



University of San Agustin
General Luna St., 5000 Iloilo City, Philippines
www.usa.edu.ph



COLLEGE OF TECHNOLOGY

**World Navigation Hat - Development of a Wearable Navigation Aid using AIoT for
the Visually Impaired**

An Undergraduate Design Project Proposal Presented to Faulty of the Computer
Engineering Department of College of Technology University of San Agustine

In Partial Fulfillment of the Requirement of the Course

CPE 413 - CpE Practice and Design I

By:

Daywan, Vince Ginno B.

D'Souza, Jason C.

Pabito, Ethel Herna C.

Wang, ChenLin

Engr. Glenda S. Quanzon

November 14, 2025



COLLEGE OF TECHNOLOGY

Table of Contents

Title Page	i
Table of Contents	ii
List of Figures	iv
List of Tables	v
Introduction	1
Background of the study	1
Rationale	3
General Objective	4
Specific Objectives	4
Significance of the study	5
Conceptual Framework	7
Theoretical Framework	8
Scope and Delimitation	10
Review of Related Literature	11
Materials and Methods	28
Research Design/Methods	28
Materials and Requirements	31
Data Presentation and Analysis	39



University of San Agustin
General Luna St., 5000 Iloilo City, Philippines
www.usa.edu.ph



COLLEGE OF TECHNOLOGY

Ethical Considerations	76
References	80
Appendices	I
Researchers and Adviser's Profile	I
Total Budget	V
Gantt chart for Project Proposal Timeline	VI
.	VI



COLLEGE OF TECHNOLOGY

List of Figures

1	IPO Model of the Conceptual Frame of the Study	7
2	Engineering Design Process diagram	29
3	Block diagram of the System	39
4	Architecture of the System	41
5	3D model of the system	43
6	Complete Circuit Diagram of the System	45
7	Wireframe of the Prototype	48
8	Measurements of the Prototype	49
9	Complete Schematic Diagram of the System	50
10	Camera 1 Schematic Diagram	51
11	Camera 2 Schematic Diagram	52
12	Flowchart of the three major pipelines	63
13	Flowchart of Server-sided Interface	66
14	Flowchart of Server's API	69



COLLEGE OF TECHNOLOGY

List of Tables

1	Differentiation table in Related Literature in Relation to the Study	22
2	Existing Publications from 2019–2025	23
3	Table of pin connections	53
4	Confusion Matrix of Hand Gesture Recognition Results	74
5	Performance Evaluation Metrics for Hand Gesture Recognition	75
6	Total budget	V
7	Gantt chart Aug 2025 - Nov 2025	VI
8	Gantt chart Dec 2025 - Mar 2026	X



COLLEGE OF TECHNOLOGY

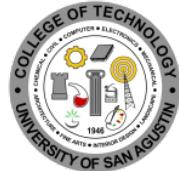
Introduction

Background of the study

Visual impairment is a global health problem that significantly affects individual's daily lives by impeding independent navigation, social interaction, and overall quality of life (Theodorou et al., 2023). There is an estimation of 2.2 billion people globally who are identified as visually impaired and this number could still increase to 2.5 billion by 2050 as stated by the World Health Organization (World Health Organization, 2023). In the Philippines, 2.17 million Filipinos are identified as visually impaired as quantized by reports from the Philippines Eye Research Institute (PERI) and Department of Health (DOH) (Shinagawa Lasik & Aesthetics, 2025). Visual impairment does not pertain to total blindness. According to World Health Organization, 2023, visual impairment can be identified and categorized based on the presenting visual acuity

- (1) Mild Vision Impairment, visual acuity is better than 6/18
- (2) Moderate Vision Impairment, visual acuity is worse than 6/18 but better than 6/60
- (3) Severe Vision impairment, visual acuity is worse than 6/60 but better than 3/60
- (4) Blindness, visual acuity is worse than 3/60
 - (4.a) Blindness with light perception, individuals can only perceive light
 - (4.b) Total Blindness, individuals who have no light perception

Other common types include astigmatism, near-sightedness (myopia) and far-sightedness (hyperopia), which account for a large share of impairments, alongside conditions like



COLLEGE OF TECHNOLOGY

cataracts, glaucoma, and age-related macular degeneration (AMD), with AMD alone affecting 8.06 million people globally in 2021 (Yon, 2025).

Navigation for visually impaired individuals increasingly relies on assistive devices that utilize sensory substitution, converting visual information into auditory or tactile cues to enhance spatial awareness (Skulimowski, 2025). These devices often incorporate IoT sensors to capture real-time environmental data, providing navigational assistance in complex environments (Real & Araújo, 2023) (Mohamadi et al., 2024). In addition, visually impaired individuals also face safety concerns and challenges such as accidents, falls, and collisions, as well as difficulties with navigation, including road crossings and destination location (Ikram et al., 2024; Gao et al., 2025; Muhsin et al., 2023). Beyond these practical difficulties, social isolation and reduced access to information further compound the challenges, often limiting educational and employment opportunities (Arvind, 2023). Current existing solutions or aid for visually impaired individuals are guide dogs, white canes, and electronic travel aids (Muhsin et al., 2023).

Sensory substitution devices (SSDs) allow users to perceive their environment by converting sensory information from one modality to another, particularly aiding those with visual impairments (Mishra et al., 2025). These devices help with obstacle detection, navigation, and object recognition, enhancing independence for users relying on traditional aids (Tokmurziyev et al., 2025). However, global adoption is limited by issues such as cognitive overload, extensive training needs, ergonomic discomfort, and the lower processing bandwidth of non-visual senses compared to vision (Hou et al., 2025).



COLLEGE OF TECHNOLOGY

Rationale

Visual Impairment encompasses a spectrum of conditions, from moderate vision loss to complete blindness, which significantly impacts individual's ability to perceive their environment and perform daily activities (Kumar et al., 2025). This condition also affects the individual's social interaction and overall quality of life (Que et al., 2025). More than 2.17 million people in the Philippines live with visual impairment, which restricts their perception, mobility, navigation, and accessibility to information—depriving them of work opportunities and independence (Chavarria et al., 2025). To mitigate the challenges encountered by the VIP, assistive technologies and rehabilitation strategies have been developed (Skulimowski, 2025). Traditional assistive aids like white canes offer limited feedback, initiating the development of electronic travel aids that deliver richer environmental information and improve autonomous navigation (Chandra et al., 2025) (Kim, 2024). Another solution is a wearable SSD, an assistive technology that an individual can wear on their body to help users navigate both their physical surroundings and digital applications, and emerging as a potential avenue to restore or gain a sense of spatial perception or awareness and navigational ability (Skulimowski, 2025). Additionally, SSDs promotes social interaction that aligns with one of the United Nations' Sustainable Development Goals for health and equality (Xue et al., 2025).

In previous studies, SSDs have been limited in their applicability, as existing devices often perform poorly in real-world situations, encountering cognitive overload problems that overwhelm users with excessive information, making navigation difficult (Casanova et al., 2025). Another identified concern is that the SSD design is uncomfortable, as it is bulky and requires extensive training, which limits its practicality (Olaosun et al., 2024).



COLLEGE OF TECHNOLOGY

In addition, many of these devices lack digital integration as capabilities, focusing mainly on obstacle detection and not connecting well with digital tools (Makati et al., 2024). These deficiencies underscore the pressing need for higher SSDs that can outperform and enhance existing devices. Therefore, a need for development of optimized smart navigation device improves the functional efficiency and performance of SSDs without overlooking the previous concerns like overloading data processing.

In this study the researchers aim to address these matters through the development of a wearable navigation hat integrated with AIoT, which introduces a smart navigation hat that uses 3D point cloud data to give audio feedback and connects to IoT services for navigation in both physical and digital environments. Furthermore, it enhances sensory processing, focusing on reducing cognitive overload in the system and making navigation simpler for VIP users. Additionally, the hat features customizable options, including a modular operating system that allows users to choose between gesture or voice commands.

General Objective

To develop an AIoT visual sensory substitution hat device that converts real-time environment information through emulating human sensory processes into audio cues for the visually impaired people (VIP)

Specific Objectives

- (1) To develop a prototype main system that converts environmental information to audio cues that includes:
 - (1.a) A photogrammetry (2D to 3D) system using dual OV5640 stereo cameras and Inertial Measurement Unit (IMU) to generate depth maps and point cloud



COLLEGE OF TECHNOLOGY

calculatio

- (1.b) Toggable modes to prioritize what information is cued into audio
- (1.c) Allows insertion anywhere in the pipeline for more custom preferences
- (2) To test the accuracy and speed of the AI models
 - (2.a) Determine the frames per seconds of AI or Image processes
 - (2.b) Compare depth predictions with actual depth
 - (2.c) Failure cases of Language Models fulfilling request
- (3) To develop the device to be portable and comfortable to the user
 - (3.a) The device is powered through UPS allowing portability and chargeable
 - (3.b) All packed into a hat to avoid putting strain on sensitive areas like eyes, nose, and ears
 - (3.c) Using the device becomes subconscious to the user's experience

Significance of the study

The proposed AIoT navigation hat has the potential to escalate the performance of existing assistive devices by:

- (1) Enhancing independence and mobility
- (2) Improving cognitive comfort and ease of use
- (3) Bridging physical and digital accessibility participation
- (4) Promoting Social Inclusion and Employment Opportunities for the VIP



COLLEGE OF TECHNOLOGY

In addition, this system innovates the research for Assistive Technology in the Philippines.

- (1) VIP can navigate safely and confidently with less assistance
- (2) The system simplifies understanding spatial information, reducing mental strain (cognitive overload) and allowing quick interpretation without long training
- (3) With AIoT integration, users can access navigation apps and online services hands-free using voice or gesture commands
- (4) As VIP gain independence, they can engage more in education, work, and community activities, boosting their self-esteem and reducing reliance on others

The project's success may inspire further studies on brain-inspired computing and wearable assistive devices, paving the way for more innovative tools. This project device will specifically benefit to the following:

- (1) **Visually Impaired and Blind Individuals** - They are the primary beneficiaries, gaining increased mobility, safety, and access to digital and social spaces through sensory substitution and IoT assistance.
- (2) **Families and Caregivers** - With the device promoting user independence, caregivers will experience reduced physical and emotional strain while maintaining peace of mind about the user's safety.
- (3) **Medical and Rehabilitation Specialists** - Eye health professionals and rehabilitation centers can use the system as a tool for sensory training and mobility rehabilitation programs, especially in fill in aspects the patient lacks or have trouble with.



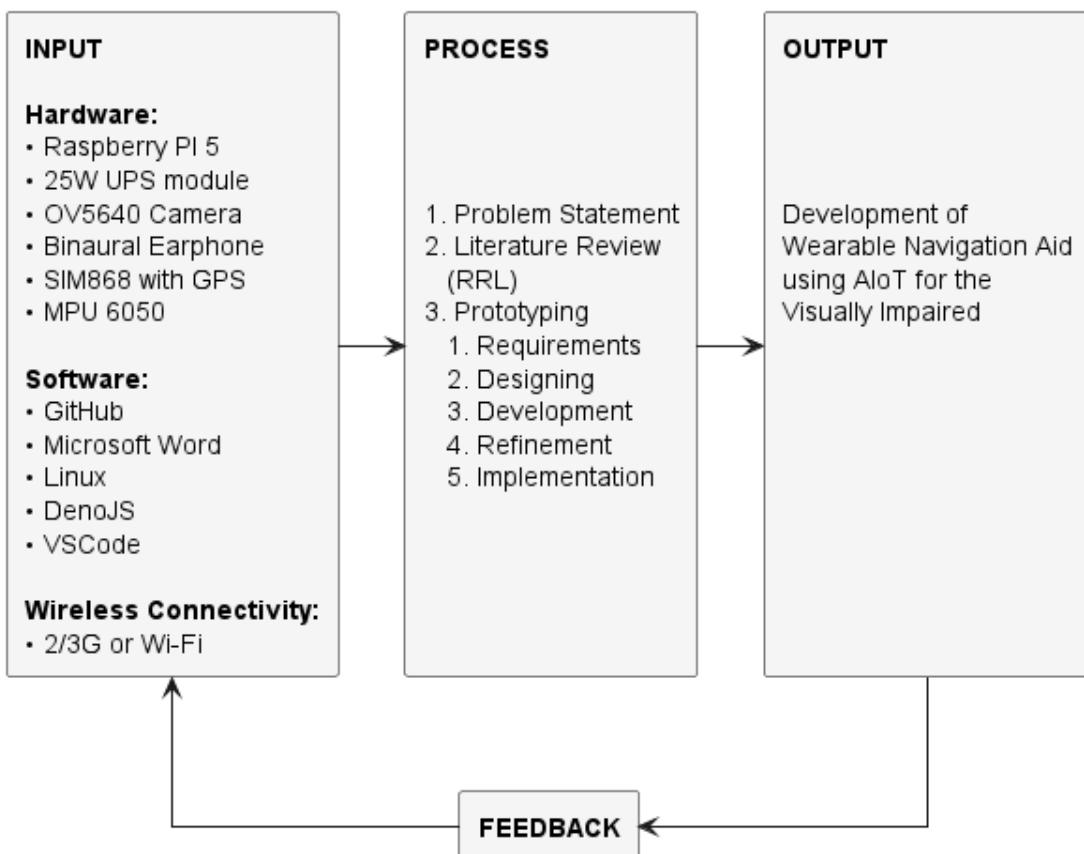
COLLEGE OF TECHNOLOGY

(4) Researchers and Engineers in Assistive Technology - The project offers a novel framework that combines virtual world modeling, IoT, and sensory emulation, providing a foundation for future innovations in human-computer interaction and embedded systems.

(5) Government and NGOs for Disability Support - Organizations involved in disability welfare can adopt or fund similar low-cost solutions to support national accessibility programs and meet inclusive development goals.

Conceptual Framework

Figure 1: IPO Model of the Conceptual Frame of the Study





COLLEGE OF TECHNOLOGY

Figure 1 states the conceptual framework of the research in the form of an Input-Process-Output (IPO) model. Starting with three kinds of required input such as the hardware components Microcontroller, Sensors, Actuators, and SIM Module, as the Software for Project management, Client-side device, and Server-side device, and the Wireless Connectivity used on to the device which can be either 2/3G used for deployment doors, or Wi-Fi as used in deployment indoors or development. The method of the research mimics typical development cycles with addition to literature review and research, such as literature review, prototyping, designing, testing/evaluating, and implementation. By the end of each iterative development the researchers expects to be closer to the end goal, with feedback from users and professionals to start the cycle again until the end goal of a wearable navigation aid is developed.

Theoretical Framework

This research attempts to solve an issue facing most Visual Sensory Substitution Devices which is primarily about the issue on translating a high bandwidth visual information to low bandwidth audio information without causing overstimulation, our novel solution is the emulate the human sensory mechanism to carry the processing load from the user and ensure that substituted information is minimal and nessary to the user. The theoretical principles includes:

- (1) **Focus and Spatial Resolution (Foveation)** - The human vision has a tiny high-resolution fovea covering around 1 to 2 degrees of the visual field but accounts for around half of visual cortex (Krantz, 2012), outside this region is a much coarser peripheral vision where acuity drops rapidly (Iwasaki & Inomata, 1986).

Our device attempts to mimic this by using a focus point system where the user can



COLLEGE OF TECHNOLOGY

contain the focus radius indicanting that area of the environment the user wants to be translated into audio and at what detail.

(2) **Selective Attention and Daliency** - Has the visual system cannot process all details at once, it selectively attends to salient or task-relavant features, this is done by implementing a bottom-up saliency and top-down goals to filter out redundant/irrelevant information (Kristjánsson et al., 2016). This is to mean that people focus on key objects ignoring uniform backgrounds as the nervous system "tunes out" repeated stimuli and amplifies novel/focused ones (Gershman, 2024).

Has the device translates environmental information, any stagnating/constant information must be tuned out over time leaving behind changes indicating motion.

(3) **Scene Gist and Gestalt Organization** - Human vision rapidly extract the gist (background) and figures (foreground) from a scene within the first fixation (36ms) with around 80% accuracy (Loschky, 2025). This process is done through Gestalt principles that groups elements to simplify complex images such as by similarity, proximity, common region, continuity, etc. (UserTesting, 2024)

This indicates that the device should be able to generally/primitively seperate the environmental data into background and foreground categories where foreground can then be seperated into groups, this is information to indicate the translated audio.

(4) **Parallel Motion vs. Detail Pathways** - Human visual systems process motion and detials in two seperate channels, the **magnocellular pathway** for fast motion and size but in less detail, and the **parvocellular pathway** for static object in more



COLLEGE OF TECHNOLOGY

detail (Zeki, 2015).

This suggest that when our device switches to motion detection it should prioritize the speed, size, and direction of that motion.

(5) Temporal Dynamics and Scanning - Vision is not a singular static snapshot but rather continuous sample of the world via eye movements has humans typically shift gaze 3-4 times a second (Kristjánsson et al., 2016).

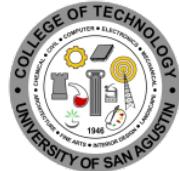
Thus our device should rapidly provide updating audio to allow temporal integration rather than one large static soundscape all at once, this also solves the issue of cognitive overload and sensory fatigue.

(6) Multisensory Integration - The brain integrates additional senses like auditory, vestibular, and proprioceptive senses to form a coherent representation of the environment (Kristjánsson et al., 2016).

Thus our device should align with those senses to prevent conflicting senses that could often cause nausea. This could be in a form of aligning the virtual environment with IMU to align with vestibular senses, or lower the output audio when the user is focusing on something else like talking to others.

Scope and Delimitation

The research focuses on Sensory Substitution Devices, more specifically on Vision Substitution. Our device aims to fill in what different variety of visually impaired individuals lacks, like if fully blind our device should be able to act as an artificial eye providing information the person wants, or if the person is near sighted it could read text from afar from the user and warn of objects rapidly approaching the user, and so on.



COLLEGE OF TECHNOLOGY

It does this by having a modular system running the main pipeline, the main pipeline inputs images from the stereo camera, creates a point cloud, maps it to a virtual physical map, follows an algorithm to emulate human vision processes, and finally generate the necessary audio cues for the individual. This pipeline allows applications to take into account more varied needs such as hand gesture recognition to control the parameters of the main pipeline so the person can navigate the device without seeing anything, image recognition to read text near or far, face recognition to recognize friends and family, and even more personalized interest like keeping track of expenses.

