Enhancing Accessibility: A Raspberry Pi-Based Image-to-Speech System for the Visually Impaired

Kosaraju Swathi[1], Bandela Rajani[2], P. L. Siri keerthika[3], G. Mattaiah[4], V. Madhusudhan[5]
[1],[2]Assistant Professor,[3],[4],[5]Undergraduate. Department of ECE,
PSCMR College Of Engineering And Technology, Vijayawada, Andhra Pradesh, India swathi7201@gmail.com[1], rajani.bandela@gmail.com[2], sirikeerthikapanchangam@gmail.com[3], matthewsgarnepudi9@gmail.com[4], Madhusudhanvakkalanka2002vms@gmail.com[5]

Abstract—Visual impairment poses considerable obstacles in the text-centric communication environment of the twenty-first century. In response to this concern, we introduce an innovative apparatus that has the ability to convert text contained within images into audio, thereby enhancing accessibility for individuals with visual impairments. By means of an embedded system architecture, our solution incorporates a Raspberry Pi as well as a Raspberry Pi camera. By employing a sequence of image pre-processing methods, the system effectively discerns the textcontaining areas from the background noise of captured images. The process of selective extraction improves the effectiveness and precision of text recognition . The extracted text is then rendered into speech , facilitating the dissemination of information orally . Through the utilization of sophisticated hardware components and image processing algorithms, our device provides a pragmatic and economical resolution to enhance the legibility of textual material for individuals who have visual impairments.

Index Terms—Visual impairment, text-to-speech, image processing, accessibility, Raspberry Pi, embedded system, text recog-nition, image pre-processing, audio rendering, hardware compo-nents, visual impairment aid

I. Introduction

Object detection is a critical component of computer vision, with the objective of recognizing and locating recognized objects in a scene, preferably in three-dimensional (3D) territory. Regaining the precise pose of 3D objects is of the utmost significance, especially for robotic control systems, because it permits accurate spatial comprehension and manipulation. The goal of endowing machines with intelligence and augmenting robotic autonomy has historically been a pioneering motivation in the field of technology [1]. Our overarching goal is to transfer monotonous, dangerous, or mundane duties to auto-mated systems, thereby liberating human capabilities for more imaginative endeavors. The present constraints in machine intelligence, nevertheless, hinder the achievement of this aspiration. In addition to hardware advancements, sophisticated software frameworks are required to attain such autonomy. However, the fundamental aspect of object recognition is the capacity to differentiate between objects, even if they are members of the same category. This is a formidable obstacle for machines that do not possess an exhaustive understanding of object variability. In order to address this disparity, it is imperative to devise resilient algorithms and machine learning methodologies that endow robots with the cognitive capacities necessary for precise object differentiation and detection. The integration of intelligent software and hardware innovation is of the utmost importance for expanding the capabilities of autonomous systems and achieving the potential for greater automation across multiple domains. Object Detection is an essential procedure within the field of computer vision that identifies and locates tangible entities, including furniture, vehicles, and pedestrians, within videos or images. This methodology allows for the concurrent identification of numerous objects, which streamlines operations such as image retrieval, surveillance, and autonomous driving systems. A multitude of methodologies comprise object detection techniques, such as Feature-Based Object Detection, YOLO V4 Object Detection, and Deep Learning Object Detection. Object detection is of the utmost importance, especially in video surveillance, as it enables the identification and tracing of objects across multiple frames [2]. Feature extraction, pre-processing, and segmentation are customary steps in the procedure to distinguish foreground objects from the background. Humans possess an exceptional ability to detect objects by virtue of their visual system's rapid and precise operation. However, in order to emulate this capability in machines, advanced algorithms and deep learning methodologies are necessary. Object detection continues to be a fundamental component in the progression of computer vision applications across various domains, serving as a link between human perception and the intelligence of machines.

