NAVIGATION SYSTEM FOR VISUALLY IMPAIRED PEOPLE USING IMAGE PROCESSING AND AUGMENTED REALITY

Sushma Dube
sushmasinnor@gmail.com

Vaishnavi Chavan
chavanvaishnavi174@gmail.com

Soundarya A. N.
soundarya11naganahalli@gmail.com

Under guidance of

Sayyada Fahmeeda Sultana
sayyadafahmeeda@pdaengg@gmail.com

ABSTRACT

Vision is most important part of human physiology as 83% of information human being gets from the environment is via sight. Among all of the human senses, vision is one of the most important and it plays an important role in comprehending the surrounding environment. It is difficult for visually impaired people to move around outside without assistance. The proposed project aims to create an system for people with visually impaired, which will help them to detect the hurdles and guides to navigate around the unfamiliar environments. The concept of Indoor Navigation using a smartphone- based Augmented Reality technology is explored in this proposed project. The proposed project built an Augmented Reality-based framework to assist users for indoor navigation using ARWAY, a software development toolkit. The project is implemented using ARWAY augmented reality algorithm and SSD in image processing for object detection, CNN for Object recognition. The proposed project creating and deploying an Android application that uses the phone's camera to detect things in the environment of a visually impaired person by using Image processing and Augmented Reality techniques to facilitate visually impaired people for indoor navigation with single or multiple objects as hurdles. Using an audio device such as the phone's speaker, the application will inform the visually impaired about the object's name, direction, distance between that object and the user.

CHAPTER 1

1 INTRODUCTION

Visually Impaired people, at some point in their life, may find difficult to navigate. Applications for indoor navigation for mobile devices are now popular and people need them to detect hurdles within rooms and with many other places. Several indoor navigation applications use multiple techniques, such as Wi-Fi Fingerprinting and Bluetooth to make this possible. In order to direct the users to navigate and to detect hurdles, these applications use image processing and Convolutional neural network technologies.

The proposed project is being implement on an android app, that detects diverse objects in live video feed along with a real-time text reader. In proposed project, an object recognition android app was developed used tensor flow, object detection API model, which implemented using SSD algorithm and real-time text reader feature which is using google. TTS engine and google played services mobile vision API which describes the use of text recognizer class to detect text from a real-time video feed. SSD algorithm for object detection model is to use for real-time and offline object detection. The program will notify the visually impaired user of the object's name, direction, distance between user and the object, using an audio device such as headphones or the phone's speaker.

1.1 BACKGROUND

The advancement of indoor positioning and navigation systems and solutions is a hot subject, owing to its role in future industrial approaches, among other things. Large companies have numerous advanced systems for indoor localization in most of their planned technology, architectural designs, and patterns, besides those that have already upgraded their applications, or an application programme interface (API, which is a series of routines, protocols, and resources for building software applications), according to the research on this subject.

In outdoor settings, navigation systems are commonly used, but indoor navigation systems are still in the early stages of growth. GPS, Bluetooth, Wi-Fi, RFID and Sensor Chip technologies make use of the latest available devices. The technology for Bluetooth is affordable but has limited range and accuracy. We need to set up Bluetooth hotspots in the building for efficient position monitoring and navigation using this technology. For other technologies, such as Wi-Fi, RFID and Sensor chips, the same applies. The expense of installation and repair is outrageously high with all these tools, and a shift in weather conditions will affect the signal intensity of the hotspot if one system breaks down.

Existing infrastructural services are inevitable for the perfect operation of all of these current structures. In the proposed project, by implementing an augmented reality-based indoor navigation technology to help people navigate in indoor environments, put forward a relatively inexpensive, reliable and efficient approach to achieve the same objective.

Usually, an augmented reality navigation system work in the following way:

- Obtain the real-world experience from the viewpoint of the user.
- Generate knowledge about the virtual world based on the real-world view and information about the venue.
- Register the visual knowledge created with the real-world view and present the information to the user, thus providing an augmented reality.

Augmented Reality (AR) is a technology that superimposes digital information, such as images, videos or 3D models, onto the real-world environment. This is typically achieved through devices like smartphones, smart glasses, or headsets. AR enhances the user's perception of the real world by overlaying digital content, allowing

for an interactive and immersive experience. Applications range from gaming and education to navigation systems, where AR can provide additional context, information, or guidance in real-time.

1.2 PROBLEM DEFINITION

The problem at hand revolves around the limited mobility and independence of visually impaired individuals in navigating unfamiliar environments. Traditional mobility aids often fall short in providing real-time, context-aware information about obstacles and points of interest. This creates challenges and safety concerns for individuals with visual impairments, impacting their ability to confidently explore new surroundings.

To address this, the proposed project seeks to overcome the limitations by combining augmented reality (AR), image processing and convolutional neural networks (CNN). The challenge lies in developing a robust and efficient solution that can seamlessly identify and interpret the visual environment in real-time using CNN, and then present this information in an accessible and meaningful way through AR overlays. The system must navigate complex scenarios, adapting to dynamic changes in the environment and providing accurate, timely feedback to the user.

Key aspects of the problem include designing an intuitive user interface that caters to diverse user preferences, ensuring the system is responsive to real-world challenges, and optimizing the integration of AR and CNN technologies for a reliable and user-friendly navigation experience. The solution aims to enhance the independence, safety, and overall quality of life for visually impaired individuals, addressing a critical need for more advanced and adaptive assistive technologies in this domain.

After implementing the proposed project, it was expected for a speech feedback for the object which was being detected. But the same object was getting called out multiple times as it got detected. But it will be undesirable to speak out the same object name even if the detection result is the same. Also, it was undesirable if two object names spoken are overlapping or very closely that the user would not be able to distinguish.

h. To solve this problem, if one object was getting detected in the first frame and was speaking out. Then the program will not speak out its class for next five seconds, even if it gets detected. By this, the problem of detecting a single object multiple times was being solved. Upon testing on several objects it found that the results may vary sometimes and the accuracy for detecting objects depends on several parameters. In order to improve the accuracy, the model needs to be trained by using objects under different scenarios and test cases.

1.3 EXISTING METHODS

There are two existing indoor mapping solutions, which are Bluetooth Beacons-based indoor mapping and RSSI-based Wi-Fi Fingerprinting indoor mapping. The Bluetooth Beacons are transmitters for hardware (Satan et.al. 2018). They are low energy Bluetooth systems that broadcast their identifier to neighbouring portable electronic devices. When they are close to the beacons, the hardware in these beacons enables laptops, smartphones, and

other gadgets to perform acts. Bluetooth Beacons use Bluetooth low energy proximity sensing to relay a globally specific identifier. A compliant program or an operating system then takes up these identifiers. Several bytes are also sent along with identifiers that can be used to decide the physical location of the device, monitor clients, or can also activate a location-based device behaviour, such as a social media check-in or a route warning. Another current method is Wi-Fi Fingerprinting using RSSI. The acronym for Obtained Signal Intensity Indicator is RSSI (Chabbar and Chami et.al. 2017).

They are generally invisible to the naked eyes. Using RSSI, the power present inside the received radio signal may be determined. Standard printing of fingers is also based on RSSI, although there is a minor distinction. In conventional fingerprinting, there are many access points and it was clearly focused on capturing the signal intensity from these access points, which are in range. This information is then stored in a database along with the known coordinates of the client computer in an offline process. This knowledge that is preserved can be probabilistic or deterministic.

For the distribution and implementation of Wi-Fi routers and Bluetooth beacons in the indoor environment. Some application systems often have high latency during the detection process, such as Bluetooth and infrared approaches. Since these Systems used for indoor positioning can be usually divided into two categories, wireless communication methods and computer vision methods. Technologies such as Ultra-wide Band (UWB), Wireless Local Area Networks (WLAN), and Radio Frequency Identification (RFID) are used in wireless communication methods to localize a unit.

These technologies also require physical infrastructures, such as technologies that are common solutions for localization, have difficulty estimating the direction of the user, and are thus not suitable for AR applications. In contrast, machine vision methods are more suited for AR-based applications, and recent tests have shown computer vision solutions to be more reliable in addition to Wi-Fi based fingerprinting. A widely studied indoor positioning solution focused on vision includes picture recognition from live camera feed of the real world.

1.4 OBJECTIVE

The proposed project is about proposing a methodology for navigating unfamiliar environments for visually impaired people using an augmented reality (AR) application. For this, the software is designed, which is using image processing based Augmented Reality Tool Kit and SDK based on the real-world view direction to quickly detect hurdle and make ease to navigate.

ARWAY is used because of its efficiency, ease to use, scalability which removes the hardware dependency and leverages mobile cameras to safely interpret and navigate the GPS-denied locations in indoors. ARWAY uses the Mapping Engine, which captures unique ‘feature points’ in each camera keyframe and then by matching those features continuously in each frame, it can estimate the camera position in real-time. The proposed project create AR based indoor navigation application using ARWAY (a software development toolkit), which is used because of its efficiency, ease to use, scalability which removes the hardware dependency and leverages mobile cameras to safely interpret and navigate the GPS denied locations in indoors.

The primary objectives of developing a navigation system for visually impaired individuals revolve around enhancing their mobility and independence. The proposed project aims to utilize CNN for real-time object recognition, identifying obstacles and landmarks in the user's environment. By integrating augmented reality, the project aims to provide intuitive and context-aware navigational guidance through auditory or haptic feedback, ensuring a seamless and user-friendly experience.

The overall goal is -

- To empower visually impaired individuals with a reliable and adaptive tool that enables them to navigate unfamiliar environments, faltering a greater sense of autonomy and safety in their daily lives.
- To reduce the use of multiple devices for object detection and object recognition.
- To accomplished by providing blind people with commands about the things and movements which should be taken all over surroundings.

1.5 PROPOSED SYSTEM

The proposed project is being implement on an android app, that detects diverse objects in live video feed along with a real-time text reader. In proposed project, an object recognition android app was developed used tensor flow, object detection API model, which implemented using SSD algorithm and real-time text reader feature which is using googles. TTS engine and google played services mobile vision API which describes the use of text recognizer class to detect text from a real-time video feed. SSD algorithm for object detection model is to use for real-time and offline object detection. The program will notify the visually impaired user of the object's name, colour, direction, distance, speed, and steps to approach that thing or destiny, using an audio device such as headphones or the phone's speaker.

