JSS MAHAVIDYAPEETHA

JSS Science and Technology University

"AI POWERED VISION ASSISTANCE SYSTEM"

A technical project report submitted in partial fulfillment of the award of the degree of

BACHELOR OF ENGINEERING

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

BY

Gururaj Deepak Todurkar	**01JST20EC033**
B P Rashmi	**01JST20EC014**
Vadiraj B Hayagreeva	**01JST20EC111**
Vishakantamurthy S	**01JST21EC404**

Under the guidance of

Prof. Halesh M R
Assistant professor
Department of Electronics and Communication Engineering
SJCE, JSS S&TU, Mysuru.

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
2023-2024

JSS MAHAVIDYAPEETHA

JSS Science and Technology University

CERTIFICATE

This is to certify that the work entitled “AI powered vision assistance system” is a Bonafide work carried out by Gururaj Deepak Todurkar, B P Rashmi, Vadiraj B Hayagreeva, Vishakantamurthy S in partial fulfilment of the award of the degree of Bachelor of Engineering in Electronics and Communication Engineering for the award of Bachelor of Engineering by JSS Science and Technology University, Mysuru, during the year 2023-2024. The project report has been approved as it satisfies the academic requirements in respect to project work prescribed for the Bachelor of Engineering degree in Electronics and Communication Engineering.

Under the guidance of

Prof. Halesh M R
Assistant Professor,
Department of ECE, SJCE,
JSS STU, Mysuru.

Head of the Department

Dr. U.B. Mahadevaswamy
Professor and Head,
Department of ECE, SJCE,
JSS STU, Mysuru.

Examiners:

1. ....

2. ....

vision assistance

by Gururaj T

Submission date: 27-May-2024 11:47AM (UTC+0530)
Submission ID: 2389037067
File name: vision_assistance.pdf (4.32M)
Word count: 6821
Character count: 40316

vision assistance

ORIGINALITY REPORT

11%	6%	9%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Sohal Islam, Muhammad Jakaria Rahimi, Kazi Arham Kabir. "The Third Tab (3T): Assistant for Visually Impaired People : ATmega based automated blind cane", 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019 Publication	2%
2	www.researchgate.net Internet Source	1%
3	www.ijeast.com Internet Source	1%
4	link.springer.com Internet Source	1%
5	www.snoqap.com Internet Source	1%
6	Bipendra Basnyat, Neha Singh, Nirmalya Roy, Aryya Gangopadhyay. "Vision Powered Conversational AI for Easy Human Dialogue Systems", 2020 IEEE 17th International	<1%

Conference on Mobile Ad Hoc and Sensor Systems (MASS), 2020
Publication

7	Wan-Jung Chang, Liang-Bi Chen, Cheng-You Sie, Ching-Hsiang Yang. "An Artificial Intelligence Edge Computing-Based Assistive System for Visually Impaired Pedestrian Safety at Zebra Crossings", IEEE Transactions on Consumer Electronics, 2021 Publication	<1 %
8	"IEEE Transactions on Consumer Electronics - Front cover", IEEE Transactions on Consumer Electronics, 2017 Publication	<1 %
9	utpedia.utp.edu.my Internet Source	<1 %
10	Submitted to Yeditepe University Student Paper	<1 %
11	researchr.org Internet Source	<1 %
12	www.sciencegate.app Internet Source	<1 %
13	Wan-Jung Chang, Liang-Bi Chen, Ming-Che Chen, Jian-Ping Su, Cheng-You Sie, Ching-Hsiang Yang. "Design and Implementation of an Intelligent Assistive System for Visually Impaired People for Aerial Obstacle	<1 %

Avoidance and Fall Detection", IEEE Sensors Journal, 2020
Publication

14	Submitted to Liverpool John Moores University Student Paper	<1 %
15	Submitted to INTI Universal Holdings SDM BHD Student Paper	<1 %
16	Qian Zhang, Xiaopei Cai, Yilin Zhong, Xueyang Tang, Tao Wang. "Dynamic response prediction of high-speed train on cable-stayed bridge based on genetic algorithm and fused neural networks", Engineering Structures, 2024 Publication	<1 %
17	Submitted to Jose Rizal Memorial State University - Main Student Paper	<1 %
18	Minjoon Kim, Junwon Mun, Jaeseok Kim. "Design of 270 Mbps 2×2 MIMO-OFDM system for video streaming in internet of things", 2015 International SoC Design Conference (ISOCC), 2015 Publication	<1 %
19	arxiv.org Internet Source	<1 %

20	www.adafruit.com Internet Source	<1 %
21	Hafeez Ali A., Sanjeev U. Rao, Swaroop Ranganath, T. S. Ashwin, Guddeti Ram Mohana Reddy. "A Google Glass Based Real-Time Scene Analysis for the Visually Impaired", IEEE Access, 2021 Publication	<1 %
22	Submitted to University of Sheffield Student Paper	<1 %
23	civil.uonbi.ac.ke Internet Source	<1 %
24	eurchembull.com Internet Source	<1 %
25	www.mindmeister.com Internet Source	<1 %
26	Huizi Mao, Song Yao, Tianqi Tang, Boxun Li, Jun Yao, Yu Wang. "Towards Real-Time Object Detection on Embedded Systems", IEEE Transactions on Emerging Topics in Computing, 2016 Publication	<1 %
27	Jeonghun Lee, Kwang-il Hwang. "YOLO with adaptive frame control for real-time object detection applications", Multimedia Tools and Applications, 2021	<1 %

Publication

28 Elvaretta Dian Detiana Yucky, Aji Gautama Putrada, Maman Abdurohman. "IoT Drone Camera for a Paddy Crop Health Detector with RGB Comparison", 2021 9th International Conference on Information and Communication Technology (ICoICT), 2021 <1%

Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

vision assistance

GRADEMARK REPORT

FINAL GRADE

/100

GENERAL COMMENTS

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

DECLARATION

We do hereby declare that the project titled “AI powered vision assistance system” is carried out by the project group, under the guidance of Prof. Halesh M R, Assistant Professor, Department of Electronics and Communication Engineering, JSS Science and Technology University, Mysuru, in partial fulfilment of requirement for the award of Bachelor of Engineering by JSS Science and Technology University, Mysore, during the year 2023-2024.We also declare that we have not submitted this dissertation to any other university for the award of any degree or diploma course.

Date:

Place: Mysore

Gururaj Deepak Todurkar
B P Rashmi
Vadiraj B Hayagreeva
Vishakantamurthy S

ACKNOWLEDGEMENT

An endeavor is successful only when it is carried out under proper guidance and blessings. We would like to thank a few people who helped us in carrying out this work by lending invaluable assistance. We express our sincere regards to His Holiness Jagadguru Sri Shivaratri Deshikendra Mahaswamiji who has showered his blessings on us for framing our career successfully. We here by thank Dr. C Nataraju, Principal, SJCE, JSSTU, Mysuru and Dr. U B Mahadevaswamy , HOD, Department of Electronics and Communication Engineering, SJCE, JSSTU, Mysuru who encouraged us at this venture. It is our foremost duty to thank our project supervisor Prof. Halesh M R for his encouragement, effective guidance, and valuable suggestions right from the beginning of this project till its completion. We thank panel members for their support and guidance throughout the project. We also extend our regards to all the teaching and non-teaching members of the Department of Electronics and Communication Engineering for their direct or indirect support towards the completion of this project. We would also like to thank our family and friends for their constant support.

