Dean&Francis
ISSN 2959-6157

Design and Implementation of a Wearable Intelligent Tourism Guide System

Shuhe Yuan[1, *]

[1]School of Glasgow, University of Electronic Science and Technology of China, Chengdu, 611731, China

* Corresponding author: 2720925Y@student.gla.ac.uk

Abstract:

The substantial growth of the global tourism industry has increased the demand for advanced and personalized travel services. In response, the "Changyou" intelligent tour guide device was developed, integrating technologies such as the Global Navigation Satellite System (GNSS), speech and image recognition, and real-time internet connectivity. This study explores the device's design and evaluates the performance of "Changyou" in various scenarios, focusing on speech recognition accuracy, image identification efficiency, system response time, and navigation precision. The findings indicate that the device performs effectively under controlled conditions; however, its performance is affected by factors such as background noise and low lighting. The study also analyzes the system's stability and scalability. The results suggest that "Changyou" has the potential to evolve into a comprehensive smart tourism service product; however, further research is needed to improve its adaptability and expand its functionality.

Keywords: Wearable technology; Smart tourism; Intelligent navigation; IoT

1. Introduction

The global tourism industry has grown rapidly in recent years. Travel has become an essential part of modern life. As demand for tourism increases, travelers are seeking more intelligent and personalized services [1]. Traditional guide devices mainly offer basic navigation and information. They often do not meet the varied needs of today's travelers. This is especially true in areas like real-time information, accurate route planning, and emergency handling [1, 2].

There is a rising demand for intelligent guide devices. These devices should integrate multiple functions. They need to provide real-time navigation using GNSS, voice recognition, and detailed sightseeing information. Additionally, they should offer dining and accommodation recommendations and emergency assistance. Wearability, ease of use, and long battery life are also becoming more important.

However, several problems and challenges persist in traditional research. Firstly, many devices lack the ability to update information in real-time, which is crucial for travelers navigating unfamiliar environments. Secondly, traditional guide devices offer limited personalization and cannot be easily adapted to meet individual preferences, such as language settings or customized sightseeing suggestions. More-over, the accuracy of navigation in complex environments, especially indoors or in densely populated cities, remains inadequate, failing to meet users' expectations for precise guidance.

To address these issues, this study develops the "Changyou" intelligent guide device. This device combines the latest GNSS technology with voice and visual recognition, information retrieval, and internet connectivity. It provides not only real-time navigation but also personalized services based on user preferences. By integrating advanced features, "Changyou" aims to make travel more convenient, safe, and enjoyable.

The "Changyou" intelligent guide device was developed to meet these needs. It combines the latest GNSS technology with voice and visual recognition, information retrieval, and internet connectivity. By integrating these advanced features into a portable device, "Changyou" aims to make travel more convenient, safer, and enjoyable. The innovation of this study lies in integrating GNSS technology, voice and visual recognition, information retrieval, and internet interaction. By offering a more precise and intelligent travel experience, this research paves the way for the design and application of smart guide devices, significantly enhancing travel convenience and safety.

2. Methodology

2.1 Hardware Design

2.1.1 Main Control Chip

Built on ARM® Cortex®-M7 architecture, it boasts 32-bit single-core processing capability and a maximum frequency of 480MHz. Support features include brown-out detection, DMA, PWM, WDT, and LCD display; the chip combines a 2x12-bit D/A converter with a 36x16-bit A/D converter. For managing plenty of data, it comprises 1MB of RAM and 2MB of Flash program memory. Effective communication with external modules is made possible by the peripheral interfaces CANbus, I2C, IrDA, LINbus, MMC/SD/SDIO, QSPI, SPI, UART/USART, and USB OTG [3].

2.1.2 Positioning Module

The device has a GPS module (AGTM336H), and it takes signals from global navigation satellite systems like GPS and GLONASS and Galileo as well as other kinds of satellite systems. The signals received by this module are processed, giving location information that is geographically accurate. The module does give sub-meter positioning accuracy because it takes signals from multiple sources to process them. It is useful for things such as navigation, route planning, and related tasks. It also can work in several environments and give continuous location updates to help with navigation system operation being maintained normally [4].

