Journal of
Electrical & Electronic Systems

Baskaran and Lakshmi, J Electr Electron Syst 2018, 7:3
DOI: 10.4172/2332-0796.1000268

Review Article Open Access

Image to Audio Conversion using Portable Camera

Baskaran S[1]* and Lakshmi D[2]

[1]Department of Electrical Engineering, S.A Engineering College, Tamil Nadu, India
[2]Department of Electrical Engineering, Sree Sastha Institute of Technology, India

Abstract

This method proposes a camera based assistive text reading for the visually impaired persons (partially or completely blind) to read the text in the label and from the products packaging in their daily lives. This proposes a efficient and effective based method to define the region of interest (ROI) in which objects can be isolated from the cluttered backgrounds. This ROI extracts the text localization and recognition to acquire text information. Localizing the text from the object ROI is done using the novel text localization algorithm by learning the features of stoke orientations and edge pixels in the ad boost model. The text characters in the localized text regions are recognized by optical character recognition (OCR) software. The codes from the OCR are recognized and it is an output to the blind user in speech.

Keywords: Raspberry; Camera; Speaker; Raspbian OS; Python

Introduction

It is a fact that all over the world that the visually impaired (partially or completely blind) people face a lot of difficulties in reading, identifying a product, and avoiding the obstacles. According to the development in today's technology towards the computer vision, digital camera and portable computers it is feasible to develop a camera-based technology that combines computer vision technology with other commercial products such as OCR systems.

Reading is very essential in today's society. Everywhere the printed text is in the form of Reports, bank statements, receipts, restaurant menu's etc. so the blind users face a difficulty in reading these forms. In order to reduce the frustrated problem the method Text to Voice Adaption Using Portable Camera is referred. The method which is already existed a carries major drawback in size and not portable. To reduce this drawback, we choose an embedded platform raspberry pi (Model 3) which acts as a mini Cohere the camera is interfaced to the raspberry pi board and the captured images is processed to the Rpi board. ROI method is used to localize and recognize the text.

The text codes from the ROI are recognised by the Optical Character recognition (OCR) and the captured image is processed using the python programming language. The text from the OCR is compared with the text in the Open CV library to identifying the orientations and edge pixels. Therefore the captured images are converted in to the text. The text codes are processed to the pytxx library and it is output to blind users in speech. And in addition we add a ultrasonic sensor to alert the blind users by the speaker in avoiding the obstacles.

Litreature Survey

X Chen et al. [1] automatic detection and recognition of signs from natural scenes. To extract the hand-held object from the camera image, this system going to develop a motion-based method to obtain a region of interest (ROI) of the object. Then, perform text recognition only that ROI.

K. Kim et al. [2] detecting text in natural scenes with stroke width transform, to post process the image, support vector machines (SVM) had been proposed to do classification on the extracted features. Some kernel functions which are being tested are second degree polynomial, radial basis function (RBF), exponential radial basis function (ERBF), sigmoid, and odd-order Bsp line. RBF and ERBF achieved the best classification accuracy compare to other kernels used. SVM also automatically helps RBF kernel to define the centers during optimization. Brodatz texture album is used in this study to test out the result. In the study, a combined GLCP with SVM post-processing showed a marked improvement over other classifier in terms of classification accuracy.

Saleous H et al. [3] a Cloud- based reading aid for the visually impaired, in this work, we discuss the design and Implementation of two assistive platforms, in one, we combine today's' Smartphone capabilities with the advantages offered by the rapidly growing cloud resources, and the other utilizes a more economical approach making use of cost-effective microcontrollers. Both approaches make us of an Optical character Recognition (OCR) engine on the cloud and use local resources for the Text-to-Speech (TTS) conversion. Prototypes are successfully developed and tested with favorable results.

Gopinath J et al. [4] text to Speech Conversion System using OCR. The main idea of this project is to recognize the text character and convert it into speech signal. The text contained in the page is first pre-processed. The preprocessing module prepares the text for recognition. Then the text is segmented to separate the character from each other. Segmentation is followed by extraction of letters and resizing them and stores them in the text file. These processes are done with the help of MATLAB. This text is then converted into speech.