II. Literature Survey

A multitude of sophisticated methodologies are utilized to augment navigation assistance for individuals who are visually impaired. The aforementioned categories consist of Electronic Orientation Aids (EOAs), Place Locator Devices (PLDs), and Electronic Travel Aids (ETAs), each of which provides distinct advantages (Barathkumar, K.). As an illustration, an advanced ETA incorporates ultrasonic sensors to identify obstacles within designated ranges. Wearable devices utilize semantic and three-dimensional data representations to convey information via haptic feedback or text-to-speech systems. (Rathi, A.) The utilization of visible light frequencies for wireless transmission enables communication between wearable smart eyewear and signs. Furthermore, (LATHA, L.) data analytics is employed by Internet of Things (IoT) frameworks to improve navigation aides by providing haptic and vocal feedback to identify obstacles. (Deshmukh, S.) Vision impairment can be assisted through the utilization of specialized cameras and infrared technology, which employ machine learning techniques to caption images. Preliminary Data Requirements (PDR) and the Global Positioning System (GPS) facilitate indoor navigation through the utilization of inertial sensors that ensure accurate positioning. Deep learning algorithms generate semantic maps for indoor and outdoor navigation by analyzing RGB images (Bissacco, A). In addition, the implementation of convolutional neural networks in smart eyewear enables the process of auditory drug identification. Embedded systems, including Raspberry Pi, augment the func-tionality of walking poles by incorporating GPS tracking and obstacle warning functionalities. These developments highlight a deliberate attempt to utilize technology in order to alleviate the difficulties encountered by those with visual impairments, offering the potential for enhanced autonomy and security when performing navigational duties.

III. Methodology

In this section, we are going to discuss about the block diagram and algorithm for the object detection as follows.

A. SYSTEM BLOCK DIAGRAM

In order to enhance functionality for individuals with visual impairments, our system integrates judiciously chosen hardware components alongside effective algorithms. When making the hardware selection, consideration is given to weight, battery life, and night vision capability. For comprehensive functionality, our configuration includes Raspberry Pi 4B , a camera, ultrasonic sensor, and Arduino as shown in fig.1. Two fundamental aspects comprise the operation of the system: obstacle detection and image processing. In order to detect ob-stacles, the Arduino-programmed ultrasonic sensor computes distances and emits beeping noises to indicate the presence of impediments along the trajectory. Through the manipulation of bleep quantity and variety, the system discerns the proximity and trajectory of obstacles. The Raspberry Pi camera acquires scene images concurrently in preparation for preprocessing. The Viola Jones and Tensor Flow Object Detection algorithms are applied to these images, respectively, for face detection and generic object recognition. The Viola Jones algorithm is designed to process grayscale images by employing Ada Boost classifiers to classify features extracted using Haar-like patterns [3]. This procedure identifies features within the image with efficiency [4]. To accurately detect objects, the Tensor Flow Object Detection API partitions images into small bounding boxes, extracts features, and combines overlapping boxes. overlapping boxes.

Hardware implementation entails the assembly of various components, such as a rechargeable battery, Raspberry Pi 4, ultrasonic sensor, camera, and servo actuator, onto a stick [5]. The camera is affixed to the stick in order to record video in real-time, whereas the ultrasonic sensor, which is affixed to the ground, assists in the detection of obstacles such as stairs and potholes. To clear pathways, a servo motor located beneath the sensor facilitates rotational tracking. By functioning as a power source, the battery guarantees continuous operation. The Raspberry Pi board functions as the central processing engine, coordinating the operations of connected devices. In its entirety, our system integrates sophisticated hardware components with optimized algorithms, thereby guaranteeing effective navigation assistance for individuals with visual impairments across diverse environments [6].

