# An Affordable Intelligent Navigation Backpack for the Visually Impaired: Deep Learning-Based Obstacle Detection and Real-Time Navigation with RGB-D Integration

*Cheng-Li Luo[1], Hai Xu[1], Min Liu[1], Shu-Chuan Chu[2*]*

*[1]School of Information Engineering, Yango University, China*
*[2]School of Artificial Intelligence, Nanjing University of Information Science and Technology, China*
*clluo@ygu.edu.cn, xuhai_fafu@163.com, lumiaqkzj@live.com, scchu0803@gmail.com*

## Abstract

According to the World Health Organization (WHO), approximately 285 million people worldwide suffer from some form of visual impairment, including 39 million who are blind and an additional 246 million experiencing severe visual impairment. Existing navigation aids often fail to provide a user-centric perspective, relying on secondary judgment and leading to inconvenience. High-tech devices, such as smart glasses and robots, offer more effective solutions but are frequently cost-prohibitive. This study presents an affordable, first-person perspective intelligent navigation backpack for the visually impaired, utilizing deep learning. The system integrates RGB images and depth maps via alignment algorithms, extracts obstacle contours through binary image processing, and detects obstacles in real-time using the YOLO model. Experimental results demonstrate that the navigation depth camera significantly outperforms traditional ultrasonic and LiDAR sensors, achieving up to 98% measurement accuracy.

**Keywords:** Assistive navigation system, Image recognition, Deep learning, Image morphology

## 1 Introduction

The World Health Organization (WHO) reports that approximately 285 million people worldwide have vision impairments, with 39 million being blind and 246 million experiencing severe visual impairment. In China, around 17 million individuals are visually impaired, of which about five million are blind, constituting 18% of the global blind population [1]. This large number of visually impaired people urgently needs support to lead normal lives. To meet this need, researchers have developed a variety of navigation aids. Simultaneously, the rapid advancement of technology, especially in artificial intelligence (AI), has brought significant progress in speech and visual recognition, offering new possibilities for improving the lives of the disabled.

AI's capabilities in visual recognition have led to the creation of innovative technologies that convert visual information into auditory signals, benefiting those with limited vision. For example, the combination of Raspberry Pi 4B and Convolutional Neural Networks (CNNs) has been used for object classification and optical character recognition (OCR) [2-3]. Google Glass, with Microsoft Azure's customized visual API service, provides real-time image-to-speech feedback [4]. A system integrating wearable gloves and Android smart devices, equipped with cameras and tactile feedback, can inform users of environmental details like pedestrian crossings [5]. Lightweight neural network models such as YOLO and MobileNet, SSD also play a crucial role in quickly and accurately recognizing objects in real-time [6].

Speech recognition technology, another key area of AI, has also evolved significantly. Starting with the traditional Continuous-Time Convolution (CTC) model for speech transcription, the field has advanced with the adoption of deep learning, especially Recurrent Neural Networks (RNNs) and end-to-end (E2E) training frameworks, greatly improving transcriber performance [7]. Recent advancements in vision transformers (ViTs) and self-supervised learning have further enhanced obstacle detection robustness. Models like Swin Transformer and DINOv2 demonstrate superior performance in low-light and occluded scenarios, offering potential for future integration with lightweight YOLO architectures [8-9].

Alongside AI's progress, sensor application research has thrived. Ultrasonic sensors, valued for their non-contact measurement, low cost, and environmental adaptability, have found wide use in obstacle detection, object localization, and vehicle positioning [10-13]. By introducing improved genetic algorithms in path planning, novel ultrasonic sensor array designs, and advanced positioning algorithms, it seeks to boost system performance in obstacle detection, object localization, and vehicle positioning, offering more robust solutions for practical use [14-17]. Recent work by Lin et al. demonstrated a smartphone-based navigation system using ultrasonic sensors and GPS, achieving 90% accuracy in obstacle avoidance [18].

