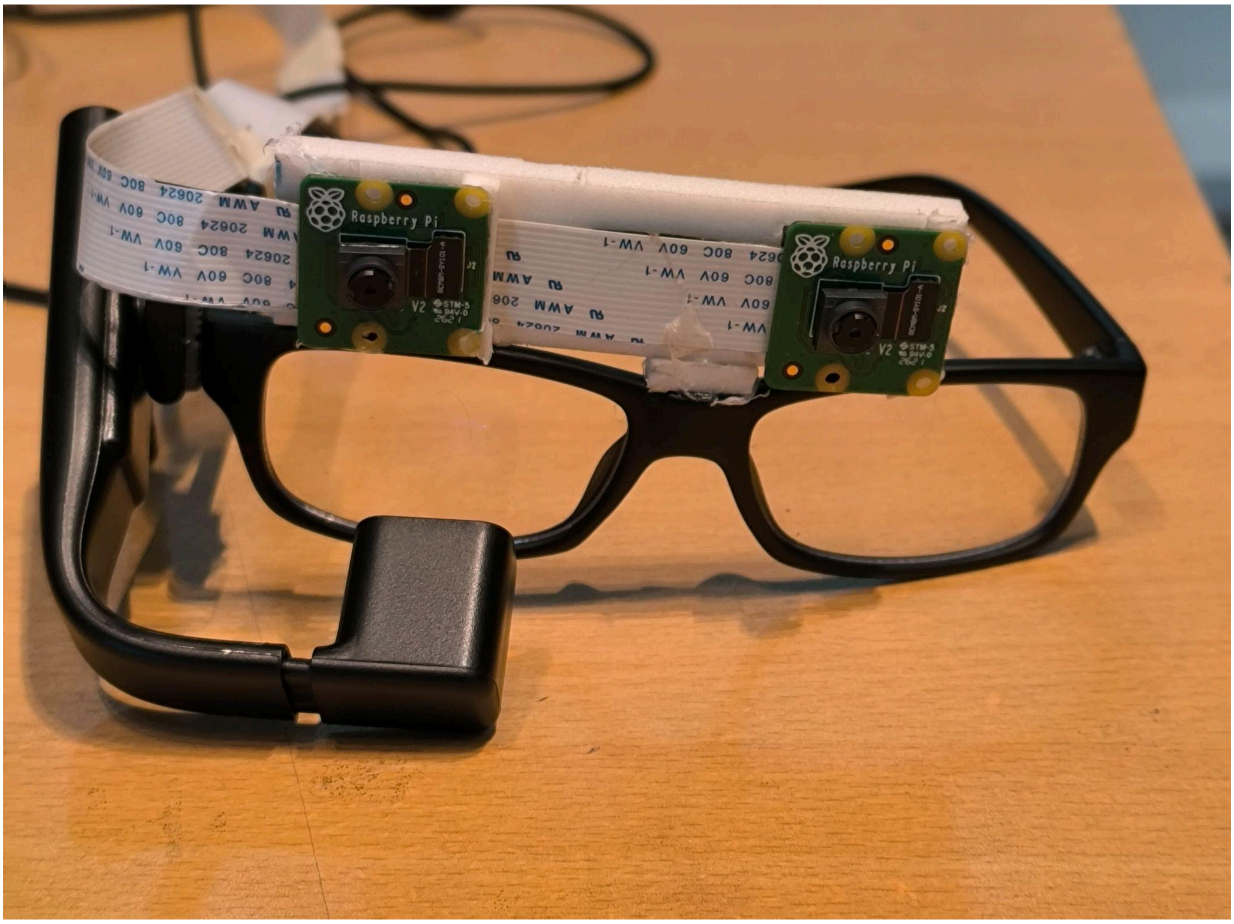


*Fig. 15: The eye glass section*

**The processing Section:**


*Fig. 16: The processing Section*

**The whole hardware:**

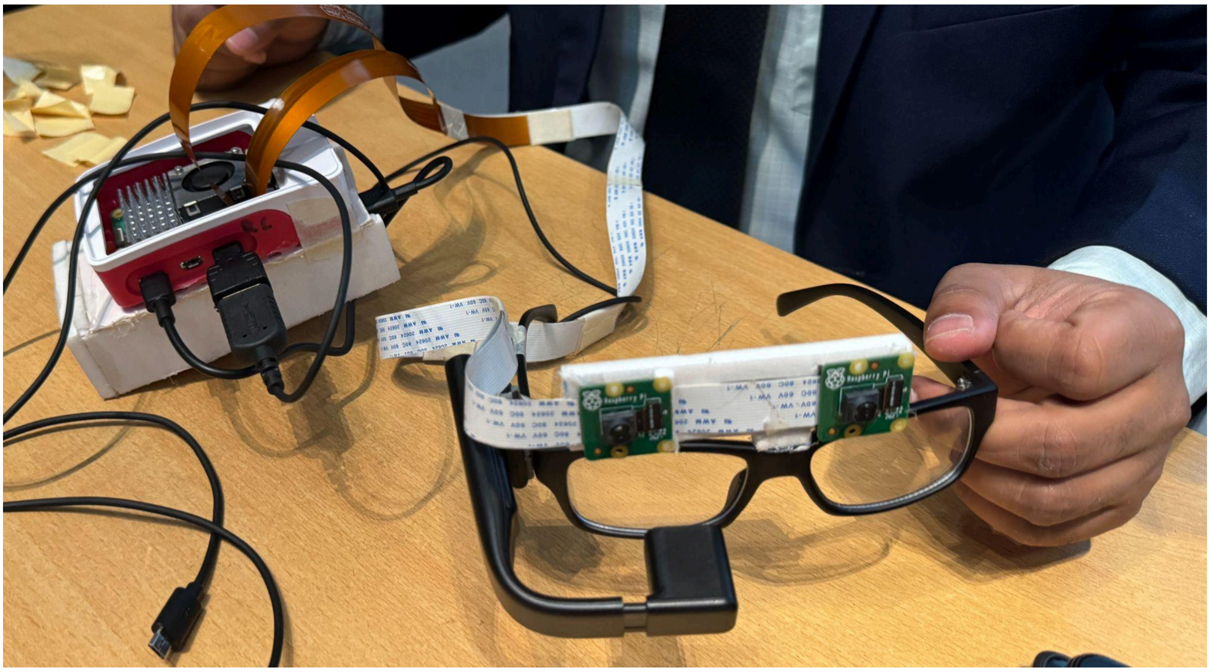

*Fig. 17: The whole hardware section*

### 5.3 Evaluate the solution to meet desired need

This design ensures portability and efficiency, suitable for users requiring assistive technology, such as the visually impaired, or for applications like augmented reality and navigation. The lightweight nature of the system allows it to run on low-powered hardware, making it practical for edge computing environments.

The device processes input images continuously to identify objects and measure their distances, providing real-time feedback. The synergy between YOLOv11's high detection accuracy and stereovision's depth estimation ensures that the user receives timely and reliable information about their surroundings, enhancing both safety and situational awareness.

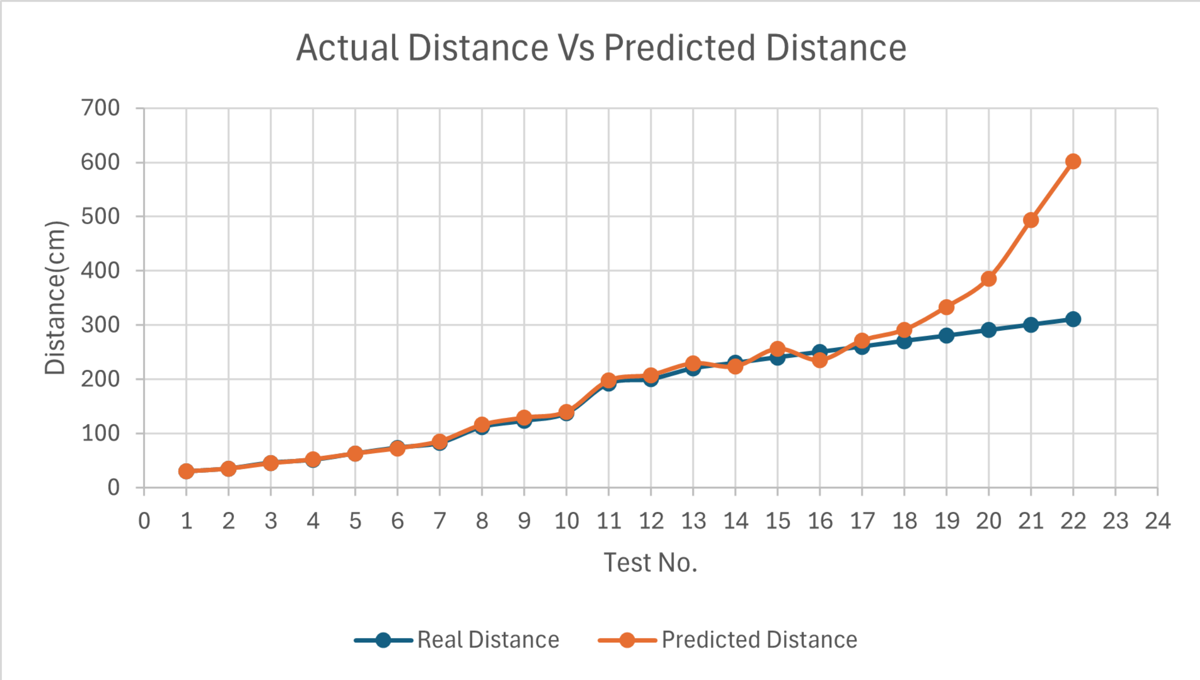


*Fig. 18: Actual vs Predicted distance with design approach 1*

figure: Actual vs Predicted (YOLOv11)