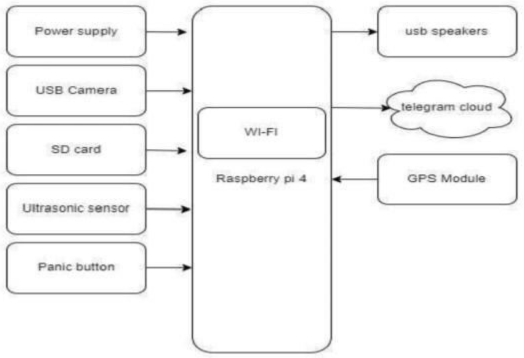


Fig. 1. Block Diagram of Proposed System

B. CLASSIFICATION THROUGH VIOLA JONES ALGORITHM

Developed by Viola and Jones in 2001, the Viola-Jones object detection framework utilizes rectangular features that are generated by subtracting darkened rectangles from plain ones and summing pixel sums within defined regions. This process results in the identification of unique attributes. While initially developed for face detection, the model has since been adapted to perform object recognition tasks on a broader scale. Its efficacy has been enhanced since its integration with Open CV. Through the implementation of a cascaded classifier, which effectively merges specific features, the framework attains precise detection outcomes for a wide range of objects. In order to optimize efficiency, Ada Boost is employed to bolster edge-type horizontal features and line features [7]. This procedure combines weak classifiers into a single strong classifier, resulting in enhanced accuracy and decreased processing time. Significantly, this methodology facilitates the accurate identification of upper body regions, specifically augmenting facial characteristics and discerning eyes by means of Haar-like rectangular line attributes. The model's efficacy is enhanced by its low noise and high accuracy, which enable it to achieve remarkable results in object detection tasks, especially when applied to visually impaired individuals .

1) DETECTION SCHEME: The algorithm is executed as shown in fig.2

1) Image importation.
2) Construct pre-processing.
3) Implement the Viola-Jones algorithm in order to detect faces.
4) Categorize found facial characteristics.

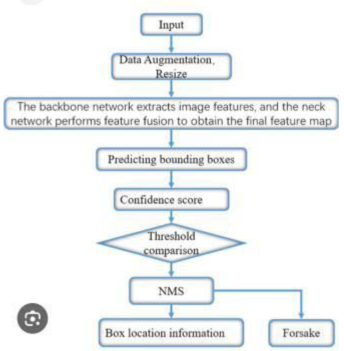

Input
Data Augmentation, Resize
The backbone network extracts image features, and the neck network performs feature fusion to obtain the final feature map
Predicting bounding boxes
Confidence score
Threshold comparison
NMS
Box location information
Forsake

Fig. 2. Working of YOLO V4 Algorithm

5) Eliminate images devoid of identified features.
6) Iterate over the remaining characteristics.
7) Define terminate if no further features are discovered.

Among the steps of the algorithm is

1) Haar feature selection.
2) Image integral generation.
3) Ada boost instruction.
4) Classifiers that cascade.

By effectively discerning and categorizing facial characteristics within images, this methodology maximizes the utilization of computational resources in the domain of facial recognition.

2) HAAR FEATURES: Haar features are capable of match-ing common facial features, including darker eye regions in comparison to the upper cheeks and whiter nasal bridges in contrast to the eyes [8]. To discern facial features through di-rected gradients of pixel intensities, Viola and Jones proposed two-rectangle features, which encompassed three-, four-, and two-rectangle varieties.

F(Haar) = F white–F black

where FWhite represents pixels in white area and FBlack represents pixels in black area

3) CREATING AN INTEGRAL IMAGE: Real-time representation of the integral image facilitates the examination of rectangle characteristics. The proposed method effectively calculates three-rectangle features and obtains four-rectangle features by utilizing six, eight, and nine array references, respectively. This significantly improves the computational efficiency of facial characteristic identification [9]

4) ADABOOST TRAINING: A variant of AdaBoost is utilized by the object identification framework in order to choose and train classifiers. AdaBoost constructs a resilient classifier through the linear combination of weighted weak classifiers, thereby augmenting the system's capacity to efficiently discern valuable attributes.

IV. Results and Discussions

The object identification framework integrates a multilayer cascaded classifier that has been trained with the specific purpose of identifying frontal upright features as shown in figures.3,4. The procedure entails training the classifier with a blend of facial and non-facial training images utilizing the AdaBoost method. This approach constructs a robust classifier by systematically integrating weighted weak classifiers.