MobileNet-SSD combines two main components: MobileNet, which is a lightweight convolutional neural network (CNN), and the Single Shot MultiBox Detector (SSD) architecture for object detection. Let's break down each component:

- MobileNet: MobileNet is a CNN architecture designed for mobile and embedded devices. It's characterized by its depth-wise separable convolutions, which significantly reduce the number of parameters and computational cost while maintaining reasonable accuracy. Traditional convolutions involve both spatial convolutions (across channels) and depth-wise convolutions (across spatial dimensions). In MobileNet, these are separated, resulting in fewer computations. This makes it suitable for deployment on devices with limited computational resources like smartphones and IoT devices.
- Single Shot MultiBox Detector (SSD) : SSD is a popular object detection architecture known for its efficiency and accuracy. It performs object detection in a single pass through the network, hence the name "single shot." SSD divides the input image into a grid of cells and predicts bounding boxes and class probabilities for objects within each cell. It uses multiple convolutional feature maps at different scales to detect objects of various sizes.

- The proposed project combines MobileNet with SSD (Single Shot MultiBox Detector) for real-time object detection in a navigation system designed for visually impaired individuals, MobileNet-SSD achieves both efficiency and accuracy in object detection tasks, making it suitable for real-time applications on mobile and embedded devices.

The proposed project aims to create an innovative navigation solution for visually impaired individuals by integrating augmented reality (AR), Image processing and convolutional neural networks (CNN). Leveraging the power of CNN, the system will perform real-time object recognition using the camera input, identifying obstacles, landmarks, and other relevant elements in the user's surroundings. This data will then be processed to create a spatial understanding of the environment. This augmented view will serve as a navigational guide, providing intuitive and context-aware feedback to the user. Auditory or haptic cues will be employed to convey information, ensuring accessibility for individuals with visual impairments. The system's user interface will be designed for simplicity and ease of use, with customizable settings to accommodate individual preferences. The navigation system will be equipped to offer route planning, real-time updates, and dynamic adjustments based on the user's movements and changing environmental conditions.

Advantages :

- Helps to ease the difficulty of performing the task of object identification and motion detection.
- Provide a portable solution to the problem of object and motion detection without the need to carry any additional devices dedicated to the same

The goal of the proposed project is to empower visually impaired individuals with a reliable, efficient, and user-friendly tool that significantly improves their mobility, independence, and overall quality of life. The fusion of AR, Image processing and CNN technologies in this navigation system represents a cutting-edge approach to addressing real-world challenges faced by individuals with visual impairments.

CHAPTER 2

LITERATURE SURVEY

When compared to the outdoor positioning, indoor positioning faces multiple problems and has specific requirement. Satellite signals such as GPS are more difficult to track in indoors because radio frequency signals are attenuated by barriers such as walls in a building. In comparison, indoor positioning is commonly associated with more specific locations than outdoor positioning (Mautz, 2009). For e.g., GPS has an accuracy of 10-20 metres measured as a root mean square error (Dardari et al., 2015), and on the other hand some indoor positioning systems can achieve accuracies of few centimetres only. In addition, indoor positioning systems besides the planar location, are usually involved in the user's altitude which indicates both discrete details such as the floor of the building, as well as the exact height of the unit from the actual floor of the user. If consumer mobile devices are used for indoor real-time positioning systems, then the positioning algorithm's efficiency, speed, accuracy, and computational load play a crucial role in the system's functionality and flexibility.

2.1 Literature review on indoor navigation systems

Types of Indoor Positioning Methods

Broadly, there are various methods developed by the researchers as methods for indoor positioning which have been discussed in this chapter.

2.1.1 Radio Frequency (RF)

Cellular towers and WI-FI are used today by mobile positioning systems to help position GPS which is also known as assisted GPS or AGPS (Filonenko et al., 2013). When compared to using GPS alone, this method increases the precision and efficiency of indoor positioning, but it is still inadequate for realistic indoor positioning use.

As reference points, Wi-Fi connection points or Bluetooth beacons may be used. Indoor positioning accuracy based on RF signal intensity is usually in the range of a few metres. The radio frequency based indoor positioning system under consideration in this research are Wi-Fi and Bluetooth/BLE which are discussed in detail below along with advantages and disadvantages.

2.1.2 Wi-Fi Access Points

The use of WIFI access points as reference points for indoor positioning has the significant advantage of having no extra equipment. For connectivity as well as localization, the access points may be used simultaneously. Bielsa et al. (2018) examined how well an indoor positioning system can work where publicly accessible WIFI access

points are used without changing their standard communication protocols in which 60 GHz range millimetre waves were used in an office setting with a combination of human traffic. Related meter-level accuracies can be accomplished with commercial WIFI infrastructures, according to other studies. Martin et al. (2010) reached an accuracy of 1.5 metres using fingerprinting, thus showing that both the offline and the online process can be carried out with a smartphone. Husen and Lee (2014) reached an accuracy of 1.8 metres, while still recording the incredibly rugged orientation of the user (one of four main facing directions).

Advantages:

There are various advantages of Wi-Fi Indoor Positioning which are listed below.

- It provides high positioning accuracy, especially in over-crowded places
- Fast speed
- Surrounding WIFI can be positioned even if it is not connected.

Disadvantages:

There are various disadvantages of Wi-Fi Indoor Positioning which are listed below.

- WIFI dependency – It cannot be located without opening WIFI
- It is must be in a networked state.

2.1.3 Bluetooth

Bluetooth Beacons that transmit signals are lightweight, battery-powered devices that broadcast Bluetooth signals. This signal can carry information which includes the ID, its calibrated strength at a known distance, and a URL or an arbitrary string connection, depending on the protocol. A beacon can be left unattached to something in a location, or connected either by screws or by using the sticker of the unit. If required, a beacon may be placed upside down on a ceiling or concealed from view. Subedi and Pyun (2017) suggested a fingerprinting indoor positioning device based on BLE beacons. Additionally, the authors used a technique in the method called Weighted Centroid Localization (WCL) which is equivalent to multilateration. The authors were able to minimise the required number of deployed BLE beacons by eliminating noisy RSS values and using both fingerprinting and Weighted Centroid Localization approaches at the same time to achieve the same degree of precision compared to the case where only fingerprinting was used. Furthermore, the offline process needs less effort for this hybrid strategy. This machine was able to obtain an accuracy of 1–2 metres in a hallway and an office room, where the BLE beacons were fitted to the ceiling, tests were carried out.

Advantages:

There are various advantages of Bluetooth based Indoor-Positioning which are listed below.

- It is cost-effective, unobtrusive hardware
- It has low energy consumption

- It is flexible integration into the existing infrastructure (battery-powered or power supply via lamps and the domestic electrical system)
- It works where other positioning techniques do not have a signal
- It is compatible with iOS and Android
- It has high accuracy compared to Wi-Fi (up to 1m)

Disadvantages:

There are various disadvantages of Bluetooth based Indoor-Positioning which are listed below.

- It costs additional hardware
- An application is required for client-based solutions
- It has relatively small range (up to 30m)

2.1.4 Optical wireless positioning

Optical wireless communication (or optical free-space communication) is a series of techniques for exchanging data wirelessly using infrared or visible light. Visual Light Communication (VLC) is a subset of wireless optical communication in which data is conveyed through visible light. Visible light has larger frequencies than radio waves and higher bitrates can theoretically be obtained. Because of the ubiquity of indoor light sources and because optical light does not suffer as badly from interference as radio waves, VLC is an enticing solution to Wi-Fi data sharing in the future and an active field of study. (2017, Kaushal et al.)

Advantages:

Optical Wireless Indoor Positioning has a number of benefits, which are described below.

- In general, an optical and vision-based positioning system makes use of a mobile device's sensor and computing power. Inertial sensors, such as accelerometers, gyroscopes, and magnetometers, are integrated into modern electronic devices. As a result, infrastructure installation is reduced (Möller et al., 2012).
- Furthermore, as opposed to other positioning systems, this system is considerably less expensive (Mulloni et al., 2011).

Disadvantages:

Optical Wireless Indoor Positioning has a number of drawbacks, which are described below.

The device has a poor level of precision.

- The device is subjected to interference from a variety of sources, including bright light and motion blur, as well as major accumulative errors that can result in bad results (Klopschitz et al., 2010; Möller et al., 2012).
- Since a server saves location information for monitoring and navigation purposes, privacy issues can arise in certain situations.

2.1.5 Computer vision

Rosenfeld (1988) defines computer vision (CV) as a set of techniques that infer information about a scene by analysing images of that scene. A computer vision algorithm takes a 2D image as an input and produces a description as an output. Often, the goal of this process is to recognize objects from the image. Rosenfeld (1988) describes the general process of 2D object recognition as follows. The process of detecting specific parts of an image is known as segmentation, or feature recognition.

2.1.6 Discrete positioning with computer vision

Mulloni et al. (2009) illustrated how to use mobile cameras for indoor positioning in a conference guide implementation in 2009. The authors attached rectangle-shaped contractual markers to signposts in the conference grounds (similar in purpose and appearance to QR codes). Every marker had a specific visual ID associated with the signpost's position. The application was able to determine the 3D location and direction of the user with regard to the markers in real time with centimetre-level precision. The user had to point the camera directly at the marker, and they had to be close enough to the signpost so that the marker was visible to the came in necessary detail.

However, given a marker's known real-world scale and how perspective projects it onto a 2D image, the camera's relative direction and orientation to the marker can be easily determined using linear algebra (Fusco and Coughlan, 2018). However, the output knowledge from this method is discrete in the sense that it defines which signpost the consumer is looking at. This technique cannot place the viewer if there are no markers in the camera's vision. As a consequence, the question is whether machine vision can offer continuous indoor navigation.

Continuous positioning with computer vision

Fusco and Coughlan (2018) created an indoor visual positioning system utilising this approach to support visually disabled individuals (the system could position the mobile device, but not yet navigate it). Visual-inertial odometry (VIO) was used by their device to place the user in between visual signs. With regard to the environment, augmented reality technologies will place the mobile device locally with a high update rate.

Indoor Positioning is also possible without impromptu visual markings. A typical computer vision technique for scene recognition is to catch and use 3D point clouds. A camera takes photographs of a 3D scene in normal or regular conditions in a passive point cloud acquisition process.

