CNN-Based Object Recognition and Tracking System to Assist Visually Impaired People

Nandashree V

Student  
*Department* *Of* *Electronics and Communication*  
*Sapthagiri College Of Engineering  
(Affiliated to Visvesvaraya Technology University,Belagavi )*Banglore,India  
[nandashreev2002@gmail.com](mailto:nandashreev2002@gmail.com)

Tejaswini B

Student  
 *Department* *Of* *Electronics and Communication*  
 *Sapthagiri College Of Engineering  
(Affiliated to Visvesvaraya Technology University,Belagavi)*Banglore,India tejugowdatejugowda919@gmail.com Prerana C K

Student  
 *Department* *Of* *Electronics and Communication*  
*Sapthagiri College Of Engineering  
(Affiliated to Visvesvaraya Technology University,Belagavi)* Banglore,India  
 preranack6@gmail.com

Padmavathi C

Assistant Prof.

*Department of Electronics and*

*Communication*

*Sapthagiri College Of Engineering*

*(Affiliated to Visvesvaraya Technology*

*University, Belagavi)*

Bangalore, India

padmavathic@sapthagiri.edu.in@gmail.com

Supritha M

Student  
 *Department* *Of* *Electronics and Communication*  
 *Sapthagiri College Of Engineering  
(Affiliated to Visvesvaraya Technology University,Belagavi)*Banglore,India  
 suprithamurali02@gmail.com

*Abstract*—

This study introduces a smart system tailored for visually impaired persons (VIPs) to enhance their mobility and safety. Leveraging technology, the system offers real-time navigation through automated voice guidance. While VIPs may not see their surroundings, they can sense and comprehend their environment with this assistance. Additionally, a web-based application ensures their security by enabling location sharing with family members on demand, balancing privacy concerns. Through this application, families can track VIPs' movements remotely, fostering peace of mind. This comprehensive solution addresses a critical gap in existing literature. The system utilizes the Mobile Net architecture for its computational efficiency, enabling it to operate on low-power devices. Six pilot studies validate the system's effectiveness, achieving an 83.3% accuracy in object detection and recognition across a dataset with over 1000 categories. Furthermore, a comparative analysis demonstrates the system's superiority over existing devices, scoring 9.1/10, an 8% improvement over the nearest competitor.

Keywords—Raspberry pi, Voice conversion, object detection (key words)

# Introduction

The World Health Organization (WHO) has reported that 285 million people worldwide are blind or visually impaired, with 39 million being blind [1]. The leading causes of visual impairment include refractive error, glaucoma, trachoma, corneal opacities, cataracts, diabetic retinopathy, and unaddressed presbyopia [2]. Visually impaired individuals (VIPs) encounter challenges in performing daily activities such as working, attending school, navigating their surroundings, interacting with the environment, and identifying common objects both indoors and outdoors, whether independently or with assistance [3]. The primary challenges for VIPs include object detection and recognition, currency identification, interpretation of textual information (signs, symbols), translation, mobility, navigation, and safety [3].Various approaches, systems, devices, and applications have been developed in the field of assistive technology to address the needs of VIPs, enabling them to perform tasks they were previously unable to accomplish [4]. These solutions typically involve electronic devices equipped with cameras, sensors, and microprocessors capable of making decisions and providing tactile or auditory feedback to the user.The primary objective of this study is to develop a system for VIPs with the following features:1. Real-time object detection and recognition using a deep learning framework.2. Audible pronunciation of the names of objects visible through the camera, i.e., objects present in the current frame.

