Personal Assistant for Visually Impaired Using Neural Networks

[1] T.Sriharsha, [2] J.Vikram, [3] Subramaniam A, [4] J.Dolly Irene

[1 2 3] UG Student,[4] Assistant Professor

[1] Dept. of Electronics and Communication Engineering,

[1] Prathyusha Engineering College, Chennai, India.

Abstract : An assistive system is one which can be used to serve in many applications for the people with physical impairments to alleviate their sufferings in their day-to-day lives. For the visually challenged community it is tedious to find an object of interest in the surroundings and the time taken for such process is too long and is very difficult to deal in unknown environments as they are unaware of the location of the desired object. To overcome this problem, we present a “Neural networks” based approach called “**Automatic Object Detection System**” which can detect the object requested by the user and the system can navigate the user to the object. This is achieved by processing frame by frame video information, of the surroundings, through a “**Convolutional neural network(CNN)**” and obtaining the spatial information which can be prompted back to the user via audio feedback. Our system is a non-haptic complete audio controlled and is not just limited to objects but can also detect and recognize faces of known people and can recognize different types of Indian Currency. The system provides some personal assistance functionalities such as time keeping, note taking, gathering information, such as weather,news,directions from place-to-place ,from the internet using web scraping techniques. On a whole our system can be considered as a pocket sized virtual eye powered by a raspberry pi.

***IndexTerms* - Convolutional Neural Networks, Movidius Vision Processing Unit, Python, Tensorflow, Web-scraping.**

I. Introduction

A study by lancet global health indicates that around 253 million people across the globe live with visual impairment in which 36 million people are estimated to be blind this number seems to triple to touch 115 million by 2050, where the most affected will be from the developing nations. In India, in the span of 25 years (1990-2015) the count of blind people rose from 7.2 million to 8.8 million constituting a 25% of the whole blind community around the world.

The contribution of the development community to the society has been focused on a number of developmental prospects in the fields of healthcare and education. It's important not to forget that if one cannot see, then there is a lot he/she can't do even with the best education.

In general, the term “assistive technology” is used where users, who are disabled require some form of assistance. One of the most challenging and important tasks in creating such technologies is to develop a device that is best suited for the potential of blind users, both in the aspects of the user providing the input and interpreting output/ feedback of the device.

The main objective is to provide a system for detection and avoidance of obstacles that assists visually impaired/disabled persons to reduce their risk of collision in their movement. Artificial intelligence and neural networks have a key role in the adaptive learning of the dynamically changing known or unknown environment. The application of hardware implemented artificial neural networks for achieving an intelligent feedback device which is then used to communicate results effectively to the user is proposed.

The proposed system employs a camera for real-time object detection whose main objective is to determine all the object instances of the any desired object from user regardless of scale,view and position with the camera.

In the following sections of this paper we briefly discuss the techniques and methodologies implemented in the design of the proposed system.

II. Literature Survey

2.1 Guiding (or) Smart canes

Ayat A. Nada et.al. [1] in 2015 proposed a smart cane based system for the effective navigation of the visually challenged people. There are many more systems as this one but they are not completely satisfactory when tested.These systems come equipped with distance detecting sensors and ultrasonic range finders which are mere extensions for the basic white cane and does not provide much spatial information regarding the environment. They just alert the user if there is an obstacle which the existing canes can be done by bumping into the obstacle.These systems are the pioneers in aiding systems for the visually challenged people.

2.2 RFID based detection systems

Jalila Al Kalbani et.al. [2] in 2015 developed a **Radio frequency identification(RFID)** based Bus detection system in the same theme of developing assistive systems for the blind. Many systems similar to this one do exist and are considered for specific applications.

These systems are better only in known environments and predefined areas where RFID tags are placed everywhere but these systems are not widely used and difficult to apply in real life situations and people with visual disabilities can adapt easily to a known environment with the help of their enhanced senses besides RFID is a complex mechanism to be maintained by a blind person. Sometimes these systems are not too compact to carry around easily.