A computer vision algorithm is able to capture a range of 3D positions on the surfaces of objects in the picture, either by providing several views or by rotating the camera in the scene (structure from motion), and the current camera poses in reference to these locations. It is possible to recreate the geometry of the inner space in such a way. If the same indoor space has been scanned before, the newly scanned and previously scanned point clouds can be matched to approximate the latest camera pose in real-world indoor coordinates. In addition to the location, the orientation of the client computer can be calculated. (2016 Weinmann)

Using computer vision, a number of techniques for localization and navigation have been developed. Simultaneous Localization and Mapping is a common technique for autonomous vehicles that originated from the robotics community (2015, Gerstweiler, Vonach and H. Kaufmann). The Simultaneous Localization and Mapping method aims to gather spatial data from the atmosphere such as Signal Intensity and 3D Point Clouds in order to create a global reference map when monitoring the subject's location (2006, Bailey and Whyte). Simultaneous Localization and Mapping algorithms are used in a range of applications, including Wi-Fi, Bluetooth, feature detection, and image recognition (2015, Gerstweiler, Vonach and H. Kaufmann). For Simultaneous Localization and Mapping, either of these data types may be used.

Any of the innovations mentioned above has its own set of advantages and disadvantages. 2D positioning systems (GPS, beacons, ceiling antennas) are successful, but they have a long latency, poor accuracy, and no height detail. It's difficult to scale markers and visual recognition, and it necessitates user interaction.

2.2 Literature Review on Alternate Indoor Positioning methods

The optical and vision-based positioning system uses a mobile sensor or camera in a user's mobile device to assess the location of an individual or an object within a building by locating a marker or image that is within range (Mautz & Tilch, 2011; Klopschitz, Schall, Schmalstieg, & Reitmayr, 2010). A marker is a static target with markings that can be used as a reference within the field of view of an imaging sensor like a cell camera (Mautz, 2012). Barcodes, QR codes, and fiducials are only a few examples of identifiers. Marker-based and Augmented

Reality (Augmented Reality) are the two most popular methods for optical and vision-based positioning.

According to Chang, Tsai, Chang, and Wang (2007), Mulloni, Wagner, Schmalstieg, and Barakonyi (2009), and Raj, Tolety, and Immaculate (2009), a cell phone camera gets visual information using identifiers, such as QR codes (2013). A mobile computer with a monitor, a QR code, and a server make up the machine (Raj et al., 2013). The mobile device's camera is used to collect data by scanning the QR code's pattern, while the server is used for tracking and storing information including floor plan map data for retrieval as appropriate (Barberis et al., 2014; Chang et al., 2007; Mulloni et al., 2009).

While Chang et al. (2007) focused on monitoring people with neurological impairments in smart settings, Mulloni et al. (2009) focused on searching and reviewing real-time location details in an area to aid continuous navigation. To put it another way, Chang et al. (2007)'s research would not allow for real-time navigation like Mulloni et al. (2009)'s, but it can monitor users' movements over time. Scanning the QR code in the case of Raj et al. (2013) yields the floor map URL and geo-location information. As a consequence, the building's floor map is obtained and used for navigation (Raj et al., 2013).

However, in this situation, navigation is not real-time. In comparison to the previous positioning systems mentioned, the QR code's ease of use makes it a feasible indoor positioning system. It is simple to deploy, according to Chang et al. (2007), because of its low cost. Furthermore, user privacy is covered because certain implementations, such as Raj et.al. research do not include real-time positioning and alerts via a server (2013). Despite the fact that the mobile device is being monitored, the consumer location is not real-time in certain other solutions (Chang et al., 2007; Mulloni et al., 2009). The marker's location is decided by the user's position.

The markers are spread across the navigation area, and their location is determined by positioning the mobile device next to the marke r (J. Kim & Jun, 2008). Real-time navigation is still possible with some other solutions, as shown by Barberis et. al. report's (2014). Furthermore, the system's accuracy is measured by the distance between the marker and the device, which is determined by the device's camera's resolution (Raj et al., 2013). If the camera resolution on the platform is insufficient, it may have a negative impact on the system's accuracy and performance (Ibid.).

Barberis et al. (2014), on the other hand, may not need knowledge of the device's camera's resolution or properties. However, because of the extra infrastructures, these networks become more complicated and costly. As a result of these problems, the AR approach was established as an alternative.

Augmented Reality (AR), like marker-based approaches, uses a handheld system with a camera, a marker, and a server (Raj et al., 2013). The data is captured by scanning the pattern of the marker with the mobile device's sensor, while the server is used for location estimation, position determination, and real-time monitoring and navigation (Chang et al., 2007; Mulloni et al., 2009). AR (Möller, Kranz, Huitl, Diewald, & Roalter, 2012) is the overlay of simulated objects with the physical world using visual markers or photographs for orientation, tracking, and navigation.

Klopschitz et al. (2010), Mulloni, Seichter, and Schmalstieg (2011), and Möller et al. (2012) propose that AR obtains visual knowledge by smoothly overlaying a user's vision with position information connected to an image store in a centralised location or server. Optical marker identification, image sequence matching, position recognition, and location annotation are all performed by the server (Klopschitz et al., 2010; J. Kim & Jun, 2008). According to Mautz and Tilch (2011), the server sends the recognised location data to the mobile user,

allowing for real-time positioning and navigation. Mulloni et al. (2011) and Möller et al. (2012) based their research on improving efficiency focused on the interface of an AR indoor navigation system so that navigation in indoor environments can be made better.

While navigating, users are guided by activity-based guidance and information points such as markers positioned in the area to assist accurate positioning and efficiency. As a consequence, navigation errors are greatly minimised. Robustness, simplicity, and accessibility are considerations that are taken into account during the deployment process to accomplish this (Möller et al., 2012; Mulloni et al., 2011). While most positioning and navigation systems use a marker-based approach, Klopschitz et al. (2010) use a new “markerless-based” approach in their research. For matching and monitoring purposes, this method makes use of existing image features.

Since matching image features in real time can be difficult, the markerless approach makes certain assumptions about the mobile device's sensor (Klopschitz et al., 2010). Furthermore, when more markers and a server are used, real-time positioning and navigation is accurate with AR (Möller et al., 2012). Despite the advantages of AR over marker-based systems, image matching can demand a large amount of computational power, raising the difficulty and affecting efficiency (Klopschitz et al., 2010). Furthermore, updating the server could result in an increase in both the cost and the cost of maintenance.

[1] "Trends on Object Detection Techniques Based on Deep Learning," - D. Choi, and M. Kim,

Electronics and Telecommunications Trends, Vol. 33, No. 4, pp. 2332. Published in Aug. 2018. Object detection is a challenging field in the visual understanding research area, detecting objects in visual scenes, and the location of such objects. The paper has recently been applied in various fields such as autonomous driving, image surveillance, and face recognition. In traditional methods of object detection, handcrafted features have been designed for overcoming various visual environments; however, they have a trade-off issue between accuracy and computational efficiency. Deep learning is a revolutionary paradigm in the machine-learning field. In addition, because deep-learning-based methods, particularly convolutional neural networks (CNNs), have outperformed conventional methods in terms of object detection, they have been studied in recent years. The paper provides a brief descriptive summary of several recent deep-learning methods for object detection and deep learning architectures and also compare the performance of these methods and present a research guide of the object detection field.

[2] “Real-Time Object Detection App for Visually Impaired : Third-Eye” - Salman TOSUN, Eni KARARSLAN. IEEE Conferences 2018.

Visually impaired people can’t move safely outdoors because they cannot perceive the outside obstacles as normal people. The prototype application aims to make the visually impaired people's lives easier with the mobile devices. The mobile application with the designed chest apparatus will be like a virtual third eye that is not too costly and easily accessible. All the design and codes is shared with GPL free software licence. The application is developed for the Android platform. Image processing and machine learning technologies are used.

[3] “Android Application Development: A Brief Overview of Android Platforms and Evolution of Security Systems” - Anirban sarkar, Ayush goyal, David hicks and Saikathazra. (2019 Third International conference on I-SMAC).

With the advent of new mobile technologies, the mobile application industry is advancing rapidly. Consisting of several operating systems like symbian OS, iOS, blackberry, etc., Android OS is recognized as the most widely used, popular and user-friendly mobile platform. This open-source linux kernel-based operating system offers high flexibility due to its customization properties making it a dominant mobile operating system. Android applications are programmed in java language. Google android SDK delivers a special software stack that provides developers an easy platform to develop android applications. Developers can make use of existing java IDEs which provides flexibility to the developers. Java libraries are predominant in the process of third-party application development. Cross-platform approaches make sure that developers do not have to develop platform-dependent applications. With the help of these approaches, an application can be deployed to several platforms without the need for changes in coding. Android is more prone to security vulnerabilities which the majority of the users do not take into account. Any android developer can upload their application on the android market which can cause a security threat to any android device. These applications do not have to go through rigorous security checks. The authors of the paper aims to layered approach for android application development along with various cross-platform approaches is discussed.

[4] “(YOLO) You Only Look Once: Unified,Live Object Detection” - Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. Cornell University. Published in Jun 2015

YOLO, an approach to object detection, Prior work on object detection repurposes classifiers to perform detection. Instead, the paper frames object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance. The unified architecture is extremely fast and base YOLO model processes images in real-time at 45 frames per second. A smaller version of the network, Fast YOLO, processes an astounding 155 frames per second while still achieving double the MAP of other real-time detectors. Compared to state-of-the-art detection systems, YOLO makes more localization errors but is less likely to predict false positives on background. YOLO learns very general representations of objects. It outperforms other detection methods, including DPM and R-CNN, when generalizing from natural images to other domains like artwork.

[5] “SSD: Single Shot Multi Box Detector,” – by W. Liu et al., European Conference on Computer Vision, Sept. 2016.

Method for detecting objects in images using a single deep neural network. The author’s approach, named SSD, discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. The network combines

predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes. SSD model is simple relative to methods that require object proposals because it completely eliminates proposal generation and subsequent pixel or feature resampling stage and encapsulates all computation in a single network. This makes SSD easy to train and straightforward to integrate into systems that require a detection component. Experimental results on the PASCAL VOC, MS COCO, and ILSVRC datasets confirm that SSD has comparable accuracy to methods that utilize an additional object proposal step and is much faster, while providing a unified framework for both training and inference. Compared to other single stage methods, SSD has much better accuracy, even with a smaller input image size. For 300×300 input, SSD achieves 72.1% mAP on VOC2007 test at 58 FPS on a Nvidia Titan X and for 500×500 input, SSD achieves 75.1% mAP, outperforming a comparable state of the art Faster R-CNN model.

