ISSN No: - 2456 - 2165

Portable Camera Based Assistance for Visually Impaired People

Muhammed Shabeer Salih P.K
Mohammed Najaaf, Moideen Unais K
Muhammed Thanzeer M.T, Rajna K.K.
B.Tech Students of Dept of ECE
MESCE,Kuttippuram

Rejiram R.,
Asst. Professor
Dept of ECE,
MESCE,Kuttippuram

***Abstract*—The ability of people who are blind or have significant visual impairments to read printed labels and product packages will enhance independent living and foster economic and social self-sufficiency. Today, there are already a few systems that have some promise for portable use, but they cannot handle product labeling. We propose a Camera-based assistive text reading framework to help blind persons read text labels and product packaging from hand-held objects in their daily lives. The proposed system will provide a complete assistance to visually impaired people. System consist of following features camera-based assistive text reading framework to help visually impaired persons read text labels and product packaging from hand-held objects in their daily lives. Face recognition and detection feature which allows the visually impaired people to capture and store the faces of the people they want to remember and hence they can easily identify or recognize the face of that person over a crowd. The system also includes navigation and an obstacle détection system for the visually impaired people to guide them to the destination.**

Index Terms*—*OCR, OpenCV, Face Recognition, Visually Impaired.

I. Introduction

Visual impairment afflicts approximately 285 million people worldwide according to recent estimates by the World Health Organization (WHO) and, without additional interventions, these numbers are predicted to increase significantly.

Reading is obviously essential in today’s society. Printed text is everywhere in the form of reports, receipts, bank statements, restaurant menus, classroom handouts, product packages, instructions on medicine bottles, etc. And while optical aids, video magnifiers, and screen readers can help blind users and those with low vision to access documents, there are few devices that can provide good access to common hand-held objects such as product packages, and objects printed with text such as prescription medication bottles. The ability of people who are blind or have significant visual impairments to read printed labels and product packages will enhance independent living and foster economic and social self-sufficiency.

In this project we propose a Camera-based assistive text reading framework to help blind persons read text labels and product packaging from hand-held objects in their daily lives. And a face recognition and detection feature which allows the visually impaired people to capture and store the faces of the people they want to remember and hence they can easily identify or recognize the face of that person over a crowd. The system also includes navigation and an obstacle detection system for the visually impaired people to guide them to the destination.

II. EXISTING TECHNOLOGY

The most basic and widely used method for reading by a blind person is by Braille. For supporting that, the technology used is Computer Driven Braille Printer, Paperless Braille Machines, Optacon etc. These technologies use different techniques and methods allowing the person to read or convert document to Braille. From Nowadays Computers are designed to interact by reading the books or documents. Synthesized voice is used to read the content by the computers. Some devices are available which can scan the documents and use interfaced screen to allow the blind to sense the scanned documents on the screen either in Braille or by using the shape of the letters itself with help of vibrating pegs. Various phone applications are also developed to help in reading or helping the blind in other ways**.** But, We propose a hand held portable module which is controlled by a camera placed on the spectacle to make a complete assistance to the visually impaired people.

III. FRAMEWORK AND SYSTEM OVERVIEW

The Block diagram of the proposed system is given in fig.3.1.The brain or the head of system is Raspberry Pi which is a credit card-sized computer powered by the Broadcom BCM2835 system-on-a-chip (SoC). The webcam captures the frames which are processed by Raspberry pi and corresponding audio signals are given to the headphone.

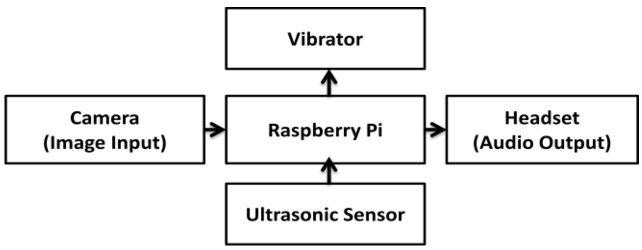

Fig.3.1 Block Diagram

The main 3 features of the system are:

- The system having a camera-based assistive text reading framework to help visually impaired persons read text labels and product packaging from hand-held objects in their daily lives.

- System having a face recognition and detection feature which allows the visually impaired people to capture and store the faces of the people they want to remember and hence they can easily identify or recognize the face of that person over a crowd.
- Obstacle detection and alert system.

A. Text Recognition

- *Image acquisition*

Fig.3.2 indicates the image to speech conversion process, here initially captures the real time image by using a webcam and stored in the sd card, which is for further process. OpenCV is used as the tool for processing of images. OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library.

