ICEST-2K24 [International Conference on Engineering, Science and Technology]
In Association with
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN - 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Application to Help the Visually Impaired By Converting Images to Audio Descriptions

Mr. Shubham Shejwal, Mr. Abhishek Jadhav, Mrs. Deepa Mahajan, Mr. Abhay Rajput
*Pimpri Chinchwad College of Engineering and Research, Ravet Pune, Maharashtra, India

ABSTRACT
In a world increasingly reliant on visual information of the surroundings, visually disabled people usually face significant challenges in their daily lives even for simple tasks. The absence of real-time, environment-apprehensive audio descriptions of their environment hinder their mobility and therefore they struggle to engage with the world, unlike other humans. This application design aims to develop an innovative AI-powered application to bridge this availability gap and increase their quality of life. The primary functionality of this operation is based on using state-of-the-art image recognition technology to give visually impaired individuals accurate and intuitive audio descriptions of their immediate surroundings. By using artificial intelligence power, this application aims to deliver real-time, detailed, and user- friendly information in audio format about objects and other applicable visual essentials within the user's surroundings.

Keywords:

I. INTRODUCTION

Visually impaired individuals often encounter barriers to independent mobility and participation in daily life due to the lack of accessible tools that provide real-time, context-aware descriptions of their surroundings. Our application tends to revolutionize their interaction with the visual world by converting the images into audio descriptions. Taking advantage of advanced algorithms and user-friendly usage, our application encourages to enrich the standard of living for the visually impaired fostering independence and facilitating an inclusive world.

II. METHODOLOGIES

A. Image Processing

Image processing acts as a virtual seeing-eye canine for the visually impaired, wielding the energy of cameras to bridge the visual gap. This record is then translated into a consumer-friendly layout, be it clean audio descriptions, intuitive vibrations on a hand-held device, or even tactile maps for spatial knowledge. Imagine a visually impaired individual being capable of independently examining store signs and symptoms or perceiving items on cabinets – this era empowers exploration and fosters an experience of independence. As algorithms end up extra sophisticated, destiny promises even richer studies. We can assume real-time scene evaluation describing whole environments, item reputation differentiating between a parked vehicle and a dangerous pothole, and seamless integration with navigation apps for flip-by-means-of-turn guidance. Image processing is revolutionizing the way visually impaired people interact withthe sector, providing a brighter route toward a more independent and enjoyable life.

B. Feature Extraction

In helping visually impaired people with navigation, feature extraction plays a crucial role in transforming raw data into information. Imagine a gadget that could take the visual scene as input and extract the maximum vital information applicable to navigation. This is where characteristic extraction plays an essential position. By making use of algorithms to camera pictures, the device can discover and isolate key elements in the photograph. These capabilities can embody a huge variety, relying on the specific software. Common examples include Obstacle detection: Extracting capabilities like edges, depth variations, and unexpected adjustments in brightness lets the device discover ability limitations like curbs, uneven pavement, or stray objects, helping customers keep away from collisions. Landmark popularity: By analyzing shapes, textures, and coloration patterns, the device can understand landmarks like crosswalks, street signs and symptoms, building entrances, or shops. These records can be relayed via audio descriptions, empowering customers to orient themselves within their surroundings. Object classification: Features like length, form, and color may be used to categorize gadgets of hobby, which include visitors' lights, parked cars, or maybe unique types of vegetation. This can provide valuable context for visually impaired customers, enhancing their information about the environment. The strength of function extraction lies in its ability to distill sizable quantities of visible records right into a concise and meaningful representation. Furthermore, the choice of capabilities extracted may be tailor-made to desires and environments. For instance, navigating an indoor shopping mall would possibly prioritize identifying store signage and product displays, while out-of- door navigation might focus on extracting features associated with visitors' lighting fixtures, pedestrian crossings, and potential risks like uneven terrain. As machine mastering algorithms continue to evolve, function extraction strategies become even more state-of-the-art. We can count on the ability to extract more and more complex capabilities, leading to a richer know-how of the surroundings. Imagine systems that cannot most effectively identify gadgets but additionally parent their capability (a parked automobile vs. A shifting car) or even understand emotional expressions on people's faces. This level of element would in addition beautify situational consciousness and provide precious information for navigating social interactions.

