Journal of
Computer & Electrical and Electronics Engineering Sciences
Original Article
DOI: 10.51271/JCEEES-0020

Vision assistant for visually impaired individuals

İrem Yıldırım, İlyas Demir, İsa Peker, Mehmet Emin Yılmaz, Ozan Güneyli, Ezgi Şevval Aktaş, Elif Özer, Emrah Kantaroğlu

Department of Electronics and Automation, Faculty of Engineering and Natural Sciences, Kırıkkale University, Kırıkkale, Turkiye

Cite this article: Yıldırım, İ., Demir, İ., Peker, İ., Yılmaz, M.E., Güneyli, O., Aktaş, E.Ş., Özer, E., & Kantaroğlu, E. (2024) Vision assistant for visually impaired individuals. *J Comp Electr Electron Eng Sci*, 2(2), 62-66.

Corresponding Author: İrem Yıldırım, iremyldrm5706@gmail.com

Received: 25/09/2024 ◆ **Accepted**: 18/10/2024 ◆ **Published**: 29/10/2024

ABSTRACT

Visually impaired individuals face significant challenges in many areas of their daily lives, such as accessing environmental information, social interaction, and safety. This article aims to enhance the independence of visually impaired individuals and facilitate their daily lives through an artificial intelligence-supported image and sound recognition system. The developed prototype consists of glasses equipped with a Raspberry Pi 4 microprocessor, integrated headphones, touch controls, and a robust battery management system. Its electronic schematic includes the integration of the camera and the microprocessor. The artificial intelligence algorithms developed successfully perform tasks such as object recognition, text reading, and face recognition. A visual question answering (VQA) model was utilized in the glasses to enable visually impaired individuals to communicate more effectively with their surroundings. Following a demonstration conducted with visually impaired individuals, the usability and effectiveness of the prototype were evaluated, and improvements were made based on user feedback. The accuracy of the algorithms was tested, resulting in a prototype with a 78% accuracy score, which was close to the 82% accuracy score of GPT-4. Additionally, tests conducted with 20 users indicated an 85% user satisfaction rate.

Keywords: Glasses, artificial intelligence, image processing, mini computer, visually impaired individuals, 3D CAD

INTRODUCTION

According to the World Health Organization (WHO), approximately 45 million people globally are completely blind, and 135 million have partial vision loss (WHO, 2003). As of 2022, the Turkish Ministry of Family and Social Services reported that there are 1,039,000 visually impaired individuals in Turkey, with 478,000 men and 561,000 women (Ministry of Family and Social Services, 2022). Common problems for both fully and partially visually impaired individuals include independence, education, transportation, social life, and societal pressure.

Visually impaired individuals often rely on healthy individuals until they can handle the challenges they face throughout life. This reliance creates problems for both the visually impaired person and the helper. Moreover, a helper may not always be available. Visually impaired individuals use the Braille alphabet in the field of education. However, the limited number of resources available in Braille and the lack of educators who can teach using this alphabet make accessing information a challenge.

Transportation is another problem for visually impaired people. The lack of sufficient tactile guide paths in public spaces, the absence of guiding structures within buildings, and obstacles placed on sidewalks by unaware individuals prevent visually impaired people from moving safely. Additionally, societal exclusion and lack of participation in social events negatively impact their psychological health, causing some to withdraw from social life.

As a solution to these problems, we designed an AI-powered vision assistant in the form of glasses. The user can take photos with the glasses using a button on the side, and the camera captures visual information quickly and in high resolution. The camera system provides a wide field of view and allows detailed image processing. The photo is analyzed by AI using an image processing algorithm, which includes both image recognition and scene understanding. The system recognizes objects, reads text, and identifies faces and environments, then communicates this information to the user via headphones. The speed, tone, and language of the descriptive sounds can be adjusted according to user preference. The glasses feature a user-friendly interface with touch and voice control systems, ensuring accessibility and comfort during prolonged use.

The system is designed with privacy in mind, using encryption to protect user data and providing offline data processing options. This ensures that user data is processed securely without reliance on cloud-based systems, maintaining privacy.

LITERATURE REVIEW

Several studies have been conducted on smart glasses for visually impaired individuals. These studies examine the benefits of such technologies in enhancing daily life and promoting independence. In the literature, four prominent studies highlight how different approaches and technologies assist visually impaired people. These studies address the functions of smart glasses, such as facial recognition, object detection, and navigation.