2.3 PLD based navigation systems

As proposed by Harsha Gawari et.al. [3] in 2014, **Position locating devices(PLD)** are those which identify the position of the user based on Global Positioning System and they help the users to navigate from one place to another. This puts the blind user into considerable disadvantage compared to sighted individuals navigating through familiar environments as this system can only help in navigation but it is difficult to identify/detect objects/obstacles which makes it unsuited for safe navigation for the blind community. This system can be helpful if properly integrated into a helpful assistive system.

Most of the solutions in assistive systems for people with improper visual conditions are not the systems that can detect and recognize daily use objects and to overcome this problem we propose our system.

III. Proposed System Overview

3.1 Prerequisites

3.1.1 Hardware Requirements

a. Raspberry pi Model 3B
b. Memory Card
c. Webcam with integrated Microphone
d. Power Source
e. Movidius Neural Compute Stick
f. Headphones

3.1.2 Software requirements

a. Raspbian Stretch Operating System
b. Python 3.4 +
c. Tensorflow 1.4.0 +
d. NCSDK 1.11.0 +
e. OPENCV 3.3.0 +
f. Caffe
g. Putty
h. VNC server

3.1.3 Setting up the raspberry pi

Our system has been configured to accelerate the video processing and make the detections faster.This system runs on raspbian stretch installed headless and "**Opencv**" installed is optimized for **neon** and **vfpv3** floating point operations and this will provide a 30% increase in processing speed compared to normal opencv installations.Similarly ncsdk is installed in api only mode since it is computationally expensive to make a full-blown installation on a raspberry pi.The model is retrained on a GPU and compiled on a ubuntu based system and deployed on the raspberry pi in a graph format.

As the raspberry pi's compute power which is 1GB RAM is not enough for a real time processing.So to accelerate this process we use Intel's **Movidius Neural Compute Stick(NCS)** which uses a unique "**Vision Processing Unit**"**(VPU)**.This Vision processing unit consists of shave processors and Intelligent memory fabric which increases the processing **Frames per second(FPS)** by 10 folds approximately.The reason for choosing a Vision processing unit is to perform quasi real-time processing on the system.Alternatively the capturing can be done on raspberry and processing can be done on a different platform but the system has to be online all the time and data transfer rates must be very high.Unfortunately the latency exists to transfer the frames to be processed to a server and all the time taken for such process is high too. The requirement that the system has to be online all the time eliminates the working of that system at places where the reception is low. In the long run the costs of online processing will definitely exceed the cost of movidius compute stick.

3.2 Block Diagram

The Block diagram for the proposed system can be observed in Fig.1

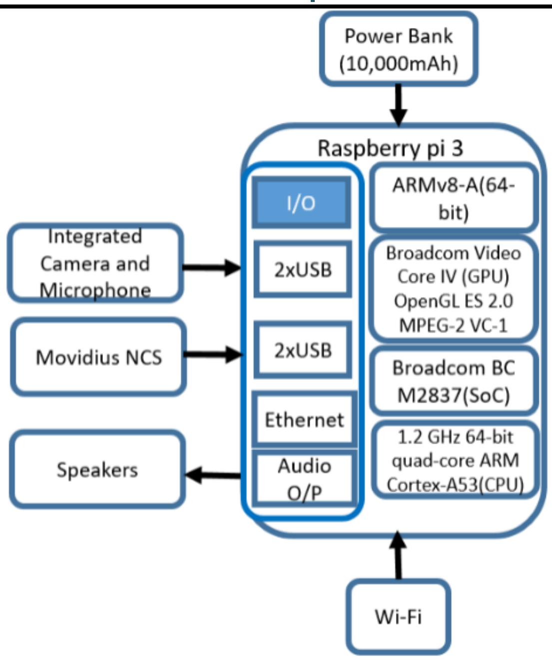

Figure 1. Block Diagram of the proposed system

3.3 Workflow

The workflow of the system is represented in Fig.2.As it is clearly depicted the system sits idle until the button press happens and using Google's speech to text api the user's request is converted into string.Using natural language processing we identify the user's command and the functionality for that command is triggered.

