

# Blind Assistance Device using AI

Safnaz<sup>1</sup>

Dept. of IS&E, YIT, Moodbidri  
Mangalore, India

Nayana Manju Jogi<sup>2</sup>

Dept. of IS&E, YIT, Moodbidri  
Mangalore, India

Shreya Shetty<sup>3</sup>

Dept. of IS&E, YIT, Moodbidri  
Mangalore, India

Prasad Rathode<sup>4</sup>

Dept. of IS&E, YIT, Moodbidri  
Mangalore, India

Mrs. Deeksha K R<sup>5</sup>

Assistant professor, Dept. of IS&E  
YIT, Moodbidri  
Mangalore, India

**Abstract**— This paper we present emotion and gender detection and also document reader for visually impaired people. Our aim is to provide an assistance device for blind and visually impaired people. In this project, we develop and implement a framework for designing real-time convolutional neural networks (CNNs). We test our models by building a real-time vision system that does face detection, emotion categorization, and gender detection. Using our suggested CNN architecture, we can do both at the same time in a single blended step. This project focuses on person expression detection and classification on pictures which are captured by the device and also text to voice conversion mounted on a device. This project contains two main parts, an Artificial Intelligence and OCR (Optical Character Recognition). There will be two buttons one for Artificial Intelligence and another for OCR. When the user hits the appropriate button on the gadget, the camera begins to take a photo and analyses the image for emotion recognition using Tensor Flow (an open-source toolkit for numerical computation) and the Raspberry Pi is used to convert and recognize what the visual is about, and then the gadget will vocally help the person utilizing a speaker or headset.

**Keywords**— AI (Artificial Intelligence), Tesseract, TensorFlow, CNN (Convolution Neural Network), OCR (Optical Character Recognition), Raspberry Pi.

## I. INTRODUCTION

From reading a book to walking down the street, blind persons face numerous hurdles in their daily lives. Many tools exist to assist people in dealing with their problems, yet they are insufficient. Vision is the most important thing a human may have, and it plays a crucial role in a person's life whether or not they can see. The life of a blind people is difficult because they can't analyze the situation like how the normal person does. Our gadget is used to assist blind and visually impaired persons with emotion recognition and reading of documented text in photos.

Humans communicate mostly through voice, but they also use body movements to highlight specific sections of their speech and to express emotions. Facial expressions, which are an integral element of communication, are one of the most

important ways humans communicate emotions. Even though no words are spoken, there is much to learn about the messages we send and receive through nonverbal communication. Facial expression recognition is a crucial component of natural human-machine interfaces and can also be used in behavioral research. We provide a facial recognition approach based on Convolutional Neural Networks in this paper (CNN). The data that goes into our system for face expression recognition is an image. Following that, we use CNN to predict the facial expression label, which should be one of the following: anger, disgust, fear, happiness, neutral, sadness, or surprise. A method for extracting text from photographs has been presented. Machine-readable text from typed text Create the text files, then process them with Digital. To translate the text into audio output, using pyttsx3 which is a voice engine. Our goal is to improve the ability of blind people. Offering a solution for them to feed the information in the form of a voice message to them.

## II. LITERATURE SURVEY

Ashwani Kumar, Ankush Chourasia [1], In this paper they focused on the assistive devices for visual impairment people in order to provide a hearable environment for blind people. It translates visual input into an alternate rendering mode suitable for blind users through image and video processing. Auditory, haptic, or a combination of both senses can be used as alternate modalities. As a result, artificial intelligence is being used to convert from one modality to another.

Ms. Rupali, D Dharmale, and Dr. P.V. Ingole [2], They proposed a Raspberry Pi prototype for extracting text from photos. The photographs were taken with a webcam and processed with Open CV and the OTSU algorithm. The collected photos are initially transformed to grayscale colour mode. By adjusting the vertical and horizontal ratios, the images are rescaled and are applied. After performing some adjustments, pictures are subjected to OTSU's adaptive thresholding technique. Contours for the photos are formed using specific functions in Open CV after thresholding, and

bounding boxes are drawn around the objects and text in the images using these contours. Each character in the image is retrieved using these created bounding boxes, which is then fed into the OCR engine, which recognizes the text in the image.