Isewon I et al. [5] design and Implementation of Text To Speech Conversion for Visually Impaired People, A Text-to-speech synthesizer is an application that converts text into spoken word, by analyzing and processing the text using Natural Language Processing (NLP) and then using Digital Signal Processing (DSP) technology to convert this processed text into synthesized speech representation

***Corresponding author:** Baskaran S, Assistant Professor, Department of Electrical Engineering, S.A Engineering College, Poonamallee - Avadi Road, Veeraraghavapuram, Thiruverkadu, Chennai, Tamil Nadu 600077, India, Tel: 044-26801999; E-mail: baskaran@saec.ac.in

Received June 22, 2018; **Accepted** July 30, 2018; **Published** August 06, 2018

Citation: Baskaran S, Lakshmi D (2018) Image to Audio Conversion using Portable Camera. J Electr Electron Syst 7: 268. doi: 10.4172/2332-0796.1000268

of the text. Here, we developed a useful text-to-speech synthesizer in the form of a simple application that converts inputted text into synthesized speech and reads out to the user which can then be saved as an mp3 file. The development of a text to speech synthesizer will be of great help to people with visual impairment and make making through large volume of text easier.

Detailed Study

This paper presents a prototype system of camera-based assistive text reading. This system consists of three functional components, screen capture, image processing, audio output. The screen capture component collects the objects of interest in the form of image or video. In our system, it corresponds to a camera attached to a pair of sunglasses. The Data processing system is used for our proposed algorithms, including 1) Region-of-interest to collects the objects from the cluttered backgrounds or from other surroundings; 2) text localization to attain the image regions to containing text and text recognition to transform the image-based readable codes. The audio output component is to inform the visually impaired persons of recognised text codes. The 3.5 mm audio jack is interfaced with the Rpi board which receives the audio signals and it is amplified and comes out as a speech.

In order to extract the text recognition from the objects, a motion based objects detection is applied to determine the user's object of interest while recording the video ground and it calculates the foreground at each frame. Then proposed text localization algorithm is applied to the object of interest to extract text regions. Text regions are generated by layout analysis of colour uniformity and text alignment. After text region localization, OCR is performed to recognise text in the text localization region. Then the recognised words or transformed into speech for the blind users.

Object region detection

To identify the text from the handheld objects in the camera veiw. We propose a resonably wide angle in our system. However it may result in capturing all other images which has been appearing in the camera veiw. For example it the user is shopping in an supermarket. To extract the object of interest from the objects we ask the user to shake the objects which contains text they wish to identify, then we employ a motion based method to localize the text from the cluttered backgrounds. (BGS) background subtraction is the effective approach to detect the locomotive objects for the video survillance with the stationary cameras [6].

Automatic text extraction

To automatically localize the text in the image regions we design a learning-based algorithm. To handle the complex backgrounds we propose two text features based on stroke orientations and edge distribution.

Text stroke orientation: The basic structure of text characters consists of stroke with constant or variable orientations. To describe the local structure of text characters we propose a new type feature, stroke orientation. It analyses the pixel-level and from the stroke orientations is perpendicular to the gradient orientation at pixel of stroke boundaries.

Distribution of edge pixels: In this feature the text characters appear in the form of stroke boundaries. It also describes the characteristic structure of text. Edge density of text region is the most commonly used feature. It measures the edge density but does not give any spatial information of edge pixels. It is used for distinguishing text regions from relatively clean background.

Text recognition and audio output

Text recognition is perfomed by Optical character recognition (OCR) to provide output of informative words from the text regions. These text regions are compared with the library and they are sent as audio signals in the audio jack and amplified and delivered as output to the blind users.

