AN ENHANCEMENT OF TEXT TO SPEECH (TTS) SYSTEM USING RASPBERRY PI

Khaldoon Ibrahim Khaleel[1], Dr. Ku Nurul Fazira Binti Ku Azir[2]

[12]Master of Science (Embedded System Design Engineering), Computer and Communication Engineering, UNIVERSITI MALAYSIA PERLIS

[1]Khaldoon.ibrahim1980@gmail.com , [2]fazira@gmail.com

Abstract: The blind man suffered greatly until the famous French inventor Louis Braille developed Braille, a form of writing invented to enable the blind to read. It did this by making characters prominent symbols on paper, allowing reading to become simple. There are already many systems that read images and give voice output. More recently, since the computer appeared in the twentieth century, things became easier for our daily life. This coincided with the beginning of the disappearance of the book as a hard copy, and electronic copy books became an urgent need in the development of reading, writing to help the visually impaired. The main problem encountered with this project was: the difficulty in the recognition of handwriting, and changes in the forms of handwriting, size and the number of paragraphs, as well as the recognition hard to extract letters containing noise from the document. The goal of this work-design system was to design a system able to extract and recognize handwriting for the visually impaired, with improved portability, to recognize the font size of lines from 8 to 10 PT handwriting using the hardware of a Raspberry Pi3 device. Another goal was to increase the valid ability of the blind people to read handwritten text. The method of work first the hardware platform used: Raspberry Pi 3 model B with inbuilt 5 MP camera, headphone; a 3D printed reader arm; second the software used: python language for after image capture of handwriting, grey scale image processing, followed by a Tesseract algorithm for sampling characters, followed by OCR and TTS, adding pronunciation capability to recognized characters. The result of this project is the ability to recognize handwriting types successfully without errors, on the condition that they are regular, as well as the ability to read capital letter and small letter handwriting without errors, and to read size 8pt text successfully.

Keywords: Text to speech(tts), OCR, Raspberry Pi.

Introduction

The advanced character recognition system is very popular nowadays in current computer technologies (Chukwunazo and Onengiye 2016). The Raspberry Pi 3 is more popular in current technology because it is reconfigurable and the latest technology with the support of Python. However, the Raspberry Pi 3 challenge is still present upon implementation and we will discuss. The subject related to the fields and research will be studied in this chapter. There are methods used to design a handwriting recognition system such as using a Tesseract algorithm that can meet the required recognition requirements. In this paper we implement OCR technology, using Tesseract. After image processing, the obtained text is converted to speech using a Text- to-speech engine that speaks processed text through a loudspeaker or speaker that can be connected to the Raspberry Pi audio jack. (Gurav et al. 2017).

The inability to read has a significant negative impact on the quality of life of the visually impaired population. Handwriting remains a large part of the knowledge that this group needs to have unrestricted access to methods of accessing and reading the text aloud to the blind.

A text to speech (TTS) is a Raspberry Pi-based and digicam gadget which could study the text, regardless of whether or not the textual content is delivered by a laptop input flow or a scanned input submitted to an Optical character recognition (OCR) engine. A speech synthesizer may be carried out by way of both hardware and software program. Speech is frequently based at the concatenation of herbal speech devices that are taken from natural speech, put together to shape a word or sentence. Many TTS structures are advanced based totally on the principle of corpus-based speech synthesis; very popular for its excessive quality and herbal speech output. The first electronics-based TTS system become designed in 1968. The concatenation technique changed into evolved by way of 1970. Many pc operating structures have included speech synthesizers since the early 1980's.

According to previous studies, the recognition system found it difficult to extract letters containing noise from the document. In addition, the recognition of handwriting is one of the most difficult things because of the processing of writing is complex, and it has proved even more difficult to improve system recognition below point size (Clausner et al., 2016).

Many visually impaired access development systems are based on the two rudimentary structure blocks of: OCR software and Text-to-Speech (TTS) engines, where OCR is used for the translation of captured text images into machine-encoded text. It is defined as the method of changing pics into a manner-capable layout. OCR is also valuable in devices for the visually reduced who cannot read the text but need to admission the gratified of the text leaflets

(Saleous et al. 2016). Optical Character Recognition is utilized after the user places the document below the camera, whereupon the camera captures and image of the text.

The camera is mounted on a stand, so that if a paper, for example, is placed in between the area marked by braces: it captures a full view of the paper into the system so long as certain pre-conditions are met; adequate lighting, and that the paper's contents are composed in the English language. When all these requirements are met, the computer takes a picture, processes it and, if it recognizes a written document, converts the contents of the document to speech, and then it speaks out the text that has been translated from processing the document's image into a text format. This is done through the use of a Raspberry Pi device which allows the visually impaired reader to understand the text unaided (Bhargava et al. 2015).

In modern society many people suffer from various diseases, including those they are born with, and some of whom are affected within days by the condition of blindness; a disease which deprives them of some of life's pleasure and beauty, does not distinguish by age and leaves most with the inability to read or write. According to the World Health Organization's 2002 estimates that blindness was one of the most common medical conditions worldwide. The blind man suffered greatly until the famous French inventor Louis Braille developed Braille, a form of writing invented to enable the blind to read. It did this by making characters prominent symbols on paper, allowing reading to become simple (Cumberland and Rahi 2016).

