ORIGINAL RESEARCH PAPER
Volume - 9 | Issue - 12 | December - 2020 | PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

OCR BASED FACILITATOR WITH EMOTION DETECTION FOR THE VISUALLY CHALLENGED

Computer Science

Srividhya Ravichandran

ABSTRACT

According to the World Health Organization (WHO), around 285 million people are estimated to be visually impaired worldwide and 90 percent of them live in developing countries. Visually impaired people report numerous difficulties with accessing printed text using existing technology, including problems like alignment, focus and accuracy. There is a need for a portable text reader that is readily available and affordable to the community. Inclusion of specially enabled people in the IT revolution is both a social obligation as well as a computational challenge in the rapidly advancing digital world today. This paper addresses a complete image processing system built on an embedded platform that enables text conversion to speech to help the visually challenged. In addition, we also discuss approaches to extract emotion from text. The field of emotion detection has been applied in applications such as emotional retrieval, capturing emotions in multimedia tagging, detecting insulting sentences in conversations and so on.

KEYWORDS

1. INTRODUCTION

Optical Character Recognition (OCR) is the alphanumeric recognition of printed or hand-written characters. It is an active field of research in Computer Vision and Pattern Recognition. The proposed system consists of a webcam interfaced with a Raspberry Pi board which accepts a page of printed text. This document is then sent to the Image Processing Module, which performs various tasks like skew correction, thresholding, feature extraction, contour detection and segmentation. The OCR Module scans this page into a digital text document. The processed text file is then fed into an audio amplifier for speech output. The emotions are identified and converted to speech.

In this paper, we utilize complex algorithms to implement emotion detection from the text. This is done by recognizing patterns of emotions and utilizing them to understand the overall emotion in the statement.

This system finds interesting applications in libraries, auditoriums and offices where instructions and notices are to be read. This project is a step in developing a prototype for the blind and visually challenged.

2. RELATED WORK

Optical Character Recognition has been an active subject of research since a decade. The rapid growth of digital libraries worldwide pose new challenges for document image analysis research and development.

Nagaraja L, Nagarjun R.S et al. [1] presents a vision based text recognition system using Raspberry Pi. This paper proposes a complete optical recognition system. Image pre-processing techniques are applied where unwanted noise is removed by applying threshold morphological transformations like morphological, dilation, black hat and discrete cosine. Text areas are extracted by drawing bounding boxes around the required text and slant is corrected. The horizontal and vertical ratios are adjusted to eliminate unwanted high frequency components.

Carlos Merino-Gracia et al. [2] reports a wearable text recognition tool that employs Maximal Stable External Regions (MSERs) as the basis for real-time text detection. The proposed method exploits the hierarchical structure obtained from MSERs to yield more steady regions compared to the adaptive threshold method. It outperforms other published approaches computationally while maintaining equivalent text detection performance on the ICDAR dataset.

N. Venkata Rao et al. [3] presents a Neural Network (NN) based technique for optical character recognition and handwritten character recognition. This paper also discusses other techniques like matrix matching, fuzzy logic, feature extraction and structural analysis. The neural network model used here is the Multi Level Perception model (supervised network).

George Naggy [4] proposes a new algorithm for optical character recognition which takes advantage of the typographic uniformity of paragraphs or other layout components. The essence of adaptive OCR is to reduce multi-font classification by taking advantage of the normal occurrence of long strings of characters in the same typeface. Typeface homogeneity also helps increase the throughput.

Kahan S et al. [5] proposes a system that recognizes printed text of various fonts and sizes for the Roman Alphabet. Shape extraction is performed directly on the graph of the run-length encoding of a binary image using shape clustering approach. This is then fed into a Bayesian Classifier. Finally, layout and linguistic context are applied.

3. EMOTION DETECTION FROM TEXT

The image is converted to text through OCR (Optical Character Recognition) and is analyzed to determine emotions. The proposed work makes use of a database which has all the words which are allowed to use on the sentence. A word could be associated with more than one emotion. Hence, the emotional audio for a word is stored in a multi-media database which could be recovered whenever required. Emotion is analyzed by identifying the words in the text and identifying the sentence pattern and strength of the words. Once this is done, the words are compared with the words in the multi-media database, and the matching emotional audio for the given text is returned. Then all these audio files are combined to produce speech. This helps to incorporate emotions in the speech output, and greatly enhances the experience.

4. WORKING PRINCIPLE

An image of printed text is captured through the Raspberry Pi camera module. Once the image is captured, it is processed in order to enhance image quality. This process includes skew correction, image sharpening, thresholding the image and segmentation of the image to meaningful boundaries. The processed image is then fed to the Optical Character Recognition Engine which converts it to text.

To retrieve emotions from the text, each word is matched with the corresponding word in the multimedia database. Since a word can have more than one emotion, the entire sentence is first analyzed to determine the overall emotion. Once this is done, we retrieve the corresponding emotions for each word from the multi-media database. The audio files for each word is then concatenated as speech output for the end user. This helps in better understanding of the text.

