UDC 004.932.2:004.8:004.58 DOI https://doi.org/10.35546/kntu2078-4481.2025.3.2.3

O. YU. BARKOVSKA
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0001-7496-4353

O. V. NECHITAILO
Postgraduate Student at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0009-0005-7299-2875

L. D. BUHAROVA
Postgraduate Student at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0009-0003-0635-8695

EMBEDDED VOICE ASSISTANCE MODULE FOR THE VISUALLY IMPAIRED BASED ON VISUAL SCENE ANALYSIS

The article discusses a hardware module designed to assist visually impaired individuals at public transport stops. The relevance of this research is determined by the increasing number of people with visual impairments in Ukraine, caused both by global trends and by the consequences of military actions that have led to eye injuries among both military personnel and civilians. Since urban infrastructure is predominantly oriented toward visual perception, visually impaired individuals face significant difficulties in orientation, especially at public transport stops. The use of an autonomous device that does not depend on a smartphone or Internet access is of great importance for ensuring mobility and safety in urban environments.

The purpose of the study is to develop and analyze the effectiveness of a hardware-software complex based on Raspberry Pi 5, which provides automatic recognition of public transport types (bus, trolleybus, tram) and their route numbers, followed by real-time voice output of the results. The hardware part includes a Raspberry Pi 5 single-board computer with active cooling, a Logitech Web Camera Carl Zeiss Tessar 2.0/3.7 2MP connected via USB 3.0, a Logitech G Pro X external sound card for audio output through headphones, and a portable power bank to ensure autonomous operation.

Experimental studies have confirmed high detection accuracy of public transport in daytime conditions and acceptable performance on the embedded platform. At the same time, under low-visibility conditions (twilight, fog, image blurring), a decrease in classification accuracy of transport types and recognition of route numbers was observed. It has been determined that further system development may include expanding the dataset with images collected under real weather conditions, improving video preprocessing algorithms (low-light enhancement and denoising), as well as integrating with mobile applications and GPS navigation. The obtained results demonstrate the practical applicability of the developed solution and its potential for scaling to improve the inclusiveness and mobility of visually impaired individuals.

Key words: *Raspberry Pi, YOLOv8, Tesseract, OpenCV, OCR, pyttsx3, intelligent module, visually impaired, hardware implementation, public transport.*

О. Ю. БАРКОВСЬКА
кандидат технічних наук, доцент,
доцент кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0001-7496-4353

О. В. НЕЧІТАЙЛО
магістрант кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0009-0005-7299-2875

Л. Д. БУХАРОВА
магістрант кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0009-0003-0635-8695

АПАРАТНИЙ МОДУЛЬ ГОЛОСОВОГО СУПРОВОДУ ДЛЯ ОСІБ З ПОРУШЕННЯМ ЗОРУ НА ОСНОВІ ВІЗУАЛЬНОГО АНАЛІЗУ ПАРАМЕТРІВ ОТОЧУЮЧОГО СЕРЕДОВИЩА

У статті розглянуто апаратний модуль супроводу осіб із вадами зору на зупинках громадського транспорту. Актуальність теми зумовлена зростанням кількості громадян з порушеннями зору в Україні, що обумовлено як загальносвітовими тенденціями, так і наслідками воєнних дій, які призводять до травмування органів зору у військових та цивільних. Міська інфраструктура здебільшого зорієнтована на візуальне сприйняття, тому слабозорі особи стикаються зі значними труднощами при орієнтації, особливо на зупинках громадського транспорту. Використання автономного пристрою, який не залежить від смартфону чи мережі Інтернет, має важливе значення для забезпечення мобільності та безпеки таких осіб у міському середовищі.

Метою дослідження є створення та аналіз ефективності програмно-апаратного комплексу на базі Raspberry Pi 5, що забезпечує автоматичне розпізнавання типу громадського транспорту (автобус, тролейбус, трамвай) та його маршрутного номера з подальшим голосовим озвученням результату в режимі реального часу. Апаратна частина включає одноплатний комп'ютер Raspberry Pi 5 з системою активного охолодження, камеру Logitech Web Camera Carl Zeiss Tessar 2.0/3.7 2MP, підключену через USB 3.0, зовнішню звукову карту Logitech G Pro X для відтворення аудіо через навушники та портативний акумулятор, що забезпечує автономність роботи.