Many researchers have tried to use the computer to help the blind read, however there are still many unresolved problems in previous studies on the recognition of handwriting where researchers struggled to develop solutions to related problems such as changes in the forms of handwriting, sizes and the number of paragraphs. The problem of recognition therefore, is a difficult problem (Elmannai et al., 2017).

Embedded System

The microprocessor-based systems are developed for controlling a specific function or a range of different functions and they are not designed for being user- controlled in the manner similar to a PC. These systems are known as the embedded system, and they are designed for carrying out a specific task, along with several other options and choices.

The embedded systems are seen to be used in daily life and are popular globally. It must be inexpensive, and their memory must be the major component of the complete system, which would further decrease the total costs. The complexity and the performance requirements for the embedded systems are seen to rapidly grow, which would lead to further use of memory and more power usage (Garousi et al. 2018).

The embedded systems are comprised of processing cores which are either digital signal processors or microprocessors. Each microprocessors is known as a "chip" that can be packaged by itself or along with other microcontrollers to form a hybrid of the Application Specific Integrated Circuits (ASIC). Generally, the input for the microprocessors comes from the detector or from sensors in the form of a specific word, while the output is passed on to the activator that may initiate or halt the operation of the whole operating system. The embedded system comprises both hardware and software components.

Material and Methods

The framework is the implementation of character recognition technology in a built-in system based on Raspberry Pi3. The drive behind the design is that previous studies with low vision people, which are small in size and portable, enables more intuitive process with little preparation. In this paper, we future a system for reading for the visually impaired. The proposal is a completely combined system that has a 5 mega pixel Digicam as an enter tool to feed the printed text report for the scanned report to be processed by means of the OCR Module (OCR). Recognition method is carried out to the letter collection and reading line. As part of software development, most accessibility skill tools designed for people with sightlessness and incomplete dream are mainly built on OCR and text-to-speech software.

Optical Character Recognition (OCR) is the conversion of images taken from published text into mechanism coded text. Digitizing texts also helps in reducing storage space. It is widely used to convert books and documents in electronic documents for use in storage and document evaluation. Optical man or woman reputation makes it feasible to use technologies along with gadget translation, and textual content to speech. The ultimate identified text document is fed to the output devices relying on the consumer's choice. The output tool may be a headphone-linked to a berry pad or a speaker that may spell the textual content document out loud. It offers an set of rules to hit upon and examine

text in herbal pics for use through the blind and visually impaired subjects. Generally, the set of rules has a success price of over ninety% in the check institution, and the amount of unread textual content is normally very small.

Project Workflow

A properly knowledge of present and relevant information become an crucial primer to begin the paper.

In the first step, the selection of the Raspberry Pi 3 is done. The suitable platform and technologies for designing device character recognition chosen to meet the requirements and achieve the objective of this project. Raspberry Pi 3 with 5 MP camera and python3 is chosen for this project due to its cutting-edge. Raspberry Pi 3 has several desirable features such as high performance, and lower power consumption.

In the second step, the image is captured, the handwritten document is placed between the base of the device and the camera, the inherent camera captures the images of the text, with a distance of 30cm between the camera and the base. The 5 MP camera of the Raspberry Pi has high-resolution and quality that aids image capture and will allow for high quality, fast and clear recognition of the image.

In the third, the letters are extracted; the photo is then transformed to grayscale. A text image is provided as an Input to the Tesseract set of rules, that's a command- primarily based tool with a four dimensional feature vector (x, y-role, path, period) derived from every issue of the polygonal estimate and combined to be processed by the Tesseract.

In the fourth step, image processing: One element on this photo processing module is OCR. Using the OCR engine required specific steps a good way to get the high-quality input for OCR. The output of this, processing a minimum mistakes charge is also a short processing time. This module does not exchange the OCR algorithm, but gives a further kingdom to get the nice enter for OCR.