In the domain of assisting the visually impaired, ultrasonic sensors are a popular research focus for safe navigation. Systems using these sensors can detect

*Corresponding Author: Shu-Chuan Chu; E-mail: scchu0803@gmail.com
DOI: https://doi.org/10.70003/160792642025032602011

obstacles and warn the visually impaired [19-23]. However, challenges like signal interference and inaccurate positioning remain, calling for further improvement.

The visually impaired community has limited access to innovative guide products due to their low per capita income, leading to a lack of research and development. Consequently, the available products are both limited in variety and expensive. Compared to the widely used ultrasonic obstacle avoidance method, this intelligent navigation system has overcome technical challenges. Traditional ultrasonic obstacle avoidance systems often face limitations in achieving a larger monitoring range, requiring the addition of multiple ultrasonic sensors arranged in an array, with distance measurements typically limited to 2-3 meters. This hinders effective obstacle recognition and environmental assessment, resulting in a bulky system structure. The intelligent guide system in this study excludes the traditional ultrasonic detection scheme and uses the RGB image and depth map alignment algorithm to accurately identify obstacles and the current environment. The primary innovative contributions of this navigation system are as follows:

1. Based on a first-person perspective navigation system, it aligns with individuals' customary habits and does not require specialized learning of operational methods, ensuring low usage barriers and practicality.

2. Compared to traditional ultrasonic detection methods, it uses a depth camera to effectively recognize obstacles and the surrounding environment, with a simple system structure and high cost-effectiveness.

3. Leveraging the characteristics of a drone's first-person perspective, the software workflow of the navigation system is designed. Utilizing multi-sensor data fusion technology on drones and advanced algorithms used in aerial imaging analysis can more accurately identify road signs, pedestrian crossings, public transportation facilities, and other elements in the surrounding environment.

The remaining sections of this paper are structured as follows. Section 2 elaborates on the guide system structural design. Section 3 details the proposed algorithms including depth image alignment algorithm, image morphological processing and YOLO deep network model. The simulation results are discussed in Section 4. Lastly, Section 5 presents concluding remarks summarizing the key findings.

## 2 Guide System Structural Design

To provide a first-person perspective for the blind, we have developed a pioneering guide backpack based on robotic learning. It includes a standard backpack, a high-definition camera, a GPS locator, and a high-frequency vibrator, as depicted in Figure 1. The hardware circuit integrated into the backpack utilizes an STM32F407VGT6 microcontroller (STMicroelectronics, 32-bit ARM Cortex-M4 core, 168 MHz clock speed) as the core hardware controller and facilitates data transmission and interaction through various communication methods such as serial ports and GPRS, as illustrated in Figure 2.


**Figure 1.** Structure diagram of the smart assistive blind backpack

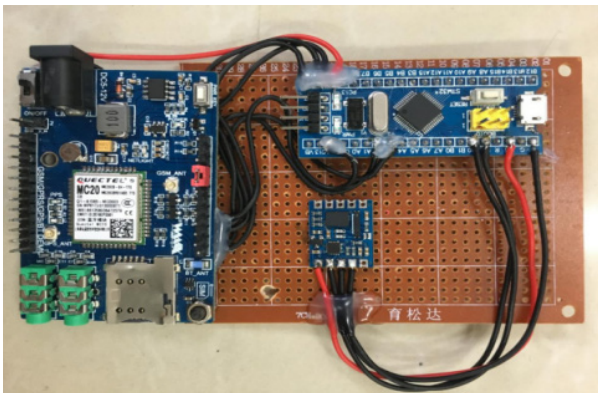

**Figure 2.** Hardware circuit embedded in the smart assistive blind backpack

The first-person perspective guide system is built around an embedded real-time operating system (RTOS) and incorporates features like voice announcements, GPS/IMU positioning, vibration feedback, and remote monitoring, as depicted in Figure 3. The hardware system diagram (Figure 3) illustrates the integration of key components. The STM32F407VGT6 microcontroller (STMicroelectronics) serves as the central processing unit, coordinating data from the RGB-D camera (Intel RealSense D435i), GPS module (U-blox NEO-M8N), and vibration feedback module. The RGB-D camera captures synchronized color and depth streams at 30 FPS, transmitted via USB 3.0 to the microcontroller. The GPS module provides real-time location updates through UART communication, while the vibration motor (Precision Microdrives 310-101) generates directional tactile feedback based on obstacle proximity.

