ABSTRACT

The idea presented on this paper is proposed for an application of the ocr. The main purpose of this project is to recognize the printed text character from any natural image which contains text and do spelling corrections and generate text file and convert it into speech signal. Along with this the application should also provide special features like spell correction and voice selection options dictionary and image to text output saving capabilities. In this project Image text recognition and converting to speech involves several steps. First we have to preprocess the input image like converting the image into grayscale image and adding some threshold values to make text recognition easier. Then the preprocessed image is given as input to an OCR module to extract text from the image. The recognized text is given to name detection algorithm, here the native words are collected and text is thereby given to the spell correction algorithm ,then spell checking will be done after this algorithm, and then the native words are replaced with previous collected words and the corrected words will be given as input to the TTS module to get the final output. Based on these features the author predicts that the application gives more accuracy and the accuracy is 96 .21%

Key words: OCR, TTS, Name Detection, Symspell Correction.

LIST OF FIGURES

CONTENTS

1. INTRODUCTION

Digital Image Processing means processing digital images by means of a digital computer. We can also say that it is a use of computer algorithms, in order to get enhanced images to extract some useful information. Digital image processing deals with manipulation of digital images through a digital computer. It is a subfield of signals and systems but focuses particularly on images. DIP focuses on developing a computer system that is able to perform processing on an image. The input of that system is a digital image and the system processes that image using efficient algorithms, and gives an image as an output.

Image processing mainly include the following steps:

1. Importing the image via image acquisition tools.
2. Analyzing and manipulating the image.
3. Output in which the result can be altered or a report which is based on analyzing that image.

Advantages:

1. Remove noises.
2. Correct image density and contrast.
3. Helps to easily store and retrieve in computers.
4. Images can be made available in any desired formats like black and white, negative images.

Disadvantages:

1. Initial cost is high depending upon the system used.
2. Once the system is damaged the image will be lost.

1.1 INTRODUCTION

There are close to 39 million blind people and around 285 million visually impaired people globally[1]. There is a huge impact on the lives of visually disabled people due to this. Although there have been several attempts made for helping visually disabled to see objects via other alternating means such as sound and touch, the

development of text reading device is still at a nascent stage.Therefore, we need a cost effective and truly efficient system that will be able to automatically identify and recite text aloud to visually challenged user base[2].The scope of this application can be widened to specially - abled people with learning disabilities, young children and various other segments of society.

1.2 MOTIVATION OF THE WORK :

It is designed for people with mild or moderate visual impairment by providing the capability to listen to the text. It can also act as a learning aid for people suffering from dyslexia or other learning disabilities that involve difficulty in reading or interpreting words and letters. We wish to enable these people to be independent and self-reliant as they will no longer need assistance to understand printed text. Such people will always have access to information hence they will never feel at a disadvantage. The impact of the development and introduction of our system into the technological world will be a revolutionary beneficial to modern civilization.

1.3 PROBLEM STATEMENT :

Visual impairment people use the braille system for writing and reading purposes. The visually impaired person feels the arrangement of the raised dots which conveys the information and is very difficult so keeping these things in mind we have designed our system in such a way that reading any book for blind people becomes easier. The braille system is very difficult and time consuming so if we can convert a text to audio then it would be much faster and easier.

2. LITERATURE SURVEY

Literature survey is the most important step in the software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the

programmers start building the tool the programmers need a lot of external support. This support can be obtained from senior programmers, from books or from websites. Before building the system the above considerations are taken into account for developing the proposed system.

2.1 TEXT DETECTION

S.K.Singla,R.K.Yadav[3] proposed that Knowledge extraction by just listening to sounds is a distinctive property. Speech signals are more effective means of communication than text because blind and visually impaired persons can also respond to sounds. This paper aims to develop a cost effective, and user friendly optical character recognition (OCR) based speech synthesis system. TheOCR based speech synthesis system has been developed using Laboratory virtual instruments engineering workbench[3]. The merits in this paper is that Text detection from general background or video image and Bangla optical character recognition (OCR) system[3].The demerits in this paper is that some words will be spelled wrong as OCR output is getting with some spelling mistakes[3].