In the fifth step, Text to speech: this module is initiated with the belief of the previous module. The module performs the venture of conversion of the transformed textual content to the audible form. The Raspberry Pi has an on-board audio jack; the on-board audio is generated with the aid of a PWM output and is minimally filtered. A USB audio card can greatly enhance the sound pleasant and quantity by using headphone as shown Figure 1.
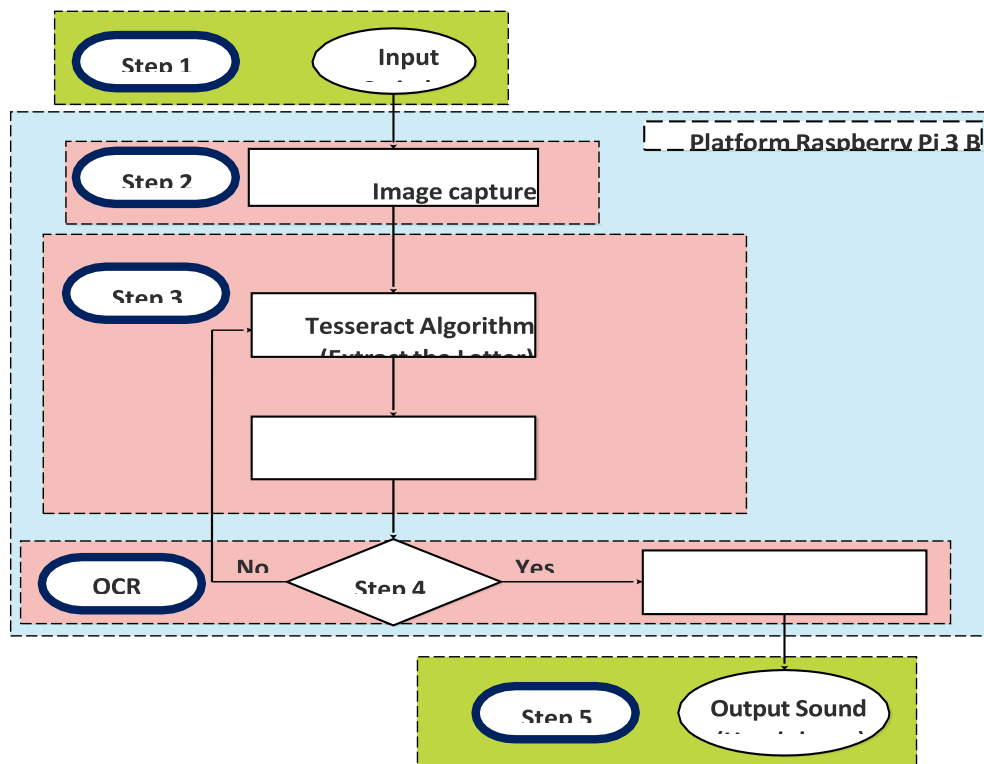

Figure 1 Flow chart of the proposed system.

Block Diagram of the System

In this work, the main Raspberry Pi 3 B platform contains the important tools to make this project work: connection with components headphones, power supply 5MP camera, the SMPS change; which have characteristics and contact specifications discussed, and that contact between components is shown in Figure 2 .
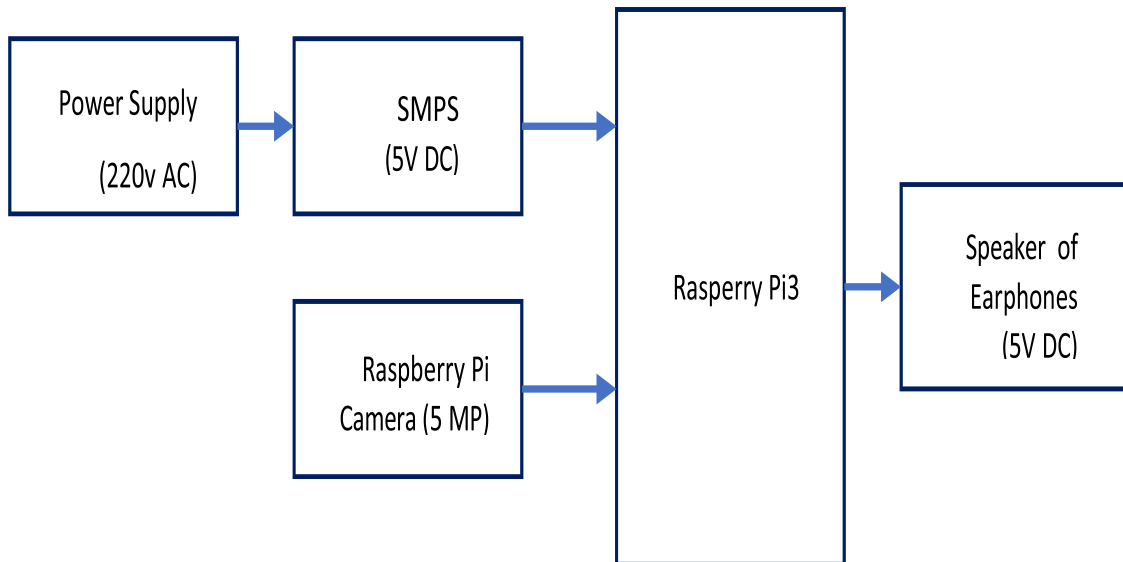

Figure 2 Block Diagram System.

System Specifications

It is a system used to collect speech from text. To convert text to speech, the engine that changes written text into language acts as the voice: converting it to the wavelength that can give the sound output. It is a system that can be effortlessly rummage-sale in Raspberry Pi 3 and 5 MP camera. The text is converted to an audio file via the TTS engine.

Hardware Specifications

The text handwriting reader system consists of a Raspberry Pi board, power supply, connecting earphones to the 3.5 mm audio port and a camera interface. The board of Raspberry Pi is powered by a USB cable or battery. It's a single machine with a credit card size. System (SoC) is a system by which all electronics are mounted on a single chip to operate a Raspberry Pi. The CPU, the GPU, the USB controller, the RAM; all bundled into a small package. An operating system is used to start the Raspberry Pi. In order to reduce costs, the Raspberry Pi skips any non-volatile on-board memory use, as seen in many embedded devices, file systems, and boot loaders. The compartment of the SD / MMC card is inserted for this purpose (Eldin Gamal et al. 2017).

Software Specifications

Extracted speech from the text that each text is analyzed by character and converted into an equivalent speech signal that is processed at the predefined frequency to obtain the audio output. This introduces the TTS and OCR system, which executes all operations in a step by step manner. The Raspberry Pi control system can deliver sound output. The required audio output can be received through the socket through earphones or speakers.