[6] "AR4VI: AR as an Accessibility Tool for People with Visual Impairments"- October 2017

Getting quality assist to information is one of the most significant tissues facing blind and mutual impaired people with assess to facial information being particularly challenging AR4vi is a general approach to tightly coupling labelling and denotation information with special location improving assist information for people and visual disabilities in both local and global context technology and the others are part of the growing trend in accessibility that rather than requiring specialised materials or modifications to the build environment takes advantage of accelerating improvements in processing power connectivity and cloud based resources, making accessible special in information available more quickly in more contacts and at lower cost and ever before with the increasing availability and reliability of such graphical resources as open street map and Google places as as well. As the broad adoption of sensory and highly connected smartphones are ability to your accessible and relevant information on the environment is reaching more applied and visually impaired users than ever before the general question model of Air force demonstrated by over there is already being adopted by other accessible GPS apps and Microsoft is unlocked project as main stream There are technology is continue to improve in quality and affordability we anticipate continue expansion of the accessible way fine model world similarly the trains and tools of year 4 we that support assist to maps model subjects and other materials points to significant expansion of the approach in coming years in the quality will comes webcams and graphics processing power are improving simultaneously the paving the way for faster and more accurate hand tracking for interactive object based for we at the same time. There is a significant growth in the availability of human and computer based resources for generating diaper you notations ultimately is rapidly expanding field with strong potential not only in assessable for the blind and visually impaired but also for mainstream commercial application and wide array of inter disciplinary HCI related research.

[7] "A Navigation and Augmented Reality System for Visually Impaired People" - Alice Lo valvo, Deniele Croce. Domenico garlist. Published in 28 April 2021.

By using AR Frameworks, Machine Learning, Voice Recognition Services, Indoor Positioning Technologies, a navigation system has been developed. The paper measures the accuracy of outdoor and indoor navigation, considering the system's ability to provide precise location information and guidance. Evaluate the effectiveness of the augmented reality overlay in providing navigational cues and information without causing distraction or confusion. Used datasets are Geospatial Data, Indoor Maps, Object Recognition Datasets, User Feedback Data. Authors suggested the future work as to investigate ways to enhance the AR overlay, potentially incorporating 3D mapping and more context-aware visualizations. Stay abreast of technological advancements, especially in machine learning, AR, and geospatial technologies, to continuously improve the system.

[8] "Navigation system for blind and visually impaired" - Santigo real, aluaro ararjo.

Published in 2 august 2019.

The aim of the author of the paper is to enhance accessibility for blind and visually impaired individuals, allowing them to navigate both indoor and outdoor environments with greater independence and safety. Datasets used are OpenStreetMap (OSM), Indoor Mapping Databases, Image Databases for Object Recognition, Public Transportation Databases. By use of Global Positioning System (GPS), Voice Recognition and Natural Language Processing, Indoor Positioning Systems (IPS), provide real-time navigation feedback, including information about the user's surroundings, obstacles, and the best routes to reach their destination. The author of the paper could focus on improving indoor navigation technologies, refining algorithms for navigating complex indoor spaces like shopping malls, airports, and large public buildings.

[9] "Navigating users with visual impairments in indoor spaces using tactile landmarks" - N. Fallah,I. Apostolopoulos, K. Bekris, E. Folmer, published in 10 may 2012.

Datasets used are Indoor Maps, User Interaction Data, Sensor Data. The goal of the paper is metrics on how well the system performed in assisting visually impaired users in navigating indoor spaces, results from any user studies or feedback sessions. By use of Tactile Landmark Technology, User Interaction Technology, Navigation System, evaluate the effectiveness of tactile landmarks in the navigation process. future work is to Suggestions or plans for improving the current system, considerations for extending the system to different indoor environments or even outdoor spaces.

[10]. "A computer vision system that ensures the autonomous navigation of blind people" - R. Tapu,B. Mocanu,T. Zaharia in IEEE conference 23 November 2013.

Using the Computer Vision Algorithms, Sensor Technologies, Navigation System Architecture, this will results regarding the accuracy of the system in providing autonomous navigation for blind individuals. If available, user feedback and evaluations of the system's usability. They Proposed for improving the current system, including potential technological advancements. Plans for conducting additional user studies to gather more insights into the system's effectiveness.

[11] “Blind Navigation Assistance for Visually Impaired based on Local Depth Hypothesis from a Single Image”-

by Santosh Divvala, Ross Girshick in 2018.

Assisting the visually impaired along their navigation path is a challenging task which drew the attention of several researchers. A lot of techniques based on RFID, GPS and computer vision modules are available for blind navigation assistance. The author of the paper proposed a depth estimation technique from a single image based on local depth hypothesis devoid of any user intervention and its application to assist the visually impaired people. The ambient space ahead of the user is captured by a camera and the captured image is resized for computational efficiency. The obstacles in the foreground of the image are segregated using edge detection followed by morphological operations. Then depth is estimated for each obstacle based on local depth hypothesis. The estimated depth map is then compared with the reference depth map of the corresponding depth hypothesis and the deviation of the estimated depth map from the reference depth map is used to retrieve the spatial information about the obstacles ahead of the user.

[12] “Stereo Vision Based Sensory Substitution for visually impaired” by Otilia Zvoris ,Teanu, Simona Caraiman, Robert-Gabriel Lupu, Nicolae Alexandru Botezatu and Adrian Burlacu, Faculty of Automatic Control and Computer Engineering, “Gheorghe Asachi” Technical University of Iasi. D. Mangeron 27, 700050 Iasi, Romania. Published on 6 July 2021 by MDPI.

Scene understanding and safe mobility are key aspects in the effort of improving the lifestyle of visually impaired people. Vision restoration through multimodal representations of the environment (sound and haptics) represents one of the most promising solutions for aiding the visually impaired. Sound of Vision is a concept that goes beyond the state of the art of the visual sensory substitution systems having the potential of becoming an affordable commercial device that will actually help blind people. It was designed by giving high importance to the possible limiting factors: Wearability, real time operation, pervasiveness, training, evaluation with end-users. To achieve these challenging requirements, the SoV SSD relies on complex computer vision techniques and a fusion of sensors to provide users with a scene description in any type of environment (indoor, outdoor, irrespective of illumination conditions).

The author of the paper has presented a robust and accurate approach for the reconstruction and segmentation of outdoor environments, based on the data acquired from the SoV stereo camera and IMU. The architecture design for stereo-based reconstruction and segmentation is a major contribution. Ground processing, dynamic obstacles detection and scene segmentation can be considered novelties in the vision pipeline. Due to their original purpose, other algorithms such as disparity estimation and camera motion estimation were tuned for the particular cameras used in our system and for the degrees of freedom induced by the current application. At the core of our approach is a fusion of stereo depth maps in a 3D model (point cloud) that is efficiently represented and processed on the GPU. The system provides a reliable 3D representation of the environment even with noisy depth data and in the presence of dynamic elements. It demonstrated system’s performance on a large custom dataset recorded with the SoV cameras.

The author proposed a future as it would also add a semantic layer upon our obstacle detection system. Preliminary results with such an approach already indicate that, besides providing the users with more specific information regarding the environment, it can also be used to further improve the segmentation results (by filtering out false positive objects and merging over-segmented regions). The paper could extend the technical and usability evaluation in outdoor environments to a larger sample and in more diverse conditions.

[14] “Mobile augmented reality using deep learning for visually impaired people” – by Md. Nahidul Islam Opu, Md. Rakibul Islam, Muhammad Ashad Kabir. Published by MDPI in 27 December 2021.

Blind and visually impaired people cannot accurately judge the changes in their surroundings due to eyesight limitations, which increases the risk of accidents indoors and outdoors. The authors propose mobile ARML based on Mobile-Net Single-shot Detection (MobileNet-SSD), Augmented Reality (AR), and a Voice Interaction system for object detection and distance calculation. ARML aimed to aid visually impaired people to avoid obstacles in daily life and quickly query the user-specified items utilizing computer vision and Integrates Lidar for both AR/VR experiences reducing additional equipment and improving detection accuracy of distance between users and obstacles. The system safely improves their ability to identify obstacles in their environment and improves the quality of life for visually impaired people. The NAAD system is based on Mobile-Net Single-shot Detection (MobileNet-SSD), Augmented Reality (AR), and LiDAR for implementing obstacle detection, target object detection, distance calculation, and navigation of user-specified lost objects. It includes a mobile application that uses simple voice and gesture controls to aid navigation. The paper aims to

- Help visually impaired and blind (VIB) people avoid obstacles in daily life.
- Use computer vision to find user-specified objects quickly.
- Integrate LiDAR for AR/VR experiences, reducing additional equipment and improving the distance accuracy between the obstacle and the user.

The Author’s research based on the NAAD system, which designs the safe mode and query mode and upgraded the system and added the navigation function in the query mode. The authors of the paper proposed the object detection integration and introduced the query mode of the NAAD system. This virtual assistant provides different functions such as obstacle detection, distance estimation, and an alarm system that analyzes the environment in real time and alerts the user to avoid obstacles.

[15] “An Indoor and Outdoor Navigation System for Visually Impaired People” by Croce D.,Giarré L.,Pascucci F.,Tinnirello I, Galioto G.E, Garlisi D, Lo Valvo A. in the year 2019

The goal of the authors of the paper is to use metrics, which measures the system's performance, accuracy, and effectiveness. If available, user feedback and evaluations of the system's usability and user satisfaction. The author proposed future work as Proposals for improving the current system, potentially with new technologies or methodologies. Plans for conducting additional user studies to gather more insights into the system's effectiveness and areas for improvement. Ideas for integrating the system with emerging technologies or methodologies to

enhance navigation capabilities and would also add a semantic layer upon our obstacle detection system. Preliminary results with such an approach already indicate that, besides providing the users with more specific information regarding the environment, it can also be used to further improve the segmentation results.

[16] “Indoor Mapping Based on Augmented Reality Using Unity Engine” by Ajith et al. (2020),

The researchers implemented an indoor navigation app for both android and iOS using a place note SDK. However, there is no software available at the time of publishing where we can access the software code. So, by implementing an AR indoor navigation application using the ARWAY software development toolkit, we conducted a comparative study of 2D physical map and AR applications to determine which navigation method would navigate indoor in less time.

[17] “Indoor Positioning and Navigation with Camera Phones” by Alessandro Mulloni, Daniel Wagner, István Barakony, Dieter Schmalstieg. Published by IEEE Pervasive Computing in April 2009

The authors illustrated how to use mobile cameras for indoor positioning in a conference guide implementation in 2009. The authors attached rectangle-shaped contractual markers to signposts in the conference grounds (similar in purpose and appearance to QR codes). Every marker had a specific visual ID associated with the signpost's position. The application was able to determine the 3D location and direction of the user with regard to the markers in real time with centimetre-level precision. The user had to point the camera directly at the marker, and they had to be close enough to the signpost so that the marker was visible to the came in necessary detail.

