ISSN 2584-041X(P)
ISSN 3048-7730(E)
VOL.06 ISSUE 02

Computer Vision Assistance System for Enhanced Mobility in Visually Impaired Individuals

Tathagat Karan [1] Ashwin Kumar [2] Krishan Kumar [3] Suryansh Gupta [4]
(B. Tech Scholar,)
Department of Electrical Engineering Netaji Subhas University of Technology,
Dwarka, New Delhi **Email:** tathagat.karan.ug21@nsut.ac.in
Email: ashwin.kumar.ug21@nsut.ac.in
Email: krishan.kumar.ug21@nsut.ac.in
Email: suryansh.gupta.ug21@nsut.ac.in

Abstract

This paper introduces a Computer Vision-Based Assistance System to help people with visual impairments in their everyday lives. The system uses cutting-edge image processing and machine learning methods to spot objects, obstacles, and text, giving instant feedback through sound or touch. It aims to boost users' ability to move independently by helping them navigate tricky environments and recognize key info like street signs or product labels. The system we suggest can be built into wearable devices, offering smooth interaction for users with visual impairments. Our approach shows great promise to improve accessibility and overall life quality for people with trouble seeing. This paper describes the assisted vision project that aids visually impaired individuals in recognizing items and others within their field of view.

Keywords: Computer Vision 'Object Distance Estimation; Object Detection and Recognition

Introduction

Computer vision is a discipline focused on enabling computers to perceive their surroundings like human vision. This field encompasses techniques for acquiring, processing, analyzing, and extracting valuable insights from single images or sequences of images. Numerous computer vision technologies have been created to support individuals who are blind or have visual impairments.

Some computer vision techniques are:

- Object Detection
- Facial Recognition
- Feature Detection

The proposed work is easy to use; an assisted vision system has been developed and put into use to assist blind and VI people in moving around in any given area. The system informs the user of the object's distance from them, which it calculates, enabling them to move at a high speed. The system will alert the user when items are approaching him, preventing mishaps.

Past Studies: Review of Literature

ISSN 2584-041X(P)
ISSN 3048-7730(E)
VOL.06 ISSUE 02

As discussed in their paper, Mahendran et al. (2021) articulated that bypasses the need for power-hungry GPU-based equipment needed for the majority of deep learning algorithms to perform inference real-time uses edge AI emulates the like the neural compute stick-2, model optimizer applications such as Open Vino and Tensor Flow Lite and Smart depth-sensing devices such as the OpenCV AI kit depth (OAK-D). It has a system design of a perception module and a user interface. The perception module consists of primitive and advanced perception modules and localized modules. The user interface uses text-to-speech and SMS. In their paper,

Sivan & Darsan (2016) suggested a device with object-detecting components that used text recognition, door detection, BRISK, and an intruder-detection security system. Because it may be used in real-time, the gadget can be used as a mobile assistive device for the blind and visually impaired. The gadget also features an integrated service-learning feature.

According to the findings of this comprehensive review, Anthony & Kusnadi's 2021 study shows that computer vision has a lot of potential for use by blind people. The main issues developers face are assessing proficiency and expediting response times. Eventually, we hope to have a complete package or solution encompassing mapping, indoor and outdoor navigation, object and obstacle recognition, person recognition, human crowd behavior, crowd human counting, study/reading, entertainment, and more for blind and vision-impaired individuals to use on handheld devices (like Android).

In the work of Jafri et al. (2013), we see that one of the biggest challenges the visually impaired have is being able to recognize generic objects in their environment, and to date, there are not very many assistive technology solutions for assisting the visually impaired and those with low vision. However, several computer vision-based opportunities have emerged in the last couple of years to help with this by placing visual tags on objects. What makes these approaches unique is capitalizing on commercial off-the-shelf hardware, mobile, and other technologies to develop an affordable, portable, easy-to-use, and, therefore, attractive solution to an immediate need.

Gaps Identified

After studying the previous paper, we found some gaps in their study. The gap we discovered that

- No talk of facial recognition. There is no talk of distance detection from the user.

ISSN 2584-041X(P)
ISSN 3048-7730(E)
VOL.06 ISSUE 02

- Sometimes, blurring of the images. The processor speed sometimes compromises system performance.

OBJECTIVES

Text recognition: The text recognition feature detects and extracts text from video frames captured by the webcam. This text is then processed and announced using a text-to-speech engine. The feature aims to enhance the script's functionality by recognizing and vocalizing text in the video feed.

GPS: GPS is a satellite-based navigation system that uses satellites to calculate one's location, speed, and time anywhere on the Earth. GPS uses signals from distantly located satellites to know precisely (Loomis et al.,2001) where the receiver is on Earth's surface. It is often used for navigation, mapping, tracking, and many other applications.