Main Component of Recognition System

A pinnacle-degree architecture design block diagram is provided. The photo received from the digicam sensor is processed via photo processing and converted right into a virtual signal of the signal received. After processing, the information is sent to the crucial computing unit of the Raspberry Pi3 controller as an input character from the document, and for conversion to the speech, the correct operation of the paper, one needs:

I. Raspberry Pi 3 Model B

II. Camera Raspberry Pi 5 MP

III. Earphone

IV. Connection Cables

V. Implementation

Initially, Raspberry Pi3 should be installed in the SD card, which should store all the program files needed by Raspberry Pi. The SD card should be formatted before installing the OS, then the SD card is introduced in the slot and the Network connectivity is provided. Raspberry Pi is linked to the PC using the VNC viewer. Installation of the necessary software and packages, configuration, camera, and earphones. The program design process is executed, and the icons are executed. Thus, after all these steps of installation, the machine is run independently without connecting with the computer as shown at Figure 3
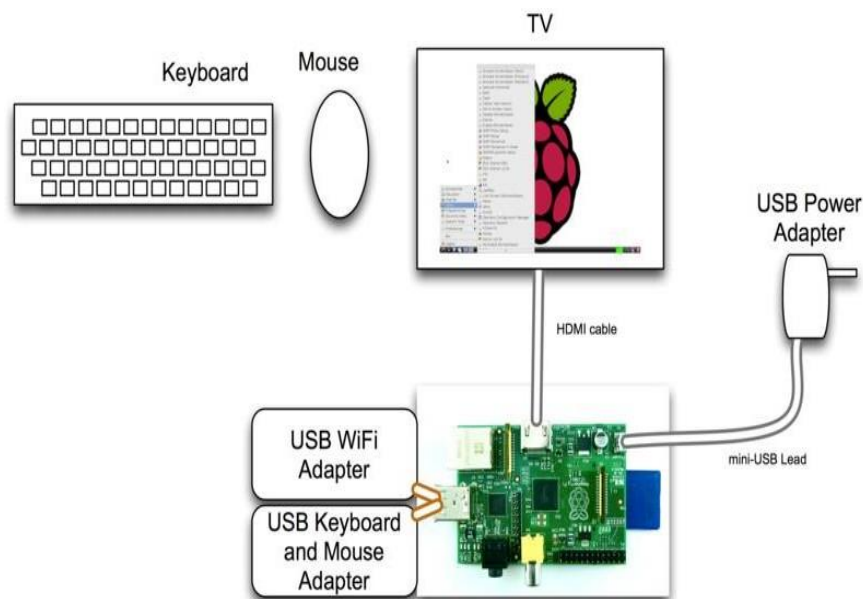
.

Figure 3 Implementation System.

Python

There are two main models of Python in use, Python 3, and Python 2. Python 2 is more compatible with existing technology currently bundled with Raspberry Pi. In this paper, Python 3 is used and is typically used to enclose parts of the code. This is achieved due to Python use of indentation to organize code. This is a good practice in most languages, because it makes the code more readable. If Python does not find the spacing correct, the code will not

work. While this can be a shift for experienced programmers, it quickly becomes natural and leads to a clear and hypothetical code.

Python provides a good initial language because it represents a simple implementation of object-oriented programming and is mitigated on how to create and manage variables. A net result is a specially produced programming environment. Python code is usually run through an interpreter rather than aggregated, but the final results are nevertheless very fast. Python applications can be run on all major operating systems. (Rossum et al., 2018).

Software Design of Recognize Text

The phases involved in the design by using python language, which write step to learn of this system, starts with the 5 MP camera taking an image capture of the handwritten text and transferring to Raspberry Pi for image processing by using the algorithm, Teseract which has instructions to separate letters from each other's and makes the system able to recognize the text by OCR, then saving the extracted text as a symbol file where the Tesseract algorithm analyzes the symbols and then converts them to the compent (TTS) and converting this to voice output using the Voice app (TTS) and producing speech through headphones as shown at Figure 4.
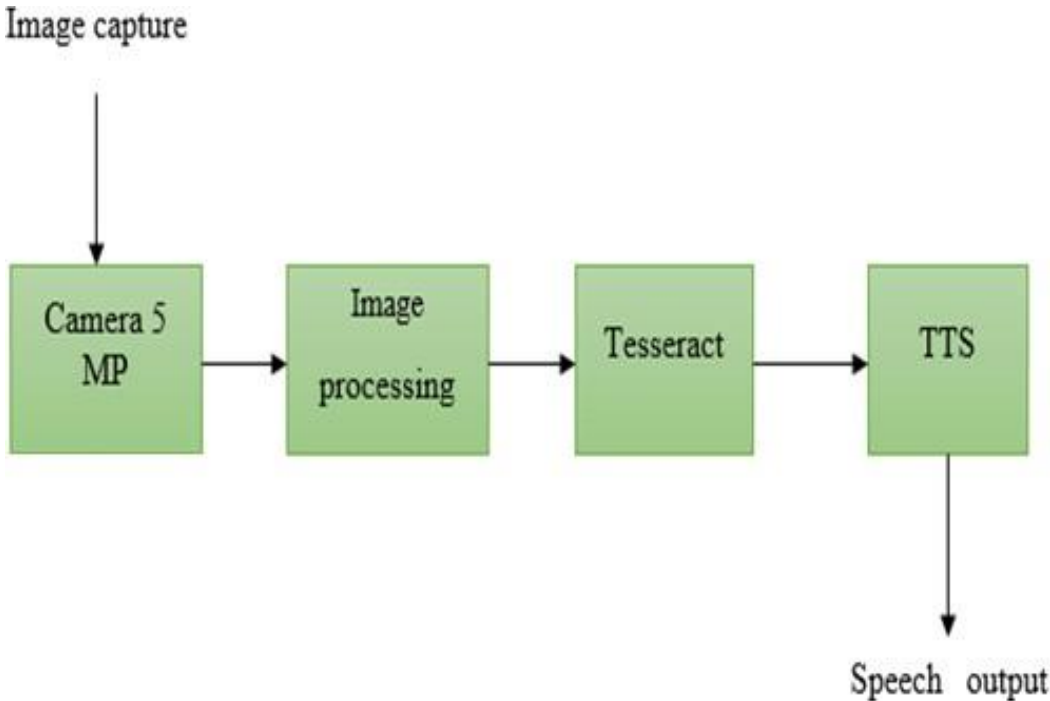