Bastola et al, focused on the development of multifunctional glasses for visually impaired individuals, allowing users to receive auditory and tactile notifications about objects, people, and environmental obstacles. The glasses feature text recognition technology for reading signs and written materials, and navigation tools to help users find their way in unfamiliar environments. Despite the advanced features, high cost and battery life issues are identified as disadvantages (Bastola et al., 2023).

Nazim et al, developed smart glasses using computer vision technologies. Their glasses use deep learning methods for object recognition, facial recognition, and Optical Character Recognition (OCR). Users can activate the glasses with simple commands and receive detailed descriptions of their surroundings. The glasses offer high accuracy but suffer from high power consumption and environmental sensitivity (Nazim & Firdous, 2023).

Mustafa et al, integrated various models into smart glasses designed for visually impaired individuals. These glasses provide features such as facial recognition, object detection, and navigation, powered by Raspberry Pi 4. The glasses offer low-cost solutions and are equipped with vibration alerts for obstacle avoidance. However, technical expertise is required for maintenance, and vibration alerts may discomfort the user (Mustafa, Omer, & Mohammed, 2023).

Waisberg et al, evaluated the potential benefits of second-generation smart glasses developed by Meta in collaboration with Ray-Ban for visually impaired individuals. These glasses feature a 12-megapixel wide-angle camera and advanced audio recording, aimed at improving the independence of visually impaired individuals in daily life. Although the glasses offer high accuracy and advanced technology, high cost and battery life issues are identified as challenges (Waisberg et al., 2023).

Visual data processing and machine learning play a significant role in various fields, such as medical diagnosis processes and interactive systems developed for visually impaired individuals. In particular, deep learning-based models used in medical imaging provide critical data for the accurate detection and diagnosis of diseases.

The processing and interpretation of visual data are applied in various fields, similar to their use in tuberculosis diagnosis. For example, deep learning-based models like U-NET provide effective segmentation in medical imaging data, assisting experts in their decision-making processes. Similarly, VQA-based models can offer auditory feedback for visually impaired individuals, allowing them to have a more interactive experience in their environment (Türk, 2024).

The importance of machine learning algorithms in medical diagnosis is increasingly growing, and their performance significantly improves when combined with optimization methods. In a study conducted by Türk, the accuracy of machine learning models in classifying heart disease was increased to 99.08% using the dominant feature detection (DFD) method. Likewise, VQA-based models can help visually impaired individuals perceive and identify objects in their surroundings, enabling them to gain more independence in their daily lives (Türk, 2024).

This article presents a user-friendly design, low cost, long battery life, durability against environmental factors, Turkish language support, a modular structure, and easy integration, distinguishing it from other glasses. Instead of complex technological structures, we accelerated users' adaptation process with an intuitive and easy-to-use interface. By utilizing lower-cost components, we increased accessibility and facilitated the access of visually impaired individuals to these technologies. We extended battery life with advanced battery management systems and energy-efficient components. We developed systems that are minimally affected by environmental factors such as light, noise, and weather conditions. We provided a modular structure that allows for easy replacement and upgrading of system components.

METHODOLOGY

This smart glasses design, developed based on the needs of visually impaired individuals, stands out with its user-friendly and ergonomic structure. The prototype includes hardware such as a camera system, a microprocessor (Raspberry Pi 4), integrated headphones, and touch control sensors. Figure 1 shows the electronic components on the glasses.

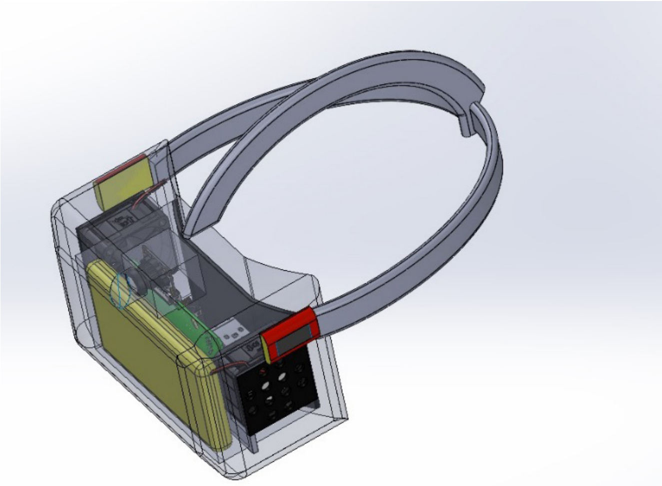

Figure 1. The electronic hardware of the glasses

To provide clear and high-quality images, the optical image stabilization camera integrated into the upper center of the glasses can capture sharp images even during movement, thanks to its dual-axis stabilization technology. This camera, with a resolution of 5 MP, has a 70° field of view, allowing it to cover a wide area.