Експериментальні дослідження підтвердили високу точність детектування транспортних засобів у денних умовах та прийнятну швидкодію системи на вбудованій платформі. Водночас в умовах низької видимості (сутінки, туман, ефект замилення) спостерігається зниження точності класифікації типу транспорту та визначення маршрутного номера. Визначено, що подальший розвиток системи може відбуватися шляхом розширення датасету зображеннями, отриманими у реальних погодних умовах, вдосконалення алгоритмів попередньої обробки відео (застосування методів low-light enhancement та denoising), а також інтеграції з мобільними додатками та GPS-навігацією. Отримані результати демонструють практичну придатність розробленого рішення та його потенціал для масштабування з метою підвищення рівня інклюзії та мобільності осіб із порушеннями зору.

Ключові слова: *Raspberry Pi, YOLOv8, Tesseract, OpenCV, OCR, pyttsx3, інтелектуальний модуль, слабозорі, апаратна реалізація, громадський транспорт.*

Formulation of the problem

The relevance of the problem of rehabilitation and social integration of people with visual impairments in Ukraine is determined both by global trends in the growing number of people with vision disorders and by the direct consequences of military actions, which lead to an increase in eye injuries among both military personnel and civilians [1]. According to official data, the number of people with severe visual impairments in Ukraine is on the rise, creating an urgent need for the development of an inclusive environment. However, modern urban infrastructure is primarily oriented toward visual perception of information: road signs, traffic lights, billboards, and navigation indicators. Because of this, people with low vision face significant difficulties in spatial orientation, especially at public transport stops, where it is important not only to identify the vehicle itself but also to obtain information about its route number and direction of movement.

To address these problems, a few technical solutions already exist worldwide, including mobile navigation applications, sensory systems for spatial orientation, as well as traditional means of assistance such as guide dogs or specialized "smart" canes. Mobile applications (for example, BlindSquare) use GPS navigation and voice guidance; however, their operation depends on stable Internet access and a functioning smartphone. Sensory systems based on ultrasonic technologies (WeWALK Smart Cane 2, Sonic Pathfinder) allow the detection of obstacles but cannot provide users with semantic information, such as the type of vehicle or route number. The most advanced camera-oriented solutions (e.g., OrCam MyEye 3 Pro) demonstrate high accuracy in object and text recognition but remain extremely expensive and inaccessible to the wider public.

To generalize and systematize the available technologies in the field of assistance for the visually impaired, Table 1 presents a comparison of modern technical solutions. Special attention is paid to hardware-based approaches, as they are the most promising for the development of autonomous intelligent orientation systems.

This paper proposes a fundamentally different approach-an autonomous software-hardware module built on the single-board computer Raspberry Pi 5. The system integrates a camera for video data acquisition, an external sound card with headphones for voice output, and portable power supply to ensure full autonomy. By employing modern deep learning algorithms (YOLOv8 for vehicle detection and Tesseract OCR for route number recognition), the module can provide real-time voice output through a local speech synthesizer (pyttsx3). A key advantage of this solution is that it does not require an Internet connection or smartphone, making it independent and suitable for everyday use in Ukrainian cities.

Analysis of the latest research and publications

The issue of applying artificial intelligence and computer vision technologies to support people with visual impairments has been attracting increasing attention among researchers in recent years. Contemporary studies emphasize the need to develop systems capable of providing spatial orientation, object recognition, and text information retrieval in urban environments in real time.

Table 1

Aids and support solutions for the blind: advantages and disadvantages of each one

№	Solution	Advantages	Disadvantages
1	WeWALK Smart Cane 2	Obstacle detection from above; Voice AI assistant; Navigation; White cane and accompanying functions in one device	For full functionality, must have access to the network
2	Sonic Pathfinder 	Obstacle detection from above; A boost to future technology development	Load on the head; Secondary
3	BlindSquare	Does not require additional devices; Navigation; Voice guidance	Requires network access; Does not allow you to use other smartphone applications
4	OrCam MyEye 3 Pro 	Works without internet access; Recognizes a large number of objects; Voice AI assistant	Price $4000; Doesn't help with navigation

For example, O. Yu. Barkovska and V. Seredniy proposed a universal model of an intelligent assistance system that enables automatic detection of static and dynamic obstacles, prediction of their trajectories, and the use of a voice interface for user interaction [2]. Such an approach makes it possible to integrate various sensors and modules into a single adaptive system.