Initially, the voice announcement module communicates received information to the embedded RTOS, which the STM32 then processes. Simultaneously, the STM32 receives data from the GPS/IMU, sending it to the RTOS for processing. Once processed, the RTOS returns the results to the STM32, which controls the buzzer or vibration motor to activate corresponding functions based on this data. Additionally, the STM32 transmits processed data to the server for remote monitoring and management. Meanwhile, a mini-program can retrieve relevant information from the server and display it on the user interface.

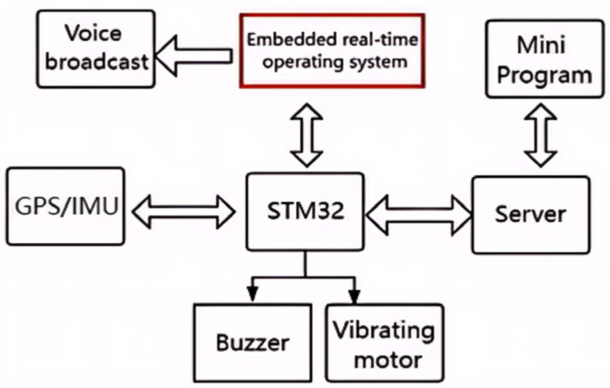


**Figure 3.** Hardware system diagram

# 3 Image Processing Algorithm

A navigation system relies on image processing algorithms to ensure the safety and environmental perception of visually impaired individuals during travel. These algorithms can analyze the real-time image information captured by cameras or visual sensors. Through a series of technical means, including environmental recognition, obstacle detection, and edge and color analysis, they accurately assess critical information such as pedestrian pathways, traffic signs, road conditions, and obstacle positions. This helps visually impaired users avoid risks and stay on the correct path. Additionally, these algorithms can effectively capture and interpret motion states for dynamic scenarios like moving people or vehicles, providing timely obstacle warnings. Furthermore, by combining advanced image processing and machine learning technologies, a navigation system can achieve augmented reality assistance. This involves converting visual information into auditory or tactile feedback, allowing users to "hear" road conditions ahead or "sense" changes in direction or obstacle distances. Even in indoor environments, image processing algorithms can assist visually impaired individuals in precise positioning and path planning through methods such as landmark recognition, QR codes, or RFID tags. Therefore, applying image processing algorithms significantly enhances the adaptability, response speed, and multi-functionality of navigation systems, providing strong support for the independent, safe, and efficient mobility needs of visually impaired users in various scenarios.

## 3.1 RGB Image and Depth Image Alignment Algorithm

Given the different spatial coordinate systems corresponding to RGB image data and depth image data, where the origin coordinates of RGB images are based on RGB cameras and those of depth images originate from infrared cameras, a certain degree of error exists when integrating and processing these two types of data. To address this issue, the system employs the RANSAC (RANdom SAmple Consensus) point set alignment algorithm for calibration and the process is illustrated in Figure 4. This algorithm initially converts the 2D points on the depth image to 3D space in a unified world coordinate system and then accurately projects these 3D points back onto the color image, achieving effective alignment and fusion of the two types of image data, with the logical relationship depicted in Figure 5.

