DESIGN AND DEVELOPMENT OF TESSERACT-OCR BASED ASSISTIVE SYSTEM TO CONVERT CAPTURED TEXT INTO VOICE OUTPUT

G.ELUMALAI[1],J.SUNDAR RAJAN[2], P.SURYA PRAKASAH[3], V.L.SUSRUTH[4],P.K.SUDHARSANAN[5]

[1] *Associate Professor, Dept. of Electronics and Communication Engineering, Panimalar Engineering College, Tamil Nadu, India*

[2,3,4,5] *UG students, Dept. of Electronics and Communication Engineering, Panimalar Engineering College, Tamil Nadu, India*

---***---

Abstract - *The biggest challenge faced by the blind people, is their inability to view real life objects and to read. The only efficient system that exists so far, is the braille system, that enables the blind to read. This system is time consuming and the time taken to recognize the text is long. Our aim here is to reduce the time taken to read. In our work, using a Raspberry Pi, we have designed a smart reader, so that the blind people may read. The module that we have designed, either uses a webcam or a mobile camera that is linked with a Raspberry Pi, to focus on a range of printed text. The OCR (Optical Character Recognition) package installed in raspberry pi tests it into a digital article which is then subjected to skew modification, segmentation, before feature extraction to perform sorting. Our proposed project automatically focuses the regions of text in the object, after which the text characters are then localized using a localization algorithm, that uses feature recognition and edge pixel distribution using artificial neural network. The text characters are then binarized into a machine readable form by using an OCR algorithm called as Tesseract. The recognized characters of the text are then converted into audio format, so that they may be recognized by the blind users.*

***Key Words*: Embedded devices, Image processing, OCR, Neural Networks, Raspberry Pi.**

1. INTRODUCTION

The importance of reading in our society, today, cannot be ignored. We could find printed text everywhere in the form of medical prescription, newspapers, books, handouts, medicine bottles and tablets and the like. While video magnifiers, optical aids, and screen readers can help with low vision and blind users to access documents, few devices do exist that provide optimum access to daily-life hand held objects such as fruit juice cartons and invention packages. The most important causes for visual degradation include diabetic retinopathy, aging factor, eye diseases such as ocular tumors and accidents, which lead to an increase in the number of visually disabled people, nowadays. Cataract is leading cause of blindness and visual impairment. Mobile applications are available these days, that aid in visualization and they form an integral part of the blind peoples lives. Recent advances in mobile technology, digital camera, computer vision and camera-based application make it possible to support visually challenged persons by developing camera-based application that combine computer vision with other existing technology such as optical character recognition (OCR) system. To detect text information from image, practical difficulties exist, such as non-uniform backgrounds, due to the large variations in character font, size, texture, color, background, orientations, and many other reasons. Text detection from scene/text camera images is possible due to high resolution camera.

2. PROJECT OBJECTIVE

The biggest challenge faced by the blind people today is their inability to read and recognize the real-world objects. The only efficient method that exists so far is the braille method, which enables them to read. It is somewhat time consuming to read the text using a braille and also it might be not be economically feasible. The objective of the project is to create a portable, light-weight module that enables the blind to read. This module uses a webcam to capture the image of the focused text, extract separate text characters from the scene and then provide a speech output using the Tesseract OCR.

