

Development of a Smart Device for Blind & Visually Impaired Person using Computer Vision

¹Mohit Sarin, ²Aayush Sharma, ³Sumit Kumar Sar

¹ Student (B.Tech), Electronics & Telecommunication, NIT Raipur ²Student (B.E), Computer Science, BIT Durg

³Associate Professor, Computer Science, BIT Durg

¹Electronics and Telecommunication Department,

¹National Institute of Technology, Raipur, Chhattisgarh, India

Abstract: *Developing a tool for the visually impaired people is a trivial area of concern. However, use of computer vision in these devices is a fairly new trend and recent topic of interest. There are several works using computer vision techniques which focus on specific requirements of blind people. But there is no existing methods that help to solve all the basic needs of blind person. In this paper, we propose an idea of a device which can capture images in real time and comprehend its equivalent speech directly into the ear of visually impaired person. This new system may solve some of major problems of blind persons that are still exists. The prototype facilitates object detection using image recognition techniques. Open Computer Vision (OpenCV) performs image analysis, classification of images and prediction of the type of object on household object dataset. This prediction is converted to audio output using gTTS (Google Text to Speech Converter). This audio output can be fed to earphones connected to the device. A computer vision technology device provides greater accessibility for those with vision impairments. The results produced by this device will be more accurate because of the novelty of the machine learning algorithms used in this.*

Index Terms - Object Detection, Image Recognition, gTTS, OpenCV, Computer Vision, Blind navigation.

Introduction

Blindness and visual impairment confines the ability to move around freely and in an uninterrupted manner. The disability significantly interferes with one's ability to function independently and to perform daily activities safely and easily. Navigation hindrances become a real challenge most especially for persons with eye-related disabilities. Traditional means of navigation included guide dogs, pointing sticks and walking canes.

The walking cane is a simple and purely an object dedicated to detect static obstacles on the ground, uneven surfaces, holes and steps [1]. This device is light, portable, but range is limited to its own size and not suitable for dynamic obstacles. Based on the interdependence between the disabled owner and a dog, the training and the relationship to the animal is another way for the blinds to walk around. The dog is able to detect and analyze complex situations like stairs, detect potential danger and know paths [2]. The blind is able to feel the attitude of his dog, analyze the situation and also give orders. But guide dogs are still far from being affordable, and their average living is seven years. The ideal correction of blindness and visual impairment may never be attained which consequence in an-over-or under-correction of vision resulted from expensive eye surgeries and transplants. The most optimum solution to all these problems is thus the use of modern technologies such as computer vision.

1.1 Objectives and Scope of Study

According to an article, people who have been blind from birth make use of the visual parts of their brain to refine their sensation of sound and touch, according to an international team of researchers led by neuroscientists at Georgetown University Medical Center (GUMC). The researchers selected computer vision as an area of research to address the needs of the blind and visually impaired individuals. Various hardware and software components made up the system.

The main objective of this study is to develop a prototype that will serve as assistance to the blind and visually impaired. The study emphasizes on the novelty of machine learning algorithms on which the data set is trained. Here, we are performing object detection on household objects only. The main task performed by this prototype would be detection of object through real time image capture & processing, object recognition and spelling the name of object directly into the ear of the visually impaired person.

1.2 Significance of the Study

Estimating the magnitude and geographical distribution of visual impairment and causes is an important work to enable an appropriate allocation of resources across the various health areas of work.

Nowadays, most of the commercial solutions for visually impaired localisation and navigation assistance are based on the Global Positioning System (GPS). However, these solutions are not suitable for the visually impaired community mainly due to low accuracy, signal loss and the impossibility to work on indoor environments. Moreover, GPS cannot provide local information about the obstacles in front of or in the near surroundings of the person [3]. Furthermore, other commercial products available in the market present limited functionalities have low scientific value and are not widely accepted by the users.

Computer vision-based approaches offer substantial advantages with respect to those systems and constitute a promising alternative to address these problems. But till now there is no existing tool that satisfies major needs of blind persons. The goal of this prototype is to combine the technologies of recent interests such as image processing, object recognition and text to speech conversion and produce a viable alternative to all the traditional solutions [4]. Architecture of this prototype is shown (Figure 1).