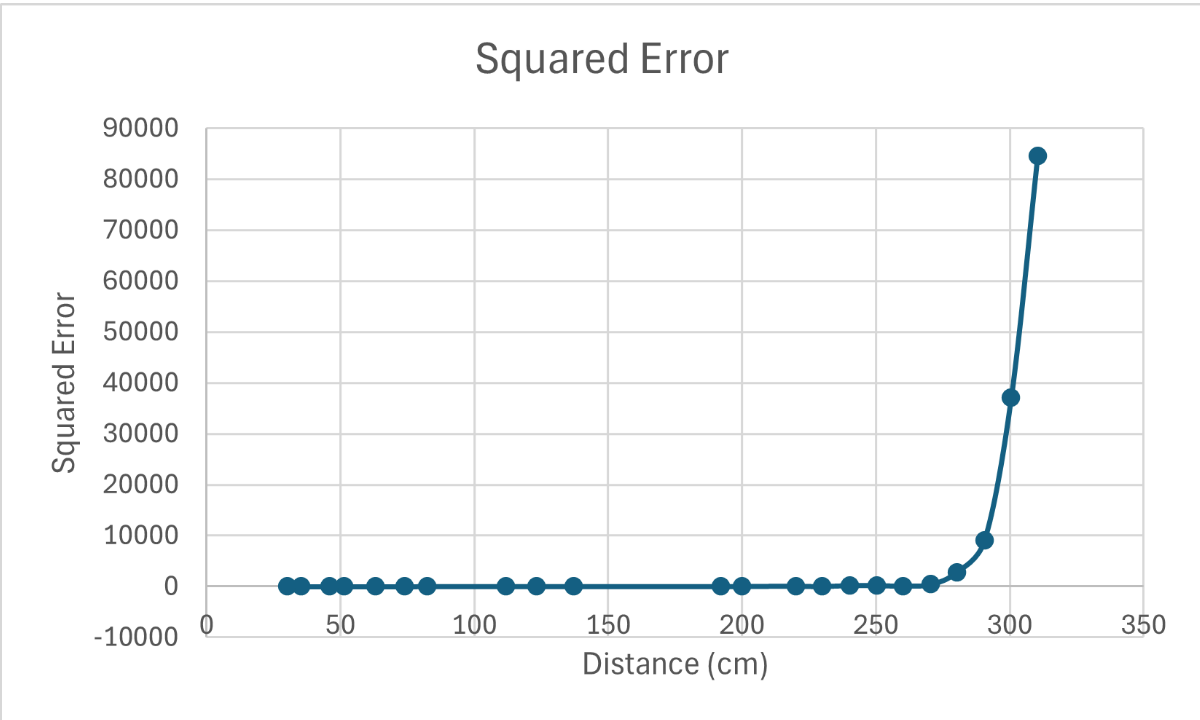


*Fig. 19: squared error for design approach 1*

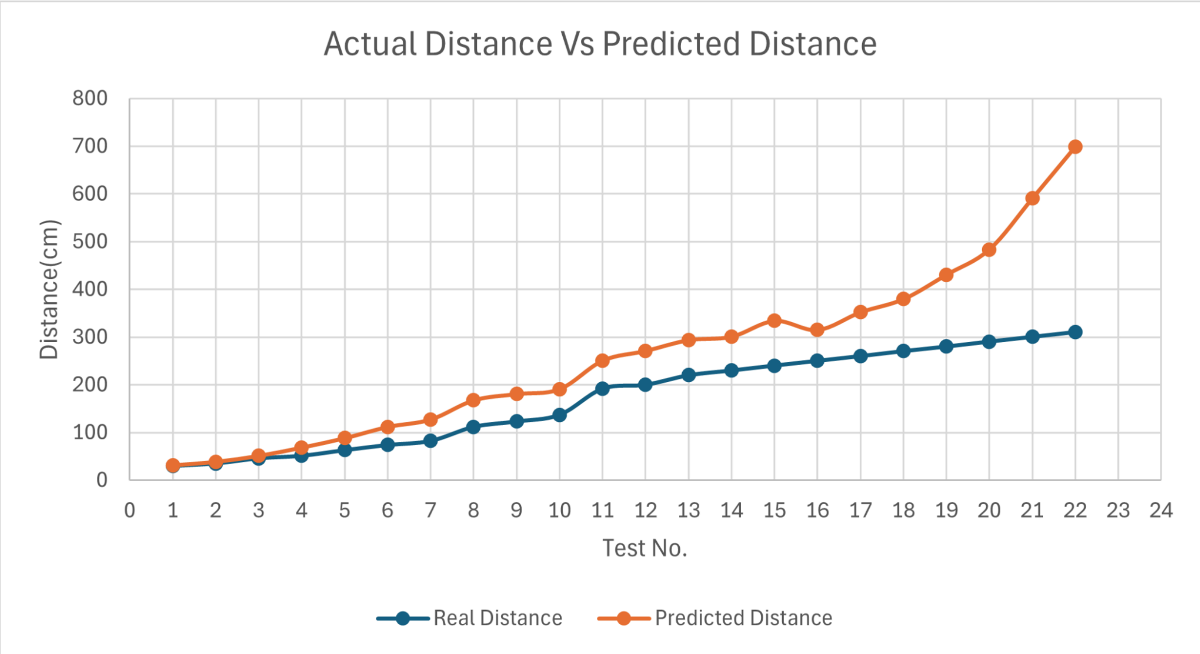


*Fig. 20: Actual vs Predicted distance with design approach 2*

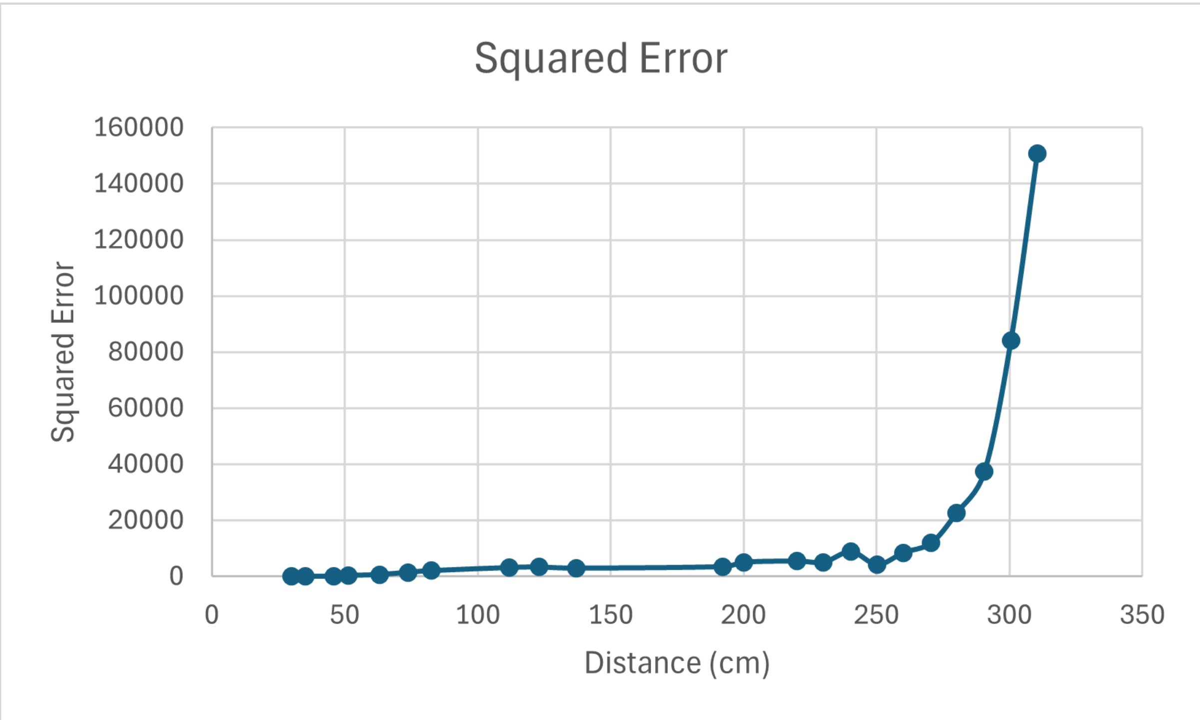


*Fig. 21: squared error for design approach 2*

**Statistical Analysis of the Data**

The graph labelled "Actual Distance vs Estimated Distance" shows a comparative analysis of real distances (in blue) versus the predicted distances by the system (in red). A total of 22 tests were conducted. The actual distances are measured by measurement tape and the collected data is from the stereo vision algorithm or from the device output. Initially, test 1 is performed at a smaller distance. Gradually the distance of the object is increased.

1. Trend Analysis:

Both the actual and predicted distances increase along the x-axis, indicating that the system can measure distances effectively. Initially, the predicted distances closely follow the real distances, suggesting good accuracy for objects closer to the camera. By analysing the figures 18,19 respectively Where we have used YOLOv11 with stereo vision algorithm with 22 datasets we can clearly see that for a higher index the predicted the value and the calculated distance values are perfectly match but absorted point it gets distorted which is between 250 to 300 cm. Then in case of Fig 20,21 respectively The data points does not match from 50 to 100 cm After that we can see that predicted value and estimated values are different from the very beginning and the letter on it gets destroded even more. This particular analysis rules that YOLOv11 is much better algorithm then Mask RCNN.

2. Error Growth:

As the distance increases beyond approximately 15 units on the x-axis, the predicted values tend to deviate from the real distances, showing a slight overestimation. This could indicate that the depth estimation algorithm becomes less precise at longer ranges, a typical challenge in stereovision systems.

3. Performance Metrics:

For closer objects (up to 15 units), the error between predicted and actual distances is minimal, ensuring reliability for short-range applications. However, for distances beyond 20 units, the prediction error increases noticeably. This suggests that the calibration or disparity calculation might require fine-tuning for improved performance at longer ranges.

| **Real Distance** $y_i$ | **Predicted Distance** $\hat{y}_i$ | **Error** $e_i = y_i - \hat{y}_i$ | **Squared Error** $e_i^2$ |
|---|---|---|---|
| 30 | 30.01 | -0.01 | 0.0001 |
| 35.3 | 35.1 | 0.2 | 0.04 |
| 45.98 | 44.92 | 1.06 | 1.1236 |
| 51.33 | 52.13 | -0.8 | 0.64 |
| 63.1 | 63 | 0.1 | 0.01 |
| 73.85 | 72.62 | 1.23 | 1.5129 |
| 82.5 | 85.51 | -3.01 | 9.0601 |
| 111.87 | 115.89 | -4.02 | 16.1604 |
| 123.2 | 128.9 | -5.7 | 32.49 |
| 137.18 | 139.95 | -2.77 | 7.6729 |
| 192 | 197.45 | -5.45 | 29.7025 |
| 200 | 207.77 | -7.77 | 60.3729 |
| 220 | 229.16 | -9.16 | 83.9056 |
| 230 | 223.54 | 6.46 | 41.7316 |
| 240.3 | 255.88 | -15.58 | 242.7364 |
| 250.2 | 235.13 | 15.07 | 227.1049 |
| 260.1 | 271.19 | -11.09 | 122.9881 |
| 270.5 | 291.04 | -20.54 | 421.8916 |
| 280.1 | 333.27 | -53.17 | 2827.0489 |
| 290.5 | 385.76 | -95.26 | 9074.4676 |
| 300.5 | 493.02 | -192.52 | 37063.9504 |
| 310.4 | 601.23 | -290.83 | 84582.0889 |

*Table 8: Real vs predicted distances and errors for design approach 1*

The mean squared error (MSE) is 16.692. Which is performed by using the equation,

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 \quad (9)$$

$n = 22$

The error is quite smaller. we can observe the trend that if we increase the distance the performance of the integrated stereo vision algorithm is decreased. Model near to good:

| **Real Distance** $y_i$ | **Predicted Distance** $\hat{y}_i$ | **Error** $e_i = y_i - \hat{y}_i$ | **Squared Error** $e_i^2$ |
|---|---|---|---|
| 30 | 31.66 | -1.66 | 2.7556 |
| 35.3 | 38.95 | -3.65 | 13.3225 |
| 45.98 | 51.41 | -5.43 | 29.4849 |
| 51.33 | 68.75 | -17.42 | 303.4564 |
| 63.1 | 88.45 | -25.35 | 642.6225 |
| 73.85 | 111.97 | -38.12 | 1453.1344 |
| 82.5 | 127.6 | -45.1 | 2034.01 |
| 111.87 | 167.67 | -55.8 | 3113.64 |
| 123.2 | 180.73 | -57.53 | 3309.7009 |
| 137.18 | 191.04 | -53.86 | 2900.8996 |
| 192 | 250.75 | -58.75 | 3451.5625 |
| 200 | 271.02 | -71.02 | 5043.8404 |
| 220 | 293.66 | -73.66 | 5425.7956 |
| 230 | 300.85 | -70.85 | 5019.7225 |
| 240.3 | 334.02 | -93.72 | 8783.4384 |
| 250.2 | 314.9 | -64.7 | 4186.09 |
| 260.1 | 351.95 | -91.85 | 8436.4225 |
| 270.5 | 379.51 | -109.01 | 11883.1801 |
| 280.1 | 430.62 | -150.52 | 22656.2704 |
| 290.5 | 483.59 | -193.09 | 37283.7481 |
| 300.5 | 590.58 | -290.08 | 84146.4064 |
| 310.4 | 698.66 | -388.26 | 150745.8276 |

*Table 9: Real vs predicted distances and errors for design approach 1*

The mean squared error (MSE) is 27.30.

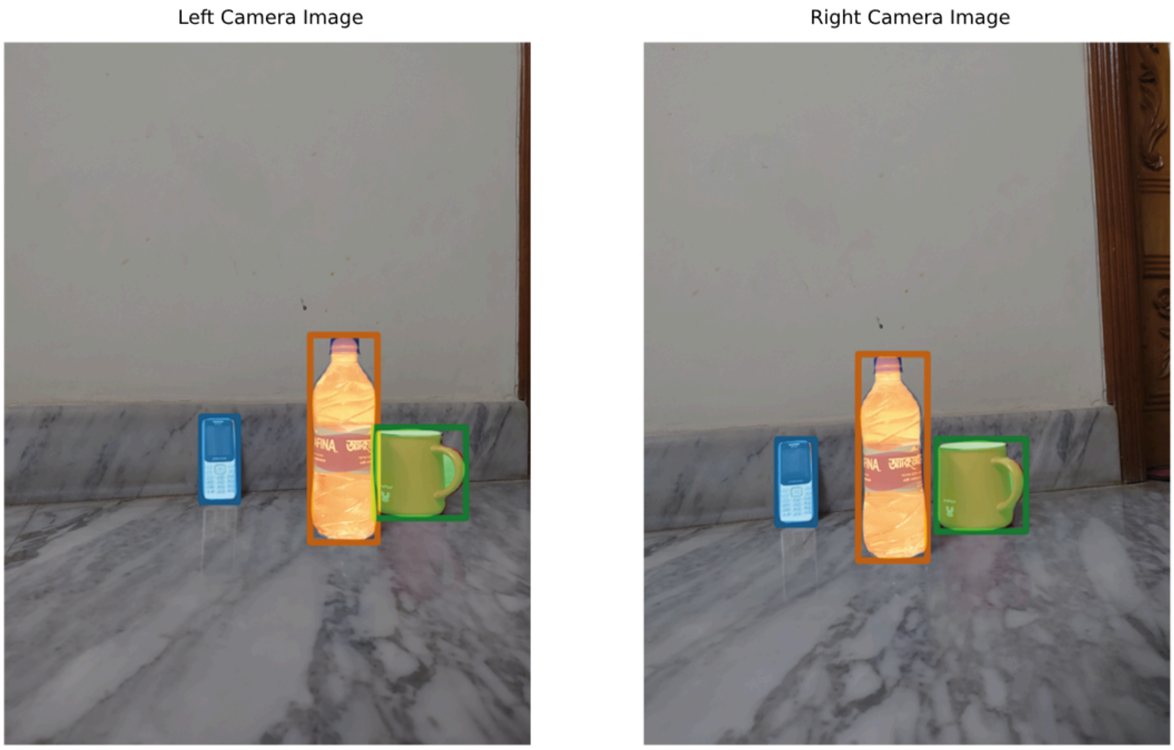


*Fig. 22: Image Segmentation and Object Detection*

Fig. 22 represents the segmented image of the detected objects that the YOLOv11 algorithm found in the input images from the cameras. In the illustration, it is visible that the object detection has been done properly.

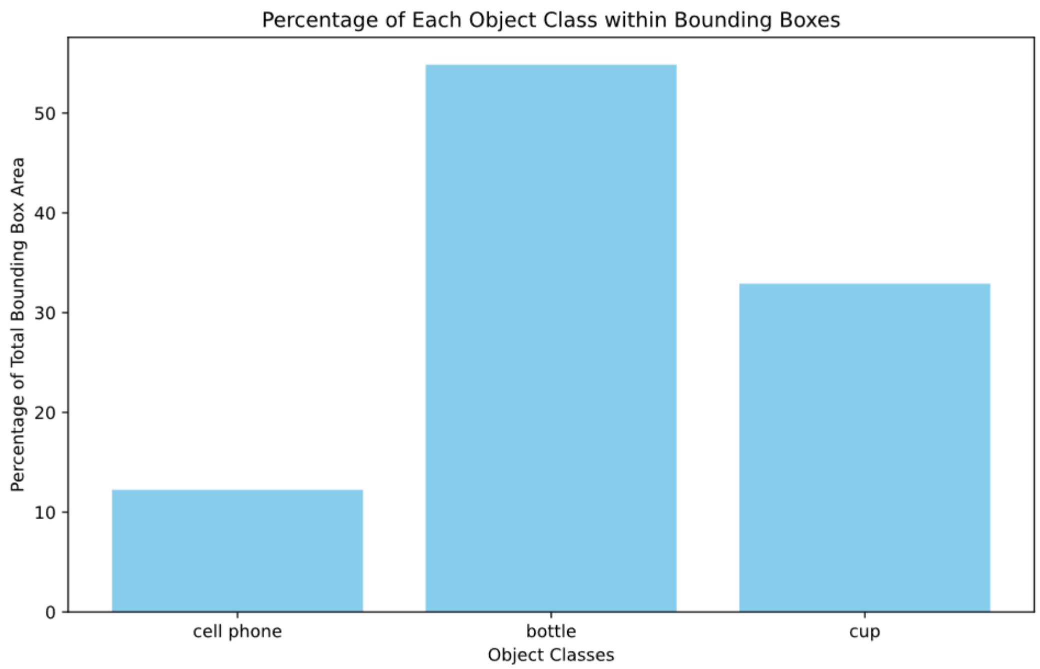


*Fig. 23: Percentage of total bounding box area*

In this study, two images were first captured, as illustrated in Fig. 25, for use in the stereovision algorithm. The two cameras were positioned 10 cm apart to obtain the left-eye and right-eye images. These images were then placed within bounding boxes for further analysis. Using the YOLOv11 algorithm, objects within the Region of Interest (ROI) were detected from a selected dataset. Next, the distance between the detected objects and the cameras was measured using the stereovision algorithm. This process involved solving two equations to find the unknown parameters, a1 and a2. The comparison of the two images provided the necessary depth information. The algorithm successfully detected various objects and calculated their distances from the cameras. Specifically, a cell phone located 74.82 cm away, a bottle at 50 cm, and a cup at 60.32 cm were identified within the defined bounding boxes.