Distance Measurement: Distance detection using a camera. Using a single-camera approach, this method estimates the distance between a camera and an object (such as a human face). By leveraging the known real-world width of an object and its apparent size in the captured frame, the system calculates distance based on the principles of the pinhole camera model.

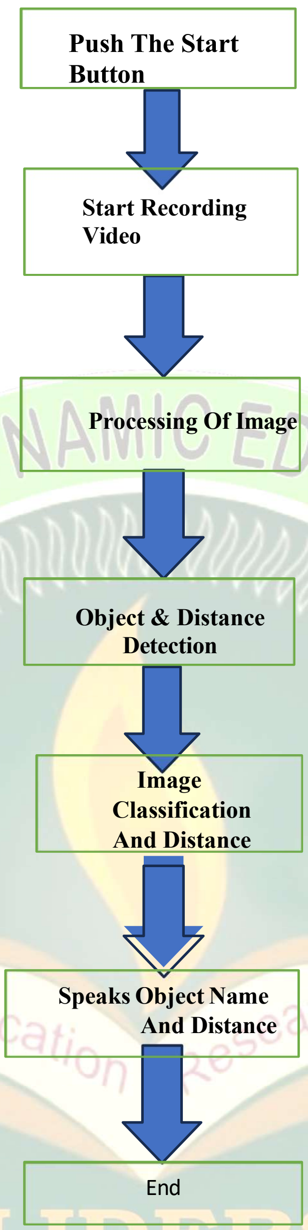
PROBLEM STATEMENT

Visually impaired individuals face significant challenges navigating their surroundings and identifying objects and people in their environment. Current assistive technologies cannot often provide seamless, real-time, accurate recognition of faces and objects.

METHODOLOGY ADOPTED

This proposed work contributes to several things, including text recognition, object detection, and obstacle distance estimation. The Object detection structure is shown in Figure 1.

ISSN 2584-041X(P)
ISSN 3048-7730(E)
VOL.06 ISSUE 02



(Figure 1)

METHODOLOGY ADOPTED

Text recognition

- **Convert Frame to Grayscale**: The video frame is converted to a grayscale image to simplify text recognition.
- **Extract Text Using Tesseract**: The grayscale image is passed to Tesseract OCR to extract text. Tesseract is an open-source OCR engine that can recognize text in images.
- **Process Extracted Text**: The extracted text is cleaned by removing extra spaces and ensuring it meets a minimum length requirement. This step helps in avoiding false

ISSN 2584-041X(P)
ISSN 3048-7730(E)
VOL.06 ISSUE 02

positives and irrelevant text.

- **Announce Detected Text**: If the processed text is sufficiently long and a specific time delay has passed since the last announcement, the text is announced using the text-to-speech engine.
- **Display the Frame**: The video frame is displayed with any detected text, faces, and objects. This provides a visual representation of the detections.

Object detection

The primary component of the device is object detection, which is the action of locating objects. Many algorithms in deep learning and machine learning focus on object detection. In this situation, we will be training the device to detect objects in a way that it will identify an object when it is in front of the device. Thus, object detection is essential to knowing what object is in front of the user.

Single Shot Detector

SSD classifies classes by one feedforward convolutional network structure without requiring a two-stage classification. It eliminates bounding box proposals, such as those used in R-CNNs. It uses a trainable stack of convolutional filters that progressively downsample spatial dimensions to detect object classes and bounding box offsets in prediction. VGG-16 serves as the foundation network of SSD, with several multi-box convolutional layers that follow. Eight thousand seven hundred thirty-two bounding boxes take SSD to realize the highest attainable scale, aspect ratio, and location capture. SSD never marks an object when the confidence level is lower than 0.001. The loss function includes confidence loss mixed with localization loss.

Distance detection

Detect faces using a Haar Cascade classifier, compute focal length using a reference image, estimate distance in real-time, and overlay the calculated distance onto the video feed.

Result and discussion

The experiment results were achieved with a trained Microsoft Coco dataset, SSD MobileNet (Howard et al., 2001). Figure 2 shows the start button of the proposed application. Every object detected will be assigned a score, which will show the accuracy of the matched object when identifying the trained object. In Figure 2, one notices that the camera determines the object (bed) in front of it, and the object is given a score based on the accuracy of object identification. In the example of the object bed, the object's given score is 79%; therefore, it states that it is higher than the object. Along with Figure 3, the objects identified through the object detection

ISSN 2584-041X(P)
ISSN 3048-7730(E)
VOL.06 ISSUE 02

method approach (backpacks and beds) will also be highlighted. The proposed application can also detect mobile phones; as seen in Figure 5, a cell phone is identified with a score of 84%.