Figure 4 Software Design.

Architectural Design

The architecture design of the system has been described at a high level, with a focus on the main modules that compose the system and the communication between these components; the architecture design of the system. This design has several external components such as the camera that captures the image, the internal memory that has been stored the data, isolation of the image, conversion to grey image and defining the desired character to be displayed on the monitor. The proposed internal component that was used was the raspberry pi 3 platform to control the retention of the data in the internal memory (SDRAM Size 32 Mb), which according to the manual was most suitable for this design. There is a pre-processing step to remove the noise in characters that appears during the image processing.

Part of the SOC Development platform, which offers a robust architecture based on Tesseract on Raspberry Pi, contains the basic Tesseract processor and specifications in this platform. The Tesseract processor as a combination of ARM and GPU.

Image Capturing: the first step is to turn the device on after connecting it to the 5 V DC power source, then placing the handwritten report between the camera stand at a distance of 30cm between the camera and the base, with the device base parallel to the base corners and then pressing the picture button allowing users to download photos from the camera directly connected to Raspberry Pi. No organizational tools were used, but it is useful to collect images from a variety of documents without the need for drivers. The color scanners convert the existing two-dimensional analog images into digital form and then transmit them into the Raspberry Pi. Image Acquisition: the camera takes images of the text in this step by pressing the button. The accuracy of the image captured depended on the lighting. The better lighting, the better the image quality. A program model interface and an API that enables Photos software to communicate with imaging devices inclusive of scanners and virtual cameras are implemented as an on-demand service in the operating system. An extensive set of support for static digital photography provided by static image engineering provides an interface for basic data transfers to and from the device (in addition to invoking the image scanning process on the Raspberry Pi device), and applications can take advantage of these features programmatically.

Working Principle

The letters are recognized by the system result. Data is recognized in the document and converted into speech. The assembly of adjacent characters in the text is made to be ready for conversion. This paper is useful for modern digitization systems that are generally used in daily life and reduce manual efforts and do not require Braille. The photo of a handwriting text is captured thru the Raspberry Pi digital camera module; the picture is difficulty to pre-processing which incorporates correcting skew angles, sharpening of the photograph, thresholding and segmentation. The processed photograph is sent to the TTS. The steps for converting an image to text are discussed below:

Step 1: Processing text image captured from a Raspberry Pi camera.

Step 2: Text Conversion.

Feature Extraction

The extraction of the function refers to extracting symbol features from the image. In this stage, to find only important attributes and disregard any unnecessary attributes: this technique takes the abstract features of the character into account. Some of the abstract features are spaces, lines, and intersections. The Tesseract algorithm is used for the extraction of the functionality shown in Figure 5.
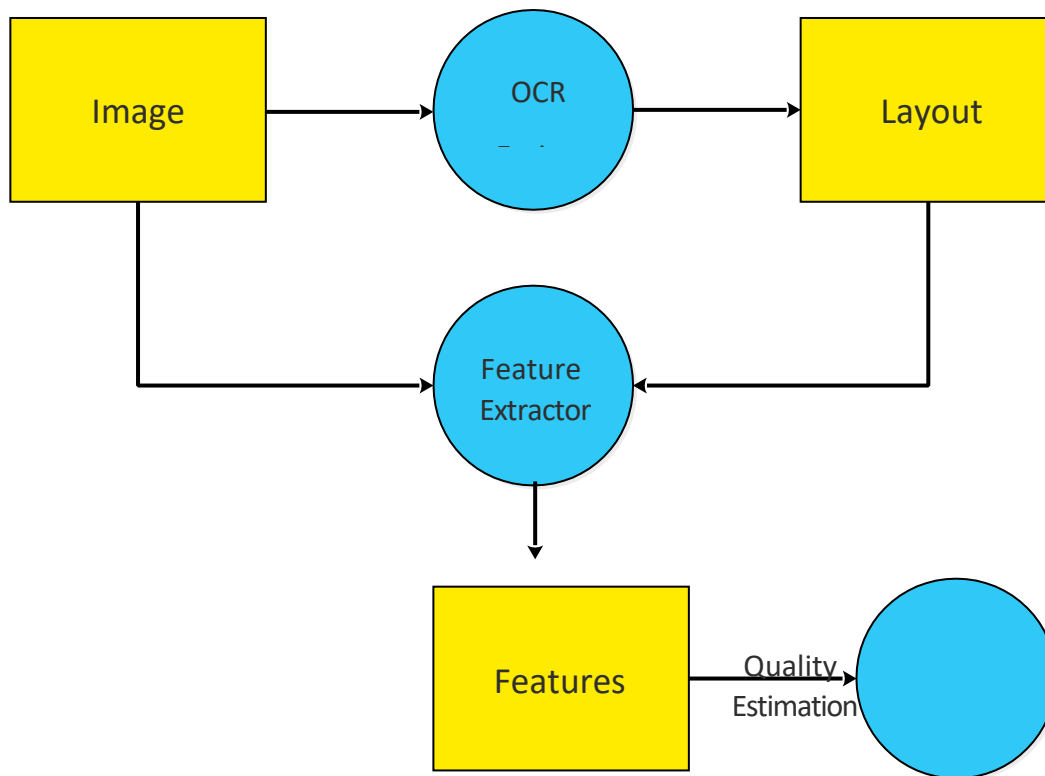
Figure 5 Feature extraction from a page image

