JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Implementation of Smart Cap for Visually Impaired Person using Raspberry Pi

[1]Nynaru. Krishna Priya, [2]Dr. S. Swarna Latha
[1]M. Tech Student, Department of E.C.E S.V. University College of Engineering, Tirupati, Andhra Pradesh, India.
[2] Associate Professor, Department of E.C.E, S.V. University College of Engineering, Tirupati, Andhra Pradesh, India

Abstract: **Most communications in today's culture are performed via speech as well as text. The purpose of the project is to offer assistive technology to help visually impaired persons in emergency situations. Our paper's purpose is to design a cap for blinds that will guide blind people from point A to B. Smart Cap's approach would be to aid visually challenged people, and it's a low-cost wearable smart cap. Smart Cap will be a conversational interface that integrates the Internet of Things with Deep Learning, with capabilities such as facial recognition, item detection and identification, textual recognition and detection, and online reading of newspaper. The hardware configuration includes a Raspberry Pi, earbuds, USB microphone, USB camera, power source, as well as extension cords. The user may interact with the Smart Cap by providing commands to activate the corresponding feature module, which provides an audio outcome. The face detection and identification module depend on OpenCV (Open-Source Computer Vision Library) face recognition, while real-time object detection would be provided by the YOLO (You Only Look Once) technique. For text detection and identification, the Optical Character Recognition (OCR) Online API service is employed. There is also the option of reading online newspapers, which keeps a blind person informed of the latest events.**

Index Terms: **– Raspberry Pi, web Camera, Face Recognition, Object Detection, Text Recognition, OCR, Online Newspaper.**

I. INTRODUCTION

Visually handicapped people are either completely or partly blind. As per the World Health Organization (WHO), 285 million individuals have vision impairment, 39 million individuals are blind, and around 3% of all age groups are vision impaired in a nation. Tumors, trachoma, glaucoma, and deficiencies such as Vitamin A, onchocerciasis, and leprosy are the primary causes of blindness. People who are blind or visually impaired face numerous challenges in their everyday lives, such as navigating unfamiliar streets and following directions to unfamiliar areas. Eyeglasses, for example, cannot heal or enhance eyesight. We attempted to remove this black dot using Machine Learning and Artificial Intelligence Learning. The three digital pillars of today's industry are IoT, cloud services, and artificial intelligence. Using all these technologies analysis of the data has become increasingly easier.

The Smart Cap presented in this research article would assist visually impaired individuals in several ways, including identifying objects, recognizing known faces, reading aloud papers, and giving recent data via a newspaper website.

Face recognition is mostly dependent on OpenCV face recognition as well as image deblurring. This divides pixels in an image to squares according to their function by using Haar Cascade classifier, that is dependent on the Haar Wavelet method. The identified "features" are calculated employing "integral image" principles. The Ada-boost learning method is employed in Haar Cascades, which selects a small number of appropriate attributes from a large collection to generate an effective result of classifiers, that are then employed to recognize faces in images utilizing cascade approaches. Finally, the obtained facial feature vector would be compared to the information within user's database for detecting a face.

Object detection is the process of detecting, recognizing, and locating objects in a photograph. The image categorization approach is critical in digital picture analysis. Images will be allocated to class labels during image classification. It refers to the steps of processing the input (such as an image) and outputting a class label, such as "cat," or displaying the probability that the input image matches to a class (the input has a 90% chance of being a "cat"). As a result, pixels within the image will be assigned to groups or classes of relevance. It is generally seen as a number-to-symbol mapping process. The YOLOv3 method is used in the Object Detection module. You Only Look Once, or YOLO, is among the finest real-time object identification algorithms (45 frames a second), and it employs a neural network to forecast bounding box coordinates and their possibilities throughout the whole image.

Text detection and identification are implemented using an online API. By combining standard deviation, z-score, as well as k-mean clustering, the results have enhanced even more. Whenever the user wants to utilize any of the features, he provides voice input to enable the associated module. In accordance with the instructions, the Smart Cap camera captures an image and transfers it as input to the appropriate component. These four modules produce text which is then converted to audio. As a result, our study has established a simple, minimal cost, and user-friendly smart assistant platform to enhance the everyday routines of the blind and the visually impaired.