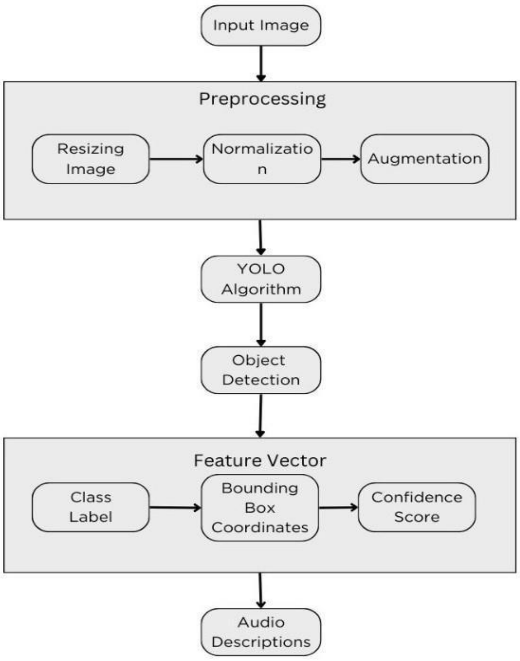

Figure1:Feature Extraction

C. Text-to-speech

Text-to-speech (TTS) technology emerges as an effective companion for visually impaired people, remodeling the visual world into a realm of accessible statistics. Imagine navigating a town road, where a constant circulation of visible cues like street symptoms, keep displays, and traffic alerts bombard sighted pedestrians. For the visually impaired, this environment can be overwhelming. However, TTS steps in, bridging the space by changing visible statistics into clean, concise audio descriptions.

This technology operates by using a combination of sophisticated algorithms and meticulously curated record sets. Cameras or other picture devices first gather visible information about the surroundings. This record is then fed into picture processing and characteristic extraction systems, as discussed earlier. These systems extract key details like the presence of a crosswalk, the type of shop across the street, or maybe the contemporary traffic mild fame.

Once those features are diagnosed, TTS takes centre level. The system interprets the extracted facts into herbal-sounding audio commands or descriptions by way of leveraging pre- recorded audio samples or dynamically synthesizing speech. Imagine a visually impaired individual drawing near an intersection. The TTS machine, having analyzed the scene via image processing, can announce "You are coming near a crosswalk with a site visitor mild. The light is presently crimson." These clear and concise facts empower customers to make informed selections about navigating their surroundings accurately and independently.

The advantages of TTS increase past simple navigation. It can offer actual-time data approximately adjustments within the surroundings, which includes describing the appearance of public transportation or describing capacity boundaries at the sidewalk. Additionally, TTS may be integrated with cellphone programs, allowing customers to access information approximately their surroundings simply using pointing their smartphone's camera. For instance, a consumer might point their digital camera at an eating place sign, and the TTS machine might study aloud the restaurant's call or even announce its customer score.