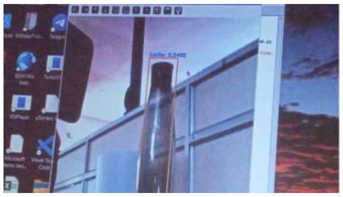


Fig. 3. Detection Of Bottle

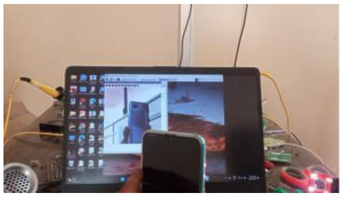
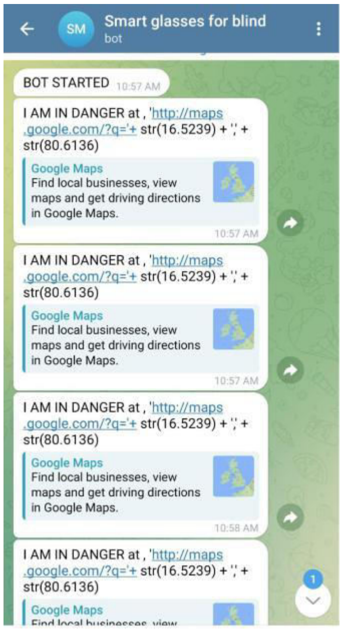


Fig. 4. Detection Of Mobile

Furthermore, Haar features are utilized in the classification process, whereby the grayscale values of image pixels determine whether they are Fwhite or FBlack, thereby facilitating the classification of binary objects. Following this, a range of facial feature detectors are evaluated with an emphasis on more minute patterns present in faces. The assessment consists of detecting the left eye, right eye, both eyes in conjunction, nose, and mouth, and validating the obtained results against the algorithm's accuracies as shown in fig.6. The Coco dataset is employed for object detection purposes [10]. It provides a wide variety of images that can be utilized to identify objects. The workflow includes image segmentation, key point detection, and captioning. In order to assess the performance of the system, images are captured and its accuracy in object identification is evaluated as shown in fig.5.

Significantly, the system attains remarkable outcomes, as evidenced by the high probabilities assigned to identified objects, including backpacks, keyboards, and suitcases; this translates to efficiencies varying from 70% to 97%Nevertheless, complications ensue when objects are not accurately positioned or display diverse reflections, which compromise the precision of the system. Additionally, detection accuracy is influenced by image clarity; images that are livelier and more distinct produce superior outcomes [11]. Notwithstanding these obsta-cles, the system accurately discerns objects, thereby furnishing visually impaired individuals with invaluable support through the detection of impediments and facilitation of navigation as shown in fig.6, along designated routes. Furthermore, the

Fig. 5. Detection Of Person



SM Smart glasses for blind
bot

BOT STARTED 10:57 AM

I AM IN DANGER at , 'http://maps.google.com/?q='+ str(16.5239) + ',' + str(80.6136)

Google Maps
Find local businesses, view maps and get driving directions in Google Maps.
10:57 AM

I AM IN DANGER at , 'http://maps.google.com/?q='+ str(16.5239) + ',' + str(80.6136)

Google Maps
Find local businesses, view maps and get driving directions in Google Maps.
10:57 AM

I AM IN DANGER at , 'http://maps.google.com/?q='+ str(16.5239) + ',' + str(80.6136)

Google Maps
Find local businesses, view maps and get driving directions in Google Maps.
10:58 AM

I AM IN DANGER at , 'http://maps.google.com/?q='+ str(16.5239) + ',' + str(80.6136)

Google Maps
Find local businesses, view

Fig. 6. Navigation through GPS

system exhibits the capacity to identify numerous objects contained within a solitary image, attaining notable efficacy in the detection of furniture and mattresses. Enhanced detection accuracy is a direct result of clearer images, underscoring the potential of the system to aid visually impaired users by effectively identifying obstacles and facilitating seamless navigation [12]. In general, the system's commendable efficacy in distinguishing a multitude of objects highlights its capacity to augment the autonomy and security of individuals with vi-sual impairments by providing indispensable aid in efficiently navigating their environment.