2.1.3 Internet Access Module

The gadget allows a Wi-Fi connection using the ESP8266 module. Highly integrated wireless communication module ESP8266 offers data transfer speeds of up to 11 Mbps. By use of this module, the gadget can contact the internet at any moment, therefore enabling information retrieval, map updates, and cloud data synchronization. The real-time internet connection function guarantees that the gadget can give consumers the most recent travel information and services, so smoothing out their whole travel experience and increasing their convenience [5].

2.1.4 Voice Recognition Module

The LD3320 module is utilized in the voice recognition component of this device. It is a module utilized for voice identification and processing, particularly for vocal interaction. It can interpret the user's vocal commands, and accurately recognize and execute them, achieving a voice recognition precision of 97% and a reaction time under 200 milliseconds [6]. This hardware satisfies the design specifications for the product.

2.1.5 Visual Recognition Module

The apparatus primarily employs the OpenMV module for visual recognition. OpenMV is an open-source embedded image processing platform with efficient capabilities, making it fit for this product's requirements. It can rapidly implement diverse picture recognition algorithms and support several visual recognition jobs. The OpenMV module features different interfaces for connecting cameras, sensors, and other peripherals, supports diverse image formats and resolutions, and facilitates real-time image processing. Utilizing OpenMV, the application may identify visual data, extract features, and furnish picture information for the system. The OpenMV module satisfies the specifications of this product [7].

2.1.6 Touchscreen Module

The device is equipped with an ATK-MD0430RGB LCD touch screen as the user interface. Users can connect to Wi-Fi, set navigation paths, and perform other operations through the UI interface configured on the touch screen. The touch screen design can provide the product with a more flexible operation method as a supplement to voice input.

2.1.7 Power Module

The device is equipped with a high-capacity lithium battery to ensure prolonged usability. The battery supplies power to the device, enabling continuous operation during

ISSN 2959-6157

extended travel without the need for frequent recharging, thus improving battery longevity. The key parameters of the power module and other system components are listed in Table 1.

Table 1. Key parameters of the intelligent guide device system

Item	Value
Main control chip	STM32H743IIT6
Positioning accuracy	1.25 m (ATM336H Module)
Navigation update frequency	5 times/sec (Gaode Map API)
Internet connection speed	11 Mbps (ESP8266 Wi-Fi)
Touch screen response time	<20 ms (ATK-MD0430RGB LCD)
Voice recognition accuracy	97% (LD3320 Module)
Voice response speed	<200 ms (LD3320 Module)
Visual recognition accuracy	92% (OpenMV Module)
Review information retrieval time	<1 sec (Dazhongdianping API)
Battery life	48 hours (Low Power Design)
Device weight	Approx. 750 g
Operating temperature range	-20°C to 60°C
External communication protocol	Wi-Fi
Internal communication protocol	UART/UASRT

2.2 Software Architecture

The software architecture of the “Changyou” intelligent guide device is designed to facilitate seamless interaction between various modules, including GNSS, voice recognition, visual recognition, and communication. These modules work together to ensure real-time navigation, information retrieval, and intelligent services.

As shown in Fig. 1, the functional block diagram illustrates the interactions between the system’s modules, highlighting how the communication module collaborates with cloud-based databases and external APIs to provide users with accurate and timely information. This architecture supports several core functions, ensuring a smart and convenient travel experience for users.