ABSTRACT

This project presents an advanced AI-powered vision assistance system designed to aid visually impaired individuals by providing real-time scene understanding and object recognition. The system initiates its process by capturing images upon user activation and transmits these images to a server for processing. The preprocessing of images involves techniques such as thresholding and edge detection to enhance image quality and highlight significant features. Following this, the images are analyzed through two distinct pathways: one for object detection and the other for optical character recognition (OCR). The system employs the EasyOCR model for OCR tasks, achieving high accuracy rates in text recognition across multiple languages, including Kannada, and is capable of handling handwritten text with moderate accuracy. For object detection and image captioning, the BLIP model is utilized, generating relevant and descriptive captions. These outputs are then refined using a natural language processing (NLP) component, which constructs coherent and meaningful sentences. The integration of the LLama3 model ensures that the final text descriptions are clear, accurate, and free from grammatical errors, significantly improving the user experience. The system also features multilingual support, making it accessible to a broader audience. User feedback has highlighted the system's high accuracy in object and text recognition and the clarity of auditory feedback, contributing to a positive overall experience. Despite its successes, the system faces challenges in complex environments such as cluttered or low-light settings, which will be addressed in future improvements. Additionally, future work will focus on expanding language support and integrating further assistive technologies to enhance the system’s functionality and user experience. This AI-powered vision assistance system represents a significant advancement in providing real-time, accurate assistance to visually impaired users, enhancing their independence and interaction with their environment. The project underscores the potential of integrating advanced AI and computer vision technologies to create accessible solutions that significantly improve the quality of life for visually impaired individuals.

Keywords: AI-powered vision assistance, Optical character recognition (OCR), Natural language processing (NLP), Real-time scene understanding, Object detection, Multilingual support

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Preamble

The advancement of AI-powered vision assistance technologies has revolutionized accessibility for visually impaired individuals, offering innovative solutions that significantly enhance their quality of life. This project aims to integrate and build upon cutting-edge research to develop a comprehensive vision assistance system that leverages machine learning. Drawing inspiration from successful implementations such as "Med Glasses" for drug identification and smart guiding devices for indoor navigation, our project seeks to create a versatile wearable solution. This system will utilize high-resolution image sensors, real-time object recognition, and text extraction to provide users with immediate, actionable insights about their surroundings. By incorporating edge machine learning for efficient data processing and leveraging affordable, camera module, our goal is to deliver a cost-effective and highly reliable assistance tool. Moreover, the project will address challenges related to privacy, accuracy, and user-friendly design to ensure broad usability and acceptance. Ultimately, this initiative aims to empower visually impaired individuals, facilitating safer navigation, better medication management, and enhanced independence in daily activities through advanced AI and innovative hardware integration.

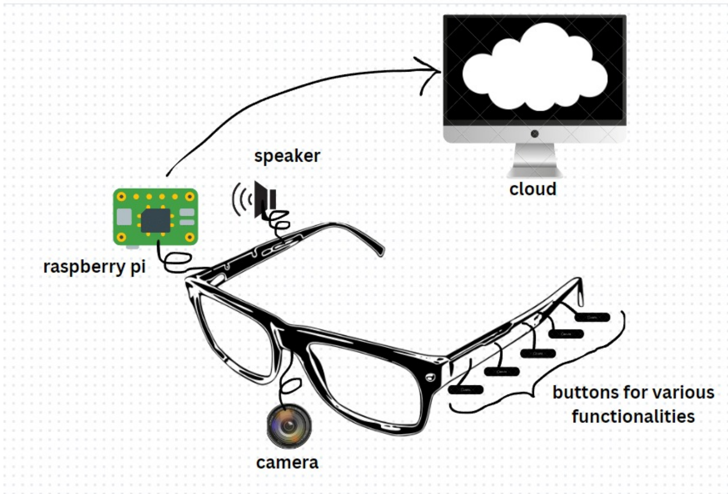

Figure 1.1: Prototype of the project

1.2 Vision assistance system

A vision assistance system using AI refers to a technology that utilizes artificial intelligence (AI) algorithms to assist individuals, particularly those with visual impairments, in perceiving and understanding their surroundings. This system typically incorporates cameras or other image sensors to capture visual information, which is then analyzed and processed by AI algorithms to provide relevant feedback or assistance. The AI algorithms can perform tasks such as object recognition, text extraction, scene understanding, and gesture recognition, enabling users to navigate their environment, recognize objects and text, and interact with their surroundings more independently. Overall, a vision assistance system using AI aims to enhance accessibility, improve quality of life, and promote independence for individuals with visual impairments or other related challenges.

1.3 Need for Vision assistance system

AI-powered vision assistance systems address the essential need for inclusivity and independence among individuals with visual impairments. By employing advanced machine learning algorithms, these systems interpret visual information, offering real-time assistance such as object recognition and navigation guidance. Traditional assistive technologies often fall short in providing comprehensive support, highlighting the necessity for more sophisticated solutions. AI's increasing prevalence and affordable hardware like Raspberry Pi make these systems accessible to a broader user base. Ultimately, AI-powered vision assistance systems have the potential to significantly enhance the quality of life for individuals with visual impairments, fostering greater participation and autonomy in society.

1.4 Motivation

The motivation for an AI-powered vision assistant system is rooted in improving accessibility and feasibility in vision Assistance. AI-driven vision assistance systems offer real-time object recognition, text extraction, and scene understanding, enhancing independence and accessibility for individuals with visual challenges, ultimately fostering a more inclusive society. By combining artificial intelligence and computer vision, the system enhances user experience, aids individuals with disabilities, and finds applications in education, healthcare, and security, contributing to a more inclusive, efficient, and innovative future.

1.5 Problem Statement

Developing an AI-powered vision assistance system aims to empower visually impaired individuals by providing real-time object recognition, text extraction, and scene understanding. Current solutions lack comprehensive features for independent mobility and object recognition. Through audio or haptic feedback, the goal is to empower visually impaired individuals, enhancing their overall mobility and independence. This project seeks to bridge the existing gap and significantly improve the daily lives of individuals with visual impairments.

1.6 Block diagram

The block diagram of the proposed project is given in 1.2. The block diagram shows a system that converts images to text using a camera, Raspberry Pi, and server. The camera captures images and sends them to the Raspberry Pi. The Raspberry Pi can perform either image captioning or OCR (Optical Character Recognition) depending on the user's selection. The result is then sent to the server, which converts the text into speech using Google's natural language processing technology, LLAMA3. Finally, the converted speech is sent back as an audio file.

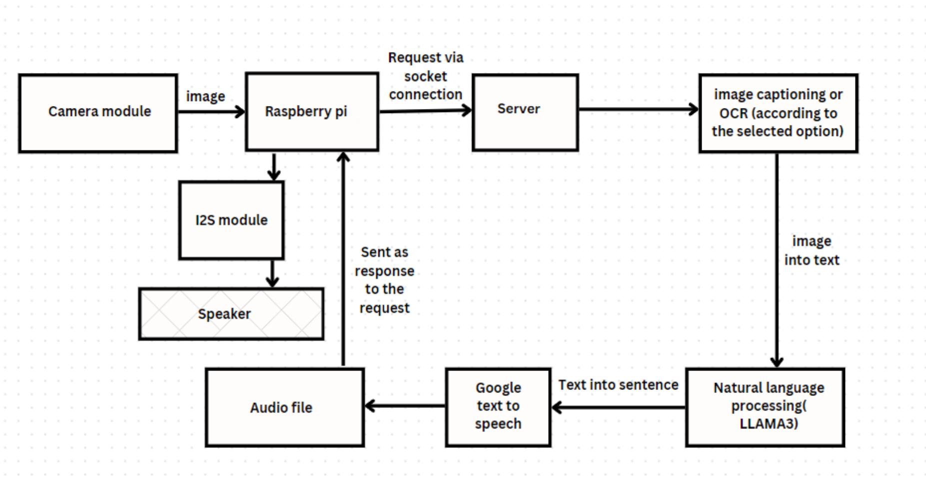

Figure 1.2: Generic block diagram

1.7 Objectives

The objectives for proposed project are as given below:

1. Integrating the camera module with the microcontroller.

2. To develop machine learning models specifically tailored for achieving the Object identification.
3. To develop machine learning models specifically tailored for achieving the Optical character recognition.

