Journal of Applied Research and Technology
ISSN: 1665-6423
jart@aleph.cinstrum.unam.mx
Centro de Ciencias Aplicadas y Desarrollo Tecnológico
México

Singla, S. K.; Yadav, R. K.
Optical Character Recognition Based Speech Synthesis System Using LabVIEW
Journal of Applied Research and Technology, vol. 12, núm. 5, octubre, 2014, pp. 919-926
Centro de Ciencias Aplicadas y Desarrollo Tecnológico
Distrito Federal, México

Optical Character Recognition Based Speech Synthesis System Using LabVIEW

S. K. Singla*[1] and R.K.Yadav[2]

[1] Electrical and Instrumentation Engineering Department
Thapar University, Patiala,Punjab
*sunilksingla2001@gmail.com
[2] Electronics and Instrumentation Engineering Department
Galgotias College of Engineering and Technology
Greater Noida, U.P

ABSTRACT
Knowledge extraction by just listening to sounds is a distinctive property. Speech signal is 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. The OCR based speech synthesis system has been developed using Laboratory virtual instruments engineering workbench (LabVIEW) 7.1.

Keywords: Optical character recognition, Speech, Synthesis, Recognition, LabVIEW.

1. Introduction

Machine replication of human functions, like reading, is an ancient dream. However, over the last few decades, machine reading has grown from a dream to reality. Text is being present everywhere in our day to day life, either in the form of documents (newspapers, books, mails, magazines etc.) or in the form of natural scenes (signs, screen, schedules) which can be read by a normal person. Unfortunately, the blind and visually impaired persons are deprived from such information, because their vision troubles do not allow them to have access of this textual information which limits their mobility in unconstrained environments. The OCR based speech synthesis system will significantly improve the degree to which the visually impaired can interact with their environment as that of a sighted person [1].

This work is related to existing research in text detection from general background or video image [2-7], and Bangla optical character recognition (OCR) system [8-9]. Some researchers published their efforts on texture-based [10] text detection also. OCR based speech recognition system using LabVIEW utilizes a scanner to capture the images of printed or handwritten text, recognize that text and translate the recognized text as voice output message [11] using Microsoft Speech SDK (Text To Speech). This paper is organized as follows: section 2 of this paper deals with recognition of characters based on LabVIEW. The speech synthesis technique has been explained in section 3. Section 4 discusses experimental result and the paper concludes with section 5.

2. Optical Character Recognition

Optical character recognition (OCR) is the mechanical or electronic translation of images of hand-written or printed text into machine-editable text [12]. The OCR based system consists of following process steps:

a) Image Acquisition

b) Image Pre-processing (Binarization)

c) Image Segmentation

d) Matching and Recognition

2.1 Image Acquisition

The image has been captured using a digital HP scanner. The flap of the scanner had been kept open during the acquisition process in order to

obtain a uniform black background. The image had been acquired using the program developed in LabVIEW as shown in the Figure 1. The configuration of the Image has been done with the help of Imaq create subvi function of LabVIEW. The configuration of the image means selecting the image type and border size of the image as per the requirement. In this work 8 bit image with border size of 3 has been used.

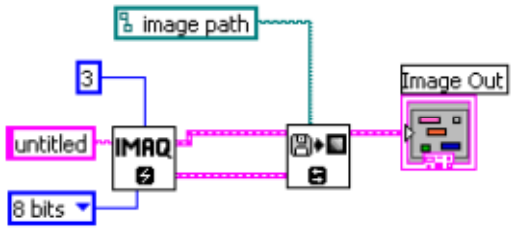

Figure 1. Image acquisition.

2.2 Image Pre-processing (Binarization)

Binarization is the process of converting a gray scale image (0 to 255 pixel values) into binary image (0 to1 pixel values) by using a threshold value. The pixels lighter than the threshold are turned to white and the remainder to black pixels. In this work, a global thresholding with a threshold value of 175 has been used to binarize the image i.e. the values of pixel which are from 175 to 255 has been converted to 1 while the of pixel which have gray scale value less than 175 have been converted to 0. The LabVIEW program of binarization has been shown in Figure 2.