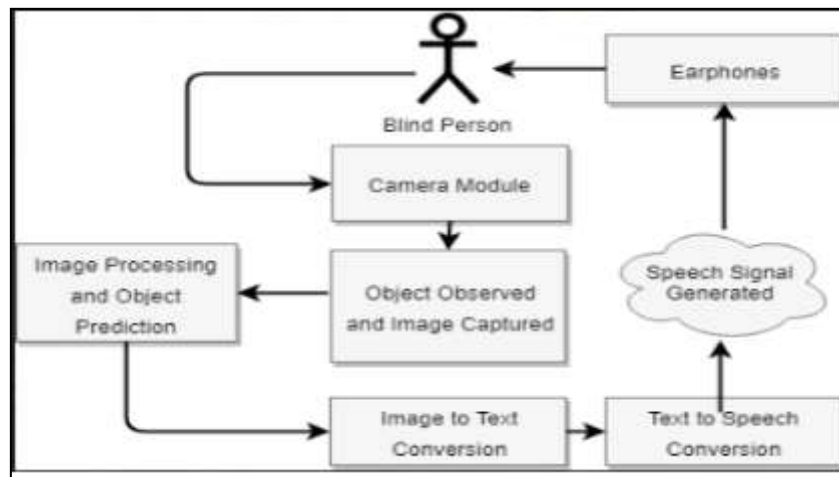


Fig. -1: Architecture of Prototype Device

I. DESCRIPTION OF DEVICE

The device will consist of JPEG Camera, AVR processor, Bluetooth module, Polymer Lithium-ion battery, and push button switch. These components will be mounted on an assembly and will be placed in front of eye as a wearable cap with using head support. The core concept is mainly based on the Object Detection and Identification using state-of-the-art machine learning algorithms and application of Text to Speech Synthesis technologies.

The computer vision technology is based on extraction of information from images or videos. This field includes methods for acquiring, processing, analyzing, and understanding images and, in general, high dimensional data from the real world in order to produce numerical or symbolic information.

The proposed device will firstly capture an image and this image will be subjected to computer vision algorithms to extract its detailing and information from it. This information extracted from image set will be utilized to predict the type of object. Following this process would be a comparison of the predicted object with its text name. Once the text is determined, this will be fed to the text to speech synthesizer through which the output will be extracted.

The text-to-speech synthesizer is used to provide audio feedback and is also a less disruptive interaction for sighted and visually impaired people. When used for the first time, the device must be paired with a smart phone; however, this is done only once and henceforth a Bluetooth connection will be automatically established. Bearing in mind that the device should support both sighted and visually impaired users, we completely rely on non-visual interaction for all usage of the system. The process flow representation is described (Figure 2).

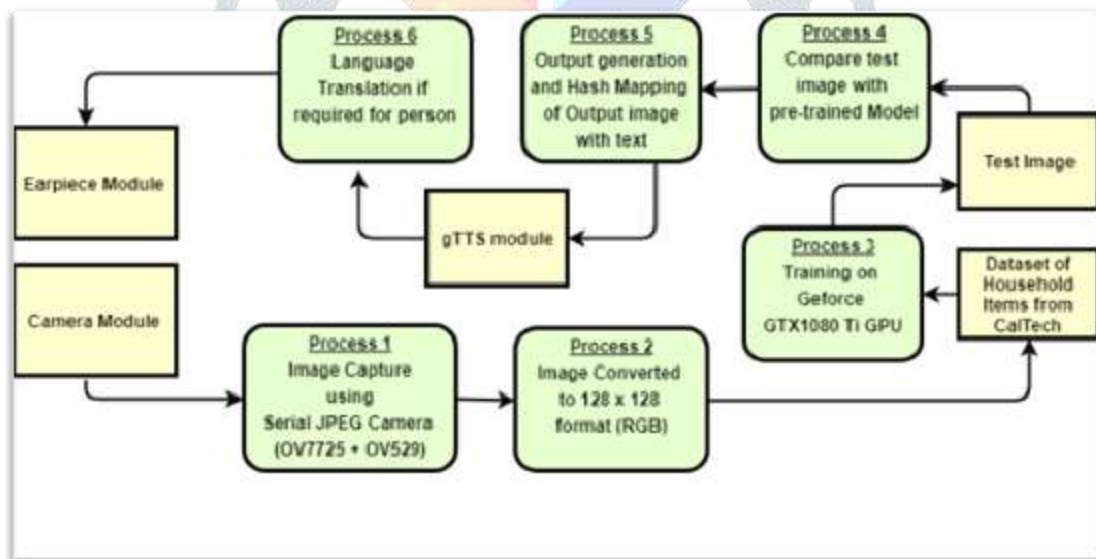


Fig. -2: Process Flow Diagram

2.1 Hardware Design

The hardware components to be used in the device would broadly include a micro-controller unit, an image acquisition module and a wireless transmission module. The description of their usage is described below.

2.1.1 Micro-controller

An Atmel 8-bit AVR (ATmega32U) can be used in the device. The device will require only basic peripherals. The digital i/o pin is configured as an i/p and connected to a push button for user interaction.

Two UART communication (universal asynchronous receiver/transmitter)- An image acquisition module and Bluetooth communication channel setup is required. UART is type of embedded interface. A UART is usually an individual (or part of an) integrated circuit used for serial communications over a computer or peripheral device serial port.