1.8 Organisation of the report

To help readers better understand the realized work, the report's seven chapters can be explained as follows:

Chapter 1 of this project describes the main elements, such as the introduction to vision assistance system and its needs, problem statement, Motivation and objectives of the proposed work. Chapter 2 of this project describes the literature review that was conducted to become familiar with the current condition and the project's scope. This chapter also summarizes the value of the practices that are now in use. Chapter 3 of this project offers a full description of the hardware and software requirements for the system that will be most effective for the project and gives the details the system architecture and methodology employed in our project, illustrates how unique the proposed research is, and goes into further detail about the technique and system design and also this chapter describes the implementation approach chosen to achieve the objectives based on the modules listed in the preceding chapter. Chapter 4 of this project tells the study's findings are presented in the results and discussion section, which also analyses how they can be interpreted. It provides a summary of the data gathered, identifies any noteworthy patterns, or trends, and discusses their significance considering the study's goals. Chapter 5 of the project tells the main conclusions and findings of a project are outlined in the conclusion, along with their importance and ramifications. The future scope section suggests areas for additional study or development and offers prospective directions for additional research and development. Chapter 6 of this project gives comprises all the citations and references that assisted us much in completing this project.

Chapter 2

Literature Review

The proposed wearable smart-glasses-based drug pill recognition system, Med Glasses, has successfully attained a notable recognition rate of 95.1 percentage. The researchers improved safety for chronic visually impaired individuals by accurately identifying drugs through the use of deep learning algorithms. Carers can access real-time information through the system, which consists of smart glasses, a cloud-based platform for medication records, a smartphone app, and an AI-driven recognition box. This important development successfully reduces the possibility of medication interactions. Continued work involves improving the system's look, simplifying the circuitry, and working with hospital pharmacies to increase the system's capacity for medication detection. This paper introduces a smart guiding device for visually impaired individuals, facilitating safe navigation in complex indoor environments. The device employs multi-sensor fusion algorithms and depth pictures to tackle obstacles such as translucent and small-sized obstacle avoidance. For users who are completely blind, three auditory cues—specifically, instructions based on beep sounds—have been found to be effective. Augmented reality (AR) improves visual information for visually impaired individuals by including navigable directions into binocular pictures for safer and faster navigation. Fast obstacle detection is made possible by the system's computational efficiency, which also has the added benefit of having inexpensive, easily reusable sensors that may be used by a large number of visually impaired people. This paper addresses video streaming within the context of the Internet of Things (IoT), taking into account the unique requirements and mechanisms associated with IoT devices. In particular, the difficulties posed by the poor processing and memory capacity of smart objects linked to the Internet over subpar wireless channels are taken into account. The H264 AVC standard is examined in detail, with an emphasis on how it supports error resilience. Future directions in both encoding and transmission systems for smart objects are recommended by the discussions. Subsequent research endeavours to investigate, formulate, and evaluate the suggested guidelines to augment video streaming effectiveness in Internet of Things contexts. The study introduces a wearable assistive equipment for those with vision impairments that consists of smart walking cane, smart waist-mounted box, and glasses. The gadget's main goals are to assist people at crosswalks and stop people from straying from zebra crossings. Zebra-crossing detection in real-time and with high accuracy is achieved through the use of deep learning and the Single Shot Multi-Box Detector (SSD) method. The walking cane incorporates vibration motor, GPS, G-sensor, Bluetooth, LPWAN connectivity, and MCU components. For obstacle identification and video recording, smart glasses come equipped with an IR sensor, an image sensor, and a G-sensor. In testing conducted in Tainan, Taiwan, the device achieved a 90percentage detection rate for zebra-crossing, demonstrating its successful recognition capabilities in a variety of settings. The research explores an edge computing architecture for real-time surveillance, utilizing an ESP32-based Wi-Fi camera and a Raspberry Pi 4B as an edge server. PlatformIO along with the ESP-IDF framework are used to programme the ESP32 module. Gun detection is trained using a bespoke TensorFlow model that uses the CocoSSD pre-trained model. In terms of features and cost, the cloud comparison takes into account Google Cloud, AWS, and Microsoft Azure. Data preparation, model training, validation, and deployment are all steps in the deployment process. Cloud connectivity is facilitated by Google

IoT Core, while Firebase uses Cloud Functions to process data. The architecture illustrates possible uses in situations where monitoring and prompt notifications are necessary. The paper addresses the common issue of image noise, which is often caused by factors like inadequate lighting, low-quality cameras, and image compression. It points out that most object recognition techniques and benchmarks focus on high-quality images, while object detection in noisy images is a significant challenge, especially in fields like surveillance. The paper introduces a new low-cost technique for detecting objects in noisy images, leveraging the Single Shot MultiBox Detector (SSD) approach. Extensive experiments are conducted using conventional detectors retrained on noisy images, and the method's performance is evaluated on the Pascal Visual Object Classes benchmark. The results show that the proposed technique performs well in detecting objects in noisy images, making it potentially useful for industries dealing with object detection challenges in low-light and low-quality image scenarios. This paper addresses the challenge of integrating object detection methods into embedded systems with limited computing resources and energy supply. The authors propose a pipelined object detection implementation on an embedded platform, aiming to strike a balance between speed and accuracy. They conduct a thorough analysis of state-of-the-art object detection algorithms and choose Fast R-CNN as a potential solution. Modifications are made to adapt Fast R-CNN for the specific platform, and a multi-stage pipelined implementation is presented for efficient use of limited computational resources. The resulting system is highly energy-efficient and achieves close to real-time performance. In a Low-Power Image Recognition Challenge (LPIRC), their system outperforms others with a specific metric (mAP/Energy) on the Jetson TK1 CPU+GPU platform. This paper addresses the challenge of achieving real-time object detection using the popular You Only Look Once (YOLO) framework on AI embedded systems with resource constraints. While YOLO is known for its ease of use and high object detection precision, it typically requires high-end hardware for real-time detection. The paper discusses the issues related to real-time processing in YOLO with network cameras and proposes a novel YOLO architecture with Adaptive Frame Control (AFC) to efficiently address these problems. Through experiments, the paper demonstrates that AFC can maintain YOLO's precision and convenience while providing real-time object detection by minimizing the total service delay, which is a limitation of the traditional YOLO approach Among many of human disabilities, blindness is a very common and unendurable one. Variety of techniques have been implemented to indicate the path of visually impaired people. To minimize the sufferings of visually impaired people various research works have been carried out in recent years. But still, there are a lot of aspects yet to be explored in this field. Our motive of the research work is also to lessen the sufferings by developing a cheap and easy to use the device to assist the visually impaired persons. As soon as an obstacle is detected, it generates a voice message to warn the blind instantaneously. The device re-charging facility is the most interesting feature of our research work. An arrangement of raw magnets produces mechanical rotation which in turn produces sufficient backup power from a small DC motor. In addition to this, they used GSM module with a panic button for getting the current location of crisis situations. In this paper, they present a design of a wearable equipment that helps with the perception of the environment for blind and visually impaired people in indoor and outdoor mobility and navigation. Their prototype can detect and identify traffic situations such as street crossings, traffic lamps, cars, cyclists, other people and low and high obstacles. The detection takes place in real time based on input data of sensors and optical cameras; the mobility of the user is aided with audio signals.

