SMART WEARABLE ASSISTIVE SYSTEM USING AI AND DEEP LEARNING FOR THE VISUALLY IMPAIRED AND DYSPHONIC INDIVIDUALS

Project Reference No.: 49S_BE_5900

College : J.S.S. Science and Technology University, Mysuru
Branch : Department of Electronics and Instrumentation Engineering
Guide(s) : Prof. Rajeswari. P
Dr. Rathnakara. S
Student(s) : Mr. Mohammed Suhaib Shariff
Ms. Srividhya H
Ms. Disha Samyami

Keywords:

Multimodal AI System, Assistive Computing, Blind Assistance, Assistive Devices.

Introduction:

Individuals with visual impairments face challenges in accessing information and communicating effectively, impacting their independence in daily life. Existing assistive tools are limited, as they often cannot read text in real time or lack practical, AI-driven solutions suitable for everyday environments. There is a need for a portable, intelligent system that integrates computer vision and audio technologies to provide accurate, real-time assistance.

Objectives:

1. Capturing text and color information, converting the extracted content into speech, and supporting a memory mode that stores visual information and later retrieves it by matching live camera images for real-time assistance through earphones.
2. Reconstruction of dysphonic (impaired) speech into text and generate normal speech from the processed text, providing improved and understandable vocal output.

Methodology:

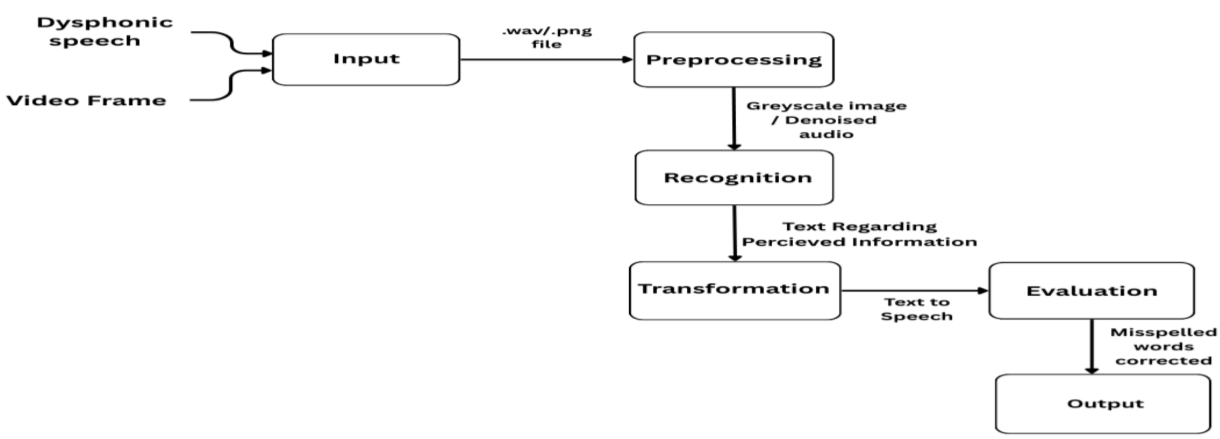

Figure 1. Block diagram of wearable smart spectacles based assistive system for blind and dysphonic individuals

The system begins by capturing inputs in the form of video frames and audio signals through the camera and microphone. These inputs are converted into suitable formats (.png for images and .wav for audio) and passed to the preprocessing stage. In preprocessing, images are converted to grayscale and enhanced, while audio signals are denoised to improve quality. The processed data is then sent to the recognition module, where object detection and text extraction are performed. The recognized information is converted into meaningful textual data representing the surroundings. This text is then passed to the transformation stage, where it is converted into speech output. An evaluation step is included to refine the generated output, such as correcting errors in text. Finally, the system delivers clear audio feedback to the user. This pipeline ensures real-time assistance for visually impaired individuals.

The current implementation of the project has been successfully developed and tested on a laptop environment. The system uses camera and microphone inputs to process real-time data. The vision module employs YOLOv8 to detect objects and identify documents within the surroundings. For text recognition, PaddleOCR is used to extract textual information from captured frames, which is then converted into speech output. A colour detection module based on predefined HSV ranges is also implemented to identify dominant object colours. Additionally, a memory feature using the EfficientNet-Lite0 model has been developed to store and match visual embeddings, enabling recognition of previously seen objects.