[18] “Augmented Reality Awareness and Latest Applications in Education: A Review” by Soraia Oueida, Pauly Awad, Claudia Mattar [July 2023]. Published by International Journal of Emerging Technologies in Learning (iJET) 18(13):21-44.

Several advances in technology have touched the educational sector, where improvements and enhancements are always a surge for innovative learning. Augmented Reality (AR) is considered an efficient and promising technology capable of enhancing the educational sector. AR is a visualization technology that allows human interaction by providing users with a perception of reality using virtual information. In other words, AR adds to the existing real-world environment some extra virtual information generated by computer techniques that can enhance the overall experience. This promising technology can be applied in different sectors such as education, healthcare, banks, tourism, etc. Despite the great work available in the literature about this topic, deep insights and extensive research are still needed before the actual implementation in the real world. The aim of authors work is to highlight the importance of AR in improving the educational sector and increase awareness of using such an approach. A review is conducted to prove the validity of AR in increasing learning outcomes, students’ motivation, and engagement in classrooms, as well as improving the understanding level. Several papers were analyzed based on different application areas under various interdisciplinary databases. As a result, the papers were collected in a table-based format to provide researchers with better and easier insights on how to improve

the use of AR in education before real implementation. The paper helped in highlighting potential future work, discussing limitations, advantages, disadvantages, and the latest advances in this technology.

[19] "A marker-based augmented reality system for mobile devices" by Alexandru Gherghina, Alex Olteanu, Nicolae Tapus. Conference: Roedunet International Conference (RoEduNet), January 2013 .

Given the rate at which mobile devices are adopted by the general public, now this generation is at a point where many people own a smart phone. Currently, this makes them the most affordable interfaces for augmented reality. With respect to augmented reality, smartphones present some limitations: location tracking is, on a large scale, supported only for outdoor environments and the content types that can be displayed to the user are generally limited. The author therefore propose a marker-based tracking system which detects QR-codes on the camera capture and overlays rich media obtained from a server. And further analyse the performance implications of such a system, by breaking down the processing into pipelined stages and providing a mechanism of skipping frames from the capture in order to give a fast on-screen response.

[20] "Augmented Reality based Indoor Navigation using Point Cloud Localization" by Vishva Patel , Dr. Ratvinder Grewal. Published Online January 2022 in IJEAST.

People of various ages may find it difficult to navigate complex building structures as they become more prevalent. The future belongs to a world that is artificially facilitated, and Augmented Reality will play a significant role in that future. The concept of Indoor Navigation using a smartphone-based Augmented Reality technology is explored in this paper. Using readily available and affordable tools, the author of the paper proposes a solution to this issue. The authors built an Augmented Reality-based framework to assist users in navigating a building using ARWAY, a software development toolkit. To find the shortest paths, we used the Point Cloud Localization and A* pathfinding algorithms. The author of the paper proposed a methodology for navigating an indoor area in student living apartments using an augmented reality (AR) application.

CHAPTER 3

SYSTEM REQUIREMENTS

3.1 Hardware Requirements

Processor : Intel Core I3 and above

Processor Speed : 1.0GHZ or above

RAM : 4 GB RAM or above

Hard Disk : 500 GB hard disk or above

Set up a device with a camera to capture real-time visual input.

Ensure compatibility with augmented reality features, considering factors such as processing power and sensor capabilities.

3.2 Software Requirements

Operating System : Windows 10/11 or above

Front End : Android Studio

Back End : Java

3.3Packages Used for the proposed project

Object detection : org.tensorflow.lite.example.detection

Object recognition : SSD-MobineNet_COCO model

Speech Synthesis : TextToSpeech tt

CHAPTER 4

METHODOLOGY

4.1 Detailed explanation about technologies

Augmented Reality (AR) is a technology that superimposes digital information, such as images, videos, or 3D models, onto the real-world environment. This is typically achieved through devices like smartphones, smart glasses, or headsets. AR enhances the user's perception of the real world by overlaying digital content, allowing

for an interactive and immersive experience. Applications range from gaming and education to navigation systems, where AR can provide additional context, information, or guidance in real-time.

To provide a user-friendly interface that allows users to access and utilize the system's features efficiently, android application is developed.

Android application : Android application is software that runs on android platform. Because the android application is built for mobile devices a typical android application is designed for smartphones or tablet pc that runs on android OS. Android application is very easy, convenient, portable and impressive in use. Also because of its feature and operation abilities more functions could be performed. It is a very powerful and upcoming platform.

Developing an Android application is integral to this project as it serves as the platform through which visually impaired individuals interact with the navigation system. Here's how the Android application can be utilized in this project:

- User Interface : The Android application provides a graphical user interface (GUI) designed specifically for visually impaired users. The interface is optimized for accessibility, featuring large buttons, high contrast visuals, and support for screen readers and other accessibility features.
- Integration of Sensors and Cameras : The Android application integrates with the device's sensors and cameras to capture visual and environmental data in real-time. This data is processed and analyzed by the navigation system to provide users with relevant information and guidance.
- Speech Synthesis : The Android application incorporates speech synthesis capabilities to convert textual information into spoken words or sentences. This allows the system to deliver auditory instructions, alerts, and feedback to users, enhancing their navigation experience.
- Augmented Reality (AR) Overlay : The Android application overlays augmented reality (AR) information onto the user's camera feed, providing real-time visual cues and guidance. This AR overlay may include directional arrows, distance markers, and object labels to assist users in navigating their surroundings.
- Navigation and Route Planning : The Android application facilitates navigation and route planning by allowing users to input their destination, select preferred routes, and receive turn-by-turn directions. Users can interact with the application to adjust route preferences, avoid obstacles, or reroute if necessary.
- Customization and Preferences : The Android application allows users to customize their navigation experience based on personal preferences and needs. This may include adjusting speech synthesis settings, choosing preferred navigation modes, or configuring interface elements for optimal usability.

Overall, the Android application serves as the primary interface for visually impaired individuals to access and interact with the navigation system. It provides a seamless and intuitive user experience, empowering users to navigate their surroundings safely, independently, and with confidence.

Image Processing is technique used to preprocess the images captured by cameras or sensors, optimizing them for CNN-based object detection. This preprocessing may involve tasks like noise reduction, contrast enhancement, and edge detection, which help improve the accuracy and efficiency of object recognition.

Image Processing involves manipulating or analyzing an image to extract information, enhance certain features, or prepare it for further analysis. The process encompasses various techniques to improve image quality, extract relevant information, and derive meaningful insights. The goal is to ensure that the processed images contribute to an effective and adaptive navigation aid for users with visual impairments.

Image processing is a field of study and practice that involves the manipulation of digital images to enhance or extract information from them. It encompasses a wide range of techniques and methods for analyzing, modifying, and interpreting images. Image processing is used in various applications, including computer vision, medical imaging, remote sensing, and multimedia systems. Here are some key concepts and techniques related to image processing:

- Image Acquisition: The process of obtaining digital images from various sources, such as cameras, satellites, or scanners. Image acquisition in the context of navigation systems for visually impaired individuals involves the process of capturing visual information from the surrounding environment using cameras or sensors integrated into the user's device or wearable technology. This visual data typically consists of images or video streams that contain crucial information about obstacles, landmarks, signs, and other relevant features essential for navigation. The image acquisition system is designed to optimize image quality and performance under various environmental conditions, ensuring clear and reliable data capture. This data is continuously monitored and processed in real-time by the navigation software, allowing for dynamic responses to changes in the environment and providing users with accurate guidance and assistance as they navigate through their surroundings. Overall, image acquisition serves as the foundation for enabling visually impaired individuals to access and interpret visual information, thereby enhancing their mobility, independence, and safety.

- Image Representation: Digital images are represented as a matrix of pixel values. Each pixel corresponds to a tiny element of the image. Image representation refers to the process of encoding visual information from images into a format that can be understood and processed by computers or algorithms. In navigation systems for visually impaired individuals, image representation involves converting raw image data captured by cameras or sensors into a structured and meaningful representation that facilitates analysis and interpretation. This representation often involves converting pixels into numerical values, where each pixel's color and intensity are represented by numerical attributes. Additionally, image representation may include techniques such as feature extraction, where key visual features, such as

edges, textures, or shapes, are identified and encoded into a more compact and informative format. The goal of image representation is to transform raw visual data into a format that can be effectively analyzed and utilized by machine learning algorithms, computer vision techniques, or other image processing methods employed in navigation systems. This structured representation enables the system to recognize objects, identify obstacles, and extract relevant spatial information from images, ultimately supporting the navigation and orientation needs of visually impaired users in their everyday environments.

- Image Enhancement: Techniques to improve the visual quality of an image, such as adjusting brightness, contrast, and sharpness. Image enhancement refers to the process of improving the visual quality of an image to make it more suitable for analysis or interpretation. In navigation systems designed for visually impaired individuals, image enhancement plays a pivotal role in optimizing the visual information captured by cameras or sensors. Through techniques such as contrast enhancement, noise reduction, sharpness adjustment, colour correction, and dynamic range adjustment, image enhancement aims to enhance the clarity, detail, and fidelity of the captured images. By improving the quality of visual input, image enhancement enables more accurate object detection, recognition, and scene interpretation, which are crucial for providing effective navigation assistance to visually impaired users. Ultimately, image enhancement contributes to a more reliable and informative navigation experience, empowering visually impaired individuals to navigate their surroundings with greater confidence and independence.

- Image Filtering: The application of filters to an image to highlight or suppress specific features. Common filters include blurring, sharpening, and edge detection filters. Image filtering is a fundamental technique used in image processing to modify the characteristics of an image by altering the pixel values according to specific rules or mathematical operations. In navigation systems tailored for visually impaired individuals, image filtering is employed to enhance the quality and clarity of visual information captured by cameras or sensors. This process typically involves applying various filters, such as smoothing filters for noise reduction, edge detection filters for highlighting object boundaries, and sharpening filters for enhancing image details. These filters help improve the visual appearance of the captured images, making it easier to identify and interpret important features within the environment. By effectively filtering visual data, navigation systems can provide more accurate object detection, obstacle recognition, and scene analysis, thus facilitating safer and more efficient navigation for visually impaired users. Overall, image filtering is a crucial component of navigation systems, enabling the enhancement and optimization of visual information to support the navigation needs of individuals with visual impairments.