II. Related work

Amjed S et al. [1] introduced a smart infrared micro - controller electronic traveling aid (ETA) which emits and reflects infrared rays to examine the current area surrounding the blind person. PIC microprocessor identifies the shape, material, and motion of the barrier then notifies the user of its shape, material, and motion. Furthermore, these two initiatives are just useful for navigation and don't contain the most current cutting-edge algorithms and cloud services.

Nishajith et al. [2] offer a system that employs object recognition and identification in real time to allow the blind to walk freely. The project provides a TensorFlow object classification API using a Raspberry Pi. The eSpeak Textual to Speech Synthesizer (TTS) application is used to convert the identified object's text data into audio speech.

The suggested method is based on visual key point extraction as well as matching. The number of frames and database information are checked, and an effort is made to identify the item in each frame [3]. The audio communicating the item's information is activated whenever an object is detected.

Manduchi R. [4] suggested an experimental study of a sign-dependent navigation system for the visually impaired. To help the blind individual in navigating, they utilize a cell phone camera to recognize certain color markers utilizing sophisticated algorithms for computer vision.

A technique for identifying objects designed for the visually handicapped [5]. This technique will benefit both the visually handicapped and the larger community. It aids in determining the direction of greatest brightness and bright colors. For object recognition, a hybrid approach is presented that combines Artificial Neural Networks with Euclidean Distance measurements.

As the blind fatality rate has continued to increase due to their inability to perceive and control circumstances in congested traffic, the [6] structure serves as a crucial precaution to avoid accidents for blind individuals in high-traffic areas and to warn them by having established a high propensity to detect obstacles and barriers. The technology assists visually challenged persons in reaching their destination by instructing them using Bluetooth speech recognition system.

The research [7] examines the deployment of a Text Read-out system designed for blind people. A page of printed text is accepted by the system. Before categorization, the OCR software placed upon Raspberry Pi analyses it and turns it into a data file, that is then processed to skew corrections, segmentation, as well as feature extraction. Following categorization, the text is typically read using a text-to-speech device placed on Raspberry Pi. Prior to being read out, the results were processed through an audio amplifier. We will employ an ultrasonic sensor to determine the distance between blind people and the barrier, and the measurement will be broadcast via earbuds.

Another technology discussed in [8] is the smart walking stick, which offers health monitoring capabilities. These health monitoring features allow users to determine individual pulse rate and their body temperature. GPS and GSM modules will be embedded inside the stick to track the user's whereabouts and send an alarm message to another individual. The voice alert system on this cane is an additional vital feature. Users will be alerted to any obstacles, pits, wet surfaces, or health issues through an audible alarm. This will improve blind people's mobility since they will become more attentive after hearing voice alerts.

The system's goal is to offer important functionalities for the visually handicapped, such as home automation, as mentioned in [9]. Different equipment, like fans, lights, and so on, that are already connected with the system, may be turned on/off using various movements and IR sensors placed on the user's fingers. For communication between equipment and sensors, RF signals are employed. Other features include GSM Modem two-way connectivity with the system, the Gas Sensor that detects fire, as well as a Temperature Sensor that gauges the surroundings.

Another technology named SVETA which is explained in [10] and the idea of using stereo cameras, computing devices and stereo devices which are built in a helmet. A sonification technique is proposed for translating the disparity image onto stereo musical which incorporates data from the scene that is presented to the user. The sound is delivered to the blind person through stereo earphones.

III. Proposed System

The smart cap would be a device that receives audio instructions and images and produces audible outputs. The suggested system is essentially comprised of Raspberry Pi-3 CPU running open CV, NumPy, a text-to-voice synthesizer, as well as speech recognition software. The system and its application in crisis circumstances are both cost-effective solutions that precisely address the demands of visually impaired people. The proposed solution is a wearable assistive 'Smart Cap' that assists persons who are blind engage and find their way to safety by wearing a smart cap. In this system, we save certain photographs or pictures in .cvs file format on an SD card. The system features a basic design that uses a Raspberry Pi to translate visual data taken using a camera for voice information. Input devices include the microphone, push button, as well as camera module. This camera module receives user input and output devices such as speakers provide audio output to the user. This approach is simple for persons who live alone. The architecture diagram (Fig. 1.1) shows how different system components work with each other.