2.1 Previous research

This paper has attained a notable recognition rate of 95.1%. The researchers improved safety for chronic visually impaired individuals by accurately identifying drugs through the use of deep learning algorithms [1]. Carers can access real-time information through the system, which consists of smart glasses, a cloud-based platform for medication records, a smartphone app, and an AI-driven recognition box. This important development successfully reduces the possibility of medication interactions. Continued work involves improving the system's look, simplifying the circuitry, and working with hospital pharmacies to increase the system's capacity for medication detection. The device employs multi-sensor fusion algorithms and depth pictures to tackle obstacles such as translucent and small-sized obstacle avoidance [2]. Fast obstacle detection is made possible by the system's computational efficiency, which also has the added benefit of having inexpensive, easily reusable sensors that may be used by a large number of visually impaired people. Video streaming within the context of the Internet of Things (IoT), taking into account the unique requirements and mechanisms associated with IoT devices. In particular, the difficulties posed by the poor processing and memory capacity of smart objects linked to the Internet over subpar wireless channels are taken into account [3]. The H264 AVC standard is examined in detail, with an emphasis on how it supports error resilience. Future directions in both encoding and transmission systems for smart objects are recommended by the discussions. Subsequent research endeavours to investigate, formulate, and evaluate the suggested guidelines to augment video streaming effectiveness in Internet of Things contexts. The study introduces a wearable assistive equipment for those with vision impairments that consists of smart walking cane, smart waist-mounted box, and glasses. The gadget's main goals are to assist people at crosswalks and stop people from straying from zebra crossings. Zebra-crossing detection in real-time and with high accuracy is achieved through the use of deep learning and the Single Shot Multi-Box Detector (SSD) method. The walking cane incorporates vibration motor, GPS, G-sensor, Bluetooth, LPWAN connectivity, and MCU components. For obstacle identification and video recording, smart glasses come equipped with an IR sensor, an image sensor, and a G-sensor [4]

The research explores an edge computing architecture for real-time surveillance, utilizing an ESP32-based Wi-Fi camera and a Raspberry Pi 4B as an edge server. Platform IO along with the ESP-IDF framework are used to programme the ESP32 module [5] [6]. This paper addresses the challenge of integrating object detection methods into embedded systems with limited computing resources and energy supply. The authors propose a pipelined object detection implementation on an embedded platform, aiming to strike a balance between speed and accuracy. They conduct a thorough analysis of state-of-the-art object detection algorithms and choose Fast R-CNN as a potential solution. Modifications are made to adapt Fast R-CNN for the specific platform, and a multi-stage pipelined implementation is presented for efficient use of limited computational resources [7] [8] The prevalent issue of blindness by developing an affordable device for visually impaired individuals. It focuses on instant obstacle detection through voice messages, complemented by a unique recharging feature using raw magnets and a small DC motor. The project also integrates a GSM module with a panic button for crisis location retrieval. The research aims to contribute to enhancing technology for the visually impaired, emphasizing affordability, ease of use, and innovative functionalities [9] [10]. The content discusses the challenges faced by visually impaired individuals, such as difficulties in reading, driving, and navigating due to visual impairment or blindness. The text introduces Sense It, an Android-based application designed to assist blind or visually impaired users by providing a text-reading interface. The application utilizes a mobile camera to capture the environment, and the captured image undergoes preprocessing to eliminate noise. Optical character recognition (OCR) is employed to extract text from the image, with object segmentation playing a crucial role in enhancing OCR results. Once the text is extracted, the application converts it into human-understandable audio using speech synthesis. The primary goal is to empower blind individuals to read various textual content, such as newspapers, menu cards, invitations, tickets, and books written in English script. The keywords associated with the project include Optical Character Recognition (OCR), Android, Cloud Vision API, Text-to-Speech (TTS) conversion, and Computer Vision (CV). The application's innovative approach leverages technology to improve accessibility and independence for individuals with visual impairments [11]. The paper introduces an end-to-end goal-oriented conversational AI agent, named Flood Bot, designed to offer contextual information from potential hazard sites. This Flood Bot is equipped with the ability to visually assess hazard conditions, utilizing state-of-the-art deep learning models, particularly in computer vision and natural language processing. The study outlines the domain-specific design of Flood Bot and shares insights gained from its real-time deployment in a flash flood-affected city. The agent's capacity to perceive, evaluate, and engage in conversations about hazard conditions is highlighted as a key feature [12].

In the realm of text recognition, Manwatkar and Yadav (2015) investigated text recognition from images, addressing the broader challenge of accessing textual information [13]. Zhang et al. (2010) presented OCRdroid, a framework for digitizing text using mobile phones [14]. Other applications, such as those by Ma et al. (2011) and Sonth et al. (2017), have explored mobile camera-based text detection and translation [15] and OCR-based facilitators for the visually challenged [16], respectively.The exploration of character recognition systems for Android smartphones by Kwon et al. (2012) and surveys of OCR applications by Singh et al. (2012) further highlight the ongoing efforts to make technology accessible to visually impaired individuals [17] [20]. The literature also reflects a commitment to linguistic diversity, as seen in Rajan et al.'s (2016) work on Android applications for optical character recognition specifically tailored to the Malayalam language [22].

2.2 Summary of Literature Review

Paper Name	Author(s)	Technology Used	Advantages	Disadvantages
MedGlasses: A Wearable Smart-Glasses-Based Drug Pill Recognition	W. -J. Chang, L. -B. Chen, C. -H. Hsu, J. -H. Chen, T. -C. Yang and C. -P. Lin	AIoT, Deep learning	Improves medication safety for visually impaired with high recognition accuracy.	Complex design and circuitry require further simplification and refinement.
Smart Guiding Glasses for Visually Impaired People in Indoor Environment	J. Bai, S. Lian, Z. Liu, K. Wang and D. Liu	Depth sensor	Utilizes multi-sensor fusion and AR for efficient, safe indoor navigation.	Relies on auditory cues which might not suit all users.
Video Streaming Considerations for Internet of Things	R. Pereira and E. G. Pereira	HTTP, TCP, LoRA	Enhances error resilience in IoT video streaming using H264 AVC standard.	Limited by IoT devices' poor processing and memory capacity.
An AI Edge Computing Based W earable Assistive Device for Visually Impaired People Zebra-Crossing Walking	W. -J. Chang et al.	Edge Computing	Achieves real-time zebra-crossing detection with 90% accuracy using deep learning and SSD.	Relies on multiple components, potentially increasing complexity and maintenance.
YOLO with adaptive frame control for real-time object detection applications	Lee, Jeonghun & Hwang, Kwang-il	YOLO architecture with Adaptive Frame Contro	Maintains YOLO precision with real-time detection using Adaptive Frame Control.	Still face limitations on extremely resource-constrained AI embedded systems
AI Powered Vision Assistance system	Gururaj, Rashmi, Vadiraj, Vishakanthamurthy	OCR, Cloud Computing	High resolution image capturing	Not specified right now