Review of Related Literature

Zvorășteanu et al. (2021) presented the **Sound of Vision (SoV) system**. It is a wearable SSD designed to aid VIPs in spatial cognition and navigation, using stereo-vision-based 3D reconstruction to interpret the environment and translate spatial data into audio and haptic cues. The device integrates a stereo camera, an infrared depth sensor, and an inertial measurement unit (IMU) to track the motion of the user's head and the installed camera mounted on the headgear, using these as its inputs. Furthermore, the output audio and haptic feedback are delivered via headphones and a haptic belt. The device processes visual data in real time using GPU-accelerated computation to reconstruct the 3D Environment and identify obstacles via segmentation algorithms based on disparity and histogram analysis. The depth of each pixel is determined by the stereo disparity formula $Z = \frac{f \cdot B}{d}$, where Z is the depth, B is the baseline, f is the focal length, and d is the disparity. The SoV converts environmental data into audio-haptic cues, allowing users to detect obstacles, open spaces, and open areas. In testing the device's performance with the visually impaired participants, it showed a 10% improvement in depth accuracy



COLLEGE OF TECHNOLOGY

and an 88.5% task success rate compared to standard stereo vision methods. The results highlight the device's effectiveness in real-world navigation and its potential to enhance mobility and spatial awareness compared to traditional aids such as white canes.

Sami et al. (2025) developed a **smart wearable (hat) assistive device** that integrates object detection with voice-assisted navigation to support VIP in real-time spatial awareness. The device uses an ESP32-CAM module to capture live video and transmit it via Wi-Fi to a mobile application that employs a TensorFlow-based deep learning model for efficient object recognition. The camera-detected objects are then processed into audio prompts that send alerts to the user when there are nearby obstacles and provide navigation directions. The programmed system produced a seamless object-detection-to-voice pipeline. The Smart Hat's system architecture demonstrates the integration of an IoT processing method to enable wireless communication between the smart hat and the user interface for remote processing and real-time updates for the user. Testing results show an object detection accuracy of over 91%, highlighting the reliability in dynamic environments. In addition, the Smart Hat device offers compact, low-cost, and user-friendly features. It demonstrates how IoT-enabled computer vision can transform traditional assistive technologies into intelligent, voice-guided navigation tools for the visually impaired.

RealSense AI (2025), in collaboration with Eyesynth, developed a wearable sensory-substitution device designed to enhance spatial awareness and independent navigation for visually impaired users by converting real-time 3D spatial data into bone-conduction audio cues. The device is called **Non-Invasive Image Resynthesis into Audio (NIIRA)**. The device employs the RealSense D415 depth camera to capture real-time 3D depth data



COLLEGE OF TECHNOLOGY

calculated using the stereo disparity formula $Z = \frac{f \cdot B}{d}$. Then it is processed to generate a point cloud representation of the surroundings, where each depth pixel being mapped to 3D coordinates using the equation $X = (u - c_x) \frac{z}{f_x}$, $Y = (v - C_y) \frac{Z}{f_y}$, $Z = Z$ to provide a complete spatial surrounding rather than a 2D view. The point-cloud data conjoint with the inertial and head motion tracking, is then processed through an onboard ASIC and Simultaneous Localization and Mapping (SLAM) modules to identify object orientation, translating them into audio signals such as pitch (height), volume (distance), and stereo panning (lateral position), allowing users to perceive objects and obstacles up to 5 meters away without external infrastructure. In testing NIIRA, it demonstrated improved independence and navigation in both indoor and outdoor environments, adapting to users' preferences, and displayed high obstacle-detection accuracy, minimal processing latency, and enhanced independent navigation. The participants reported that the bone-conduction audio feedback provided a clear environmental awareness without blocking external sounds.

Udayakumar et al. (2025) designed a **Smart Vision Glasses (SVG)** that processes environmental input data using AI, LiDAR depth mapping, and computer vision algorithms. The device's front-facing camera captures visual data, while the LiDAR sensor measures real-time depth information using the principle of Time-of-Flight (ToF) – where the distance (D) is calculated using the formula $D = \frac{c \cdot t}{2}$, where c represents the speed of light and t the time taken for emitted light to return after reflection. These depth and image data are then processed through an AI-based recognition model that uses Convolutional Neural Networks (CNNs) to classify objects, text, and faces within the user's field of view. The recognized elements are subsequently analyzed and prioritized

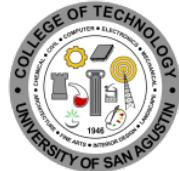


COLLEGE OF TECHNOLOGY

based on proximity and relevance using the LiDAR-assisted spatial mapping, which produces a semantic understanding of the scene; implementing Volume of Interest (VOI), a controllable spatial region or focus radius that limits feedback, to prevent sensory overload, enhancing user focus in dynamic environment. This processed information is then transmitted to a smartphone-based application converting the recognized data into voice audio cues through text-to-speech (TTS) engine. The device applies natural language processing (NLP) principles to ensure that the audio output is contextually meaningful and user-friendly. The input-output process follows a sequential pipeline:

- (1) capture of the image depth data
- (2) AI-driven detection and classification
- (3) distance estimation through the ToF equation
- (4) conversion of results into voice audio output describing nearby objects, text, or faces

The SVG through hearing supports the four primary modes – “Thighs Around You”, “Reading”, (Walking Assistance), and “Face Recognition”. The SVG is also equipped with gesture and voice control interfaces, allowing users to switch modes or issue commands hands-free, enhancing accessibility and user interaction. A multicenter usability study was conducted across five rehabilitation centers in India, involving 90 participants with a mean age of 23.5 years, to test the device functionality. The participants tested the four primary modes by completing real-world navigation and recognition tasks, and their feedback on the device’s usability and helpfulness was also recorded. Results showed that 72.9% indicated that ”Reading” mode is helpful, 44.7% for “The Things Around

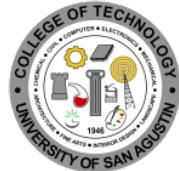


COLLEGE OF TECHNOLOGY

You”, 36.5% for “Face Recognition”, and 22.4% for “Walking Assistance”, showing that the device effectively enhances environmental awareness, reading ability, and object identification.

Ruan et al. (2025) developed a multifaceted sensory substitution wearable device that uses an audio-based curb detection to improve real-time awareness and navigation safety for individuals who are blind or have low vision. The device integrates a stereo camera, ultrasonic range sensors, and inertial motion units (IMUs) to identify curbs and ground-level transitions. Adapting the stereo disparity formula $Z = \frac{f \cdot B}{d}$, the system calculates the curb height and distance, while the ultrasonic sensors confirm the surface proximity of the environment for redundancy and accuracy. The sensory data are processed and converted into audio and vibrational cues, where the pitch and repetition rate of the sound correspond to the curb’s distance and height, providing intuitive, time-sensitive feedback. In the pilot testing, 12 participants were involved, 8 with blindness and 4 with low vision. The participants tested the navigation feature of the device in both indoor and outdoor test environments, including simulated sidewalks, ramps, and descending curbs. After testing, the performance statistics of the device showed an average curb detection accuracy of 94.3%, with a false-negative rate of 3.7%, and an average alert response time of 1.2 seconds, indicating the system’s reliability in detecting and signaling curbs in real time. Participants reported that it also improved confidence in mobility and reduced the risk of missteps or trips. This demonstrates that the audio-based curb detection approach driven by LiDAR-like depth estimation and multimodal feedback effectively enhances real-time alerting and safe navigation for the VIP.

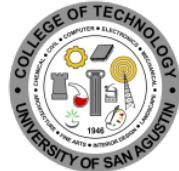
Viancy V et al. (2024) introduced AuralVision, a wearable assistive device that is



COLLEGE OF TECHNOLOGY

designed as eyewear to improve navigation for visually impaired individuals by incorporating object detection, scene classification, and reinforcement learning. The device's system architecture includes a camera sensor to capture the user's environment, image-processing and object-detection software to classify obstacles and identify significant objects, and an audio module to translate detection results into auditory cues for navigation. Utilizing a convolution neural network (CNN) approach or deep learning-based vision algorithms trained on the Object Net 3D dataset to interpret the surrounding environment and convert the visual information to auditory cues that convey spatial awareness. AuralVision uses 3D visualization and sound-based feedback to identify objects, classify road scenes, and assist users with obstacle avoidance and pathfinding. The device adapts to dynamic environments through continuous learning, thereby improving navigation decision-making.

Hamilton-Fletcher et al. (2021) developed a mobile sensory-substitution device called SoundSight. This device translates input color, depth, and temperature data into output audio cues, allowing the visually impaired users to sense their environment. The system captures environmental input using a mobile with an RGB-D camera and a thermal sensor. Then it processes the data stream through a multi-feature mapping pipeline before integrating it into a composite audio stream. Depth is calculated using the stereo disparity formula to determine the distance of the object from the user's perspective. At the same time, the RGB camera's color hue values are mapped to sound pitch using the equation $f_c = k_c \times H$, where H represents the hue intensity and scaling constant. Temperature measurements are converted to amplitude modulation through $A_t = k_t \times T$, where T is the sensed temperature and k_t is a sensitivity coefficient. These parallel mappings



COLLEGE OF TECHNOLOGY

are integrated using the data fusion algorithms, synchronizing depth, color, and thermal inputs before generating the composite audio streams that encode environmental depth through rhythm, color through pitch, and temperature through loudness. In testing the device's functionality, 15 blind and 10 low vision participants are involved. The device demonstrated rapid user learning and over 85% recognition accuracy, demonstrating that the multi-feature sonification can reliably convey complex environmental information.

Han et al. (2026) designed a Multi-Path Sensory Substitution Device (MSSD) that enhances low-vision mobility and virtual navigation of the visually impaired people by integrating real-time depth mapping, IMU-based motion tracking, and depth optimization algorithms. The system processes input from a depth camera and inertial sensors to create a 3D spatial model of the environment, using the stereo disparity equation to calculate object distance and the projection model $X = (u - c_x)\frac{z}{f_x}$, $Y = (v - c_y)\frac{z}{f_y}$ to map each pixel to world spatial coordinates. After acquiring the data from the processed input, it generates a 3D point cloud, which then is converted into an occupancy grid for spatial awareness, while the IMU readings refine user orientation using the equation $\theta_t = \theta_{t-1} + \omega t \Delta t$. A modified A* (A-star) algorithm is used to compute the optimal navigation routes using the cost function $f(n) = g(n) + h(n)$, balancing actual distance and heuristic estimates to avoid obstacles. The selected virtual path is then translated into audio-haptic feedback, where the vibration intensity and sound frequency indicate direction and proximity, using speakers or headphones to create output data directional audio cues and vibration motors/haptic actuators for tactile feedback. To check the device's functionality, the researchers tested it with 20 low-vision participants. The system attained a result of 91% path following accuracy and a 38% reduction in navigation



COLLEGE OF TECHNOLOGY

errors. As the device demonstrates a satisfactory functional performance, this indicates that combining 3D mapping, motion sensing, and heuristic path planning effectively supports real-time, safe, and virtual navigation for the visually impaired individuals.

Commère and Rouat (2023) evaluated five depth-to-sound sonification methods to compare their performance and effectiveness. One of the methods evaluated is the LiDAR-to-repetition-rate sonification model. The identified sonification model converts visual information into rhythmic auditory cues, helping visually impaired individuals to develop a sense of spatial awareness through sound. Using the Time-of-Flight (ToF) equation $D = \frac{c \cdot t}{2}$, the model measures the distance of an object and maps it inversely to the beep repetition rate $R = \frac{k}{D}$, where R is the repetition rate (Hz), D is the distance (meters), and k is a scaling constant which is determined experimentally to maintain the perceptible tempo differences between near and far objects which results if an object is closer it generates faster repetition rates or beeps and far detected objects have slower rhythm beeps which the model's spectrogram design reflects this relationship. Whereas the temporal density (number of pulses per second) indicates proximity, while frequency and amplitude encode additional object features such as height and reflection intensity, creating a dynamic “rhythmic depth map” that represents distance through time spacing between sound bursts—in experimental testing with 28 sighted but blindfolded participants, a three-phase protocol consisting of Depth Estimation, Azimuth Estimation, and Retention Assessment confirmed that repetition-rate sonification produces the lowest mean absolute error (MAE) in distance perception and was the most intuitive and accurate among the other mapping methods like frequency or amplitude-based cues. The results finding revealed a strong inverse correlation between the perceived distance and repetition rate



COLLEGE OF TECHNOLOGY

($R \propto \frac{1}{D}$), indicating that LiDAR-driven rhythmic approach effectively translates the spatial depth into comprehensive sound patterns, offering an enhanced basis for real-time auditory sensory substitution systems for the visually impaired.

Zhao et al. (2025) conducted a clinical evaluation of an assistive device, a wearable electronic navigation aid (ENA), for visually impaired individuals through a prospective, non-randomized, single-arm, open-label trial involving 30 participants (each participant was given five trials for each functional task), selected through purposive sampling to ensure that individuals with blindness or severe visual impairment could assess the performance of the assistive device. The study primarily focused on the functional efficacy of the device, assessing how effectively it supports navigation and daily tasks of the visually impaired users. The assistive device used in testing integrates ultrasonic sensors, vibration motors, and audio output components. The study implemented a structured, protocol-guided or functional evaluation framework that combined quantitative measures including accuracy rate ($\frac{\text{Successful Trails}}{\text{Total Trails}} \times 100$), error rate ($\frac{\text{Navigation Errors}}{\text{Total Trails}} \times 100$), reaction time ($RT = T_{\text{response}} - T_{\text{stimulus}}$), task completion time ($TCT = T_{\text{end}} - T_{\text{start}}$), and performance improvement rate ($PIR = \frac{(Post\ test - Pre\ test)}{Pre\ test} \times 100$) to assess navigation accuracy, responsiveness, and efficiency objectively, supported by qualitative user feedback on comfort, usability, and confidence. The study reveals that the ENA device effectively improved safe navigation and mobility performance for visually impaired users.



COLLEGE OF TECHNOLOGY

Summary

Recent studies (2019–2025) on assistive technologies for the VIP highlight wearable and mobile devices utilizing a sensory substitution approach to convert visual/spatial data into audio, haptic, or multimodal cues, applying stereo vision (e.g., disparity formula $Z = \frac{f \cdot B}{d}$), depth sensors like LiDAR/ToF (e.g., $D = \frac{c \cdot t}{2}$), AI-driven computer vision (e.g., CNNs, TensorFlow), and IMUs. This approach expresses the outcome produced by the device such as real-time obstacle detection, object recognition, and navigation, similar to the Caraiman et al.'s Sound of Vision (SoV), which addresses sensory overloading via a controllable focus radius (VOI) and modular components for partial generalizability; Sami et al.'s Smart Hat for cost-effective voice output; RealSense's NIIRA point-cloud processing to semantic audio; Commère and Rouat's LiDAR sonification for rhythmic depth cues; Hamilton-Fletcher et al.'s SoundSight for mobile timbre/volume mappings; Ruan et al.'s curb detector for 94.3% accuracy; Han and Li's Multi-Path Sensory Substitution Device (MSSD) for 91% path accuracy; Zhao et al.'s electronic navigation aid (ENA); and Udayakumar and Gopalakrishnan's Smart Vision Glasses (SVG) for AI-LiDAR voice modes. However, systems like AuralVision that utilizes ToF lasers for bone-conduction stereo sound, lab-effective but range-limited, Eyesynth's NIIRA object recognition process prone to overload without human emulation, Depth Sonification repetition rates for intuitiveness, SoundSight LiDAR/thermal to audio hwich is app-constrained, Smart Hat that is non-modular, and multi-faceted SSDs with 85% curb accuracy often face issues like sensory overload, training demands, limited real-world adaptability, and lack of digital-physical navigation. These studies proposes an idea that work closely aligns with SoV by tackling sensory overloading through VOI and



COLLEGE OF TECHNOLOGY

modular components but advances it by explicitly mimicking human perception via a modes system and a modular IoT pipeline that applications can embed for navigation in both physical and digital worlds, addressing SoV's training needs and real-world performance gaps while incorporating brain emulation for superior overload reduction and versatility, building on these predecessors to offer a more intuitive, scalable solution for VIP independence. In addition, evaluations across studies ranging from 12 to 90 participants showcase 85% to 95% accuracies and qualitative benefits like reduced errors (e.g., 38% in MSSD). These studies introduce an innovative and improved SSD by adding brain emulation, AI, and IoT for advanced overload reduction and versatility.