Nagaraja L [3], "Vision based text recognition using raspberry PI." In this paper they Presented a camera-based assistive text reading framework to help visually impaired people read text labels and product packaging from hand-held objects in everyday life. The approach presents a motion-based method for defining a Region Of Interest (ROI) in the camera image, which isolates the object from untidy backdrops or other surrounding objects. To extract the moving object region, a Gaussian-based background removal technique is applied. Text localization and recognition are used to obtain text details from the ROI. The object ROI's text regions are then increased. The filters of a convolution neural network will improve the model's performance layer by layer with the same number of network layers. At the same time, the 1 h 1 convolution layer improves the expression ability of the models by increasing the non-linear representation of the input, deepening the networks, and increasing the non-linear representation of the input. However, this study has a number of flaws, like neglecting the profound emotions hidden beneath the face and having low accuracy, which will be rectified in future research. Novel Text localization approach performs gradient features of stroke orientations and distributions of edge pixels in an AdaBoost model. Self-optical character identification software binarizes and recognizes text characters in localized text regions.

Ms.Kavya.S[4] Visually impaired people encounter challenges and are at a disadvantage as visual information is what they lack the most. The visually handicapped can be helped with the use of innovative technologies. They use the concept is using an Android mobile app that includes features such as voice assistant, image recognition, currency recognition, e-book, and chat bot. The software can recognize items in the environment using voice commands and do text analysis to recognize text in a hard copy document.

Padma Shneha[6] They used Artificial intelligence and development of computer systems which are capable of perception, reasoning, and action. Their goal is to create a machine that can think for itself. Learning, thinking, making decisions, and solving problems are all aspects of intelligence. Where AI is a collaborative discipline that necessitates expertise in a variety of fields, including psychology, computer science, engineering, logic, and mathematics.

Ani R1, [7] This paper consists of a prototype system of assistive text reading. The concept of proposed system is the idea of developing specs reader-based text reading system for visually impaired persons. There are three different modules in this system: Camera module, Optical Character Recognition Module and Text-To-Speech Module This explains the text reading system for visually impaired users for their self-independent.

Xiaofang Jin,[8] By comparing the four neural networks models for facial expression recognition, we can

see that the accuracy may not increase if the number of network layers is increased. In the same number of network layers, the filters of convolution neural network will improve the performance of the model by increasing layer by layer. At the same time, it can be seen that the 1 h 1 convolution layer increases the non-linear representation of the input, deepens the networks and improves the expression ability of the models. However, this study still has many shortcomings, ignoring the deep emotions hidden under the expression, and the accuracies are not very high, which will be improved in the future research.

Ashwani Kumar, [8] In this paper, they used object detector model that uses deep learning neural networks for or even webcam feeds. The accuracy of the model is more than 75%. The training time for this model is 5-6 hours. This model uses deep neural networks to extract feature information and then perform feature mapping, the SSD Mobile Net FPN performs localization and classification of neural networks and hence it is fast compared to other models. In this work, we have used a Single Shot MultiBox Detector (SSD) algorithm to achieve high accuracy and IOU in real time for detection of the objects for the blind person. A convolutional neural network is used to extract feature information from the image and then perform feature mapping to classify the class label. This paper points out that the algorithm uses truth box to extract feature maps. The metrics of our approach indicates that the algorithm achieves a higher MAP and acceptable accuracy for detecting the objects from color images for blind person.

### III. METHODOLOGY

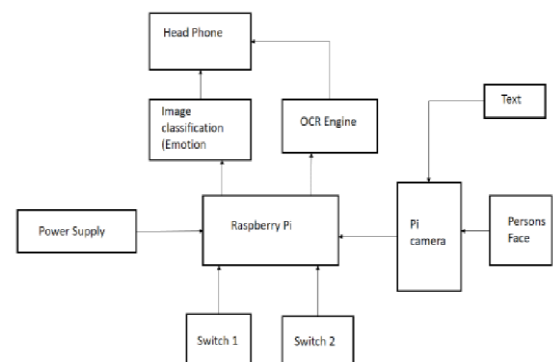


Fig 1. Methodology of Blind Assistance Device

Project entitled as blind assistance device using AI, mainly consist of two divisions in which one is "facial expression recognition" and the other one is "text to speech conversion". We are combining two of these and deploying to raspberry pi device along with a pi camera. We arranging two switches for each operation respectively. For both the operation image will be the input which is scanned through the camera. Once the input is being scanned, for the face expression program switch one need to be pressed and it will perform facial expression operation similarly switch 2 will perform the text recognition to speech conversion. The

output is then relayed to the user through head phone/speaker/audio device.