V. Conclusion

In summary, our intelligent eyewear signifies a substantial progression in assisting individuals with visual impairments, promoting enhanced self-assurance and autonomy in their day-to-day activities. Through the utilization of Optical Character Recognition (OCR) technology to identify adjacent objects and decipher text blocks, our prototype system empowers users to effortlessly navigate securely and retrieve printed information. In subsequent iterations, the reading 1 experience for the visually impaired will be enhanced through the resolution of user interface challenges and the improvement of text localization algorithms. Furthermore, potential developments encompass the translation of texts into multiple languages [13], the integration of grammatical structures to enable sign language recognition, and gesture-controlled input for alpha-bets and numerals. Moreover, the potential of the system to effectively handle handwritten notes is encouraging in terms of wider usability and functionality. This portable solution func-tions without reliance on internet connectivity, guaranteeing uninterrupted efficacy for users who have visual impairments

[14]. Our continuous endeavors are focused on enhancing the capabilities and inclusivity of our device so that visually impaired individuals can interact with their environment and information to a greater extent.

References

[1] Barathkumar, K., Balaji, S., Desikan, K. N., Iyappan, S. P. (2019). Rasp-berry Pi based Smart Reader for Visually Impaired People. International Journal of Research in Engineering, Science and Management, 2(2), 204-207.

[2] Rathi, A., Nikalje, A. V. (2019). Review on portable camera based assistive text and label reading for blind persons. Int Res J Eng Technol (IRJET), 6(12), 879-882.

[3] LATHA, L., GEETHANI, V., DIVYADHARSHINI, M. HELLA-A SMART READING BOT. Dalmeda, J., Raju, G. K., Reddy, D. N. (2022). Reader and Object Detector for Blind. International Research Journal of Modernization in Engineering Technology and Science, 2530-2536.

[4] Deshmukh, S., Rede, P., Sharma, S., Iyer, S. (2021, December). Voice-Enabled Vision For The Visually Disabled. In 2021 International Confer-ence on Advances in Computing, Communication, and Control (ICAC3) (pp. 1-6). IEEE.

[5] Zaman, H. U., Mahmood, S., Hossain, S., Shovon, I. I. (2018, October). Python based portable virtual text reader. In 2018 Fourth International Conference on Advances in Computing, Communication Automation (ICACCA) (pp. 1-6). IEEE.

[6] Kumari, K. N., Meghana Reddy, J. (2016). Image text to Speech conversion using OCR technique in Raspberry pi. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 5(5).

[7] Bissacco, A., Cummins, M., Netzer, Y., Neven, H. (2013). Photoocr: Reading text in uncontrolled conditions. In Proceedings of the ieee international conference on computer vision (pp. 785-792)

[8] Koo, H. I., Kim, D. H. (2013). Scene text detection via connected component clustering and nontext filtering. IEEE transactions on image processing, 22(6), 2296-2305.. Trilla, A., Alias, F. (2012). Sentence-based sentiment analysis for expressive text-tospeech. IEEE transactions on audio, speech, and language processing, 21(2), 223-233.

[9] Tang, H., Zhang, X., Wang, J., Cheng, N., Xiao, J. (2023, June). Qi-tts: Questioning intonation control for emotional speech synthesis. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

[10] Schnell, B., Garner, P. N. (2021, August). Improving emotional tts with an emotion intensity input from unsupervised extraction. In Proc. 11th ISCA Speech Synth. Workshop (pp. 60-65).

[11] An, S., Ling, Z., Dai, L. (2017, December). Emotional statistical parametric speech synthesis using LSTM-RNNs. In 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) (pp. 1613-1616). IEEE.

[12] Tang, H., Zhang, X., Wang, J., Cheng, N., Xiao, J. (2023). Emomix: Emotion mixing via diffusion models for emotional speech synthesis. arXiv preprint arXiv:2306.00648.

[13] Zandie, R., Mahoor, M. H., Madsen, J., Emamian, E. S. (2021). Ryanspeech: A corpus for conversational text-to-speech synthesis. arXiv preprint arXiv:2106.08468.

[14] Tan, X., Chen, J., Liu, H., Cong, J., Zhang, C., Liu, Y., ... Liu, T. Y. (2024). Naturalspeech: Endto-end text-to-speech synthesis with human-level quality. IEEE Transactions on Pattern Analysis and Machine Intelligence.