Next, an attempt was made to show the percentage of detected objects that matched the dataset. The x-axis represents the object classes, while the y-axis represents the percentage of the total bounding box area. This plot revealed how well the objects fit within their respective bounding boxes (cell phone = 12%, bottle = 58%, cup = 35%). Additionally, Fig. 23 identifies patterns or biases in object size and shape, informing further refinements in object detection strategies to improve overall detection performance. By analyzing Fig. 5, it was observed that the detection consistency for the cell phone is the lowest, whereas for the water bottle, it is the highest.

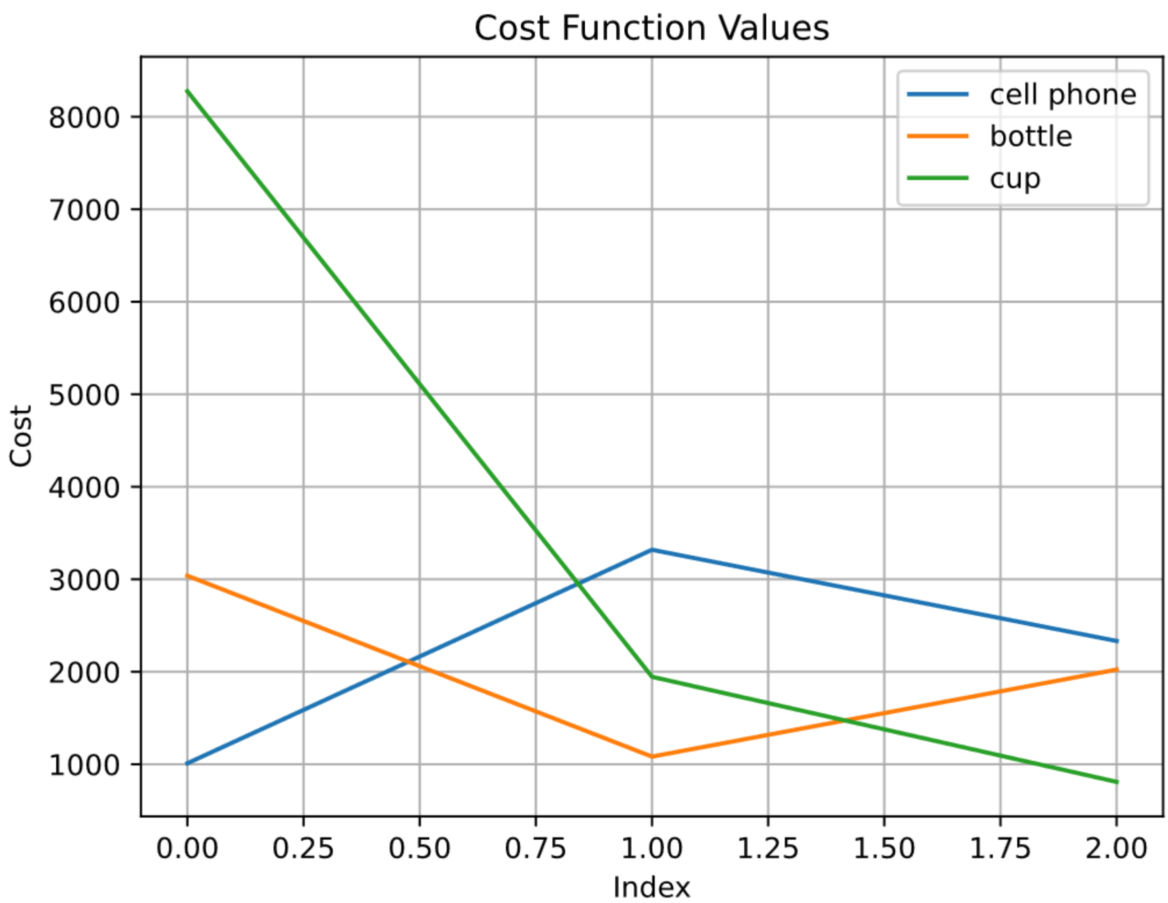


*Fig. 24: Cost function analysis graph*

In the case of Fig. 24, the associated cost of matching detected objects across different frames from the algorithm is analyzed. Each line on the graph represents a different object category, indicated by the color of the bounding box in which the object has been detected.

The x-axis represents the different indices, while the y-axis represents the cost values. From the observations, it is evident that the green line (representing the object 'cup') shows a challenging match for the cup but drops significantly at index 1, suggesting an easier match or better detection at that point. In contrast, the blue and orange lines (representing the cell phone and bottle, respectively) show more moderate and fluctuating costs.

The images provided show segmented outputs from the left and right camera feeds. In both images, the YOLOv11 algorithm has successfully identified multiple objects and highlighted them with bounding boxes of different colors:

1. Objects Detected:

- Orange Box: A plastic bottle (likely the central object).
- Green Box: A green cup adjacent to the bottle.
- Blue Box: A mobile phone on the left side of the scene.

2. Segmentation Accuracy:

- The detected objects are properly localized, with the bounding boxes closely fitting the shape and size of each object. This indicates a high level of segmentation precision by YOLOv11.
- Both the left and right images display consistent detection results, ensuring that the stereovision algorithm receives matching input for accurate disparity and depth computation.

3. Stereo Matching and Depth Estimation:

The consistency in object positioning across both camera images ensures that the stereo matching algorithm can accurately compute the disparity. This disparity is essential for calculating the distance between the camera and each object. Any slight shifts in the relative position of objects between the two images are processed by the stereovision system to estimate their depth and provide real-time distance measurements.

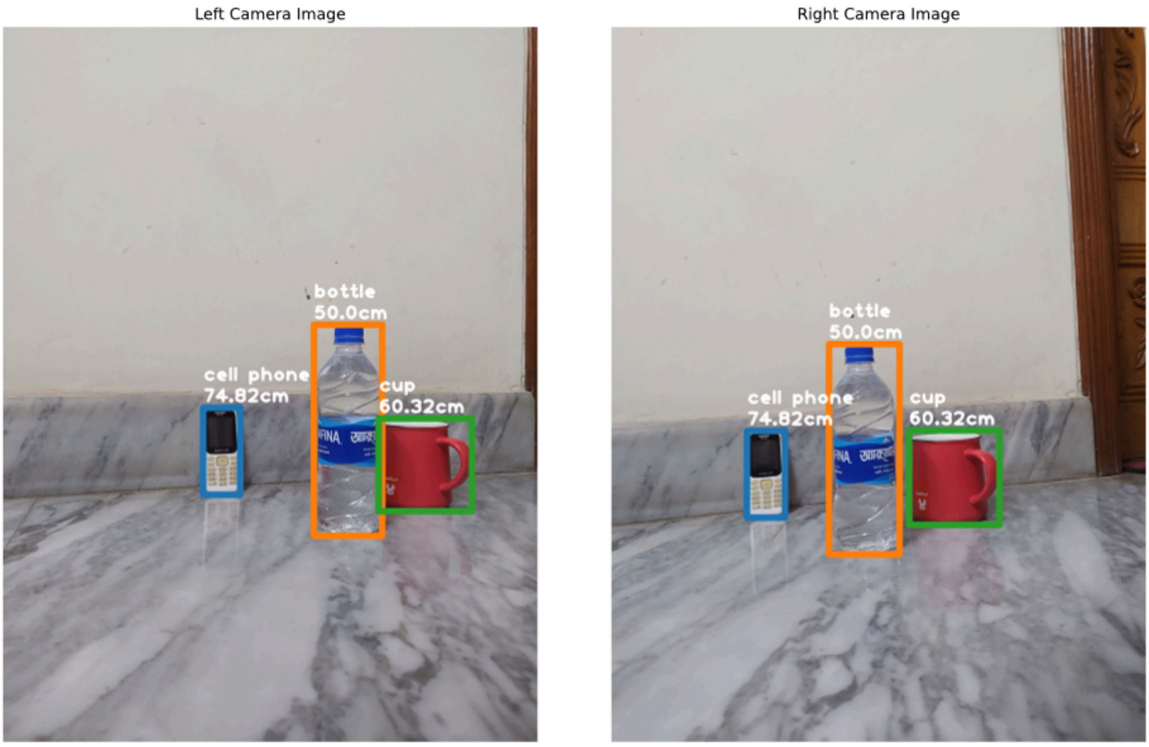


*Fig. 25: The final output*

This project highlights the potential of wearable technology for assistive and navigation purposes, showcasing a practical application of advanced computer vision algorithms in a compact, user-friendly form.

### 5.3.1 Evaluate the solution to meet desired need

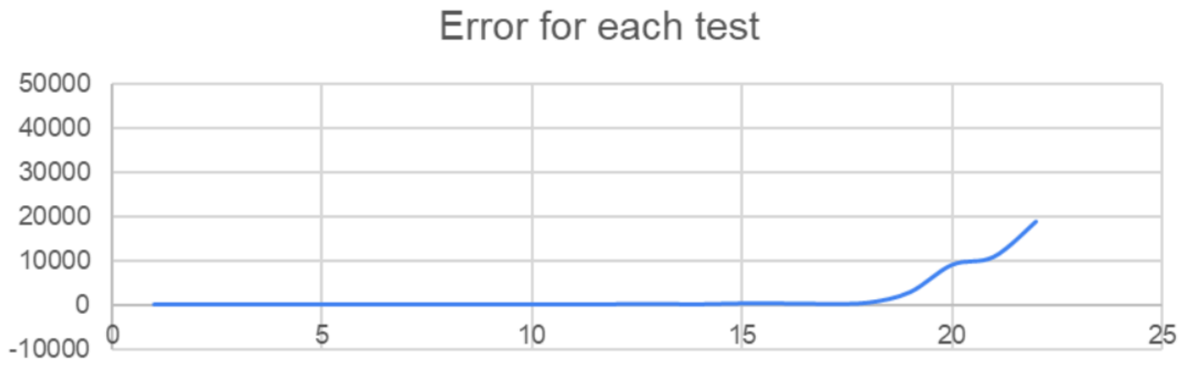


*Fig. 26: Error for each test*

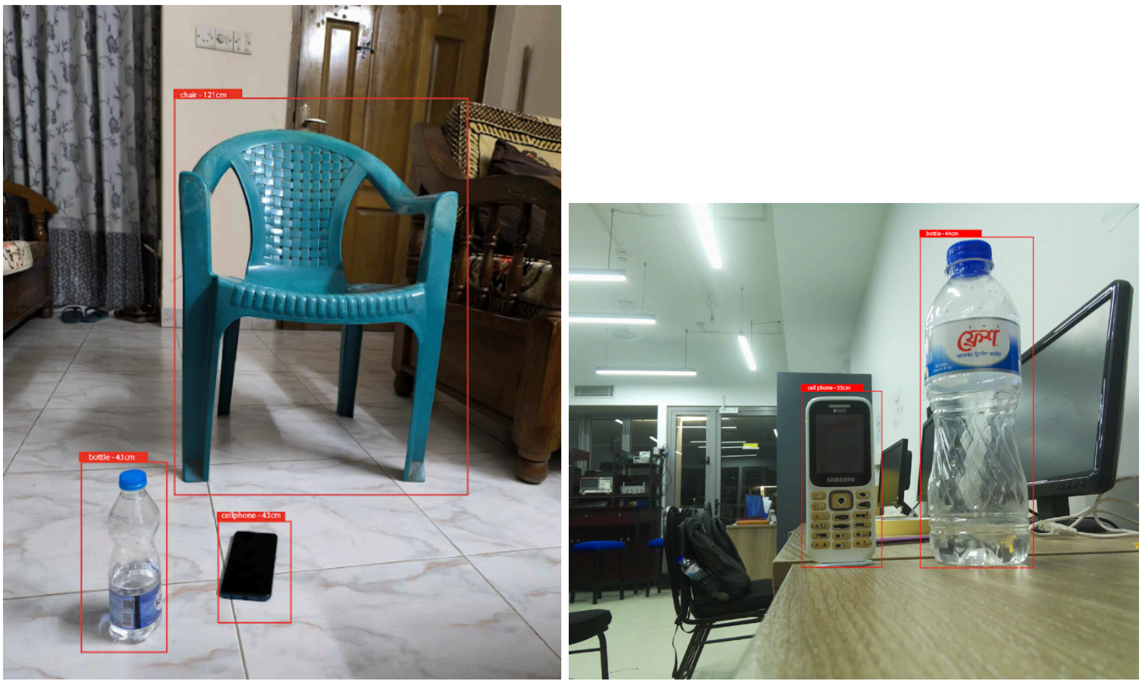


*Fig. 27: The Final Output*

## 5.4 Conclusion

Our project successfully developed a smart eyeglass using the YOLOv11 algorithm integrated with a stereovision system for object detection and distance measurement. The enhanced architecture of YOLOv11 enabled real-time, accurate identification of objects, while the stereo vision algorithm provided precise depth perception, allowing users to gauge the distance to surrounding objects effortlessly. This combination proved highly effective in delivering spatial awareness, making the smart eyeglass particularly beneficial for visually impaired individuals, indoor navigation, and other assistive applications. With its lightweight design and efficient performance, the smart eyeglass demonstrates the potential for wearable technology to enhance independence and safety, paving the way for more advanced and user-centric solutions in the future. The final design of the smart eyeglass demonstrates the effectiveness of integrating the YOLOv11 algorithm with stereovision for accurate object detection and depth estimation. The statistical analysis reveals strong performance at shorter distances, though further calibration might be required for more accurate predictions at longer ranges. The image segmentation results confirm that the system can consistently detect multiple objects with precision, providing the foundation for reliable distance measurement and situational awareness.