Table 2.1: Summary of Papers

2.3 Summary of the chapter

The literature survey explores technological advancements aiding visually impaired individuals through deep learning, wearable devices, and edge computing. A system achieving a 95.1% recognition rate uses deep learning to accurately identify drugs, enhancing safety by integrating smart glasses, a cloud platform, a smartphone app, and an AI-driven recognition box, with future improvements focusing on design enhancements, simplified circuitry, and expanded medication detection. Another notable development employs multi-sensor fusion and depth images for obstacle avoidance, ensuring computational efficiency and cost-effectiveness. Real-time video streaming within the IoT context addresses the limitations of smart objects' processing and memory capacities over subpar wireless channels, focusing on the H264 AVC standard's error resilience. A wearable assistive device comprising a smart walking cane, a waist-mounted box, and smart glasses aims to aid at crosswalks and prevent straying from zebra crossings, using deep learning and the SSD method for high-accuracy detection. The cane includes various sensors and connectivity features, while the glasses facilitate obstacle identification and video recording. Edge computing for real-time surveillance uses an ESP32-based Wi-Fi camera and Raspberry Pi 4B, with a pipelined object detection implementation balancing speed and accuracy on embedded platforms with limited resources, using a modified Fast R-CNN.

Addressing blindness, an affordable device offers instant obstacle detection through voice messages and features a unique recharging system and GSM module for crisis location retrieval. The Sense It application assists users by reading text captured by a mobile camera, processed using OCR, and converted to audio via speech synthesis. Flood Bot, an AI agent, provides hazard site information using deep learning for computer vision and natural language processing, highlighting its real-time deployment in flood-affected areas. Overall, these innovations significantly enhance accessibility for visually impaired individuals, showcasing substantial strides in leveraging technology to improve their independence and safety.

Chapter 3

Present work carried out

The operation of the vision assistance system begins with the user pressing a predetermined button to activate the device. This operation starts the image capture using the Raspberry Pi camera, which is connected to the Pi via the MIPI Camera Serial Interface (CSI) protocol, ensuring efficient communication and high-quality image capture. Once recorded, the image is sent to the server via TCP (Transmission Control Protocol) sockets, ensuring reliable data transfer. When the server receives the image, it first preprocesses it, using techniques like thresholding or binarization at both the global and local levels to improve image quality. Following that, edge detection methods are used to emphasize the edges of the image, which is important for later analysis.

The preprocessed image then takes two different processing pathways. One approach is object detection, in which powerful deep learning models identify and classify items in a picture. The second method includes Optical Character Recognition (OCR) with the EasyOCR model, which extracts text from images. In addition, for picture captioning, the BLIP model with fine-tuning capabilities creates descriptive captions for the scene shown.The outputs of the object identification and OCR models are then processed by the NLP model, which analyzes the keywords and generates sentences. These phrases, which contain refined and cohesive information about the scene, are

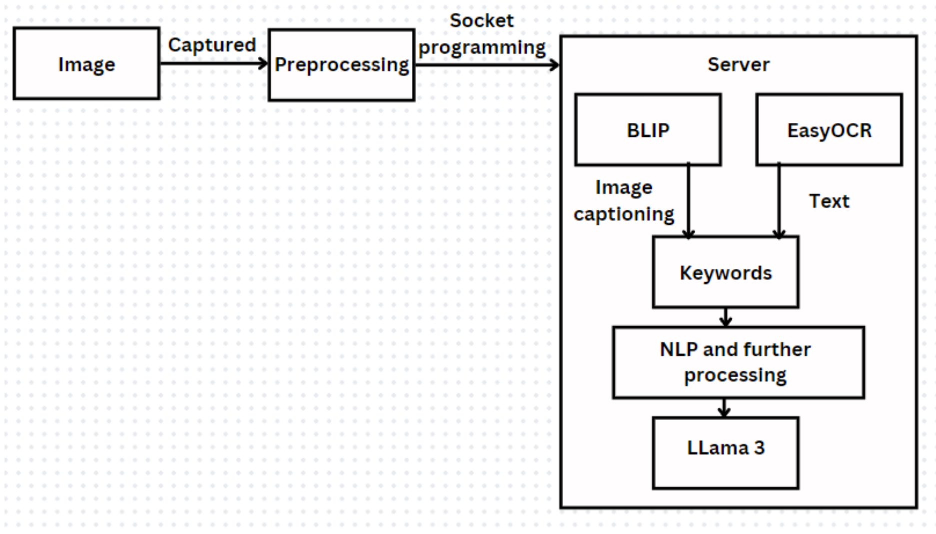

Figure 3.1: Block diagram of proposed project

further improved with the LLama3 model, which ensures that the language is clear, correct, and error-free.

Finally, the improved textual information is sent to the user via a speaker, offering a thorough and clear explanation of the scene, improving the user's ability to navigate and comprehend their surroundings efficiently.

EasyOCR

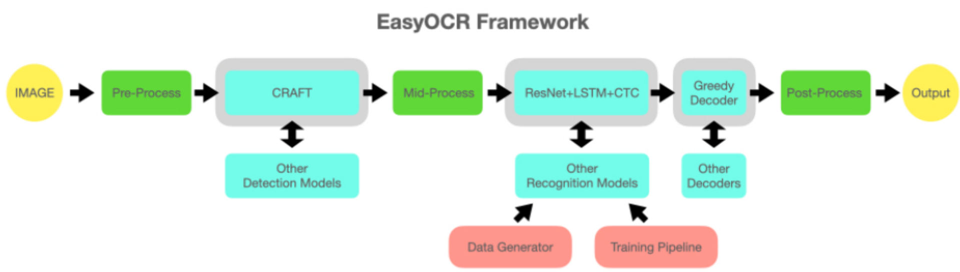

Figure 3.2: Block diagram of EasyOCR

The EasyOCR workflow for text extraction begins with text detection using

the CRAFT model, which employs Convolutional Neural Networks (CNNs) to identify text regions within an image and generates bounding boxes around these areas. Once the text regions are detected, they are cropped from the image and processed by the CRNN model. The CRNN model first uses CNN layers to extract features from the cropped images, then applies Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, to understand the sequence of characters. Finally, Connectionist Temporal Classification (CTC) is used to transcribe the recognized text. Thus, CRAFT is responsible for detecting and linking character regions, while CRNN handles the feature extraction, sequence modeling, and transcription of the detected text.
Here is a comparison table given below in 3.3 which tells EasyOCR is better.