COLLEGE OF TECHNOLOGY

Table 1: Differentiation table in Related Literature in Relation to the Study

Component	R1	R2	R3	R4	R5	R6	R7	R8	R9	Ours
Zero form factor						✓	✓			✓
SBCs										
LiDAR Sensor		✓	✓			✓	✓	✓		
Depth RGB Camera	✓	✓	✓	✓	✓		✓			✓
IMU	✓	✓		✓	✓	✓	✓			✓
Computer Vision / AI	✓	✓	✓	✓	✓	✓	✓	✓		✓
Audio Output	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Image Recognition	✓	✓	✓	✓	✓	✓	✓	✓		✓
Mobile Wearable Form	✓	✓	✓	✓	✓		✓	✓		✓
Point Cloud Projection & 3D Mapping		✓				✓				✓

(1) Zvorișteanu et al. (2021)

(2) RealSense AI (2025)

(3) Udayakumar et al. (2025)



COLLEGE OF TECHNOLOGY

(4) Ruan et al. (2025)

(5) Viancy V et al. (2024)

(6) Hamilton-Fletcher et al. (2021)

(7) Han & Li (2025)

(8) Commère and Rouat (2023)

(9) Zhao et al. (2025)

Table 2 : Existing Publications from 2019–2025

Patent Title	Authors	Parameters	Publica-tion No.	Publi-cation Date
Stereo Vision-Based Sensory Substitution for the Visually Impaired	Simona Caraiman, Ovidiu Zagan	Stereo vision cameras, IMU integration, depth computation ($Z = \frac{f \cdot B}{d}$), obstacle-to-audio/haptic mapping	PMC6630 569	June 20, 2019

Continued on next page



COLLEGE OF TECHNOLOGY

Table 2: Existing Publications from 2019–2025 (Continued)

Patent Title	Authors	Parameters	Publica-tion No.	Publi-cation Date
Evaluation of Short-Range Depth Sonifications for Visual-to-Auditory Sensory Substitution	Louis Commère, Jérôme Rouat	LiDAR depth sensing, LiDAR-to-repetition rate mapping ($R = \frac{k}{D}$), temporal density sonification	arXiv:2304.05462	April 11, 2023
Aural Vision – Envisioning the Independence of the Visually Impaired	V. V., N. R. C. Mouli, M. M., S. S.	Dual cameras, IMU sensors, reinforcement learning, ObjectNet3D dataset, visual-to-audio translation	IEEEDOI: 10.1109/I CICT6015 5.2024.10 545004	April 2024

Continued on next page



COLLEGE OF TECHNOLOGY

Table 2: Existing Publications from 2019–2025 (Continued)

Patent Title	Authors	Parameters	Publica- tion No.	Publi- cation Date
SoundSight: A Mobile Sensory Substitution Device That Sonifies Colour, Distance, and Temperature	Giles Hamilton-Fletcher, J. Arias	RGB-D camera, thermal sensor, multi-feature translation (color, depth, temperature), formulas ($f_c = k_c H$), ($A_t = k_t T$)	DOI:10.1007/s12193-021-00376-w	July 2, 2021
Multi-Path Sensory Substitution Device for Navigation	Zaidao Han, S. Li	Depth camera, IMU, A* pathfinding ($f(n) = g(n) + h(n)$), 3D occupancy mapping, audio-haptic feedback	ScienceDirectS0141938225002379	August 27, 2025

Continued on next page



COLLEGE OF TECHNOLOGY

Table 2: Existing Publications from 2019–2025 (Continued)

Patent Title	Authors	Parameters	Publica-tion No.	Publi-cation Date
Evaluation of the Efficacy of an Assistive Device for Blind People	Yiming Zhao, S. Y.	Functional efficacy evaluation, accuracy ($\frac{\text{Successful}}{\text{Total}} \times 100$), error rate, reaction time, 5-trial testing, 30 participants	Taylor&Fr ancisOnlin eDOI:10.1 080/02713 683.2025. 2495212	April 29, 2025
AI-Powered Smart Vision Glasses for the Visually Impaired	Devi Udayakumar, S. Gopalakrishnan	LiDAR, RGB camera, AI fusion, gesture/voice control, mobile app integration	PMC1217 8407	May 30, 2025
Multi-Faceted Sensory Substitution Using Wearable Technology for Curb Alerting	Ligao Ruan, G. Hong, R. Fernández-Robles	Stereo disparity ($Z = \frac{f \cdot B}{d}$), ultrasonic sensors, audio-based curb detection, 94.3% accuracy	DOI:10.10 80/174831 07.2025.24 63541	February 1, 2025

Continued on next page



COLLEGE OF TECHNOLOGY

Table 2: Existing Publications from 2019–2025 (Continued)

Patent Title	Authors	Parameters	Publica-tion No.	Publi-cation Date
Efficient Object Detection and Voice-Assisted Navigation: The Smart Hat Approach	Memoona Sami, D. Abro	Raspberry Pi 4, ultrasonic sensor, camera, object detection-to-voice IoT system	SukkurIB AJournal,S J-CMS	December 2024
NIIRA: Non-Invasive Image Resynthesis into Audio (Sonic Vision for the Blind)	RealSense & Eyesynth Team	RealSense D415 camera, point-cloud depth mapping, SLAM, bone-conduction sound, obstacle detection up to 5 m	IntelRealSenseCaseStudy	May 2025

All entries are peer-reviewed or pre-print publications.



Materials and Methods

Research Design/Methods

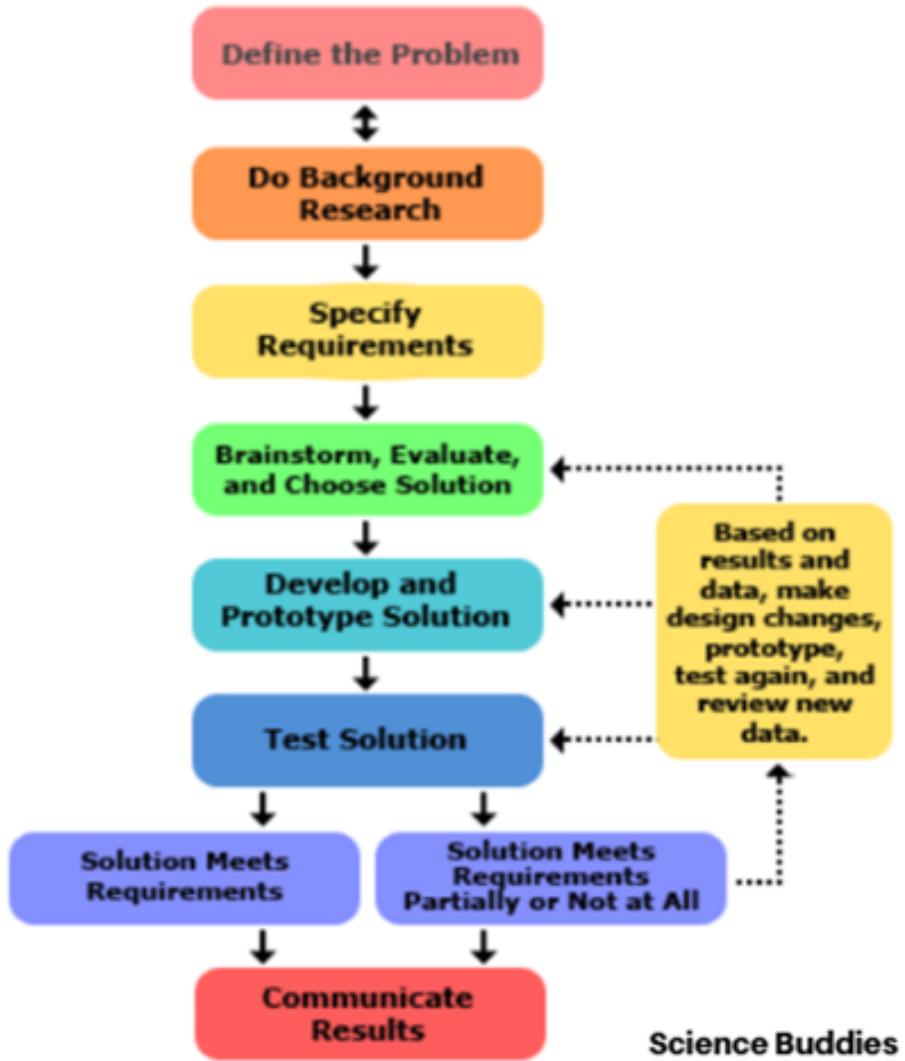
This research follows a mixed method design in data collection such as using quantitative data in measuring the accuracy and speed of the device in conveying information, and qualitative data in measuring the comfortability and acclimation of users towards using the device.

The study to develop a wearable AIoT navigation hat for visually impaired individuals (VIP) will use the Engineering Design Process (EDP), a systematic, iterative framework adapted from sources such as Science Buddies, to guide the project's methodology from problem identification to prototype optimization. This approach ensures a structured path for designing, building, and testing the device, incorporating user feedback and ethical considerations to effectively address VIP needs.



COLLEGE OF TECHNOLOGY

Figure 2: Engineering Design Process diagram



Source: sciencebuddies.org

Identify the Problem and Constraints

Identify the key problems and constraints VIP faces, including difficulties navigating their environment, safety risks, and cognitive overload from existing aids such as white canes and basic sound-sensing devices. Identify the needs for advanced technology that can effectively navigate and replicate human vision without requiring extensive user training. Establish the requirements for the hat and outline the project's necessary functions,



COLLEGE OF TECHNOLOGY

including converting 3D environmental data into audio cues and ensuring portability through features such as battery power and ergonomic design. Constraints such as budget considerations, available components (e.g., the Raspberry Pi Zero 2W or other zero-form-factor single-board computers), and ethical guidelines (e.g., obtaining Institutional Review Board (IRB) approval). The focus remains on specific VIP challenges such as social isolation, navigation, mobility, and employment barriers, to align with evaluating AI frame rate (FPS) and user adaptation.

Brainstorm Possible Solutions

Illustrate various possible ideas for wearable designs, drawing inspiration from existing technologies such as Sound of Vision and Smart Hat. Consider emphasizing the use of IoT and AI for photogrammetry and inertial measurement units (IMUs), with a creative approach to enhance human senses and thereby reduce cognitive load. The goal is to connect physical and digital experiences to benefit a broader range of users. Select a Solution and Develop a Design Select the best design for implementation. This entails creating a detailed development plan that includes component selection, circuit diagrams, prototyping timelines, and considerations for ethical data collection and risk management. The prototype will focus on creating a wearable hat that integrates AIoT, providing audio cues while ensuring comfort and customizability for VIP users.

Build a Prototype

In the prototyping stage, the researchers will assemble the components according to specifications and conduct initial tests to verify functionality. Document the entire process, including any challenges faced, to streamline improvements. The development will focus on a user-friendly, efficient, ergonomically designed prototype for VIPs,



COLLEGE OF TECHNOLOGY

integrating AIoT features for real-time feedback.

Evaluate the Prototype

Evaluation of the prototype involves conducting both quantitative and qualitative tests. This includes measuring AI FPS, depth accuracy, and other collected data, along with user trials to assess comfort and understanding. Feedback from focus groups and careful analysis of results will inform further refinements, helping identify and document any issues encountered during prototype testing.

Redesign and Optimize

Based on the feedback received, the prototype design will undergo iterative refinement. Changes may include enhancements to audio algorithms or modifications to extend battery life, ultimately striving to optimize the prototype for practical use by visually impaired individuals.

Materials and Requirements

The World Navigation Hat operates through a multi-stage processing pipeline that begins with data acquisition and culminates in audio feedback delivery. The Orange Pi Zero 2W microcontroller, powered by a 5V/3A UPS module, serves as the central processing unit that orchestrates all system operations. The process initiates with stereo image capture from dual OV5640 USB cameras positioned 10 cm apart, providing 5MP resolution at 160° field of view for photogrammetry operations. Concurrently, the MPU6050 Inertial Measurement Unit tracks head motion and orientation to align the virtual environment with the user's vestibular senses, preventing sensory conflicts. Audio input is captured through an earphone microphone for voice command recognition. Five primary applications run simultaneously on the device:



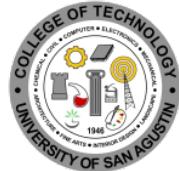
COLLEGE OF TECHNOLOGY

- (1) Photogrammetry processing that converts 2D stereo images into depth maps and 3D point clouds using stereo disparity calculations
- (2) Human visual sense emulation pipeline that implements foveation, selective attention, and scene gist extraction to generate necessary audio cues while minimizing cognitive overload
- (3) MediaPipe AI for real-time hand gesture recognition enabling touchless device control
- (4) WhisperX voice recognition activated by specific hand gestures for command input
- (5) modular framework allowing custom applications to integrate into the main pipeline with connection to the main server through standard APIs

Processed data is transmitted via the SIM868 module (supporting 2G/3G cellular or Wi-Fi connectivity) to a NodeJS server infrastructure that handles additional computational tasks including Ollama for large language model processing, YOLO for advanced object detection, Puppeteer for web automation, external API integration (Google Maps, Meta platforms), SQL database management, and HTTP server operations for client-server communication.

Hardware Material

The hardware architecture integrates several critical components optimized for portable wearable computing. The Orange Pi Zero 2W single-board computer features a quad-core ARM Cortex-A53 processor with 512MB/1GB RAM options, providing

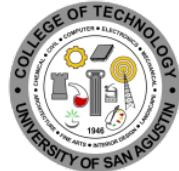


COLLEGE OF TECHNOLOGY

sufficient computational power for real-time computer vision tasks while maintaining low power consumption suitable for battery operation. Two OV5640 USB cameras serve as the stereo vision system, each delivering 5-megapixel resolution with a 160° field of view and positioned with a 10 cm baseline to enable accurate depth estimation through stereo disparity calculations. The MPU6050 6-axis Inertial Measurement Unit combines a 3-axis gyroscope and 3-axis accelerometer, providing motion tracking data at high refresh rates to synchronize the virtual environment with user head movements. The SIM868 development board integrates GSM/GPRS/GPS functionality, enabling both outdoor navigation through GPS positioning and cellular data connectivity (2G/3G) for server communication when Wi-Fi is unavailable. An Onyehn TRRS audio module facilitates bidirectional audio communication, connecting both speaker output for audio cues and microphone input for voice commands through standard 3.5mm earphone jacks. Power management is handled by a custom UPS module accepting 5V/3A input, featuring lithium battery charging circuits (supporting 18650 or similar rechargeable cells) and voltage regulation to ensure stable operation during mobile use. All components are mounted on a custom-designed hat frame measuring 333.38 mm in length with integrated cable management and component housings to maintain ergonomic comfort during extended wear.

Software Specification

The software ecosystem operates across two primary environments: client-side processing on the Orange Pi Zero 2W and server-side operations on a NodeJS platform. The client-side implementation utilizes Python 3.x as the primary programming language,



COLLEGE OF TECHNOLOGY

leveraging OpenCV (cv2) for image acquisition, preprocessing, and stereo vision algorithms including stereo calibration, rectification, and disparity map generation. NumPy provides optimized numerical computing for point cloud transformations, matrix operations in photogrammetry calculations (applying the stereo projection matrix Q to depth maps), and efficient array manipulations for real-time processing. Matplotlib enables visualization during development and debugging phases, generating spectrograms for audio cue analysis and 3D scatter plots for point cloud verification. MediaPipe framework implements hand gesture recognition through pre-trained machine learning models optimized for mobile deployment, detecting key hand landmarks and classifying gestures with minimal latency. PyTorch serves as the deep learning backend for custom neural network implementations, including depth estimation refinement models and audio generation networks that emulate human auditory processing. Additional Python libraries include SciPy for signal processing (audio synthesis, filtering), librosa for audio feature extraction, and pyserial for communication with peripheral modules (SIM868, MPU6050). The server-side infrastructure runs NodeJS with Express framework for HTTP server operations, handling REST API endpoints for client requests, managing WebSocket connections for real-time bidirectional communication, and coordinating database operations through SQL query execution. Ollama integration enables on-device large language model inference for natural language understanding and context-aware assistance, while YOLO (You Only Look Once) models provide high-speed object detection and classification when additional visual recognition is required beyond the client-side capabilities. Puppeteer automates web browser interactions, enabling the device to fetch real-time information from web services and interact with digital en-



COLLEGE OF TECHNOLOGY

vironments for enhanced navigation support. The development workflow integrates multiple productivity and collaboration tools to streamline project organization, documentation, and version control. Visual Studio Code (VSCode) serves as the primary integrated development environment (IDE), providing syntax highlighting for Python and JavaScript, integrated terminal access for remote SSH connections to the Orange Pi Zero 2W, Git integration for version control operations, and extensive extension support including Python linters (pylint, flake8), formatters (black, autopep8), and remote development extensions for direct editing on embedded systems. GitHub hosts the project repositories, implementing version control through Git for both client-side Python applications and server-side NodeJS code, facilitating collaborative development among team members through branch management, pull requests, code reviews, and issue tracking for bug reports and feature requests. LaTeX document preparation system generates the formal thesis documentation with precise mathematical typesetting for equations including stereo disparity formulas, point cloud projection matrices, and depth-to-audio mapping functions, ensuring professional formatting standards for academic submission while maintaining version-controlled .tex source files. Google Docs provides real-time collaborative editing for preliminary research documentation, literature review organization, meeting notes, and draft content development, enabling simultaneous multi-user editing with comment threads for peer feedback and revision tracking. Microsoft Office suite (Word, Excel, PowerPoint) supports supplementary documentation tasks including formatted research proposals, data analysis spreadsheets for experimental results (accuracy measurements, user feedback surveys, performance benchmarks), and presentation materials for design reviews and project demonstrations.

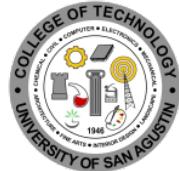


COLLEGE OF TECHNOLOGY

Claude AI assists throughout the development lifecycle as an AI-powered coding assistant and research tool, providing algorithm optimization suggestions, debugging support for complex stereo vision calculations, documentation generation for API endpoints and function specifications, literature synthesis for theoretical framework development, and technical writing refinement for thesis chapters. This integrated software toolchain ensures systematic project management from initial concept through final implementation, maintaining code quality through version control, facilitating team collaboration across distributed work environments, and producing comprehensive documentation that meets both academic standards and technical reproducibility requirements.

Library and Board Managers

Development and deployment require specific software dependencies managed through multiple package systems across different platforms. For the Orange Pi Zero 2W running Armbian or Ubuntu-based distributions, the APT package manager handles system-level dependencies including Python 3.x development packages (`python3-dev`, `python3-pip`), OpenCV system libraries (`libopencv-dev`, `python3-opencv`), audio system requirements (`libasound2-dev`, `portaudio19-dev`), and USB camera support utilities (`v4l-utils`). Python package management through pip3 installs critical libraries with specific version constraints: `opencv-python` ($\geq 4.5.0$) for computer vision operations, `numpy` ($\geq 1.21.0$) for numerical computing, `mediapipe` ($\geq 0.8.0$) for hand tracking AI models, `torch` ($\geq 1.10.0$) with ARM-compatible builds for deep learning inference, `sounddevice` and `soundfile` for audio I/O operations, `pyserial` (≥ 3.5) for hardware communication, and `requests` for HTTP client functionality. Arduino IDE or Platform IO manages firmware



COLLEGE OF TECHNOLOGY

for any auxiliary microcontrollers if implemented, including board definitions for custom boards and library dependencies for sensor communication protocols (I2C for MPU6050, SPI for peripheral modules). The NodeJS ecosystem uses npm (Node Package Manager) to install server-side dependencies: express ($\geq 4.18.0$) for web server framework, ws for WebSocket implementation, sequelize or knex for SQL database ORM, axios for HTTP client operations to external APIs, and child_process management for spawning Ollama and YOLO processes. Board-specific configurations require Orange Pi GPIO libraries (OrangePi.GPIO or wiringOP) for direct hardware pin control if needed, and kernel module loading for camera interfaces (ensuring proper v4l2 driver support for OV5640 USB cameras through modprobe configurations). Git version control manages source code across development iterations, with repositories configured for both client and server codebases, while systemd service configurations enable automatic startup of critical applications on boot for production deployment.