2.3 Image Segmentation

The segmentation process consists of line segmentation, word segmentation and finally character segmentation.

2.3.1 Line segmentation

Line segmentation is the first step of the segmentation process. It takes the array of the image as an input and scans the image horizontally to find first ON pixel and remember that coordinate as y1. The system continues to scan the image horizontally and found lots of ON pixel since the characters would have started. When finally first OFF pixel has been detected the system remembers that coordinate as y2 and check the surrounding of the pixel to find out required number of OFF pixels. If this happens then the system clips the first line (fl) from input image between the coordinate y1 to y2. In this way, all the lines have been segmented & stored to be used for word and character segmentation. The LabVIEW program of Line segmentation is shown in Figure 3.

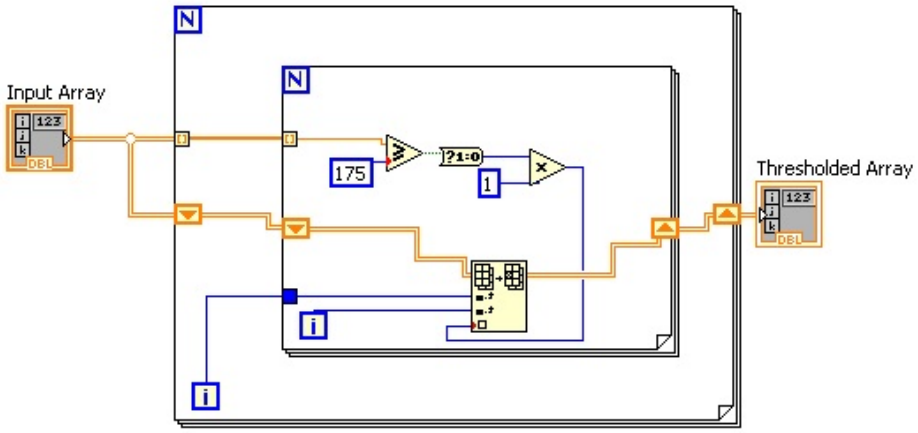

Figure 2. Binarization process.

2.3.2 Word Segmentation

In the word segmentation process the line segmented images have been vertically scanned to find first ON pixel. When this happen the system remember the coordinate of this point as x1. This is the starting coordinate for the word. The system continues the scanning process until fifteen (this is assumed word distance) successive OFF pixels have been obtained. The system records the first OFF pixel as x2. From x1 to x2 is the word. In this way all the words have been segmented and these segmented words have been used in next step for character recognition. The Figure 4 shows the LabVIEW programs of word segmentation.

2.3.3 Character Segmentation

Character segmentation has been performed by scanning the word segmented image vertically. This process is different from the word segmentation in following two ways:

i) Number of horizontal OFF pixels between the different characters are less in comparison to number of OFF pixels between the words

ii) Total number of characters and their order in the word has been determined so as to reproduce the word correctly during speech synthesis.

The LabVIEW program of character segmentation has been shown in Figure 5.

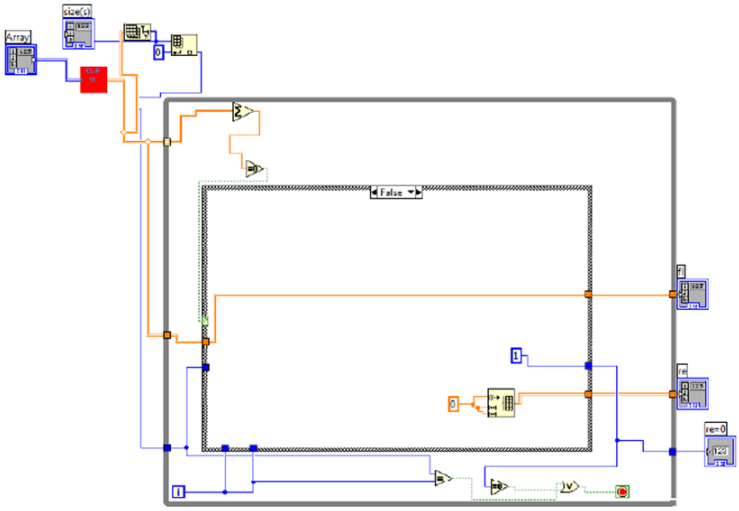

Figure 3. Line Segmentation.