- Image Restoration: Methods to remove noise, distortions, or artifacts from an image to restore it to a cleaner state. Image restoration involves the application of various techniques to improve the quality of an image by reducing degradation caused by factors such as noise, blur, or distortion. In navigation

systems for visually impaired individuals, image restoration plays a vital role in enhancing the clarity and reliability of visual information captured by cameras or sensors. This process encompasses a range of algorithms and filters designed to mitigate degradation effects and restore the image to its original or desired quality.

One common form of degradation addressed by image restoration techniques is noise, which can arise from factors such as low light conditions or sensor limitations. Noise reduction algorithms aim to suppress or remove unwanted noise while preserving essential image details, thus improving overall image quality. Additionally, image restoration techniques may include deblurring algorithms to compensate for motion blur or defocus blur, ensuring sharper and more focused images for accurate object recognition and scene interpretation.

By applying image restoration techniques, navigation systems can provide visually impaired users with clearer and more reliable visual information, facilitating better object detection, obstacle recognition, and navigation guidance. Ultimately, image restoration contributes to a more robust and effective navigation experience, empowering visually impaired individuals to navigate their surroundings with greater confidence and independence.

- Image Segmentation: Dividing an image into meaningful regions or segments based on certain criteria, such as colour, intensity, or texture. mage segmentation is a fundamental technique in image processing that involves partitioning an image into multiple segments or regions based on certain criteria, such as color, intensity, texture, or spatial proximity. In navigation systems designed for visually impaired individuals, image segmentation is utilized to identify and delineate different objects, obstacles, and features within the environment captured by cameras or sensors.

Through image segmentation, visually impaired users can obtain a clearer understanding of their surroundings by distinguishing between various elements in the scene, such as sidewalks, roads, buildings, pedestrians, and vehicles. This segmentation enables the navigation system to provide more precise and contextually relevant guidance and assistance tailored to the user's immediate surroundings.

Various algorithms and techniques are employed for image segmentation, including thresholding, clustering, edge detection, and region-based segmentation methods. These techniques analyze the visual characteristics of the image to group pixels or regions with similar attributes, effectively separating distinct objects and structures from the background.

By leveraging image segmentation, navigation systems can enhance the accuracy and effectiveness of object detection, obstacle recognition, and spatial understanding, ultimately facilitating safer and more efficient navigation experiences for visually impaired individuals. Through clear delineation of the environment's features, image segmentation empowers users with valuable spatial awareness and aids in navigating complex urban environments with confidence and independence.

- Object Recognition: Identifying and classifying objects or patterns within an image using techniques from machine learning and computer vision.

Programming languages like Java and libraries such as OpenCV are commonly used for image processing tasks. Additionally, deep learning techniques, particularly convolutional neural networks (CNNs), have become increasingly important in solving complex image processing problems.

To recognize various objects and obstacles Convolutional Neural Networks (CNNs) can be trained.

Convolutional Neural Networks (CNNs) are a class of deep learning algorithms specifically designed for processing and analysing visual data. Trained on a comprehensive dataset, CNNs architecture allows for the extraction of intricate features from visual input, contributing to a more nuanced understanding of the user's environment. By leveraging CNNs, this project enhances the system's ability to interpret and respond to visual cues, providing valuable insights crucial for creating an effective and reliable navigation aid for visually impaired individuals.

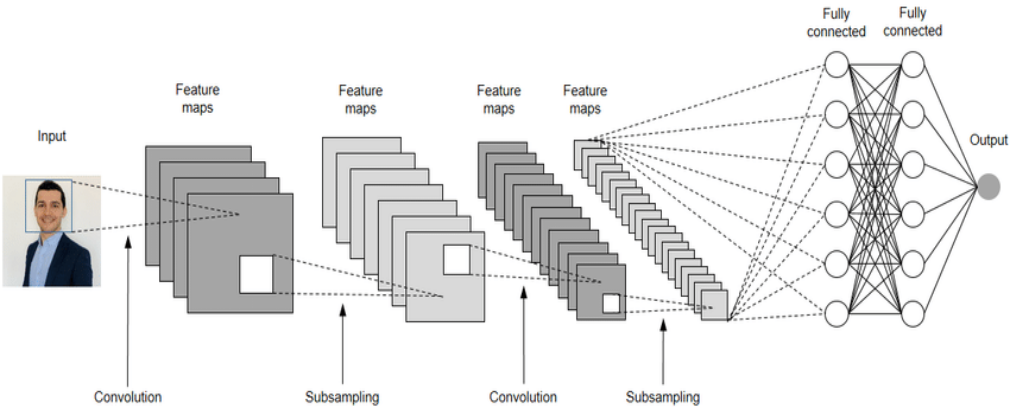


Fig 1 : CNN Architecture

A standard CNN architecture involves several layers, commonly organized as follows:

i. Input Layer: This is where the network receives the raw input data, usually in the form of images. Each element in the input layer represents a pixel in the image.
ii. Convolutional Layers: These layers perform convolution operations, applying filters or kernels to the input data. This process helps identify patterns and features in the data, capturing spatial hierarchies.
iii. Activation (ReLU) Layers: After convolution, ReLU (Rectified Linear Unit) activation functions are often applied to introduce non-linearity, allowing the network to learn more complex representations.

iv. Pooling (Subsampling or Down-sampling) Layers: Pooling layers reduce the spatial dimensions of the data by down sampling. Common pooling operations include max pooling, which extracts the maximum value from a group of values, aiding in retaining essential features.

v. Flattening Layer: This layer converts the multidimensional data into a one-dimensional vector. It prepares the data for input into the fully connected layers.

vi. Fully Connected (Dense) Layers: These layers connect every neuron from one layer to every neuron in the next layer. They consolidate the learned features from previous layers for final decision-making.

vii. Output Layer: The final layer produces the network's output. The number of nodes in this layer depends on the task (e.g., binary classification, multi-class classification).

Activation functions like SoftMax are commonly used for classification problems. This general structure is a simplified overview of a CNN, and specific architectures may have additional layers or variations to suit particular tasks.

SSD stands for Single Shot MultiBox Detector, which is a type of object detection algorithm used in computer vision and image processing tasks. SSD is known for its efficiency and effectiveness in real-time object detection, particularly in scenarios where speed and accuracy are essential, such as autonomous vehicles, surveillance systems, and augmented reality applications.

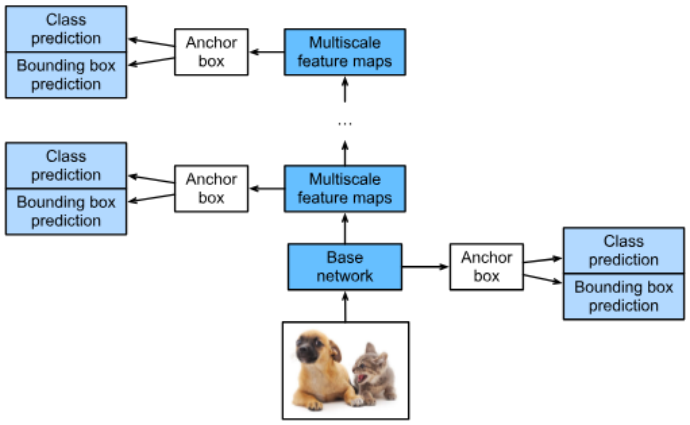


Figure 2. SSD Working model

SSD is a popular deep learning architecture for object detection. It's known for its efficiency, simplicity, and effectiveness.

Here is a detailed explanation of how SSD works:

i. Base Convolutional Network: SSD starts with a base convolutional network (e.g., VGG, ResNet, or MobileNet), which serves as a feature extractor. This network processes the input image and extracts feature maps at multiple spatial scales.

ii. Multi-scale Feature Maps: SSD then applies a series of convolutional layers on top of the base network to produce feature maps at multiple resolutions. These feature maps capture semantic information at different levels of abstraction. Each feature map represents a different scale of the input image.

iii. Anchor Boxes: SSD uses anchor boxes or default boxes to predict object bounding boxes and their corresponding class labels. Anchor boxes are predefined boxes of different aspect ratios and scales, evenly distributed across each spatial location in the feature maps. These anchor boxes serve as reference templates for detecting objects of different sizes and shapes.

iv. Predictions: For each spatial location in the feature maps, SSD predicts two types of information:
   - Object Class Scores: For each anchor box, SSD predicts the probability scores for each object class. These scores indicate the likelihood of the presence of each class within the anchor box.
   - Bounding Box Offsets: SSD predicts adjustments (offsets) to the dimensions and location of the anchor boxes to better fit the objects in the image.

v. Multi-scale Predictions: SSD makes predictions at multiple scales by using feature maps from different layers of the network. Predictions from higher-resolution feature maps are more accurate for detecting small objects, while predictions from lower-resolution feature maps are better for detecting larger objects.

vi. Loss Function: SSD uses a combination of localization loss (smooth L1 loss) and confidence loss (softmax loss) to train the network. The localization loss measures the accuracy of predicted bounding box coordinates, while the confidence loss measures the accuracy of predicted class probabilities.

vii. Non-maximum Suppression (NMS) : After making predictions, SSD applies non-maximum suppression to filter out redundant detections. This post-processing step removes overlapping bounding boxes with lower confidence scores, keeping only the most confident detections for each object.

By combining these components, SSD achieves real-time object detection with high accuracy. It's widely used in applications such as autonomous driving, surveillance, and image-based search. Additionally, SSD can be adapted to different tasks and datasets by adjusting the network architecture and training on domain-specific data.

To provide auditory guidance and instructions to visually impaired people, Speech synthesis is used.

Speech synthesis: Speech synthesis is the artificial production of human speech. Synthesized speech can be created by concatenating pieces of recorded speech that is stored in database. This is generated after data extraction and detection in speech synthesis. The output will be in form of speech which will give the impaired users an idea of object detected.

Speech synthesis also known as text-to-speech (TTS), it converts textual information, such as navigation directions, distance of object from user and object name into spoken words or sentences that are then delivered to the user through audio output devices, such as headphones or speakers.

Speech synthesis enhances the accessibility and usability of the navigation system for visually impaired users by providing them with real-time auditory feedback and guidance. It enables users to navigate their surroundings more effectively, without the need to rely solely on visual cues. Additionally, speech synthesis can be customized

to suit the user's preferences, such as voice type, speaking rate, and language, ensuring a personalized and user-friendly navigation experience. Overall, integrating speech synthesis into the navigation system enhances the user's independence, safety, and confidence during travel.