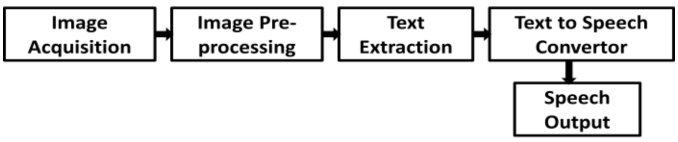

Fig.3.2 Image to audio conversion

- *Image Pre-processing*

First of all we segment the image using Thresholding . It is a non-linear operation that converts a gray-scale image into a binary image where the two levels are assigned to pixels that are below or above the specified threshold value. In other words, if pixel value is greater than a threshold value, it is assigned one value (may be white), else it is assigned another value (may be black). Fig.3.3 shows the thresholding process.

We apply a fixed-level thresholding to a single-channel array. Then convert the grayscale to a bi-level (binary) image. For removing a noise, filtering out pixels with too small or too large values. Then computed threshold value and thresholded image.

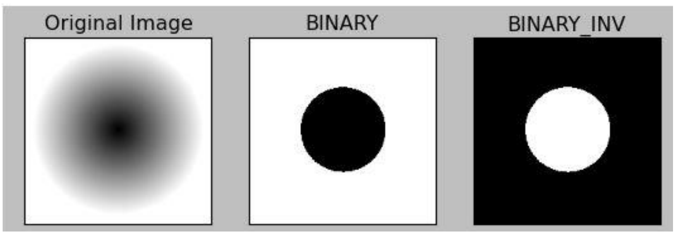

Fig.3.3 Image Thresholding

- *Text Extraction*

For the text extraction, The pre-processed image is given to Tesseract OCR engine. Tesseract OCR works in step by step manner as per the block diagram shown in fig. 3.4 First step is Adaptive Thresholding , which converts the image into binary images. Next step is connected component analysis , which is used to extract character outlines. This method is very useful because it does the OCR of image with white text and black background.

Tesseract was probably first to provide this kind of processing. Then after, the outlines are converted into Blobs. Blobs are organized into text lines, and the lines and regions are analyzed for some fixed area or equivalent text size . Text is divided into words using definite spaces and fuzzy spaces . Recognition of text is then started as two-pass process as shown in fig 3.4 In the first pass, an attempt is made to recognize each word from the text. Each word passed satisfactory is passed to an adaptive classifier as training data . The adaptive classifier tries to recognize text in more accurate manner. As adaptive classifier has received some training data it has learn something new so final phase is used to resolve various issues and to extract text from images.

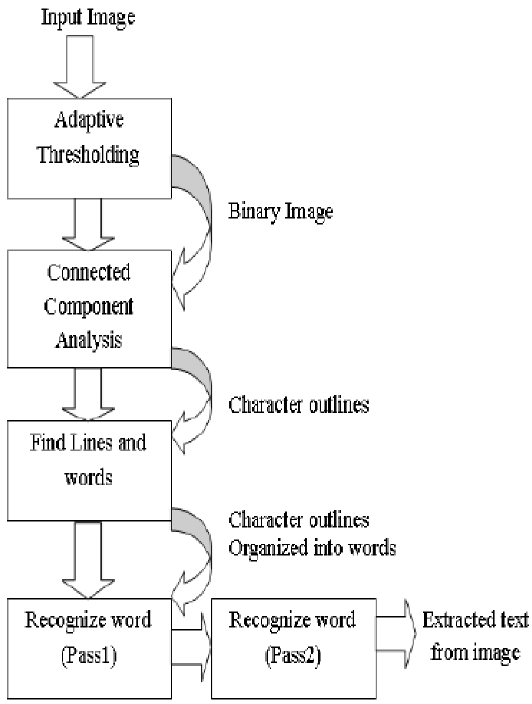

Fig 3.4 Architecture of Tesseract OCR

After processing is completed, the content of the output file is a text file. In simple images with or without color (gray scale), Tesseract provides results with 100% accuracy. But in the case of some complex images Tesseract provides better accuracy results if the images are in the gray scale mode as compared to color images. To prove this hypothesis, OCR of same color images and gray scale images is performed and in both cases different result are achieved.

B .Text to Speech Conversion

The OCR converted text is processed by the eSpeak TTS engine. It is composed of two parts, The front-end has two major tasks. First, it converts raw text containing symbols like numbers and abbreviations into the equivalent of written-out words. This process is often called *text normalization*, *pre-processing*, or *tokenization*. The front-end then assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units, like phrases, clauses, and sentences. The process of assigning phonetic transcriptions to words is called *text-to-phoneme* or *grapheme-to-phoneme* conversion. Phonetic transcriptions and prosody information together make up the symbolic linguistic representation that is output by the front-end. The back-end—often referred to as the *synthesizer*—then converts the symbolic linguistic representation into sound. In certain systems, this part includes the computation of the *target prosody* (pitch contour, phoneme durations) which is then imposed on the output speech.Fig.3.5 shows the overview of atypical TTS system.

