Cartable Camera Based Assistive Text Recognition for Visually Impaired

A. Karthikeyan[1] U.Kripanya[2] M.Manish[3] S.Nivetha[4] H.Prabanjan[5] K.Ramkumar[6]

[1]Assistant Professor, [23456] UG Scholar, Department of ECE
SNS College of Technology
Coimbatore, India

Abstract : Decades ago, visually impaired people faced a lot of challenges in their daily routine. Development and enhancement in technology enabled them to live confidently. Our proposed work consists of camera based assistance for the visually impaired students to improve their academics to read texts from hand held objects in daily routine works. In our proposed system we use OpenCV to filter and recognize characters from texts. This process is framed into three stages. The first stage consists of a mini camera which captures the image and the text is extracted from the image by text recognition algorithm. The second stage consists of the filtered text which is converted to speech as an audio output. The audio output is same as that of normal human speech. The third stage consists of speech to text conversion. The entire application will run on Raspberry-Pi kit. The camera module is used to capture the required image which will be later processed in OpenCv platform by using the OCR, the text will be extracted from the image captured. The audio to text conversion is achieved by means of Google Voice Talks. The text to audio output will be provided using eSpeak.

IndexTerms **- OpenCV, OCR, Google Voice Talks, eSpeak.**

I. INTRODUCTION

World Health Organization released a survey report on 10 facts regarding blindness which stated that there are around 314 million visually impaired and blind people are leading their life across the world. Out of which 45 million are with visual impairment. The National Health Interview released a valuation report that 25.2 Million adult Americans are blind or visually impaired. The National Census of India stated that there are around 21.9 Million disabled people in the country out of which more than 15 Million people are blind. In today's society everything is in the form of printed text. Printed text appears everywhere in the form of receipts, bank statements, menus in restaurants, classroom notes, product labels and even medicine bottles with instructions. So reading is very much essential. The purpose of our proposed system is to make a significant change in their life by enabling them to read printed labels and product packages. This will enhance their independent living, economic standard and social sufficiency. This paper presents a prototype for assistive text reading. This prototype consists of three main stages. They include image capturing based on the requirement as an audio output from the user. The image is captured in the form of video or image when the scene images matches with the object of interest from the user. The data processing component uses object-of-interest detection and text localization to extract image regions containing text, and text recognition. The recognized text codes is received by audio output component to the user. A Bluetooth earpiece with mini microphone is employed for speech output.

II. LITERATURE SURVEY

1.Fast Image Gradients Using Binary Feature Convolutions

St-Charles, P.L., Bilodeau, G.A. and Bergevin, R., deals with the recent increase in popularity of binary feature de-scriptors has opened the door to new lightweight computer vision applications. Most research efforts thus far have been dedicated to the introduction of new large-scale binary features, which are primarily used for key point description and matching. In this paper, we show that the side products of small-scale binary feature computations can efficiently filter images and estimate image gradients. The improved efficiency of low level operations can be especially useful in time-constrained applications. Through our experiments, we show that efficient binary feature convolutions can be used to mimic various image processing operations, and even outperform Sobel gradient estimation in the edge detection problem, both in terms of speed and F-Measure. Binary feature descriptors have recently experienced a surge in popularity in computer vision tasks such as object matching, visual correspondence, and texture analysis. Their compact and discriminative nature allows them to per-form just as well as traditional handcrafted image feature descriptors like SIFT's or SURF's yet at a much lower computational cost Typically, binary features are created by comparing low-level image characteristics based on a pre-determined pattern. The Boolean comparison results obtained are then concatenated

into bi-nary strings, and the "feature-space" distance between these descriptors can finally be defined as the Hamming distance between their binary strings.In this paper, we show that dense binary features can be used to efficiently approximate image processing operations that rely on convolutions, for instance the generation of an image pyramid, and local image gradient estimation. These operations are commonly used to solve computer vision problems, but to our knowledge, no previous work has studied how the side products of local binary descriptor computations can be used. To quantitatively assess the performance of binary feature convolutions and compare them to their traditional counterparts in a computer vision task, to test our approach on the edge detection problem. An existing solution to use our proposed convolution approach for gradient estimation, and measure how well it performs. Many solutions have been recently proposed for edge detection.