RANSAC point set alignment algorithm

Input: Point set and maximum tolerance distance
Output: Optimal rigid body transformation T(R, t)
$\Theta \leftarrow \varnothing$, $c_{\max} \leftarrow 0$
**repeat**
Take 3 pairs of matching points to calculate R and t
    Inliers $\leftarrow \varnothing$, $\forall\, U_i : c_i \leftarrow 0$
    **foreach** $\langle p_i, q_i\rangle \in \{p_i\}, \{q_i\}$ **do**
        **if** $\| q_i - (R\cdot p_i + t) \| < d_{\max}$
            AND $| z_p - z_q | < \zeta$ **then**
            Inliers $\leftarrow$ Inliers $\cup \{i\}$
            $c_j \leftarrow c_j + 1$, iff $p_i \in U_j$
        **end**
    **end**
    $C \leftarrow \sum_i \min(c_i, \eta)$
    **if** $C > C_{\max}$ **then**
        $\Theta \leftarrow$ Inliers, $c_{\max} \leftarrow C$
    **end**
**until** (Iteration > maxIterations)
$(R, t) \leftarrow \underset{R,t}{\arg\min} \sum_{i\in\Theta} \| M\cdot(q_i - (R\cdot p_i + t))\|^2$
**return** $(R, t)$

**Figure 4.** RANSAC point set alignment algorithm

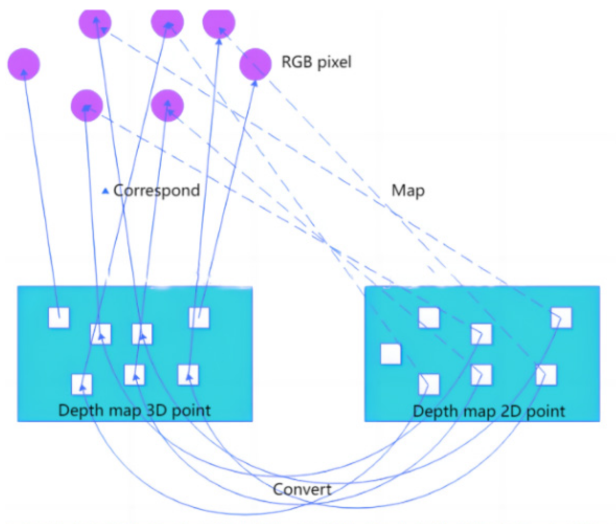


**Figure 5.** Mapping relationship between RGB image and depth image

### 3.2 Image Morphological Processing

Morphological operations are simple yet effective image processing techniques based on the shape features of images, typically applied to binarized image data. When performing such operations, two key parameters are required: the original binarized image to be processed and the structuring element or kernel, which determines the specific operation mode and characteristics. The two fundamental operations in morphology are erosion and dilation, and these operations can be combined in various ways to derive more complex processes such as opening, closing, and gradient operations.

After capturing depth information with the depth camera, alignment algorithms were used to extract obstacle images from the data. Subsequently, morphological processing was employed to optimize and stabilize the overall contour of the obstacles. This process helps determine the central position and overall size of the obstacles and removes redundant and irrelevant background information, thus accelerating overall computation and processing. To achieve this goal, the open-source computer vision library OpenCV was utilized. Through a series of morphological operations such as boundary detection, dilation, and erosion on the real-time RGB video stream data, relatively stable and accurate obstacle data were effectively extracted.

### 3.3 YOLO Deep Network Model

In this study, the YOLOv11 algorithm is adopted for obstacle detection in the navigation system, leveraging its remarkable real-time object detection capabilities. YOLOv11 builds upon the YOLO series by integrating a hybrid attention mechanism and adaptive feature fusion, enabling improved detection accuracy for multi-scale objects. Compared to YOLOv5, YOLOv11 reduces computational complexity by 15% while maintaining comparable precision in real-time scenarios [24-26].

YOLOv11 can efficiently handle objects of diverse scales with both high accuracy and speed. During the implementation, the COCO2014 dataset was utilized, and the model was trained for 100 epochs with an initial learning rate of 0.002, following the online YOLO reference framework. The network structure of YOLOv11, illustrated in Figure 6.

