International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 2, No. 3, September - November, 2017 | Pages 63-68

NEURO-OCR BASED ASSISTIVE SYSTEM FOR TEXT RECOGNITION WITH VOICE OUTPUT

D.Maheswari[1], S. Ohmshankar[2]
[1,] Sr Assistant Professor, Department of Electronics and communication Engineering, Agni College of Technology, Chennai, Tamil Nadu
[2,]Assistant Professor, Department of Electronics and communication Engineering, Chennai, Tamil Nadu
[1]maheswarid.ece@act.edu.in [2] ohmshankar.ece@act.edu.in

ABSTRACT:
A designing of a system, which captures the text automatically of a Median filter combination and these details, were verified using Raspberry Pi processor for authentication. This system captures the text of median filter further processing for the character recognition. This paper makes use of an onboard computer, which is commonly termed as Raspberry Pi processor. It acts as heart of the project. This onboard computer can efficiently communicate with the output and input modules, which are being used. The device, which is able to perform the task, is a Raspberry Pi processor. When any image passes by the system, the image of the text is captured using camera. The image of the text details are fed as input to the Raspberry Pi processor. The Processor takes responsibility to check the authentication details of every median filter. Once the details are recognized then the processor operates it detects an unauthorized image of text is detected. To perform this task, Raspberry Pi processor is programmed using embedded 'Raspbian'.

Keywords-Median filter, Raspbian, Raspberry PI

INTRODUCTION

To design a Text Recognition System Using Raspberry Pi in the application of the Median filter combination.Generally, an automatic text recognition system is made up of five modules; Text Detection, Segmentation, Classification, Text Recognition, OCR Segmentation modules.

- Firstly, text localization from filtered images
- Secondly, character segmentation from localized images.
- Finally, optical character recognition of extracted characters.

PROPOSED SYSTEM

The proposed system uses Raspberry PI instead of PC.Python is used as programming language instead of MAT lab.Neuro OCR technique is used for image processing.Ultrasonic sensor is used to intimate the person regarding the distance.The recognized text will be given to the person as audio output through headphone. Background Subtraction and Feature Extraction algorithms process the text.

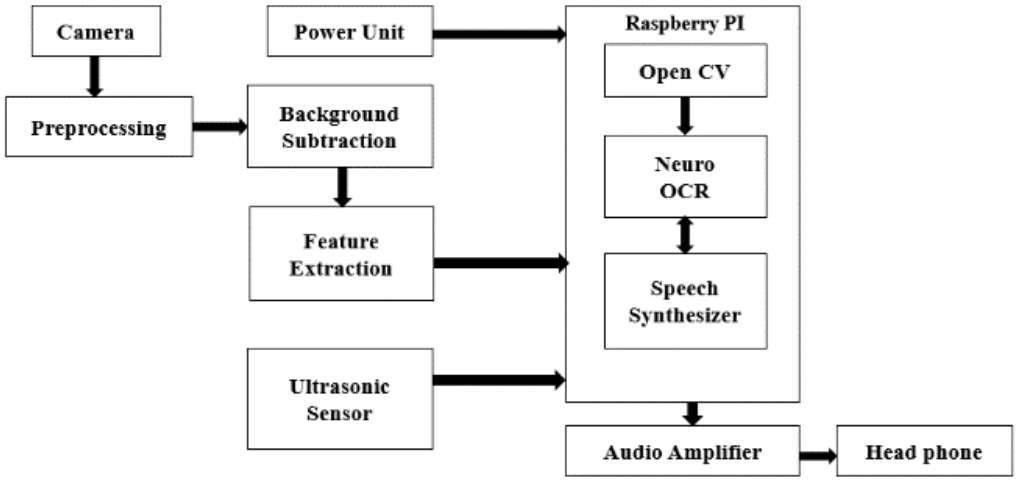

Fig 1: Proposed system block diagram

SYSTEM FLOW

Phase 1: In this step, we have to detect all the texts in the current camera frame. Two broad categories in which they can be defined are:

- Segmentation
- Classification

Segmentationis the process of dividing an image into multiple segments. This process is to simplify the image for analysis and make feature extraction easier. One important feature that can be exploited from text is the high number of vertical edges. However, before that, we need to do some handy pre-processing of the current image namely:

Conversion to Gray Scale: The Red, Green, and Blue components are separated from the24-bit color value of each pixel. To convert the image captured into Gray Scale code can be written in Python IDLE:

```
img = cv2.imread ('2.jpg',0)
```

Blur the Image: The result of blurring an image by a Gaussian functions. Using this technique image looks sharper or more detailed if we are able to perceive all the objects and their shapes correctly in it. Using Code:

```
Blur=cv2.GaussianBlur(img(5,5),0)
```

Classificationtechnique is utilized to distinguish the text region from the given picture. The principle target of such sort of systems is to confine the text region from images of the Median filter that are captured from the camera mounted on the Raspberry Pi. The quality of the image forms an important part of this technique so preprocessing the image helps in improving the quality.

The text usually appear to have high contrast areas in the image (black-and-yellow or black-and-white). The numbers and letters are placed in the same row (i.e. at an identical vertical level), which results in frequent changes in intensity horizontally .This provides the basis for detecting the changes in the horizontal intensity horizontally. This provides the basis for detecting the changes in the horizontal intensity, as the rows that will contain the texts are expected to show sharp variations. The reason for this sharp variation is the contrast between the letters and its background.

Phase 2: Edge Detection helps to characterize the boundaries and therefore are a problem of fundamental importance while processing the image. Edges in images are the areas where strong intensity contrasts are present ,a sudden variation in the intensity from one pixel to the next Detecting the edges of an image significantly reduces the amount of data and it helps in filtering out the useless information, while preserving the important structural properties of an image. There are many ways to perform the edge detection. However, the majority of various methods can be grouped in to two different categories, gradient and laplacian. The gradient methods detect the edges by finding out the maximum and minimum in the first derivative of the image. The Laplacian method

searches for the zero crossings in the second derivative of the image to find the edges. An edge has the one-dimensional shape of the ramp and calculation the derivative of the image can highlight its location.

Phase 3: Character Segmentation

The character segmentation process acts as bridge between text detection and optical character recognition modules. Its main function is to segment the characters in the selected candidate region (extracted text) such that each character can be sent to the optical character recognition module individually for recognition.

Normalized or standardized are of a fancy format the conditions of the texts are important criteria for efficient segmentation because if numbers are of a fancy format the conditions of the text as described .Once the text is localized we proceed to obtain the individual characters .A text as described above has high intensity variation regions. This forms the basis for character segmentation. Sometimes it is observed that along with text, various texts may be present, which have to be removed. By various observation, we observed that for the text regions the amount of white on black is specific for the number regions and falls within a certain range. Morphological technique are used to remove small white areas, which escape range corrections. Finally, individual characters are extracted to pass on through the optical character recognition system. Segmentation is one of the most important processes in the automatic text recognition, because all further steps rely on it. If the segmentation fails, a character can be improperly divided in two pieces, or two characters can be wronged merged which would lead to the failure of following stages of recognition.The second phase of the segmentation is an enhancement of segments.

Preprocessing stage

Before we can proceed with the segmentation stage, we must ensure that the text obtained is cleared off most of the unwanted characters or graphics like state name or flags etc. We proceed to do so by scanning the text vertically and horizontally and ignoring those rows and columns, which have too much white and black. This is justified as those areas containing the numbers have black areas which be in a particular range .This range by experiments was found to be between 0.2 to 0.8 times the number of pixels horizontally and vertically.The character segmentation process takes the extracted text region from the preceding module, as the input .The input is a colored JPEG image. For our process, we work only binary images and thus the first part of segmentation is banalization of the image.