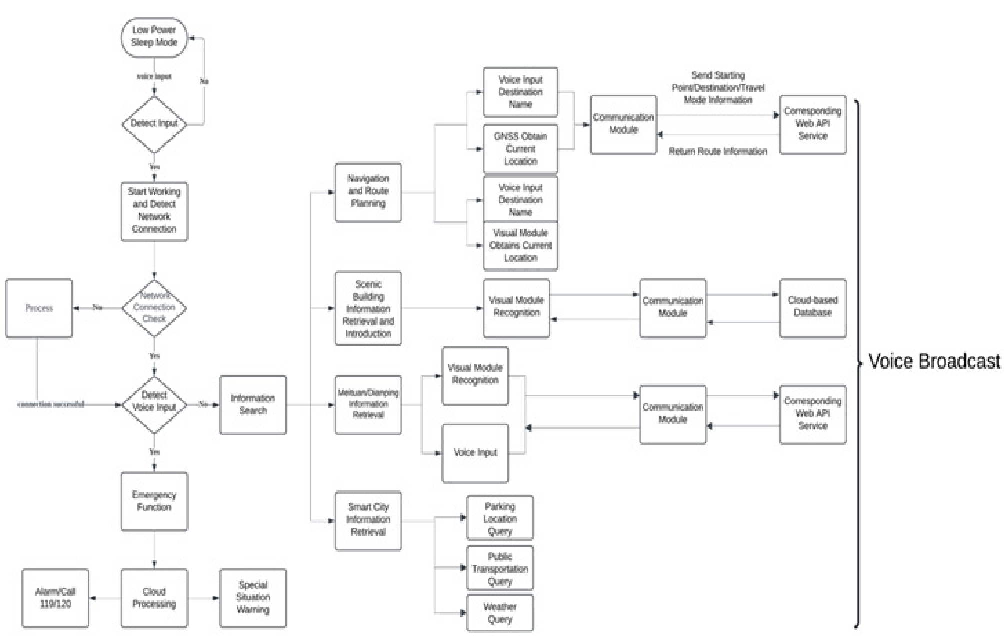

Fig. 1 Functional block diagram. (Photo/Picture credit: Original)

2.2.1 Voice Recognition Module

2.2.1 .1 System Initialization Module

This module initializes the necessary hardware and software environment for speech recognition, ensuring the system is operational at startup. The initialization procedure entails resetting the LD3320 chip to eliminate any prior states, executing standard initialization for the LD3320 by configuring fundamental parameters, and doing specialized initialization for Automatic Speech Recognition (ASR) to establish the recognition environment. Upon completion of initialization, the system emits a signal signaling readiness to commence the voice recognition task.

2.2.1 .2 Main Program Logic Module

This module constitutes the foundation of the speech recognition capability, managing user voice command processing and delivering recognition results. Upon receiving a voice command, the system endeavors to initiate the ASR (Automatic Speech Recognition) process, retrying up to five times if required. Upon initiation of the recognition process, the system loads the target keywords and verifies whether the LD3320 chip is in an idle condition. When the chip is inactive, the system adjusts the microphone settings and initializes the pertinent registers before commencing the identification process. Upon completion, the system acquires the recognition results for subsequent processes.

2.2.1 .3 Interrupt Handling Module

Upon receiving an audio signal, this module initiates, processes the recognition results, and changes the system status. The system reads the recognition result, assesses its validity, modifies the internal state variables, reconfigures the LD3320's registers in order to get ready for the subsequent recognition task, and then sends the main program the final recognition result.

2.2.1 .4 Auxiliary Function Module

This module is utilized to ascertain the condition of the speech recognition system and get the recognition outcomes. The system initially verifies if the LD3320 is in an idle condition to ascertain the feasibility of initiating a fresh recognition task. This module is in charge of acquiring the final recognition results and distributing them to other modules in the system for use.

2.2.2 Visual Recognition Module

2.2.2 .1 System Initialization Module

The visual recognition system is initialized using this part of the system. The first step in the process is to return the camera sensor to its initial state. The frame size and pixel format are then modified to satisfy the particular needs of the application. In order to complete the design, the system then determines the size of the sensor window and defines the region for image processing. The module announces that the visual recognition system is ready for use by sending out a signal after initialization is complete.

2.2.2 .2 Model Loading Module

The module loads the pre-trained model and label file for object detection. The system initializes the TensorFlow Lite model, encompassing the weights and architecture for image recognition, and accesses the label file to correlate identified items with their corresponding labels. Upon loading the model and label file, the system initiates the object detection process.

2.2.2 .3 Main Program Logic Module

This module oversees visual recognition, orchestrating the complete procedure of image acquisition and object identification. The system utilizes the current frame as input and employs a pre-loaded model to identify items inside the image. Background classes are omitted to optimize compute resources, and the central coordinates of the identified items are computed. The detection findings are annotated on the image and either output to the console or transmitted over serial transmission for subsequent processing or display.

2.2.2 .4 UART Communication Module

This module outputs detected labels and coordinates via the UART interface for use by external devices or systems. Object labels, center coordinates, and frame rate information are transmitted through the UART to ensure accurate data delivery to external devices or systems, enhancing the system's interactivity and functionality.

2.2.3 Positioning and Navigation Module

2.2.3 .1 System Initialization Module

This module initializes the hardware and software environment for the positioning and navigation module, ensuring that the GNSS and IMU modules function correctly. During initialization, the system first sets up the GNSS and IMU modules, configuring necessary communication interfaces and parameters to ensure efficient communication with the main control chip. After initialization, the system outputs a signal indicating that the positioning and navigation module is ready to begin acquiring location information.

