Implementation of Text to Speech Conversion

Department of Computer Science, Graphic Era Hill University

Abstract- **is a computer based system that can be able to read any text aloud is known as Text-To-Speech (TTS) conversion, whether it was directly introduced in the computer by an operator or scanned and submitted to an Optical Character Recognition (OCR) system. The main aims of this paper are to study on Optical Character Recognition with speech synthesis technology and to develop a cost effective user friendly image to speech conversion system using MATLAB. While in text to speech, there are many systems which convert normal language text in to speech. In this work, the OCR system is implemented for the recognition of capital English character A to Z and number 0 to 9. Each character is recognized at once. The recognized character is saved as text in notepad file. In this work a text-to-speech conversion system that can get the text through image and directly input in the computer then speech through that text using MATLAB.**

INTRODUCTION

Speech synthesis is the artificial result of human talk. A computer plan second hand for this purpose is called a talk synthesizer, and maybe implemented in spreadsheet or fittings [2]. A textto-talk (TTS) arrangement converts normal accent paragraph into speech; different wholes show symbolic semantic likenesses like phonetic transcriptions into talk. Manual-to-speech (TTS) practice revolutionizes linguistic news stocked as data or theme into talk. It is widely used in visual and audio entertainment transmitted via radio waves learning devices for blind nation immediately a days [6]. In the last few age nevertheless, the use of document-to-speech change electronics has grown far further the incapacitated community to enhance a important adjunct to the expeditiously increasing use of digital voice depository for telephone message system and voice response schemes. Further developments in speech synthesis

The Speech Application Programming Interface or SAPI is an API developed by Microsoft to allow the use of speech recognition and speech synthesis within Windows applications.

PROPOSED ALGORITHM

In this work, there are two main parts:

- Optical Character Recognition System for Paper Text

- Text to Speech Conversion

Optical character recognition system

OCR stands for Optical Character Recognition. As the name suggests OCR is used to detect the normal human readable language which may be present in the form of textual matter present in image or any documents or pdf files and convert it into editable formats.

In this part, there are three portions as described in the follow:

- Creating template file
- Neural Network creation
- Recognition of characters

2.1.1. Template file creation. Letter A to Z and number 0 to 9 images are collected. Each image is

changed into 5 x 7 character representation in single vector by using step 1 to 5 as described in the character recognition section. These data are saved as data file for training in neural network.

2.1.3. Character recognition. Figure 1 shows the flowchart of OCR system.

electronics for various accents have engaged place.

Submitted to: Chandradeep Bhatt by Rohan Kukreti

```
        ┌──────────┐
        │  Start   │
        └────┬─────┘
             ▼
 ┌──────────────────────────┐
 │ Image Acquiring and Reading │
 └────────────┬─────────────┘
              ▼
 ┌──────────────────────────┐
 │    RGB to Gay Image      │
 └────────────┬─────────────┘
              ▼
 ┌──────────────────────────┐
 │   Gray to Binary Image   │
 └────────────┬─────────────┘
              ▼
 ┌──────────────────────────┐
 │      Sementation         │
 └────────────┬─────────────┘
              ▼
 ┌──────────────────────────┐
 │   Feature Extraction     │
 └────────────┬─────────────┘
              ▼
┌───────────┐  ┌──────────────┐
│ Templates │◄─►│ Classification│
│trained in │  └──────┬───────┘
│    NN     │         ▼
└───────────┘  ┌──────────────┐
               │ Convert E-Text│
               └──────┬───────┘
                      ▼
               ┌──────────────┐
               │ Open text.txt │
               │as file for write│
               └──────┬───────┘
                      ▼
               ┌──────────────┐
               │Write in the text file│
               └──────┬───────┘
                      ▼
               ┌──────────────┐
               │     End      │
               └──────────────┘
```
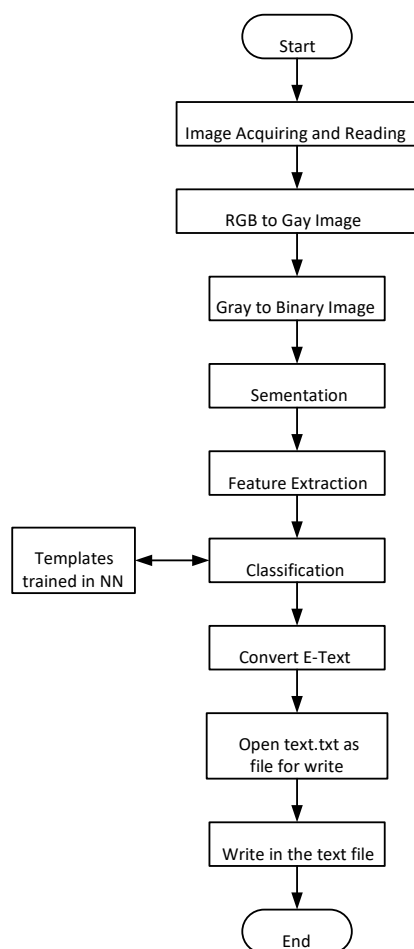
Figure 1. Flowchart of OCR system