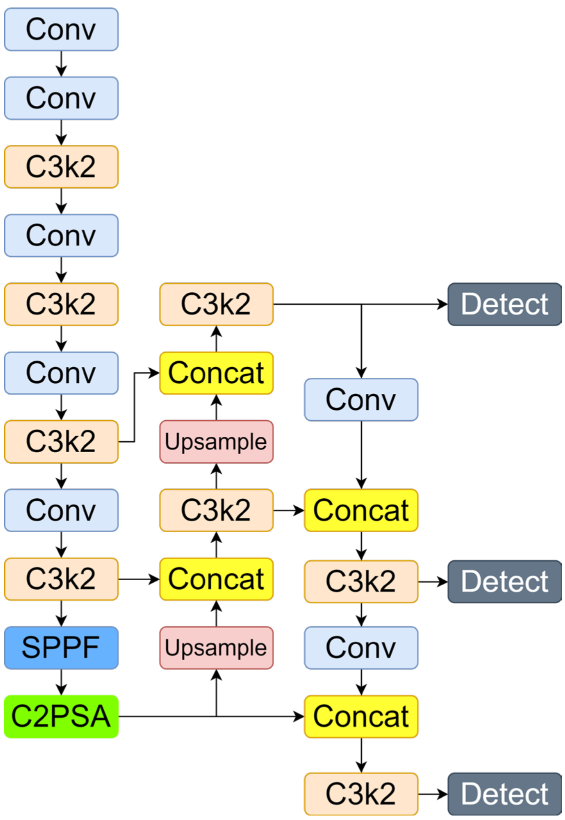


**Figure 6.** YOLOv11 network structure

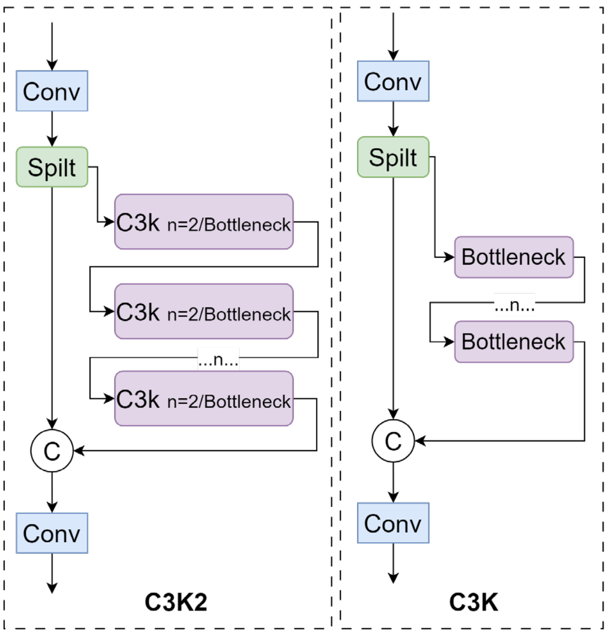


**Figure 7.** The structure of C3K2 module

The model architecture of YOLOv11 consists of three parts: the backbone network, the neck architecture, and the head network, collaboratively achieving efficient and accurate object detection. At the core of YOLOv11's backbone network is the C3k2 module, an evolution of the Cross Stage Partial (CSP) bottleneck introduced in earlier versions. The C3k2 module optimizes the information flow within the network thanks to its smaller kernel convolutions, which, while retaining the basic image feature processing capabilities, are faster and have lower computational costs compared to larger kernel convolutions. The neck architecture connects the backbone network with the rest of the system, gathering and combining information from different parts of the image.

YOLOv11's neck architecture includes components such as the C3k2 module, Spatial Pyramid Pooling-Fast (SPPF) module, and the C2PSA mechanism. Among these, the SPPF module aims to pool features from different areas of the image at various scales, enhancing the network's ability to capture objects of different sizes, especially small ones. The C2PSA mechanism embeds a multi-head attention mechanism inside the C2 framework, increasing the model's sensitivity and accuracy towards features, allowing for more precise capturing of the target's detailed characteristics. YOLOv11 employs multi-scale prediction heads to detect objects of varying sizes. By inserting two DWConv into the classification detection head, YOLOv11 significantly reduces the number of parameters and computations. The detection heads output predictions on three feature maps according to different granularity levels in the image to ensure that smaller objects are detected with finer details. YOLOv11 introduces the innovative C3k2 network architecture and optimized loss function, ensuring the accurate capture of tiny and morphologically diverse targets, making YOLOv11 an ideal choice for real-time obstacle detection tasks. The structure of C3K2 is illustrated in Figure 7.