2.2.3 .2 GNSS Positioning Module

This module uses the GNSS system to obtain the current geographical location information, including latitude and longitude. After a user triggers the positioning operation

ISSN 2959-6157

via voice input, the system retrieves the current location information from the GNSS module and parses the received GPS data into actual latitude and longitude coordinates. The parsed coordinate information is then outputted for use by the navigation module, ensuring accurate positioning and navigation information during the user's journey.

2.2.3 .3 IMU Module

This module gathers current directional and acceleration data to assist in precise navigation over short distances. The system obtains the user's attitude data (e.g., roll, pitch, and yaw angles) and gyro and accelerometer data through the IMU module, updating the user's directional information in real-time to support more accurate navigation in confined spaces.

2.2.3 .4 API Calling Module

This module calls the Amap API to retrieve the user's destination coordinates and returns detailed navigation route information. Upon receiving the user's voice command and current coordinates, the system sends an API request to obtain the destination coordinates and generate the corresponding travel route. The parsed route information is then provided to the user to guide them during their journey, ensuring they reach their destination smoothly.

2.3 System Integration and Coordination

As shown in Fig. 2, this image represents the integration of the system's hardware components.

Fig. 2 Integration diagram. (Photo/Picture credit: Original)

3. Results and Discussions

3.1 Voice Recognition Accuracy Test

In the voice recognition accuracy test, the controlled environment of the laboratory and campus resources were utilized to simulate three typical testing scenarios. First, the test was conducted in a soundproofed laboratory, ensuring a completely quiet environment to evaluate the system's performance under ideal conditions. Next, testing was performed in an open area on campus with moderate background noise, including natural sounds, conversations, and distant traffic. Portable recording devices provided by the lab were used to play additional background noise, further simulating a typical outdoor environment. Finally, a high-noise environment, such as a busy street or construction site, was simulated using high-powered speakers and audio equipment outside the laboratory, and the system's voice recognition performance was recorded under these extreme noise conditions [8].

As shown in Table 2, the system performed best in a quiet indoor environment, achieving a recognition accuracy of 97%. However, as noise levels increased, performance declined. In a moderately noisy outdoor environment, accuracy dropped to 92%. In a high-noise outdoor environment, accuracy further decreased to 83%. These results indicate that while the system is highly effective under ideal conditions, its performance is significantly impacted by higher noise levels.

Table 2 Voice recognition accuracy test results

Test Environment	Recognition Accuracy (%)
Quiet indoor	97
Outdoor (moderate noise)	92
Outdoor (high noise)	83

3.2 Visual Recognition Effectiveness Test

The visual recognition tests were conducted both in the laboratory and under various lighting conditions around the campus. For the bright light conditions, high-brightness LED lighting in the laboratory was used to ensure ample and evenly distributed light, testing the system's recognition capability in optimal lighting. In low-light conditions, the laboratory's lighting equipment was adjusted to reduce light intensity, and blackout curtains were used to further diminish light, simulating a dim environment and evaluating the system's performance in low-light scenarios. For complex background environments, various backgrounds with different colors and textures were set up in the lab, placing test objects against these challenging backdrops to observe the system's accuracy and stability when handling visually complex scenes.

As shown in Table 3, under bright light conditions, the system achieved a recognition accuracy of 94%, demonstrating strong performance. However, in low light conditions, accuracy dropped to 83%, indicating the impact of insufficient lighting on the system's effectiveness. In complex background environments, the accuracy was 80%, showing that the system faces challenges when dealing with visually complex scenes.

Table 3 Visual recognition effectiveness test results

Test Environment	Recognition Accuracy (%)
Bright Light	94
Low Light	83
Complex Background	80

3.3 System Response Speed and Battery Life Test

The system response speed and battery life tests were conducted in the laboratory, covering both normal and high-intensity usage scenarios. In the normal usage scenario, the laboratory provided a stable testing environment to simulate typical operations, such as navigation, voice command input, and information retrieval. Monitoring devices were used to record the system's response times and tracked battery consumption with power monitoring equipment [9]. For the high-intensity usage scenario, automated testing equipment was employed in the lab to simulate continuous navigation, frequent voice recognition tasks, and high-frequency data processing, recording the system's response times and battery life under these heavy workloads.