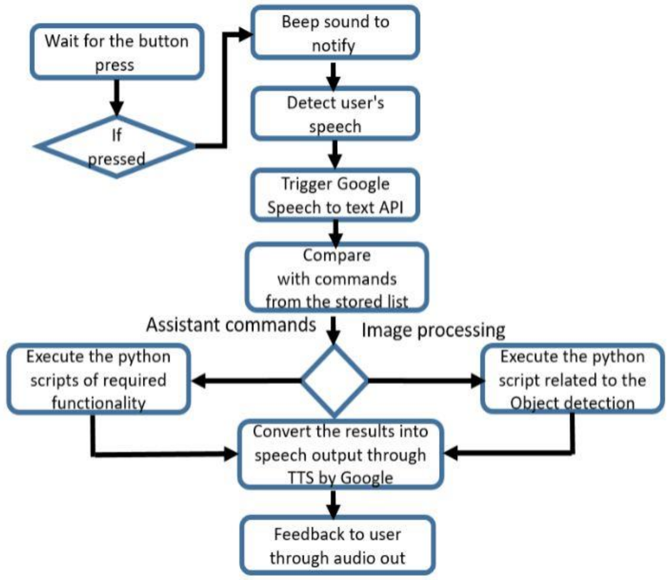

Figure 2. Flow Process of the proposed system

There are two different types of functionalities are being provided

a. Image Processing
b. Personal Assistant

3.3.1 Image Processing

In the image processing functionality the system can perform three different types of image processing as shown in Fig.3

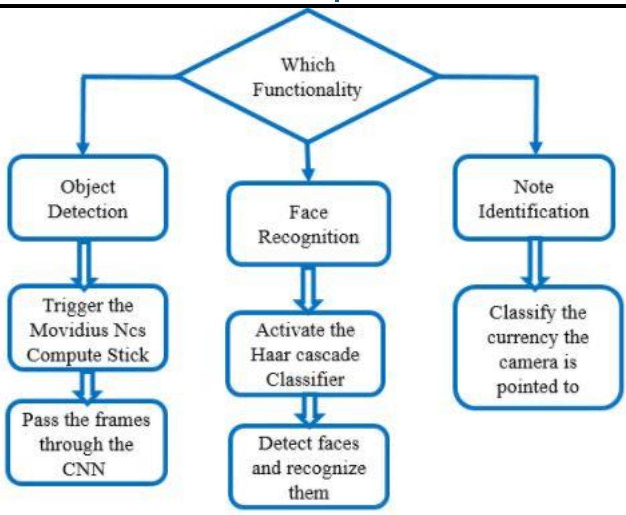

Figure 3. Image Processing Functionalities

(i) CNN based Object detection.
(ii) Haar Cascade based Face Detection and recognition.
(iii) **Scale Invariant Feature transform (SIFT)** based feature matching for currency detection.

(i) CNN based Object detection

For the Visually impaired people it is a difficult task to determine or locate an object unless they have a physical contact with that object and this is a very time-consuming process.In recent times there have been a lot of advancements in the fields of artificial intelligence and Visual Computing based deep learning applications and due to advent of GPUs faster training of the networks is also possible.

Before the era of convolutional neural networks all the features are hand extracted and the filters for the kernels are manually calculated .This process is tedious and needs a lot of technically able human labour which is extremely difficult to employ or find. Convolutional Neural networks started out as being a replacement to deeper neural nets but they turned out to be better feature extractors i.e, instead of hand labelling data or images the network itself understands and identifies the most evident object in the frame. This way CNNs became the go-to approach for any visual processing applications based on neural networks.