A. Flow Chart

The basic workflow is shown in Figure 1.2. It offers a high-level summary of the suggested system's workflow.

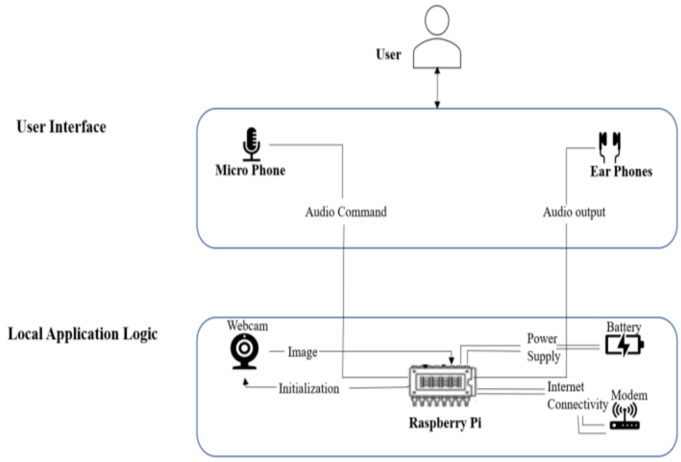

Fig. 1.1. User flow diagram depicting the working flow of the system

The following are the procedural stages for the system's operation.

1. When the Raspberry Pi is switched on the smart cap system, it wakes up and greets the user with whatever message we prepared during the development process.
   The system is now prepared to process our orders and carry out the necessary actions.
2. Following greetings, the program's flow is directed to waiting for user input in the form of voice, which is converted to text by the speech recognition module, after which the flow is directed to detecting faces, obstacle detection, text recognition, and online newspaper reading.
3. If the user's input command describes a scene or an object, the camera will capture the available frame (image) and store it to the captured frame folder on the SD card, first performing image deblurring and then object detection on that frame.
4. If the system finds objects in the scene (i.e., frame) then it finds the object names and the system will create the sentence which describes all the objects that are present in the frame.
5. The system has now completed one branch of the flowchart. If we wish to interrupt the system now, we can do so; otherwise, it will remain in the loop and switch to Listening mode.
6. Furthermore, voice instructions like "who is in front of me" will activate the facial recognition unit, "read me the text" will activate the OCR unit, and "tell me the news" will activate the newspaper website read capability. When turned on, the smart cap constantly listens to the commands of the user. A speech-to-text library is employed to transform the audio command into text.
7. Raspberry Pi does all the functionality. Eventually, its audio output is transmitted to the user through the Raspberry Pi's earphones/speaker.