The following steps are implemented for character recognition.

- Firstly acquire the character image and the image was read.

- Second step is preprocessing step. In this step firstly the image is converted into gray scale. Then this gray image is converted into black and white image (binary image). Firstly the threshold is counted in gray image then according to that threshold it is converted into black and white image.
- Find the boundary of the character image.

Crop the image to the edge.

- Character is extracted and resized in this step. Letters are resized according to templates size.
- The resized binary image is changed into 5 x 7 character representation in single vector.
- Load templates that it can be matched the letters with the templates.
- Open the text.txt as file for write.
- Write in the text file and concatenate the letters.

Feature extraction and categorization are the heart of Scanning in of documents. The figure countenance is plan to a higher level by eliciting distinguished traits and patterns of the figure in the feature extraction state. The classifier is therefore prepared accompanying the extracted visage for categorization task. The categorization stage labels each input integrity countenance by taking everything in mind the discovered features. As Classifiers, Motif Corresponding and Affecting animate nerve organs Networks are second hand.

Text to speech conversion

The character image is converted into text and then text into speech. The algorithm is followed.

- Firstly check the condition that if Win 32 SAPI is available in the computer or not. If it is not available then error will be generated and Win 32 SAPI library should be loaded in the computer.
- Gets the voice object from Win 32 SAPI.

- Compares the input string with Win 32 SAPI string.
- Extracts voice by firstly select the voice which are available in library.
- Choose the pace of voice.
- Initializes the wave player for convert the text into speech.
- Finally get the speech for given image. Text to speech conversion for the e-text input that directly typed in computer is also executed by the above steps.

METHODOLOGY USED :

Our goal is to convert a given text image into a string of text, saving it to a file and to hear what is written in the image through audio.

For this, we need to import some Libraries

Pytesseract(Python-tesseract) : It is an optical character recognition (OCR) tool for python sponsored by google.
1.pyttsx3 : It is an offline cross-platform Text-to-Speech library
2.Python Imaging Library (PIL) : It adds image processing capabilities to your Python interpreter
3.Googletrans : It is a free python library that implements the Google Translate API.

SIMULATION RESULTS

In this work, the OCR system is implemented for the recognition of capital English character A to Z and number 0 to 9. Each character is recognized at one time. The recognized character is saved as text with notepad file. There are two portions in program; in the first portion it gives the text output according to input image , then it convert that text into the speech. In the second portion, the e-text is directly input in computer, then it is converted into speech. Firstly the input image of time new romance, font size