Feature	EasyOCR	Tesseract OCR
Language Support	Supports over 100 languages	Supports over 100 languages
Accuracy	95-99% accuracy for common fonts	85-95% accuracy for common fonts
Performance	Fast processing speed	Processing speed can vary based on setup
Handwritten Text	Can handle handwritten text with moderate accuracy	Not well-suited for handwritten text

Figure 3.3: Comparison between EasyOCR and other technology

Blip model

The BLIP (Bootstrapping Language-Image Pre-training) model block diagram outlines the process of generating descriptive captions for images. It starts with an input image fed into a vision encoder, typically a Convolutional Neural Network (CNN), which extracts image features. These features are then combined with a language model, such as a Transformer, that processes textual data. The combined model uses fine-tuning to align visual and textual modalities, enabling it to generate coherent and contextually relevant

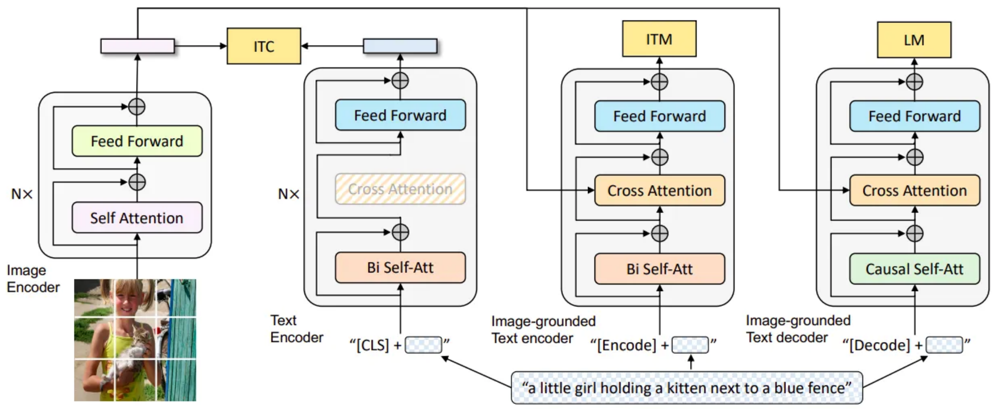

Figure 3.4: BLIP model

captions. The final output is a meaningful text description of the image, refined through iterative training to enhance accuracy and relevance.

Image Pre-processing

Binarisation

To achieve image binarization, we utilized Python with two popular libraries: OpenCV and PIL. The process involved loading the image in grayscale to simplify the data to a single channel. For OpenCV, we applied a binary thresholding technique where each pixel's value was compared to a defined threshold; pixels above the threshold were set to white (255), and those below were set to black (0). For PIL, we similarly converted the image to grayscale and applied a lambda function to map pixel values based on the threshold, producing the binary result. The binarized images were then saved and displayed for comparison.

The flow of the working process is given in 3.5

The process starts with initializing the server and waiting for a client connection. Once a client connects, the connection is accepted. The server then reads the current image number from 'num.txt' and receives the image data from the client. The received image is saved using the current image number,

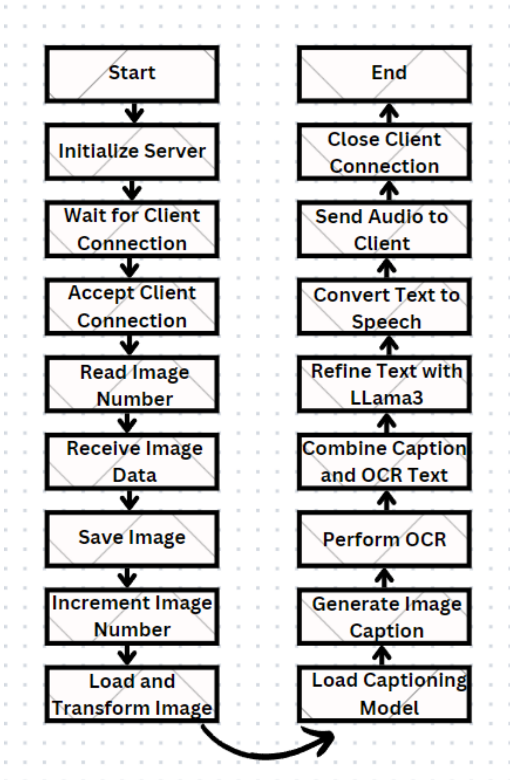

Figure 3.5: Flow chart of the project

and the image number is incremented and updated in 'num.txt'. The saved image is then loaded and transformed. The BLIP model is loaded to generate a caption for the image, followed by performing OCR to extract text from the image. The generated image caption and OCR text are combined, and LLama3 is used to refine this combined text. The refined text is converted to speech, and the generated audio data is sent back to the client. Finally, the client connection is closed, and the process repeats for the next client.

3.1 Hardware requirements

The hardware requirements specification, which includes all the specifications required for the project's development are listed below:

3.1.1 Raspberry Pi Zero 2W

Figure 3.6: Raspberry pi zero 2W

The Raspberry Pi Zero 2 W is a compact and affordable single-board computer with wireless connectivity. It's ideal for IoT projects, embedded systems, and DIY electronics. With its small form factor, low power consumption, and GPIO pins, it's perfect for tasks like automation, sensing, and small-scale computing projects.

3.1.2 8 ohm speaker

Figure 3.7: 8 ohm speaker

An 8-ohm speaker is a common type of audio transducer used in various electronic devices. Its impedance of 8 ohms affects the electrical resistance it presents to an audio signal, influencing its performance and compatibility. This specification is crucial for ensuring optimal sound quality and efficient power delivery in audio systems.

3.1.3 Raspberry pi camera 5 mp

Figure 3.8: Raspberry pi camera 5 mp

The Raspberry Pi camera module, with a 5-megapixel resolution, offers high-quality imaging capabilities for Raspberry Pi projects. Its compact design and compatibility with the Raspberry Pi ecosystem make it ideal for various applications, including surveillance, robotics, and IoT projects, providing crisp and clear image capture in a wide range of scenarios.

3.1.4 Adafruit I2S 3W Class D Amplifier

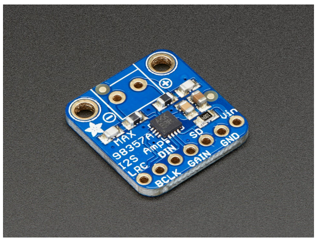

Figure 3.9: I2S module

The I2S 3W Class D Amplifier Breakout is a compact audio module that interfaces with digital audio sources via I2S protocol. It delivers 3 watts of power efficiently, ideal for small speaker systems. With its breakout form factor, it simplifies integration into projects requiring high-quality audio amplification in a space-saving design.

3.1.5 Rechargeable battery

Integrating rechargeable batteries into AI projects ensures uninterrupted operation, enhancing mobility and reliability. It enables sustained performance and extended usage, crucial for on-the-go applications and remote deployments.

3.1.6 TP4056 charging module

Figure 3.10: TP4056 charging module

The TP4056 charging module efficiently charges lithium-ion batteries with its integrated charging and protection features. With precise voltage and current regulation, it ensures safe and reliable charging. Its compact design and ease of use make it ideal for portable electronics and DIY projects requiring battery power.

3.1.7 Push buttons

Push buttons are simple yet essential components in electronics, initiating actions upon activation. They consist of a switch mechanism that completes an electrical circuit when pressed. Widely used in control panels, keyboards, and electronic devices, they provide tactile feedback and enable user interaction, facilitating seamless operation in various applications.

Figure 3.11: Push buttons

3.2 Software requirements

The software requirements specification, which includes all the specifications required for the project's development are listed below:

3.2.1 Python

Python's simplicity, versatility, and rich libraries make it indispensable. Its readable syntax aids efficiency and accessibility. With TensorFlow and PyTorch, it excels in machine learning. Its strong community fosters collaboration. Cross-platform compatibility ensures seamless deployment. Python empowers developers to build robust, scalable solutions efficiently, a cornerstone of modern software development.