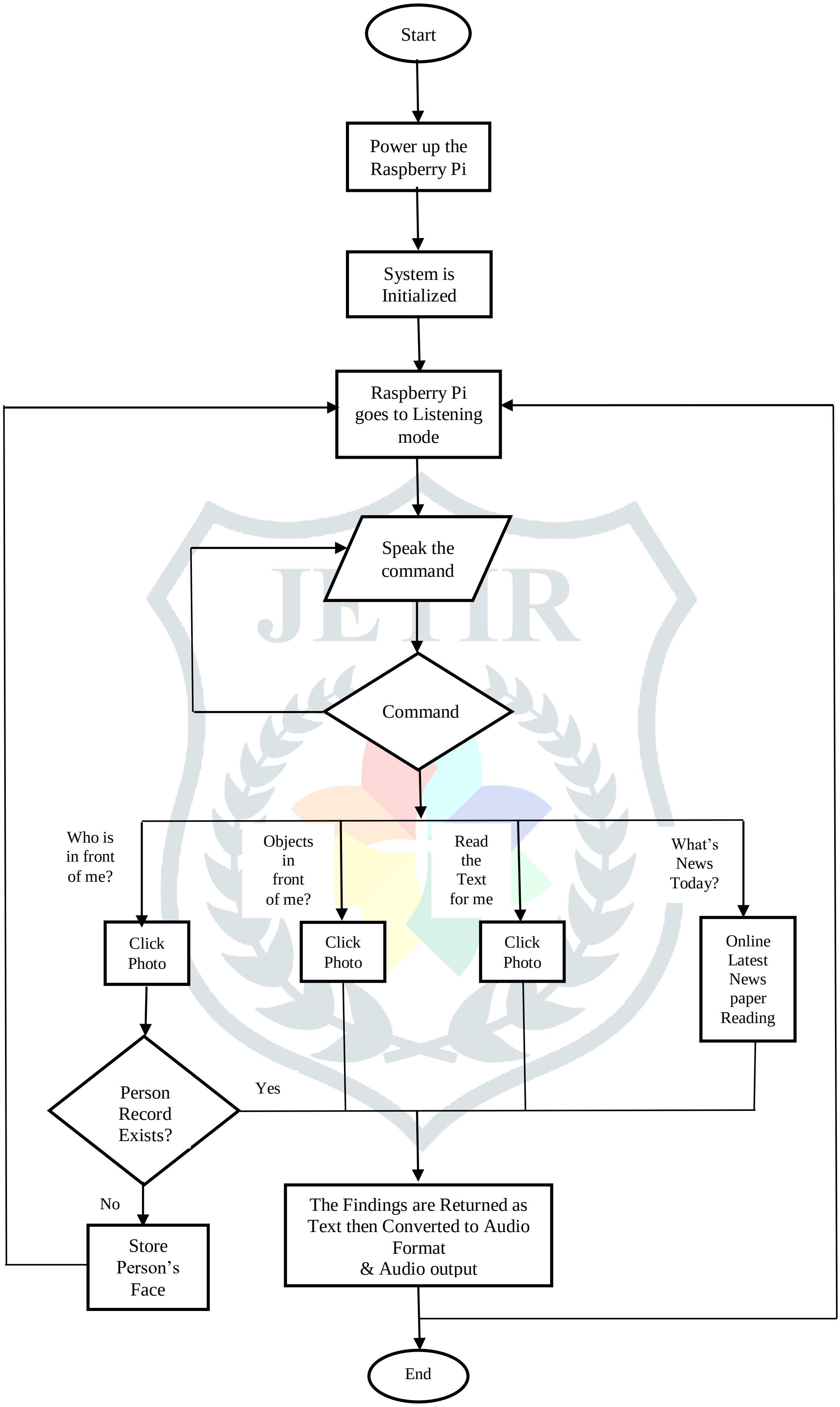

Fig. 1.2. Illustration of Smart cap architecture

B. Hardware Used

The Raspberry Pi is a miniature computer chip on a single board. There are several Raspberry Pi models available on the market, including the Raspberry Pi1 Model B, Raspberry Pi1 Model B+, Raspberry Pi2, and Raspberry Pi3 Model B. These all differ in memory capacity and hardware capabilities, such as the Raspberry Pi3's integrated Bluetooth and Wi-Fi modules, which were not available in prior models. As seen in Fig.1.3, it contains a 1.2 GHz 64-bit four core ARMv8 CPU with 1 GB of RAM.

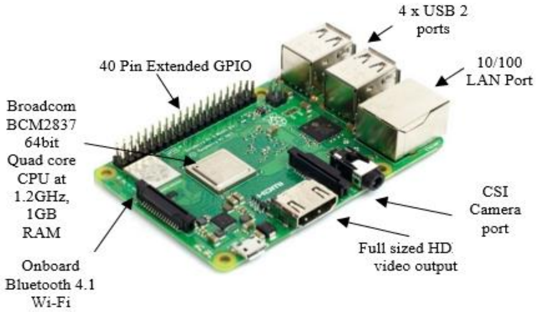

Fig 1.3 Raspberry Pi Controller

S. No	Components	Specifications
1	Raspberry Pi Model 3b+	1 GB Ram, 1.2 GHz ARM CPU
2	Web Camera	Logitech HD C270
3	USB Headphones	Boat Bass-heads 100
4	Power bank	Voltage of 5.1v and 5000mA
5	USB Dongle/ Wi-Fi	Minimum of 1mbps

Table 1 Hardware Components of the System

C. Software Used

1. Python Software

Python would be a widely employed high-level language of programming in industry and academia. Python IDLE comes in a variety of versions for programming in the Python language.

2. Open-Source Computer Vision (Open CV)

It contains a set of programming functions, primarily for computer vision. It offers over 2500 optimized algorithms for a combination of classical techniques and cutting-edge computer vision methods. In this research, it is utilized for face identification, object detection, image recognition, tracing, as well as other activities. It is mainly employed for image analysis.