Proposed Architecture

This proposed system contains Raspberry pi as the central unit. Camera, speaker, ultrasonic are the major components interfaced with the raspberry pi. The power supply of 5VDC is supplied to the board through the transformer and rectifier circuit else an separate power bank can be used for the board. The system has been divided into two sections. Transmitter section collects the images from the products or objects from the camera and it has been proposed through the camera speaker interface as shown in the Figure 1a.

Receiver section has been shown in the Figure 1b. Power supply of 5v dc is supplied through the transformer and rectifier circuit to the rasperry pi board and separate 5v power suppy for amplifier and ultrasonic sensor.

Power Supply

LPC2148 works on 3.3V Power Supply, So LM117 a 1A low dropout regulator designed to provide 3.3V from a 5V supply. It is ideally suited for systems which contain both 5V and 3.3V logic, with power provided from 5V bus. Because the LM3940 is a true low dropout regulator, it can hold its 3.3V output in regulation with input voltages as low as 4.5V [7].

A Power supply unit is assigned as follows:

- Transformer
- Rectifier
- Filter
- Regulator etc,

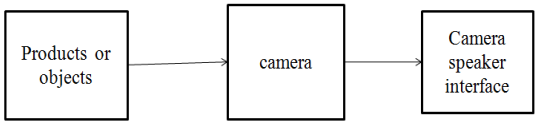

Figure 1a: Block diagram of Transmitter Section.

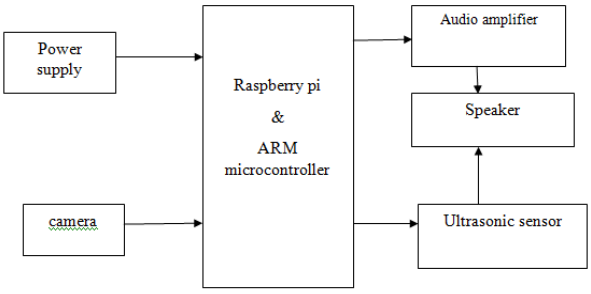

Figure 1b: Block diagram of receiver section.

As shown in the block diagram above Figure 2.

Transformer: A transformer is a static device that transforms electric power from one circuit to electric power of the same frequency in another circuit. The voltage can be raised or decreased in a circuit, but with proportional variations in the current ratings [8].

In this system it is used to step down 230V AC to 9V AC supply and provides isolation between power grids and circuit.

Rectifier: A rectifier is a device that converts alternating current (AC), which periodically reverses direction, to direct current (DC), which flows in only one direction. The process is known as rectification.

Filter: The output from the rectifier is pulsating D.C. These pulsations are due to the presence of A.C. component in the rectifier output. The filter circuit removes the A.C. component, thus steady D.C. voltage is obtained across the load.

Regulator: A regulated power supply consists of an ordinary power supply and voltage regulating device. The output of ordinary power supply is fetched to the voltage regulator which produces the final output. The output voltage remains constant whether the load current changes or there are fluctuations in the input A.C. voltage.

Transistor model are connected to the common terminals. The output terminals Y0-Y3 (PNP) model is connected to the terminals UP and ZP.

Raspberry Pi

In our proposed prototype system we have used a raspberry pi advanced board model 3 which acts a mini computer. Built on the latest Broadcom 2837 ARM 11, 64bit processor, the new generation Raspberry Pi 3 is advanced and more powerful than its predecessors. With built-in wireless and Bluetooth connectivity, it becomes the ideal IOT ready solution Figure 3.

Specifications

SoC: Broadcom 2837

CPU: ARM 11, 1.2GHz

GPU: Broadcom Video Core IV

RAM: 1GB LPDDR2 (900 MHz)

Networking: 10/100 Ethernet, 2.4GHz 802.11n wireless lan.