# 4 Experimental Results

This study commenced with the experimental design and implementation of image contour extraction and obstacle recognition analysis. In the initial phase of the experiment, advanced RGB image and depth image alignment techniques were employed, successfully mapping the two-dimensional information from the depth image to the corresponding color image through precise algorithms. This ensured seamless integration and efficient fusion of the two types of image data. During the image preprocessing stage, specialized image extraction work was conducted on the obstacle parts and image morphological methods to finely process the extracted obstacle images. This leads to effectively acquiring clear and complete contour features of the obstacles. Various types of sample data were collected to further validate the accuracy and reliability of the image recognition system, and rigorous testing was conducted to assess recognition accuracy. Additionally, to investigate the detection depth capability of the image recognition system, depth data collection work was carried out in various complex environments, and thorough comparisons were made with field measurement values.

## 4.1 Image Contour Extraction

Firstly, the alignment algorithm between RGB images and depth images transformed the two-dimensional points on the depth image into the color image in three-dimensional space, achieving effective alignment and fusion of the two types of image data. Secondly, obstacle images were extracted and subjected to image morphological processing to obtain the overall contour of the obstacles, as shown in Figure 8. The depth camera (Intel RealSense D435i) provides 1280×720 depth maps with a maximum range of 10 meters. Depth accuracy is ±2% up to 4 meters, validated through structured light calibration. Depth data is fused with RGB via RANSAC alignment, enabling pixel-wise correspondence between color and depth channels. This integration allows precise 3D obstacle localization within ±5 cm accuracy in controlled environments.

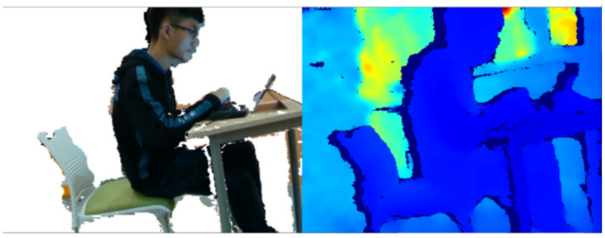

**Figure 8.** Extraction of obstacles by aligning RGB with depth images

## 4.2 Obstacle Recognition Analysis

To assess image recognition accuracy, 150 sample data points were collected and divided into three classes, each with 50 samples. The results are presented in Table 1 and Figure 9, showing an overall recognition accuracy of over 94%.


**Figure 9.** Image recognition effect of intelligent guide backpack

To understand the depth detection capability of the image recognition system, depth data were collected from the camera seven times and were compared with actual measurement values.

To understand the depth detection capability of the image recognition system, depth data were collected from the camera seven times and were compared with actual measurement values. The results indicate that image recognition detection not only supports measurement operations in complex environments but also has a significant advantage in measurement area compared to ultrasonic ranging and laser radar ranging. The detection depth completion rate is over 98%, as shown in Table 2.

**Table 1**. Image recognition data

| Sample categories | Sample size | Correct recognition | Recognition error | Recognition accuracy |
|---|---|---|---|---|
| 1 | 50 | 47 | 3 | 94% |
| 2 | 50 | 49 | 1 | 98% |
| 3 | 50 | 47 | 3 | 94% |

**Table 2**. Completion rate of depth camera measurements

| Number of tests | Actual distance (cm) | Measured depth distance (cm) | Completion rate of depth measurements (%) |
|---|---|---|---|
| 1 | 50 | 50 | 100% |
| 2 | 60 | 59 | 98% |
| 3 | 100 | 98 | 98% |
| 4 | 120 | 117 | 98% |
| 5 | 150 | 153 | 98% |
| 6 | 200 | 201 | 99% |
| 7 | 500 | 495 | 99% |