3. LITERATURE SURVEY

In the field of Text To Speech conversion, research is focused on the identification of the input text and the ability to express it in a natural manner. Sentimental analysis procedure(positive/neutral/negative) is used in order to input features for expressive speech synthesis, where identification of the input text is a front end task. Different combinations of classifiers and textual features are evaluated for the determination of the most appropriate adaptation procedure. Semeval 2007 dataset and Twitter corpus are used for the evaluation of the effectiveness of this scheme of Sentiment Analysis, which is appropriate for an expressive Text To Speech scenario. Conducted experiments validate the procedure proposed with respect to Sentiment Analysis [1]. MD-TTS TC not only considers not only the contents of the text but also its structure, in contrast to the topic of text classification tasks. The proposal is validated by the experiments that were conducted in terms of both subjective(synthetic speech quality) and objective(TC efficiency) evaluation criteria. [2] For the visually disabled people, travel aids have been developed in the recent years in the form of stereovision-capable devices. They are however, bulky and assist the blind person in avoiding obstacles only and they are head-mounted. We researched about a wearable system that can not only assist in avoiding obstacles, but at the same time supports live streaming of videos over the 3G network. The module is comprised of an eye-glass and an embedded device that is power efficient. Two miniature cameras are used for imaging, on one end, by the eye-glass. For the combination of stereo images, FPGA and FIFO buffers are used. Also, live stereo matching on an embedded processor is achieved by means of parallel processing. Live streaming over 3G network, of the video that is captured by the system is also supported. A person with healthy eyesight can guide a visually impaired person, with the help of this live video streaming [3]. We researched about a system that helps the blind people to read from the labels and product packages of common hand held devices in their daily lives, by means of a camera. The Region Of Interest(ROI) in the camera view is obtained by using an innovative motion-based approach, where the users were asked to shake the object. This is done in order to isolate the intended image from cluttered backgrounds and other objects in the camera view. A mixture of Gaussian based, background subtraction methods were used by this method. Text information is obtained from the extracted ROI, by a combination of localization and recognition. Feature extraction and the distribution of edge pixels are studied, using Adaboost model, to automatically localize the text characters in the ROI. Localized text characters are then identified by using the OCR algorithms. Recognized text is conveyed to the blind users using audio output. The effectiveness of the system hardware is evaluated by using the dataset that is obtained from ten blind people[4]. In TTS technology, the ability is provided to a computer to speak. A synthesizer is used to convert the text into speech, thus making it eligible for further processing. By learning various textbooks related to languages, language mastery was obtained, traditionally. Due to the implied cost factor, moving along this method was very difficult. Various dictionaries were used by the existing system, to pronounce the identified text with correct pronunciation . Only for a set of words which are available in the dictionary, the operation is supported. [5] Studies in blind people have shown the activation of cortical areas that sub serve the vision, but it has long been controversial whether blind people have tactile acuity. We compared the passive tactile acuity of blind and sighted subjects on a fully automated grating orientation task and used multivariate Bayesian data analysis to determine predictors of acuity. Irrespective of the amount of childhood vision, braille reading or light perception level, acuity was superior in blind people. Acuity is better in women than in men and it also decline with age. Acuity is chiefly dependent on the force of contact between the stimulus surface and the subject . The difference between blind people and healthy sighted people varied a lot in spite of the intragroup variation: the acuity of both sighted and blind people of the same gender remained the same, except that the acuity of a blind person was similar to that of a sighted person who was 23 years younger. The results suggest that cross modal plasticity may underlie tactile acuity enhancement in blindness. [6]

4. EXISTING SYSTEM

A few systems already exist that help the blind people have access to daily-life hand held objects, but they are largely ineffective with respect to the focus area of the labelling. For example, portable bar-code readers do exist, but it is very difficult for a blind person to locate a bar code and then point the laser beam on that bar code. Also, to enable the blind people to read, braille system exists. In spite of ignoring the manufacturing cost, the system can be said to be largely ineffective with regard to the processing speed, that is, the speed at which the blind subject can identify the information and then assimilate it is too low.

5. PROPOSED SYSTEM

The proposed system uses a Raspberry Pi board, an ultrasonic sensor and a webcam to recognize the text in the scene. In this, the webcam is focussed on the scene. A video streaming is obtained, from which the images are captured frame by frame. The images are refined in order to eliminate any noise that is present in it. A feature called segmentation is used in order to separate each character from other in the text. Graphical details such as icons or logos, if any, are eliminated. Each obtained character are compared with the datasets that are created as a part of the Tesseract library. The Tesseract OCR is the most efficient algorithm available that checks for the obtained character in ten dimensions. Once, the character is recognized, it must be made available as an audio output. For this, we use a software called festival. The festival is used to provide the audio output for the recognized character. Apart from these features, an extra feature is added, that enables the blind to know the type of object that he/she interacts with.(a menu, newspaper and the like). An ultrasonic sensor is included as a part of the project, that makes the project obtain characters only within a particular distance.