Result and Conclusion:

Figure 2. Real-Time Color Identification

The system takes a live camera input where an object (perfume bottle) is detected using YOLO, along with a person in the frame. Simultaneously, color detection identifies the dominant color (blue) of the object and displays it as output as shown in figure 2.

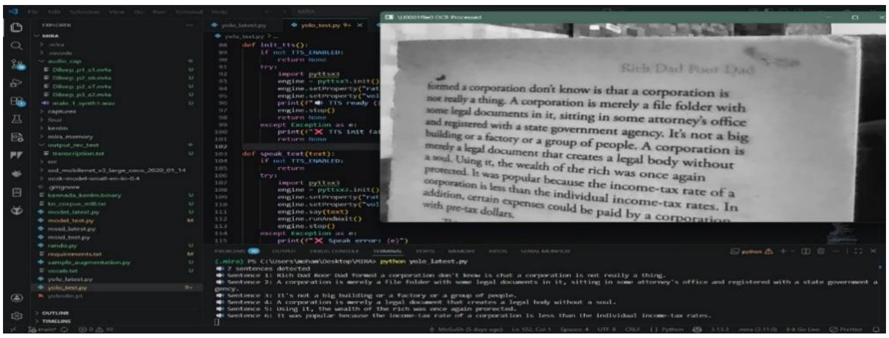

Figure 3. Text Extraction and Speech Output

The system captures an image of printed text (book page) and processes it using OCR to extract readable sentences. The extracted text is displayed and converted into speech output, enabling auditory assistance for visually impaired users as shown in figure 3.

The system demonstrates a low-cost wearable assistive solution that integrates OCR, colour detection, and memory-based object recognition, operating efficiently on a Raspberry Pi 5 with near real-time performance. It functions offline, making it practical for real-world deployment, and the current progress includes a fully functional prototype with a complete vision module codebase.

References:

[1] W. Wang *et al.*, "From object detection to text detection and recognition: A survey," *IEEE Access*, pp. 4–6.
[2] S. Subramani *et al.*, "A survey of deep learning approaches for OCR and document understanding," in *Proc. Int. Conf.*, pp. 10–15.
[3] S. Sinha, "Digitization of document and information extraction using OCR," *Int. J. Res. Appl. Sci.*, p. 5.
[4] S. Long, J. Ruan, W. Zhang, X. He, W. Wu, and C. Yao, "TextSnake: A flexible representation for detecting text of arbitrary shapes," in *Proc. Eur. Conf. Comput. Vis.*, pp. 3–4**.**
[5] B. Shi, X. Bai, and S. Belongie, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, p. 1.
[6] D. Garcia-Lamont *et al.*, "Segmentation of images by color features: A survey," *Pattern Recognition*, pp. 3–7.
[7] M. Bambhania and R. Patel, "Color-based object tracking with OpenCV: A survey," *Int. J. Eng. Res.*, p. 4.
[8] Y. Che *et al.*, "Speaking without vocal folds using a machine-learning-assisted wearable system," *Nature Communications*, pp. 5–6.
[9] M. Herold *et al.*, "Accelerometer-based real-time voice activity detection from neck surface vibrations," *IEEE Sensors Journal*, p. 4.
[10] F. Cesarini *et al.*, "Machine learning-based study of dysphonic voices," *Biomedical Signal Processing*, p. 6.
[11] J. Konczak *et al.*, "Wearable neuromodulation device for spasmodic dysphonia," *Clinical Trial Reports*.
[12] A. Tsanas *et al.*, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, p. 1.
[13] J. Orozco-Arroyave *et al.*, "Characterization methods for the detection of multiple voice disorders," *IEEE J. Biomed. Health Inform.*, p. 3.
[14] P. Little *et al.*, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, p. 2.

Future Scope: The future scope of this project focuses on improving both performance and usability. The object detection module can be enhanced using advanced deep learning models to achieve higher accuracy and detect a wider range of objects in real-world environments. Similarly, the OCR system can be improved to handle complex conditions such as poor lighting, varied text orientations, and multilingual text. Alongside software advancements, the hardware can be miniaturized into a compact and lightweight wearable form using efficient embedded systems. Optimizing power consumption and battery life will further enhance practicality. These improvements will make the system more reliable, efficient, and suitable for everyday use by visually impaired individuals.