Other Materials/Equipment/Devices

Beyond the core electronic components, the project requires supplementary materials and equipment for fabrication, testing, and evaluation. The physical enclosure utilizes 3D-printed components designed in CAD software, with wireframe dimensions of 333.38 mm length, 197.50 mm width, and component-specific mounting features including camera housings with precise 10 cm spacing alignment, ventilated sections for heat dissipation from processing components, and cable routing channels. The 3D printing process uses PLA or PETG filament for structural components requiring strength and thermal stability, with print settings optimized for layer adhesion and surface finish.



COLLEGE OF TECHNOLOGY

to ensure user comfort during extended wear. A standard baseball cap or adjustable hat frame serves as the foundation for component integration, providing familiar ergonomics and adjustable sizing through standard strap mechanisms. Micro SD card (minimum 32GB Class 10 or UHS-I) stores the operating system, application code, and temporary data cache for the Orange Pi Zero 2W, while a micro SIM card with active data plan enables cellular connectivity through the SIM868 module. Lithium-ion battery cells (18650 format, 3.7V nominal, minimum 2500mAh capacity) power the UPS module, with appropriate battery holders and spot-welded connections ensuring reliable electrical contact and mechanical stability. USB-C cable provides charging interface for the UPS module, allowing users to recharge using standard mobile phone chargers or power banks. Testing and evaluation equipment includes a reference measuring device (laser rangefinder or calibrated ruler) for depth accuracy validation, comparing predicted distances from stereo disparity calculations against ground truth measurements across various ranges (0.5m to 5m typical operating distance). A standardized navigation course with marked obstacles, varying textures, and elevation changes facilitates user testing protocols for evaluating device effectiveness. Audio analysis equipment (reference microphone, audio interface, spectrum analyzer software) characterizes the generated audio cues for frequency response, dynamic range, and spatial positioning accuracy. Additionally, development workstations (laptop/desktop computers with Linux or Windows OS) run cross-compilation toolchains, debugging interfaces (SSH, serial console), and simulation environments for algorithm validation before hardware deployment. Safety equipment for user testing includes protective padding for obstacle courses, first aid supplies, and emergency communication devices in compliance with IRB-approved



COLLEGE OF TECHNOLOGY

testing protocols outlined in the ethical considerations section.

Data Presentation and Analysis

Figure 3: Block diagram of the System

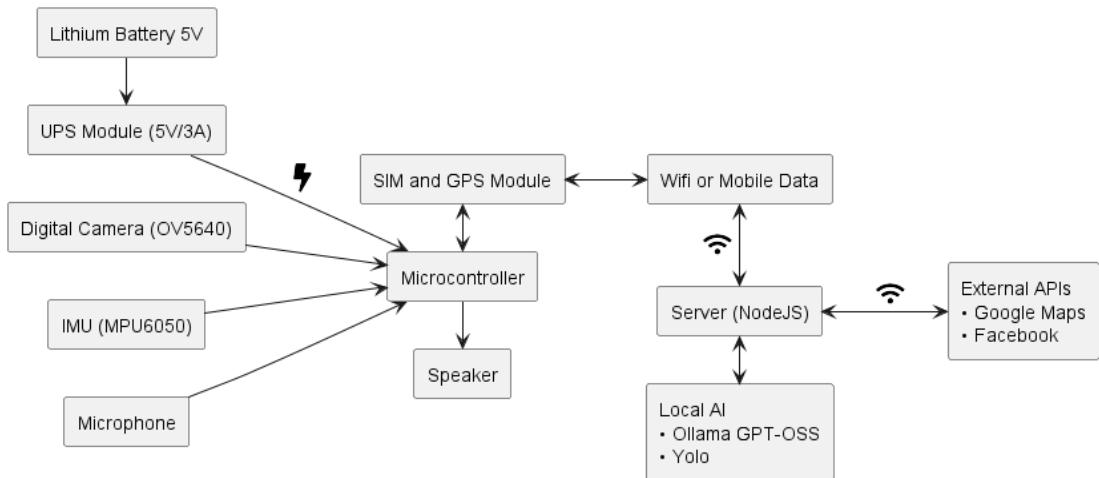


Figure 3 illustrates the simplified hardware-software interaction flow of the World Navigation Hat system, showing the essential component connections and data pathways. The architecture begins with power distribution from a 5V Lithium Battery that feeds into the UPS Module (5V/3A), which provides regulated power (indicated by a lightning bolt symbol) to the central Microcontroller (Orange Pi Zero 2W or similar embedded system). Multiple input sensors connect directly to the microcontroller: the Digital Camera (OV5640) captures stereo visual data for photogrammetry and depth mapping, the IMU (MPU6050) provides 6-axis motion tracking data for head orientation and movement detection, and the Microphone captures audio input for voice command recognition through WhisperX AI processing. The microcontroller outputs audio feedback through the Speaker, delivering spatial audio cues that convey environmental information to the visually impaired user. Bidirectional communication flows between the microcontroller and the SIM and GPS Module, allowing for real-time location updates and connectivity to external networks and services.



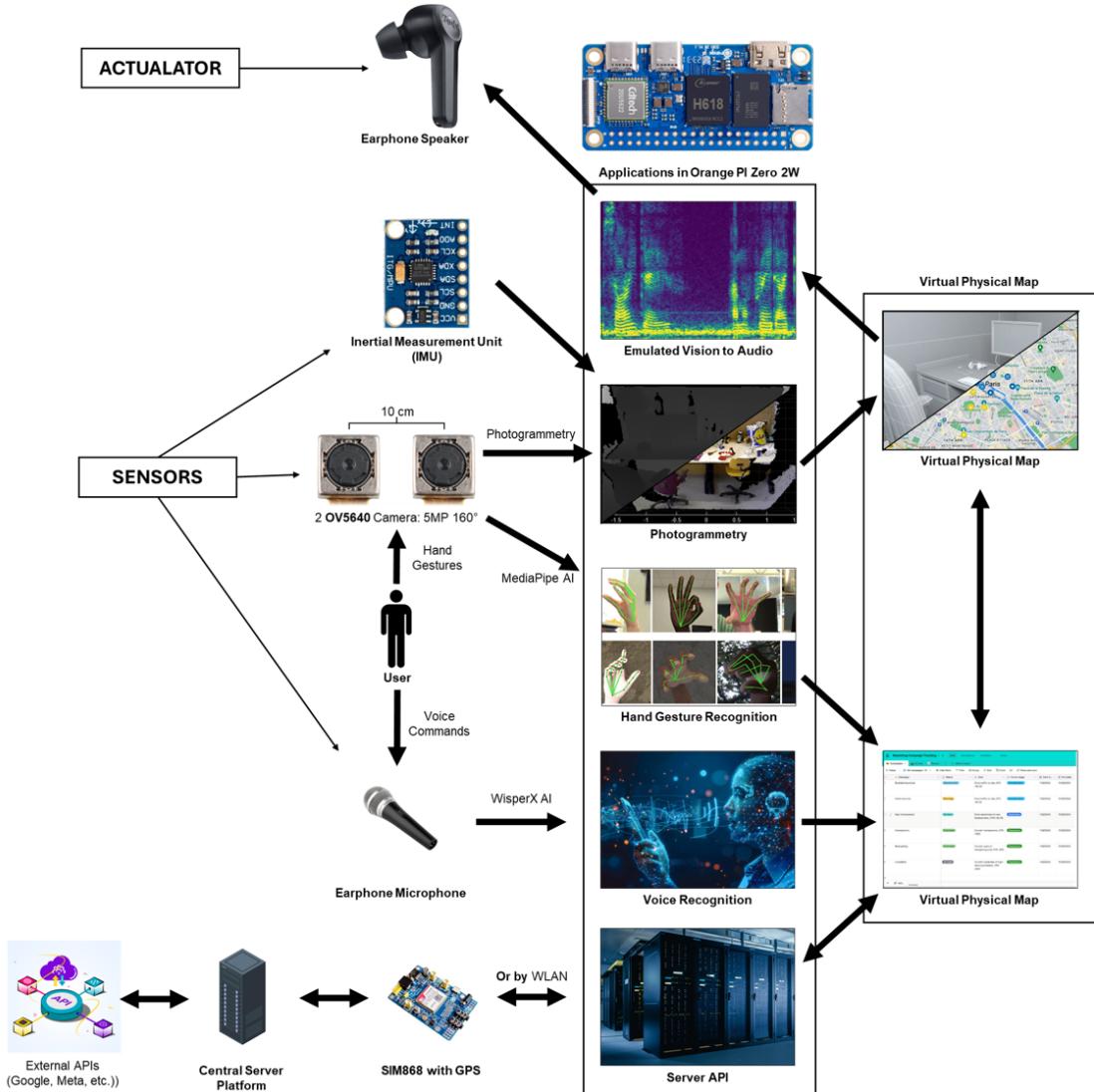
COLLEGE OF TECHNOLOGY

and the SIM and GPS Module, which provides both GPS positioning data for outdoor navigation and wireless connectivity options (WiFi or Mobile Data, indicated by wireless symbols) to enable internet access. This connectivity links the device to a remote Server (NodeJS), which hosts computationally intensive processes including Local AI models (Ollama GPT-OSS for natural language processing and Yolo for advanced object detection) and facilitates integration with External APIs such as Google Maps for navigation assistance and Facebook/Meta platforms for social connectivity and digital environment access. The server acts as the computational backbone for tasks beyond the microcontroller's capacity, processing complex AI inference requests and coordinating with cloud services, while the embedded system handles real-time sensor fusion, basic computer vision operations, and immediate audio feedback generation—creating a distributed computing architecture that balances on-device responsiveness with cloud-enhanced intelligence for comprehensive assistive navigation capabilities.

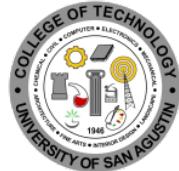


COLLEGE OF TECHNOLOGY

Figure 4: Architecture of the System



This system architecture diagram of figure 4 illustrates the complete data flow and component interaction of the World Navigation Hat, showing how sensory input is processed and transformed into actionable navigation assistance for visually impaired users. The process begins with three primary sensor categories: the dual OV5640 cameras (5MP at 160° FOV, spaced 10 cm apart) capture stereo images for photogrammetry, the Inertial Measurement Unit (IMU) tracks head motion and orientation, and the earphone



COLLEGE OF TECHNOLOGY

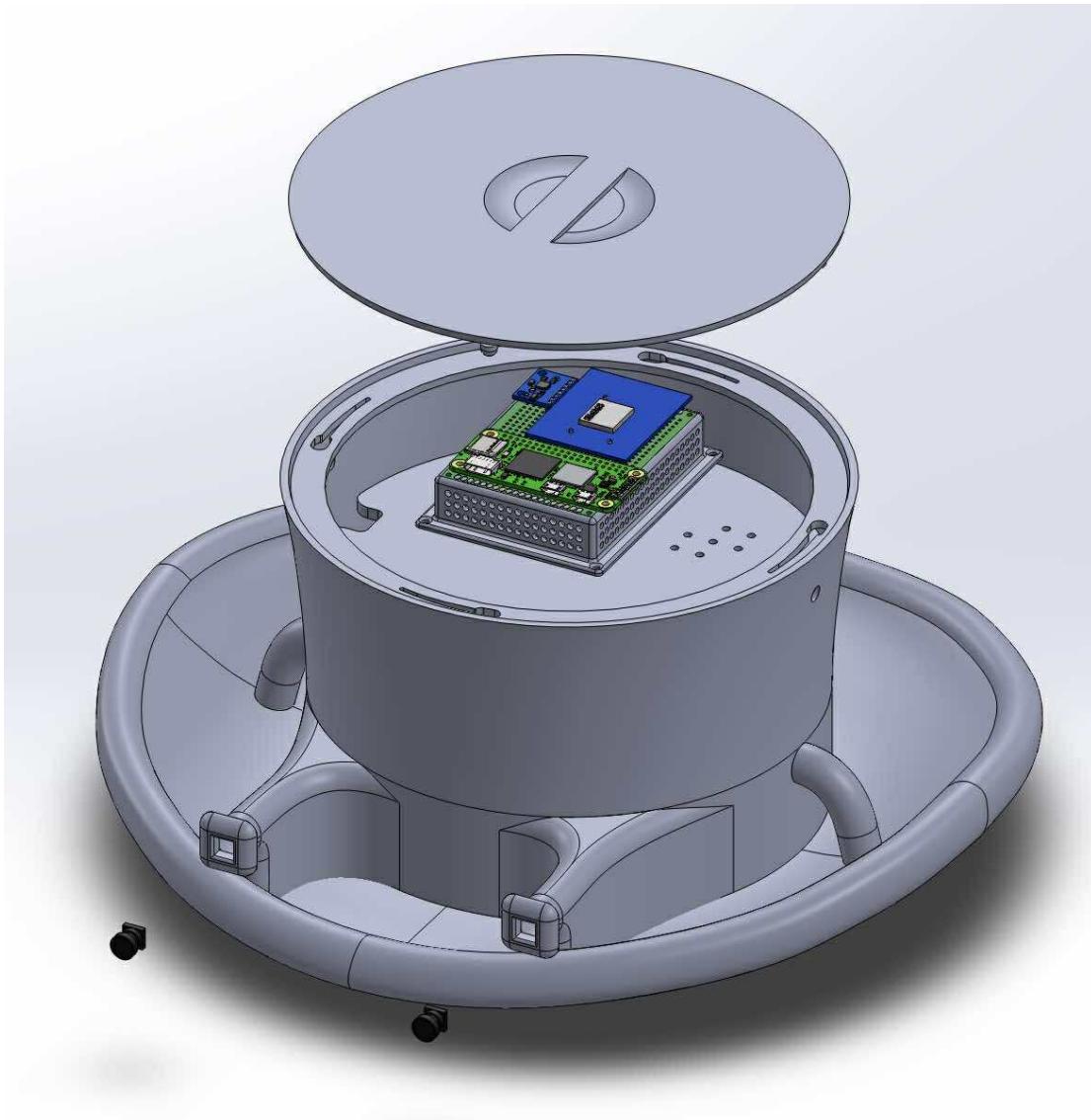
microphone captures voice commands from the user. These inputs feed into the Orange Pi Zero 2W microcontroller, which runs five core applications concurrently: (1) Emulated Vision to Audio—converting stereo images into 3D point clouds and depth maps, then applying human visual processing emulation (foveation, selective attention, scene gist extraction) to generate spatially-aware audio cues delivered through the ear-phone speaker; (2) Photogrammetry—reconstructing 3D environmental geometry from stereo disparity calculations; (3) Hand Gesture Recognition via MediaPipe AI—detecting user hand gestures for touchless device control and parameter adjustment; (4) Voice Recognition through WhisperX AI—processing spoken commands when activated by specific gestures; and (5) Server API integration—connecting to external computational resources. Data flows bidirectionally between the Orange Pi Zero 2W and a central server platform (accessed via the SIM868 module with GPS using either cellular data or WLAN), where NodeJS processes coordinate with external APIs (Google Maps, Meta platforms) to enhance navigation capabilities. The server infrastructure manages three interconnected virtual map representations: the Virtual Physical Map derived from real-time photogrammetry and GPS positioning showing the user's immediate surroundings with obstacle locations, the Virtual Digital Map providing access to online navigation services and digital content, and a unified spatial model that bridges physical and digital navigation contexts. Throughout this pipeline, the IMU continuously synchronizes the virtual environment with the user's head movements to maintain vestibular alignment and prevent disorientation, while the modular architecture allows applications to access any stage of the processing pipeline—enabling customization for specific user needs such as text reading, face recognition, or expense tracking—all working cohesively to



COLLEGE OF TECHNOLOGY

transform complex visual-spatial information into intuitive, non-overwhelming audio feedback that empowers independent navigation for visually impaired individuals.