3. Optical Character Recognition (OCR)

It is a method for detecting text within digital images. It is often used to recognize text in scan documents or images. OCR software could convert a hardcopy document or image into a digital text counterpart.

IV. PROPOSED SYSTEM

A. Face Recognition Module

Face recognition would be a two-step process consisting of face detection and facial recognition. Haar Cascades employ machine learning techniques to train a function from many positive and negative pictures. This algorithm technique is known as feature extraction. Haar features essentially comparable to convolution kernels, such as those employed to determine the presence of image features. The very first two are 'edge characteristics' that used identify edges. The third characteristic is a "line feature," and the fourth component would be a "four rectangle feature," that is most probably used to detect an angled line. Each feature returns a specific value, which would be computed by subtracting the number of pixels within the white rectangle as from the total number of pixels inside the black rectangle. Each Haar feature corresponds to a facial feature.

$$\text{Value} = \sum \text{(pixels in white area)} - \sum \text{(pixels in black area)}$$

Examining all imaginable Haar feature attributes like location, size, as well as type yields approximately 160,000+ features. Therefore, for each 24*24 PX, we should examine a significant number of attributes.

To prevent this, we can avoid duplicated features and focus on the ones that are really useful to us. AdaBoost might be used to do this.

B. Object Detection Module

This Object Detection module takes an image as an input then executes image classification as well as object localization. YOLOv3 leverages the Darknet-53 framework, that contains 53 convolution layers that are trained on the dataset of ImageNet. The YOLO algorithm splits the image as an S x S grid structure.

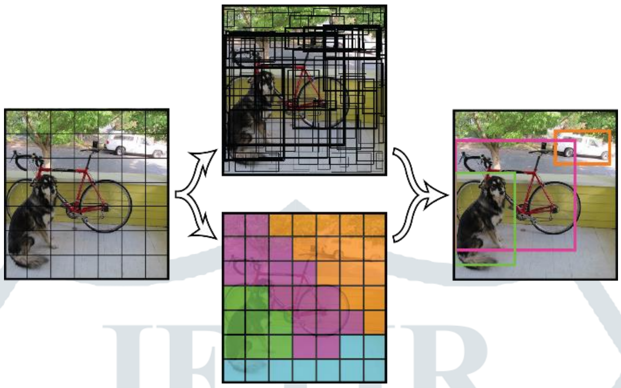

Fig 2 The example of working of Object Detection using Yolov3

Every grid within the input image may appear to be responsible for recognizing an object. Usually, the grid cell anticipates the number of bounding boxes for every item. Every bounding box consists of five parts (i.e., x, y, w, h, and confidence score). 'x', 'y' are coordinates of an object inside the input image, while 'w', 'h' are the width, height of the object respectively. The confidence score shows the likelihood that the bounding box has been associated with an object and its precision.

Confidence = probability (object) * IoU

Position accuracy is measured using IoU. YOLO employs non-maximal suppression to reduce copies of identical objects. Non-maximal suppression deletes the prediction with the lowest confidence score if there is an IoU threshold between any of the predictions in the picture. For each ground truth item, YOLOv3 assigns one bounding box anchor.

C. Text Recognition Module

Recent advancements in image processing, computer vision, and cloud services have made it simpler for assistive devices to identify and interpret text. The image processing techniques of Otsu binarization, Sobel edge detection, as well as connected component extraction are not applicable to all text-filled images. Due to the Raspberry Pi's limited processing power, deep learning algorithms like the integration of VGG16 and Connectionist Text Proposal Network (CTPN) & EAST (abbreviation for Efficient and Accurate Scene) are not appropriate for implementation. As a solution, an online API that provides reliable data while minimizing latency is used. The API can recognize text in both printed documents and photos of natural environments. It gives a hierarchical representation of the identified text that is divided into pages, sections, sentences, words, and symbols with their x and y coordinates.

D. Online Newspaper Reading

In the twenty-first century, the relevance of newspaper websites has expanded. Numerous Python libraries, particularly plugins allowing online newspaper scraping, have been created over the last decade. NewsAPI.org offers a simple API to retrieve news from more than 30,000 global sources. The API is accessible to all non-commercial (including open source also) and methodology to develop apps. To receive an API key, registration is required. This may be accomplished simply and quickly at "https://newsapi.org/register." After accessing the API, the requested specifics are translated to audio so that the user might just hear to remain abreast of the most recent news.