2. Automatic Voice Generation System After Street Board Identification For Visually Impaired

Dhulekar, P.A., Prajapatr, N., Tribhuvan, T.A. and Godse, K.S.,deals with the importance of communication. Communication is a very important tool in a human life. It is an essential requirement in this world to survive. To overcome the difficulties faced in communication; this Paper has been implemented to achieve a real time system with feature/character/symbol extraction , text to speech conversion and then speech conversion into different languages.
The Project is based on design & implementation of smart hybrid system for street sign boards recognition , text and speech conversions through character extraction and symbol matching. The default language use to pronounce signs on the street boards is English . Here we are proposing a novel method to convert identified character or symbol into multiple languages like Hindi , Gujarati , etc. This Project is helpful to all starting from the visually impaired , the tourists , the illiterates and all the people who travel.

The system is accomplished with embedded platform for real time conversion , speech pronunciation and to display on intended devices. This Project has a multidisciplinary approach as it belongs to the domains like computer vision , speech processing and embedded system. Computer vision is used for character and symbol extraction from sign boards . Speech processing is used for text to speech conversion and then to multiple language conversion of original speech signal . Embedded platform is used for real time pronunciation and displaying of desired output . This paper is an effort to implement for character extraction and text to voice conversion of different images using optical character recognition and text to speech synthesis technology . A user friendly , cost effective to all and applicable in the real time system is achieved. Using this methodology , we can read text from a document, street sign boards, web page or also an e-Book and can generate synthesized speech through any portable system i.e. computer's speakers or phones speaker. In developed software, the use of computer vision has set all policies of the characters corresponding to each and every alphabet extraction, matching, its pronunciation and the way it is used in grammar and dictionary. Speech processing has given robust output for various images. This saves users time by allowing him to listen background materials. Other application of this system includes such as making information browsing for people who do not have the ability to read or write.

This approach can be used in part as well. If requirement is only for text conversion then it is possible or else text to speech conversion is also done easily. People with vision impairment or visual dyslexia or complete blindness can use this approach for reading the documents books and also while travelling. People with vocal impairment or complete dumb person can utilize this approach to turn typed words into vocalization. Tourists having language barrier can also use this approach for understanding different languages. People travelling in cars and buses can save their time using this feature. Experiments have been performed to test the text and speech generation system and good results have been achieved. Still the work is in progress for symbol extraction.

III. EXISTING SYSTEM

In the existing system, a camera-based assistive technology helps blind persons to read text from hand-held objects in their daily lives. The earlier system is not that efficient because the image or video captured is along with noise. In the earlier system, the images with multiple backgrounds can not be captured and processed. The product packaging labels containing text labels with different font styles and different colour cannot be processed. The object in motion is not also processed.To isolate the object from multiple backgrounds or other surrounding objects from the camera view, we first propose an efficient and effective motion based method to define a region of interest (ROI) in the video by asking the user to shake the object to the camera's sight. This method extracts moving object region by a mixture-of-Gaussians-based background subtraction method. In the extracted ROI, text localization and recognition are conducted to acquire text information.

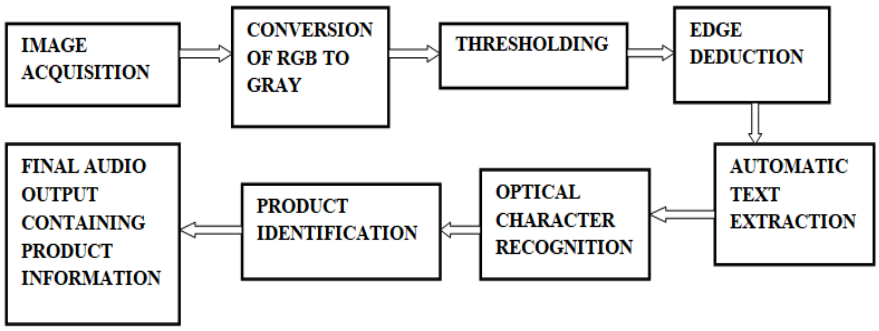

Fig.1 Existing System

IV. PROPOSED SYSTEM

The need for equity for all humans has become an important aspect and own surveillance is essential.This work mainly aims to provide confident lifestyle for blind .Many works have proposed to provide various devices for their daily life. In this paper an advanced cartable text reading device has been implemented.