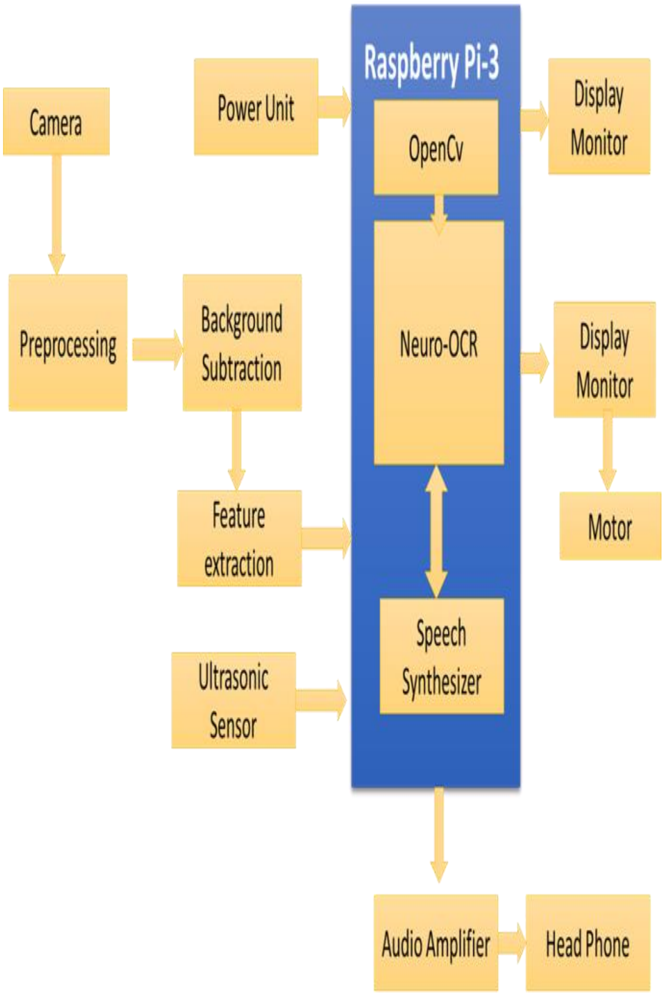


Fig-1: Block diagram of the proposed system

6. SYSTEM DESIGN SPECIFICATION

6.1. Hardware specification

1. RASPBERRY PI: The Raspberry Pi,using the ARM1176JZF-S core, is a credit card sized single computer on chip or SoC. System on chip is a method by which all the electronic components that are necessary to run a computer system are placed on a single chip. In order to start up, the Raspberry Pi needs an OS. On board non-volatile memory, which stores the bootloaders are eliminated in the Raspberry Pi, with the aim of cost reduction. Traditional embedded systems use Linux kernels and file system. For this reason, an SD/MMC card slot is provided. Raspberry Pi will be executed according to the application program and after the completion of successful boot loading.

2. ULTRASONIC SENSORS: Ultrasonic sensors provide cost effective, unique sensing methods, which are not provided by any other technology, thus serving the market at large. Application problems that are either cost prohibitive or those that couldn't be solved using other sensors can be solved by an ultrasonic sensor, by applying a wide variety of ultrasonic transducers and a wide variety of frequency ranges. Detection over distance are required by a variety of applications in industrial sensing. The detection range of ultrasonic sensors is quite high, as high as up to forty feet, which limit switches and inductive switching cannot do. Photo electric sensors can detect long distances, but they lack a wide area coverage or a wide area coverage which can be obtained by using a number of sensors. Migatron's ultrasonic sensor cover both narrow and wide areas. Proper ultrasonic transducer selection is what it all takes. Since there are wide range of target materials available, only ultrasonic sensors are independent of the target material composition . The target material may be solid, liquid, porous, soft or woody in nature and irrespective of its target composition, it can still be detected. Distance measurement is easy and it can be calculated by using the time from which the ultrasonic signal leaves the signal to the time the ultrasonic returns back to the sensor. This method is accurate to .05% of range which equates to +or- .002 of an inch at a distance of 4 inches.