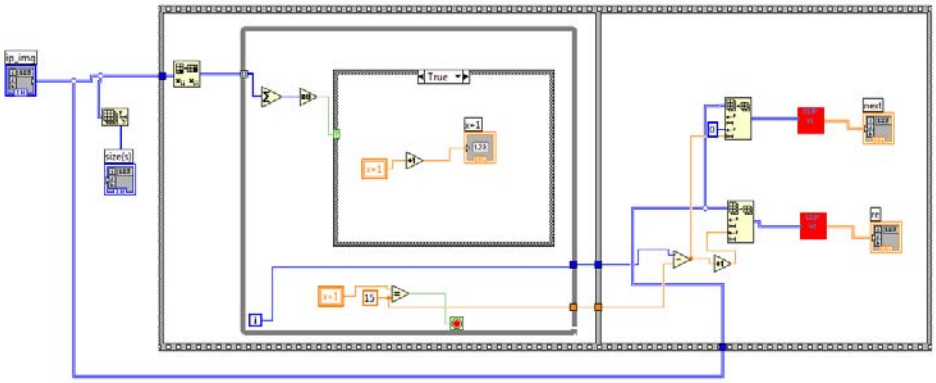

Figure 4. Word segmentation.

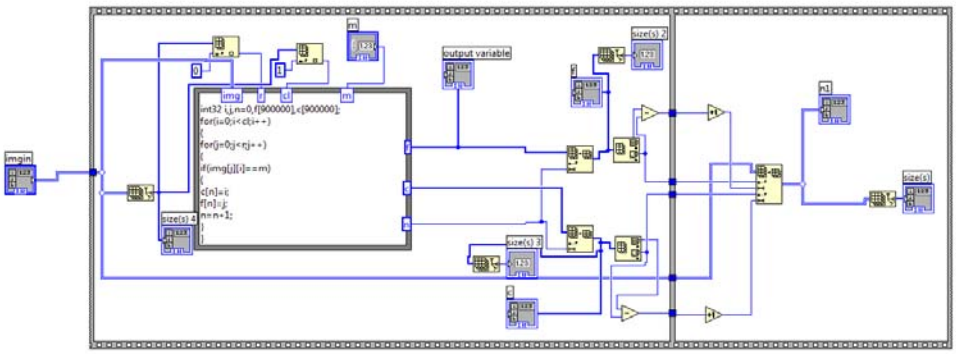

Figure 5. Character segmentation.

2.4 Matching and Recognition

In this process, correlation between stored templates and segmented character has been obtained by using correlation VI. The correlation VI determines the correlation between segmented character and stored templates of each character. The value of the highest correlation recognizes a particular character. In this way in order to recognize the character every segmented character has been compared with the pre defined data stored in the system. Since same font size has been used for recognition, a unique match for the each character has been obtained. Figure 6 shows the LabVIEW program of correlation between two images.

The flow chart of OCR system has been shown in Figure 7.