V. EXPERIMENTAL RESULTS

The results and analysis of the four modules mentioned above are presented in this section.

A. Face Recognition Results

OpenCV facial recognition is fast and precise, including for small faces, exhibiting greater accuracy of faces collected from a variety of angles.

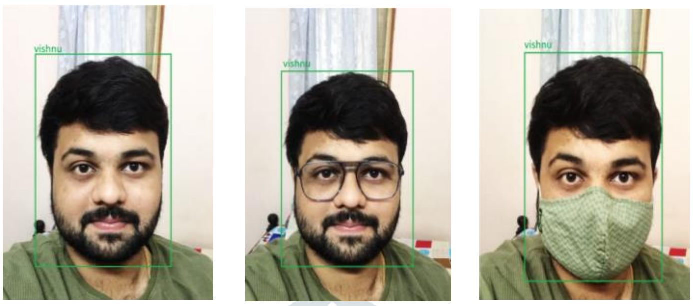

Fig3.1 Face Detection of a known person in different scenarios

The Haar Cascade classifier can be used to identify the presence of a human inside the input images. Fig 3.1 depicts the face detection output in the captured frame, which recognizes the individual whose data is saved in our database and converts to audio speaks the person's name as output.

Fig3.2 Unknown Face

B. Object Detection Results

The primary purpose of computer vision is to grasp observed situations. Scene comprehension comprises a variety of tasks, including as recognizing the sorts of items present, localizing numerous objects in 3D and 2D, recognizing the objects and scene elements, defining the connections between objects, and finally summarizing the scene verbally. The Common Objects in Context (COCO) data set is used to evaluate the performance of the proposed model. COCO is a huge dataset that is used for object recognition, humans key point detection, captioning, and segmentation. Objects are seen in the following photos, while Figure 5 demonstrates the correctness of the preceding result. We thus have two types of threshold levels: threshold as well as NMS Threshold. The fixed NMS threshold value equals 0.30, and the threshold value varies between 0 - 1.

Fig3.3 Detected Objects in different use case scenarios

C. OCR Results

The OCR module utilizes the Online API, that is sophisticated enough even to recognize text within images using different fonts & orientations.

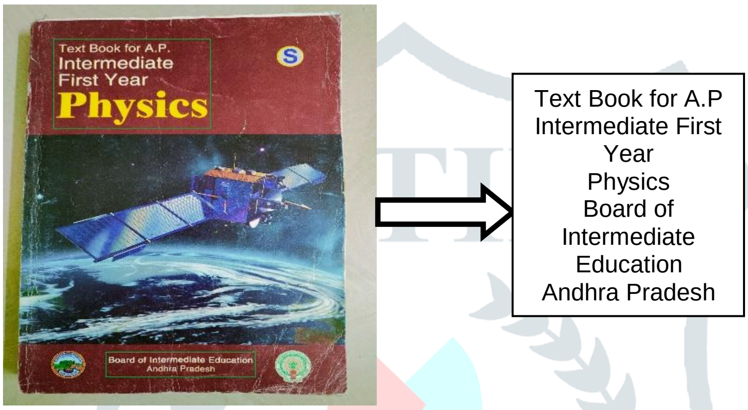

Fig3.4 OCR Output Image with Simple Text

Fig3.5 OCR Output Image with Simple Text

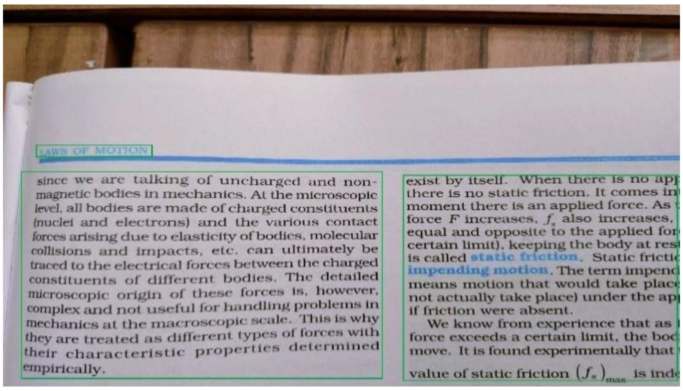

Fig3.6 OCR result image containing text in two columns

Considering character recognition via the documents, this one is designed so the vision challenged user may hold the paper in the right orientation and in close proximity to the Smart cap.