J.N. Balaramakrishna, J. Geetha proposed that the main aim of this project is to convert the text in the textual image into the speech efficiently. For this project we are using the Raspberry Pi 3 processor, which supports OpenCV libraries and some image processing algorithms[4]. The merits in this paper is that it is an efficient text to speech conversion technique by using the Raspberry Pi 3 processor. When the text image was captured by the camera, the synthesizer used to separate the text from the image and then the Optical Character Recognition algorithm was implemented to recognize the characters in the text and then the Raspberry Pi 3 was the responsible to convert that text into speech by using the OpenCV libraries[4]. The demerits in this paper is that some words will be spelled wrong as OCR output is getting with some spelling mistakes[4].

Nidhi Kalidas Sawan proposed that this paper text to speech conversion be done for Marathi printed text. To obtain the required output the two techniques are implemented that are Optical Character Recognition (OCR) and Text to Speech (TTS) system. OCR is utilized to convert the text from an image into editable text which is done using the multiclass Support Vector Machine (SVM) and Text to Speech system gives the audio output[5]. The merits in this paper is that the correct recognition of the segmented characters feature extraction is very important. Here for character recognition histogram of oriented gradients (HOG) feature is used[3].The demerits in this paper are some words will be spelled wrong as OCR output is getting with some spelling mistakes[5].

2.2 EXISTING SYSTEM

In the existing system the text recognized by ocr is directly converted to speech. The drawback of this system is, there is no post-processing step to check and correct the misspelled or wrongly recognized words[6].

3. METHODOLOGY

3.1 START

This module is created to automate the execution of individual programs(modules) in cascading manner.Since each module takes input from the previous module and generates output which is input to the next module, all modules need to run in cascading manner as shown in fig. 2.

This module runs the project

- import os [7]
- #run preprocessor
- os.system('python prepro.py')

Let us take an example of a 5 letter word that has about 3 million possible spelling errors for a maximum edit distance of 3, but SymSpell generates only 25 deletes to cover all of them, both at pre-calculation and at lookup time [21].

Works:

The string is broken into individual tokens thus getting suggestions from the symmetric delete spell correction algorithm for each token and finally merging all individual tokens to a corrected sentence.

This process is performed in three steps:

1)Text preprocessing:

Firstly, to normalize a text, it was important to define a valid word (token) within a computer-readable text or speech collection (corpus, plural corpora) [22]. We cannot decide that the tokens can count (or not) punctuation marks, it depends on the final application.

Among the text preprocessing methods, we can highlight:

- tokenization means it is the task of breaking a sentence into words or characters.
- filtering, means it is the task to remove words of little information (stopwords).
- lemmatization, means it is the task of grouping several inflected forms of the same word so that they can be analyzed as a single token and
- stemming, means it is the task of obtaining the stem of derived words, also as a way of analyzing it as a single token[22].

2)Error Detection:

After tokenizing the sentence into individual words to find whether the word is misspelled or not is done by dictionary lookup where it compares the token with the dictionary. If there is no match for the token then error correction is done. Else, the token remains unchanged.

3)Error Correction:

After detecting a spelling error in a text, the correction process consists of generating correction candidate keys and comparing them with token or misspelled word. Symmetric delete spelling correction algorithm generates candidate keys using delete only upto 2 edit distances. The types of checks are performed in

- delete (dictionary term) = incorrect term
- dictionary term = delete (incorrect term)
- delete (dictionary term) = delete (incorrect term)

after checking the dictionary lookup by candidate keys the dictionary word is replaced in the error word.