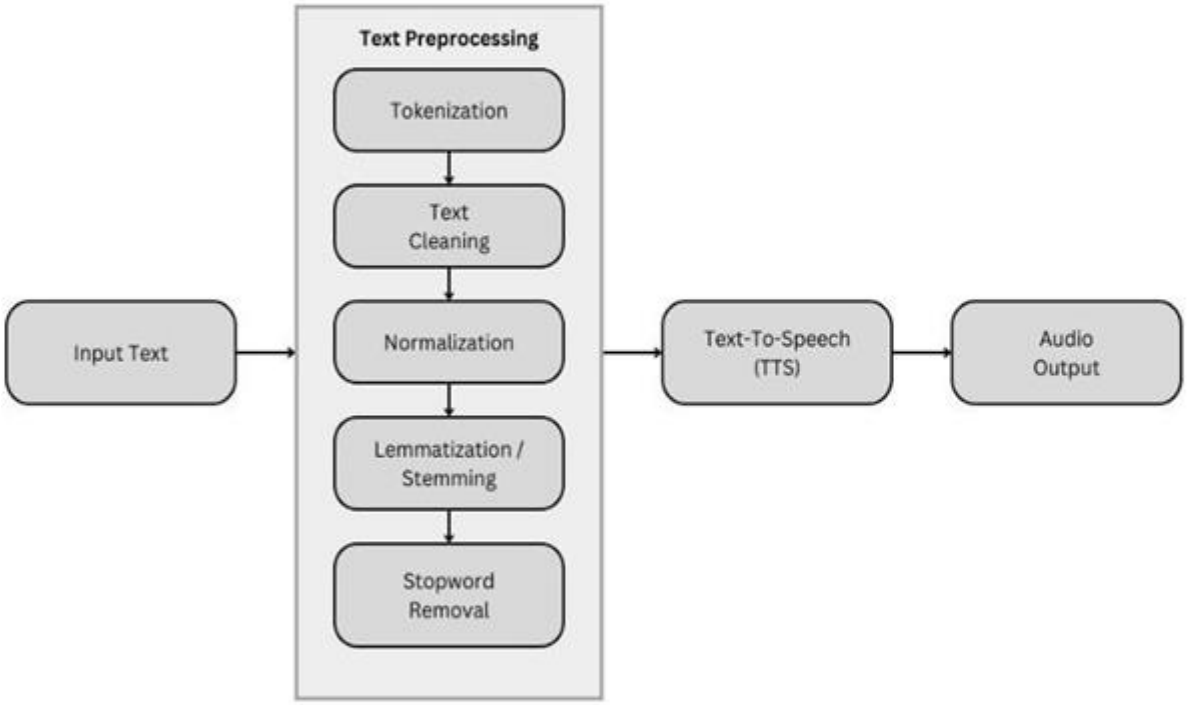

Figure2:Text to Speech

As TTS technology continues to evolve, we can expect even more natural-sounding voices with improved inflection and emotion. Furthermore, the integration of artificial intelligence opens doors for real-time scene analysis. Imagine a system that not only describes static objects but also narrates dynamic events, such as a child playing in a park or a street performer entertaining a crowd. This level of detail would create a richer and more immersive experience for visually impaired users.

In conclusion, text-to-speech technology is a digital narrator for the visually impaired, transforming visual information into a stream of clear and concise audio descriptions. This empowers them to navigate their surroundings independently, fostering a sense of inclusion and confidence in exploring the world around them.

D. Image Acquisition

Image acquisition serves as the foundation, capturing and translating the visual world into a data stream that can be interpreted and utilized. Imagine a visually impaired person navigating a busy intersection. Traditional methods might rely on cumbersome cane sweeps or pre-recorded audio descriptions with limited applicability. However, image acquisition offers a dynamic solution, capturing real-time visual details of the environment.This technology relies on specialized cameras or sensor devices specifically designed for navigation. Unlike traditional cameras used for photography, these devices prioritize capturing the most relevant information for safe and efficient navigation. Common examples include:

E. Lightweight Head-Mounted Cameras

These compact cameras are worn comfortably on a user's head, capturing a wide field of view of the immediate surroundings. This allows for real-time analysis of obstacles, landmarks, and potential hazards within the user's path.

F. Smart cane Cameras

Integrating a camera into a traditional cane adds a new dimension to navigation. By capturing information about the ground level, such as uneven pavement, curbs, or dropped objects, these smart canes empower users to avoid potential tripping hazards.

G. Environmental Sensor Arrays

In specific scenarios like high-traffic areas or public transportation hubs, a network of strategically placed cameras can create a comprehensive picture of the environment. This allows users to receive information about their surroundings even when not directly facing a particular direction.

The captured visual data then undergoes processing to extract the most critical details for navigation. This processing might involve techniques like:

H. Object Detection

Algorithms identify and locate objects within the image frame, such as pedestrians, traffic signals, or obstacles.

I. Depth Perception

Technologies like LiDAR (Light Detection and Ranging) can provide depth information, allowing the system to distinguish between a flat surface and a potential obstacle like a staircase.