3.2.2 PyTorch

PyTorch is vital for deep learning, providing a flexible framework for neural networks. Its dynamic computational graph system aids seamless model creation, crucial for research. With GPU acceleration, it enables rapid training for complex tasks like image recognition and NLP. PyTorch's rich ecosystem simplifies implementation, empowering innovation in AI across domains.

3.2.3 EasyOCR

EasyOCR offers seamless integration and efficient text extraction from images in multiple languages. With pre-trained models and a user-friendly API, it simplifies OCR tasks, accessible to developers of varying expertise. Its versatility enables rapid deployment in diverse applications, enhancing productivity and facilitating innovative solutions in document digitization, image analysis, and augmented reality.

3.2.4 LLama3

LLama3 is a sophisticated large language model developed by OpenAI, designed to advance natural language understanding and generation. Building on the capabilities of its predecessors, LLama3 incorporates cutting-edge deep learning techniques to achieve higher accuracy and efficiency in various language tasks, including text completion, translation, summarization, and conversational AI. It is trained on a vast corpus of data, enabling it to understand and generate human-like text with context awareness and coherence. LLama3 aims to provide enhanced AI-driven solutions across diverse applications, from customer support to content creation, while addressing ethical considerations such as bias and responsible use.

3.2.5 PIL

PIL (Python Imaging Library), also known as Pillow, is essential for image processing tasks in Python. It offers a wide range of functionalities, including image opening, manipulation, and saving in various file formats. With PIL, developers can perform tasks such as resizing, cropping, filtering, and enhancing images, making it a versatile tool for both basic and advanced image processing needs. PIL's ease of use and extensive documentation make it a go-to library for handling images efficiently within Python applications.

3.2.6 Socket programming

Socket programming in Python enables real-time data exchange between networked devices, facilitating client-server applications. Its versatility supports tasks like data transmission, distributed systems, remote communication, and peer-to-peer networking. With socket programming, developers build robust network applications for file transfers, messaging, remote command execution, and beyond, essential in network programming.

3.3 Novelty of the work

The addition of the LLama3 model to the vision assistance system represents a big step forward in improving communication capabilities. LLama3's powerful natural language processing (NLP) capabilities enable the system to deliver scene descriptions with unparalleled clarity and coherence. Using LLama3's cutting-edge linguistic analysis, the system provides accurate and error-free text-to-speech conversion, removing spelling and grammatical faults that could impede user comprehension. This revolutionary feature not only improves the overall user experience, but it also demonstrates the system's commitment to diversity and accessibility. Furthermore, the addition of Kannada language support to the optical character recognition (OCR) functionality broadens the system's linguistic range, allowing it to correctly recognize and process Kannada text. This new feature caters to customers that want or require Kannada language support, which improves the system's functioning and usability. Together, these developments strengthen the vision aid system's ability to provide intuitive, error-free communication adapted to its users' different linguistic demands and preferences, thereby increasing their independence and engagement with the world around them.

3.4 Summary of the chapter

This chapter outlines the development and implementation of a vision assistance system leveraging the BLIP (Bootstrapping Language-Image Pretraining) model for generating descriptive captions for images. The process involves using a CNN for image feature extraction, combined with a Transformer-based language model for textual data processing, resulting in refined and contextually relevant captions. The system hardware includes a Raspberry Pi Zero 2W, 8-ohm speaker, Raspberry Pi 5 MP camera, I2S 3W Class D Amplifier, rechargeable battery, TP4056 charging module, and push buttons. Software requirements encompass Python, PyTorch, EasyOCR, LLama3, PIL, and socket programming. The integration of the advanced LLama3 model enhances natural language understanding and generation, offering improved scene descriptions and error-free text-to-speech conversion, with added support for the Kannada language in OCR tasks, thus broadening accessibility and functionality for users.

Chapter 4

Results and Discussion

4.1 Results

The primary aim of this project was to develop an assistance system for visually impaired individuals, providing them with an aid that converts visual scenes into text and subsequently into speech. This system enables users to hear descriptions of their surroundings, enhancing their ability to navigate and understand their environment. Additionally, the project aimed to implement Optical Character Recognition (OCR) to identify and read out text from various objects, converting it into speech.

To achieve these objectives, we employed a variety of modern technologies and frameworks, including EasyOCR for vertical recognition and the BLIP (Bootstrapped Language-Image Pretraining) model for image captioning and processing. These tools were chosen for their accuracy and efficiency in real-time applications.

Additionally, we fine-tuned the BLIP model with multiple datasets to achieve more accurate results in OCR. This process involved extensive training and validation to ensure the model's robustness in recognizing and interpreting text from various objects.

In terms of practical achievements, the project successfully integrated scene analysis and OCR capabilities into a cohesive system that can provide real-time audio descriptions to visually impaired users. This accomplishment marks a significant step towards improving accessibility and independence for these individuals. The final prototype of the proposed project is as given in 4.1

Figure 4.1: Prototype of proposed project

4.2 Discussion

In developing this assistance system, several additional implementations were incorporated to enhance its functionality. One of the key features added was multi-language support, including support for Kannada. This feature allows the system to recognize and translate text into multiple languages, making it accessible to a broader audience. The multi-language capability was implemented using advanced natural language processing techniques and supported languages such as Kannada, [other languages].

This enhancement significantly improved the user experience by providing accurate and timely information in the user's preferred language. The system's ability to switch between languages seamlessly ensures that users from different linguistic backgrounds can benefit equally from the technology.

In particular, the inclusion of Kannada support makes the system especially useful for Kannada-speaking visually impaired users.

Throughout the project, we encountered several challenges. One major challenge was ensuring real-time processing with minimal latency. To address this, we optimized our algorithms and utilized efficient processing techniques. However, some limitations remain, particularly in varying lighting conditions and complex backgrounds, which can affect the accuracy of both scene analysis and OCR.

Looking forward, there are multiple avenues for future work to enhance this assistance system further. One potential improvement is the integration of advanced deep learning models, such as transformers, to increase the accuracy and speed of scene analysis and OCR. Additionally, expanding the system to include more languages and dialects could greatly benefit a broader range of users. Exploring wearable devices, like smart glasses, could also offer a more seamless and hands-free experience for the visually impaired.

These future enhancements have the potential to significantly improve the system's performance and usability, making it an even more valuable tool for visually impaired individuals.

Chapter 5

Conclusion

5.1 Conclusion

The AI-powered vision assistance system leverages advanced technologies to provide a seamless and effective solution for visually impaired individuals. The process begins with the user triggering a predefined button, prompting the Raspberry Pi camera to capture an image. This image is transmitted to a server via TCP using socket methodology, ensuring reliable and efficient data transfer. On the server, the image undergoes preprocessing, including thresholding or binarization at both global and local levels, followed by edge detection to highlight significant features.

The preprocessed image then diverges into two distinct paths: one for object detection and another for optical character recognition (OCR). For OCR operations, the EasyOCR model is employed to accurately extract text, while the BLIP model, enhanced with fine-tuning capabilities, is used for image captioning to describe the visual content. The outputs from these models are processed by the NLP component, which evaluates keywords and constructs meaningful sentences.

Finally, the refined textual information is further processed by the LLama3 model to ensure error-free, coherent communication. The result is a clear and informative description of the scene, delivered audibly to the user through a speaker. This innovative system not only enhances the independence of visually impaired individuals by providing detailed environmental descriptions but also underscores the potential of integrating advanced AI models to improve accessibility and quality of life.