Results and Discussion

There are already many structures that study photographs and supply voice output. But this device offers voice output with out the need for community get entry to via the person. This is accomplished by capturing the photograph to document that's to be read using a Raspberry Pi 3 camera module. Raspberry Pi is a credit-card-sized single-board laptop. The working system used is Raspberry Pi 3. A 50 cm ribbon cable is used to connect the camera module to the raspberry pi. The coding is achieved the usage of python language. The Optical person popularity engine converts the images of textual content into device-encoded textual content and saves it in a textual content file. Tesseract is the OCR engine that is used for extracting the English text from the record and storing it in a textual content record the textual content to speech engine converts textual content to speech.

The Results

In this part of paper we discussed how the method covered. This work presents the results that were performed in selected laboratories in an ideal environment. After connecting the components correctly, the power supply is activated and the device is powered by 5V (DC) method after which the camera, as shown Figure 6 is working, and when the button is pressed, the camera starts to capture pictures of the text and the system works to pronounce the words.
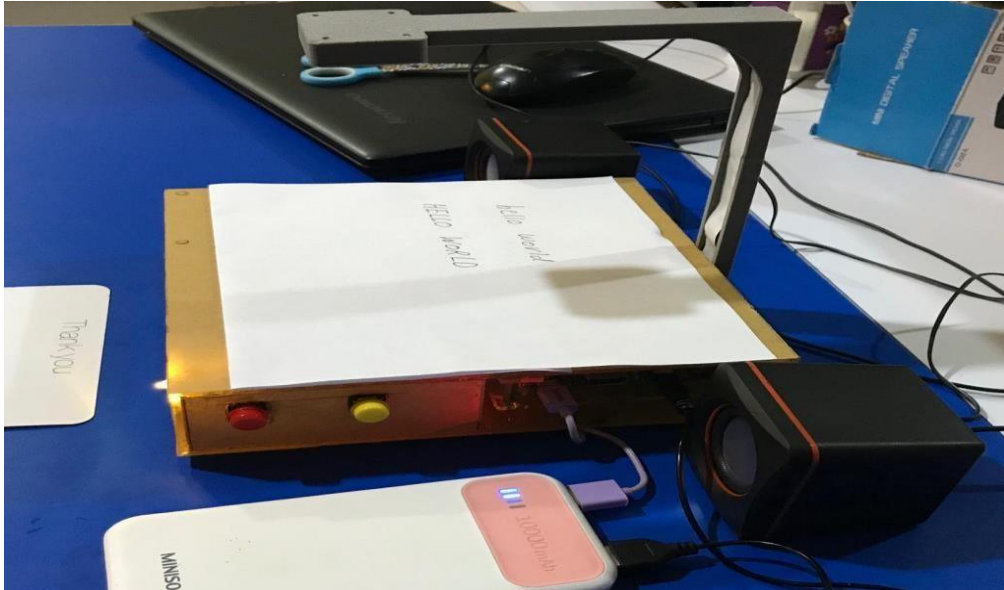
Figure 6 Turn on the Device.

In Figure 6 shown the simple of handwriting capital letters in the device that have been tested.

Experimental Outputs

The text document which to be read out is placed at a considerable distance; 30 CM from the camera so that the image is clear enough with proper illumination. The proposed system Is tested with different enter sets as, Text in the form of Black and White, Text with distinctive font patterns; using four documents the device is tested by the OCR process with examples of 100 words which harvests an correctness of around 98%. The displayed text is read out by the text to speech.