J. Scene Recognition

Advanced algorithms might even analyses the broader context of the scene, recognizing landmarks like bus stops, building entrances, or specific stores, providing valuable information for orientation.

The key advantage of image acquisition lies in its ability to capture real-time visual data specific to the user's immediate environment. This dynamic approach surpasses the limitations of pre-recorded information or static maps, providing a more accurate and adaptable navigation experience.Looking ahead, advancements in image acquisition technology promise even greater benefits. Smaller, more discreet cameras will offer increased comfort and ease of use.

III.LITERATURE SURVEY

TABLE I LITERATURE SURVEY

Author	Title	Methodology	Findingsand Limitations
K.C.Shahira, Sagar Tripathy, A Lijiya(2019)	Obstacle Detection, Depth Estimationand Warning System for Visually Impaired People.	Yolov2, TTS	The accuracy of distance measurement was found to vary for some objects.The Executiontimefor some inputs is higher upto 17.98 secondsper frame.
Selman Tosun,EnisKaraarslan(2019)	Real-time Object Detection Applicationfor Visually Impaired People.	Yolo,Image Processing	Used on Tiny- Yolo Dataset which lowered the mean Average Precision(map).Ori ented towards detection on mobile devices, particularly Android.
Shaoqing Ren,Kaiming He,	FasterRCNN:	Regional	RPNwasmerged withFasterRCNN in

Ross Girsichk, Jian Sun(2016)	TowardsReal- time Object Detectionwith Regional Proposal Networks	Proposal Network(RPN)	the single network which enabled nearly cost-free region proposal.
DuyThanh Nguyen, TuanNghia Nguyen, Hyun Kim(2019)	AHigh Throughputand Power EfficientFGPA Implementation ofYOLOCNNforObject Detection	YOLOv2model using1-bit weights	Results are based on FPGA implementationandresultsonother hardware like ASCIs may differ. Performance is evaluated only on thePASCALVOC dataset, therefore limitedanalysison the impact of different bit- widths.

The cited works showcase diverse approaches to addressing the challenge of object detection and warning systems for visually impaired individuals. Shahira et al. (2019) developed an obstacle detection and warning system using YOLOv2 and gTTS but encountered accuracy and execution time issues. Meanwhile, Tosun and Karaarslan (2019) focused on real-time object detection with YOLO, particularly emphasizing its application on mobile devices like Android. Ren et al. (2016) proposed Faster RCNN, integrating a Regional Proposal Network for efficient region proposal generation. Nguyen et al. (2019) contributed a high throughput and power-efficient FPGA implementation of YOLOv2 using 1-bit weights. While each work presents valuable contributions, comparing them reveals distinct strengths and limitations. Tosun and Karaarslan's approach stands out for its real-time capabilities and mobile device focus, offering practical benefits for everyday use. Ren et al.'s integration of RPN with Faster RCNN enhances efficiency, while Nguyen et al.'s FPGA implementation prioritizes hardware optimization.