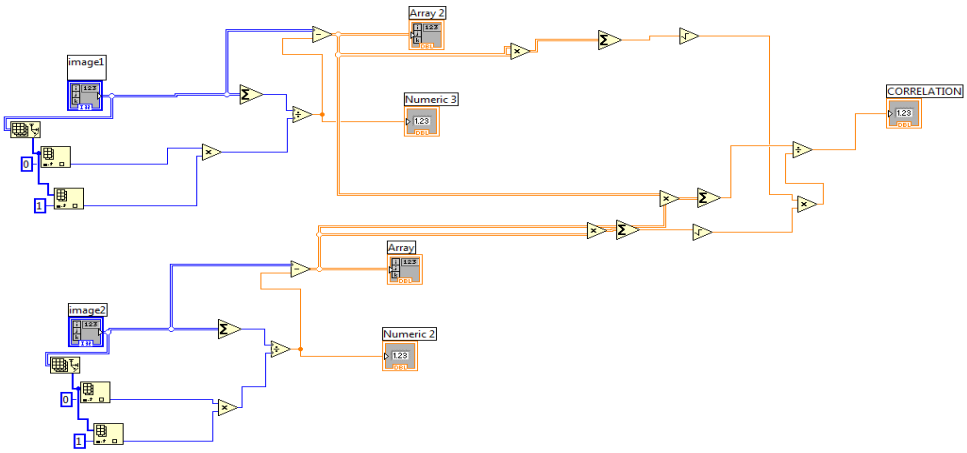

Figure 6. Correlation between two images.