5.2 Advantages and Disadvantages

5.2.1 Advantages

- Accessibility: With AI-powered vision assistance on Raspberry Pi, visually impaired individuals gain access to a portable and affordable solution for interpreting visual information, enhancing their independence and quality of life.

- Cost-effectiveness: Utilizing Raspberry Pi as the hardware platform offers a cost-effective solution compared to dedicated assistive devices, making AI-powered vision assistance accessible to a wider demographic.

- Real-time Feedback: The system provides immediate feedback on the surrounding environment, enabling users to navigate unfamiliar spaces, recognize objects, and read text in real-time, promoting safety and autonomy.

- Portability: Raspberry Pi's small form factor enables the creation of portable vision assistance devices, facilitating on-the-go use in scenarios like assistive technology for the visually impaired.

5.2.2 Disadvantages

- Delay: As it involves client and server connection establishment in the socket it generates a small delay between the trigger and the response.

- Heat Dissipation: Intensive network computations may cause Raspberry Pi to generate heat, requiring additional cooling mechanisms to prevent overheating and potential damage to the hardware.

By pinpointing bottlenecks, we optimize algorithms, streamline data flow, and leverage hardware acceleration to enhance system efficiency and ensure real-time responsiveness in visual recognition tasks.

5.3 Future scope

Increasing the scope of our AI-powered vision assistance system utilizing Raspberry Pi opens avenues for innovative enhancements. Integrating advanced machine learning models for real-time object detection, scene recognition, and gesture recognition could enhance its functionality. Additionally, incorporating natural language processing capabilities could enable interactive voice commands, expanding accessibility and usability. Further exploration into edge computing optimizations and integration with IoT devices could amplify its applicability in diverse settings, ranging from smart homes to industrial automation. Collaboration with healthcare institutions could lead to the development of assistive technologies for visually impaired individuals, fostering inclusivity and societal impact. Continual refinement and adaptation to emerging technologies ensure our vision assistance system remains at the forefront of AI-powered solutions, driving transformative change in human-computer interaction.

References

[1] W. -J. Chang, L. -B. Chen, C. -H. Hsu, J. -H. Chen, T. -C. Yang and C. -P. Lin, "MedGlasses: A Wearable Smart-Glasses-Based Drug Pill Recognition System Using Deep Learning for Visually Impaired Chronic Patients," in IEEE Access, vol. 8, pp. 17013-17024, 2020, doi: 10.1109/ACCESS.2020.2967400.

[2] J. Bai, S. Lian, Z. Liu, K. Wang and D. Liu, "Smart guiding glasses for visually impaired people in indoor environment," in IEEE Transactions on Consumer Electronics, vol. 63, no. 3, pp. 258-266, August 2017, doi: 10.1109/TCE.2017.014980.

[3] R. Pereira and E. G. Pereira, "Video Streaming Considerations for Internet of Things," 2014 International Conference on Future Internet of Things and Cloud, Barcelona, Spain, 2014, pp. 48-52, doi: 10.1109/FiCloud.2014.18.

[4] W. -J. Chang et al., "An AI Edge Computing Based W earable Assistive Device for Visually Impaired People Zebra-Crossing Walking," 2020 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2020, pp. 1-2, doi: 10.1109/ICCE46568.2020.9043132.

[5] M. L. N, A. E. Rao and M. P. Kalyan, "Real-Time Object Detection with Tensorflow Model Using Edge Computing Architecture," 2022 8th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2022, pp. 01-04, doi: 10.1109/ICSSS54381.2022.9782169.

[6] A. -A. Nayan, J. Saha, K. Raqib Mahmud, A. Kalam Al Azad and M. Golam Kibria, "Detection of Objects from Noisy Images," 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2020, pp. 1-6, doi: 10.1109/STI50764.2020.9350521.

[7] H. Mao, S. Yao, T. Tang, B. Li, J. Yao and Y. Wang, "Towards Real-Time Object Detection on Embedded Systems," in IEEE Transactions on Emerging Topics in Computing, vol. 6, no. 3, pp. 417-431, 1 July-Sept. 2018, doi: 10.1109/TETC.2016.2593643.

[8] Lee, Jeonghun & Hwang, Kwang-il. (2021). YOLO with adaptive frame control for real-time object detection applications. Multimedia Tools and Applications. 81. 10.1007/s11042-021-11480-0 .

[9] S. Islam, M. J. Rahimi and K. Arham Kabir, "The Third Tab (3T): Assistant for Visually Impaired People: ATmega based automated blind cane," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICASERT.2019.8934623.

[10] B. Sveny, G. Kovcs, and Z. T. Kardkovcs, "Blind guide - a virtual eye for guiding indoor and outdoor movement," in 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom), pp. 343–347, Nov 2014.

[11] Gupta, Rahul & Choudhary, Sachi & Sharma, Rashmi & harsh,. (2018). SenseIt: Text Recognition & Text to Speech Conversion for Visually Challenged People using Google Cloud Vision.

[12] B. Basnyat, N. Singh, N. Roy and A. Gangopadhyay, "Vision Powered Conversational AI for Easy Human Dialogue Systems," 2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Delhi, India, 2020, pp. 684-692, doi: 10.1109/MASS50613.2020.00088.

[13] R. Chen, Z. Tian, H. Liu, F. Zhao, S. Zhang and H. Liu, "Construction of a voice driven life assistant system for visually impaired people," 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2018, pp. 87-92, doi: 10.1109/ICAIBD.2018.8396172.

[14] P. M. Manwatkar, S. H. Yadav: —Text recognition from images‖. IEEE Sponsored 2nd International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2015, pp. 1-6.

[15] M. Zhang, A. Joshi, R. Kadmawala, K. Dantu, S. Poduri, and G. S. Sukhatme: —OCRdroid: A framework to digitize text using mobile phones‖. Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng., vol. 35 LNICST, pp. 273–292, 2010.

[16] D. Ma, Q. Lin, T. Zhang: —Mobile Camera Based Text Detection and Translation‖. Stanford.Edu, 2011.

[17] S. Sonth, J. S. Kallimani: "OCR based facilitator for the visually challenged". IEEE International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques. (ICEECCOT), Mysuru, 2017, pp.

[18] B. Sanjana, J. RejinaParvin: —Voice Assisted Text Reading System for Visually Impaired Persons Using TTS Method‖. IOSR Journal of VLSI and Signal Processing (IOSR-JVSP), vol. 6, no. 3, pp. 1–5, 2016.

[19] S. Kwon, H. An, Y. Cho: —Character Recognition System Based On Android Smart Phone‖. International Journal of Modern Engineering Research (IJMER), vol. 2, no. 6, pp. 82–87, 2012.

[20] A. Singh, K. Bacchuwar, A. Bhasin: —A Survey of OCR Applications‖. Int. J. Mach. Learn. Comput., vol. 2, no. 3, pp. 314–318, 2012.

[21] P. A. C, A. K. Rajan, J. Keerthana, P. B. Lubna, N. P. Pavana, A. R. Roshny: —Android Application Based Optical Character Recognition For Malayalam Language‖. IJRCCT, Vol 5, Issue- 3, 2016.

[22] R. Neto, N. Fonseca: —Camera Reading for Blind People‖. Procedia Technol., vol. 16, pp. 1200–1209, 2014.

[23] D. Mulfari, A. Celesti, M. Fazio, M. Villari, A. Puliafito: —Using Google Cloud Vision in assistive technology scenarios‖. Proc. – IEEE Symp. Comput. Commun., vol. 2016–August, pp. 214–219, 2016."