The OCR system's reliability depends on the camera used to capture the text's raw image. With different factors affecting the quality of the camera focus, the amount of noise present, resolution of the picture, the Tesseract algorithm achieved 100% accurate. To estimate the OCR accuracy, the OCR output can then be compared to the text of the authentic record. The very last system of the above technique is whole because the above text is synthesized and transformed into audio format and performed the use of earphones or mini speakers connecting to the onboard audio jack of Raspberry Pi as shown in Figure 7.
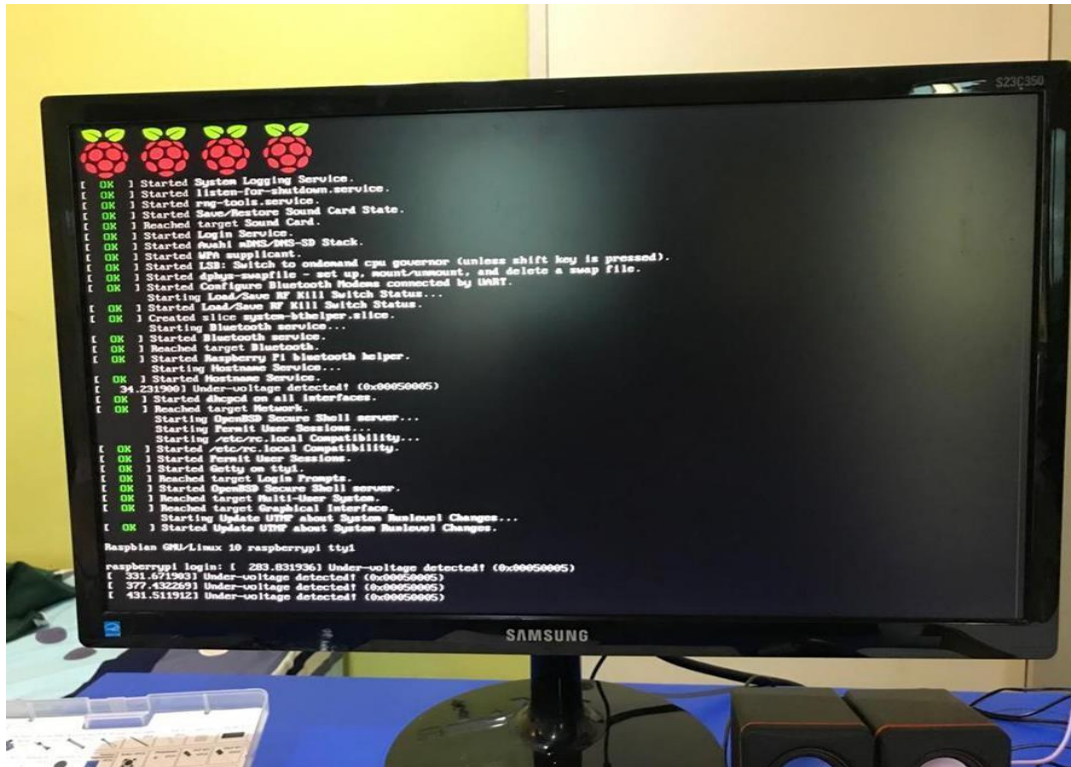
Figure 7 Whole Setup for Text to Speech Conversion.

The device was able to recognize the small letters as well and process them and then turn them into speech successfully. As observed from the experimental results, Tesseract fares reasonably well with respect to the core recognition accuracy on user- specific handwritten samples of isolated free-flow text. The performance of the system is validated on a four-user recognition engine.

The device is designed to compensate for Braille and other traditional methods of reading documents and complex books that make them less available as an image, with text given as inputs into the TTS engine, a command-based tool, t is processed by the TTS command. A number of different samples of handwritten documents has been identified from several people and a font size of fewer than 10 points to 8 points has been identified. The Tesseract algorithm provides results with 100% accuracy and provides better accuracy results if the images are in grayscale mode compared to color pictures.
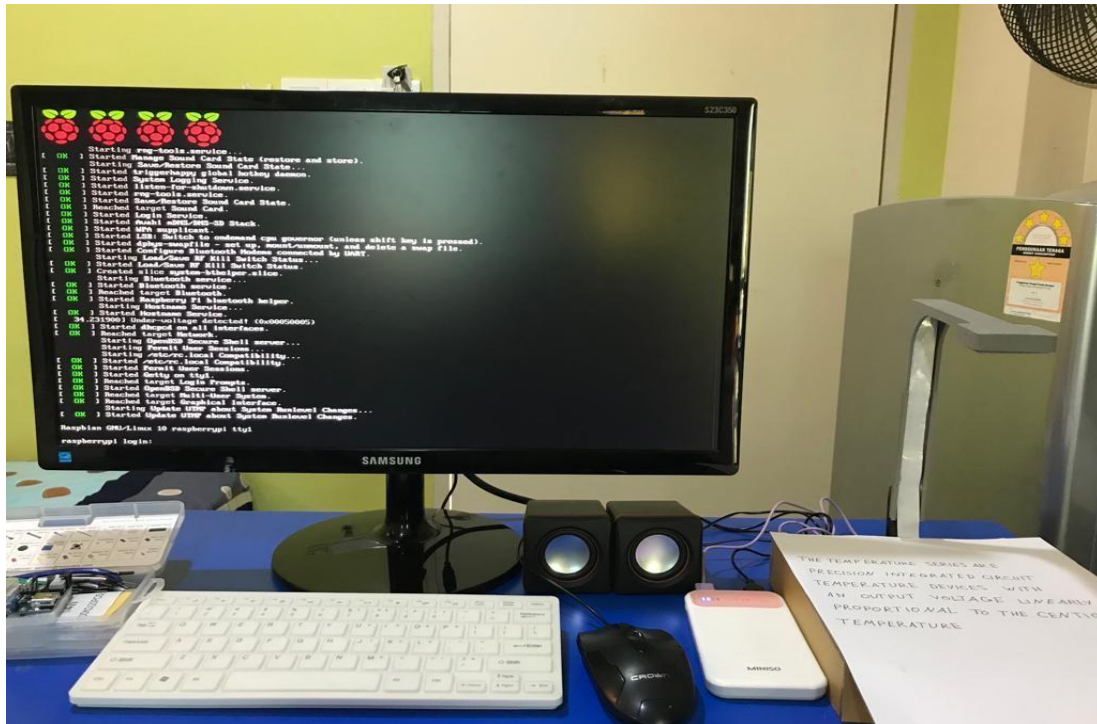
Figure 8 Hardware Setup.