# Chapter 6 Impact Analysis and Project Sustainability

## 6.1 Introduction

The word "impact" describes the 'influence' that research has or its 'effects on people, their communities, the development of policies, or even the discovery of new devices or services. In essence, it refers to how research impacts our lives in a range of dimensions, from our health through social organization and legal rules to cultural life. As such, this project could have a multitude of impacts: social, legal, health, economic, among others. We shall discuss a number of these dimensions briefly.

## 6.2 Assess the impact of the solution

<u>Social Impact</u>

Social impact refers to the overall effect of an activity on a community and individual and well-being of families. We need to ensure the good bits are greater than the bad bits and also endeavour to reduce the bad bits as much as possible.

- ➢ Visual Spatial Agnosia persons will feel more empowered and they will not feel incompatible anymore.
- ➢ It will be more easier for them to socialize, reduce the stress related to spatial perception difficulties and improve their ability to recognize objects.
- ➢ The outlook of society will change towards visual spatial patients.

<u>Cultural Impact</u>

The term "cultural impact" means the effects that people have from any public or private policy and activity, which alter their norms, values, beliefs, practice, and institution. These also affect how they live, work, socialize, and organize themselves in their cultural life.

- At present, there is no specific cure available for this but there are some therapy like Occupational Therapy, Vision Therapy, Cognitive Rehabilitation Therapy (CRT) etc. But these are not actually sufficient for the patient to lead a relaxed life. In this scenario, if they start to use smart assistive glass then it will be a huge cultural change for that particular community.

Impact on Health

In our project, we are following all the rules set by ISO. Even with this, this smart assistive glass have some bad effects on human health. That is,

- **Eye Strain and Cognitive Fatigue:** In the case where wearable screens or elements of augmented reality are involved, the long-time working may result in eye strain or headaches. Similarly, the imminent auditory feedback may eventually lead to cognitive fatigue or stress unless it is specifically fitted to avoid overloading on the user.
- **Overdependence on Technology:** Users may depend too much on the device. This may lower their motivation to carry out brain exercises or participate in other rehabilitation activities. If they are using the device instead of putting any effort into it, their thinking skills may worsen over time.

### 6.3 Evaluate the sustainability

Environmental sustainability refers to responsibility for protecting natural resources and preserving a livable world ecosystem to ensure health and quality of life for present and future generations. Since most decisions on the environment involve no immediacy, another important characteristic of environmental sustainability is its forward-looking nature [13].

We will be using Raspberry Pi 5 as a microcomputer. Inside it, there is a heatsink which is used in electronic devices to dissipate heat and prevent overheating. It is typically made of a metal such as aluminum or copper. It absorbs and transfers heat from the device's central processing unit (CPU) or other heat-generating components. The Aluminium metal is normally termed green due to the numerous environmental benefits it offers. Recycling of aluminum cuts off around 95% of the energy required to make it from new resources [14]. This means it creates a positive impact on the human environment.

Aluminum production is linked to many negative environmental effects. This metal is extracted from a group of mineral known as Bauxite, that contains red dirt and clay combined together. It is mainly sourced from countries such as Australia, Brazil, and India. This Bauxite is mined utilizing open-pit mining, therefore it needs to be dug out from the ground. After being extracted, Bauxite has to be refined and treated with many chemical reactions to turn into an Aluminum metal. This conversion often requires a huge amount of energy. It hence generates a lot of environmental problems. Both open and underground mines can lead to lots of hazards with the local flora and fauna, which would impart these effects for several generations to come. Cutting trees and grasslands causes loss of biodiversity, habitat destruction, increased carbon emissions, and soil erosion. More recently, mining companies have taken this damage more seriously, and they propose plans for habitat restoration and replanting following the completion of mining operations. Also, various refining processes involve different intensities in the use of electricity and water, leading to excessive carbon release; together with air and water emissions, noise and heat contamination [15].

To power the system, a power source must be utilized, hence a battery. Mario Grosso, Professor at the Polytechnic Institute of Milan and researcher focused on the environmental impact of different types of batteries underlines the fact that the greenhouse gas emissions during rechargeable battery production are much smaller compared to disposable ones. This is because rechargeable batteries can be reused; hence, reduce production. Grosso also added that should a battery be recharged approximately up to 50 times, it is all over the point. On the contrary, if the battery use is shallow and is gotten rid of, this might damage the environment even further [16].

We decided to use rechargeable batteries instead of disposable ones in our project. Among all the different types of rechargeable batteries, lithium batteries are much better because they are lighter and cleaner than other forms, hence ideal for portable gadgets such as cordless tools, wearable technology, and wireless medical devices. Besides, they are designed to have longer endurance, something that would come in really handy with highly computerized smart gadgets and gizmos, because replacing batteries is quite a nuisance [16]. Therefore, we use lipo batteries for our entire projects.

**SWOT Analysis**

| STRENGTHS | WEAKNESSES |
|---|---|
| 1. Addresses a critical need. | 1. High development cost. |
| 2. Innovative technology. | 2. Limited awareness and data privacy concerns. |
| 3. Market potential and cost-effectiveness. | 3. Potential for bias and limited accessibility. |
| 4. Positive social impact and SDG alignment. | 4. Ethical considerations. |
| **OPPORTUNITIES** | **THREATS** |
| 1. Partnerships, funding, and customization. | 1. Leveraging strengths to address weaknesses. |
| 2. Data-driven research, global expansion, and platform integration. | 2. Capitalizing on opportunities to mitigate threats. |
| | 3. Developing strategies for sustainable and impactful development. |

*Table 18: SWOT Analysis*

**6.4 Conclusion**

To conclude, this section showed how our project related to sustainability and its general impact. We wanted to understand how our device would eventually affect society and the environment. We also studied issues such as social aspects, health and safety, legal, and cultural issues. The greatest impact we saw was in society because our project affects people's and families' lives directly. There is also a great cultural impact because it might change the way society treats the visually spatial agnosia community. We realized that collaboration with various groups and organizations, considering the rules that will apply, would be important. Next, we considered the environmental impacts of our solution and adapted our design to mitigate these negative effects. Herein, we learnt that rechargeable batteries produced high levels of greenhouse gases, thus opting for lithium batteries. These are lighter, greener, and longer-lived-all useful features in electronic devices where changing batteries is most awkward. Finally, we did a SWOT analysis to point out the strengths, weaknesses, threats, and opportunities that our device could face on the current market.

# Chapter 7 Engineering Project Management

## 7.1 Introduction

Since engineering is a complex and ever-evolving industry, engineers must acquire and modify management skills in addition to their engineering expertise in order to effectively manage any project, which is essential to their professional success. Although most engineers view themselves as technical experts in their domains, managing budgets, timelines, and personnel in a professional setting necessitates a fundamental grasp of project management techniques.

## 7.2 Methodology

In order to conduct this engineering capstone project we have to maintain a generalized method throughout the whole year. The road map of our project plan shows how we will do our Research and do the necessary activities throughout the whole given timeline. After starting our project at first we are going to define a problem which will be a real life problem and this problem can be solved using the knowledge of engineering of electrical and electronics. Then we are going to do our background Research and literature review where we will search through research papers on the internet and also collect primary data if possible. Then after analyzing the information we have we are going to create a complete project plan and successfully submit our project proposal. Finally, we made a plan for all the things we needed to do and wrote a detailed proposal explaining everything that is necessary for our project.

Next, we target the development of the different portions of the system. Then we will try to integrate the whole system. This approach will start our simulation and testing of the system where we will go on a process that if the simulation works properly or not if it does not work then we have to remodel our design and try it again and again and if it does work in our approach model then we will step into our design and constructing the hardware prototype of the project.

Now comes the most important part of our project which is testing the prototype which we have just made then we will observe if our prototype gives the desired outcome which we need through our project if it does not then we have to change the design in our prototype or check if there any important approach which we missed in our design calculation and match it with the simulation that we have made. After this if our prototype design works then we will check if there are any potential errors in our design . If there are Errors then we will try to remove them with proper engineering knowledge of debugging and the rivers engineering process. Then if our photo type works perfectly without any errors then we need to make proper documentation and give a presentation of our project that we have worked on for 1 year.

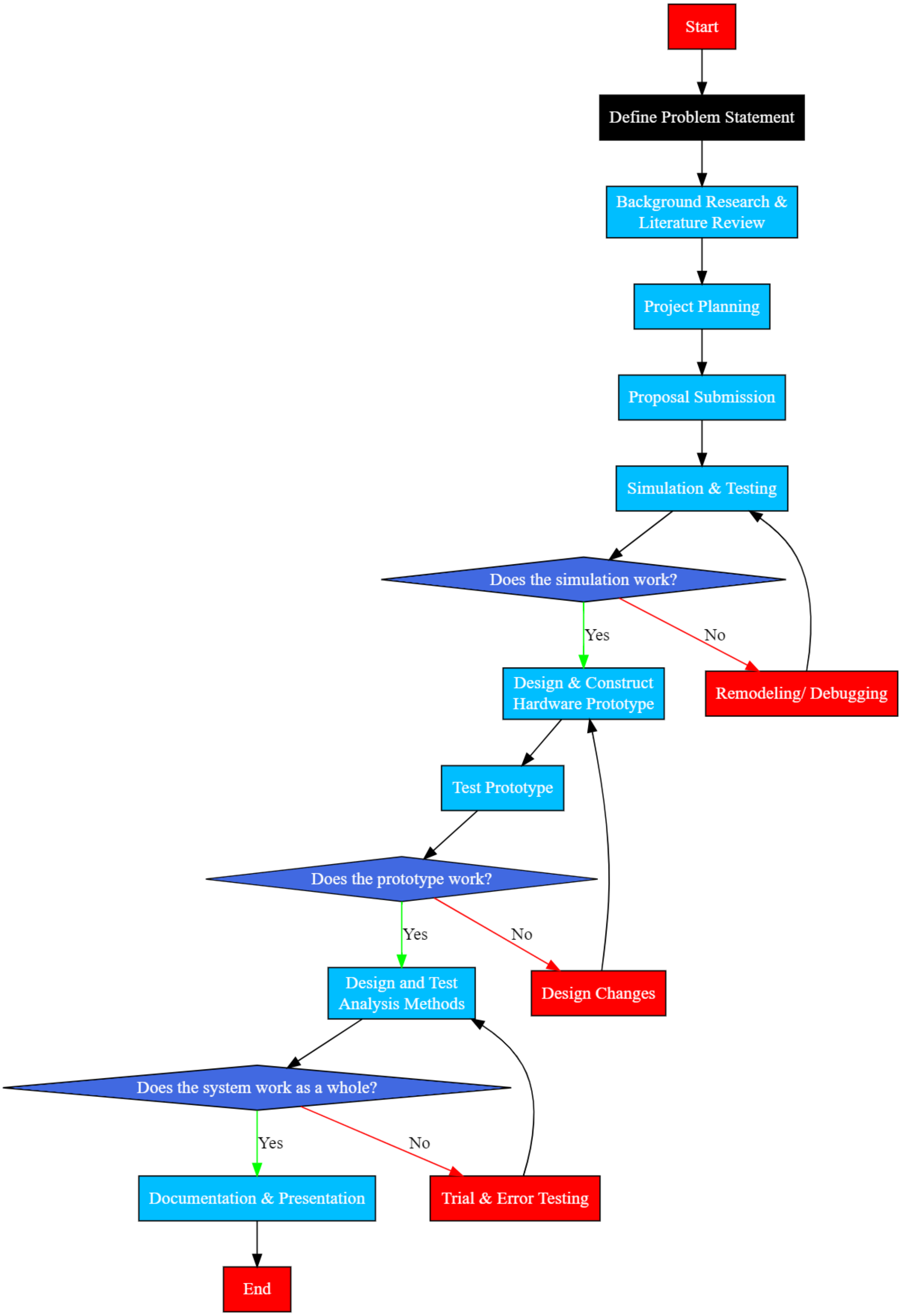


*Fig. 28: Flowchart of Methodology*

### 7.3 Define, plan and manage engineering project

The methodical preparation and execution of such strategy by a team of engineers is known as engineering project management. It employs the same techniques and procedures as traditional project management, with a primary focus on engineering projects. It entails learning the project objectives and developing several situations using their backup plan. The project for engineers, management is a crucial stage since unforeseen issues might occur at any time. It impedes several people's work and might impede the project's advancement.

Initialisation, development, and implementation are the three phases into which our engineering project was separated. EEE-400P (initialisation), EEE-400D (development), and EEE-400C (implementation) are the three courses to which these three phases are attributed. A logbook has been kept to record the agendas and meetings in order to oversee our project. The group members were given equal work duties at each step. Gantt charts were created for each of the three phases of our project in order to monitor its development. The Gantt chart and the logbook have been used to assess our project's progress in relation to the project plan. We created rough Gantt charts for 400P, 400D, and 400C during our EEE-400P course. Our estimation served as the basis for the task and deadline suggested in those Gantt charts.