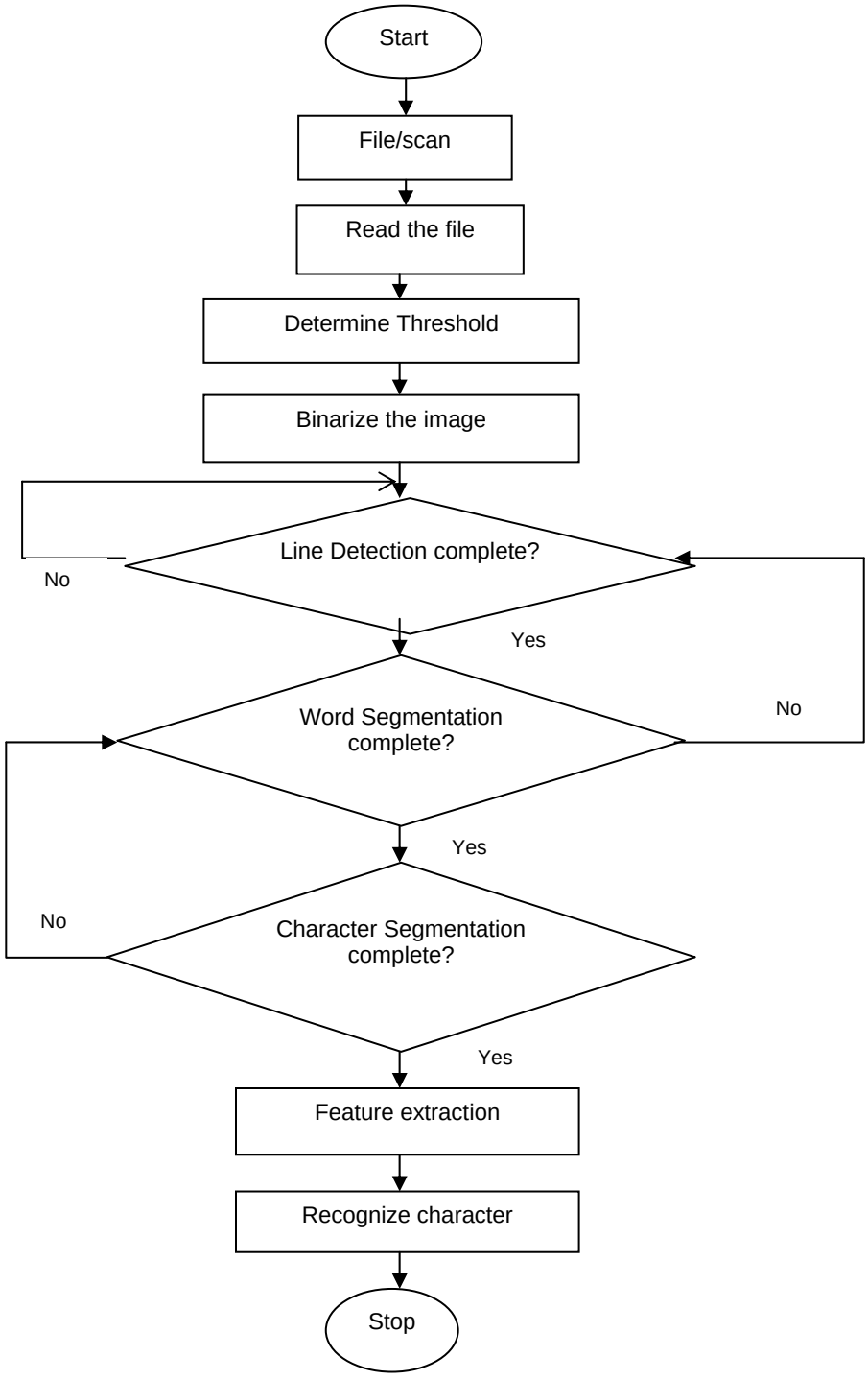

Figure 7. Flow chart OCR system.

3. Text to speech synthesis

In text to speech module text recognised by OCR system will be the inputs of speech synthesis system which is to be converted into speech in .wav file format and creates a wave file named output wav, which can be listen by using wave file player.

Two steps are involved in text to speech synthesis

i) Text to speech conversion

ii) Play speech in .wav file format

3.1 Text to speech conversion

In the text speech conversion input text is converted speech (in LabVIEW) by using automation open, invoke node and property node. LabVIEW program of Text to speech conversion is shown in Figure 8 and Flow chart text speech conversion has been shown below in Figure 9. Figure 10 shows the flow chart of playing the converted speech signal while Figure 11 shows its LabVIEW program.

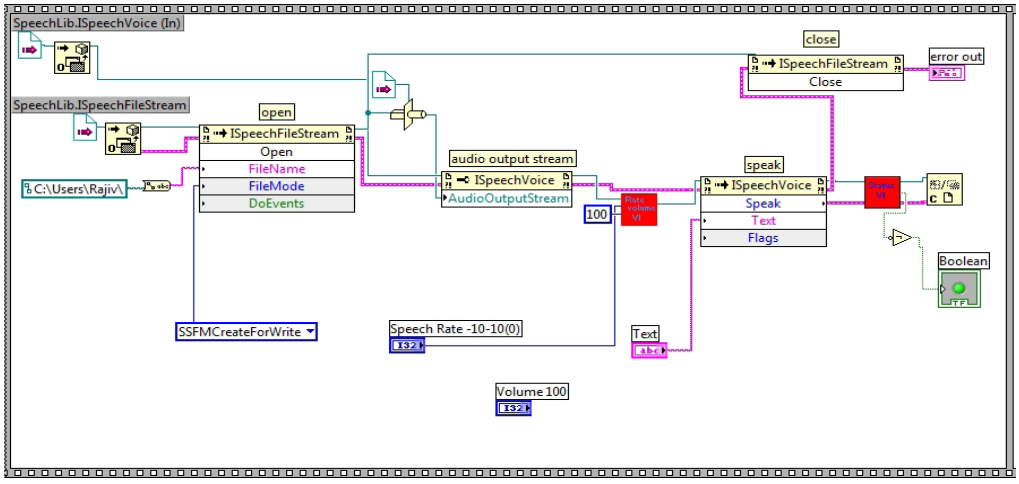

Figure 8. Text to speech conversions.

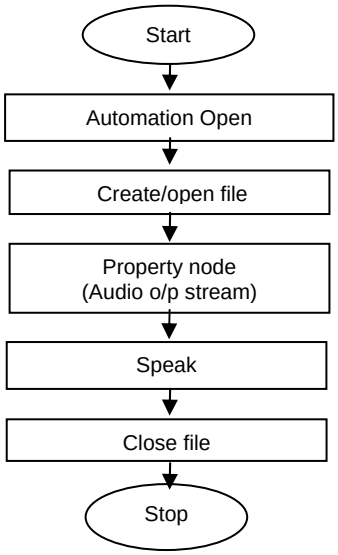

Figure 9. Flow chart text speech conversion.