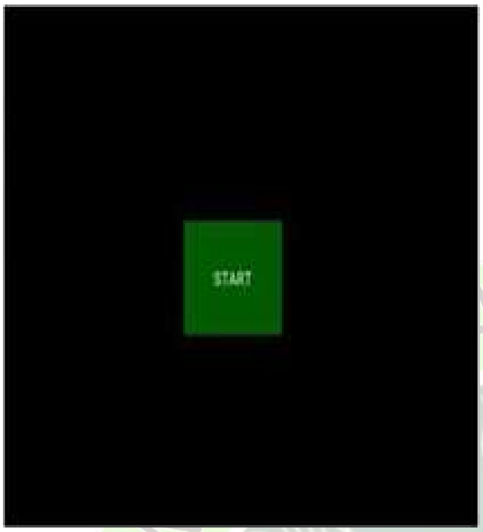


Figure 1. Start Button

Figure 2. Detecting Objects with 79%

Figure 3: Detecting a backpack with 71%.

Figure 4: Detecting a backpack with 75%

ISSN 2584-041X(P)
ISSN 3048-7730(E)
VOL.06 ISSUE 02

Figure 5: Detecting mobile phones with 84%

Conclusion

Calculating distance was crucial as it offered the most significant advantage to the user. When the user engages with the device, any item in front of them will be recognized and conveyed to the user, along with the object's distance from their position. This will increase its utility since he can now comprehend the object's position before him rather than just identifying which object is before him. The assisted vision project assists visually impaired individuals in recognizing objects and individuals within their sight range. It employs object detection, object recognition, and measurement of distance. We anticipate integrating this system into glasses worn by visually impaired individuals heading towards him.

For example, suppose a vehicle approaches the user closely and its behavior indicates that it may collide with the user. In that case, the system will then prompt the user to move in a particular direction as a way for the user to escape from these actions.

References

Mahendran, J.K., Barry, D.T., Nivedha, A.K. and Bhandarkar, S.M., 2021. Computer vision-based assistance system for the visually impaired using mobile edge artificial intelligence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2418-2427).

Sivan, S. and Darsan, G., 2016, July. Computer vision-based assistive technology for blind and visually impaired people. In *Proceedings of the 7th International Conference on Computing Communication and Networking Technologies* (pp. 1-8).

Anthony, E.J. and Kusnadi, R.A., 2021. Computer Vision for Supporting Visually Impaired People: A Systematic Review. *Engineering, Mathematics and Computer Science Journal (EMACS)*, *3*(2), pp.65-71.

ISSN 2584-041X(P)
ISSN 3048-7730(E)
VOL.06 ISSUE 02

Rahman, M.A. and Sadi, M.S., 2021. IoT enabled automated object recognition for the visually impaired. *Computer methods and programs in biomedicine update*, *1*, p.100015.

Jafri, R., Ali, S.A., Arabnia, H.R. and Fatima, S., 2014. Computer vision-based object recognition for the visually impaired in an indoors environment: a survey. *The Visual Computer*, *30*, pp.1197-1222.

lzahrani, N. and Al-Baity, H.H., 2023. Object recognition system for the visually impaired: a deep learning approach using Arabic annotation. *Electronics*, *12*(3), p.541.

Caraiman, S., Morar, A., Owczarek, M., Burlacu, A., Rzeszotarski, D., Botezatu, N., Herghelegiu, P., Moldoveanu, F., Strumillo, P. and Moldoveanu, A., 2017. Computer vision for the visually impaired: the sound of vision system. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 1480-1489).

Patel, K. and Parmar, B., 2022. The assistive device uses computer vision and image processing for the visually impaired, as well as review and current status. *Disability and Rehabilitation: Assistive Technology*, *17*(3), pp.290-297.

Loomis, J. M., Golledge, R. G., & Klatzky, R. L. (2001). GPS-based navigation systems for the visually impaired. *Fundamentals of wearable computers and augmented reality*, *429*, 46.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, *37*(9), 1904-1916.

Tapu, R., Mocanu, B., & Zaharia, T. (2017). A computer vision-based perception system for visually impaired. *Multimedia Tools and Applications*, *76*, 11771-11807.

Aduri, K. R., & Naganjaneyulu, P. V. (2024, December). Advancements in computer vision-based assistive technologies for visually impaired. In *IET Conference Proceedings CP908* (Vol. 2024, No. 30, pp. 465-474). Stevenage, UK: The Institution of Engineering and Technology.

Arakeri, M. P., Keerthana, N. S., Madhura, M., Sankar, A., & Munnavar, T. (2018, September). Assistive technology for the visually impaired using computer vision. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1725-1730). IEEE.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.