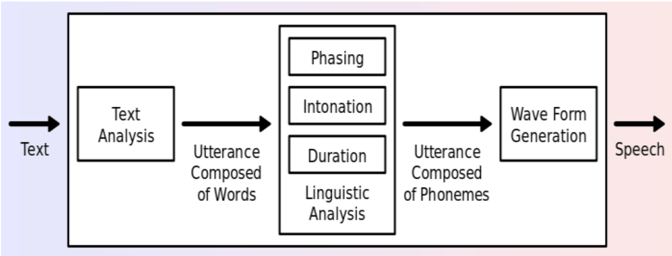

Fig 3.5 Overview of a typical TTS system

Rhythm is an important factor that makes the synthesized speech of a TTS system more natural and the prosodic structure provides important information for the prosody generation model to produce effects in synthesized speech.

- *Architecture of TTS*

The TTS system comprises of these 5 fundamental components:
A. Text Analysis and Detection
B. Text Normalization and Linearization
C. Phonetic Analysis
D. Prosodic Modeling and Intonation
E. Acoustic Processing
The input text is passed through these phases to obtain the speech.

C. Face Recognition

The second feature is the face detection and recognition feature. Here the user is allowed to take the portrait pictures of the people he wants to remember. This collection of pictures is used as the database. The faces of the people is detected using the software OpenCV. Later, when the person whose face is stored in the database comes around , can be recognized as the person he knows.

The inability to recognize known individuals in the absence of audio or haptic cues severely limits the visually impaired in their social interactions and puts them at risk from a security perspective. The difference between face detection and recognition is that in detection we just need to determine if there is some face in the image, but in recognition we want to determine whose face it is. In order to understand the methods for recognizing faces, more advanced mathematical knowledge is required; namely linear algebra and statistics.

OpenCV 2.4 comes with the very new Face Recognizer class for face recognition. The currently available algorithms are:

i. Eigenfaces
ii. Fisherfaces
iii. Local Binary Patterns Histograms

All three methods perform the recognition by comparing the face to be recognized with some training set of known faces. In the training set, we supply the algorithm faces and tell it to which person they belong. When the algorithm is asked to recognize some unknown face, it uses the training set to make the recognition. Each of the three aforementioned methods uses the training set a bit differently.

Eigenfaces and Fisherfaces find a mathematical description of the most dominant features of the training set as a whole. LBPH analyzes each face in the training set separately and independently.

D. Obstacle Detection

The final feature is the obstacle detection feature. Here, with the help of ultrasonic sensors , the blind person can easily walk around the obstacles. Thus , acts as a path guidance for the visually impaired people.

Ultrasonic Distance Sensor with ASCII serial O/P is an amazing product that provides very short to long-range detection and ranging. The sensor provides precise, stable non-contact distance measurements from about 2cm to 400 cm with very high accuracy. Its compact size, higher range and easy usability make it a handy sensor for distance measurement and mapping. The board can easily be interfaced to microcontrollers RX pin (USART) . At every 50ms sensor transmits an ultrasonic burst and send out ASCII value of distance that corresponds to the time required for the burst echo to return to the sensor. This sensor is perfect for any number of applications that require you to perform measurements between moving or stationary objects.

IV. RESULT

In this proposed scheme we have obtained these following results. Fig 4.1 shows the real time image from portable camera, it is the input to the text extraction program.

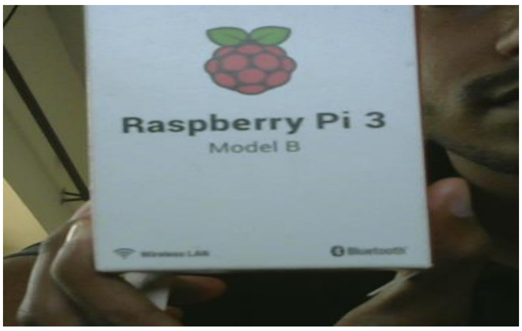

Fig.4.1 Captured Image

ISSN No: - 2456 - 2165

The input image is pre-processed by adaptive thresholding and the binary inverted image is obtained, which is shown in Fig.4.2.