In a study by an international group of authors, the MR.NAVI system was introduced, combining computer vision and natural language processing to explain situations, avoid collisions, and assist with public transport navigation. The authors emphasize that integrating multimodal interaction methods significantly increases the effectiveness of orientation for people with low vision [3].

Another approach was demonstrated in the work of Shah et al., where a Smart Real-Time Object Detection system was developed based on the YOLOv8 algorithm for object recognition and OCR technologies for text reading. Importantly, the results are immediately voiced to the user, making the solution convenient for practical use [4].

A significant contribution was made by the VISA system described by Mahmud and colleagues. It integrates augmented reality methods, deep neural networks for object recognition, scene depth estimation, and a voice interface, ensuring natural interaction with the system [5].

It is also worth mentioning a systematic review of modern technologies for visually impaired individuals conducted by Rafiq and co-authors. The review analyzed over 50 scientific works, including solutions based on mobile applications, tactile devices, and neural network navigation algorithms. The authors highlight the necessity of localizing such systems and adapting them to specific environmental conditions [6].

Thus, the literature review demonstrates that the application of artificial intelligence and computer vision for supporting visually impaired people is actively developing; however, in the Ukrainian context, this problem still requires further scientific and applied research.

Formulation of the purpose of the research

Given the relevance of the problem of the growing number of people with visual impairments in Ukraine, particularly as a result of military actions, and considering the limitations of existing solutions in providing orientation at public transport stops, this study aims to develop a hardware- and software-oriented module to assist visually impaired individuals. The main task is to ensure the compactness, autonomy, and ease of use of a device that integrates a camera for video data collection, an audio interface for voice output, and portable power supply, enabling its operation in urban environments without dependence on external infrastructure.

It should be emphasized that the implementation of such a hardware solution entails several constraints and challenges:

– ensuring water resistance of the case and durability under external conditions (rain, snow, low and high temperatures);

– limited autonomous operation time, which depends on the capacity and characteristics of the external battery (power bank);

– the need to select peripheral devices (USB camera, sound card, headphones) that are simultaneously compatible with Raspberry Pi 5 and do not create excessive power consumption;

– technical complexity of integrating cooling to ensure stable processor performance during long-term operation.

To ensure the operability of the hardware system, the following components are required:
- single-board computer Raspberry Pi 5 with a cooling system;
- USB camera (e.g., Logitech Web Camera Carl Zeiss Tessar) for video stream capture;
- external USB sound card and headphones for voice playback;
- portable power bank with PD (Power Delivery) support and an output power of at least 25–30 W;
- lightweight case or mount that allows the system to be carried on the body (waist, chest) while simultaneously protecting electronic components.

The implementation process requires the following tasks to be performed:
- selection and integration of peripherals that meet mobility and low power consumption requirements;
- development of guidelines for module placement on the user;
- configuration of automatic software launch after power-on to simplify usage;
- testing the module on different types of public transport;
- analysis of testing results when capturing frames of public transport from various distances.

Future development of the hardware component may include the integration of a built-in battery instead of an external power bank, the design of a compact case with enhanced water and dust resistance (IP54–IP67), as well as the creation of a specialized single-board system with an integrated graphics processor, which would optimize system performance and reduce energy consumption.

Research Results and Discussions

At the initial stage of developing the hardware module, a general algorithm of the system's operation was designed. Considering the technology of assisting visually impaired individuals in orientation at public transport stops, it should be taken into account that such a tool must include a hardware component. Therefore, the scheme presented below incorporates both software and hardware components (Figure 1).