The basic algorithm that will be implemented for working of this proposed system is as follows,

- **Speech Recognition**- converts the user’s voice to text.
- **OCR (Keras OCR)** - extracts text from images. If the user wants any text to be read to them, they can utilize this option.
- **Raspberry Pi 4:** It is the latest version of the low-cost Raspberry Pi Computer.
- **PI Camera** - It is faster to use.
- **Image classification** - is a process of taking an input and produces an output classification for identifying whether the person is present or not
- **Text-to-Speech (Google TTS)** - It turns the user’s textual response into a voiced response.

#### IV SYSTEM REQUIREMENT ANALYSIS AND SPECIFICATION

Software requirements and hardware requirements are given as follows:

- *Software Requirements*

- Operating system : windows 10
- Software : Tensor flow, Keras,
- Languages : Python

- *Hardware Requirement*

- Raspberry pi
- Pi camera
- Switch
- Button
- Headphone
- SD card

#### A. Raspberry Pi

Raspberry pi is a line of compact single-board computers aimed at teaching basic computer science in schools and other poor countries. The Raspberry Pi can be used for a variety of applications. It can be adjusted to meet the needs of the user.



Fig 2. Raspberry Pi

#### B. Artificial Intelligence

The field of creation of intelligent machines that work like humans and respond quickly, in computer science is known as Artificial intelligence. The core part of AI research is Knowledge engineering. Machines can react and act like humans only when they have abundant information related to the world. To implement knowledge engineering, Artificial intelligence should have access to objects, categories, properties, and relations. To initiate common sense, reasoning and problem-solving power in machines, it is a difficult and tedious task [1].

#### C. Convolutional Neural Network (CNN)

A convolutional neural network is a class of deep, feed-forward artificial neural networks that have successfully been applied to analyse the visual image. CNNs use a multilayer perceptron’s to attain minimal pre-processing. The deep convolutional neural community can gain practical overall performance on challenging visible consciousness tasks, matching or exceeding human performance in some domains. This community that we build is a very small community that can run on a CPU and on GPU properly.

The images are processed in such a manner that the faces are nearly captures the main face parts of the person, in the emotion recognition that is in AI part. In this project, we used a dataset provided via Kaggle website, that is fer2013 which is an open-source dataset. This dataset consists of 35.887 grayscale, 48x48 sized face images, dataset can be divided into 7-emotions. These facial emotional expressions have been categorised as: 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, and 6=Neutral. Each image must be assigned to one of seven classes that characterize facial expressions.

The given images in the dataset are separated into three different sets: training, validation, and test sets. About 80% of images for training, 20% of images for validation, and testing is given. After reading the raw pixel data, we normalized the raw pixel data by subtracting the average of the training images from each image in the test and validation set. Create a mirror image to strengthen the data by flipping, rotating and zooming the image of the drive assembly (in other words image augmentation). So that we get proper output while giving the real time input.

I.

- Flow chart
-

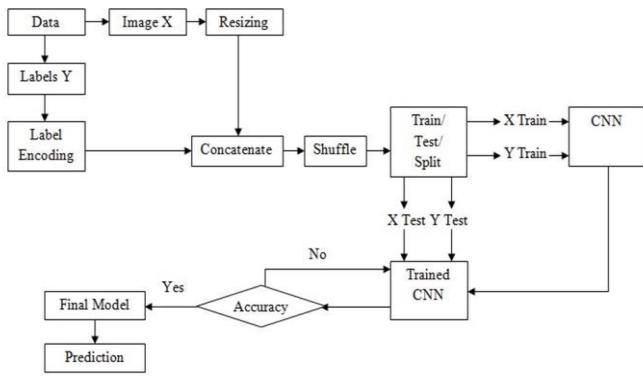


Fig 3. Flowchart Of Blind Assistance Device

In the data the property will be divided as an image x and text y the image will be resized and the text will be encoded then resized images and text will be concatenate and it will be shuffled. The shuffled data will undergo train test splitting process. The x train and y train will undergo with the CNN and x test and y test will undergo trained CNN does CNN will come to train CNN the train CNN will give the accurate result of which we expected. Otherwise, it will undergo the trained CNN if it will not satisfy our expectation and then it will train again and we will get the accurate result it will be a cyclic process then we will get the final model the model will be predicted. Software processes the input image and converted into text format. The software implementation is showed in Figure.