As shown in Table 4, in normal usage scenarios, the system's battery life was 45.3 hours, indicating excellent performance. Under high-intensity usage scenarios, battery life decreased to 18.2 hours. This reflects that the system consumes battery power more quickly under heavy usage conditions.

Table 4 System response speed and battery life test result

Usage Scenario	Response Time (ms)	Battery Life (hours)
Normal Usage	200	45.3
High-Intensity Usage	300	18.2

3.4 Navigation and Route Planning Accuracy Test

The navigation and route planning tests combined laboratory resources with real-world terrain on campus. In the lab, Geographic Information System (GIS) tools were used to simulate the complex urban environment surrounding the campus, incorporating the layout of cam-

ISSN 2959-6157

pus buildings to evaluate the system's path planning and navigation functions. Practical tests were then conducted in various areas of the campus, such as open fields, dense building areas, and hilly or stair-covered regions.

Using GPS recording tools provided by the lab, the actual routes taken were compared with the system-generated routes to analyze the system's accuracy and effectiveness in navigating complex and dynamic environments, as shown in Fig. 3. Fig. 3 illustrates the route used for testing the navigation and route planning accuracy of the system. Fig. 4 shows the navigation feedback displayed through a serial port. The system processes location data and provides real-time instructions, such as specific directions and the distance to the next point. This output is part of the navigation accuracy evaluation, allowing the user to follow the system's guidance based on current position data.