IV. PROPOSED SYSTEM

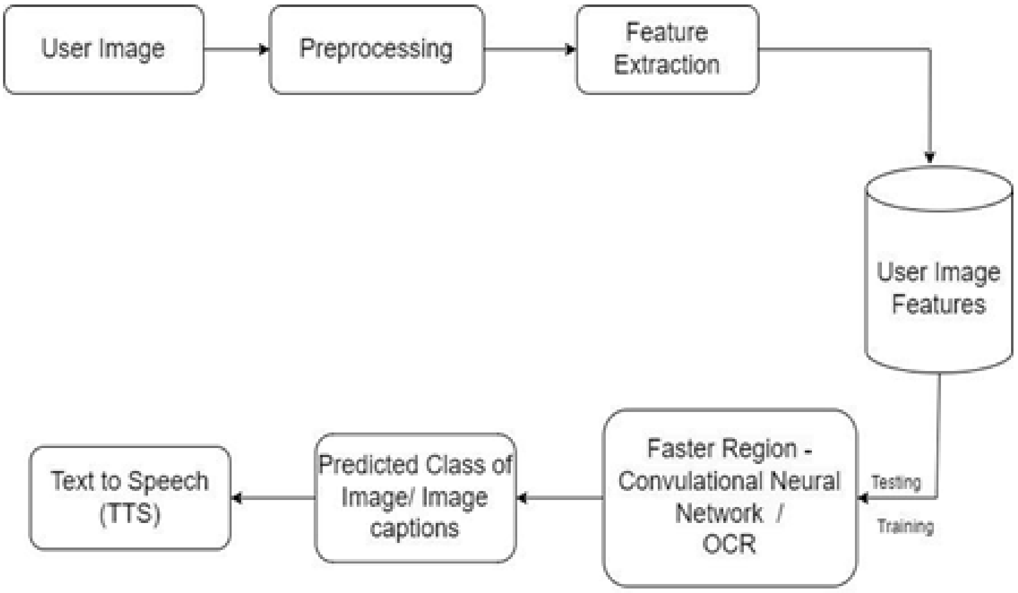

Figure3:System Architecture

A. User Image

This is the input to the system, consisting of images captured by a device such as a smartphone or camera. These images typically depict the user's surroundings, such as streets, intersections, or indoor environments.

B. Preprocessing

The user images undergo preprocessing to enhance their quality and facilitate subsequent analysis. Preprocessing steps may include resizing, noise reduction, and contrast adjustment to improve image clarity and remove irrelevant information.

C. Feature Extraction

In this step, relevant features are extracted from the pre-processed images. These features may include edges, shapes, textures, or other visual patterns that are important for navigation and object recognition.

D. User Input Features

Additional user input features are incorporated into the system to personalize the navigation experience. These features may include preferences, destination inputs, or specific instructions provided by the user through voice commands or text input.

E. RNN/OCR (Recurrent Neural Network/Optical Character Recognition)

This component employs RNN or OCR techniques to interpret text or recognize objects within the images. RNN models can analyze sequential data, making them suitable for tasks such as recognizing street signs, while OCR algorithms extract text from images, enabling the system to interpret written information such as store names or street names.

F. Predicted Class of Image

Based on the features extracted from the image and the output of the RNN/OCR component, the system predicts the class or category of the image. For example, it may identify objects such as pedestrians, vehicles, traffic signs, or landmarks relevant to navigation.

G. Text-to-Speech

The predicted class of the image is then converted into spoken text using text-to-speech (TTS) technology. This enables the system to convey information about the user's surroundings audibly, providing real-time guidance and alerts.

V. CONCLUSION

The challenge "Application to Help the Visually Impaired with the aid of Converting Images to Audio Descriptions" ends in effects that may extensively enhance accessibility and independence for visually impaired people with the aid of translating pics into audio descriptions, presenting users with all the facts about their environment gadgets. The integration of superior image recognition algorithms ensures an immoderate degree of precision which allows the visually impaired to recognize visible content material in a manner that became formerly inaccessible to them and maintain to make a meaningful difference in their lives.

VI. REFERENCES

[1]. Obstacle Detection, Depth Estimation and Warning System for Visually Impaired People K.C.Shahira, Sagar Tripathy, ALijiya(2019)https://ieeexplore.ieee.org/document/8929334

[2]. Real-time Object Detection Application for Visually Impaired People Selman Tosun Enis Karaarslan (2019) https://ieeexplore.ieee.org/document/8620773

[3]. Faster RCNN: Towards Real-time Object Detection with Regional Proposal NetworksShaoqing Ren, Kaiming He, Ross Girsichk, Jian Sun (2016) https://dl.acm.org/doi/10.5555/2969239.2969250

[4]. A High Throughput and Power Efficient FPGA Implementation of YOLO CNN for Object DetectionDuy Thanh, Nguyen, Tuan, Nghia Nguyen, Hyun Kim (2019) https://ieeexplore.ieee.org/document/8678682

[5]. Object Detection Based on the YOLO NetworkChengji Liu, Yufan Tao, Jiawei Liang, Kai Li, Yihang Chen (2018) https://ieeexplore.ieee.org/document/8740604