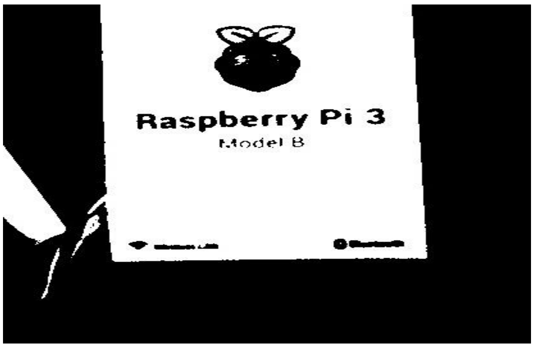

Fig.4.2 Pre-Processed Image

The thresholded image is then processed by tesseract ocr for text localization and extracted. The extracted text is stored as text file. Fig.4.3 shows the extracted text from the input image fig.4.1.

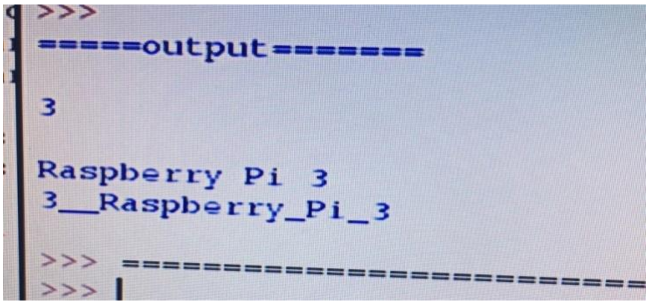

Fig.4.3 Output text

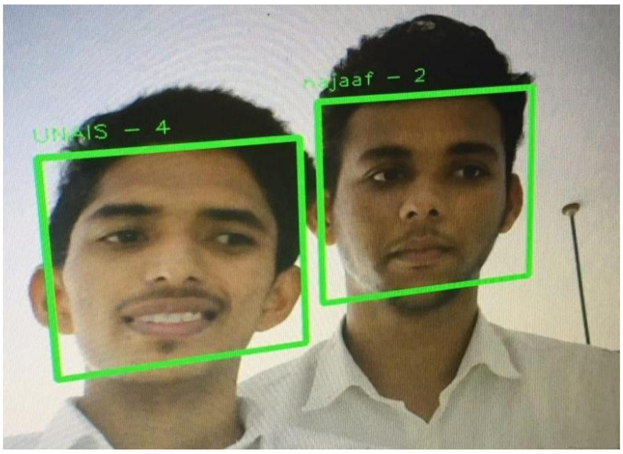

Fig.4.4 Face Recognition

Fig.4.4 shows the face detection and recognition feature, which shows the detected face, name and their accuracy value in prediction. The name is stored in a text file.

The texts from the extracted image and name from the face recognition program is converted into audio by eSpeak TTS engine.

Fig.4.5 Primitive Prototype Model

Fig.4.5 shows the prototype of the proposed system. It has the camera placed on a spectacle which is used by the visually impaired people. The system controlled by the Raspberry Pi.

V. CONCLUSION AND FUTURE WORK

In this paper, we have described a prototype system to read printed text on hand-held objects , detect and recognize known faces and obstacle detection for assisting blind persons. We have attached a portable camera to the system which will detect the faces and text labels. The final output is taken as speech and vibrator. We have done the image extraction method in two levels firstly we converted the image to text using tesseract tool and text to speech conversion using espeak tool. Face reorganization operation is performed using openCv tool. Ultrasonic sensors are being used to detect the obstacles and feeds the warning vibration to the user.

The future work will be concentrated on developing an efficient portable product that can extract text from any image enabling the blind people to read text present on the products, banners, books having complex backgrounds .

REFERENCES

[1] CHUCAI YI, STUDENT MEMBER IEEE, YINGLI TIAN, SENIOR MEMBER, IEEE, AND ARIES ARDITI , "*PORTABLE CAMERA BASED ASSISTIVE TEXT AND PRODUCT LABEL READING FROM HAND-HELD OBJECTS FOR BLIND PERSONS*", IEEE/ASME TRANSACTIONS ON MECHATRONICS, VOL. 19, NO. 3, JUNE 2014

[2] Anusha Bhargava, Karthik V. Nath , Pritish Sachdeva and Monil Samel, **"***Reading Assistant for the Visually Impaired*" in International Journal of Current Engineering and Technology Vol.5, No.2 (April 2015

[3] R. Smith, "*An Overview of the Tesseract OCR Engine*," Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), Parana, 2007, pp. 629-633.

ISSN No: - 2456 - 2165

[4] Chirag Patel , Atul Patel, Dharmendra Patel , "*Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study"* International Journal of Computer Applications (0975 – 8887) Volume 55– No.10, October 2012.