The integrated Raspberry Pi 4 in the glasses serves as a mini-computer with its quad-core ARM Cortex-A72 processor. This processor can run image recognition and description software, providing users with auditory feedback. Additionally, it has the capability to respond to user questions within the images. The Raspberry Pi 4 is equipped with modern features such as 8 GB RAM, a MicroSD card slot, USB-C power input (5V, 3A), Wi-Fi, Bluetooth, and USB ports.

The integrated headphones in the glasses allow for the auditory transmission of images processed by artificial intelligence. With a frequency range of 20,000 Hz, sounds can be clearly heard even in the noisiest environments. The headphones come with a 1.2-meter long cable and a 3.5 mm jack input, providing a widely used connection standard worldwide.

The device's power requirement is met by a 7.4 V 2S 2200 mAh 35C LiPo battery. To prevent overheating of this battery, which supplies the necessary power to all electronic systems, cooling is provided by two mini fans integrated into the ventilation holes located at the back of the glasses. The fans, measuring 40x40x20 mm, operate at 5V and offer effective cooling performance at a noise level of 30.5 dBA. This ensures the safe use of the battery.

The camera, headphones, and touch buttons are connected to the Raspberry Pi 4. The energy management of the glasses is handled by a regulator board, which reduces the voltage supplied by the battery to appropriate levels, ensuring stable operation of the electronic components. The regulator board operates at a voltage range of 1.25V to 30V.

Under the cover located on the left side, there is a processor that runs the software for recognizing objects and providing auditory descriptions. This processor processes images from the 2K resolution camera positioned in the center of the glasses and presents information to the user. Users can take photos using a button located behind the processor, and these images are audibly described by artificial intelligence. The described images are transmitted to the Bluetooth-connected headphones located at the back left of the glasses.

On the right side of the glasses, there is a battery that supplies power to all hardware. To prevent the battery from overheating, the cooling fan activates, effectively distributing heat through the ventilation holes on the glasses.

The prototypes are designed as innovative smart glasses aimed at facilitating the daily lives of visually impaired individuals. They combine high performance and ergonomics with modern hardware components and a user-friendly interface. The glasses are designed using a 3D CAD program. The prototypes of the glasses have been printed using PLA material with a 3D printer. Figures 2 and 3 showcase two different glasses designs for completely visually impaired individuals, while Figures 4 and 5 display designs for partially visually impaired individuals.

This article utilizes a visual question answering (VQA) based model that enables visually impaired individuals to communicate more interactively with their surroundings.

The operational principles and functionality of the VQA-based model have been described. The system is activated by pressing a button by the user and provides instant responses

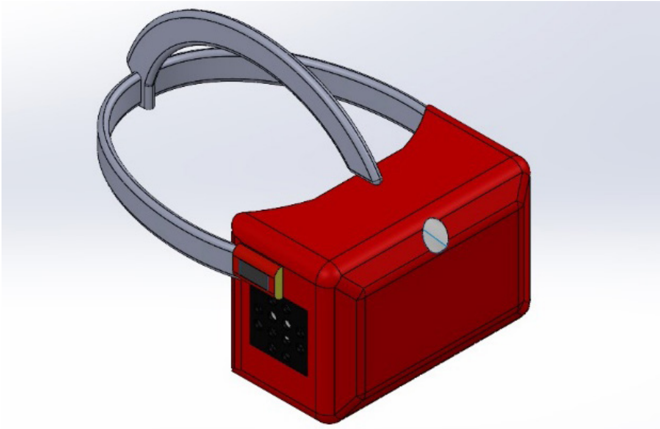

Figure 2. 3D CAD drawing of the glasses designed for completely visually impaired individuals

Figure 3. The prototype of the glasses designed for completely visually impaired individuals

Figure 4. 3D CAD drawing of the glasses designed for partially visually impaired individuals

Figure 5. The prototype of the glasses designed for partially visually impaired individuals

about environmental visuals through the individual's spoken questions. The working principles of the model are illustrated in Figure 6, and its functionality is explained step by step in the flowchart provided in Figure 7.