In our system, we used Tensorflow's Object detection API along with a CNN model called **SSD(Single shot detector)** which is trained by Google brain team for the **ILSVRC(ImageNet Large Scale Visual Recognition Challenge)**.This model is trained to detect daily use objects such as chair,car,motor cycle,mobile etc.,also this model has been trained for some objects which are not desired for this application so our system has been fine tuned to utilize only 20 objects which improves the detection rate and accuracy of our system to those particular objects.

Transfer learning is the main concept behind this system's operation.The profiling and compiling for the NCS was done on a virtualbox UBUNTU 16.04 operating system where the system was retrained for around 8 hours and the model was generated which is ready to be deployed on any system like an edge device or an embedded system like the one which was used in this project.

(ii) Haar Cascades based Face Detection and recognition

Recognizing the people around them is a socially important aspect for the blind people.In this system we have implemented a deep learning based approach to basically train known people's faces and to recognize them whenever they are spotted in the frame.

(iii) SIFT based Feature Matching for currency detection

The most difficult task for the blind is to recognize the currency for almost everything money is involved. Our system has a database of all the different kinds of Indian Currency front and back faced so that the user can point any currency to the camera and it can identify and convey the user which currency that the system is pointed at.Scale Invariant Feature transform as the name itself suggests this technique is robust and can efficiently identify and match features at faster processing rates.

3.3.2 Personal Assistant

The personal assistant functionalities are introduced in this system to improve the quality of life.There are several services in the personal assistant functionality some of them are

(i) Timekeeping

With the help of python's "time" and "calendar" modules one can replicate almost any calendar related functionalities like reminding us on specific dates etc.,

(ii) Note taking

By integrating SQLite database with python the user can ask the system to note down shopping list or some personal lists,Phone numbers with names etc.,

(iii) News

The user can get to know about the recent happenings across the globe with web scraping techniques

(iv) Weather

The user can get weather updates so that they can decide whether to go with safety measures or just to stay in home.

(v) Directions

Integrating Google's MAPS api to geo locate the user based on IP address and thereby navigating the user to his desired position without the use of any GPS module.

IV. RESULT AND ANALYSIS

All the results that are obtained by the system are audio outputs and the images represented below depict how the system perceives the frames after processing and before converting them to speech to feedback to the user.