Bluetooth: Bluetooth Classic, Low Energy

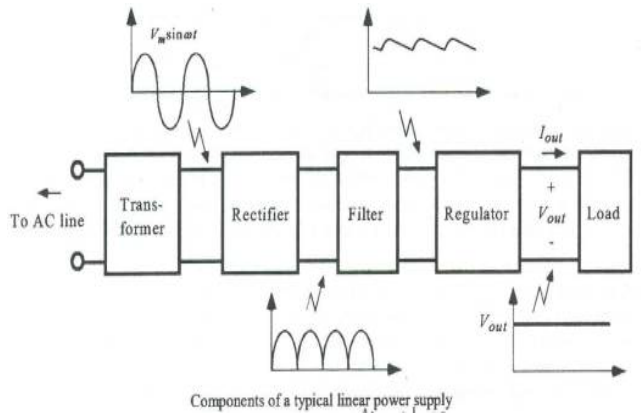

Figure 2: Block diagram of power supply.

Storage: micro SD external

GPIO: 40-pins, populated

Ports: HDMI, 3.5 mm analogue audio-video jack, 4 × USB 2.0, Ethernet, Camera Serial Interface (CSI), Display Serial Interface (DSI)

USB camera

The image is captured using camera and sent to the Raspberry Pi through USB ports. You can use a standard USB webcam to take pictures and video on the Raspberry Pi Figure 4 [9].

Camera serial interface (CSI)

The camera serial interface defines an interface between peripheral device (camera) and the processor. It built a standard interface between camera and processor for mobile applications Figure 5.

Audio amplifier and speaker

The Audio signal from the raspberry pi is taken from 3.5 mm jack and HDMI. The audio signal from the raspberry pi will not drive the speaker. So we use a external audio amplifier to amplify the signal Figure 6.

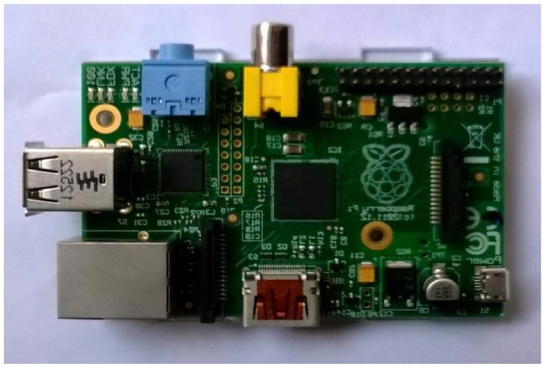

Figure 3: Raspberry pi.

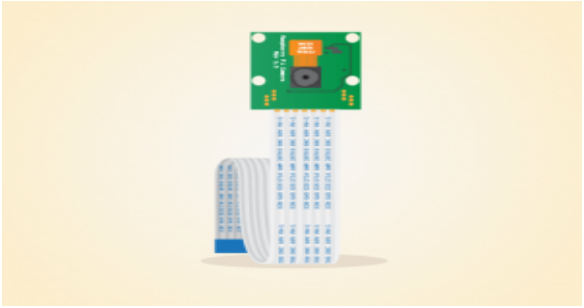

Figure 4: CMOS camera.

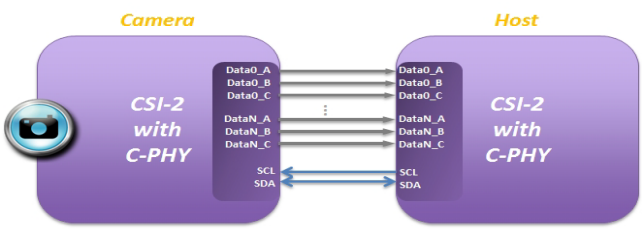

Figure 5: Camera serial interface (CSI).

Secure digital

The Rpi platform has an internal slot for a secure digital (SD). It is a non volatile memory card used in portable devices, mobile phones, cameras, computers, etc. Figure 7.

Operating system

Raspbian is the operating system implemented on our prototype. Raspbian is the free operating system optimized for raspberry pi. Raspbian has a set of basic programs and utilizes the raspberry to run. It comes over with 35,000 packages and pre-complicated software with nice format for easy installation on your raspberry pi.