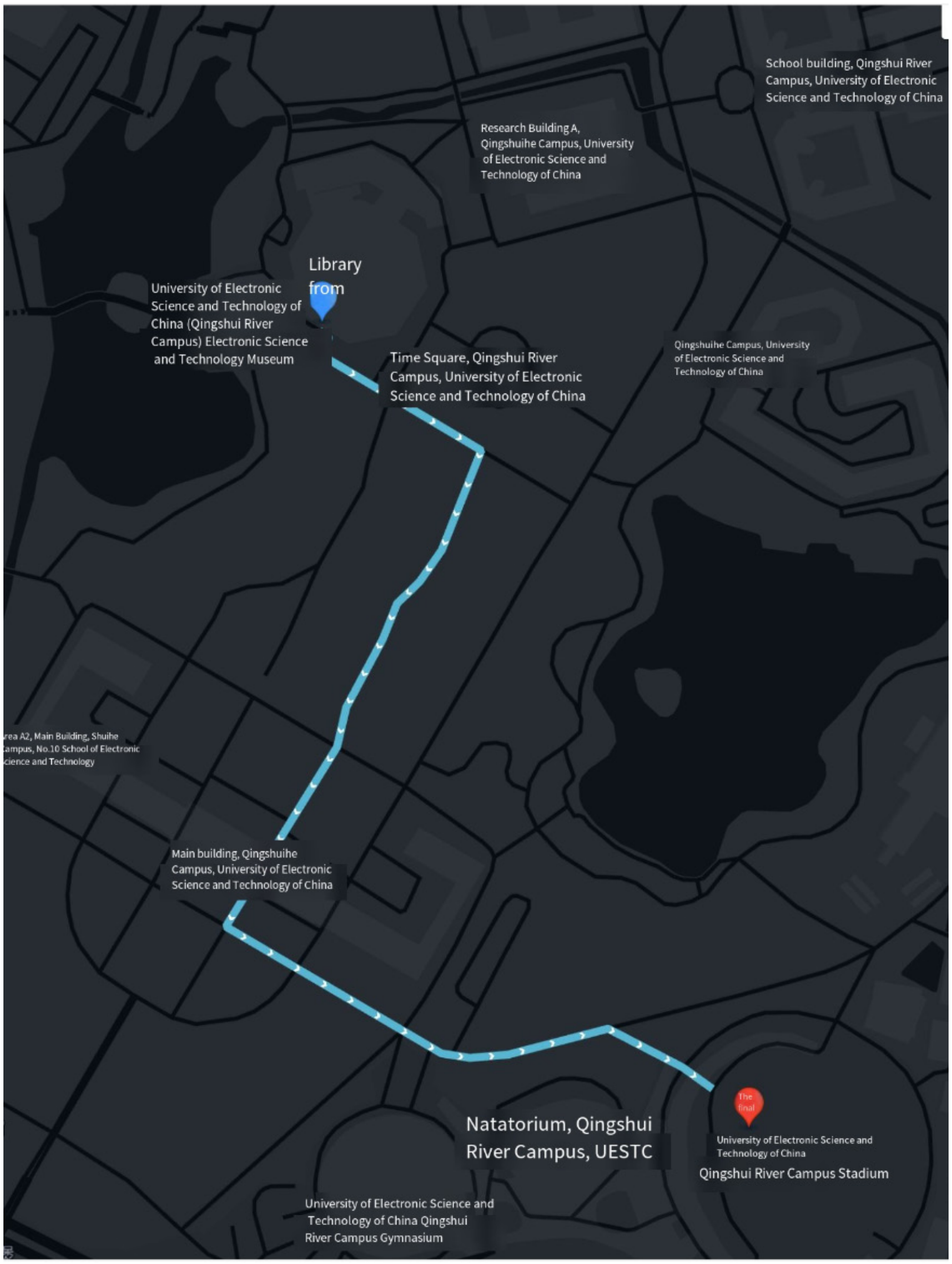

Fig. 3 Route planning map. (Photo/Picture credit: Original)

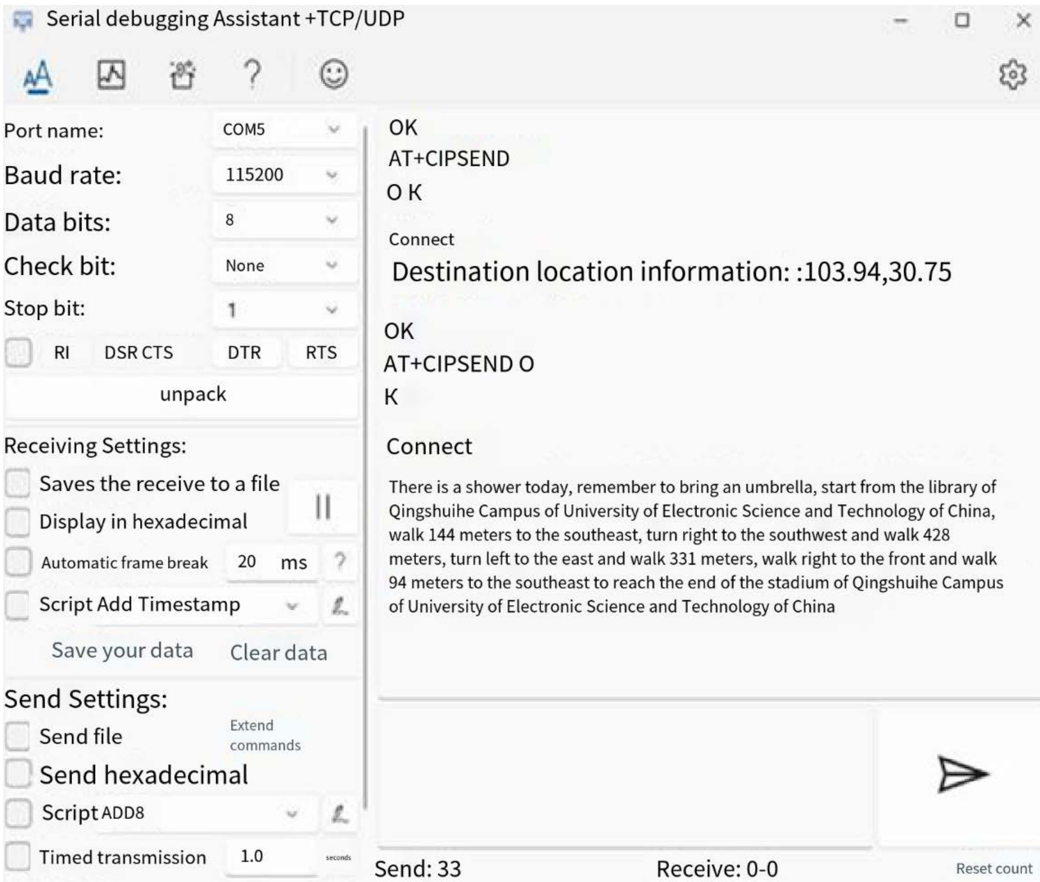

Fig. 4 Navigation feedback serial port display. (Photo/Picture credit: Original)

The navigation and route planning tests, combining laboratory simulations with real-world campus terrain, showed that the system generally provided accurate path planning and navigation guidance. However, in some complex terrains and dynamic environments, there were discrepancies between the system-generated routes and the actual paths taken, indicating the need for further optimization.