The proposed approach is designed along with a portable camera model to provide an efficient text reading for visually impaired by using Raspberry-Pi . Various techniques are proposed in the work for the simulation of audio to text and text to speech conversion . In this article, the speech to text conversion is performed using Google Voice Talks, an open source platform and the text to speech conversion is performed in OpenCV platform. The text recognition is performed efficiently using eSpeak and an audio output is provided.

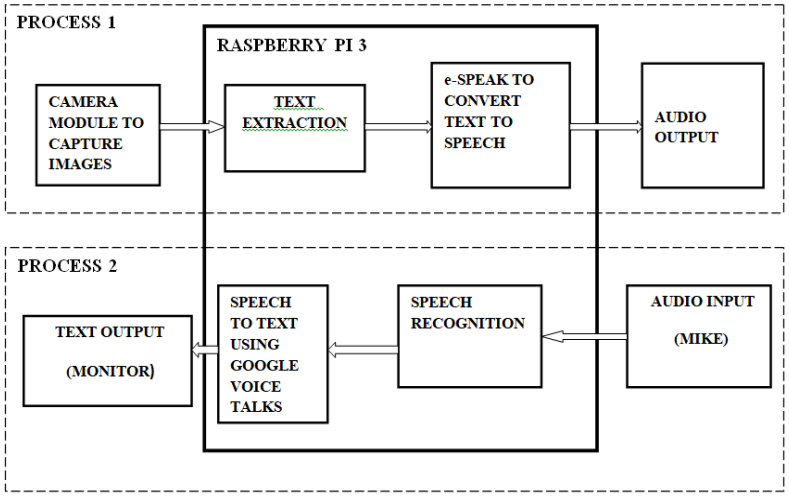

Fig.2 Proposed System

Camera acts as a main tool in capturing the text as image. Then image is processed internally and separates text from image by using OpenCV. OpenCV is available with Tesseract OCR as default. The image to text conversion is performed by image acquisition, grey scale conversion, edge detection, thresholding and automatic text extraction by the Optical Character Recognition(OCR). The OCR is used to convert the text from text regions. The extracted text to speech is done by eSpeak, It is an open source software which is used to prepare and compile phoneme data. eSpeak gives a clear data through the audio jackets. It gives the clear idea about the text.

The audio to text conversion is performed by Google Voice Talks, an open source platform. The output from the microphone is fed as an input to the Google Voice Talks as voice command. The voice may contain both background noise and human voice. The input signal is zero padded on both ends of voice spectrum to retrieve human noise. The desired spectrum is prominent relative to the noiseless zero-pads and filter out the background noise. The speech synthesizer and recognition is using Google's speech API. Since, this speech recognition is one of the best with Google Voice Talks. Speech to text is converted and the text is eventually displayed in the monitor.

Tesseract OCR

Tesseract is an open source OCR engine which can effectively extract text from black on white text and white on black text .The image is first converted to grey image to make the process easier, then the image is threshold to make it a binary image. The binary image text lines are broken according to the pitch , fixed pitch texts are broken first then the non fixed pitch are solved by measuring the gaps between the text vertically and text nearing the threshold values are taken whole and words are identified. Word recognition is done for each word separately for obtaining accuracy and also it solves particular spaces to locate small cap text and it is breaks the word. The broken characters are joined and associate each words for the right meaning. these words are later classified according to the string assigned and the accurate text gets extracted.

eSpeak

Text to Speech is an open source synthesizer. It involves the following process, First input text to pronunciation phonemes and then the pronunciation phonemes into sound. To include intonation for speech apostrophes are used which gives the stress to the syllable. Secondly, sound synthesis from intonated data.The eSpeak synthesizer adds the voiced speech to the sine waves to make the total sound.These sounds are filtered to eliminate noise and a clear speech is obtained.