Sample:

- Input text : “it wll peraps smoe distancee beyOnd”
- Output text: “it will perhaps some distance beyond”

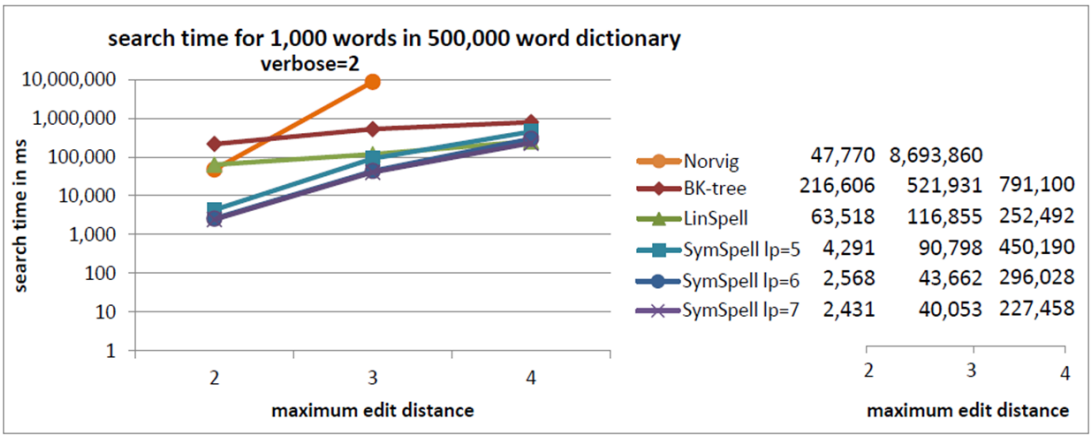

Fig. 3.5.1 comparison for search time

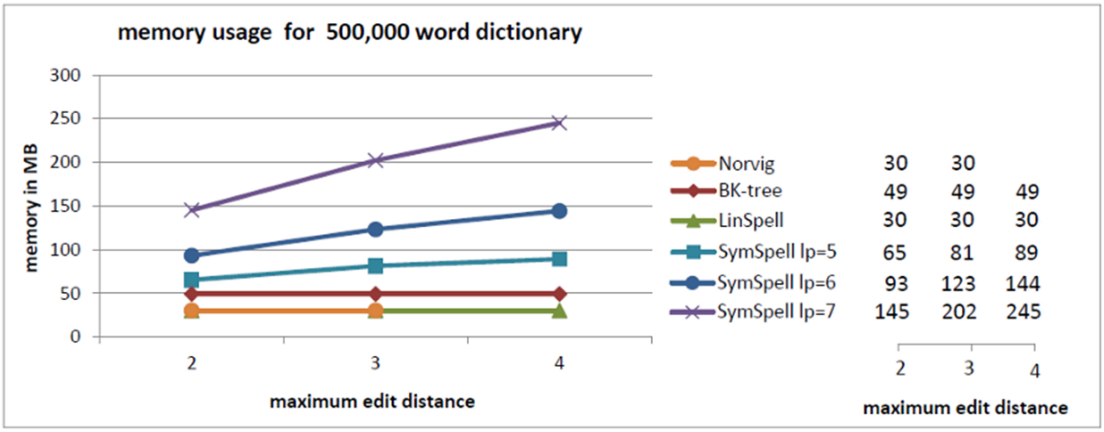

Fig. 3.5.2 comparison for memory usage

3.6 TEXT - TO - SPEECH (TTS) :

TTS can enable the reading of monitor information for the dim-sighted person, or may simply be wont to augment the reading of a text message. The present (TTS) applications consist of voice enabled e-mail and the spoken prompt's in the voice response systems respectively[23].

Formant speech synthesis method:

A formant speech synthesizer could also be a source-filter model , which models the glottal pulse train and thus the filter models the formant resonances of the vocal tract. Another quite formant-synthesis method, developed specifically for singing-voice synthesis is understood because of the FOF method.

Within speech science and phonetics, a formant could even be a broad spectral maximum that results from an acoustic resonance of the human vocal tract. In acoustics, a formant is typically defined as a broad peak, or the local maximum, within the spectrum..

Speech synthesis could even be a simple quite the output where the pc or other machine read words to you aloud during a real or simulated voice played through a loudspeaker; the technology is usually called text-to-speech (TTS).All which may