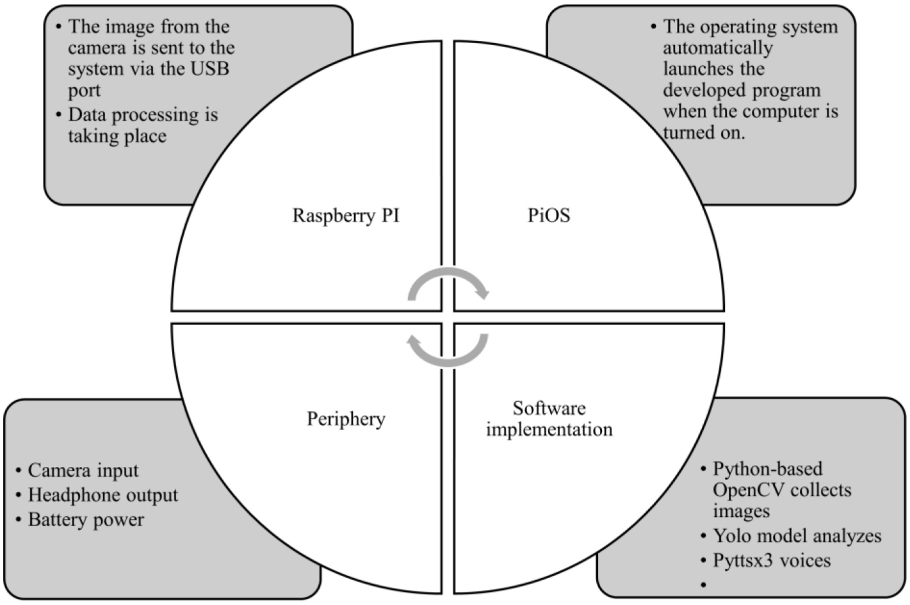


Fig. 1. Abstract of detecting the work cycle of a software-hardware solution

To formalize the system's operational logic, an IDEF0 diagram was used, which clearly illustrates the process structure and the interaction between its components. The top level of the model demonstrates the main function: the detection and voice announcement of approaching public transport, which is implemented through the sequential operation of several subsystems.

The system input includes the video stream from the camera as well as environmental factors (lighting, weather conditions). These data are processed by the computer vision module (YOLOv8), which performs transport object detection. Next, the processing block extracts from the frame the area where the route number is expected and sends it to the Optical Character Recognition (OCR, Tesseract) module. If the route number is successfully recognized, the system generates a text message and transmits it to the speech module (pyttsx3), which produces a voice signal. In case of recognition failure, only the type of transport (e.g., "bus") is announced.

An additional element of the architecture is the environmental filtering module, which applies pre-processing to frames (for example, contrast enhancement in low-light conditions), thus increasing the accuracy of subsequent processing stages.