4.2 METHODOLOGY OF PROPOSED PROJECT

In many scenarios, it is difficult to implement a physical path dedicated to visually impaired people, and even installing a simple line painted on the floor of sites can be a problem and any kind of path or line applied on a precious historical pavement or in some important cultural context might not be appropriate. To overcome these difficulties, have to exploit the potential offered by augmented reality algorithms. Instead of a physical path, proposed project apply a virtual path, that is guided through the smartphone.

The proposed project, that is developed by integrating various technologies, namely, Android SDK was used for developing the application because it is the official Integrated development environment (IDE) designed specifically for developing Android applications. The Android framework supports capturing images and video through the android. Hardware used is camera2 API or camera Intent, is a package used for capturing real-time video for object detection and reading text. TensorFlow library is used for implementing object detection models inside the android application. It provides high performance numerical computing. It has a flexible architecture, making it easy to deploy the calculation through a variety of all possible platforms. SSD-Mobile-Net-COCO model was being used for real time processing. The SSD architecture is a single convolution network that learns to predict bounding box locations and predicts the detected object in the form of limitation boxes. The system uses two object detector modules and real time text reader.

The proposed system has the following key aspects.

Dataset used in the proposed project

In this proposed project, the Common-Object-in-Context (COCO)dataset was used for training the model i.e. SSD MobileNet model, which was able to recognize 81 different categories.

Object Detection

The app is using the SSD-MobileNet-COCO model detecting objects. It utilizes only one neural network for the entire input image. The network model then separates the real time input image into various different regions and predicts the objects using a quadrilateral surrounding the object along with its probability score.

Text Reader

The text reader is using Google mobile vision API for detecting texts in real time just like an OCR and then converts the text into speech by using Googles TTS library, which was available in android SDK while making application. Using this, a user can easily read the menu cards in restaurants, texts on objects(medicines, food,etc.), hotel room no., or even read a paper document, etc.

Text Recognizer

This object processes the images and determines the text contained in it. Once initialized, it can be used to detect text in all picture types. Reading text feature was implemented using Google Text-to-Speech, which speaks the detected text and acknowledged objects.

Along the path, the user receives a tactile stimulus when walking along the virtual line and is accompanied step-by-step guidance in the right direction with precise feedback. Indeed, visually impaired people cannot see the line on the smartphone, but they can feel the tactile vibration when the line is located at the center of the camera.

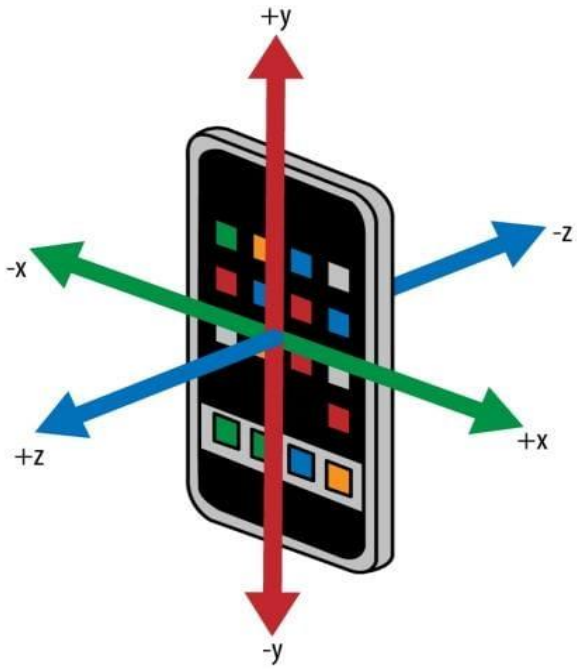


Figure 3. Camera coordinate system

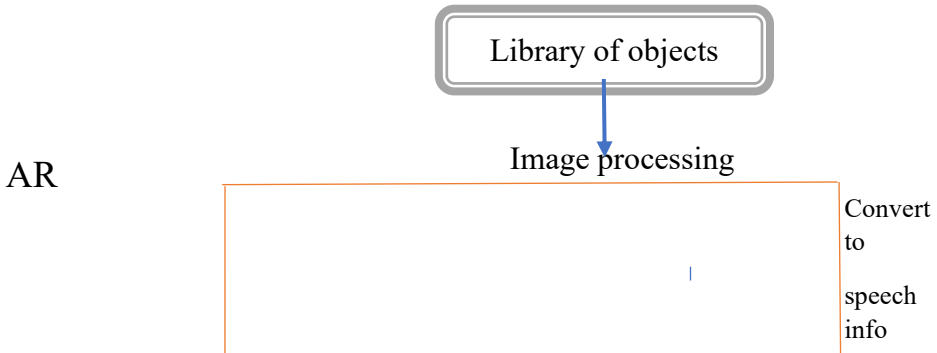
Figure(3) explains the view of mobile as axis format to navigate purpose for visually impaired individuals. Camera view coordinate system is the system that has its origin on the image plane and the Z -axis perpendicular to the image plane. It assume that +X points left, and +Y points up and +Z points out from the image plane. The transformation from world to view happens after applying a rotation (R) and translation (T).

The camera coordinate system typically has its origin at the camera's optical center, with the camera's view direction aligned along the negative z-axis. The x and y-axes then correspond to the horizontal and vertical directions in the image plane. This system is often used in computer vision and graphics for representing and manipulating images.

CHAPTER 5

DESIGN AND IMPLEMENTATION

5.1 System Design



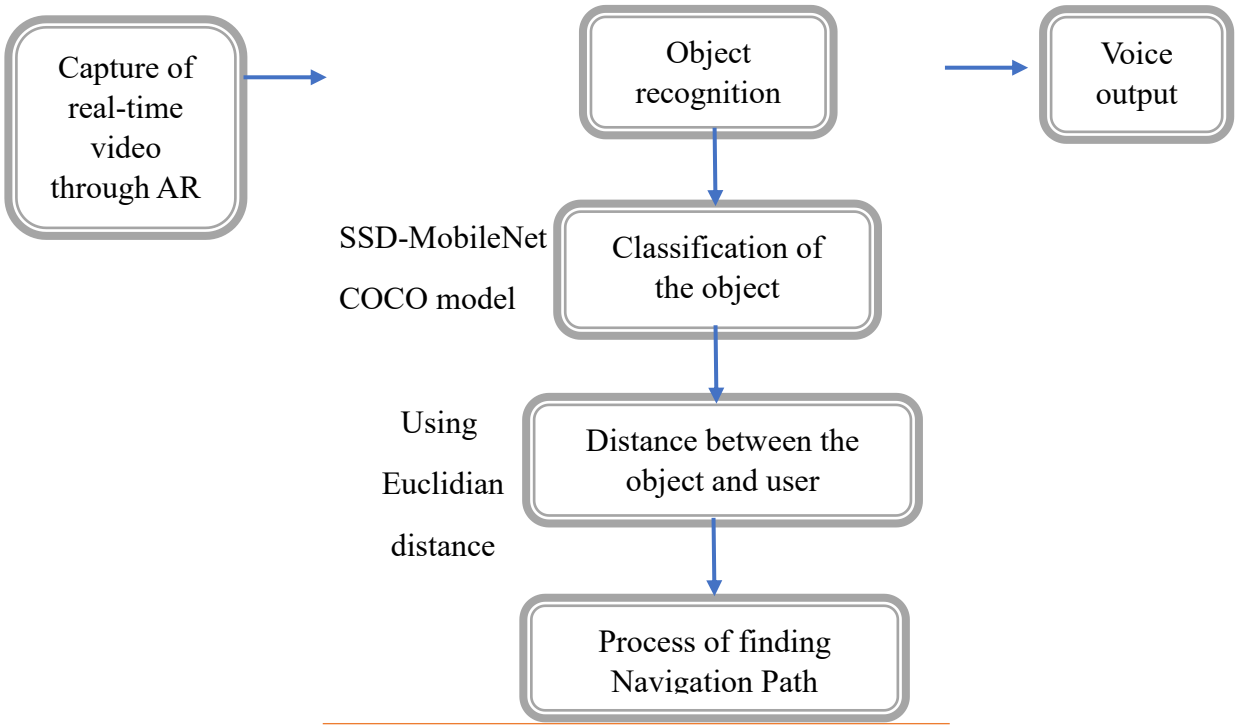


Figure 4. Block diagram of the proposed project

5.2 Implementation

This model mainly consists of a base network followed by several multiscale feature map blocks. The base network is for extracting features from the input image, so it can use a deep CNN. For example, the original single-shot multibox detection paper adopts a VGG network truncated before the classification layer (Liu et al., 2016), while ResNet has also been commonly used. Through our design we can make the base network output larger feature maps so as to generate more anchor boxes for detecting smaller objects. Subsequently, each multiscale feature map block reduces (e.g., by half) the height and width of the feature maps from the previous block, and enables each unit of the feature maps to increase its receptive field on the input image. Since multiscale feature maps closer to the top of figure are smaller but have larger receptive fields, they are suitable for detecting fewer but larger objects.

```
def cls_predictor(num_anchors, num_classes):
    return nn.Conv2D(num_anchors * (num_classes + 1), kernel_size=3,
                     padding=1)
```

In a nutshell, via its base network and several multiscale feature map blocks, single-shot multibox detection generates a varying number of anchor boxes with different sizes, and detects varying-size objects by predicting classes and offsets of these anchor boxes (thus the bounding boxes). thus, this is a multiscale object detection model.

Class Prediction Layer

Let the number of object classes be q. Then anchor boxes have q+1 classes, where class 0 is background. At some scale, suppose that the height and width of feature maps are h and w, respectively. When a anchor boxes are generated with each spatial position of these feature maps as their center, a total of hwa anchor boxes need to be classified. This often makes classification with fully connected layers infeasible due to likely heavy parametrization costs. Single-shot multibox detection uses the same technique to reduce model complexity. Specifically, the class prediction layer uses a convolutional layer without altering width or height of feature maps. In this way, there can be a one-to-one correspondence between outputs and inputs at the same spatial dimensions (width and height) of feature maps. More concretely, channels of the output feature maps at any spatial position (x, y) represent class predictions for all the anchor boxes centred on (x, y) of the input feature maps. To produce valid predictions, there must be $a(q$+1) output channels, where for the same spatial position the output channel with index $i(q$+1)+j represents the prediction of the class j ($0 \leq j \leq q$) for the anchor box i ($0 \leq i < a$). Below it define such a class prediction layer, specifying a and q via arguments num_anchors and num_classes, respectively. This layer uses a 3×3 convolutional layer with a padding of 1. The width and height of the input and output of this convolutional layer remain unchanged.