2.1.2 Image Acquisition Module

Image acquisition module uses the C329 UART JPEG compression module (Lossy compression) which acts as a video camera. User can give the snapshot command to capture full image with the help of OV7725 VGA CMOS sensor and the OV529 JPEG engine. This module has a power saving mode characterized adjustable resolution.

2.1.3 Wireless Transmission Module

RN-42 Bluetooth module is utilized in the prototype device. Bluetooth is wireless technology standardized as IEEE 802.15.1 which is used for exchanging data over very short distance. Having a frequency spectrum of 2.4 to 2.485 giga hertz (GHz), so the wireless communication between the device and mobile device is done through a Roving Networks. RN42 Bluetooth module is low power, high performance with shorter length and delivers up to a 3 mega bytes per second data rate for distances up to 20 meters.

2.2 Software Design

The software technology consists of an established connection with a mobile device or computers with a modern operating system. It also consists of a gTTS (Google Text to Speech Converter), which will convert the name of image(text) to audio signal which will be fed into the ears of the disabled person through earphones.

The device also uses YOLO/YOLO version 3 (You only look once) object detection algorithm using which the object is predicted and its name is stored as text which is then converted to audio signals. This algorithm is implemented as it highly efficient and reliable[5]. The Software design used in this prototype provides an ease for development of a variety of other applications.

The prototype designed is currently operable in household conditions only. The algorithm used is trained on home object dataset from CALTECH[6]. This can be further enhanced to function on other environment as well.

II. WORKING

The user needs to pair the device with the mobile phone application only once and henceforth a Bluetooth connection will be automatically established when both are running. After the device is switched on, it will start sending images on a real time basis. When a snapshot is taken from the camera and the image is transferred via Bluetooth to the mobile phone or Computer, an application starts analyzing the image using computer vision engine. Type of analysis and response depends on the preset mode (color, distance, currency, etc.). Upon analyzing the image data, the application uses a Text-to- Speech module to read out the information through a headset.

The YOLO algorithm used applies a single neural network to the full image. This network divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities [5]. A bounding box model of an object detected from an image is represented (Figure 3)[8]

YOLO looks at the whole image at test time so its predictions are informed by global context in the image. It also makes predictions with a single network evaluation unlike systems like R-CNN which require thousands for a single image. This makes it extremely fast, more than 1000x faster than R-CNN and 100x faster than Fast R-CNN. We ensure the novelty of the algorithm as previous results shows that the speed and accuracy of YOLO v3 is high as comparison to other object detection algorithm both for single and multiple object detection in a given image

These detected objects are then mapped with their corresponding text names to be fed into the gTTS (Google Text to Speech) module. The final output is the audio output of the name of the detected object.



Fig. -3: Bounding Box of detected Object

III. APPLICATIONS AND FUTURE SCOPE

The design of this prototype can be used in a variety other applications which are described below

1. It can be improved to act as a shopping assistant for price tag recognition. A user would simply touch point index finger to a currency note (TP gesture) and click the button. The system will voice the value.
2. It can also be converted to a finger –worn device which can read non- Braille text, and can directly convert text to speech.
3. It can also be used as an augmented and mixed reality tool to empower children in pre-reading stage (read text on their own).
4. Color Detector: This application aids a visually impaired person to understand the color of an object. The user simply touch point (TP gesture) to an object and click the button to deliver an image for processing. The system analyses the image and returns the average color via audio feedback [7].

REFERENCES

- [1] Shah, C.Bouzit, M., Youssef, M. & Vasquez, L. Evaluation of RU-Netra – Tactile feedback navigation system for the visually impaired. Page(s): 72-77 ISBN: 1-4244-0280-8
- [2] Rodrigo C. Belleza, Jr., Roselyn A. Maaño, Karl Patrick E. Camota, Darwin Kim Q. Bulawan -Development of a Computer Vision System for the Blind and Visually Impaired Person. World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering Vol:7, No:11, 2013.
- [3] Sujith B, Safeeda V -Computer Vision-Based Aid for the Visually Impaired Persons- A Survey And Proposing New Framework. IJIRCCE- Vol.2, Special Issue 5, October 2014.
- [4] F. van der Heijden, P.P.L. Regtien -Wearable navigation assistance - a tool for the blind ,Faculty of Electrical Engineering, Mathematics and Computer Science University of Twente. MEASUREMENT SCIENCE REVIEW, Volume 5, Section 2, 2005
- [5] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, University of Washington, Allen Institute for AI, Facebook AI Research-pjreddie.com/darknet/yolo
- [6] Griffin, Gregory and Holub, Alex and Perona, Pietro (2007) Caltech-256 Object Category Dataset. California Institute of Technology
- [7] Suranga Nanayakkara, Roy Shilkrot, Patricia Maes-EyeRing: An Eye on a Finger, CHI'12, May 5–10, 2012, Austin, Texas, USA.
- [8] Ammema Zainab-medium.com/mindorks/detection-on-android-using-tensorflow-a3f6fe423349