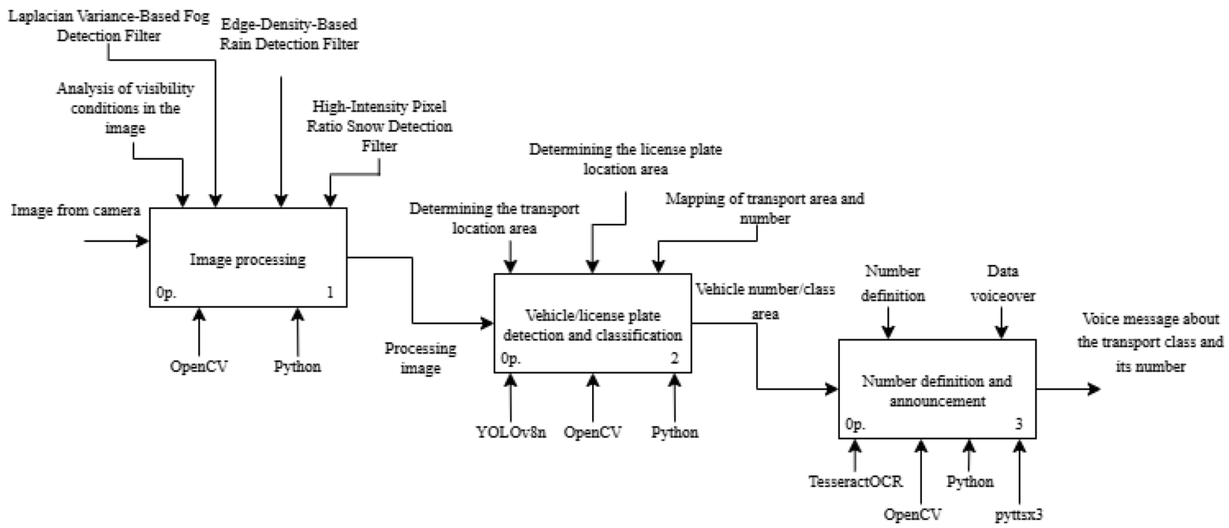


Fig. 2. IDEF0 diagram of the general architecture of the software and hardware complex

For the implementation of the hardware part of the system for assisting visually impaired individuals at public transport stops, the following hardware stack was used:

– Raspberry Pi 5 with a cooling system installed (Figure 3);
– Logitech Web Camera USB camera (Figure 4);
– Logitech G Pro X sound card with headphones (Figure 5);
– Baseus PD 20W 10000mAh power bank (Figure 6).

The central computing module of the system is the Raspberry Pi 5, equipped with a quad-core ARM Cortex-A76 processor, 8 GB of RAM, and a passive-active cooling system with a fan. The device is preinstalled with the Raspberry

Fig. 3. Raspberry Pi 5 with cooling system installed

Fig. 4. USB-Camera Logitech Web Camera

Fig. 5. Logitech G Pro X sound card with headphone

Fig. 6. Power source

Pi OS Lite operating system, while the software is stored on an internal microSD card. For stable operation under high loads, a power supply with at least 25 W output is required (Figure 3).

Video capture is performed using an external Logitech Web Camera with a Carl Zeiss Tessar 2.0/3.7 lens, providing 2 MP resolution and maintaining high image clarity even under low lighting conditions. The camera is equipped with a quality autofocus lens and UVC (USB Video Class) certification, ensuring full compatibility with Linux systems and stable operation on the Raspberry Pi platform. The device is mounted on the user's chest or upper clothing with a clip or strap, oriented in the user's line of sight (Figure 4).

Audio output is implemented through an external Logitech G Pro X USB sound card, to which standard wired headphones are connected via a 3.5 mm jack. This configuration provides stable sound quality and minimal delay in speech synthesis. It is recommended to use a single earphone that freely fits into the user's ear (Figure 5).

The system is powered by an external power supply with Power Delivery support. Capacity – 10000 mAh, output – 5V/3A. This allows uninterrupted operation of the device for 4–6 hours. In the prototype, a Baseus PD 20W 10,000 mAh power bank is used (Figure 6).

The complete system forms a wearable device, shown in Figure 7.

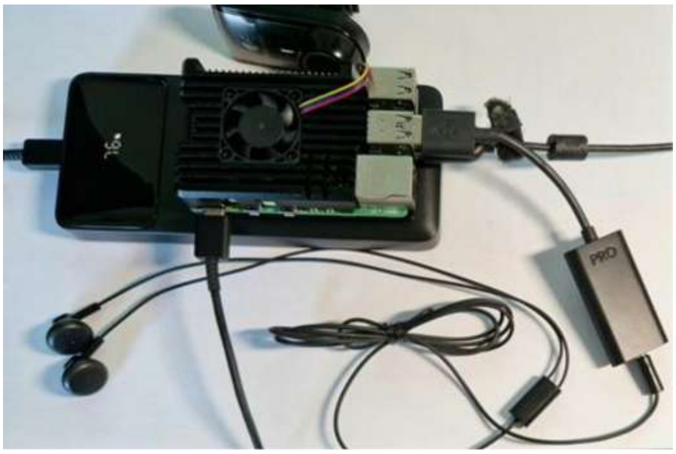


Fig. 7. General view of the module

The installation sequence of the hardware module is as follows:

– Raspberry Pi is attached to the user's belt;
– the camera is mounted on the chest;
– the earphone is placed in the ear;
– the power bank is mounted on the back of the belt.

The wiring is routed along the clothing so as not to hinder movement or obstruct the view (Figure 8).