## 7.4 Evaluate project progress

### Project Plan (EEE499P)

| ID | Task Description | Duration | Start Date | End Date |
|---|---|---|---|---|
| 1 | Research Regarding Topic Selection | 15 | Oct 01, 2023 | Oct 15, 2023 |
| 2 | Complex Engineering Problem Identification | 18 | Oct 01, 2023 | Oct 18, 2023 |
| 3 | Search Related Papers | 20 | Oct 15, 2023 | Nov 05, 2023 |
| 4 | Selection of Final Topic | 6 | Oct 20, 2023 | Oct 25, 2023 |
| 5 | Tentative Problem Statement & Objectives | 2 | Oct 22, 2023 | Oct 24, 2023 |
| 6 | Multiple Design Approaches | 3 | Oct 22, 2023 | Oct 25, 2023 |
| 7 | Draft Concept Note | 4 | Oct 20, 2023 | Oct 24, 2023 |
| 8 | Requirements, Specifications & Constraints | 14 | Oct 25, 2023 | Nov 07, 2023 |
| 9 | Applicable Standards & Codes | 1 | Oct 24, 2023 | Oct 24, 2023 |
| 10 | Slide Preparation for Progress Presentation | 2 | Oct 24, 2023 | Oct 25, 2023 |
| 11 | Progress Presentation | 1 | Oct 25, 2023 | Oct 25, 2023 |
| 12 | Final Concept Note | 21 | Oct 20, 2023 | Nov 10, 2023 |
| 13 | Backround Study & Literature Review | 15 | Nov 05, 2023 | Nov 20, 2023 |
| 14 | Methodology, Budget & Project Plan | 3 | Nov 20, 2023 | Nov 23, 2023 |
| 15 | Sustainability & Ethical Considerations | 3 | Nov 24, 2023 | Nov 27, 2023 |
| 16 | Risk & Safety Measures | 5 | Nov 20, 2023 | Nov 24, 2023 |
| 17 | Draft Project Proposal | 25 | Nov 15, 2023 | Dec 10, 2023 |
| 18 | Slide Preparation for Final Presentation | 3 | Dec 10, 2023 | Dec 13, 2023 |
| 19 | Final Presentation | 1 | Dec 14, 2023 | Dec 14, 2023 |
| 20 | Final Project Proposal | 15 | Dec 05, 2023 | Dec 20, 2023 |

*Fig. 29: Timeline of FYDP-P*

### Project Plan (EEE499D)

| ID | Task Description | Duration | Start Date | End Date |
|---|---|---|---|---|
| 1 | Market Research | 4 | Jan 24, 2024 | Jan 28, 2024 |
| 2 | Components Requirement | 4 | Jan 24, 2024 | Jan 28, 2024 |
| 3 | Optimal Solution | 10 | Jan 28, 2024 | Feb 04, 2024 |
| 4 | Analyzing the Optimal Solution | 7 | Jan 26, 2024 | Feb 01, 2024 |
| 5 | Debugging | 8 | Feb 02, 2024 | Feb 10, 2024 |
| 6 | Software Simulation | 13 | Feb 08, 2024 | Feb 21, 2024 |
| 7 | Troublshooting the Simulation | 11 | Feb 23, 2024 | Mar 06, 2024 |
| 8 | Slide Preparation | 4 | Mar 07, 2024 | Mar 11, 2024 |
| 9 | Progress Ptresentation | 5 | Mar 12, 2024 | Mar 17, 2024 |
| 10 | Draft Project Report | 7 | Mar 18, 2024 | Mar 25, 2024 |
| 11 | Breadboard Implementation | 25 | Mar 27, 2024 | Apr 21, 2024 |
| 12 | Debugging | 15 | Apr 06, 2024 | Apr 21, 2024 |
| 13 | Slide Preparation | 4 | Apr 21, 2024 | Apr 24, 2024 |
| 14 | Final Presentation | 2 | Apr 25, 2024 | Apr 26, 2024 |

*Fig. 30: Timeline of FYDP-D*

**Project Plan (EEE499C)**

| Task ID | Task Description | Duration | Start Date | End Date | May 22, 2024 | May 26, 2024 | Jun 03, 2024 | Jun 08, 2024 | Jun 09, 2024 | Jun 15, 2024 | Jul 05, 2024 | Jul 12, 2024 | Jul 13, 2024 | Jul 14, 2024 | Jul 17, 2024 | Jul 18, 2024 | Jul 19, 2024 | Jul 21, 2024 | Jul 22, 2024 | Jul 23, 2024 | Jul 24, 2024 | Aug 14, 2024 | Aug 23, 2024 | Aug 25, 2024 | Aug 27, 2024 | Aug 28, 2024 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Prototype Develepment | 17 | May 22, 2024 | Jun 08, 2024 | | | | | | | | | | | | | | | | | | | | | | |
| 2 | Programming the Analysis Unit | 5 | May 22, 2024 | May 26, 2024 | | | | | | | | | | | | | | | | | | | | | | |
| 3 | Component Testing | 8 | May 26, 2024 | Jun 03, 2024 | | | | | | | | | | | | | | | | | | | | | | |
| 4 | Development | 6 | Jun 03, 2024 | Jun 09, 2024 | | | | | | | | | | | | | | | | | | | | | | |
| 5 | Protype Testing | 28 | Jun 15, 2024 | Jul 13, 2024 | | | | | | | | | | | | | | | | | | | | | | |
| 6 | Troubleshooting | 7 | Jul 05, 2024 | Jul 12, 2024 | | | | | | | | | | | | | | | | | | | | | | |
| 7 | Final Prototype Testing | 9 | Jul 05, 2024 | Jul 14, 2024 | | | | | | | | | | | | | | | | | | | | | | |
| 8 | Debugging | 4 | Jul 18, 2024 | Jul 22, 2024 | | | | | | | | | | | | | | | | | | | | | | |
| 9 | Progress Presentation | 2 | Jul 17, 2024 | Jul 18, 2024 | | | | | | | | | | | | | | | | | | | | | | |
| 10 | Project Debugging | 4 | Jul 19, 2024 | Jul 23, 2024 | | | | | | | | | | | | | | | | | | | | | | |
| 11 | Project Report | 4 | Jul 21, 2024 | Jul 24, 2024 | | | | | | | | | | | | | | | | | | | | | | |
| 12 | Prototype Project Demonstration | 9 | Aug 14, 2024 | Aug 23, 2024 | | | | | | | | | | | | | | | | | | | | | | |
| 13 | Final Presentation | 2 | Aug 27, 2024 | Aug 28, 2024 | | | | | | | | | | | | | | | | | | | | | | |
| 14 | Project Showcase | 3 | Aug 25, 2024 | Aug 27, 2024 | | | | | | | | | | | | | | | | | | | | | | |

*Fig. 31: Timeline of FYDP-C*

Regarding the status of our project, we have finished all of the work in the initialisation stage during the first semester (EEE-400P) and all of the duties in the development stage during the second semester (EEE-400D). We tested and assessed our chosen solution during the third semester (EEE-400C), put it into practice, conducted a cost-benefit analysis, and then created the necessary paperwork and sent in the project report. Therefore, we have finished the EEE-400C implementation phase.

## 7.5 Conclusion

We have managed our engineering project inside the deadline with little to no fuss thanks to a logbook to record meetings and agendas and a Gantt chart for each of the project's three stages to assess our progress.

# Chapter 8 Economical Analysis.

## 8.1 Introduction

The evaluation of a project's costs and benefits is known as economic analysis. It is employed to ascertain the economic feasibility, allowing us to approximate the profit, that is, the return that may be fairly anticipated from the project in relation to the investment. This will also help with improved resource allocation and project welfare effect analysis. Below is a discussion of our project's economic analysis.

## 8.2 Economic analysis

For that part, we have calculated the probable cost for the project, then estimated the benefits that we will get from the project when it will be done and then compared the cost and benefits to get the idea if it is beneficial to invest in the project.

**Breakdown for Approach 1 Budget**

**Smart V-Glass Assistant Budget:**

| **Subsystem** | **Component Name and Specifications** | **Quantity** | **Cost (TK)** |
|---|---|---|---|
| Frame | Eye glass | 1 | 1500 |
| Power supply | 2-cell Lipo Battery | 1 | 2500 |
| Internal Processing Unit | Raspberry Pi 5 Model B 8GB | 1 | 22000 |
| Sensors | Micro Camera | 2 | 9280 |

| | | | |
|---|---|---|---|
| Sound system | Earbuds | 1 | 2100 |
| Visual Unit | Wearable display | 1 | 36000 |
| Storage Unit | Micro SD | 1 | 1700 |
| Others | …. | …. | 4500 |
| | | Total | 79580 |

*Table 10: Budget for design approach 1*

**Breakdown for Approach 2 budget:**

**Smart V-Watch Navigation Assistant Budget**

| **Subsystem** | **Component Name and Specifications** | **Quantity** | **Cost (TK)** |
|---|---|---|---|
| Frame | Eye glass | 1 | 1500 |
| Power supply | 2-cell Lipo Battery | 1 | 2000 |
| Internal Processing Unit + Communication module | ESP32 S2 Mini LOLIN | 1 | 1000 |
| Sensors | Micro Camera | 1 | 12000 |

| | | | |
|---|---|---|---|
| | Monochrome camera | 2 | 15000 |
| Sound system<br>+<br>Visual Unit | On smart watch | 1 | ….. |
| Others | …. | …. | 5000 |
| | | Total | 36500 |

*Table 11: Budget for design approach 2*

We have analyzed the budgets of both of the design approaches. We found that approach 1 costs much more than approach 2. Where approach 1 costs almost 80 thousand bdt, approach 2 costs only 36500 bdt. But as approach 1 fits well with our expectations and requirements, we have chosen approach 1.

**8.3 Cost benefit analysis**

We have spent almost 80 thousand on our prototype, but we could improve that. Our wearable display costs much more than the expected budget due to unavailability in the market. We have to import it from Japan and it costs almost 36 thousand, which is too much expensive. If it was done on a bigger scale and if we could arrange the electronics, used for that display, it could be made with a much cheaper amount. Also if we get the resources, we can make a device that will have fewer errors. We can also make that device fully wireless at a lower cost. We expect that our device will not only be used by visual-spatial agnosia patients, but also by other eye patients and it can also benefit blind people as it has audio output. Therefore, if it is made in a bigger scale, it will get cheaper and also more of that device will be available to the people.

### 8.4 Evaluate economic and financial aspects

While evaluating the economic and financial aspects of our project, we have considered many factors that affect both the initial investment and the long-term financial sustainability of the device:

**Initial Investment**

As we have mentioned before, approach one costs almost 80 thousand, and approach 2 costs 36500 initially. But approach one is more beneficial to invest even if it is much more costly. Key components like the wearable display, Raspberry Pi 5, and dual micro cameras are major cost drivers. Although these items are expensive in the prototype phase, we expect to realize cost savings during mass production through economies of scale, local sourcing, and design optimization.

**Long-Term Viability**

Large production of this device will reduce costs. If we can produce the components locally, it can reduce the cost of production significantly and also can reduce the dependency like we had to bring the expensive display from Japan. However, The device not only helps visual-spatial agnosia patients but can also serve a wider audience, including individuals with various visual impairments or blindness. It will increase the market size and investment will increase for those devices.

**Revenue Potential**

After the market acceptance, we can approach hospitals and organizations to distribute the device. Government support and subsidies for medical devices will boost sales.

**Cost-Effectiveness**

If we can improve our device for industrial production, it will be cheaper than the prototype. As it has a wide range of users, its market size will be bigger.

**Financial Risks**

**Technological Advancements**

There is a risk that if a new and improved technology comes to the market, our device will be less competitive in the future. To get rid of that risk, we have to improve out device and update the technology continuously.

**Funding and Investment**

A large scale funding will be needed to carry on our production.

### 8.5 Conclusion

The economic analysis ends here by showing the benefits of investing in the large-scale production of the Smart V-Glass Assistant device. In the era of technology, visual-spatial agnosia patients don't have any assistive technology yet. Though we developed a device specially dedicated to them, it also has the potential to assist some of the other diseases. Therefore, it has a larger market. The project is a strategic effort aimed at increasing the efficiency and productivity of industrial operations. Given its potential for cost savings, market opportunities, and sustainable benefits

# Chapter 9 Ethics and Professional Responsibilities

## 9.1 Introduction

The guiding rules of professional ethics and responsibilities help in decision-making during the performance of different jobs, and one has to ensure that one follows these rules to maintain the ethical standards while at work. Essential ingredients of such rules are integrity, responsibility, compassion, and kindness.

## 9.2 Identify Ethical Issues and Professional responsibility

In a professional world we must follow the professional responsibilities. In the case of our project, the responsibilities are narrowed down by following: Practice the use of references that are properly attributed or cited without plagiarising. The necessary standards and codes must also be adhered to. Besides professional ethics, ethical implications must also be put into practice. According to the paper [17],

1. Support people with a visual spatial agnosia or low vision to participate in our project and help share it with others.
2. Ensuring the data output of our project is available to every participating agent in a testing procedure.
3. Take necessary measures to ensure the protection of participants from potential harm.
4. Get free and informed consent from participants before allowing them to participate in product testing.
5. Understand and comply with all relevant legal regulations before carrying out any testing activities.

### 9.3 Apply ethical issues and professional responsibility

We had been working on this project report following the rules in giving proper credit to all our references, never duplicated an existing design from other projects or publications. In addition, our team is committed to dealing with various ethical issues: anonymity, confidentiality, and informed consent of those respondents or participants. We do have a prepared consent form regarding this issue. Before being enrolled, all participants must go through this form and agree to all the terms included therein. For those participants who may be blind or having poor eyesight, we shall ensure that their family members or friends will help them understand the consent form. Consent from participants will be requested with the assistance of this form-a sample is given below:

Please take a moment to read and sign this document.

In this usability test:

- You will get the chance to try out our assistive device.
- We will provide a brief training session to help you use device effectively.
- Your feedback and insights will be crucial for enhancing the profile of the device.

Joining this usability test is completely voluntary. All your information will remain confidential. The details and outcomes may be used to make the smart device better, but we will never share your name or any personal details. You can choose to withdraw from the study at any time. If you have any questions, feel free to reach out to Adri Shankar Das at adri.shankar.das@g.bracu.ac.bd.

I have read and fully understand the information on this form and have had all of my questions answered.

______________________________          ____________________

Subject's Signature          Date

### 9.4 Conclusion

Engineering practice entitles us to uphold the highest degree of honesty and integrity in fulfilling our professional ethics and responsibilities and to prioritize the safety, health, and welfare of the public.