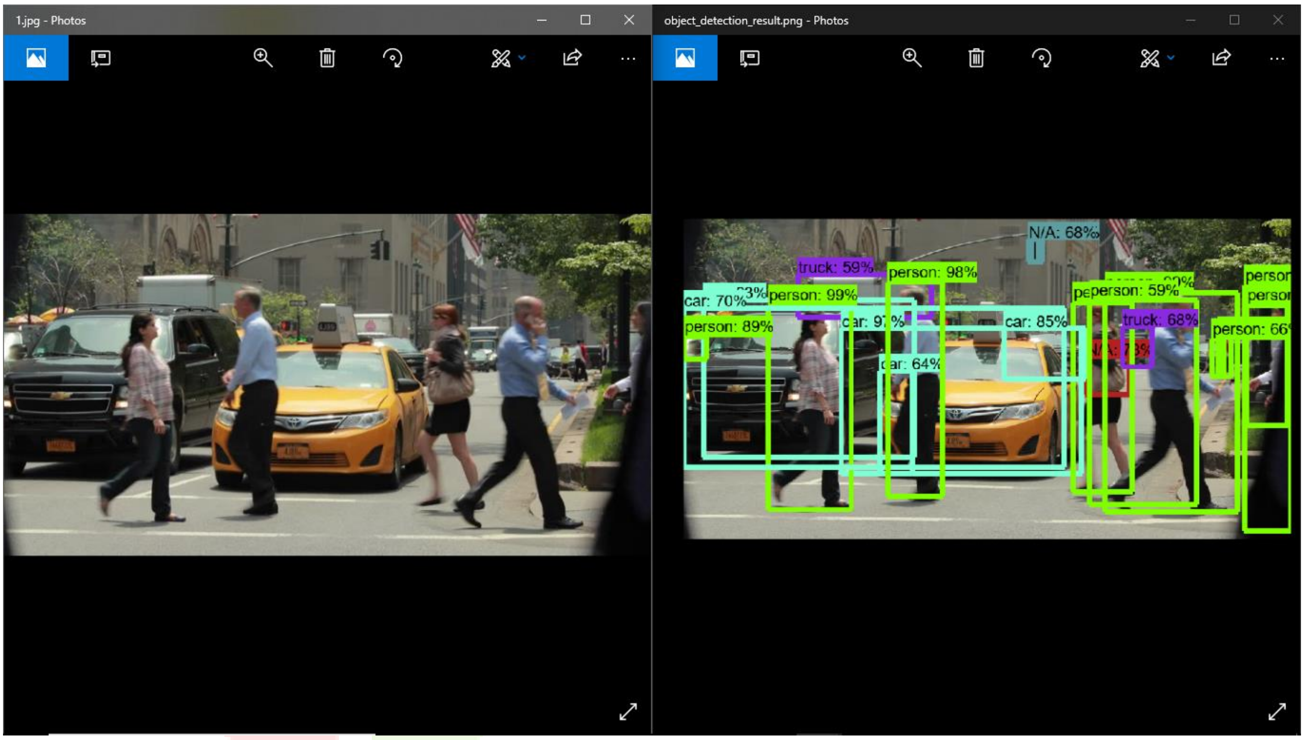

Figure 4. Objects recognized in the applied frame

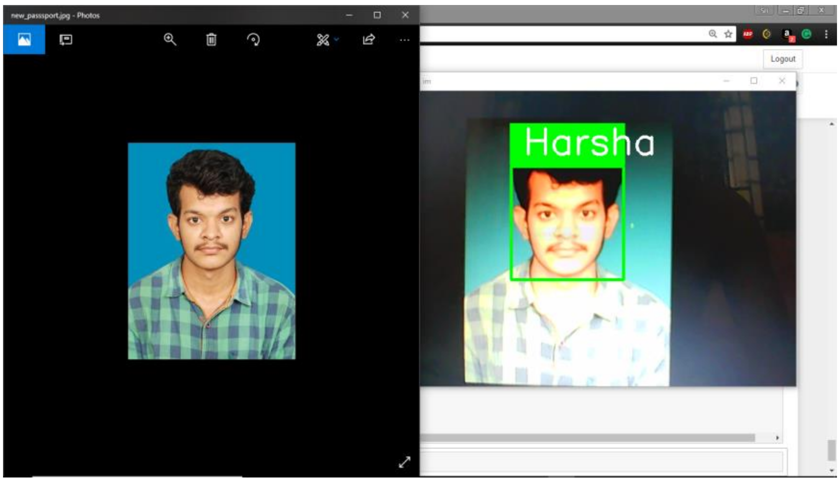

Figure 5. Face detected and recognized

Figure 6. Currency being detected

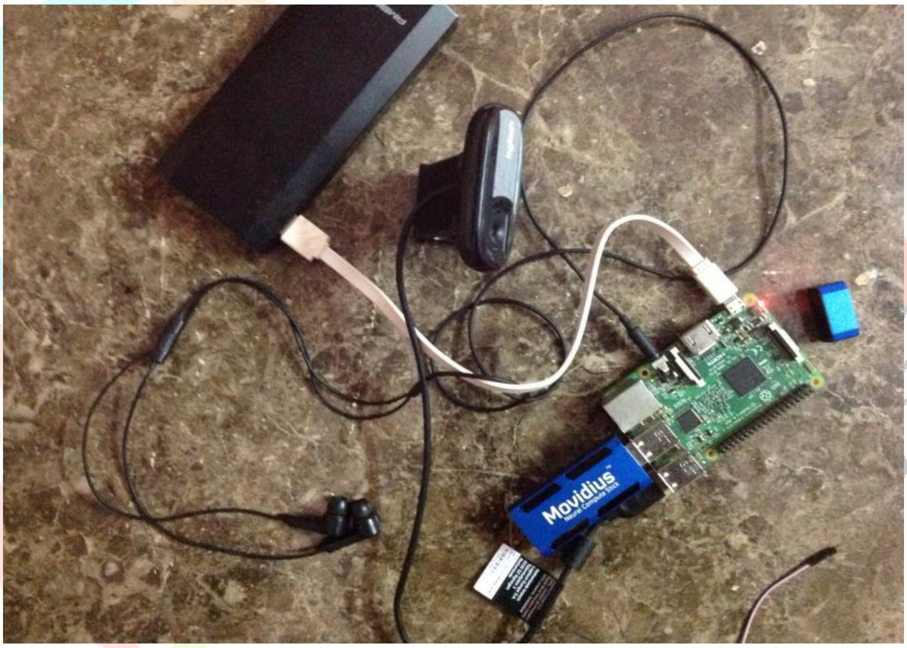

Figure 7 Setup of the Project

In Fig 4 the system detects the objects and forms bounding boxes around all the objects detected with the help of those coordinates audio prompts are generated to direct the user towards the required object.As we can notice the accuracy of the system is around 70~80% which is quite acceptable benchmark considering the inference system is a raspberry pi and the frame rate is in the terms of 16~19 FPS detecting simultaneously 20 different objects. In Fig 5 we can observe the system detecting and recognizing the known people's faces.This system provides a easy way of setting up a new face for recognizing and training happens on the go. In Fig 6 the system detects different type of currency simultaneously without any delay.The setup of the project can be observed in Fig 7.The system is made compact such that it is made to fit into a small box which can be handheld.

V. CONCLUSION AND FUTURE SCOPE

The project proposed in this paper caters the needs of the visually impaired community in their day-to-day activities.The important feature of this project is that the time needed for a blind person to locate the required object is reduced to a notable extent. The proposed system is limited to detecting objects but not obstacles in the path of navigation.