Figure 5: 3D model of the system



This 3D CAD rendering in figure 5 illustrates the physical prototype of the World Navigation Hat's core electronics enclosure, designed as a compact, wearable sensory substitution device for visually impaired individuals. The cylindrical housing features a removable top lid embossed with a stylized logo, revealing the internal Orange Pi Zero



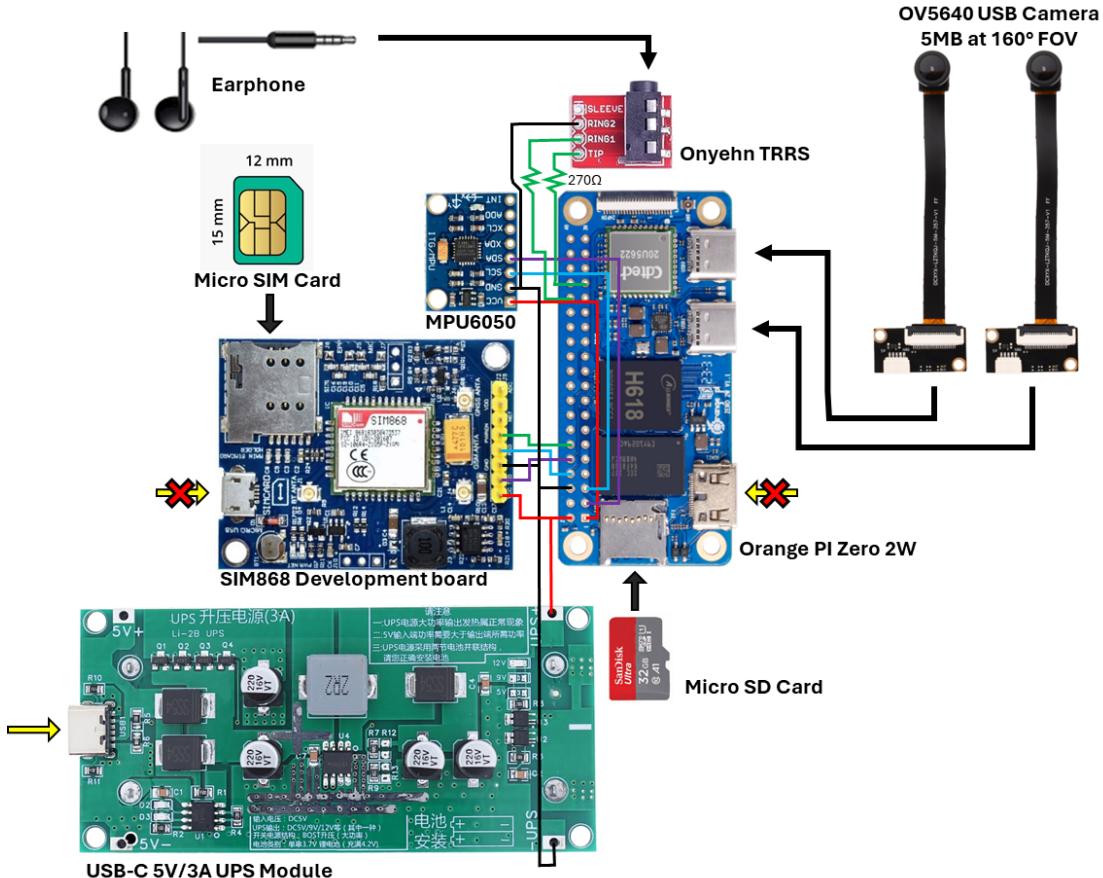
COLLEGE OF TECHNOLOGY

2W single-board computer (shown in green PCB with the prominent blue Allwinner H618 processor) mounted on a platform with ventilation holes for heat dissipation during intensive AI processing tasks. The device integrates into a modified hat brim structure (visible as the curved frame at the base) with attachment points and cable routing channels that secure cameras, sensors, and the UPS battery module while maintaining ergonomic comfort for extended wear. Small external ports (visible on the lower left) provide access for USB charging, while the sealed design protects sensitive electronics from environmental elements during outdoor navigation, embodying the project's goal of transforming bulky assistive technology into an unobtrusive, socially acceptable wearable that converts real-time visual information into intuitive audio cues through advanced photogrammetry, AI-driven scene understanding, and human sensory process emulation.



COLLEGE OF TECHNOLOGY

Figure 6: Complete Circuit Diagram of the System



This circuit figure 6 illustrates the complete hardware interconnection architecture of the World Navigation Hat system, showing how all components integrate to form a functional wearable assistive device for visually impaired individuals. Power Distribution System: The USB-C 5V/3A UPS Module serves as the primary power source, positioned at the bottom of the diagram with a yellow arrow indicating the external power input connection. This rechargeable power management unit supplies regulated 5V power to all system components, with Chinese text indicating battery charging specifications and protection circuitry. The green PCB features multiple power regulation stages with visible inductors (marked as 74Z), capacitors (both electrolytic and ceramic), and battery



COLLEGE OF TECHNOLOGY

management ICs to ensure stable operation during mobile use and support lithium battery charging cycles. Central Processing Unit: The Orange Pi Zero 2W single-board computer occupies the central position in the architecture, acting as the main computational hub. It features the Allwinner H618 quad-core processor (visible as the large silver chip), USB ports for camera connectivity, a micro SD card slot for operating system and data storage, GPIO header pins for peripheral connections, and HDMI output. The board receives 5V power from the UPS module (indicated by red power lines) and ground connections (black lines), while managing bidirectional data communication with all sensors and actuators. Stereo Vision System: Two OV5640 USB Cameras are positioned at the top right of the diagram, each providing 5MP resolution at 160° field of view. These cameras are mounted with a precise 10 cm baseline separation (as noted in the physical prototype specifications) and connect to the Orange Pi Zero 2W via USB interfaces (shown by thick black connection lines). The stereo camera configuration enables depth perception through stereo disparity calculations, capturing simultaneous images from slightly different viewpoints to generate 3D point cloud data essential for environmental mapping and obstacle detection. Motion Sensing: The MPU6050 Inertial Measurement Unit (6-axis gyroscope and accelerometer) appears as a small blue module at the center-top of the diagram. It connects to the Orange Pi Zero 2W through I2C communication protocol (indicated by green connection lines labeled with I2C pins: SCL, SDA, VCC, GND). This sensor tracks head motion and orientation in real-time, providing critical data for aligning the virtual environment representation with the user's physical movements, preventing vestibular-visual conflicts that could cause disorientation or nausea. Cellular and GPS Communication: The SIM868 Development Board is prominently displayed



COLLEGE OF TECHNOLOGY

on the left side, featuring the SIM868 module (white component with CE marking) that integrates GSM/GPRS cellular communication and GPS positioning capabilities. The micro SIM card slot (shown with dimensions 12mm × 15mm) accepts a standard cellular SIM card for mobile data connectivity. Connection lines (green for data, red for power, black for ground) link the SIM868 to the Orange Pi Zero 2W, enabling outdoor navigation through GPS coordinates and server communication via 2G/3G networks when Wi-Fi is unavailable. Two X-marked connectors on either side indicate antenna connection points for cellular and GPS signals. Audio Interface: The Onyehn TRRS audio module (small purple/red PCB) positioned at the top-center provides bidirectional audio functionality. It connects to the Orange Pi Zero 2W's audio pins and features a TRRS (Tip-Ring-Ring-Sleeve) 3.5mm jack supporting both audio output and microphone input through a single connector. The earphone illustration at the top shows standard earbud-style output for delivering spatial audio cues to the user, while the integrated microphone captures voice commands for hands-free device control through WhisperX voice recognition processing. Connection Architecture: Color-coded wiring illustrates the data flow and power distribution:

Red lines: 5V power rails from UPS to all components

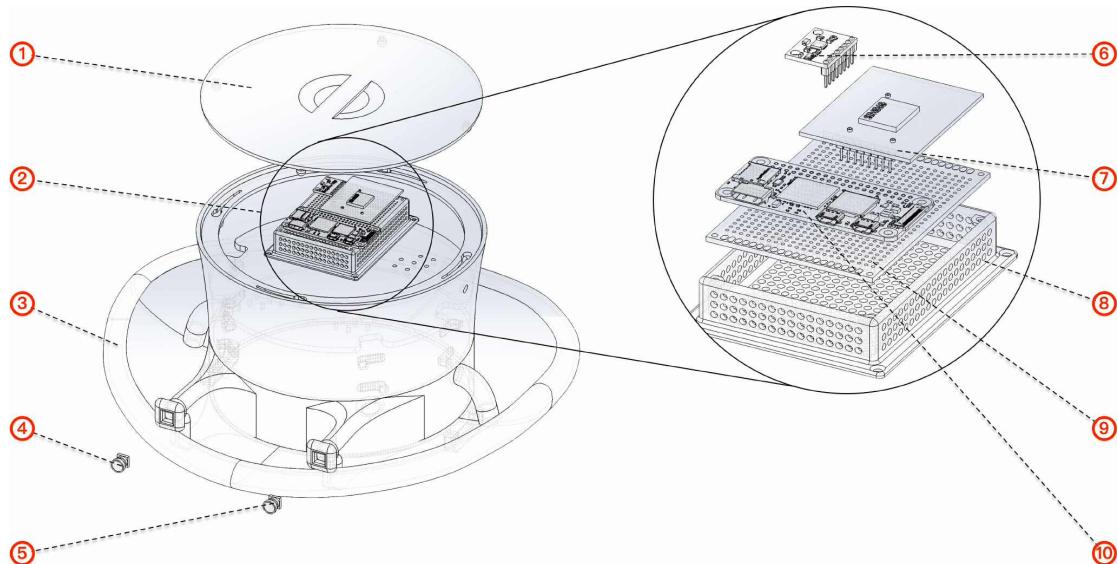
- (1) Black lines: Ground connections ensuring common reference voltage
- (2) Green lines: I2C communication (MPU6050), UART serial data (SIM868), or other digital communication protocols
- (3) Thick black lines: USB data connections for high-bandwidth camera feeds
- (4) Blue/purple lines: Audio signal paths between Orange Pi and TRRS module



COLLEGE OF TECHNOLOGY

Physical Integration: All components are designed to mount within the 3D-printed hat enclosure shown in previous wireframe diagrams, with the UPS module likely positioned at the rear for weight distribution, the Orange Pi Zero 2W and SIM868 board in the central crown area for protection, cameras mounted at the front brim for unobstructed field of view, and the IMU positioned near the cameras to accurately track head orientation relative to the visual input direction. The compact arrangement maintains the wearable form factor while ensuring proper thermal management through ventilation and component spacing.

Figure 7: Wireframe of the Prototype



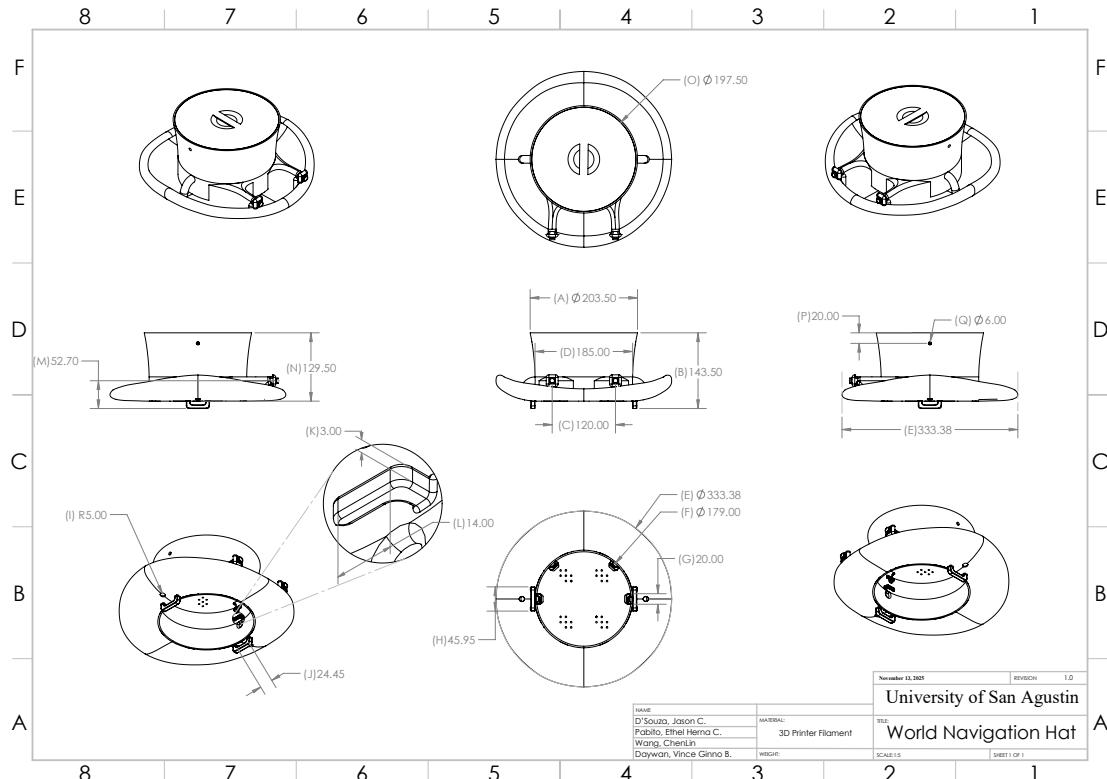
#	Component
1	Cover for the device - 3D Printed
2	Core circuit box
3	Case of Hat - 3D Printed
4	OV5640 USB Camera 5MP - 1
5	OV5640 USB Camera 5MP - 2

#	Component
6	MPU-6050
7	Pico SIM868
8	PCB Box - 3D Printed
9	PCB Board
10	Raspberry Pi Zero - W



COLLEGE OF TECHNOLOGY

Figure 8: Measurements of the Prototype



A: Hat top diameter

J: Strap handle inner length

B: Total height

K: Wire clamp width

C: Camera distance

L: Wire clamp inner length

D: Hat Trim diameter

M: Camera height - from bottom of hat

E: Hat Brim diameter

N: Total height - without strap handle

F: Hat Sweatband diameter

O: Cover diameter

G: Wire clamp outer length

P: Antenna hole position - relative to hat

H: Strap handle outer length

Q: Antenna hole diameter

I: Earphone wire channel radius

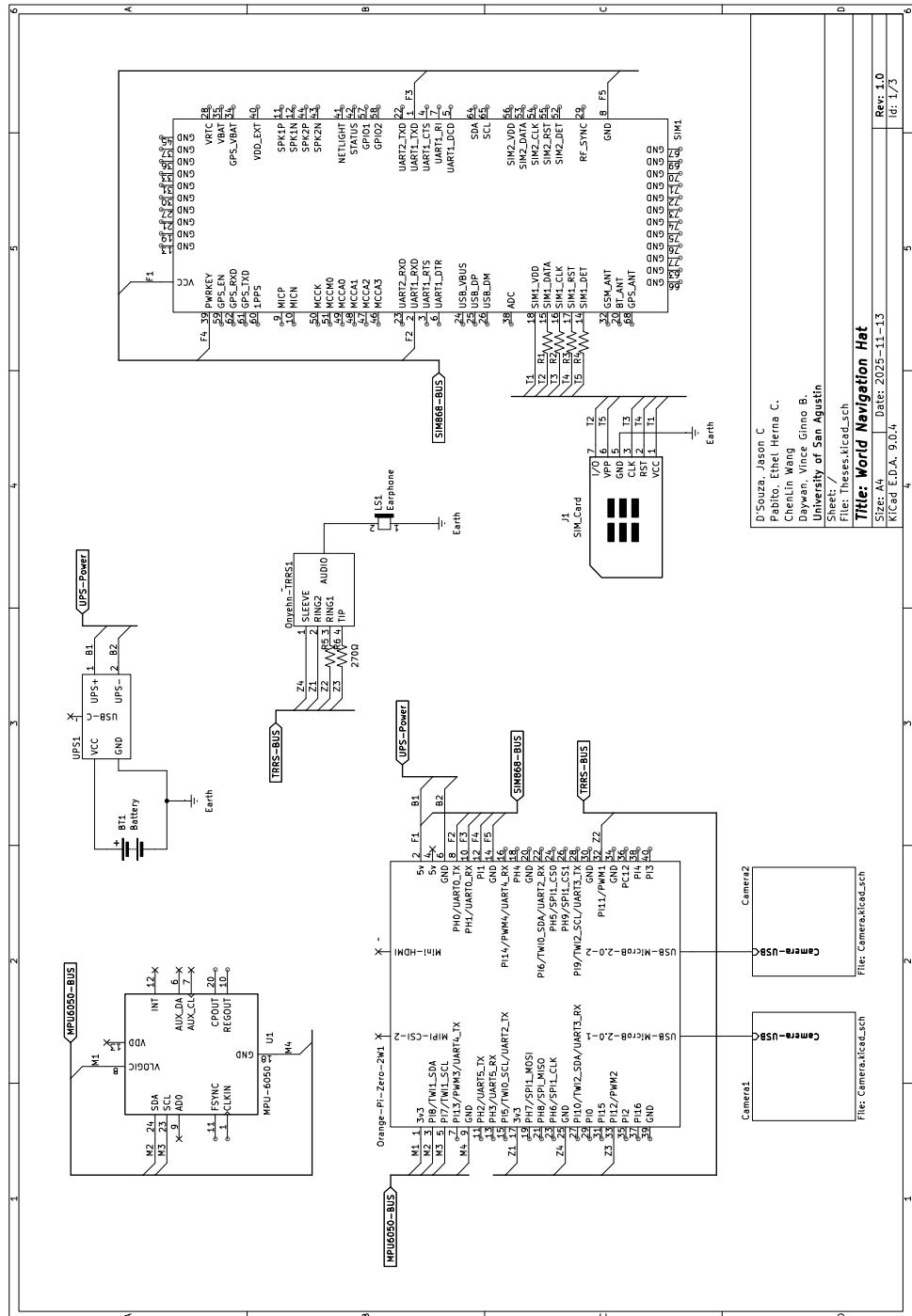


University of San Agustin
General Luna St., 5000 Iloilo City, Philippines
www.usa.edu.ph