# 5 Conclusion

This study successfully designed and implemented a blind guide depth camera image recognition system based on YOLOv11. Through the alignment algorithm between RGB and depth images, the system can effectively fuse the two types of image data, enhancing the accuracy of obstacle detection. The system extracts the overall contour of obstacles using binary image morphological processing methods. Additionally, applying YOLOv11 image processing algorithms further enhances the recognition capability of objects ahead. Experimental results demonstrate the system's outstanding performance in measurement operations within complex environments, with an accuracy of up to 98%, surpassing traditional ultrasonic and laser radar ranging methods. The system also exhibits a high accuracy rate in image recognition, achieving an overall recognition accuracy of over 94% by testing 150 sample data points. These results demonstrate the effectiveness and practicality of this system in assisting visually impaired individuals in daily travel and environmental perception. Future research may integrate GPS and other positioning technologies to develop more intelligent navigation and path-planning functionalities, providing visually impaired users with more precise and personalized travel guidance.

# Acknowledgements

Foundation items: Supported by 2020 Fujian Provincial University Innovation Team (Industrialization Special Project) "Intelligent Information and Perception Technology", Fuzhou Science and Technology Planning Project (2023-ZD-009).

# References

[1] K. Xia, X. Li, H. Liu, M. Zhou, K. Zhu, IBGS: A wearable smart system to assist visually challenged, *IEEE Access*, Vol. 10, pp. 77810-77825, July, 2022.

[2] M. M. K. Peyal, Q. M. A. U. Haque, T. Tahiat, S. Habib, A. Noor, A. A. M. Azad, Inexpensive Voice Assisted Smart Eyewear for Visually Impaired Persons in Context of Bangladesh, *2021 IEEE Global Humanitarian Technology Conference (GHTC)*, Seattle, Washington, USA, 2021, pp. 43-50.

[3] U. Masud, T. Saeed, H. M. Malaikah, F. U. Islam, G. Abbas, Smart Assistive System for Visually Impaired People Obstruction Avoidance Through Object Detection and Classification, *IEEE Access*, Vol. 10, pp. 13428-13441, January, 2022.

[4] H. A. A, S. U. Rao, S. Ranganath, T. S. Ashwin, G. R. M. Reddy, A Google Glass Based Real-Time Scene Analysis for the Visually Impaired, *IEEE Access*, Vol. 9, pp. 166351-166369, December, 2021.

[5] A. Bhattacharya, V. K. Asari, Wearable Walking Aid System to Assist Visually Impaired Persons to Navigate Sidewalks, *2021 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, Washington, DC, USA, 2021, pp. 1-7.

[6] S. Chandna, A. Singhal, Towards Outdoor Navigation System for Visually Impaired People using YOLOv5, *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 2022, pp. 617-622.

[7] Y. He, T. N. Sainath, R. Prabhavalkar, I. McGraw, R. Alvarez, D. Zhao, D. Rybach, A. Kannan, Y. Wu, R. Pang, Q. Liang, B. Dhatia, Y. Shangguan, B. Li, G. Pundak, K. C. Sim, T. Bagby, S.-Y. Chang, K. Rao, A. Gruenstein, Streaming end-to-end speech recognition for mobile devices, *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 2019, pp. 6381-6385.

[8] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, *Proceedings of the IEEE/CVF*

*international conference on computer vision*, Montreal, QC, Canada, 2021, pp. 10012-10022.

[9] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jegou, J. Mairal, P. Labatut, A. Joulin, P. Bojanowski, Dinov2: Learning robust visual features without supervision, arXiv preprint arXiv:2304.07193, April, 2023. https://arxiv.org/abs/2304.07193v1

[10] D. Thomas, A. Chandran, A. K. Babu, D. P. Charly, J. J. Daniel, S. Kumar, Depth Sensing and Object Localization Using Ultrasound 2D Traversing Scanner, *2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS)*, Kanjirapally, India, 2023, pp. 1-6.

[11] S. Dávila-Montero, S. Parsnejad, Y. Gtat, A. J. Mason, Improving Object Localization Resolution using a Wearable Ultrasonic Sensor Array, *2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS)*, Springfield, Massachusetts, USA, 2020, pp. 237-240.