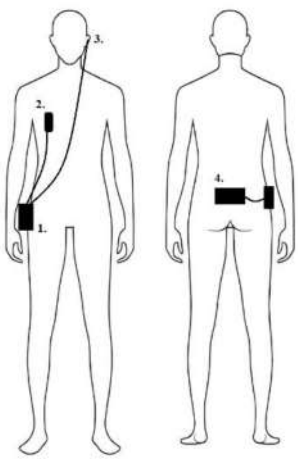


Fig. 8. Location of elements on the user

Thus, since the components are sufficiently compact and lightweight, carrying the device is simple and convenient for the user.

To verify the functionality of the developed hardware-software system, testing was conducted under real-world operating conditions. The main goal of the testing was to determine system stability, accuracy of vehicle detection, recognition of route numbers, and quality of voice output under various environmental conditions.

The results were evaluated based on the content of voice messages generated by the module. The formation of voice messages in the developed module takes place in several sequential stages (Figure 9). At the software level, after completing the transport detection algorithm (YOLOv8) and route number recognition (Tesseract OCR), a text string containing the analysis result is generated, for example: “Bus, route 25.” This text is passed to the speech module implemented with the pyttsx3 library.

Fig. 9. Voice message generation sequence

The pyttsx3 library is a Python wrapper for the standard speech synthesis engine that works locally without the need for Internet access. In the case of PiOS (Raspberry Pi OS), it uses low-level synthesis libraries such as espeak or espeak-ng as the speech generation engine, as well as pyaudio or ALSA (Advanced Linux Sound Architecture) for sending the audio signal to the sound device. Thus, pyttsx3 acts as an intermediate layer that receives text data, converts them into a command for the synthesizer (espeak/espeak-ng), and then transmits the resulting audio stream to the playback system.

At the hardware level, the Raspberry Pi 5 sends the digital audio signal via the ALSA driver to the external Logitech G Pro X USB sound card, which converts the signal from digital form (PCM) into analog. This signal is then transmitted via a standard 3.5 mm audio jack to the user’s headphones.

The messages generated by the software module have two types:

– announcement of the transport class only;
– announcement of both the transport class and its route number.

Thus, as a result of the module’s operation, six types of messages can be obtained, as presented in Table 2.

Table 2

Voice message options

Only transport detected	Transport and route number detected
“Trolleybus is coming”	“Trolleybus (№) is coming”
“Bus is coming”	“Bus (№) is coming”
“Tramway is coming”	“Tramway (№) is coming”

Testing was carried out in the city of Kharkiv over the course of one day, at public transport stops in normal summer weather. Several locations were selected where the movement of buses, trolleybuses, and trams was observed during both daytime and evening hours.

During real-world testing of the system, 10 experiments were conducted under different lighting conditions. In daytime (100–10,000 l*x*), the system demonstrated high efficiency: vehicle detection accuracy reached 100 %, transport type classification accuracy – 90 %, and correct OCR recognition of route numbers – 70 %.

It should be noted that the distance at which the transport route number was recognized directly influenced the output message generated by the module. On average, at a distance of 80 meters from the vehicle, the module announced its class. At a distance of 50 meters, the route number was recognized (Table 3). As a result, full announcements of both transport type and route number were obtained in 7 out of 10 cases.

In twilight conditions (50–100 l*x*), the system performance decreased: vehicle detection accuracy was 80 %, type classification accuracy – 70 %, and OCR recognition succeeded in 50 % of cases. Consequently, the number of successful full announcements dropped to 5 out of 10.

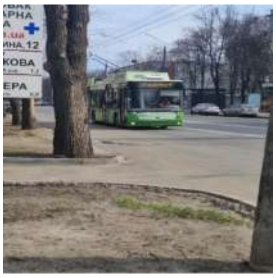
Conclusions

The developed hardware module confirmed the feasibility of using a dedicated hardware-software solution to ensure autonomy and convenience of orientation for visually impaired individuals in urban environments. The module demonstrated stable operation under daylight conditions, is sufficiently compact and portable, and allows the user to easily carry and apply it while waiting at public transport stops. Owing to its independence from Internet connectivity, the device can operate effectively even outside urban areas or in the absence of network coverage, thus broadening its scope of application.