COLLEGE OF TECHNOLOGY



Figure 9: Complete Schematic Diagram of the System



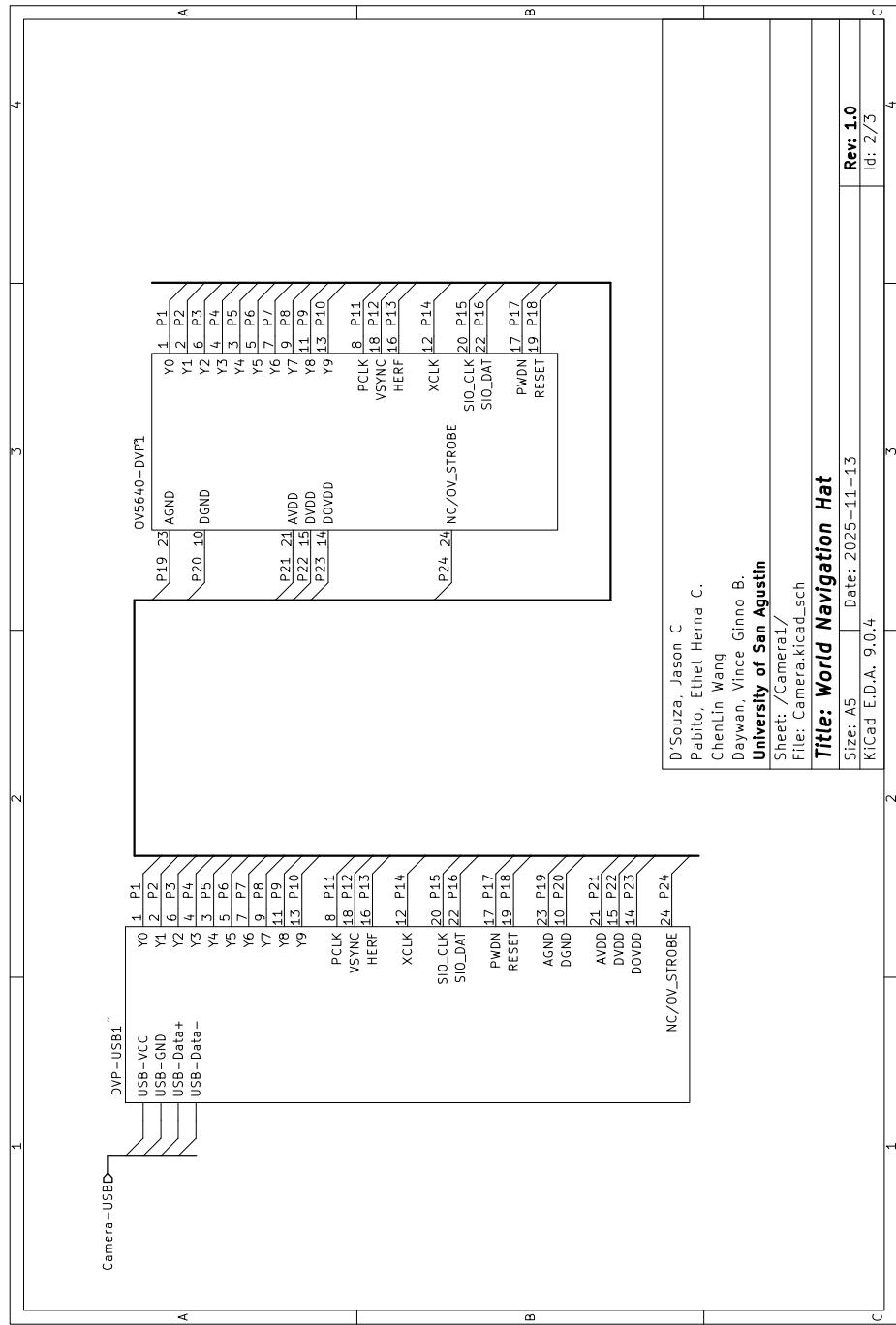


University of San Agustin
 General Luna St., 5000 Iloilo City, Philippines
www.usa.edu.ph



COLLEGE OF TECHNOLOGY

Figure 10: Camera 1 Schematic Diagram



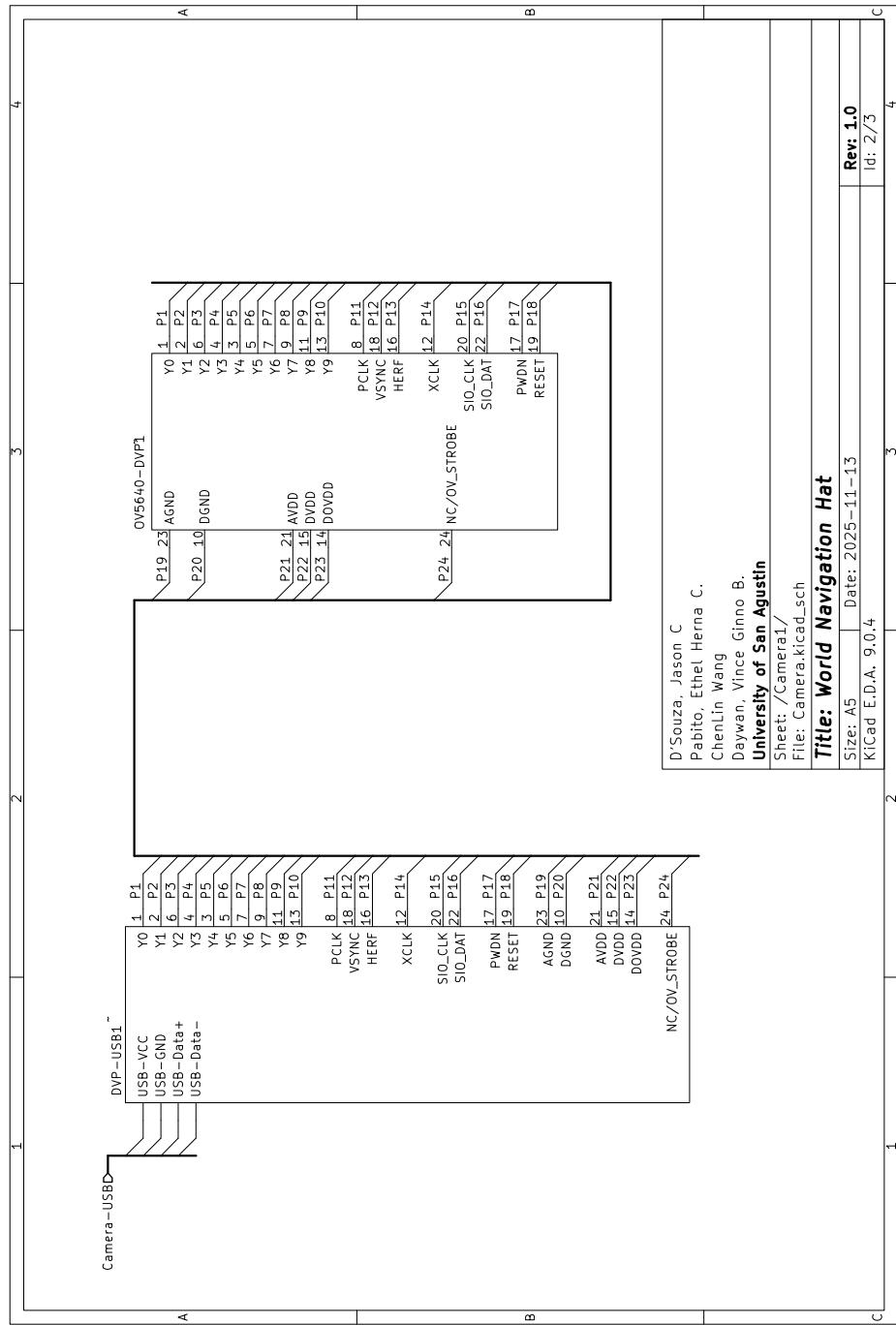


University of San Agustin
 General Luna St., 5000 Iloilo City, Philippines
www.usa.edu.ph



COLLEGE OF TECHNOLOGY

Figure 11: Camera 2 Schematic Diagram





COLLEGE OF TECHNOLOGY

Table 3 : Table of pin connections

Net Name	Code	Connections
<i>Power and Ground Nets</i>		
/B1	2	Orange-Pi-Zero-2W1(2:5v) → SIM1(VCC) → UPS1(1:UPS+)
/B2	3	Orange-Pi-Zero-2W1(6:GND) → UPS1(2:UPS-)
Earth	71	BT1(2:-) → J1(5:GND) → LS1(1:-) → UPS1(GND)
Net-(BT1-+)	73	BT1(1:+) → UPS1(VCC)
/Z1	67	Onyehn-TRRS1(2:RING2) → Orange-Pi-Zero-2W1(17:3v3)
/M1	58	Orange-Pi-Zero-2W1(1:3v3) → U1(8:VLOGIC)
<i>SIM868 Module – UART & Control</i>		
/F2	54	Orange-Pi-Zero- 2W1(8:PH0/UART0_TX) → SIM1(2:UART1_RXD)
/F3	55	Orange-Pi-Zero- 2W1(10:PH1/UART0_RX) → SIM1(1:UART1_TXD)

Continued on next page



COLLEGE OF TECHNOLOGY

Table 3: Table of pin connections (Continued)

Net Name	Code	Connections
/F4	56	Orange-Pi-Zero-2W1(12:PI1) → SIM1(39:PWRKEY)
/F5	57	Orange-Pi-Zero-2W1(14:GND) → SIM1(8:GND)
<i>SIM Card Interface</i>		
/T1	62	J1(1:VCC) → SIM1(18:SIM1_VDD)
/T2	63	J1(7:I/O) → R1(1)
/T3	64	J1(3:CLK) → R2(1)
/T4	65	J1(2:RST) → R3(1)
/T5	66	J1(6:VPP) → R4(1)
Net-(SIM1-SIM1_DATA)	78	R1(2) → SIM1(15:SIM1_DATA)
Net-(SIM1-SIM1_CLK)	77	R2(2) → SIM1(16:SIM1_CLK)
Net-(SIM1-SIM1_RST)	80	R3(2) → SIM1(17:SIM1_RST)
Net-(SIM1-SIM1_DET)	79	R4(2) → SIM1(14:SIM1_DET)
<i>MPU-6050 (IMU)</i>		

Continued on next page



COLLEGE OF TECHNOLOGY

Table 3: Table of pin connections (Continued)

Net Name	Code	Connections
/M2	59	Orange-Pi-Zero- 2W1(3:PI8/TWI1_SDA) → U1(24:SDA)
/M3	60	Orange-Pi-Zero- 2W1(5:PI7/TWI1_SCL) → U1(23:SCL)
/M4	61	Orange-Pi-Zero-2W1(9:GND) → U1(18:GND)
<i>Audio Output (TRRS)</i>		
Net-(LS1-+)	74	LS1(2:+) → Onyehn-TRRS1(AUDIO)
/Z2	68	Orange-Pi-Zero- 2W1(32:PI11/PWM1) → R5(1)
/Z3	69	Orange-Pi-Zero- 2W1(33:PI12/PWM2) → R6(1)
Net-(Onyehn-TRRS1-RING1)	75	Onyehn-TRRS1(3:RING1) → R5(2)
Net-(Onyehn-TRRS1-TIP)	76	Onyehn-TRRS1(4:TIP) → R6(2)

Continued on next page



COLLEGE OF TECHNOLOGY

Table 3: Table of pin connections (Continued)

Net Name	Code	Connections
/Z4	70	Onyehn-TRRS1(1:SLEEVE) → Orange-Pi-Zero-2W1(25:GND)
<i>Camera 1 (/Camera1/)</i>		
/Camera1/Camera-USB	4	Orange-Pi-Zero-2W1(USB-MicroB-2.0-1)
/Camera1/P1	5	DVP-USB1(1:Y0) → OV5640-DVP1(1:Y0)
/Camera1/P2	6	DVP-USB1(2:Y1) → OV5640-DVP1(2:Y1)
/Camera1/P3	7	DVP-USB1(6:Y2) → OV5640-DVP1(6:Y2)
/Camera1/P4	8	DVP-USB1(4:Y3) → OV5640-DVP1(4:Y3)
/Camera1/P5	9	DVP-USB1(3:Y4) → OV5640-DVP1(3:Y4)
/Camera1/P6	10	DVP-USB1(5:Y5) → OV5640-DVP1(5:Y5)

Continued on next page



COLLEGE OF TECHNOLOGY

Table 3: Table of pin connections (Continued)

Net Name	Code	Connections
/Camera1/P7	11	DVP-USB1(7:Y6) → OV5640-DVP1(7:Y6)
/Camera1/P8	12	DVP-USB1(9:Y7) → OV5640-DVP1(9:Y7)
/Camera1/P9	13	DVP-USB1(11:Y8) → OV5640-DVP1(11:Y8)
/Camera1/P10	14	DVP-USB1(13:Y9) → OV5640-DVP1(13:Y9)
/Camera1/P11	15	DVP-USB1(8:PCLK) → OV5640-DVP1(8:PCLK)
/Camera1/P12	16	DVP-USB1(18:VSYNC) → OV5640-DVP1(18:VSYNC)
/Camera1/P13	17	DVP-USB1(16:HERF) → OV5640-DVP1(16:HERF)
/Camera1/P14	18	DVP-USB1(12:XCLK) → OV5640-DVP1(12:XCLK)
/Camera1/P15	19	DVP-USB1(20:SIO_CLK) → OV5640-DVP1(20:SIO_CLK)

Continued on next page



COLLEGE OF TECHNOLOGY

Table 3: Table of pin connections (Continued)

Net Name	Code	Connections
/Camera1/P16	20	DVP-USB1(22:SIO_DAT) → OV5640-DVP1(22:SIO_DAT)
/Camera1/P17	21	DVP-USB1(17:PWDN) → OV5640-DVP1(17:PWDN)
/Camera1/P18	22	DVP-USB1(19:RESET) → OV5640-DVP1(19:RESET)
/Camera1/P19	23	DVP-USB1(23:AGND) → OV5640-DVP1(23:AGND)
/Camera1/P20	24	DVP-USB1(10:DGND) → OV5640-DVP1(10:DGND)
/Camera1/P21	25	DVP-USB1(21:AVDD) → OV5640-DVP1(21:AVDD)
/Camera1/P22	26	DVP-USB1(15:DVDD) → OV5640-DVP1(15:DVDD)
/Camera1/P23	27	DVP-USB1(14:DOVDD) → OV5640-DVP1(14:DOVDD)

Continued on next page



COLLEGE OF TECHNOLOGY

Table 3: Table of pin connections (Continued)

Net Name	Code	Connections
/Camera1/P24	28	DVP- USB1(24:NC/OV_STROBE) → OV5640- DVP1(24:NC/OV_STROBE)
<i>Camera 2 (/Camera2/)</i>		
/Camera2/Camera-USB	29	Orange-Pi-Zero-2W1(USB-MicroB-2.0-2)
/Camera2/P1	30	DVP-USB2(1:Y0) → OV5640-DVP2(1:Y0)
/Camera2/P2	31	DVP-USB2(2:Y1) → OV5640-DVP2(2:Y1)
/Camera2/P3	32	DVP-USB2(6:Y2) → OV5640-DVP2(6:Y2)
/Camera2/P4	33	DVP-USB2(4:Y3) → OV5640-DVP2(4:Y3)
/Camera2/P5	34	DVP-USB2(3:Y4) → OV5640-DVP2(3:Y4)
/Camera2/P6	35	DVP-USB2(5:Y5) → OV5640-DVP2(5:Y5)

Continued on next page



COLLEGE OF TECHNOLOGY

Table 3: Table of pin connections (Continued)

Net Name	Code	Connections
/Camera2/P7	36	DVP-USB2(7:Y6) → OV5640-DVP2(7:Y6)
/Camera2/P8	37	DVP-USB2(9:Y7) → OV5640-DVP2(9:Y7)
/Camera2/P9	38	DVP-USB2(11:Y8) → OV5640-DVP2(11:Y8)
/Camera2/P10	39	DVP-USB2(13:Y9) → OV5640-DVP2(13:Y9)
/Camera2/P11	40	DVP-USB2(8:PCLK) → OV5640-DVP2(8:PCLK)
/Camera2/P12	41	DVP-USB2(18:VSYNC) → OV5640-DVP2(18:VSYNC)
/Camera2/P13	42	DVP-USB2(16:HERF) → OV5640-DVP2(16:HERF)
/Camera2/P14	43	DVP-USB2(12:XCLK) → OV5640-DVP2(12:XCLK)
/Camera2/P15	44	DVP-USB2(20:SIO_CLK) → OV5640-DVP2(20:SIO_CLK)

Continued on next page



COLLEGE OF TECHNOLOGY

Table 3: Table of pin connections (Continued)

Net Name	Code	Connections
/Camera2/P16	45	DVP-USB2(22:SIO_DAT) → OV5640-DVP2(22:SIO_DAT)
/Camera2/P17	46	DVP-USB2(17:PWDN) → OV5640-DVP2(17:PWDN)
/Camera2/P18	47	DVP-USB2(19:RESET) → OV5640-DVP2(19:RESET)
/Camera2/P19	48	DVP-USB2(23:AGND) → OV5640-DVP2(23:AGND)
/Camera2/P20	49	DVP-USB2(10:DGND) → OV5640-DVP2(10:DGND)
/Camera2/P21	50	DVP-USB2(21:AVDD) → OV5640-DVP2(21:AVDD)
/Camera2/P22	51	DVP-USB2(15:DVDD) → OV5640-DVP2(15:DVDD)
/Camera2/P23	52	DVP-USB2(14:DOVDD) → OV5640-DVP2(14:DOVDD)

Continued on next page



University of San Agustin
General Luna St., 5000 Iloilo City, Philippines
www.usa.edu.ph



COLLEGE OF TECHNOLOGY

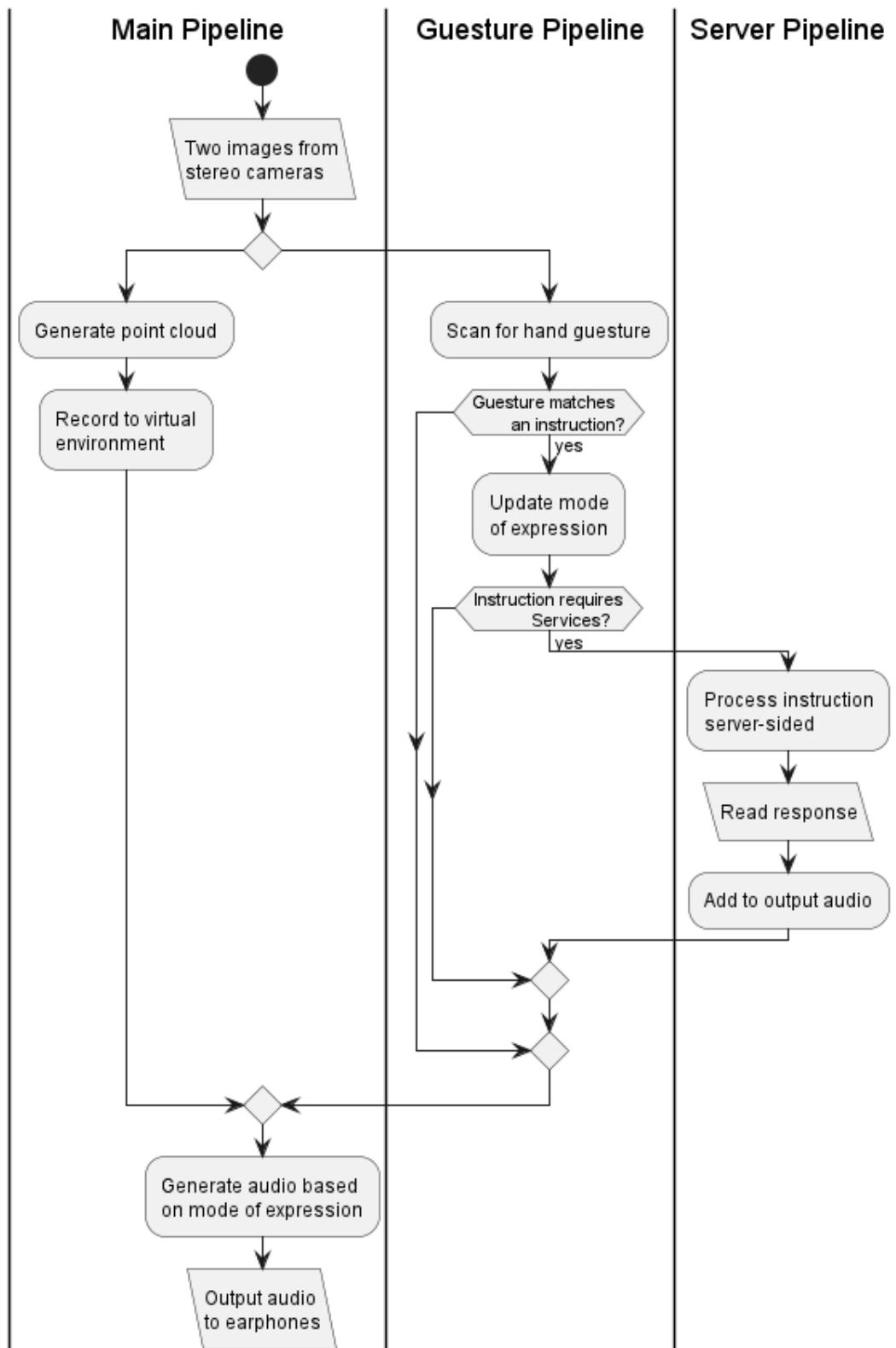
Table 3: Table of pin connections (Continued)