# LITERATURE REVIEW

Researchers have proposed various techniques to develop assistive devices for visually impaired persons (VIPs). Object detection devices utilize sensors such as laser scanners, ultrasonic devices, and cameras to gather information from the surrounding environment, process it, and provide feedback to users. These devices typically work by detecting objects around the user and conveying instructions regarding the object/obstacle and its distance through vibrations or sound waves. Saputra et al. [5] introduced an obstacle avoidance system utilizing a Kinect depth camera for VIPs. This system detects obstacles and calculates their distance using an auto-adaptive threshold. Tested on ten blind individuals aged 20-40 years, the system demonstrated promising results by detecting obstacles without collisions from any direction. Yi and Dong [6] presented a blind-guide crutch equipped with multiple sensors. A triplet ultrasonic module detects obstacles from the front, left, and right sides, identifying objects through voice and vibration waves.Kumar et al. [7] introduced an ultrasonic cane providing environmental information to ensure safe navigation. The ultra-Cane incorporates a narrow beam ultrasound system for 100% obstacle detection up to 2-4 meters away. Tested on ten individuals aged 20-26 years, volunteers effectively detected hurdles within the specified range. Petal et al. [8] developed a multi-sensor system facilitating indoor object detection. This system employs statistical parameters and the SVM algorithm for object detection. Chen et al. [9] introduced an application aiding in obstacle detection through glasses, a long stick cane, and a mobile application. If the user falls, information is uploaded to an online platform and displayed on the mobile device. Pogi and Mattacin [10] devised a wearable system for VIPs utilizing deep learning and 3D vision via a smartphone. This system utilizes an RGB camera to capture frames and CNN for obstacle detection. While boasting exceptional performance nearing 98% using the LeNet architecture, it struggles with categorizing most daily life objects. Users are guided through tactile and audio feedback.

Ⅲ PROBLEM STATEMENT

Existing assistive technologies for the visually impaired, though beneficial to some extent, frequently lack the capability to accurately detect and track objects in real-time. Additionally, these technologies may be impractical to use, costly, or inaccessible to many individuals who could derive benefits from them.

# Ⅳ PRPOSED FRAMEWORK

## Hardware components used

1. Raspberry pi
2. Camera
3. Speaker
4. Mic
5. Power Supply

## Software tools used

1. YOLO
2. OpenCV
3. Python

## System Architecture

The proposed system consists of a Raspberry Pi digital signal processing (DSP) board equipped with GSM and a global positioning system (GPS) module, along with headphones and a camera. The DSP captures a live feed from the video camera and transfers it to the object detection and recognition module, which utilizes a convolutional neural network (CNN) model. This model identifies objects in the current video frame and relays their names to the text-to-speech converter (SAPI) module, which then pronounces the names using the headphones.

## Implementation

Objective 1

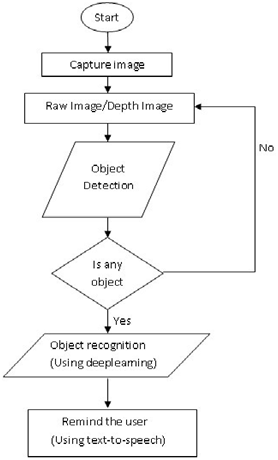


Fig.1 Flow chart for object detection

Fig.1 operates as follows: it receives input from the camera, conducts object detection and recognition, and vocalizes the names of objects visible through the camera. Notifications are conveyed to the user via headphones. The system is capable of performing multiple object detections and recognitions in a single snapshot. When multiple objects are present in one snapshot, they are arranged in descending order based on their confidence levels or probabilities of prediction. The object with the highest confidence level is pronounced first, followed by the remaining objects. If an object is detected, whether correctly or incorrectly, information regarding the closest match is provided to the user. However, if the object is unknown and cannot be identified, the system will convey to the user that there is an unknown object and it is unable to recognize it.

Objective 2

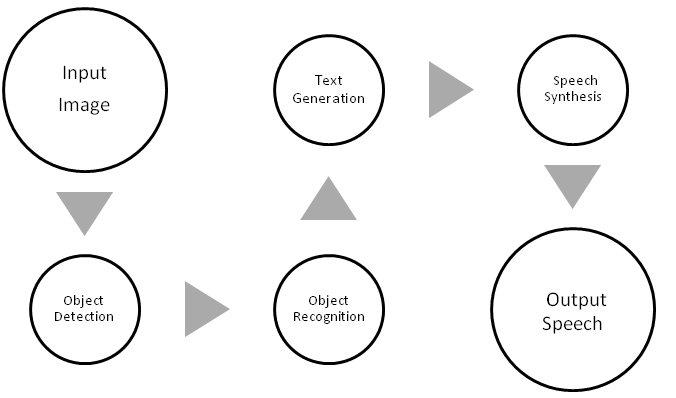


Fig.2: Recognized object to text.