3.5 Improvements for the Voice Recognition Module

Two stages of command optimization were used to enhance voice recognition performance in noisy conditions. The user input is split into two steps using this strategy. First, the system needs a predetermined keyword (like "Changyou"). When the user enters additional specific information in the second step—such as going to location A or playing music—the system will prompt them to do so. In noisy surroundings, the accuracy of recognition is enhanced by this tiered processing.

3.6 Analysis of System Stability and Scalability

While the system functions flawlessly in a test setting, when more people join or new features are added, stability issues could arise. The system is modular in nature, with independent operation for every functional module. It is possible to introduce new features individually without interfering with how current functionalities are normally used. The system can schedule tasks and manage resources, and it can function well even when there is a lot of demand on it. Through cloud computing, some processing programs can be moved to the cloud to enable dynamic resource allocation in response to user volume increases. This will provide enough support for upcoming product upgrades, enabling them to adjust to changes in the market [1,10].

4. Conclusion

The research investigates a system architecture that incorporates GNSS technology, speech recognition, visual recognition, and internet connectivity. To improve voice recognition in cacophonous settings, it utilizes a dual-tier command framework. The system intends to deliver real-time navigation and customized services. Assessments in diverse demanding contexts exhibit stability, precision, and applicability potential.

In serene, regulated settings, the system exhibits outstand-

ISSN 2959-6157

ing performance, attaining elevated precision in auditory and visual identification tests. Performance diminishes in noisy and low-light environments. The dual-tier command framework markedly enhances voice recognition in cacophonous settings, demonstrating the efficacy of this optimization approach. The overall system has commendable stability and durability, suggesting applicability in real-world scenarios.

The study has inherent limitations. Testing was conducted within a restricted spectrum of environmental conditions. Additional investigation in varied and harsh environments is essential to thoroughly evaluate robustness. The primary emphasis was on small to medium-scale experiments. Subsequent efforts should extend to larger-scale implementations to assess scalability and performance in more extensive applications. Research should advance noise canceling methods and improve adaptability to diverse illumination conditions. This will enhance adaptability and reliability in practical applications.

References

[1] Ortiz Rincon, F., Tommasini, E., Rainoldi, M., Egger, R. The Future of Wearable Devices On-Site: A Scenario Technique Approach. In: Schegg, R., Stangl, B. (eds) Information and Communication Technologies in Tourism, Springer, Cham, 2017.

[2] K. Zhang et al., Successful Cases of Smart Tourism Construction on Analysis and Research - Taking Hangzhou and Xiamen as Examples, 2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Jakarta, Indonesia, 2021, pp. 285-290.

[3] InvenSense. MPU-6050. Last accessed 2024/09/10. https://invensense.tdk.com/products/motion-tracking/6-axis/ mpu-6050/

[4] Ai-Thinker Technology Co., Ltd., ATGM336H GNSS Module Datasheet, 2019. Last Accessed: Sept. 1, 2024. https://www.ai-thinker.com/ATGM336H.

[5] Espressif Systems, ESP8266EX Datasheet, Last Accessed: Aug. 30, 2024. https://www.espressif. com/ sites/default/files/documentation/0aesp8266ex_datasheet_en.pdf.

[6] LD3320 Voice Recognition Module Datasheet, Universal Scientific Industrial (Shanghai) Co., Ltd., Last Accessed: 2024/09/01. https://www.waveshare.net/wiki/LD3320_Board.

[7] OpenMV H7 Datasheet, OpenMV, Last Accessed: September 1, 2024. https://www.sparkfun. com/datasheets/Sensors/OpenMV-H7.pdf.

[8] Frewat, G., Baroud, C., Sammour, R., Kassem A., and Hamad, M., Android voice recognition application with multi speaker feature, 2016 18th Mediterranean Electrotechnical Conference (MELECON), Lemesos, Cyprus, 2016, pp. 1-5.

[9] Chen, C., Chen, B., Wang, T., Zheng, J., and Lin, Z., Battery Health Assessment and Life Prediction in Battery Management System, 2022 3rd International Conference on Electronic Communication and Artificial Intelligence (IWECAI), Zhuhai, China, 2022, pp. 87-90.

[10] Han, W., Hubei All-Area Tourism Development under the Background of Big Data, 2023 5th International Conference on Decision Science & Management (ICDSM), Changsha, China, 2023, pp. 219-22.