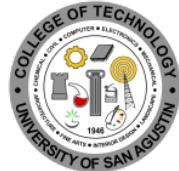
Net Name	Code	Connections
/Camera2/P24	53	DVP- USB2(24:NC/OV_STROBE) → OV5640- DVP2(24:NC/OV_STROBE)



COLLEGE OF TECHNOLOGY

Figure 12: Flowchart of the three major pipelines





COLLEGE OF TECHNOLOGY

The flowchart at figure 12 illustrates the modular software architecture of the World Navigation Hat, organized into three concurrent processing pipelines that work together to provide comprehensive assistive navigation through sensory substitution. The Main Pipeline forms the core sensory substitution system, beginning with "Two images from stereo cameras" captured simultaneously from the dual OV5640 cameras positioned 10 cm apart, which undergo photogrammetry processing to "Generate point cloud" by applying stereo disparity calculations ($Z = f \cdot B/d$) that create a three-dimensional representation of the environment where each pixel's depth is computed and transformed into 3D spatial coordinates; this point cloud data is then used to "Record to virtual environment," creating a persistent spatial map integrated with IMU-tracked head movements to form a coherent representation of the user's surroundings. The Gesture Pipeline runs in parallel, providing touchless user control through "Scan for hand gesture" using MediaPipe AI models that continuously analyze camera frames to detect hand landmarks—when a gesture is detected, the system evaluates "Gesture matches an instruction?" and if true, proceeds to "Update mode of expression," modifying parameters that affect how the Main Pipeline generates audio output (allowing users to switch between obstacle detection, text reading, or face recognition modes), then checks "Instruction requires Services?" to determine whether the command needs server-side processing, branching to the Server Pipeline if yes or merging directly back to the Main Pipeline if no. The Server Pipeline handles computationally intensive tasks, where requests are transmitted via the SIM868 module to "Process instruction server-sided" on the NodeJS platform, which may invoke Ollama large language models for natural language understanding, execute YOLO object detection for detailed scene analysis, query Google Maps APIs for navigation routes, or access



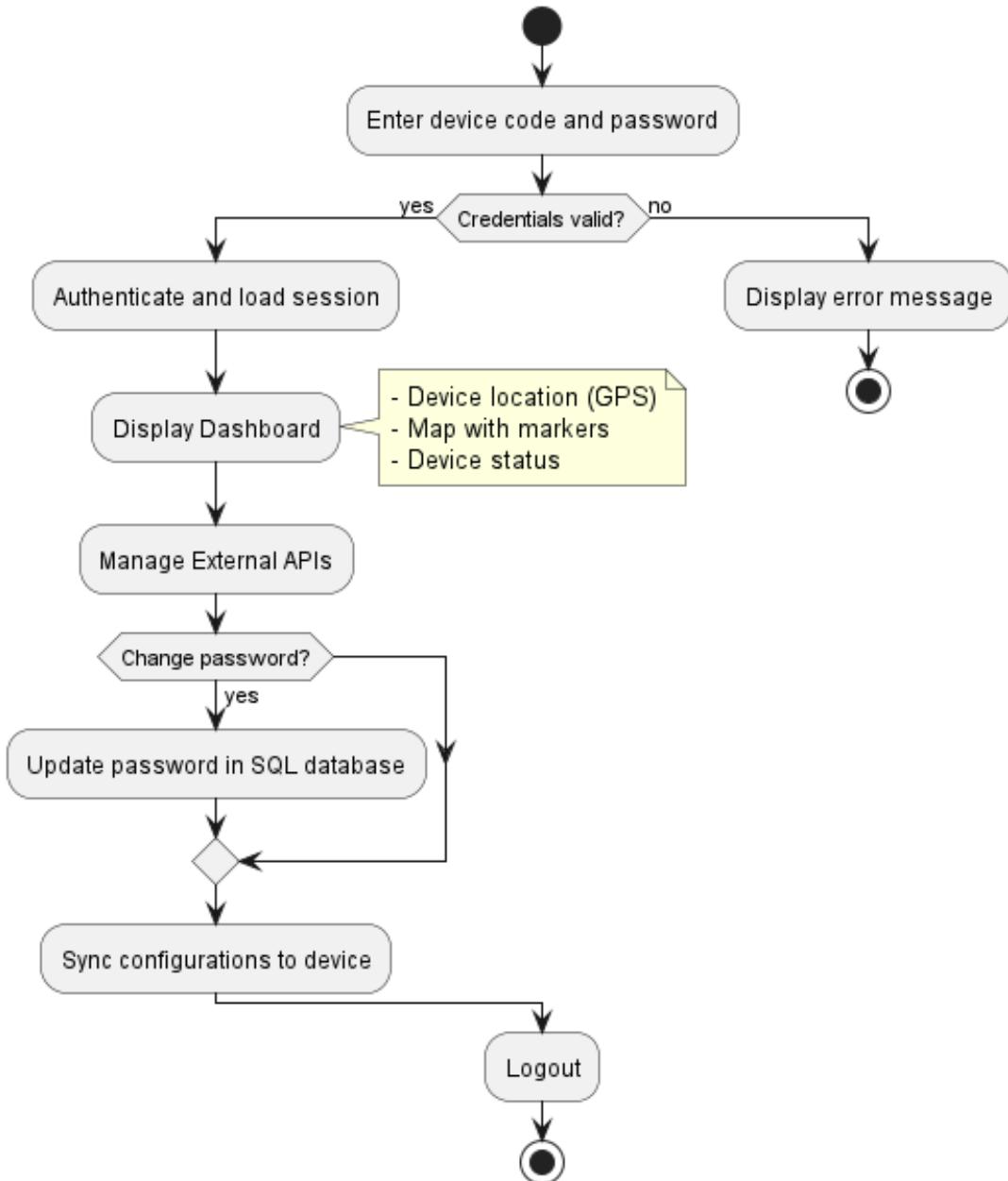
COLLEGE OF TECHNOLOGY

social media platforms, then proceeds to "Read response" to collect and format results before "Add to output audio" packages the server-processed information and transmits it back to merge with the Gesture Pipeline. All three pipelines converge through diamond-shaped merge points before the final stage where the system "Generate audio based on mode of expression," transforming consolidated environmental data, user control inputs, and cloud-enhanced intelligence into spatial audio cues through algorithms that emulate human visual processing—implementing foveation (focus-based detail levels), selective attention (filtering salient features), scene gist extraction (background/foreground separation), and temporal dynamics (progressive audio updates)—before finally proceeding to "Output audio to earphones," delivering intuitive sonic representations that respect cognitive load limitations while maximizing navigational utility, embodying the research's core innovation of solving the bandwidth mismatch between high-detail visual information and lower-bandwidth auditory perception.



COLLEGE OF TECHNOLOGY

Figure 13: Flowchart of Server-sided Interface



This flowchart at figure 13 illustrates the server-side user management and configuration system for the World Navigation Hat, enabling caregivers or visually impaired users (with assistance) to remotely manage device settings and integrate external services through a web-based dashboard. The process begins when a user enters their unique



COLLEGE OF TECHNOLOGY

device code and password for authentication—if credentials are valid, the system authenticates and loads the user session, displaying a comprehensive dashboard that shows real-time device location via GPS data from the SIM868 module, a map visualization with saved markers indicating important places and items the user has tagged during navigation, and current device status information. The dashboard provides access to the "Manage External APIs" partition where users can configure various third-party service integrations: selecting Facebook allows entry of social media credentials that are encrypted and stored in the SQL database, enabling the device to initiate hands-free voice or video calls without requiring the visually impaired user to navigate login interfaces; selecting Google Maps enables entry of API keys for enhanced navigation features; selecting Local AI Models allows configuration of Agent Models (like Ollama GPT for conversational assistance) or Custom Models (user-uploaded machine learning models for specialized tasks like recognizing specific objects or faces), with all parameters and endpoints stored in the SQL database; and selecting Other APIs accommodates additional services with credential storage. After API configuration, the system updates the API status on the dashboard to reflect active integrations, then offers users the option to change their password (which updates the SQL database upon confirmation), and finally syncs all configuration changes to the physical device via the SIM868 cellular/Wi-Fi connection, ensuring the wearable hat has updated credentials and settings for seamless operation. If authentication fails at any point, the system displays an error message and prompts re-entry of credentials, while successful sessions conclude with a logout action that securely terminates the user's access—this server-side infrastructure thus acts as the central management hub that bridges the gap between complex digital services and the



University of San Agustin
General Luna St., 5000 Iloilo City, Philippines
www.usa.edu.ph



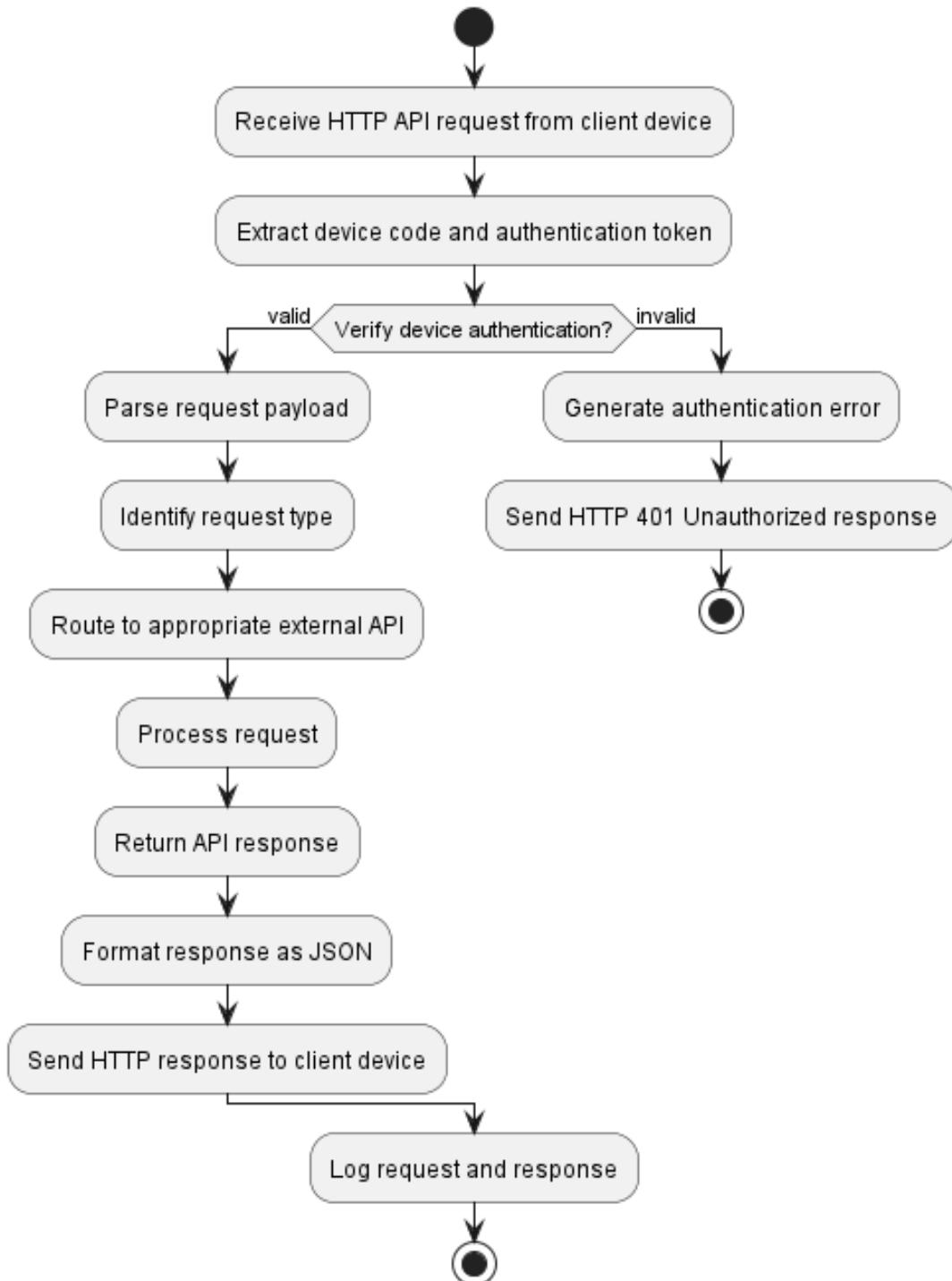
COLLEGE OF TECHNOLOGY

simplified, accessible interface needed by visually impaired users, handling credential management, data persistence, and device synchronization to enable features like social connectivity and AI-enhanced navigation without overwhelming the user with technical complexity.



COLLEGE OF TECHNOLOGY

Figure 14: Flowchart of Server's API



This flowchart at figure 14 illustrates the server-side API processing system that



COLLEGE OF TECHNOLOGY

handles HTTP requests from the World Navigation Hat client device, acting as an intelligent intermediary between the embedded system and various external services. The process begins when the server receives an HTTP API request from the client device, immediately extracting the device code and authentication token to verify device authentication—if the credentials are invalid, the server generates an authentication error and sends an HTTP 401 Unauthorized response to prevent unauthorized access. Upon successful authentication, the server parses the request payload to identify the request type and routes it to the appropriate processing pipeline: for Facebook Call requests, the server retrieves stored Facebook credentials from the SQL database, launches Puppeteer browser automation to navigate to Facebook's web interface, authenticates using the stored credentials (eliminating the need for the visually impaired user to manually log in), initiates a call to the specified contact, and returns the call status response; for Navigation Requests (such as "take me home"), the server parses the natural language command, queries the Local Agent LLM (Ollama) which interprets user intent and determines the destination (retrieving the home address from the SQL database), calls the Google Maps API to calculate the optimal route with turn-by-turn directions, formats the navigation instructions into audio-friendly text, and returns the route response that the client device converts into spatial audio cues; for YOLO Object Detection requests, the server receives image data from the client cameras, loads the YOLO model to process the image for comprehensive object detection beyond the client's computational capacity, generates detailed object labels with spatial positions, and returns the detection results for integration into the audio feedback system; for Custom AI Model requests, the server identifies the requested model from SQL configuration data, loads and executes the model with



COLLEGE OF TECHNOLOGY

provided input parameters, generates the model output (which could be face recognition, text reading, or specialized tasks), and returns the AI response. After processing any request type, the server formats the response as JSON for consistent parsing, sends the HTTP response back to the client device via the SIM868 module's cellular or Wi-Fi connection, and logs both the request and response for debugging, analytics, and system monitoring purposes—this architecture offloads computationally expensive operations from the battery-constrained wearable device to cloud infrastructure while maintaining a simple HTTP-based communication protocol that ensures reliability and compatibility across different network conditions.

Formula References

- (1) **Percentage Error Formula** - Expresses the error between a measured value and the true value (e.g., predicted vs. actual direction or depth).

$$\text{Percentage Error} = \left| \frac{\text{Measured Value} - \text{True Value}}{\text{True Value}} \right| \times 100\% \quad (1)$$

- (2) **Success Percentage Formula** - Measures the proportion of successful trials (e.g., correct predictions by user or device).

$$\text{Success Percentage} = \frac{\text{Successful Trials}}{\text{Total Trials}} \times 100\% \quad (2)$$

- (3) **Stereo Camera Matrices** - Intrinsic and extrinsic representation of left (C_1) and



COLLEGE OF TECHNOLOGY

right (C_2) cameras, where f is focal length in pixels, c_x, c_y are principal point coordinates, and T_x is the baseline.

$$C_1 = \begin{bmatrix} f & 0 & c_{x_1} & 0 \\ 0 & f & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad C_2 = \begin{bmatrix} f & 0 & c_{x_2} & T_x f \\ 0 & f & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (3)$$

(4) **Stereo Reprojection Matrix (Q)** - Used to compute 3D points from disparity:
 converts $(x, y, \text{disparity})$ into 3D coordinates.

$$Q = \begin{bmatrix} 1 & 0 & 0 & -c_{x_1} \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 0 & f \\ 0 & 0 & -T_x^{-1} & T_x^{-1}(c_{x_1} - c_{x_2}) \end{bmatrix} \quad (4)$$

(5) **Point Cloud Projection Formula** - Reprojects a pixel with disparity into 3D space using the Q matrix.

$$\begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} = Q \begin{bmatrix} x \\ y \\ \text{disparity}(x, y) \\ 1 \end{bmatrix} \quad (5)$$



COLLEGE OF TECHNOLOGY

(1) Precision Formula:

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

This formula determines how correctly the system identifies and recognizes hand gestures. It shows how many of the gestures classified by the system were actually the correct target gesture.

(2) Recall Formula:

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}$$

This formula determines how well the system detects all actual instances of each gesture type. It shows how many of the real target gestures were correctly identified and recognized.