Discussion

In this research we were able to create a device composed of Raspberry Pi and 5 MP Raspberry Pi camera installed on a stand where it helps the visually impaired to read handwritten texts such as prescriptions, lists of account and the like, and this means any document written by hand, written regularly and not randomly where the experience of several different characters and different people have been successfully identified, the without constraint in line length, whether capital letters, uppercase or small letters and the font size (8 Pt) was recognized to successfully. The device is easy to carry and low energy consumption where the device has been running for lone battery 5V.

Conclusion

The design is a portable device that is less complicated and one that can be carried in one's hands, is cheap and easy to use. Blind people and children can use it. This model is a standalone program developed for Raspberry Pi and Python 3. This design defined a system for reading handwriting and describing handheld objects to help blind people, OCR is used to perform word recognition in major written text areas and to convert this to audio output. The camera talks and acts as, with the input of paper as the image, flowing as the Raspberry Pi is switched on. The Tesseract algorithm treats the image effectively and reads it easily and provides great assistance for disabled people. It is also cost-effective as a tool for persons with visual impairment. We applied the algorithm to a number of handwritten texts and noticed that they were converted successfully. The device is easy to use and can be used by the public and children.

The main benefits of this design are that it takes less time to identify and read the text and can recognize text in lowercase letters with lower operating costs. This program may be used by partially blind individuals and older people with different eye problems. It plays an important part in educating visually impaired students. Naturally, if hearing gets a reader faster through text, also when time is of interest, it has to be considered more effective using device as the output is quicker. Certain benefits include more versatility, high accuracy; it is better adapted to different lighting conditions, and its function can be quickly performed. There is no difficult limitation on this paper. A regular handwriting font size is not important for the device, as it is able to recognize any size. The challenge is that the device can read small and large letters, so it is easy to adjust the distance between the camera and the written text by organizing the camera without the need for distance between the base and the camera.

References

1. Bhargava, Anusha, Karthik V Nath, Pritish Sachdeva, and Monil Samel. 2015. "Reading Assistant for the Visually Impaired." International Journal of Current Engineering and Technology 5(2):1050–55.
2. Chukwunazo, Ezeofor J. and Georgewill M. Onengiye. 2016. "Design and Implementation of Microcontroller Based Mobility Aid for Visually Impaired People." International Journal of Science and Research (IJSR) 5(6):680–86.
3. Cumberland, Phillippa M. and Jugnoo S. Rahi. 2016. "Visual Function, Social Position, and Health and Life Chances the UK Biobank Study." JAMA Ophthalmology 134(9):959–66.
4. Clausner, Christian, Stefan Pletschacher, and Apostolos Antonacopoulos. 2016. "Quality Prediction System for Large-Scale Digitisation Workflows." Proceedings - 12th IAPR International Workshop on Document Analysis Systems, DAS 2016 (1):138– 43.
5. Eldin Gamal, Bahaa, Ahmed Nasr Ouda, Yehia Zakaria Elhalwagy, and Gamal Ahmed Elnashar. 2017. "Embedded Target Detection System Based on Raspberry Pi System." 2016 12th International Computer Engineering Conference, ICENCO 2016: Boundless Smart Societies 154–57.
6. Elmannai, W., & Elleithy, K. (2017). Sensor-based assistive devices for visually- impaired people: Current status, challenges, and future directions. Sensors, 17(3), 565.
7. Garousi, Vahid, Michael Felderer, Çağri Murat Karapiçak, and Uğur Yilmaz. 2018. "What We Know about Testing Embedded Software." IEEE Software 35(4):62–69.
8. Gurav, Mallapa D., Shruti S. Salimath, Shruti B. Hatti, Vijayalaxmi I. Byakod, and Shivaleela Kanade. 2017. "B-LIGHT: A Reading Aid for the Blind People Using OCR and OpenCV." International Journal of Scientific Research Engineering & Technology 6(5):2278–2882.
9. Rossum, Guido Van and Fred L. Drake. 2018. "Porting Python 2 Code to Python 3." Strategy 3–8.
10. Saleous, Heba, Anza Shaikh, Ragini Gupta, and Assim Sagahyroon. 2016. "Read2Me: A Cloud-Based Reading Aid for the Visually Impaired." 2016 International Conference on Industrial Informatics and Computer Systems, CIICS 2016.