At the same time, the test results revealed several aspects requiring improvement. First of all, the hardware case requires modernization to enhance its durability and resistance to external factors. Moisture resistance, dust protection,

Table 3

Notifications depending on the distance to transport

Type of transport	Distance, m	Message
	≈80	“Trolleybus is coming”
	≈40	“Trolleybus three is coming”
	≈60	“Tramway is coming”
	≈30	“Tramway twenty is coming”
	≈70	“Bus is coming”
	≈40	“Bus is thirteen is coming”

and tolerance to temperature fluctuations and mechanical impacts must be considered, which is especially important for outdoor use. Another development direction involves optimizing the case design to ensure ergonomics and comfort during prolonged wear or use in various weather conditions.

Another important task for further enhancement is the modernization of the power supply system. While the use of external power banks provides autonomy, it has certain disadvantages, including increased size and reduced portability. A more appropriate solution is the integration of a built-in rechargeable battery with energy-saving features and fast-charging capability. This would significantly increase the device's autonomy, reduce its weight, and ensure compactness.

A promising direction for hardware development is also the optimization of the component base with the goal of reducing power consumption and improving the overall efficiency of the module. The use of energy-efficient microcontrollers and additional hardware accelerators for image processing would significantly increase the device's performance while maintaining autonomy.

Bibliography

1. Нечітайло О. В., Гаврашенко А. О. Інтелектуальний модуль навігації для реабілітації осіб з вадами зору на зупинках громадського транспорту. Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління : тези доп. 15-ї міжнар. наук.-техн. конф. (24–25 квіт. 2025 р., Баку – Харків – Жиліна). Т. 3. Харків ; Баку ; Жиліна, 2025. С. 123.

2. Barkovska O., Serdechnyi V. Development of a universal model of an intellectual system to help people with visual impairments. Information Technology and Security. 2024. Vol. 12, No. 2. P. 47–56.

3. Kumar A., Ahmed S., Lee J. MR.NAVI: Mixed Reality Navigation Assistance for Visually Impaired Using Computer Vision and Natural Language Processing. 2025.

4. Shah P., Patel K., Reddy S. Smart Real-Time Object Detection for Blind Individuals Using YOLOv8 and OCR. Journal of Emerging Technologies. 2024. Vol. 18, No. 3. P. 115–123.

5. Mahmud R., Zhang T., Li Y. VISA: A Vision-Based Spatial Awareness System for Indoor Navigation of Visually Impaired. Journal of Imaging. 2024. Vol. 11, No. 1. P. 9.

6. Rafiq M., Hussain A., Khan N. Navigation Technologies for Visually Impaired: A Systematic Review (2019–2024). Sensors. 2024. Vol. 25, No. 7. P. 2213.

References

1. Nechitailo, O. V., & Havrashenko, A. O. (2025). Intellectual navigation module for the rehabilitation of visually impaired people at public transport stops. *Modern directions of development of information and communication technologies and management tools: Proceedings of the 15th International Scientific and Technical Conference (Vol. 3, p. 123). Baku – Kharkiv – Žilina*.

2. Barkovska, O., & Serdechnyi, V. (2024). Development of a universal model of an intellectual system to help people with visual impairments. *Information Technology and Security, 12*(2), 47–56.

3. Kumar, A., Ahmed, S., & Lee, J. (2025). *MR.NAVI: Mixed Reality Navigation Assistance for Visually Impaired Using Computer Vision and Natural Language Processing*.

4. Shah, P., Patel, K., & Reddy, S. (2024). Smart real-time object detection for blind individuals using YOLOv8 and OCR. *Journal of Emerging Technologies, 18*(3), 115–123.

5. Mahmud, R., Zhang, T., & Li, Y. (2024). VISA: A vision-based spatial awareness system for indoor navigation of visually impaired. *Journal of Imaging, 11*(1), 9.

6. Rafiq, M., Hussain, A., & Khan, N. (2024). Navigation technologies for visually impaired: A systematic review (2019–2024). *Sensors, 25*(7), 2213.

Дата першого надходження рукопису до видання: 30.09.2025
Дата прийнятого до друку рукопису після рецензування: 27.10.2025
Дата публікації: 28.11.2025