Google Voice Talks

Speech recognition is done using Google Voice Talks, an open source software. Speech recognition is to decipher on spoken word. The process involved in Speech to Text conversion is to extract the sound that group together(Phoneme).English utilise 40 phonemes to represent the 500,000 words. Phonemes are extracted through a Fourier Transform by running the waveform. This waveform can be analysed at a spectrograph. Then the second process involves Markov Models .The Hidden Markov Model(HMM) is used to convert the phonemes to words and words into sentences. The HMM contains a chain of phonemes that represent a word, this chain can be extended to form a sentence.

V. HARDWARE

Raspberry Pi

The assistive text reader system includes a power supply to Raspberry Pi board , connecting earphones to the 3.5mm audio jack and interfacing camera. The Raspberry Pi board will be powered through USB cable or battery. It is a credit card sized single computer or Sock uses ARM1176JZF-S core. System on a Chip (SoC), is a method in which all electronics for running a computer are placed on a single chip. The CPU, GPU, USB controller, RAM everything is compressed down into one small package. Operating system is used to start Raspberry Pi. In the aim of cost reduction, the Raspberry Pi skips any on-board non-volatile memory used that stores the Linux Kernels, file systems and boot loaders as seen in many embedded systems. For this purpose SD/MMC card slot is inserted.

SD/MMC CARD

Secure digital is used for storage. It is a non-volatile flash memory. Gradually Multimedia card is also used to store data. SD card holds all the open source software's like Raspbian OS, eSpeak and Google voice talks. SD card is the main element in the working of the Raspberry Pi. It can be extended according to our needs. Minimum speed ranges from 2 mbps to 30 mbps.

CAMERA

Web cam is an image sensors used to capture text images. It is incorporated with noise suppression algorithms to obtain a clear image with an optical resolution 1280x960 1.2MP. Its field of view is 600 with a focal length of 4.0mm.

DISPLAY

The extracted text from speech is displayed in the monitor. Monitor can be interfaced to Raspberry Pi by two mediums wired or wireless. Wired connections use HDMI Cable while wireless uses WIFI when connected to a laptop by obtaining the IP address.

APPLICATIONS

It is highly recommended for visually impaired as assistance to perform their daily chores.
It also provides assistance for the paralysed patients.
Automatic read-write process in industries

ADVANTAGES

It is cost efficient.
The result is with high resolution.
Any language text at any angle or any style can be captured and processed by means of TesseractOCR

V. CONCLUSION

This paper is used to solve the common aiming problem for blind users. We have proposed a camera-based text reading framework to help blind persons to excel in their academics by providing an assistive cartable read-write kit with the help of Raspberry Pi.

VI. FUTURE WORK

The future works include implementing native language recognition for excelling in their own languages with more accuracy. To provide recognition of mathematics equations.

REFERENCES

[1] St-Charles, P.L., Bilodeau, G.A. and Bergevin, R., 2016. Fast Image Gradients Using Binary Feature Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 1-9).

[2] Dhulekar, P.A., Prajapatr, N., Tribhuvan, T.A. and Godse, K.S., 2016, December. Automatic voice generation system after street board identification for visually impaired. In Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), 2016 International Conference on (pp. 91-96).

[3] Karaoglu, S., Tao, R., van Gemert, J.C. and Gevers, T., 2017. Con-Text: Text Detection for Fine-grained Object Classification. IEEE Transactions on Image Processing

[4] Khan, M.N.H., Arovi, M.A.H., Mahmud, H., Hasan, M.K. and Rubaiyeat, H.A., 2015, December. Speech based text correction tool for the visually impaired. In Computer and Information Technology (ICCIT), 2015 18th International Conference on (pp. 150-155). IEEE

[5] A.Karthikeyan, P.Saranya, N.Jayashree "An Efficient VLSI Architecture For 3D DWT using Lifting Scheme" in International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 2, Issue 1, January 2013, pp 292-298 ISSN : 2319-5967

[6] A.Karthikeyan, N. Rajeswaran "Design and Implementation of multiple Fault Diagnosis on VLSI Circuits Using Artificial Neural Networks" in International Journal of Advanced Engineering & Technology (IJAET) Volume 3, Issue 2, May 2012, pp 685-695, ISSN: 2231-1963.