[12] H. Zhang, Deep neural network for object classification and optimization algorithms for 3D positioning in Ultrasonic Sensor Array, M.S. Thesis, Umeå University, Västerbotten, Sweden, 2021.

[13] X. You, D. Tian, C. Liu, X. Yu, L. Song, Vehicles positioning in tunnel: A real-time localization system using DL-TDOA technology, *Journal of Internet Technology*, Vol. 22, No. 5, pp. 967-978, September, 2021.

[14] D. Chen, Application of Improved Genetic Algorithms in Path Planning, *Journal of Internet Technology*, Vol. 25, No. 7, pp. 1091-1099, December, 2024.

[15] Y.-Z. Hsieh, Y.-Y. Fanjiang, J.-P. Hwang, Predicting the Distance of Objects in the Blind Zone of an Ultrasonic Array, *Journal of Internet Technology*, Vol. 25, No. 7, pp. 977-985, December, 2024.

[16] S. Jain, K. Singh, Signal Noise Reduction Algorithm for Object Localization and Tracking Using a 2-Dimensional Ultrasonic Sensor Array, *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*, Bengaluru, India, 2024, pp. 1-6.

[17] D. Pal, A. Banerjee, S. Warshi, R. Dutta, D. Deb, M. U. Ansari, Sonic Alert: GPS-Integrated Ultrasonic Spectacle for the Blind, *International Conference on Innovations in Data Analytics*, Kolkata, India, 2023, pp. 449-461.

[18] B. S. Lin, C. C. Lee, P. Y. Chiang, Simple smartphone-based guiding system for visually impaired people, *Sensors*, Vol. 17, No. 6, Article No. 1371, June, 2017.

[19] P. T. Kumaran, B. Shivani, V. Roshini, M. Akash, Detecting Obstacles using Image Processing and Ultrasonic Sensor, *i-Manager's Journal on Instrumentation & Control Engineering*, Vol. 9, No. 2, pp. 28-37, July-December, 2021.

[20] K. Goel, M. Nagar, M. Tiwari, A. R. Mishra, S. S. Chauhan, Enhanced Ultrasonic Cane for Visually Impaired People, *2024 2nd International Conference on Disruptive Technologies (ICDT)*, Greater Noida, India, 2024, pp. 500-505.

[21] A. Meliones, C. Filios, J. Llorente, Reliable ultrasonic obstacle recognition for outdoor blind navigation, *Technologies*, Vol. 10, No. 3, Article No. 54, June, 2022.

[22] G. Saranya, N. Krishnamoorthy, B. Vidhya, R. Thangam, A Safe and Enhanced Navigation System for the Visually Challenged, *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, Chennai, India, 2024, pp. 1-4.

[23] J. Deepa, P. M. Adeline, S. S. Madhumita, N. Pavalaselvi, Obstacle Detection and Navigation for The Visually Impaired, *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2023, pp. 905-909.

[24] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, arXiv preprint arXiv:1804.02767, April, 2018. https://arxiv.org/abs/1804.02767

[25] C. Y. Wang, A. Bochkovskiy, H. Y. M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, British Columbia, Canada, 2023, pp. 7464-7475.

[26] R. Khanam, H. Muhammad, Yolov11: An overview of the key architectural enhancements, arXiv preprint arXiv:2410.17725, October, 2024. https://arxiv.org/abs/2410.17725

## Biographies

**Cheng-Li Luo** Dean, Professor and Master Supervisor of College of Information Engineering at Yango University, and Adjunct Researcher at Monash Suzhou Institute of Science and Technology.

**Hai Xu** Lecturer of College of Information Engineering at Yango University, The main research areas: industrial IoT software development, introduction to electronic information engineering.

**Min Liu** Lecturer of College of Information Engineering at Yango University. The main research areas cover machine vision, image processing, mobile communication and object detection.

**Shu-Chuan Chu** Received Ph.D. from Flinders U in 2004. Worked at Cheng Shiu U for 9 years, then joined Flinders U in 2009. Now a research fellow at Shandong U of Science and Tech since 2019, interested in swarm intelligence, etc.