12, bold type characters is taken and then it is converted into text. As shown in Figure 2, character "A" is cropped from the image and features are extracted. After that it is converted to text, saved in notepad file and speech simultaneously. Similarly, the test results for character "T" is also illustrated in Figure 3. The recognized character can be displayed in the command widow and can be save in notepad file as shown in Figure 4.
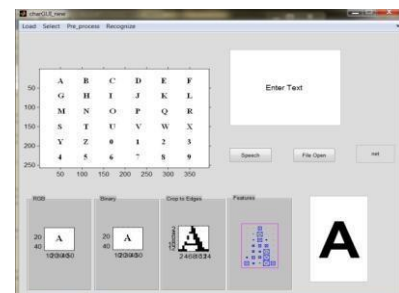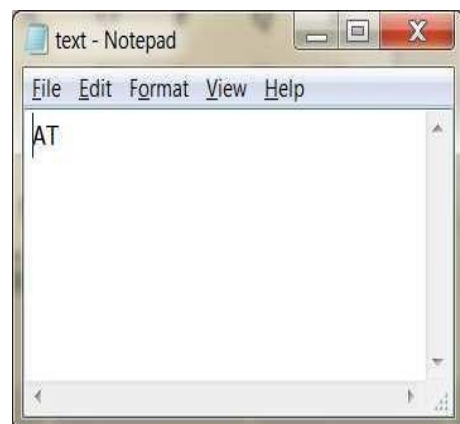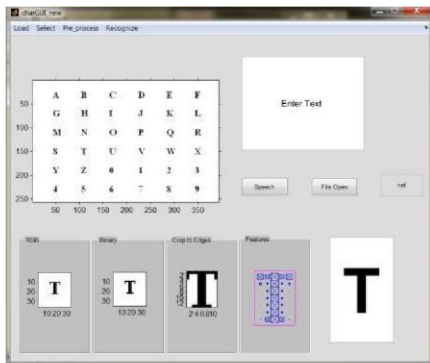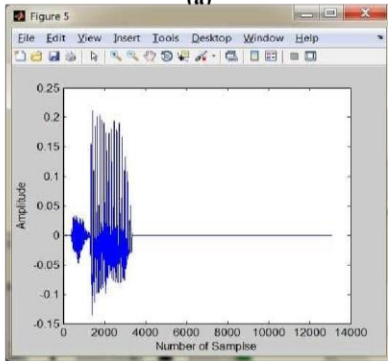

(a)
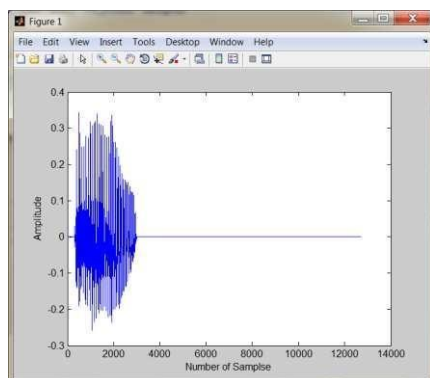
(b)

Figure 2. (a) Character "A" converted into text (b) "A" sound wave

(b)

Submitted to: Chandradeep Bhatt by Rohan Kukreti

(a)

(b)