6.2. Software specifications:

1. OPEN CV: OpenCV (Open Source Computer Vision Library) is a collection of library functions, whose main objective is to enable computer vision. The main objective of OpenCV is to provide a framework for computer vision and to improve the readability of machines over commercial products. Since it is BSD licensed, businesses can easily use and modify the code. Over 2500 sophisticated algorithms are present in the library, which include both machine learning and computer vision algorithms. Using the algorithms, the computer can identify objects, identify faces, recognize moving objects, recognize human actions, extract 3D models of objects, combine different images to produce an entire image scene, red eye removal, scenery recognition and augmented reality etc. can be performed. Around 47 thousand member user community for OpenCV exists and OpenCV is widely used by many commercial bodies, governmental institution and research organizations.

2. NEURAL-NETWORKS: Neural networks are a collection of hardware and software modules that are patterned around the neural networks of the human brain, in technical terms of information technology. They are a variety of deep learning technology that are also termed as artificial neural networks. Complex signal processing and complex patter recognition tasks are some of the commercial applications of the neural networks. Image processing and text recognition, handwriting recognition, weather prediction and facial recognition are some examples of the applications of neural networks since 2000. A neural network usually involves a large number of parallel operating processors arranged in tiers. Raw input information are received in the first tier – analogous to optic nerves in human visual processing. Instead of receiving the output from the raw input, each tier receives the output from the tier that precedes it. The last tier produces the output of the system. Every node that processes possess its own sphere of knowledge, including what it was originally programmed with, or what is has learnt or those set of rules which it has created for itself.

3. OCR: Optical Character Recognition is the conversion of either printed or handwritten characters into a machine readable format(binary 0s and 1s). The text of concern may be obtained either from an image of a document, or a scanned copy of the document or from a real life scene such as the text on signs and billboards(Fig.6 and Fig.7) or from subtitle text underneath an image (such as those that occur in a television broadcast). It is used to enter the information from different types of documents such as records, passports, cheques, invoice, bills or any other such suitable type of documentation. OCR is a widely used method for converting the printed data into digital format, so that they may be edited, searched , compressed, used in line display and in machine related activities such as data mining, TTS, machine learning and cognitive computing.

4. ESPEAK: A job is done by a computer in three distinct stages, wherein the first stage is called input (where information is often fed using a computer and a mouse), processing (where the input is responded to by the computer, say, by subtracting the numbers that you typed in or by enhancing the colors on a photo you scanned), and output (where you can see how the computer has processed the input, wherein it is obtained on a computer screen or as a printout on a paper).Espeak uses a technology called as speech synthesis, wherein a text material is loudly read out by the computer using a real or simulated voice via a loudspeaker. This technology is referred to as Text To Speech.

5. PYTHON: Python is an object oriented programming language that is widely used in Pi development boards. It is the preferred language for data analytics and data science. It was found towards the end of the 1980s as a successor to the ABC language by Guido Van Rossum in the National Research Institute for Mathematics and Computer Science. Python has a clear syntax and it is developed with the intention of making the code readable. All these has paved the way for its popularity since its inception. Python is a high-level language. The code for python is largely written in simple English, that makes it understandable to the user. The commands provided to the Pi board are mostly in English. This is in stark contrast to an assembler that more or less operates in a way that is comfortable for the machine, thus making it cumbersome for a normal human to read. Python is valued by those who want to learn programming, on account of its clear syntax and the fact that it is a high level programming language. Raspberry Pi recommends Python for those who are willing to learn embedded programming from scratch.