VI. CONCLUSION

Implementation of the proposed system "Smart Cap for Visually Impaired People" provides several advantages to users. Using a Raspberry Pi, the system's basic architecture converts visual data gathered by a camera into voice information. The system consists of a Raspberry pi-III CPU on which OpenCV, OCR, speech recognition, and NumPy have been installed. This research also suggests that adopting a more comprehensive approach that focuses on several attributes, as opposed to just one, is essential for building multipurpose and adaptive assistive technology. In addition, the research demonstrates how successfully the combination of deep learning and cloud services using inexpensive single board CPUs such as the Raspberry Pi can be utilised to create complicated assistive devices which assists visually impaired people in addressing real-world difficulties. As a result, this study proposes a single system that integrates a variety of diverse attributes. The system is inexpensive, user-friendly, and simple to setup, and its operation requires no specialized knowledge. The paradigm for the visually assistive device proposed in this research, when integrated with appropriate hardware and technology assistance, has the potential to play a significant role in assisting the blind & visually impaired.

Future Scope

The Smart Cap seen within the future of work for this subject may go in several ways. Initially, better hardware resources, such as GPUs, not only increase the device's reaction speed but also open the way for the inclusion of quicker and more accurate deep learning models. Combining OCR with Document Image Analysis (DIA) may enhance outcomes. The audio interface of the system might well be improved by adding support for many languages, allowing users to operate the smart hat in their native language. Lastly, the addition of proximity sensors enables real-time object detection.

REFERENCES

[1] Amjed S. Al-Fahoum, Heba B. Al-Hmoud, and Ausaila A. Al-Fraihat, “A Smart Infrared Microcontroller-Based Blind Guidance System,” Active and Passive Electronic Components, vol. 2013, p. 7, 2013.

[2] A. Nishajith, J. Nivedha, S. S. Nair and J. Mohammed Shaffi, “Smart Cap - Wearable Visual Guidance System for Blind” 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, 2018, pp.275-278, doi: 10.1109/ICIRCA.2018.8597327.

[3] Hanen Jabnoun, Faouzi Benzarti, Hamid Amiri, “Object recognition for blind people based on features extraction”, IEEE IPAS’14: International Image Processing Applications and Systems Conference 2014.

[4] Manduchi R. (2012) Mobile Vision as Assistive Technology for the Blind: An Experimental Study. In: Miesenberger K., Karshmer A., Penaz P., Zagler W. (eds) Computers Helping People with Special Needs. ICCHP 2012. Lecture Notes in Computer Science, vol 7383. Springer, Berlin, Heidelberg.

[5] Shagufta Md.Rafique Bagwan, Prof. L.J. Sankpal, ”Visual Pal: A Mobile App for Object Recognition for the Visually Impaired”, IEEE International Conference on Computer, Communication and Control (IC4-2015).

[6] Mala, N. Sathya, S. Sushmi Thushara, and Sankari Subbiah. “Navigation gadget for visually impaired based on IoT.” 2017 2nd International Conference on Computing and Communications Technologies (ICCCT). IEEE, 2017

[7] Vasanthi. G and Ramesh Babu Y India. “Vision Based Assistive System for Label Detection with Voice Output”, Department of ECE, DMI College of Engineering, Chennai, Jan 2014.

[8] Namita Agarwal, Anosh Iyer, Sonalakshi Naidu, Snedden Rodrigues, “Electronic Guidance System for The Visually Impaired - A Framework”, 2015 International Conference on Technologies for Sustainable Development (ICTSD-2015), Feb. 04 – 06, 2015, Mumbai, India.

[9] Aggarwal, L., Gaur, V.& Verma, P., “All-in-One Companion for Visually Impaired”, International Journal of Computer Applications, vol. 79, no. 14, pp. 37-40, 2013.

[10] G. Balakrishnan, G. Sainarayanan, R. Nagarajan and S. Yaacob, “Wearable Real-Time Stereo Vsaision for the Visually Impaired,” Engineering Letters, vol. 14, no. 2, 2007.