Bounding Box Prediction Layer

The design of the bounding box prediction layer is similar to that of the class prediction layer. The only difference lies in the number of outputs for each anchor box: here we need to predict four offsets rather than q+1 classes.

```
def forward(x, block):
    block.initialize()
    return block(x)


Y1 = forward(np.zeros((2, 8, 20, 20)), cls_predictor(5, 10))
Y2 = forward(np.zeros((2, 16, 10, 10)), cls_predictor(3, 10))
Y1.shape, Y2.shape


def flatten_pred(pred):
    return npx.batch_flatten(pred.transpose(0, 2, 3, 1))


def concat_preds(preds):
    return np.concatenate([flatten_pred(p) for p in preds], axis=1)
```

Concatenating Predictions for Multiple Scales

Single-shot multibox detection uses multiscale feature maps to generate anchor boxes and predict their classes and offsets. At different scales, the shapes of feature maps or the numbers of anchor boxes centered on the same unit may vary. Therefore, shapes of the prediction outputs at different scales may vary.In the following example, we construct feature maps at two different scales, Y1 and Y2, for the same minibatch, where the height and width of Y2 are half of those of Y1. Let's take class prediction as an example. Suppose that 5 and 3 anchor boxes are generated for every unit in Y1 and Y2, respectively. Suppose further that the number of object classes is 10. For feature maps Y1 and Y2 the numbers of channels in the class prediction outputs are $5\times(10+1)=55$ and $3\times(10+1)=33$, respectively, where either output shape is (batch size, number of channels, height, width).As we can see, except for the batch size dimension, the other three dimensions all have different sizes. To concatenate these two prediction outputs for more efficient computation, we will transform these tensors into a more consistent format. The channel dimension holds the predictions for anchor boxes with the same center. We first move this dimension to the innermost. Since the batch size remains the same for different scales, we can transform the prediction output into a two-dimensional tensor with shape (batch size, height × width × number of channels). Then we can concatenate such outputs at different scales along dimension 1.

```
def down_sample_blk(num_channels):
    blk = nn.Sequential()
    for _ in range(2):
        blk.add(nn.Conv2D(num_channels, kernel_size=3, padding=1),
                nn.BatchNorm(in_channels=num_channels),
                nn.Activation('relu'))
    blk.add(nn.MaxPool2D(2))
    return blk


def base_net():
    blk = nn.Sequential()
    for num_filters in [16, 32, 64]:
        blk.add(down_sample_blk(num_filters))
    return blk


forward(np.zeros((2, 3, 256, 256)), base_net()).shape
```

The complete single shot multibox detection model consists of five blocks. The feature maps produced by each block are used for both (i) generating anchor boxes and (ii) predicting classes and offsets of these anchor boxes. Among these five blocks, the first one is the base network block, the second to the fourth are down sampling blocks, and the last block uses global max-pooling to reduce both the height and width to 1. Technically, the second to the fifth blocks are all those multiscale feature map blocks in Figure 1.

```
def get_blk(i):
    if i == 0:
        blk = base_net()
    elif i == 4:
        blk = nn.GlobalMaxPool2D()
    else:
        blk = down_sample_blk(128)
    return blk
```

Now we define the forward propagation for each block. Different from in image classification tasks, outputs here include (i) CNN feature maps Y, (ii) anchor boxes generated using Y at the current scale, and (iii) classes and offsets predicted (based on Y) for these anchor boxes. a multiscale feature map block that is closer to the top is for detecting larger objects. thus, it needs to generate larger anchor boxes. In the above forward propagation, at each multiscale feature map block we pass in a list of two scale values via the sizes argument of the invoked multibox_prior function. In the following, the interval between 0.2 and 1.05 is split evenly into five sections to determine the smaller scale values at the five blocks: 0.2, 0.37, 0.54, 0.71, and 0.88. Then their larger scale values are given by 0.2×0.37=0.272, 0.37×0.54=0.447, and so on. It create a model instance and use it to perform forward propagation on a minibatch of 256×256 images X. As stated in earlier, the first block outputs 32×32 feature maps. Recall that the second to fourth down sampling blocks halve the height and width and the fifth block uses global pooling. Since 4 anchor boxes are generated for each unit along spatial dimensions of feature maps, at all the five scales a total of (322+162+82+42+1)×4=5444 anchor boxes are generated for each image.

5.3 Module Description

5.3.1 Input Module

-For capturing real-time camera feed, serving as the primary source of visual information for the system.

5.3.2 Processing Module (Convolutional Neural Network):

-Utilizes CNN for object recognition and spatial understanding based on the input from the real-time camera feed. The CNN processes visual data to identify obstacles, landmarks, and other relevant elements.

5.3.3 Augmented Reality Integration:

-Integrates the output from the processing module into the user's real-world view, overlaying identified objects and spatial information using augmented reality technology.

5.3.4 Feedback Module:

-Generates auditory or haptic feedback to convey navigational cues and object proximity alerts to the user. This module ensures that users receive real-time, context-aware information about their surroundings.

5.3.5 User Interface:

-Provides a user-friendly interface with customizable settings, enabling users to plan routes, make real-time adjustments, and receive updates on their navigation.

5.3.6 Adaptability and Scalability:

-Ensures the system's capability to integrate with emerging technologies, allowing for future enhancements and updates to keep the system adaptable to evolving user needs.

5.3.7 Iterative Testing and User Feedback Loop:

-Involves continuous testing and refinement based on user feedback and experiences. This iterative process is crucial for optimizing the system's performance and usability in diverse environments.

-Delivers an enhanced navigation experience for visually impaired individuals, promoting improved mobility, independence, and safety by providing meaningful, real-time information about their surroundings.

CHAPTER 6

CONCLUSION

The development of a navigation system for visually impaired individuals represents a significant stride towards improving accessibility, safety, and independence for this user group. By seamlessly integrating real-time object recognition through CNN and presenting the information in an augmented reality environment, the system addresses the challenges faced by visually impaired individuals in navigating unfamiliar surroundings. The fusion of these technologies not only enhances the accuracy of spatial understanding but also provides an intuitive and context-aware user interface. The auditory or haptic feedback mechanisms contribute to a more immersive and accessible navigation experience, ensuring that users receive timely information about their surroundings. The navigation system stands as a testament to the possibilities when advanced technologies are harnessed to address real-world challenges, ultimately empowering individuals with visual impairments to navigate the world with increased confidence and autonomy.

The future enhancements of the proposed project may focus on the development of a system-on-chip (SoC) so that the size, weight, and cost of the system can be reduced. Moreover, new methodology could be introduced to increase the number of objects to be recognized. The specified tool can be implement instead by using mobile, to make easy to use for visually impaired individuals.

REFERENCES

1. D. Choi, and M. Kim, "Trends on Object DetectionTechniques Based on Deep Learning," Electronics andTelecommunications Trends, Vol. 33, No. 4, pp. 23-32,Aug. 2018, Article.

2. Salman TOSUN, Eni KARARSLAN “Real-Time Object Detection App for Visually Impaired : Third-Eye”. IEEE Conferences 2018.

3. Anirban sarkar, Ayush goyal, David hicks and Saikathazra “Android Application Development: A Brief Overview of Android Platforms and Evolution of Security Systems” 2019 Third International conference on I-SMAC.

4. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi “(YOLO) You Only Look Once: Unified,Live Object Detection”. Cornell University, Jun 2015.

5. W. Liu et al., “SSD: Single Shot Multi Box Detector,” European Conference on Computer Vision, Vol.9905, pp. 21-37, Sept. 2016, Article (CrossRef Link)

6. AR4VI: AR as an Accessibility Tool for People with Visual Impairments October 2017 – by James Coughlan, Researchgate.

7. Alice Lo valvo, Deniele Croce.Domenico garlist -“A Navigation and Augmented Reality System for Visually Impaired People” 28 April 2021.

8. “Navigation system for blind and visually impaired” - Santigo real, aluaro ararjo. Published in 2 august 2019.

9. “Navigating users with visual impairments in indoor spaces using tactile landmarks” -N. Fallah,I. Apostolopoulos, K. Bekris, E. Folmer, published in 10 may 2012.

10. “A computer vision system that ensures the autonomous navigation of blind people” - R. Tapu,B. Mocanu,T. Zaharia in IEEE conference 23 november 2013.

11. “Blind Navigation Assistance for Visually Impaired based on Local Depth Hypothesis from a Single Image”- Santosh Divvala, Ross Girshick in 2018.

12. “Stereo Vision Based Sensory Substitution for visually impaired” by Otilia Zvoris ,Teanu, Simona Caraiman, Robert-Gabriel Lupu, Nicolae Alexandru Botezatu and Adrian Burlacu, Faculty of Automatic Control and Computer Engineering, “Gheorghe Asachi” Technical University of Iasi. D. Mangeron 27, 700050 Iasi, Romania. Published on 6 July 2021 by MDPI.

13. “Mobile augmented reality using deep learning for visually impaired people” - Md. Nahidul Islam Opu, Md. Rakibul Islam, Muhammad Ashad Kabir. Published by MDPI in 27 December 2021.

14. “An Indoor and Outdoor Navigation System for Visually Impaired People” by Croce D.,Giarré L.,Pascucci F.,Tinnirello I, Galioto G.E, Garlisi D, Lo Valvo A. in the year 2019.

15. “An Indoor and Outdoor Navigation System for Visually Impaired People” by Croce D.,Giarré L.,Pascucci F.,Tinnirello I, Galioto G.E, Garlisi D, Lo Valvo A. in the year 2019

16. “Indoor Mapping Based on Augmented Reality Using Unity Engine” by Ajith et al. (2020).

17. “Indoor Positioning and Navigation with Camera Phones” by Alessandro Mulloni, Daniel Wagner, István Barakony, Dieter Schmalstieg. Published by IEEE Pervasive Computing in April 2009

18. “Augmented Reality Awareness and Latest Applications in Education: A Review” by Soraia Oueida, Pauly Awad, Claudia Mattar [July 2023]. Published by International Journal of Emerging Technologies in Learning (iJET) 18(13):21-44.

19. “A marker-based augmented reality system for mobile devices” by Alexandru Gherghina, Alex Olteanu, Nicolae Tapus. Conference: Roedunet International Conference (RoEduNet), January 2013.

20. “Augmented Reality based Indoor Navigation using Point Cloud Localization” by Vishva Patel, Dr. Ratvinder Grewal. Published Online January 2022 in IJEAST.