7. IMPLEMENTATION OF THE PROPOSED SYSTEM

The main idea behind developing the device is the portability and ease that it could offer to the end user. Hence, it is deliberately intended to be simple, so that it could be of light weight and hence, easy to carry. The module uses a Raspberry Pi development board, which is a credit card sized SoC and is quite efficient. A webcam is interfaced with this board, in order to enable the transmission of the captured stream of video to the Pi board. An ultrasonic sensor is used as a range finder, that enables the module to detect text data only within a specified range. The intended simplicity of the module comes from the coding part, which is towards the software side. The most efficient OCR algorithm that is available is the Tesseract OCR, whose accuracy of detection is pretty high. Since portability is intended and the Pi board runs either using a PC power supply or an AC adapter, we are using a power bank to power on the Pi board. The working of the device thus goes on as follows: The Raspberry Pi board is powered ON, on startup, the code for identifying and recognizing the text gets executed(Fig.4). The camera is then focused towards the intended text. Ultrasonic sensor is used to enable the webcam to read the text only within a stipulated area. The Tesseract algorithm then identifies each character of the text and feeds it to another software module called as Espeak. Espeak converts the detected data into audio output. The audio output is made available at the 3mm earphone jack in the Pi board. The arrangement of the device(Fig.2 and Fig.3), the scanned text and the obtained output are included below(Fig.5):

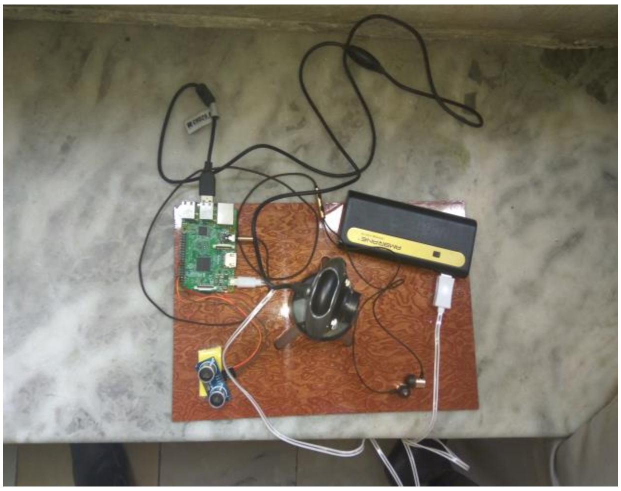


Fig-2: Top view of the proposed system

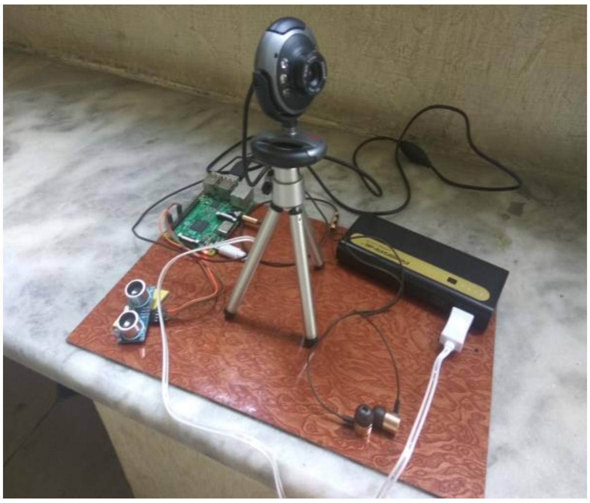
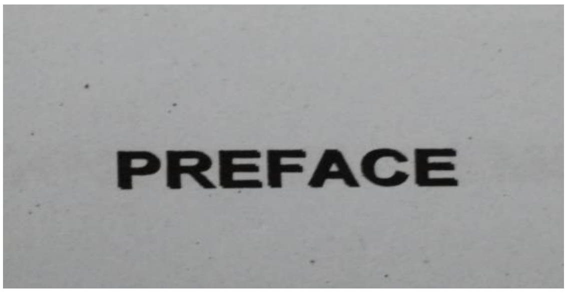
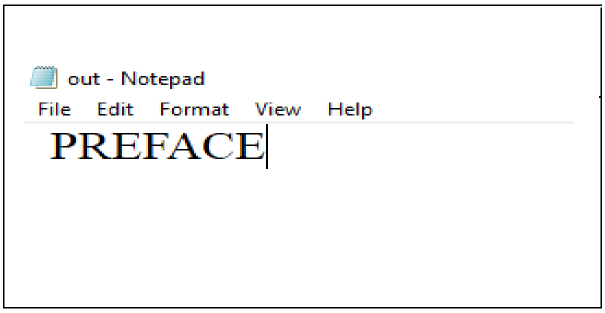