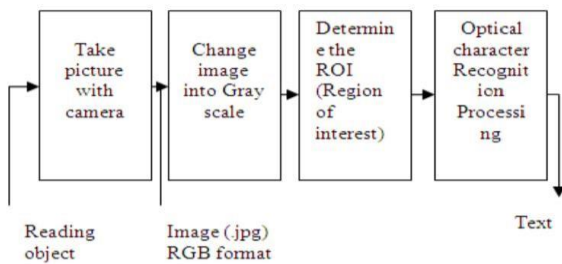


Fig 4. System Design Of Image Processing Modules

In this module text is converted to speech. The output of OCR is the text, which is stored in a file (speech. txt). Pyttsx3 is the text to speech conversion library. The pyttsx3 module supports two voices first is female and the second is male which is provided by “sapi5” for windows. Which is an open-source Text to Speech (TTS) system, which is available in English language. In this project, English language is used for reading the text. After pressing the process button computer. After pressing the process button, the acquired document picture is subjected to OCR technology.

OCR technology allows you to convert scanned images into text. The transformation of printed words or symbols into text or information that can be read. A computer application was used to understand or alter the text. In our project, we use the TESSERACT library for OCR technologies. The data will be transformed to audio using the Text-to-Speech software. The camera serves as the primary vision for detecting the picture of the placed

document, after which the image is internally processed to remove the label from the image using the open CV library, and finally the text is identified using speech. Now, the audio output of the converted text can be listened to using headphones connected to the computer.

III.CONCLUSION

Artificial intelligence Technology is advancing rapidly every day, improving other aspects of the lives of people with disabilities, autism, the elderly, the visually impaired and other people in need of care. To help these people, machines must understand the senses on their own and have the empathy to solve problems. Our artificial intelligence researchers and experts are trying to make the devices or machines real. There is no doubt that in the coming decades the machines we visit will readily understand the human senses and solve problems accordingly. At this time, psychotherapy and the diagnosis of other chronic psychological problems are often dealt with mechanically. The machine is like a great assistant for doctors and can help them. We designed a very unique text recognition and emotion detection device to extract text areas from smart backgrounds. Equivalence feature maps estimate the global structural features of text at the component level. We proposed by training and tested a common architectural project for generating CNNs in real time. We start by completely removing the fully connected layers and reducing the number of parameters in the remaining convolutional layers.

REFERENCES

- [1] Ashwani Kumar, Ankush Chourasia “Blind Navigation System Using Artificial Intelligence” Dept. of Electronics and Communication Engineering, IMS Engineering College, INDIA
- [2] Ms. Rupali, D Dharmale, Dr. P.V. Ingole, "Text Detection and Recognition with Speech Output for Visually Challenged Person", vol. 5, Issue 1, January 2016
- [3] Nagaraja, L., et al. "Vision based text recognition using raspberry PI." National Conference on Power Systems, Industrial Automation (NCPSIA 2015).
- [4] Ezaki, Nobuo, et al. "Improved text-detection methods for a camera-based text reading system for visually impaired persons." Eighth International Conference on Document Analysis and Recognition (ICDAR'05). IEEE, 2005.
- [5] Ms. Kavya. S ,Ms. Swathi Scholar “Assistance System for Visually Impaired using
- [6] Padma Shneha1, Prathyusha Reddy2, V.M.Megala3 “Artificial Intelligence For Vision Impaired People “International Journal Of Latest Trends In Engineering And Technology Special Issue April-2018, Pp. 031-036 E-ISSN:2278-621X
- [7] Ani R1, Effy Maria2, J Jameema Joyce3, Sakkaravarthy V4, Dr.M.A.Raja “Smart Specs: Voice Assisted Text Reading System For Visually Impaired Persons Using TTS Method” IEEE International Conference On Innovations In Green Energy And Healthcare Technologies(ICIGEHT'17)
- [8] Xiaofang Jin, Ying Xu “Research on Facial Expression Recognition Based on Deep Learning” Authorized Licensed Use Limited To: University of Canberra. Downloaded On June 07,2020 at 15:35:56 UTC From IEEE Xplore.
- [9] Ashwani Kumar, S Sai Satyanarayana Reddy, Vivek Kulkarni “An Object Detection Technique for Blind People in Real-Time Using Deep Neural Network” 2019
- [10] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of facial expression extracted automatically from video. Image and Vision Computing, 24(6), 2006.
- [11] P. Ekman, W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press, 1978

- [12] Amsaveni, M and Anurupa, A and Preetha, RS Anu and Malarvizhi, C and Gunasekaran, Mr Gsm based LPG leakage detection and controlling system, The International Journal Of Engineering And Science (IJES) ISSN (e) 2015
- [13] Cohen, Ira, et al. "Evaluation of expression recognition techniques." Image and Video Retrieval. Springer Berlin Heidelberg, 2003. 184- 195.