The working of the CNN is limited to the areas where proper lighting conditions prevail and the frame processing rate is less considering to a normal personal computer since tensorflow on raspberry pi is not an official build all the computing optimizations are not available making it slow.In near future, one can expect an official tensorflow build on embedded devices which can drastically alter the processing rates.Extending this system with additional components such as a GPS receiver or some infrared sensors etc. can help in better and safe navigation. Thus there are obvious advancements possible to improvise this system.

VI. ACKNOWLEDGMENT

The Authors of this paper would like to sincerely thank all of them who have provided their valuable suggestions and guidance in successful completion of this project.We wish to express our sincere gratitude to parents for valuable help,co-operation and moral support during the time of completion of this project.

REFERENCES

[1] Ayat A. Nada, Mahmoud A. Fakhr, Ahmed F. Seddik "**Assistive Infrared Sensor-Based Smart Stick for Blind People**" Science and Information Conference 2015 July 28-30.

[2] Jalila Al Kalbani,et.al."**Bus Detection System for Blind People using RFID**" Proceedings of the 8th IEEE GCC Conference and Exhibition, Muscat, Oman, 1-4 February, 2015.

[3] Harsha Gawari, Prof.Meeta Bakuli,"**Voice and GPS Based Navigation System for Visually Impaired**" International Journal of Engineering Trends and Technology (IJETT) – Volume 11 Number 6 - May 2014.

[4] Wei Liu Dragomir Anguelov, Dumitru Erhan,et.al. **"SSD: Single Shot MultiBox Detector**" ECCV 2016,29 Dec 2016.

[5] Jonathan Huang, Vivek, Google Research,et.al. "**Speed/accuracy trade-offs for modern convolutional object detectors**".

[6] Vincent Vanhoucke, Ian Goodfellow, Google Research Team,Greg S,et.al. "**TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems"** (Preliminary White Paper, November 9, 2015).

BOOK : GOODFELLOW -ET-AL**,"DEEP LEARNING"(**MIT PRESS,2016).