Phase 4: The Image Deskewing Mechanism

Either way, Image skew could severely hamper subsequent steps of character extraction and recognition. It is thus desired to deskew the image before passing the localized text on the next LPR step. Below is a sample image illustrating the need of de skewing technique.

Phase 5: Optical Character Recognition

Introduction

The neural networks are typically made up of many artificial neurons. An artificial neuron is an analogy to biological neuron. It is simply electronically modelled to the biological neuron. The number of neurons that are used depends on the task at hand. The number of neurons used can be few as two or three or large as two or several thousand. There are many ways of connecting artificial neurons together to form a neural network. Some of the ways are discussed below.

Feedforward network:

In feedforward Neural Network, each input into the neuron has its own weight associated with it. A weight can simply be a floating-point no. and it is these that we adjust when we come to train the network. The weights in most of the neural networks can be both negative and positive, therefore, it helps in providing inhibitory influences to each input. As each of the input enters the nucleus it is then multiplied by its weight. The nucleus sums up all these new input values and gives us the activation which is again a floating-point no. and can be negative or positive. The threshold value is decided and if the activation value is greater than a threshold value, the neuron outputs 1(considering

these are two outcomes 1 and 0 to the input) as a signal. If the activation is less than the threshold value the neuron then outputs zero. A neuron can take any number of inputs from one to n, here n is the total number of inputs. The inputs, therefore, may be represented as x1, x2, x3…xn. The corresponding weights for the inputs can be represented as w1, w2,w3,w3..Wn. The weighted sum ofthe links and its corresponding weights is called the activation value as discussed above.
a=x1w1+x2w2+x3w3…xnwn
Where, a is the activation value.
Each of the input is sent to every neuron of the hidden layer and each hidden layers neurons output is connected to every neuron of the next layer. There is no predefined number of the neurons to be present in a particular layer it can be arbitrary and it totally depends on the problem. Each of the input is sent to every neuron of the hidden layer and each hidden layers neurons output is connected to every neuron of the next layer. There is no predefined number of the neurons to be present in a particular layer it can be arbitrary and it totally depends on the problem.

Back Propagation Network Algorithm:

A back propagation networks learns by example various sets of datasets are provided as input. The various inputs provided helps the network to calculate and recalculate the networks weight value so that when the network is trained it can give the required output. The network is initialized by first setting random weights, which generally have very small value such as values between -1 and 1. There are two passes in the Back Propagation Algorithm. After the networks is setup with the random weights the output is calculated this is called the forward pass the result obtained in the forward pass may not be equal to the required result or the target and so the error is calculated for each neuron which is Target-Actual Output. The error calculated for each neuron is then mathematically used to change the weights so the next time the forward pass will have minimum the error. The character is recognized after training the network with various datasets of the particular character to get maximum accuracy and minimum error.

HARDWARE DESCRIPTION

A.RASPBERRY PI

The Raspberry Pi has a Broadcom BCM2835 system on a chip (SoC), which includes an ARM1176JZF-S 700 MHz processor, VideoCore IV GPU and was originally shipped with 256 megabytes of RAM, later upgraded (Model B & Model B+) to 512 MB. It does not include a built-in hard disk or solid-state drive, but it uses an SD card for booting and persistent storage, with the Model B+ using a MicroSD.

B.WEB CAMERA

Inbuilt sensitive microphone and Image Sensor High Quality CMOS Sensor.
Anti-flicker 50Hz, 60Hz or outdoor.
Resolution Hardware: 500K pixels.
Image Quality: RGB24 or I420.
Exposure: Auto or manual and Angle of view: 58 Degree.
Interface: USB2.0; Frame Rate: 30 fps (MAX).