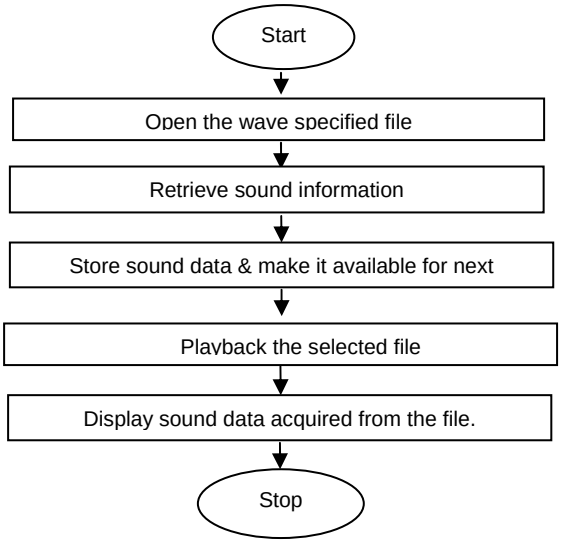

Figure 10. Flow chart of wave file player.vi.

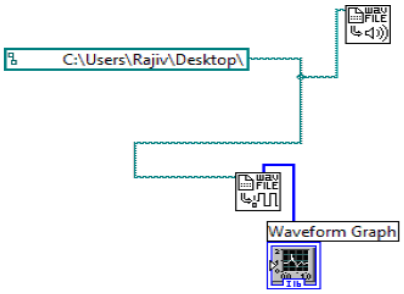

Figure 11. Wave file player.

4. Results and Discussion

Experiments have been performed to test the proposed system developed using LabVIEW 7.1 version. The developed OCR based speech synthesis system has two steps:

a. Optical Character Recognition

b. Speech Synthesis

4.1 Optical Character Recognition

Step 1. The scanner scans the printed text and the system reads the image using IMAQ ReadFile and display the image by using IMAQ WindDraw function of the LabVIEW as shown below in Figure 12.

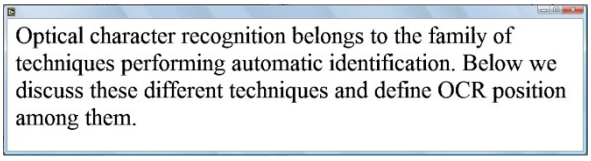

Figure 12. Read an image.

Step 2. In this step binarization of the image has been done with a threshold of 175 and the resulting image has been shown in Figure 13.

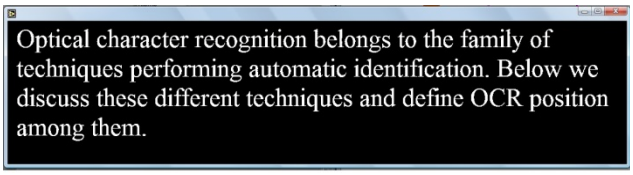

Figure 13. Image after thresholding and inverting.

Step 3. In this step line segmentation of thresholded image has been done. Figure 14 shows the result of line segmentation process.

Figure 14. Segmented line.

Step 4. In this step words have been segmented from the line. Figure 15 shows the result of word segmentation process.

Figure 15. Segmented Word.

Step 5. In this step character segmentation has been performed and all the chacter in word image window have been segmentated. The segmenatation of first three characters of word ”Optical” has been shown in Figure 16. After segmentation each character has been correlated with stored character templates, and recognition of printed text has been done by the system.

Figure 16. Segmented character image.

Step 6. Finally the output of OCR system is in text format which has been stored in a computer system. The result of recognized text can also be shown on Front panel as shown bellow in Figure 17.

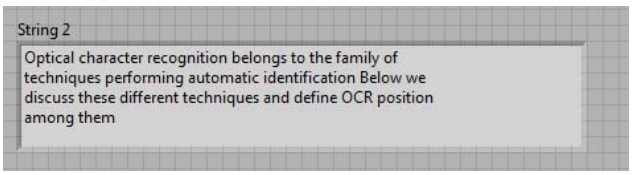

Figure 17. Final result of OCR system.