(3) Accuracy Formula:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

This formula determines the overall correctness of the system. It shows how many of the total hand gesture classifications were correctly identified and recognized.



COLLEGE OF TECHNOLOGY

Table 4 : Confusion Matrix of Hand Gesture Recognition Results

Actual Gesture Type	Classified as Open Hand	Classified as Closed Hand	Classified as Pointing	Classified as Index	Classified as Thumb	Total Actual Item
<i>Open Back Hand</i>						
<i>Closed Back Hand</i>						
<i>Pointing Index</i>						
<i>Pointing Thumb</i>						
Total						

Table 4 presents the confusion matrix of hand gesture recognition results, which shows the relationship between the actual gesture types and the classified results produced by the MediaPipe AI model. Each row represents the true gesture category, while each column represents how the system classified and recognized the gesture into its corresponding type.



COLLEGE OF TECHNOLOGY

Table 5 : Performance Evaluation Metrics for Hand Gesture Recognition

Gesture Type	TP	FP	FN	TN	Precision	Recall	Accuracy
<i>Open</i>							
<i>Back</i>							
<i>Hand</i>							
<i>Closed</i>							
<i>Back</i>							
<i>Hand</i>							
<i>Pointing</i>							
<i>Index</i>							
<i>Pointing</i>							
<i>Thumb</i>							
Overall (Average)							

Table 5 presents the performance metrics of the MediaPipe AI model for each hand gesture type. It shows the values of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN), which represent the classification outcomes of the system. Based on these values, the Precision, Recall, and Accuracy metrics were computed to evaluate the overall performance of the gesture recognition system.



COLLEGE OF TECHNOLOGY

Ethical Considerations

(1) Informed Consent and Autonomy:

All participants will be informed of their rights, such as the right to voluntary consent, particularly during FGDs involving VIPs and during prototype testing. The researchers will secure approval of implementation and evaluation and provide an accessible format, such as audio instruction or a face-to-face explanation, to present the project's purpose, duration, recording methods, and data use to ensure the participants' understanding despite communication barriers. For professional consultations with experts, the researchers will confirm their willingness to participate, disclose any conflicts of interest, and clarify their voluntary role. Participants will be informed that they have the right to withdraw at any time without facing repercussions, which is crucial in group settings to mitigate peer pressure or to avoid participating during the testing of the project's functionality. Researchers will assess participants' capacity to consent, involving caregivers as needed, while preserving participants' IP rights.

(2) Privacy and Confidentiality:

The researchers will practice data privacy and confidentiality when acquiring participants' personal experiences or device feedback data. The researchers will obtain consent, ensure data is anonymized, and securely store it to prevent unauthorized access, excluding participants' identifiable details from reports. In addition, during professional consultations, they often share information that must remain private, so the researchers will need confidentiality agreements to protect ideas and opinions. Participants will be informed that what they share in the group will



COLLEGE OF TECHNOLOGY

stay within the group. The researchers will only collect the necessary data for AI models and get clear permission for any other uses. Finally, the data acquired will be protected by encryption, and access will be exclusive to authorized researchers.

(3) Safety and Risk Assessment:

The researchers will focus on reducing and evaluating physical and mental harm. They will set up safety protocol checks and conduct pre-tests during ongoing evaluations of the navigation hat. This device has components such as batteries and sensors that might overheat, fail, or produce incorrect audio signals, which could lead to falls. During user testing, there will be emergency protocols to stop testing immediately and provide medical assistance if the device fails. Users may experience cognitive overload or sensory fatigue after prolonged device use. It is important to watch for adverse effects like disorientation or anxiety and to have psychological support available to ensure participants stay safe and well.

(4) Beneficence and Non-Maleficence:

Minimize harm while maximizing benefits. The researchers will monitor the FGDs for emotional distress when discussing navigation challenges and provide support resources. Researchers will also ensure venues are accessible to prevent physical or psychological harm to VIPs. The discussion improves the study's quality through user feedback and enhances the device design for VIPs, but researchers must address the prototype's expectations and limitations to avoid false anticipations. Improve the AI process without biased advice. Promote equity by making technology accessible and affordable for various VIP groups to avoid exclusion. Evaluate the long-term impact on overall well-being and the risk of isolation due



COLLEGE OF TECHNOLOGY

to decreased human interaction.

(5) Justice and Inclusivity:

Fair representation will be ensured when selecting participants and consultants for FGDs and when evaluating the prototype, including people from different backgrounds, with varying levels of impairment, ages, genders, and local cultural views on disability in the Philippines. This helps prevent bias and respects cultural norms. Accessibility is also essential. The researchers will use tools like audio aids for visually impaired participants and ensure that logistical issues do not exclude anyone. Our goal is to benefit persons with disabilities (PWDs), especially VIPs, beyond researchers and institutions, and to promote equal participation in research outcomes.

(6) Transparency and Integrity:

Transparency and integrity will clearly outline the study's objectives during focus group discussions (FGDs) and consultations to avoid misleading participants. The researchers will document how the input received influences the project, such as by refining the modular system. In addition, the researchers will seek neutral experts who can provide objective advice. Ethical practices will be followed, including appropriately attributing contributions while safeguarding participants' privacy to uphold research integrity.

(7) AI and Technology Ethics:

Potential biases in photogrammetry models or audio cues can happen when the training data is not representative. This can lead to inaccuracies and require



COLLEGE OF TECHNOLOGY

the researchers to conduct regular fairness checks. The researchers will clearly explain AI decision-making to participants, ensuring it aligns with human sensory experiences without being misleading. The researchers must also address dual-use concerns, such as using real-time tracking for surveillance, which requires safeguards to protect user privacy and prevent unauthorized data sharing.

(8) Inclusion and Exclusive Criteria:

The study will exclude VIPs aged 17 years old and below (minors) and 60 years old and above (senior citizens) as respondents. Only those aged 18-59 years old will participate in the study's survey, testing, and evaluation of the prototype.

(9) Regulatory and Institutional Compliance:

The researchers will submit plans for focus group discussions (FGD), consultations, and testing to the university's Institutional Review Board (IRB) or an advisor for approval. This is particularly important given the vulnerability of vulnerable populations (VIPs). The goal is to ensure compliance with ethical standards and legal frameworks, including the Philippines' Data Privacy Act, especially regarding recordings and information sharing. Same with testing the device, the researchers will ensure that sensor and camera data are anonymized and encrypted to comply with privacy laws, and establish emergency protocols for handling device failures during trials. Post-activity follow-ups, such as summaries of the FGD and consultation to build trust and allow feedback on the use of input. Researchers will disclose conflicts of interest, such as affiliations with tech companies, to ensure impartiality.



COLLEGE OF TECHNOLOGY

References

- Commère, L., & Rouat, J. (2023). Evaluation of short range depth sonifications for visual-to-auditory sensory substitution.
- Gershman, S. J. (2024). Habituation as optimal filtering. *iScience*, 27(8), 110523.
- Hamilton-Fletcher, G., Alvarez, J., Obrist, M., & Ward, J. (2021). Soundsight: A mobile sensory substitution device that sonifies colour, distance, and temperature. *Journal on Multimodal User Interfaces*, 16(1), 107–123.
- Han, Z., Li, S., Wang, X., Hu, X., Higashita, R., & Liu, J. (2026). Multi-path sensory substitution device navigates the blind and visually impaired individuals. *Displays*, 91, 103200.
- Hou, Y., Xie, Q., Zhang, N., & Lv, J. (2025). Cognitive load classification of mixed reality human computer interaction tasks based on multimodal sensor signals. *Scientific Reports*, 15(1).
- Iwasaki, M., & Inomata, H. (1986). Relation between superficial capillaries and foveal structures in the human retina. *Investigative Ophthalmology & Visual Science*, 27(12), 1698–1705.
- Krantz, J. H. (2012). Chapter 3: The stimulus and anatomy of the visual system. In *Experiencing sensation and perception*. Pearson Education.
- Kristjánsson, Á., Moldoveanu, A., Jóhannesson, Ó. I., Balan, O., Spagnol, S., Valgeirs-dóttir, V. V., & Unnþorsson, R. (2016). Designing sensory-substitution devices: Principles, pitfalls and potential. *Restorative Neurology and Neuroscience*, 34(5), 769–787.
- Loschky, L. C. (2025, October). Recognizing the gist of a scene [Accessed: 2025-11-06].



COLLEGE OF TECHNOLOGY

- Mishra, A., Bai, Y., Narayanasamy, P., Garg, N., & Roy, N. (2025). Spatial audio processing with large language model on wearable devices.
- RealSense AI. (2025). Eyesynth: Powering Sonic Vision for the Blind.
- Ruan, L., Hamilton-Fletcher, G., Beheshti, M., Hudson, T. E., Porfiri, M., & Rizzo, J.-R. (2025). Multi-faceted sensory substitution using wearable technology for curb alerting: A pilot investigation with persons with blindness and low vision. *Disability and Rehabilitation: Assistive Technology*, 20(6), 1884–1897.
- Sami, M., Agha, D., Baloch, J., Dewani, A., & Maheshwari, K. D. (2025). Efficient object detection and voice-assisted navigation for the visually impaired: The smart hat approach. *Sukkur IBA Journal of Computing and Mathematical Sciences*, 8(2), 1–12.
- Shinagawa Lasik & Aesthetics. (2025, May). Perfect vision and eye health in the philippines.
- Theodorou, P., Tsiligos, K., & Meliones, A. (2023). Multi-sensor data fusion solutions for blind and visually impaired: Research and commercial navigation applications for indoor and outdoor spaces. *Sensors*, 23(12), 5411.
- Tokmurziyev, I., Cabrera, M. A., Khan, M. H., Mahmoud, Y., Moreno, L., & Tsetserukou, D. (2025). Llm-glasses: Genai-driven glasses with haptic feedback for navigation of visually impaired people.
- Udayakumar, D., Gopalakrishnan, S., Raghuram, A., Kartha, A., Krishnan, A. K., Ramamirtham, R., Muthangi, R., & Raju, R. (2025). Artificial intelligence-powered smart vision glasses for the visually impaired. *Indian Journal of Ophthalmology*, 73(Suppl 3), S492–S497.



COLLEGE OF TECHNOLOGY

UserTesting. (2024, April). 7 gestalt principles of visual perception: Cognitive psychology

for ux.

Viancy V, N. R. C., M Mouli, M., & S, S. (2024). Aural vision - envisioning the independence of the visually impaired. *2024 International Conference on Inventive Computation Technologies (ICICT)*, 387–392.

World Health Organization. (2023, August). Blindness and visual impairment.

Zeki, S. (2015). Area v5 - microcosm of the visual brain. *Frontiers in Integrative Neuroscience*, 9.

Zhao, Y., Yang, S., Tao, Y., & Kang, H. (2025). Evaluation of the efficacy of an assistive device for blind people: A prospective, non-randomized, single arm, and open label clinical trial. *Current Eye Research*, 50(8), 865–869.

Zvorișteanu, O., Caraiman, S., Lupu, R.-G., Botezatu, N. A., & Burlacu, A. (2021). Sensory substitution for the visually impaired: A study on the usability of the sound of vision system in outdoor environments. *Electronics*, 10(14), 1619.



COLLEGE OF TECHNOLOGY

Appendices - Researchers and Adviser's Profile



JASON C. D'SOUZA

Address: Monticello, Balabag, Pavia, Iloilo City

Phone Number: 09282543420

Email Address: jdsouza@usa.edu.ph

EDUCATIONAL BACKGROUND

College University of San Agustin

Gen. Luna St., Iloilo City

2022 – Present

Secondary Central Philipine University

Lopez-Jaena St., Jaro, Iloilo City

2015 – 2021

Elementary St. Jerome Academy

Duenas, Iloilo

2011 – 2015



COLLEGE OF TECHNOLOGY

Appendices - Researchers and Adviser's Profile



CHENLIN WANG

Address: San Pedro, Molo, Iloilo

Phone Number: 09274553340

Email Address: cwang@usa.edu.ph

EDUCATIONAL BACKGROUND

College University of San Agustin

Gen. Luna St., Iloilo City

2022 – Present

Senior High School

MianYang Church Home Schooling

MianYang City, Sichuan Province, China

2017 – 2021

Junior High School

LONGMEN Middle School

TaiYuan City, ShanXi Province, China

2014 – 2017

Elementary LONGMEN Elementary School

TaiYuan City, ShanXi Province, China

2008 – 2014



COLLEGE OF TECHNOLOGY

Appendices - Researchers and Adviser's Profile



ETHEL HERNA C. PABITO

Present Address: Townsville Studios, Ma. Clara Ave.,

Aurora Subd., Iloilo City, Iloilo, 5000

Present Address: Home Address: Purok 1, Brgy. Igbalangao, Bugasong, Antique, 5704

Phone Number: 09663447178

Email Address: ehpabito@gmail.com

EDUCATIONAL BACKGROUND

College University of San Agustin

Gen. Luna St., Iloilo City

2022 – Present

Senior High School STEM

University of San Agustin

Gen. Luna St., Iloilo City

2019 – 2022

Junior High School

Pax Catholic Academy Diocese of Bacolod Inc.

CWGC+588, La Paz Street, La Carlota City, Negros Occidental, 6130

2015 – 2019

Elementary Nelia Boston Maghari Memorial Elementary School

Brgy. Igbalangao, Bugasong, Antique, 5704

2009 – 2015



COLLEGE OF TECHNOLOGY

Appendices - Researchers and Adviser's Profile



VINCE GINNO B. DAYWAN

Present Address: Sta. Justa, Tibiao, Antique

Present Address: Sta. Justa, Tibiao, Antique

Phone Number: 09991755994

Email Address: vgalitao-saan@usa.edu.ph

EDUCATIONAL BACKGROUND

College University of San Agustin

Gen. Luna St., Iloilo City

2020 – Present

Secondary University of Antique Tario-Lim Memorial Campus Laboratory High School

San Jose de Buenavista, Antique, Philippines

2014 – 2020

Elementary TIBIAO CENTRAL SCHOOL

72PP+CFX, Poblacion, Tibiao, Antique

2008 – 2014



COLLEGE OF TECHNOLOGY

Appendices - Total Budget

Table 6 : Total budget

Description	Quantity	Price	Supplier and Manufacture
Orange PI Zero 2W	1	₱1,560.00	Orange Pi Official Store
OV5640 USB Camera 5MP and 160° FOV	2	₱1,200.00	DigiKey Philippines
TRRS 3.5 Jack Breakout	1	₱25.00	MCXK-E Philippines
Type-C 5V/3A 18650 Lithium Battery UPS	1	₱121.00	SAMIORE Philippines
Stereo Earphone	1	₱101.00	Zeus Philippines
MPU6050	1	₱102.00	BBModule Philippines
SIM868 GSM GPRS GPS	1	₱660.00	BBModule Philippines
18650 Lithium Battery	2	₱70.00	BBModule Philippines
Filament	1	₱715.00	Makerlab

Continued on next page



COLLEGE OF TECHNOLOGY

Appendices - Gantt chart for Project Proposal Timeline

Table 6: Total budget (Continued)

Description	Quantity	Price	Supplier and Manufacture
Total		₱5,824	

Table 7 : Gantt chart Aug 2025 - Nov 2025

Task No.	Description	Week									
		Aug 2025	Sep 2025	Oct 2025	Nov 2025						
1	Identify Problem										
2	Brainstorming sessions to generate project ideas										
3	Evaluation of ideas based on feasibility and skills of team										
4	Review of Studies and Literature										

Continued on next page



COLLEGE OF TECHNOLOGY

Appendices - Gantt chart for Project Proposal Timeline

Table 7: Gantt chart Aug 2025 - Nov 2025 (Continued)

Task No.	Description	Aug 2025	Sep 2025	Oct 2025	Nov 2025
		Week			
5	Definition of objectives and scope of the project				
6	Identification of the hardware and software materials				
7	System Design / Block Diagram				
8	Outline Methodology & System				
9	Creation of Conceptual and Theoretical Framework				

Continued on next page



COLLEGE OF TECHNOLOGY

Appendices - Gantt chart for Project Proposal Timeline

Table 7: Gantt chart Aug 2025 - Nov 2025 (Continued)

Task No.	Description	Aug 2025		Sep 2025		Oct 2025		Nov 2025	
		Week							
10	Purchase Materials								
11	Program of Zero form factor SBC and other Hardware Components								
12	Point Cloud Projection & 3D Mapping Setup								
13	AI Model Setup and Fine-Tuning								
14	Building Prototype (Phase one – Circuit Prototype)								
15	Unit Testing								

Continued on next page



COLLEGE OF TECHNOLOGY

Appendices - Gantt chart for Project Proposal Timeline

Table 7: Gantt chart Aug 2025 - Nov 2025 (Continued)

Task No.	Description	Aug 2025	Sep 2025	Oct 2025	Nov 2025		
		Week					
16	Writing of Thesis						
	Proposal						
	Manuscript						
17	Thesis Proposal						
	Presentation						
18	Refinement of Project based on panelist feedback						
19	Design UI of Web Server						
20	3D Print Hat and Components Case Parts						
21	Building Prototype (Phase Two – Device Prototype)						



COLLEGE OF TECHNOLOGY

Appendices - Gantt chart for Project Proposal Timeline

Table 8 : Gantt chart Dec 2025 - Mar 2026

Task No.	Description	Dec 2025		Jan 2026		Feb 2026		Mar 2026	
		Week							
22	Calibration of the System								
23	Web Application Development								
24	Web Application Testing and Debugging								
25	Data Preparation								
26	IoT Sensors Integration with Web Application								
27	Debugging & Troubleshooting								
28	Refinement of prototype and manuscript								
29	Ethics Review								

Continued on next page



COLLEGE OF TECHNOLOGY

Appendices - Gantt chart for Project Proposal Timeline

Table 8: Gantt chart Dec 2025 - Mar 2026 (Continued)

Task No.	Description	Dec 2025	Jan 2026	Feb 2026	Mar 2026
		Week			
30	Letters and validation forms				
31	Implementation				
32	Interpretation and analysis of the gathered data				
33	Finalization of Manuscript for defense				
34	Final Defense				
35	Revisions based on the panelist's feedback				
36	Research Approval				
37	Hardbound				