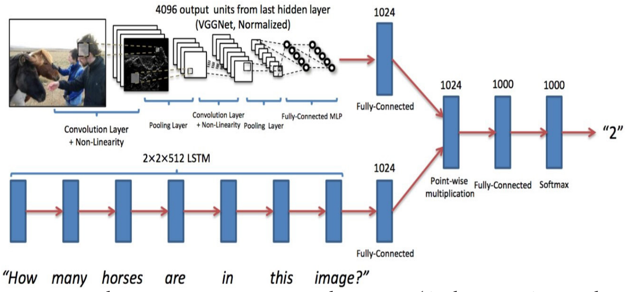

Figure 6. Visual question answering architecture (Aishwarya Agrawal et al., 2016)

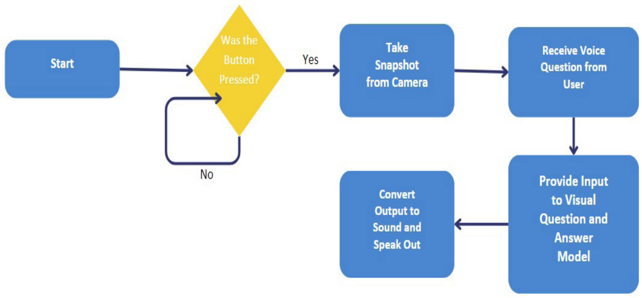

Figure 7. VQA-based visual question-answering model

This image is used as an example showing how the visual question answering (VQA) process works.

The VQA model answers the user's image-based questions by combining both visual and linguistic inputs. The key components of this process are as follows:

Image processing [Convolutional neural network (CNN)]: When an image is provided to the VQA model, it passes through a series of convolutional layers that extract meaningful features from the image. These layers analyze important details such as the edges, textures, and structures of the image. As a result, a normalized feature vector is obtained to be transferred to the fully connected layers.

Natural language processing [Recurrent neural network/ long short-term memory (RNN/LSTM)]: Simultaneously, the question posed by the user is transferred to the model's language processing section, namely the LSTM layers. LSTM processes the words of the question sequentially and learns the meaning of the question, thus extracting linguistic features.

Feature fusion: The features extracted from the image are combined with the linguistic features derived from the question. This process is typically performed using the dot product method. The fusion enables the meaningful integration of important information from both the image and the question.

Answer generation: The combined features are sent to a classifier (fully connected layer), where a probability distribution is generated among possible answers through the softmax function. The model selects the answer with the highest probability, and this answer is presented to the user.

This article utilizes a visual question answering (VQA) based model that enables visually impaired individuals to communicate with their surroundings in a more interactive way. The system becomes active when the user presses a button and provides instant responses about environmental visuals through the individual's spoken questions. The operation of the flowchart provided in Figure 6 is explained step by step below:

Initial state: The system is in an initial state and remains in passive mode until an input is received from the user.

Has the button been pressed?: The system checks whether the user has pressed the button.

- **Yes:** If the button has been pressed, the system moves to the next step.
- **No:** If the button has not been pressed, the system continues to remain in passive mode.

Capture instant image from the camera: After the user presses the button, the system captures an image of the current environment through the connected camera. This step ensures that the visual required to answer the user's question is obtained.

Receive voice question from the user: Once the visual is obtained, the system captures the user's question in a spoken format. This question can relate to any object, event, or situation in the visual.

Send visual and question to the VQA model: The system sends the voice question received from the user and the instant image obtained from the camera as input to the VQA model. The VQA model generates the correct answer based on the given question and image.

Convert output to speech and vocalize: The written response from the VQA model is converted into an audible form for the visually impaired individual and conveyed to the user through a speaker. This step allows the user to receive a direct and quick response to their question.

This system helps visually impaired individuals perceive their surroundings more independently and comfortably. By combining the instant images obtained from the camera with the spoken questions from the user, the system provides instant feedback. This system, which processes visual and auditory inputs to generate spoken responses, demonstrates the significant benefits of visual question answering (VQA) technology for visually impaired individuals.

The system employs a unique image processing algorithm that offers both image recognition and scene understanding capabilities. This advanced algorithm provides high accuracy and speed in tasks such as object recognition, text reading, facial recognition, and environmental description. With more accurate descriptions and fast response times, the user experience is significantly improved.

In addition to the standard datasets in the literature, new datasets have been created considering specific situations that visually impaired individuals may encounter. These datasets have been optimized to be most suitable for the needs of visually impaired users.