# Chapter 10 Conclusion and Future Work

## 10.1 Project Summary/Conclusion

The proposed device has been developed for those people who cannot navigate properly due to visual-spatial agnosia. Much importance was given to understanding the needs of the users during the design process. A comparative study showed which design and solution would be best for this purpose. The final setup includes two monochrome cameras which combined use as a stereo vision camera, an intelligent eyewear framework, a display facing the eyewear, Bluetooth capability for audio output, and a Raspberry Pi 5 as a micro-computer. Later on, the stereo vision algorithm is used to measure the distances while object detection makes use of YOLOv11. One of the most important parts of this device is the power supply module, which is very essential in helping this device keep working without any interruptions for hours and not frequently requires recharging. For that purpose, a power bank is used as an energizing device in this paper. Also, impact analysis, economic evaluation, discussion on the consideration of ethics, codes pertaining to the project, and the logbook regarding the development of the project are attached within the report.

## 10.2 Future work

It has made available an affordable, accessible smart glass used by persons with visual spatial agnosia to deal with their surroundings easily. Mass production will make the device cheaper hence more accessible to persons who may require them. However, much work is yet to be done in an effort to develop more meaningful services for the visually impaired community. Many problems had arisen, but the biggest had been limitation in budget. Working on the project with Raspberry Pi 5 showed that the skill in object detection and measurement was just not good enough. Perhaps with Jetson Nano, some changes will surely improve the speed in doing detection. Compared in the paper [18] with a table of major features in Jetson TX2, Jetson Nano, and Raspberry Pi, it seems that Jetson TX2 outperforms all the features, though at a higher cost, while the Jetson Nano could be a possible choice but is out of budget at the moment. We believe that the usage of either Jetson Nano or Jetson TX2 would yield more accurate results.

Wireless connectivity between different subsystems is also a possibility, then it will be more portable and handy. Again, the incorporation of Google Maps is also possible on the device. Moreover, artificial intelligence, deep learning, and image processing integration can considerably enhance the working of the product by making the product more sensitive in the detection of obstacles. Technologies like these would detect obstacles much more accurately. However, very few assistive technology devices are presently available in Bangladesh, and hence research studies are essential to understand the challenges faced by the community in daily living and to design more useful assistive devices.

# Chapter 11 Identification of Complex Engineering Problems and Activities

### 11.1: Identify the attribute of complex engineering problem (EP)

**Table 12: Attributes of Complex Engineering Problems (EP)**

| | **Attributes** | **Put tick (√) as appropriate** |
|---|---|---|
| P1 | Depth of knowledge required | √ |
| P2 | Range of conflicting requirements | |
| P3 | Depth of analysis required | √ |
| P4 | Familiarity of issues | |
| P5 | Extent of applicable codes | |
| P6 | Extent of stakeholder involvement and needs | √ |
| P7 | Interdependence | |

### 11.2: Provide reasoning how the project address selected attribute (EP)

**P1: Depth of Knowledge Required**: The project necessitates a profound understanding of computer vision algorithms, such as stereovision and YOLOv11, and their integration with the Raspberry Pi 5. This requires expertise in both hardware and software engineering to ensure optimal object detection and distance estimation.

**P3: Depth of Analysis Required**: Implementing stereovision for depth estimation and YOLOv11 for object detection involves comprehensive analysis. This includes calibrating the cameras, tuning the models for accuracy, and evaluating the system's performance to ensure that it operates effectively in real-world environments.

**P6: Extent of Stakeholder Involvement and Needs**: The project addresses the needs of patients and healthcare professionals. Gathering input from clinicians is crucial to refine the design, usability, and effectiveness of the device. We have collected data from 23 doctors and medical students.

### 11.3 Identify the attribute of complex engineering activities (EA)

**Table 13 : Attributes of Complex Engineering Activities (EA)**

| | **Attributes** | **Put tick (√) as appropriate** |
|---|---|---|
| A1 | Range of resource | √ |
| A2 | Level of interaction | |
| A3 | Innovation | |
| A4 | Consequences for society and the environment | √ |
| A5 | Familiarity | |

### 11.4 Provide reasoning how the project address selected attribute (EA)

**A1:** The project uses diverse resources, including specialized hardware (cameras, display, speakers, Raspberry Pi 5) and advanced algorithms (stereovision, YOLOv11 algorithms). It also involves collaboration among experts in medical science.

**A4: Consequences for Society and the Environment**: This project has a positive social impact by providing essential support to a vulnerable group. It aims to improve the independence and safety of individuals with visual-spatial agnosia. Additionally, using energy-efficient components, like the Raspberry Pi 5, helps to minimize the environmental footprint.

## References

[1] F. J. Rowe, L. R. Hepworth, C. Howard, K. L. Hanna, C. P. Cheyne, and J. Currie, "High incidence and prevalence of visual problems after acute stroke: An epidemiology study with implications for service delivery," *PLoS ONE*, vol. 14, no. 3, pp. e0213035–e0213035, Mar. 2019, doi: https://doi.org/10.1371/journal.pone.0213035.

[2] O. Martinaud, "Visual agnosia and focal brain injury," *Revue Neurologique*, vol. 173, no. 7–8, pp. 451–460, Jul. 2017, doi: https://doi.org/10.1016/j.neurol.2017.07.009.

[3] "Agnosia – Scottish Acquired Brain Injury Network," *Scot.nhs.uk*, 2024. https://www.acquiredbraininjury-education.scot.nhs.uk/impact-of-abi/visual-and-perceptual-impairments/agnosias/ (accessed Nov. 05, 2024).

[4] D. D. Dilks, J. T. Serences, B. J. Rosenau, S. Yantis, and M. McCloskey, "Human Adult Cortical Reorganization and Consequent Visual Distortion," *Journal of Neuroscience*, vol. 27, no. 36, pp. 9585–9594, Sep. 2007, doi: https://doi.org/10.1523/jneurosci.2650-07.2007.

[5] A. Kumar and M. Wroten, "Agnosia," *Nih.gov*, Jan. 30, 2023. https://www.ncbi.nlm.nih.gov/books/NBK493156/ (accessed Nov. 05, 2024).

[6] J. Karki, S. Rushton, S. Bhattarai, and L. D. Witte, "Access to assistive technology for persons with disabilities: a critical review from Nepal, India and Bangladesh," *Disability and Rehabilitation Assistive Technology*, vol. 18, no. 1, pp. 8–16, Mar. 2021, doi: https://doi.org/10.1080/17483107.2021.1892843.

[7] A.-F. M. Schölvinck, C. A. C. M. Pittens, and J. E. W. Broerse, "The Research Priorities of People with Visual Impairments in the Netherlands," *Journal of Visual Impairment & Blindness*, vol. 111, no. 3, pp. 201–217, May 2017, doi: https://doi.org/10.1177/0145482x1711100302.

[8] F. Liu, G. Emerson, and N. Collier, “Visual Spatial Reasoning,” *Transactions of the Association for Computational Linguistics*, vol. 11, pp. 635–651, Jan. 2023, doi: https://doi.org/10.1162/tacl_a_00566.

[9] Triphena Delight D and Karunakaran V, “Deep Learning based Object Detection using Mask RCNN,” Jul. 2021, doi: https://doi.org/10.1109/icces51350.2021.9489152.

[10] R. Sapkota, Z. Meng, M. Churuvija, X. Du, Z. Ma, and M. Karkee, “Comprehensive Performance Evaluation of YOLO11, YOLOv10, YOLOv9 and YOLOv8 on Detecting and Counting Fruitlet in Complex Orchard Environments,” *arXiv.org*, 2024. https://arxiv.org/abs/2407.12040 (accessed Oct. 05, 2024).

[11] Rostam Affendi Hamzah and H. Ibrahim, “Literature Survey on Stereo Vision Disparity Map Algorithms,” *Journal of Sensors*, vol. 2016, pp. 1–23, Dec. 2015, doi: https://doi.org/10.1155/2016/8742920.

[12] J. McFIE, M. F. PIERCY, and O. L. ZANGWILL, “VISUAL-SPATIAL AGNOSIA ASSOCIATED WITH LESIONS OF THE RIGHT CEREBRAL HEMISPHERE,” *Brain*, vol. 73, no. 2, pp. 167–190, Jan. 1950, doi: https://doi.org/10.1093/brain/73.2.167.

[13] “What Is Environmental Sustainability?,” *Sphera*, Jun. 25, 2024. https://sphera.com/resources/glossary/what-is-environmental-sustainability/ (accessed Oct. 01, 2024).

[14] B. O. Rosseland, T. D. Eldhuset, and M. Staurnes, “Environmental effects of aluminium,” *Environmental Geochemistry and Health*, vol. 12, no. 1–2, pp. 17–27, Mar. 1990, doi: https://doi.org/10.1007/bf01734045.

[15] D. Grossman, “Rechargeable or disposable batteries?,” *Yale Climate Connections*, Nov. 29, 2016. https://yaleclimateconnections.org/2016/11/which-battery-is-better-rechargeable-or-disposable/ (accessed Nov. 04, 2023).

[16] “Lithium vs Alkaline Batteries – Which is better? | Grepow,” *Grepow.com*, 2021. https://www.grepow.com/blog/lithium-vs-alkaline-batteries.html (accessed Nov. 01, 2023).

[17] “Ethical Guidance for Research with People with Disabilities: NDA Report - National Disability Authority,” *National Disability Authority*, 2024. https://nda.ie/publications/ethical-guidance-for-research-with-people-with-disabilities-report (accessed Nov. 05, 2023).

[18] A. A. Suzen, B. Duman, and B. Sen, “Benchmark Analysis of Jetson TX2, Jetson Nano and Raspberry PI using Deep-CNN,” *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, Jun. 2020, doi: https://doi.org/10.1109/hora49412.2020.9152915.

# Appendix

## Logbooks

**Table 14: ATC panel and Members details**

| | Final Year Design Project (P) Summer 2021 | | |
|---|---|---|---|
| **Student Details** | **NAME & ID** | **EMAIL ADDRESS** | **PHONE** |
| **Member 1** | Adri Shankar Das (20321035) | adri.shankar.das@g.bracu.ac.bd | 01799611544 |
| **Member 2** | Fardin Rahman Sahill (20321025) | fardin.rahman.sahill@g.bracu.ac.bd | 01707514840 |
| **Member 3** | Ababil Hossain Fahad (20321034) | ababil.hossain.fahad@g.bracu.ac.bd | 01793466616 |
| **Member 4** | Mohammed Shakib (20321029) | mohammed.shakib@g.bracu.ac.bd | 01934387978 |
| **ATC Details: ATC 2** | | | |
| **Chair** | Dr. Abu Hamed M. Abdur Rahim | abu.hamed@bracu.ac.bd | |
| **Member 1** | Tasfin Mahmud | tasfin.mahmud@bracu.ac.bd | |
| **Member 2** | Md. Mehedi Hasan Shawon | mehedi.shawon@bracu.ac.bd | |

**Table 15: FYDP (P) Fall 2023 Summary of Team Log Book/ Journal**

| Date/Time/Place | Attendee | Summary of Meeting Minutes | Responsible | Comment by ATC |
|---|---|---|---|---|
| 5-10-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | Ice-breaking session with ATC committee. | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | 1. Find research topics<br>2. Read related research papers. |
| 06-10-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | 1. Brainstorming for research topic<br>2. Reading research paper | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | |
| 08-10-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | 1. Selecting a research topic<br>2. Reviewing relevant research paper. | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | |
| 10-10-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | 1. Proposing our research topic to ATC committee<br>2. Getting feedback on our topic | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | "Find more compelling complex engineering topic" |
| 12-11-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | 1. Changing our research topic according to feedback. Selecting another topic for FYDP. Reading relevant research papers on our new topic | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | |
| 16-11-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | "1. Proposing a new topic to our ATC committee.<br>2. Feedback on our research topic by ATC committee" | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | "Approval of the topic and read additional research paper on the specific topic" |
| 18-11-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | "Submission of concept note" | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | |
| 20-11-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | "Receiving feedback through progress presentation" | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | "1. Change design approaches<br>2. Update problem statement<br>3. Specify functional requirements" |

| | | | | |
|---|---|---|---|---|
| 23-11-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | "1. Making necessary changes on our concept note<br>2. Preparing our Final Project Concept note" | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | |
| 26-11-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | "Submission of Final Project Concept note" | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | |
| 30-11-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | "1. Reading more research papers<br>2. Preparing Project Proposal" | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | |
| 03-12-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | "Consulting with ATC committee on our research topic" | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | |
| 05-12-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | "1. Submission of the Project Proposal draft<br>2. Giving Mock Presentation to our ATC committee" | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | "1. Maintain time limit during Final Presentation<br>2. Group practice for final presentation" |
| 07-12-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | "Finalizing Project Proposal" | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | |
| 09-12-2023 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | Submission of the Final Project Proposal of FYDP (P) | 1. Adri | |