Flow Chart of the Process

Steps of the process

1. Power supply of 5V DC is supplied to the raspberry pi and ARM microcontroller.

2. Camera is interfaced with raspberry pi board and captured images have been processed through the board.

3. Image processing has been done and sends audio signals to the jack and amplified gives output to the speaker.

4. Ultrasonic sensor has been interfaced with the board transmits and receives the pulse signal according to the duty cycle.

5. Distance is measured using the duty cycle Figure 8.

Simulation and Hardware

From the programming side first we have to configure the Rpi in communication mode and we have to upload the coding in the python programming language. After the compilation program is in the online simulation mode. Online simulation is used to check that how program is running step by step.

Figure 6: Headphone.

Figure 7: SD slot.

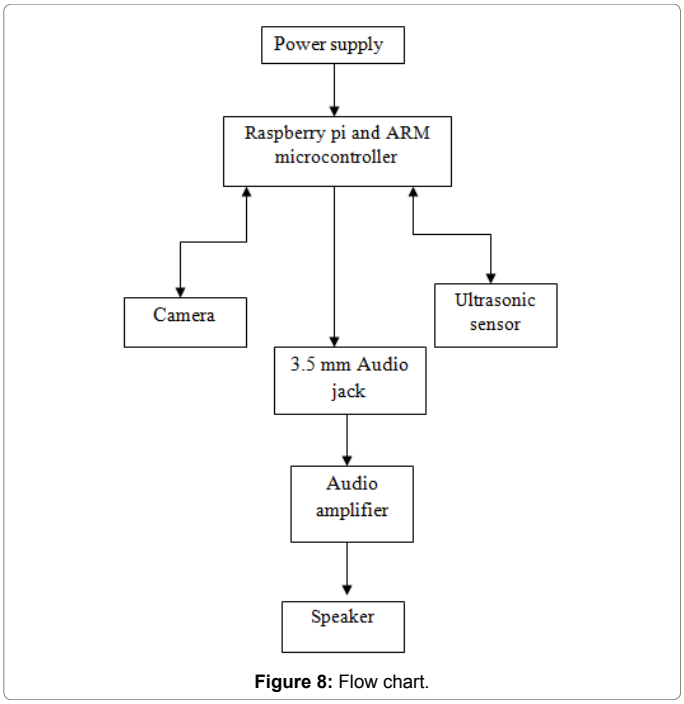

Figure 8: Flow chart.

Conclusion and Future Scope

Our prototype system is to read the printed text in the objects or products for assisting blind persons. To solve the common aiming problem we have implemented motion-based method to detect the objects of interest. Text extractions are done using stroke orientation and distribution of edge pixels. The text characters are recognized using Optical Character Recognition, the text codes are transformed as speech for blind persons. Our future work will extend the text localization algorithm with further more features and we will address the human interface issues associated with text reading by the blind user.

References

1. World Health Organization (2009) 10 facts about blindness and visual impairment.
2. http://www.cdc.gov/nchs/nhis/nhis_ad.htm
3. International Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2005).
4. Chen X, Yuille AL (2004) Detecting and reading text in natural scenes. in Proc Comput Vision Pattern Recognit 2: 366-373.
5. Chen X, Yang J, Zhang J, Waibel A (2014) Automatic detection and recognition of signs from natural scenes. IEEE Trans Image Process 13: 87-99.
6. Bilgaiyan S, James S, Bhonsle SS, Shahdeo S, Keshavamurthy (2016) Android Based Signboard Detection using Image and Voice Alert System. IEEE Int Conf.
7. Chowdhury S, Dey A, Sing JK, Basu DK, Nasipuri M (2014) A Novel Elastic Window for Face Detection and Recognition from Video. International Conference on Computational Intelligence and Communication Networks.
8. Basavaraju R, Chetana H (2015) Traffic Signal. Time Analysis and Voice - Based App for Visually Impaired Pedestrians. ICERECT.
9. Kim K, Jung K, Kim J (2014) Detecting text in natural scenes with stroke width transform. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.