Figure 3. (a) Character "T" converted into text (b) "T" sound wave
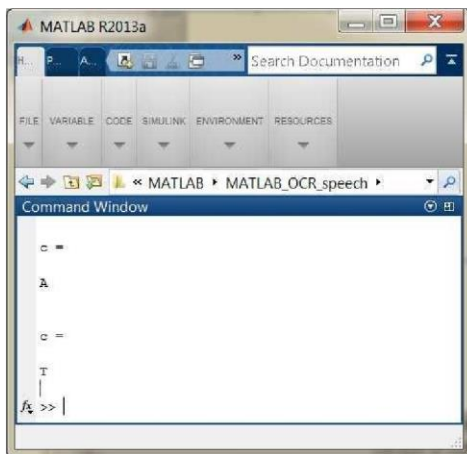
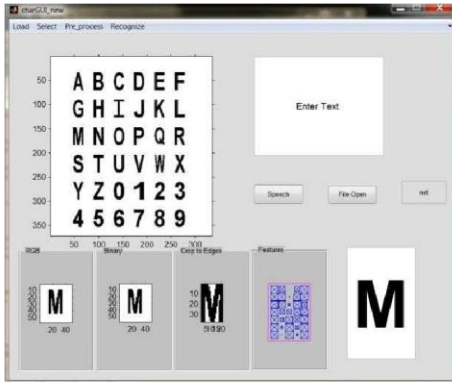The mathematical numbers are also successfully converted into text and then speech which is shown in Figure 5.
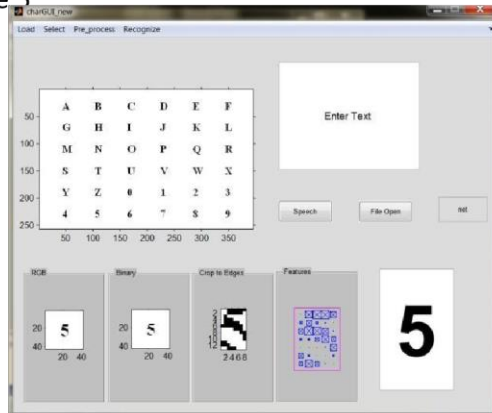
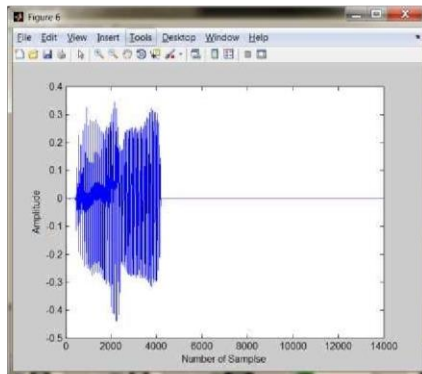(a)
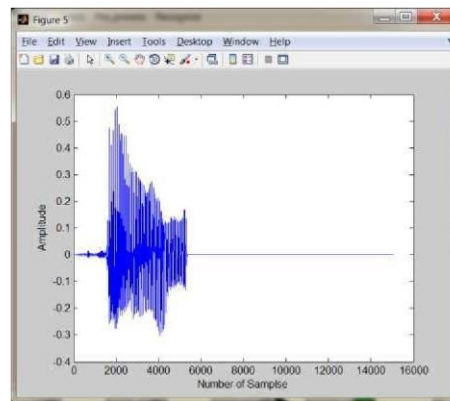
(b)

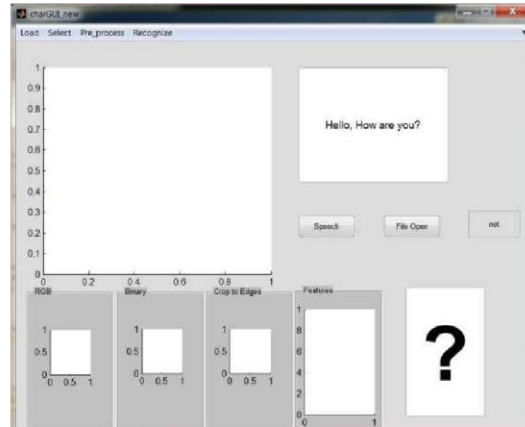Figure 5. (a) Number "5" converted into text (b) Number "5" sound wave

Submitted to: Chandradeep Bhatt by Rohan Kukreti
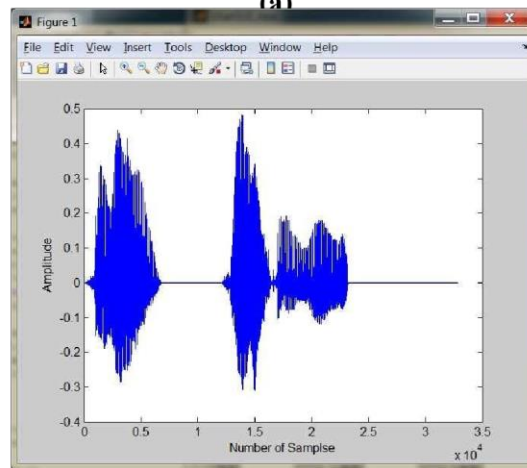
(a)

(b)

Figure 6. (a) Character "M" converted into text (b) Character "M" sound wave

(a)