In this Fig.2 the Speech API (SAPI 5.3) [11] is employed for audio synthesis based on the input text. To articulate the detected objects in the current frame, SAPI receives textual input comprising the names of objects detected by the CNN model and produces an audio signal accordingly. The resulting audio output from SAPI is directed to the headphones, enabling visually impaired persons (VIPs) to audibly perceive the names of the detected objects.

V CONCLUSION

This paper introduces an intelligent system designed to assist visually impaired persons (VIPs) with mobility and safety concerns by addressing their everyday needs. The proposed system is tailored to help VIPs visualize their environment and gain a better understanding of their surroundings. By utilizing a CNN-based low-power Mobile-Net architecture, the system enables users to recognize objects in their vicinity and perceive the natural environment.

REFERENCES:

[1] W. Elmannai and K. Elleithy, ‘‘Sensor-based assistive devices for visually impaired people: Current status, challenges, and future directions,’’ Sensors, vol. 17, no. 3, p. 565, 2017.

[2] Work Sheet. Accessed: Mar. 7, 2020. [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/blindness-andvisual-impairment

[3] R. Velázquez, ‘‘Wearable assistive devices for the blind,’’ in Wearable and Autonomous Biomedical Devices and Systems for Smart Environment. Berlin, Germany: Springer, 2010.

[4] L. B. Neto, F. Grijalva, V. R. M. L. Maike, L. C. Martini, D. Florencio, M. C. C. Baranauskas, A. Rocha, and S. Goldenstein, ‘‘A Kinect-based wearable face recognition system to aid visually impaired users,’’ IEEE Trans. Human-Mach. Syst., vol. 47, no. 1, pp. 52–64, Feb. 2017.

[5] M. R. U. Saputra, Widya wan, and P. I. Santosa, ‘‘Obstacle avoidance for visually impaired using auto-adaptive thresholding on Kinect’s depth image,’’ in Proc. IEEE 11th Int. Conf. Ubiquitous Intell. Compute., IEEE 11th Int. Conf. Auton. Trusted Compute., IEEE 14th Int. Conf. Scalable Compute. Commun. Associated Workshops, Dec. 2014, pp. 337–342.

[6] Y. Yi and L. Dong, ‘‘A design of blind-guide crutch based on multi sensors,’’ in Proc. 12th Int. Conf. Fuzzy Syst. Know. Discovery (FSKD), Aug. 2015, pp. 2288–2292.

[7] K. Kumar, B. Champaty, K. Uvanesh, R. Chachan, K. Pal, and A. Anis, ‘‘Development of an ultrasonic cane as a navigation aid for the blind people,’’ in Proc. Int. Conf. Control, Instrum., Commun. Compute. Technol. (ICCICCT), Jul. 2014, pp. 475–479.

[8] C. T. Patel, V. J. Mistry, L. S. Desai, and Y. K. Meghrajani, ‘‘Multi sensor– based object detection in indoor environment for visually impaired people,’’ in Proc. 2nd Int. Conf. Intell. Compute. Control Syst. (ICICCS), Jun. 2018, pp. 1–4.

[9] L.-B. Chen, J.-P. Su, M.-C. Chen, W.-J. Chang, C.-H. Yang, and C.-Y. Sie, ‘‘An implementation of an intelligent assistance system for visually impaired/blind people,’’ in Proc. IEEE Int. Conf. Consume. Electron. (ICCE), Jan. 2019, pp. 1–2.

[10] M. Poggi and S. Mattoccia, ‘‘A wearable mobility aid for the visually impaired based on embedded 3D vision and deep learning,’’ in Proc. IEEE Symp. Compute. Commun. (ISCC), Jun. 2016, pp. 208–213.

[11] Speech API Overview (SAPI 5.3). Accessed: Apr. 3, 2020.[Online].Available:https://docs.microsoft.com/enus/previousversions/windows/desktop/ms720151(v%3Dvs.85)