4.2 Speech Synthesis

A wave file output.wav is created containing text converted into speech which can listen using wave file player.

The waveform will vary according to the different text from OCR output in the text box and can be listened on the speaker. The wave form for above recognize text has been shown in Figure 18.

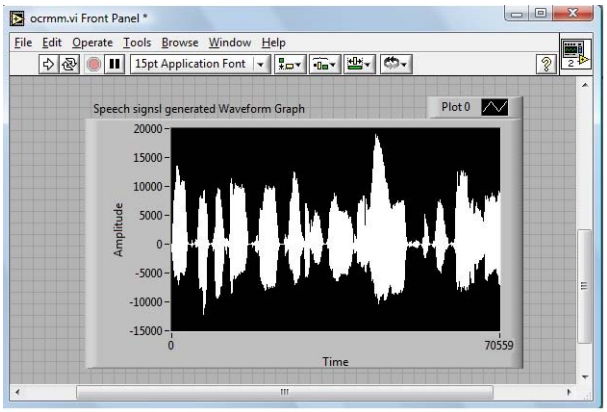

Figure 18. Output of file wave player.

5. Conclusion

In this paper, an OCR based speech synthesis system (which can be used as a good mode of communication between people) has been discussed. The system has been implemented on LabVIEW 7.1 platform. The developed system consists of OCR and speech synthesis. In OCR printed or written character documents have been scanned and image has been acquired by using IMAQ Vision for LabVIEW. The different characters have been recognized using segmentation and correlation based methods developed in LabVIEW. In second section recognized text has been converted into speech using Microsoft Speech Object Library (Version 5.1). The developed OCR based speech synthesis system is user friendly, cost effective and gives the result in the real time. Moreover, the program has the required flexibility to be modified easily if required.

References

[1] S. M. A. Haque et al., “Automatic detection and translation of bengali text on road sign for visually impaired,” Daffodil Int. Univ. J. of Sc. and Tech,vol. 2, 2007.

[2] N. Ezaki et al., “Text detection from natural scene images: Towards a system for visually impaired persons”, In Proceedings of the International Conference on Pattern Recognition, 2004, pp. 683-686

[3] T. Yamaguchi et al., “Digit classification on signboards for telephone number recognition”, In Proceedings of 7th International Conference on Document Analysis and Recognition,2003, pp. 359–363.

[4] K. Matsuo et al., “Extraction of character string from scene image by binarizing local target area,” Trans. of The Ins. of Elec. Eng. of Japan, vol. 122-C(2), pp. 232–241, 2002.

[5] T. Yamaguchi and M. Maruyama, “Character Extraction from Natural Scene Images by Hierarchical Classifiers,” In Proceedings of the International Conference on Pattern Recognition, 2004, pp. 687-690.

[6] J. Yang et al., “An automatic sign recognition and translation system”, Proceedings of the Workshop on Perceptive User Interfaces, 2001, pp. 1-8.

[7] H. Li et al., “Automatic Text Detection and Tracking in Digital Video,” IEEE Trans. on Im. Proc., vol. 9(1), pp. 147-156, 2000.

[8] B.B.Chaudhuri and U.Pal, “A complete printed bangla OCR system,” Pat. Rec., vol. 31, pp. 531-549, 1997.

[9] A.O.M. Asaduzzaman et al., “Printed bangla text recognition using artificial neural network with heuristic method,” Proceedings of International Conference on Computer and Information Technology, 2002, pp. 27-28

[10] T. Sato and T. Kanade, “Video OCR: Indexing digital news libraries by recognition of superimposed caption,” ICCV Workshop on Image and Video retrieval, 1998, pp. 1-10.

[11] T. Dutoit, "High quality text-to-speech synthesis: a comparison of four candidate algorithms," IEEE International Conference on Acoustics, Speech, and Signal Processing, 1994, pp. 19-22

[12] B.M. Sagar et al., “OCR for printed Kannada text to machine editable format using database approach,” WSEAS Trans. on Computers, vol. 7, pp. 766-769, 2008.