**Table 16: FYDP (D) Spring 2024 Summary of Team Log Book/ Journal**

| Date/Time/Place | Attendee | Summary of Meeting Minutes | Responsible | Comment by ATC |
|---|---|---|---|---|
| 11-02-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | Discussion on the progress of our project | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | 1. Work on finding relevant designs on the project.<br>2. Read research papers to get ideas. |
| 18-02-24 | 1. Fahad<br>2. Adri<br>3. Sahill | 1. Finding optimal design solutions | 1. Fahad<br>2. Adri | |

| | | | | |
|---|---|---|---|---|
| | 4. Shakib | 2. Reading more research papers to get ideas. | | |
| 25-02-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | 1. Selecting multiple designs. | 1. Sahill<br>2. Shakib | |
| 27-02-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | 1.Simulating the designs on simulation software.<br>2. Running the simulations. | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | "Find more compelling complex engineering topic" |
| 01-03-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | 1. Picking optimal two designs from multiple simulation design runs | 1. Fahad<br>2. Shakib | Change the design approach 2 as it was not showing convincing results |
| 07-03-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | 1. Proposing our optimal designs to our ATC committee<br>2. Feedback on our design by ATC committee | 1. Adri<br>2. Sahill<br>3. Fahad | |
| 15-03-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | 1. Reading more research papers to get proper understanding on the relevant project design | 1. Adri<br>2. Sahil | |
| 16-03-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | 1. Running simulations to get optimal design<br>2. Selecting the design with best results | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | |
| 17-03-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | Proposing our new simulation design to the ATC committee | | |
| 20-03-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | Preparing for progress presentation | 1. Adri<br>2. Fahad<br>3. Shakib<br>4. Sahill | |
| 28-03-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | Giving progress presentation in front of the FYDP committee | 1. Adri<br>2. Shakib<br>3. Fahad | |
| 4-04-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | Fixing the technical component specifications | 1. Shakib<br>2. Fahad<br>3. Sahill<br>4. Adri | |

| | | | | |
|---|---|---|---|---|
| 5-04-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | Reading research papers on machine language | 1. Fahad<br>2. Adri<br>3. Shakib | |
| 17-04-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | Reading research papers to get more knowledge on machine language | 1. Shakib<br>2. Fahad<br>3. Adri<br>4. Sahill | |
| 23-04-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | 1. Finding some machine language models<br>2. Looking for relevant dataset | 1. Sahil<br>2. Shakib | |
| 23-04-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | 1. Fixing the dataset and applying machine language models 2. Finding relevant results | 1. Shakib<br>2. Fahad<br>3. Adri<br>4. Sahill | |
| 24-04-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | Preparing the FYDP-D project proposal | 1. Shakib<br>2. Fahad<br>3. Adri<br>4. Sahill | "1. Maintain time limit during Final Presentation<br>2. Group practice for final presentation" |
| 24-04-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | "Preparation for Final Presentation" | 1. Shakib<br>2. Fahad<br>3. Adri<br>4. Sahill | |
| 24-04-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | "Final Presentation in front of the FYDP committee" | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | "Fix the sustainability matrix" |
| 25-04-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | Making necessary fixes on the engineering tools part | 1. Shakib<br>2. Sahill | |
| 25-04-24 | 1. Fahad<br>2. Adri<br>3. Sahill<br>4. Shakib | Submission of FYDP-D project proposal | Adri | |

**Table 17: FYDP (C) Summer 2024 Summary of Team Log Book/ Journal**

| Date/Time/Place | Attendee | Summary of Meeting Minutes | Responsible | Comment by ATC |
|---|---|---|---|---|
| 25-07-2024 | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | FYDP-C, First session attending. | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | Have to order the components fast. |
| 29-07-2024 | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | Ordering the components. | 1. Adri<br>2. Shakib | |
| 19-08-2024 | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | Received and collect all the components | 1. Adri | Check the components. |
| 09-09-2024 | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | Testing the components one by one. | 1. Fahad<br>2. Sahill | |
| 23-09-2024 | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | Using the components making the final project integrating ML in hardware. | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | |
| 07-10-2024 | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | Testing hardware after implementation. | 1. Adri<br>2. Fahad | Improve the code. |
| 14-10-2024 | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | Changing some codes. | 1. Fahad<br>2. Sahill | |
| 14-10-2024 | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | Complete the report writing and final improvements for report writing and finalized the design. | | Improve the prototype design. |

| | | | | |
|---|---|---|---|---|
| 23-10-2024 | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | 1. Report summary and design report improvements for report writing<br>2. Finalized the design report<br>3. Making poster for presentations. | 1. Adri<br>2. Fahad<br>3. Sahill<br>4. Shakib | |

**Code of YOLOv11 with Stereo Vision Algorithm (For design approach 1):**

```
from ultralytics import YOLO
from picamera2 import Picamera2
import cv2
import numpy as np
import os

# COCO class names (for YOLOv8)
coco_names = [
      'person', 'bicycle', 'car', 'motorcycle', 'airplane', 'bus', 'train',
'truck', 'boat',
     'traffic light', 'fire hydrant', 'stop sign', 'parking meter', 'bench',
'bird', 'cat', 'dog',
        'horse', 'sheep', 'cow', 'elephant', 'bear', 'zebra', 'giraffe',
'backpack', 'umbrella', 'handbag',
    'tie', 'suitcase', 'frisbee', 'skis', 'snowboard', 'sports ball', 'kite',
'baseball bat', 'baseball glove',
       'skateboard', 'surfboard', 'tennis racket', 'bottle', 'wine glass',
'cup', 'fork', 'knife', 'spoon',
      'bowl', 'banana', 'apple', 'sandwich', 'orange', 'broccoli', 'carrot',
'hot dog', 'pizza', 'donut',
        'cake', 'chair', 'couch', 'potted plant', 'bed', 'dining table',
'toilet', 'TV', 'laptop', 'mouse',
       'remote', 'keyboard', 'cell phone', 'microwave', 'oven', 'toaster',
'sink', 'refrigerator', 'book',
    'clock', 'vase', 'scissors', 'teddy bear', 'hair drier', 'toothbrush'
]

# Initialize Picamera2 for both left and right cameras
picam2_left = Picamera2(camera_num=0)
```

```
picam2_left.preview_configuration.main.size = (1280, 960)
picam2_left.preview_configuration.main.format = "RGB888"
picam2_left.preview_configuration.align()
picam2_left.configure("preview")
picam2_left.start()

picam2_right = Picamera2(camera_num=1)
picam2_right.preview_configuration.main.size = (1280, 960)
picam2_right.preview_configuration.main.format = "RGB888"
picam2_right.preview_configuration.align()
picam2_right.configure("preview")
picam2_right.start()

# Load YOLOv8 model
model = YOLO('yolo11n.pt')

# Initialize stereo matcher (StereoBM or StereoSGBM)
stereo = cv2.StereoBM_create(numDisparities=16, blockSize=15)

# Stereo vision parameters (adjust based on your setup)
focal_length = 8.378644073612918  # You will need to calibrate your cameras
to get this value
baseline = 12.9       # Distance between the two cameras in meters (adjust
accordingly)

while True:
    # Capture frames from both cameras
    frame_left = picam2_left.capture_array()
    frame_right = picam2_right.capture_array()
```

```
    frame_left = cv2.rotate(frame_left,cv2.ROTATE_90_COUNTERCLOCKWISE)
    frame_right = cv2.rotate(frame_right,cv2.ROTATE_90_COUNTERCLOCKWISE)


    # Run object detection on both frames
    result_right = model(frame_right)

    # Convert frames to grayscale for stereo vision
    gray_left = cv2.cvtColor(frame_left, cv2.COLOR_RGB2GRAY)
    gray_right = cv2.cvtColor(frame_right, cv2.COLOR_RGB2GRAY)

    # Compute the disparity map between the two grayscale images
    disparity = stereo.compute(gray_left, gray_right)

     # Calculate depth for the center of the bounding box of the detected
objects
    for obj in result_right[0].boxes:
            x1, y1, x2, y2 = map(int, obj.xyxy[0])  # Get bounding box
coordinates
        center_x = (x1 + x2) // 2
        center_y = (y1 + y2) // 2

        # Get disparity at the center of the bounding box
        disparity_value = disparity[center_y, center_x]

            if disparity_value > 0:  # Avoid division by zero or negative
disparities
            # Calculate depth in meters
            depth_meters = (focal_length * baseline) / disparity_value
            # Convert depth to centimeters
            depth_cm = depth_meters * 100
```

```
            # Get the class ID and corresponding object name
            class_id = int(obj.cls)
             object_name = coco_names[class_id] if class_id < len(coco_names)
else "Unknown"

            # Prepare the label with object name and distance
            label = f'{object_name}: {depth_cm:.2f} cm'
                    labelPrint  =  f'=========Object  Detected!=======\nObject
Name:{object_name}                                  \nDistance:{depth_cm:.2f}
cm\n=================================='
            print(labelPrint)
            speech = f"{object_name} is detected at {depth_cm:.2f} cm away."
            os.system('espeak -ven-us+f4 -s 145 "'+speech+'"')

            # Draw bounding box and label with object name and distance
            cv2.rectangle(frame_right, (x1, y1), (x2, y2), (0, 255, 0), 2)
                          cv2.putText(frame_right,  label,  (x1,  y1  -  10),
cv2.FONT_HERSHEY_SIMPLEX, 0.5, (255, 0, 0), 2)

     # Display the annotated frame from the left camera with object name and
distance labels
    cv2.imshow("Camera Left - Object Detection and Distance", frame_right)

    # Exit on 'q' key press
    if cv2.waitKey(1) & 0xFF == ord('q'):
        break

# Cleanup
picam2_left.close()
picam2_right.close()
```

```
cv2.destroyAllWindows()
```