As illustrated in Figure 8, the e-text that directly input in computer by typing from keyboard, then it is also converted into speech successfully.
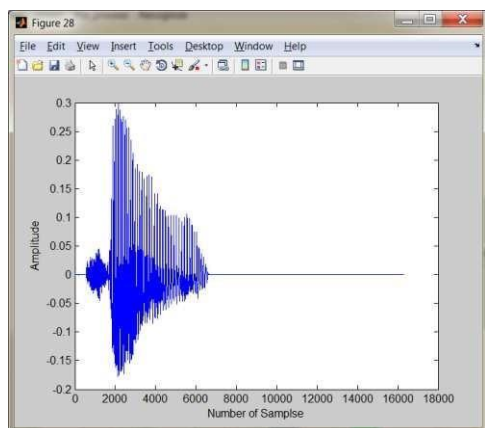

(a)

(b)

Figure 8. (a) E- text Input (b)Sound Wave "Hello, How are you?"

4. CONCLUSION

Another type of font character is taken and again it is converted into text and then speech successfully as shown in Figure 6 and 7.

Submitted to: Chandradeep Bhatt by Rohan Kukreti

(a)

(b)

Figure 7. (a) Number "2" converted into text (b) Number "2" sound wave

also converted successfully. By this approach text from a word document, Web page or e-Book can be read and can generate synthesized speech through a computer's speakers. For image to text conversion, firstly image is converted into gray image. Gray image is converted into binary image by thresholding and then it is converted into text by MATLAB. Microsoft Win 32 SAPI library has been used to build speech enabled applications, which retrieve the voice and audio output information available for computer. In this work, one character can be converted into text at once. As a further extension, OCR system can be developed for converting words or sentences image into text.

Nowadays, there is increasing demand of text information extraction from image. So, many extracting techniques for retrieving relevant information have been developed. Moreover, extracting text from the color image takes time that leads to user dissatisfaction. In this paper we have proposed a method to extract the text from image which extracts text more accurately. Using our method it is possible to extract information within short time. Although, our connected component based approach for text extraction from color image method has several features than existing method but it becomes less effective when the text is too small and if the text region is not clearly visible or the color of the text is not visible clearly. In future, this work can be extended to detect the text from video or real time analysis and can be automatically

documented in Word Pad or any other editable format for further use

REFERENCES

In this work, image into text and then that text into speech is converted by MATLAB. E-text into speech is

1. Hertz, S., "English text to speech conversion with delta," Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86. , vol.11, no., pp.2427-2430, Apr 1986

2. Ainsworth, W., "A system for converting English text into speech," Audio and Electroacoustics, IEEE Transactions on , vol.21, no.3, pp. 288-290, Jun 1973

3. S. Furui, "Speaker independent isolated word recognition using dynamic features of speech spectrum", IEEE Transactions on Acoustic, Speech, Signal Processing, Vol.34, issue 1, Feb 1986, pp. 5259.

4. Fushikida, Katsunobu; Mitome, Yukio; Inoue, Yuji, "A Text to Speech Synthesizer for the Personal Computer," Consumer Electronics, IEEE Transactions on , vol.CE-28, no.3, pp.250-256, Aug. 1982

5. Lynch, M.R.; Rayner, P.J., "Optical character recognition using a new connectionist model," Image Processing and its Applications, 1989., Third International Conference on , vol., no., pp.63-67, 18-20 Jul 1989

6. Leija, L.; Santiago, S.; Alvarado, C., "A system of text reading and translation to voice for blind persons ," Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE , vol.1, no., pp.405-406 vol.1, 31 Oct-3 Nov 1996

7. Tanprasert, C.; Koanantakool, T., "Thai OCR: a neural network application,"TENCON '96. Proceedings. 1996 IEEE TENCON. Digital Signal Processing Applications , vol.1, no., pp.90-95 vol.1, 26-29 Nov 1996

8. Breen, A.P., "The future role of text to speech synthesis in automated services," Advances in Interactive Voice Technologies for Telecommunication Services (Digest No: 1997/147), IEE Colloquium on , vol., no., pp.6/1-6/5, 12 Jun 1997

Submitted to: Chandradeep Bhatt by Rohan Kukreti