The speed, tone, and language of the descriptive voices can be adjusted according to the user's preferences. The ability for users to personalize the audio description options is a rare feature in existing products. This customization option makes the system more accessible to users with different language and tone preferences.

The glasses feature a user-friendly interface that allows for easy interaction with touch and voice control systems. With its ergonomic design, it provides comfort during prolonged use and enhances accessibility.

The system features advanced security protocols that provide encryption and offline data processing options to secure user data. This privacy-focused approach allows data to be processed locally without relying on cloud-based systems, thereby maximizing personal privacy. Thanks to the offline operation capability, the system can process user data (images and voice questions) without needing an internet connection and does not export any data. By ensuring that data is not sent to any cloud service or third-party server, potential data breaches are prevented, keeping user information completely confidential. This structure offers a significant security advantage, ensuring that the data remains safe.

DISCUSSION

The design of the glasses aligns with existing habits, providing users with ease of use. The rapid description of images captured by the cameras enables users to receive immediate information. The glasses' form factor makes them lightweight, portable, and easy to wear. With camera glasses, information can be gathered discreetly and without drawing attention to the user. As the independent mobility of visually impaired individuals increases, their quality of life improves. These glasses can be used across a wide range of daily activities, from shopping to social interactions.

However, high processing power and continuous camera use can increase battery consumption, leading to limited usage time. The use of advanced technology and components may raise costs, making the product expensive for some users. Ensuring user privacy and the security of collected data can be a significant challenge. Integrating image processing and audio description systems can be complex and may lead to technical issues.

RESULTS

We measured the accuracy and speed of object recognition, text reading, and face recognition algorithms in the glasses. The prototype was introduced to visually impaired individuals, and the ease of use and effectiveness of the prototype were evaluated. Improvements were made based on user feedback.

CONCLUSION

When compared to one of the most popular models today, GPT-4, our model achieved results that are quite close, despite having only 500 million parameters compared to GPT-4's 1.76 trillion parameters. In tests conducted on 100 samples, GPT-4 achieved an accuracy score of 82%, while our model achieved a score of 78%.

ETHICAL DECLARATIONS

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Bastola, A., Enam, M. A., Bastola, A., Gluck, A., & Brinkley, J. (2023). Multi functional glasses for the blind and visually impaired: design and development. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. doi:10.1177/21695067231192450.
2. Nazim, S., Firdous, S., Shukla, V. K., & Pillai, S. R. (2023). Smart glasses: a visual assistant for the blind. Amity University Dubai.
3. Mustafa, A., Omer, A., & Mohammed, O. (2023). Intelligent Glasses for Visually Impaired People. Sudan University for Science and Technology.
4. Waisberg, E., Ong, J., Masalkhi, M., Zaman, N., Sarker, P., Lee, A.G., & Tavakkoli, A. (2023). Meta smart glasses-large language models and the future for assistive glasses for individuals with vision impairments. Eye, 38,1036-1038. doi:10.1038/s41433-023-02842-z
5. Turk, F. RNGU-NET: a novel efficient approach in segmenting tuberculosis using chest X-ray images. *PeerJ Comp Sci*, 2024,5,10. doi: 10.7717/peerj-cs.1780
6. Türk, Fuat. (2024). Investigation of machine learning algorithms on heart disease through dominant feature detection and feature selection. *Signal Image Video Process*, 18,1-13. doi:10.1007/s11760-024-03060-0
7. Agrawal, A., Lu, J., Antol, S., Mitchell, M., Zitnick, C. L., Batra, D., & Parikh, D. (2016). VQA: Visual question answering. *arXiv*, https://arxiv.org/abs/1505.00468.
8. Papers with Code. (n.d.). Visual Question Answering. Retrieved from https://paperswithcode.com/task/visual-question-answering-1
9. OrCam MyEye 2 Pro. (n.d.). Retrieved from https://www.orcam.com/en us/orcam-myeye-2-pro
10. Envision Glasses. (n.d.). Retrieved from https://www.letsenvision.com/glasses/home
11. Berry, T. (2018, February 9). Visually stunning; aira smart glasses help the blind see airport gates. CW39. Retrieved from https://cw39.com/2018/02/09/visually-stunning-aira-smart-glasses-help-the blind-see-airport-gates.
12. Wikipedia contributors. (2023, December 9). Seeing AI. In Wikipedia, The Free Encyclopedia. Retrieved from https://en.wikipedia.org/wiki/Seeing_AI.