**Code of MaskRCNN with Stereo Vision Algorithm (For design approach 2):**

```
import copy
import math
import numpy as np
import cv2
import matplotlib.pyplot as plt
import scipy
import scipy.optimize
import torch
import torchvision
import torchvision.transforms.functional as tvtf
from torchvision.models.detection import
MaskRCNN_ResNet50_FPN_Weights,MaskRCNN_ResNet50_FPN_V2_Weights
# from torchvision.models.quantization import ResNet50_QuantizedWeights
# from torchvision.utils import make_grid
# from torchvision.io import read_image
from pathlib import Path
# from torchvision.utils import draw_bounding_boxes
# from torchvision.utils import draw_segmentation_masks
# from torchvision.utils import make_grid
# from torchvision.io import read_image
# from pathlib import Path

# import stereo_image_utils

weights=MaskRCNN_ResNet50_FPN_V2_Weights.DEFAULT
```

```
model_file_path = "maskrcnn_resnet50_fpn_v2_coco-73cbd019.pth"  # Replace
this with the path to your downloaded model file
model = torchvision.models.detection.maskrcnn_resnet50_fpn_v2(weights=None)
model.load_state_dict(torch.load(model_file_path))
_ = model.eval()
ra1=30
ra2=50
pa1=1620.2266
pa2 = 1243.642
basel = 12.9

fl=ra1-pa2*ra2/pa1

def load_img(filename):
    img = cv2.imread(filename)
    return cv2.cvtColor(img, cv2.COLOR_BGR2RGB)

def preprocess_image(image):
    image = tvtf.to_tensor(image)
    image = image.unsqueeze(dim=0)
    return image

def display_image(image):
    fig, axes = plt.subplots(figsize=(12, 8))

    if image.ndim == 2:
        axes.imshow(image, cmap='gray', vmin=0, vmax=255)
    else:
        axes.imshow(image)

    plt.show()
```

```
def display_image_pair(first_image, second_image):
    #this funciton from Computer vision course notes
    # When using plt.subplots, we can specify how many plottable regions we
want to create through nrows and ncols
    # Here we are creating a subplot with 2 columns and 1 row (i.e.
side-by-side axes)
    # When we do this, axes becomes a list of length 2 (Containing both
plottable axes)
    fig, axes = plt.subplots(nrows=1, ncols=2, figsize=(12, 8))

    # TODO: Call imshow on each of the axes with the first and second images
    #       Make sure you handle both RGB and grayscale images
    if first_image.ndim == 2:
        axes[0].imshow(first_image, cmap='gray', vmin=0, vmax=255)
    else:
        axes[0].imshow(first_image)

    if second_image.ndim == 2:
        axes[1].imshow(second_image, cmap='gray', vmin=0, vmax=255)
    else:
        axes[1].imshow(second_image)

    plt.show()
# these colours are used to draw boxes.

COLOURS = [
    tuple(int(colour_hex.strip('#')[i:i+2], 16) for i in (0, 2, 4))
    for colour_hex in plt.rcParams['axes.prop_cycle'].by_key()['color']
]
```

```
def annotate_class3(img, det, lbls, class_map, conf=None, colours=COLOURS):
    for i, (tlx, tly, brx, bry) in enumerate(det):
        txt = class_map[i]
        if conf is not None:
            txt += f' {conf[i]:.3f}'

        # A box with a border thickness draws half of that thickness to the
left of the
        # boundaries, while filling fills only within the boundaries, so we
expand the filled
        # region to match the border
        offset = 1

        # Rectangle for the text background
        cv2.rectangle(img,
                      (tlx - offset, tly - offset + 12),
                      (tlx - offset + len(txt) * 12, tly),
                      color=colours[i % len(colours)],
                      thickness=cv2.FILLED)

        # Font type
        ff = cv2.FONT_HERSHEY_PLAIN

        # Increase fontScale and thickness for thicker and wider text
        fontScale = 7 # Adjust as needed for width
        thickness = 15    # Adjust as needed for thickness

        # Put text on the image
        cv2.putText(img, txt, (tlx, tly - 1 -30), fontFace=ff,
fontScale=fontScale, color=(255,) * 4, thickness=thickness)
```

```
# Example usage
# annotate_class2(image, detections, labels, class_map, confidences)


def annotate_class2(img, det, lbls, class_map, conf=None, colours=COLOURS):
    for i, (tlx, tly, brx, bry) in enumerate(det):
        txt = class_map[i]
        if conf is not None:
            txt += f' {conf[i]:.3f}'

        # A box with a border thickness draws half of that thickness to the
left of the
        # boundaries, while filling fills only within the boundaries, so we
expand the filled
        # region to match the border
        offset = 1

        # Rectangle for the text background
        cv2.rectangle(img,
                      (tlx - offset, tly - offset + 12),
                      (tlx - offset + len(txt) * 12, tly),
                      color=colours[i % len(colours)],
                      thickness=cv2.FILLED)

        # Font type
        ff = cv2.FONT_HERSHEY_PLAIN

        # Increase fontScale and thickness for thicker and wider text
        fontScale = 7 # Adjust as needed for width
        thickness = 15    # Adjust as needed for thickness

        # Put text on the image
```

```
        cv2.putText(img, txt, (tlx, tly - 1 - 150), fontFace=ff,
fontScale=fontScale, color=(255,) * 4, thickness=thickness)

# Example usage
# annotate_class2(image, detections, labels, class_map, confidences)

def get_tracks(cost):
    return scipy.optimize.linear_sum_assignment(cost)



def get_tracks_ij(cost):
    tracks = scipy.optimize.linear_sum_assignment(cost)
    return [[i,j] for i, j in zip(*tracks)]

def get_cost_with_com(masks, lbls = None,prob_thresh = 0.8):
    alpha = 240; beta  = 10; gamma = 5

    #left masks
    mask_bool = masks[0] > prob_thresh
    mask_bool = mask_bool.squeeze(1)
    #right masks
    mask_bool2 = masks[1] > prob_thresh
    mask_bool2 = mask_bool2.squeeze(1)

    #left params
    #com1 is center of mass of height
    #com2 is center of mass of width
    mask_size = (mask_bool).sum(dim=[1,2])
    mask_com_matrix_1 = torch.tensor(range(mask_bool.shape[1]))
    com1 =
((mask_com_matrix_1.unsqueeze(1))*mask_bool).sum(dim=[1,2])/mask_size
```

```
    mask_com_matrix_2 = torch.tensor(range(mask_bool.shape[2]))
    com2 =
((mask_com_matrix_2.unsqueeze(0))*mask_bool).sum(dim=[1,2])/mask_size

    left_params = torch.stack((com1, com2, mask_size)).transpose(1,0)

    #get right params
    mask_size2 = (mask_bool2).sum(dim=[1,2])
    mask_com_matrix_12 = torch.tensor(range(mask_bool2.shape[1]))
    com12 =
((mask_com_matrix_12.unsqueeze(1))*mask_bool2).sum(dim=[1,2])/mask_size2
    mask_com_matrix_22 = torch.tensor(range(mask_bool2.shape[2]))
    com22 =
((mask_com_matrix_22.unsqueeze(0))*mask_bool2).sum(dim=[1,2])/mask_size2

    right_params = torch.stack((com12, com22, mask_size2)).transpose(1,0)

    #calculate cost function
    cost = (left_params[:,None] - right_params[None])
    #scale counts
    cost[:,:,2]=abs(cost[:,:,2])/alpha

    #can't move right, can only move left
    cost[cost[:,:,1]<0] = beta*abs(cost[cost[:,:,1]<0])

    #move up and down, take abs vals
    cost[:,:,0] = gamma*abs(cost[:,:,0])
    # print(cost.shape)
    cost = cost.sum(dim=2)
    if lbls is not None:
        for i in range(cost.shape[0]):
```

```
            for j in range(cost.shape[1]):
                if (lbls[0][i]!=lbls[1][j]):
                    cost[i,j]+=100
#                     print(lbls[0][i], lbls[1][j])
    return cost



def get_horiz_dist(masks, prob_thresh = 0.7):
    # gets the horizontal distance between the centre of mass for each object
    #left masks
    mask_bool = masks[0] > prob_thresh
    mask_bool = mask_bool.squeeze(1)
    #right masks
    mask_bool2 = masks[1] > prob_thresh
    mask_bool2 = mask_bool2.squeeze(1)

    #left params
    #com1 is center of mass of height
    #com2 is center of mass of width
    mask_size = (mask_bool).sum(dim=[1,2])
    mask_com_matrix_1 = torch.tensor(range(mask_bool.shape[1]))
    com1 =
((mask_com_matrix_1.unsqueeze(1))*mask_bool).sum(dim=[1,2])/mask_size
    mask_com_matrix_2 = torch.tensor(range(mask_bool.shape[2]))
    com2 =
((mask_com_matrix_2.unsqueeze(0))*mask_bool).sum(dim=[1,2])/mask_size

    left_params = torch.stack((com1, com2, mask_size)).transpose(1,0)

    #get right params
    mask_size2 = (mask_bool2).sum(dim=[1,2])
```

```
    mask_com_matrix_12 = torch.tensor(range(mask_bool2.shape[1]))
    com12 =
((mask_com_matrix_12.unsqueeze(1))*mask_bool2).sum(dim=[1,2])/mask_size2
    mask_com_matrix_22 = torch.tensor(range(mask_bool2.shape[2]))
    com22 =
((mask_com_matrix_22.unsqueeze(0))*mask_bool2).sum(dim=[1,2])/mask_size2

    right_params = torch.stack((com12, com22, mask_size2)).transpose(1,0)

    #calculate cost function
    cost = (left_params[:,None] - right_params[None])
    return cost[:,:,1]

#create the tracking cost function.
#consists of theree parts.
#  1. The vertical move up and down of object centre of mass. Scale this up
because we do not expect this to be very much.
#  2. The move left or right by the object. We only expect it to move right
(from the left eye image). So penalise if it moves left.
#  3. The difference in area of pixels. Area of image is width x height, so
divide by height, there for this will have max value of width

def get_cost(boxes, lbls = None, sz1 = 400):
    alpha = sz1; beta  = 10; gamma = 5

    #vertical_dist, scale by gamma since can't move up or down
    vert_dist = gamma*abs(get_vertic_dist_centre(boxes))

    #horizonatl distance.
    horiz_dist = get_horiz_dist_centre(boxes)
```

```
    #increase cost if object has moved from right to left.
    horiz_dist[horiz_dist<0] = beta*abs(horiz_dist[horiz_dist<0])

    #area of box
    area_diffs = get_area_diffs(boxes)/alpha

    cost = np.array([vert_dist,horiz_dist,area_diffs])

    cost=cost.sum(axis=0)

    #add penalty term for different object classes
    if lbls is not None:
        for i in range(cost.shape[0]):
            for j in range(cost.shape[1]):
                if (lbls[0][i]!=lbls[1][j]):
                    cost[i,j]+=150
    return cost

## get distance bentween corner and centre

# centre = sz1/2

def get_dist_to_centre_tl(box, cntr):
    pnts = np.array(tlbr_to_corner(box))[:,0]
    return abs(pnts - cntr)



def get_dist_to_centre_br(box, cntr):
    pnts = np.array(tlbr_to_corner_br(box))[:,0]
    return abs(pnts - cntr)
```

```
#get all distances from every object box to every other object box
#left image is boxes[0]
#right image is boxes[1]

#do broad casting.
#in python, col vector - row vector gives matrix:
# [a] - [c,d] = [a-c, a-d]
# [b]           [b-c, b-d]

def get_horiz_dist_centre(boxes):
    pnts1 = np.array(tlbr_to_center1(boxes[0]))[:,0]
    pnts2 = np.array(tlbr_to_center1(boxes[1]))[:,0]
    return pnts1[:,None] - pnts2[None]

def get_horiz_dist_corner_tl(boxes):
    pnts1 = np.array(tlbr_to_corner(boxes[0]))[:,0]
    pnts2 = np.array(tlbr_to_corner(boxes[1]))[:,0]
    return pnts1[:,None] - pnts2[None]

def get_horiz_dist_corner_br(boxes):
    pnts1 = np.array(tlbr_to_corner_br(boxes[0]))[:,0]
    pnts2 = np.array(tlbr_to_corner_br(boxes[1]))[:,0]
    return pnts1[:,None] - pnts2[None]

def get_vertic_dist_centre(boxes):
    pnts1 = np.array(tlbr_to_center1(boxes[0]))[:,1]
    pnts2 = np.array(tlbr_to_center1(boxes[1]))[:,1]
    return pnts1[:,None] - pnts2[None]

def get_area_diffs(boxes):
    pnts1 = np.array(tlbr_to_area(boxes[0]))
```

```
    pnts2 = np.array(tlbr_to_area(boxes[1]))
    return abs(pnts1[:,None] - pnts2[None])
#get centr, top left and bottom right of boxes

def tlbr_to_center1(boxes):
    points = []
    for tlx, tly, brx, bry in boxes:
        cx = (tlx+brx)/2
        cy = (tly+bry)/2
        points.append([cx, cy])
    return points

def tlbr_to_corner(boxes):
    points = []
    for tlx, tly, brx, bry in boxes:
        cx = (tlx+tlx)/2
        cy = (tly+tly)/2
        points.append((cx, cy))
    return points

def tlbr_to_corner_br(boxes):
    points = []
    for tlx, tly, brx, bry in boxes:
        cx = (brx+brx)/2
        cy = (bry+bry)/2
        points.append((cx, cy))
    return points

def tlbr_to_area(boxes):
    areas = []
    for tlx, tly, brx, bry in boxes:
```

```
        cx = (brx-tlx)
        cy = (bry-tly)
        areas.append(abs(cx*cy))
    return areas

# this functions returns the detections
# det is the boxes, top left and bottom right cooridinates
# lbls are the class labels
# scores are the confidence. We use 0.5 as default
# masks are the segmentation masks.

def get_detections(maskrcnn, imgs, score_threshold=0.9): #person, dog,
elephan, zebra, giraffe, toilet
    ''' Runs maskrcnn over all frames in vid, storing the detections '''
    # Record how long the video is (in frames)
    det = []
    lbls = []
    scores = []
    masks = []

    for img in imgs:
        with torch.no_grad():
            result = maskrcnn(preprocess_image(img))[0]

        mask = result["scores"] > score_threshold

        boxes = result["boxes"][mask].detach().cpu().numpy()
        det.append(boxes)
        lbls.append(result["labels"][mask].detach().cpu().numpy())
        scores.append(result["scores"][mask].detach().cpu().numpy())
#         masks.append(result["masks"][mask].detach().cpu().numpy())
```

```
        masks.append(result["masks"][mask]) #I want this as a tensor

    # det is bounding boxes, lbls is class labels, scores are confidences and
masks are segmentation masks
    return det, lbls, scores, masks

#det[0] are the bounding boxes in the left image
#det[1] are the bounding boxes in the right image
def draw_detections(img, det, colours=COLOURS, obj_order = None):
    for i, (tlx, tly, brx, bry) in enumerate(det):
        if obj_order is not None and len(obj_order) < i:
            i = obj_order[i]
        i %= len(colours)
        c = colours[i]

        cv2.rectangle(img, (tlx, tly), (brx, bry), color=colours[i],
thickness=35)
def annotate_class(img, det, lbls, conf=None, colours=COLOURS,
class_map=weights.meta["categories"]):
    for i, ( tlx, tly, brx, bry) in enumerate(det):
        txt = class_map[lbls[i]]
        if conf is not None:
            txt += f' {conf[i]:1.3f}'
        # A box with a border thickness draws half of that thickness to the
left of the
        # boundaries, while filling fills only within the boundaries, so we
expand the filled
        # region to match the border
        offset = 1

        cv2.rectangle(img,
```

```
                      (tlx-offset, tly-offset+12),
                      (tlx-offset+len(txt)*12, tly),
                      color=colours[i%len(colours)],
                      thickness=cv2.FILLED)

        ff = cv2.FONT_HERSHEY_PLAIN
        cv2.putText(img, txt, (tlx, tly-1+12), fontFace=ff, fontScale=1.0,
color=(255,)*3)

while True:
    !rpicam-jpeg --camera 1 --timeout 2 --output
"/home/Project/Project/Images/left501.jpg"

    !rpicam-jpeg --camera 0 --timeout 2 --output
"/home/Project/Project/Images/right501.jpg"



    left_eye= "/home/Project/Project/Images/left501.jpg"
    right_eye="/home/Project/Project/Images/right501.jpg"

    #down sample image to get same size as expected from esp32 cam



    left_img = load_img(left_eye)
    # left_img = cv2.resize(left_img, dsize=(sz1,sz2),
interpolation=cv2.INTER_LINEAR)

    right_img = load_img(right_eye)
    # right_img = cv2.resize(right_img, dsize=(sz1,sz2),
interpolation=cv2.INTER_LINEAR)
```

```
    sz1 = right_img.shape[1]
    sz2 = right_img.shape[0]

    tantheta = (1/(ra2-fl))*(basel/2)*sz1/pa2
    # display_image_pair(left_img, right_img)

    imgs = [left_img, right_img]

    left_right = [preprocess_image(d).squeeze(dim=0) for d in imgs]
    centre = sz1/2

    det, lbls, scores, masks = get_detections(model,imgs)

    tmp1 = get_dist_to_centre_br(det[0],centre)
    tmp2 = get_dist_to_centre_br(det[1],centre)

    cost = get_cost(det, lbls = lbls)

    dists_tl =  get_horiz_dist_corner_tl(det)
    dists_br =  get_horiz_dist_corner_br(det)

    final_dists = []
    dctl = get_dist_to_centre_tl(det[0],centre)
    dcbr = get_dist_to_centre_br(det[0],centre)
    tracks = scipy.optimize.linear_sum_assignment(cost)
    for i, j in zip(*tracks):
        if dctl[i] < dcbr[i]:

final_dists.append((dists_tl[i][j],np.array(weights.meta["categories"])[lbls[
0]][i]))
```

```
        else:

final_dists.append((dists_br[i][j],np.array(weights.meta["categories"])[lbls[
0]][i]))



    h_d =
[[np.array(weights.meta["categories"])[lbls[0]][i],np.array(weights.meta["cat
egories"])[lbls[1]][j]] for i, j in zip(*tracks)]

    fd = [i for (i,j) in final_dists]

    dists_away = (basel/2)*sz1*(1/tantheta)/np.array(fd)+fl

    cat_dist = []
    for i in range(len(dists_away)):
        cat_dist.append(f"{np.array(weights.meta['categories'])[lbls[0]][i]}
{dists_away[i]:.1f}cm")
        print(f'{np.array(weights.meta["categories"])[lbls[0]][i]} is
{dists_away[i]:.1f}cm away')

    tt=np.round(dists_away, 2)
    t=str(tt)

    fig, axes = plt.subplots(1, 2, figsize=(13, 8))

    t1 = [list(tracks[1]), list(tracks[0])]

    dislist=[]
    for i in range(len(dists_away)):
```

```
        dislist.append(str(np.round(dists_away[i],2))+"cm")

    for i, imgi in enumerate(imgs):
        img = imgi.copy()
        deti = det[i].astype(np.int32)
        draw_detections(img,deti[list(tracks[i])], obj_order=list(t1[i]))

annotate_class2(img,deti[list(tracks[i])],lbls[i][list(tracks[i])],np.array(w
eights.meta['categories'])[lbls[0]])

annotate_class3(img,deti[list(tracks[i])],lbls[i][list(tracks[i])],dislist)
        axes[i].imshow(img)
        axes[i].axis('off')
        if i==0:
            axes[i].set_title("Left Camera Image")
        else:
            axes[i].set_title("Right Camera Image")

    fig.tight_layout()
    # Do the plot code
    fig.savefig('finaloutput.svg', format='svg', dpi=300)
    plt.savefig("output.jpg")
    plt.show()
```

**Primary data collection sheet**

https://docs.google.com/spreadsheets/d/1B8UQij-P8OMdZkX_DBIsYlk-QOsU5UTeVjuHJZ7VBtg/edit?usp=sharing