5. Components

The first step in to process the images to enhance their quality. The following modules are developed:

Rescaling Module

We found that re-scaling images greatly improved the optical character recognition capabilities. We noticed that Tesseract works best with images with dots per resolution (DPI) of 300. Thus, all images are re-scaled to this value.

De-skew Module
Sometimes our image can contain a rotated block of text at unknown angle. This is when we need to correct our skew angles. This is done by first detecting the block of text in the image, and computing the angle of the rotated text. Then, we rotate the image to correct the skew. There are several techniques that are used for skew correction.

(a) Projection profile method
(b) Hough transform
(c) Topline method
(d) Scanline method

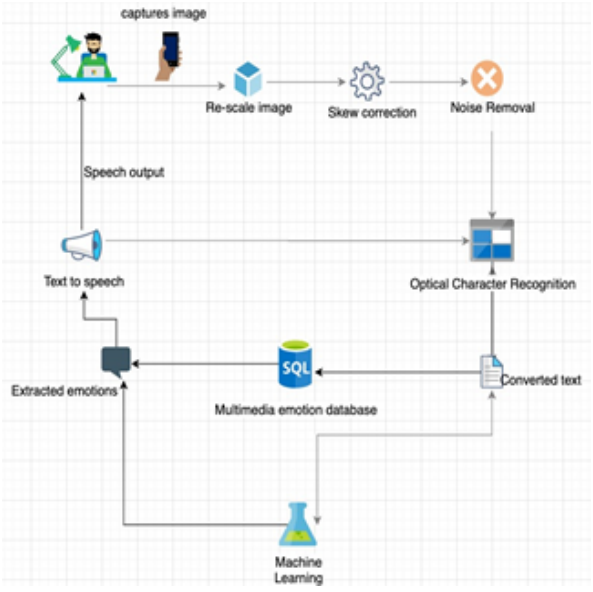
In the first method, we convert the image to black and white. Now, the image is projected vertically to get a histogram of pixels. Then, the image is rotated at various angles and the process is repeated. Wherever we find the maximum difference between peaks, that will be the best angle. The process can be repeated until we find a suitable level of accuracy.

Noise Removal Module
De-noising continues to be one of the most challenging aspects of image processing especially when images are acquired under poor conditions. There are two types of noise that we encounter in this project – Salt and Pepper Noise & Gaussian Noise. The former may contain light and dark dots that are dispersed throughout the image. The latter is statistical noise having probability distribution equal to the Gaussian function. Primary sources of Gaussian Noise during the process of image acquisition are poor illumination and(or) very high/very low temperatures. We apply the Non Local Means algorithm for image de- noising. This method adds white noise to the existing image, and computes the mean of all the pixels in the image. This is weighted by their similarity to the target pixels.

Segmentation Module
Image segmentation involves converting an image into a collection of region of pixels that are represented by a mask or a labeled image. The advantage of dividing the image into segments is that it allows us to process only the important segments instead of processing the whole region. The isolated blocks of characters are segmented and are labelled for identity.

Optical Character Recognition Module
Optical Character Recognition is a technology that enables the conversion of handwritten, typed, scanned text or text inside images to machine-readable text. Some uses of OCR include creating automated workflows by digitizing PDF documents, eliminating manual data entry, creating secure access to sensitive information by digitizing ID cards and credit cards, etc.

5. ARCHITECTURE DIAGRAM



6. CONCLUSION AND FUTURE SCOPE
This paper proposes a novel implementation of a smart OCR based reader for the visually impaired. This project has been implemented on an embedded platform, and uses technologies such as optical character recognition, image processing and speech processing engines. The objective of this project was to provide an affordable hand-held book reader device for the under-represented sections of the society.

The future scope of this project is promising. Optical Character Recognition has been applied to a number of applications such as invoice imaging, banking, optical music recognition and automatic number recognition. The current system works for both English and Hindi scripts, but we plan to extend it to other Indian regional languages.

7. REFERENCES
[1] Nagaraja L. et al, Vision Based Text Recognition using Raspberry Pi, International Journal of Computer Applications, NCPSIA 2015
[2] Merino-Gracia C., Lenc K., Mirmehdi M. (2012) A Head-Mounted Device for Recognizing Text in Natural Scenes. In: Iwamura M., Shafait F. (eds) Camera-Based Document Analysis and Recognition. CBDAR 2011. Lecture Notes in Computer Science, vol 7139. Springer, Berlin, Heidelberg
[3] N. Venkata Rao et al, Optical Character Technique Recognition Algorithms, Journal of Theoretical and Applied Information Technology, Vol.83, No. 2, 20th January 2016
[4] George Naggy, At the Frontiers of OCR, Invited Paper,pp 1093-1100
[5] Richard G. Casey et al, A survey of methods and strategies in Character Segmentation, IEEE Computer Society, July 1996