Fig-3: Side view of the proposed system



PREFACE

Fig-4: Scanned text

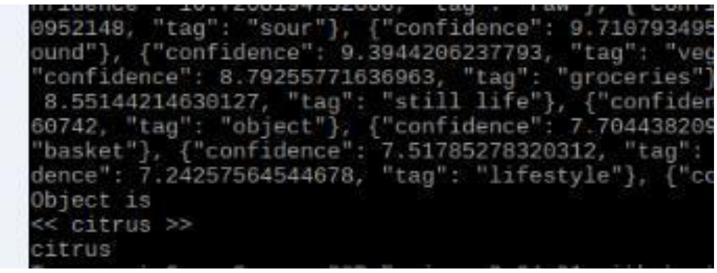


PREFACE

Fig-5: Obtained Output(since the audio output cannot be included, the output is included in visual form)

Fig-6: Scanned object

The object scanned by the web camera is orange which is identified by the name of citrus using the Imagga database (Fig.6). The tesseract detects the scanned object and detects it as citrus and displays the output(as shown in Fig.7).



```
0952148, "tag": "sour"}, {"confidence": 9.710793495
ound"}, {"confidence": 9.3944206237793, "tag": "veg
"confidence": 8.79255771636963, "tag": "groceries"}
 8.55144214630127, "tag": "still life"}, {"confiden
60742, "tag": "object"}, {"confidence": 7.704438209
"basket"}, {"confidence": 7.51785278320312, "tag":
dence": 7.24257564544678, "tag": "lifestyle"}, {"co
Object is
<< citrus >>
citrus
```

Fig-7: Output

8. CONCLUSION AND FUTURE SCOPE

A text detection and recognition with speech output system was successfully demonstrated on Android platform. This system is very handy and useful for the visually challenged persons. Compared with a PC platform, the mobile platform is portable and more convenient to use. This system will be helpful for visually challenged persons to access information in written form and in the surrounding. It is useful to understand the written text messages, warnings, and traffic direction in voice form by converting it from Text to voice. It is found that this system is capable of converting the sign boards and other text into speech. There is further scope in the development of this project. The project may further be developed into/embedded along with the walking stick, which blind people usually use to navigate. Apart from this, in the future, smaller development boards may come into existence or smaller cameras may be developed, thus further enhancing the portability of this device. Flexible PCB boards is the latest trending technology in the field of printed circuit boards. If we implement all these into the project, the size of the project will be greatly reduced, thus being a handy system for the blind people.

9.REFERENCES

[1]. Alexandre Trilla and Francesc Alías. (2013), "Sentence Based Sentiment Analysis for Expressive Text-to-Speech", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 21, Issue. 2. pp. 223-233.

[2]. Alías F, Sevillano X, Socoró J. C, Gonzalvo X. (2008), "Towards high-quality next-generation text-to-speech synthesis", IEEE Trans. Audio, Speech, Language Process, Vol. 16, No. 7. pp. 1340-1354.

[3]. Balakrishnan G, Sainarayanan G, Nagarajan R. and Yaacob S. (2007) "Wearable real-time stereo vision for the visually impaired", Vol. 14, No. 2, pp. 6–14.

[4]. Chucai Yi, Yingli Tian, Aries Arditi. (2014), "Portable Camera-based Assistive Text and Product Label Reading from Hand-held Objects for Blind Persons",IEEE/ASME Transactions on Mechatronics, Vol. 3, No. 2, pp. 1-10.

[5]. Deepa Jose V. and Sharan R. (2014), "A Novel Model for Speech to Text Conversion", International Refereed Journal of Engineering and Science (IRJES) Vol.3, Issue.1, pp. 39-41.

[6]. Goldreich D. and Kanics I. M. (2003), "Tactile Acuity is Enhanced in Blindness", International Journal of Research and Science, Vol. 23, No. 8, pp. 3439–3445.