C.ULTRASONIC SENSOR

Ultrasonic sensor emit ultrasonic pulses, and by measuring, the time of ultrasonic pulse reaches the object and back to the transducer. The sonic waves emitted by the transducer are reflected by an object and received back in the transducer. After having emitted the sound waves, the ultrasonic sensor will switch to receive mode. The time elapsed between emitting and receiving is proportional to the distance of the object from the sensor.

SOFTWARE DESCRIPTION

A.OPEN CV

OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCVwas built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. This library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects.

B. Operating System

The Raspberry Pi primarily uses Linux kernel-based operating systems. The ARM11 is based on version 6 of the ARM, which is no longer supported, by several popular versions of Linux, including Ubuntu. The install manager for Raspberry Pi is NOOBS. The OSs included with NOOBS are:

- Arch Linux ARM
- Open ELEC
- Pidora (Fedora Remix)
- Raspbmc and the XBMC open source digital media center
- RISC OS – The operating system of the first ARM-based computer.

Boot Process

The Raspberry Pi does not boot as a traditional computer. The Video Core i.e. the Graphics processor actually boots before the ARM CPU.
The boot process of the Raspberry Pi can be explained as follows:

- When the power is turned on, the first bits of code to run is stored in a ROM chip in the SoC and is built into the Pi during manufacture. This is the called the **first-stage bootloader**.
- **The physical addresses perceived by the ARM core is actually mapped to another address in the Video Core (0xC0000000 and beyond) by the MMU (Memory Management Unit) of the Video Core**.
- After starting the operating system, the GPU code is not unloaded. In fact, start.elf is not just firmware for the GPU, It is a proprietary operating system called Video Core OS (VCOS). When the normal OS (Linux) requires an element not directly accessible to it, Linux communicates with VCOS using the mailbox messaging system.

CONCLUSION

There are frequent situations in which a system able to recognize the text can be useful. This paper presents few such situations, a system designed to satisfy the requirements, Visicar, and some experimental results obtained with this system.The system has been designed using a modular approach which allows easy upgrading and/or substituting of various sub-modules thus making it potentially suitable for a large range of vision applications. The performances of the system makes it a valid choice among its competitors especially in those situations when the cost of the application has to be maintained at reasonable levels. Furthermore, the modular architecture makes Visicar extremely flexible and versatile.

REFERENCES

[1] Wisam Al Faqheri and SyamsiahMashohor, "A Real-Time Malaysian Automatic Text Recognition (M-ALPR) using Hybrid Fuzzy" ,IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.2, February 2009.
[2] Saeed Rastegar, Reza Ghaderi, GholamrezaArdeshipr&NimaAsadi, " An intelligent control system using an efficient Text Location and Recognition Approach", International Journal of

Image Processing (IJIP) Volume(3), Issue(5) 252, 2009. [3] V. Ganapathy and W.L.D. Lui, "A Malaysian Vehicle Text Localization and Recognition System", Journal of Systemic, Cybernetics and Informatics, Vol. 6, No. 1, 2008.

[3] 3D surface tracking and approximation using Gabor filters, JasperJullHenrik Sen, South Denmark University, March 28, 2007

[4] O. Martinsky, “Algorithmic and Mathematical Principles of Automatic Number Text Recognition System”, B. Sc. Thesis, BRNO University of Technology, 2007.

[5] Localization from Vehicle Images: An Edge Based Multi-stage Approach”, International Journal of Recent Trends in Engineering, Vol 1, No. 1, May 2009

[6] SatadalSaha, SubhadipBasu, MitaNasipuri, Dipak Kumar Basu “An Offline Technique for Localization of Texts for Indian Commercial Vehicles”

[7] T. L. Chien, H. Guo, K.L. Su and S.V. Shiau, "Develop a Multiple Interface Based Fire Fighting Robot," IEEE International Conference on Mechatronics, May 2007.

[8] K. L. Su, "Automatic Fire Detection System Using Adaptive Fusion Algorithm for Fire Fighting Robot," IEEE International Conference on Systems, Man and Cybernetics, Vol. 2, Oct. 2006.