International Journal of Engineering & Technology
CENTRE OF PROFESSIONAL RESEARCH PUBLICATIONS
ISSN 2049-3444
International Journal of Engineering and Technology Volume 6 No.8, August, 2016

Image Acquisition and Text To Speech Conversion for Visually Impaired People

Priyanka Rathod , Suvarna Nandyal
Computer Science and Engineering, Poojya Doddappa Appa College of Engineering Kalaburgi, Karnataka, India

ABSTRACT

In today's world, many people are facing problems due to disability in sense organs such as low eye sight, hearing issues, and many other problems. There are nearly about 161 million visually impaired and 37 million blind people worldwide. Many times they are confused in a new environment because of communication and access to information. It becomes tedious for blind people to read and walk to know the shop around due to visually impaired. Hence objective of the proposed work is to detect, extract and recognize text from images and convert them into text to speech. The work presents an algorithm for implementation of Optical Character Recognition (OCR) to translate images with standard labels with standard font sizes, into electronically editable format and then to speech to assist visually impaired users. OCR can do this by applying pattern matching algorithm the recognized character are stored in editable format. Thus OCR makes the computer read input image text label discarding noise. Image detection and extraction is done using OCR algorithm, then the extracted text of the image is converted to actually text using cross correlation algorithm. Then finally text is converted to speech/audio through the mobile via remotely.

Keywords: *Binarization, Segmentation, Templates, Optical Character Recognition (OCR), Text-To-Speech.*

1. INTRODUCTION

Every human is gifted by the nature with multiple senses. Unfortunately number of people lost or deprived their visual sense. Blind people and the visually impaired people face a lot of adverse challenges in their routine. Human has disorders in sense organs then there should be modern smart technology to assist such human beings. Hence today's world should focus on development of the technologies to assist an disabled person.For example ,we have smart phones which acts as a personal assistant, we have technologies in the form of the hardware and software solutions which help us to solve many modern day problems and needs in the present society. There are only a limited technologies which offer guidance to the differently able people of the society. Choosing products with suitable choice and reading is a challenging task for the visually impaired people. They manage this task with the help of their family members or friends using plastic braille labels. So the aim and motivation of the work is to provide an application which can act as an assistance to the visually impaired people and help them to read the text from any product ,etc provided the text contained in the images should be machine readable that is in a standard format ,for example ,bar-code and Quick Response code.

2. RELATED WORK

In the Digital Image Processing many of the authors and scholars have developed different technologies and methods to help visually impaired people following literature survey has been carried out.

Mobile travel aid for the blind[1] proposed a system which comprises of two interlinked terminals:a mobile for blind user and remote assistant .Mobile has a setup of camera , headset and GPS receiver. Blind user and remote assistant are connected through GSM internet link. The assistant watches a video transmitted from the blind's camera. They both have voice communication through a packet switched internet connection. Image Segmentation for Text Extraction [2] used Discrete Wavelet Transform (DWT) for text extracting from complex images. Morphology based text extraction in images[3] have used an approach of morphological operation for text localization and support vector machine(SVM) for recognizing character, together with some pre-processing and post-processing steps. Classification of newspaper image blocks using texture analysis[4] Page segmentation and classification[5]are the techniques uses grouping and merging of the character elements and then recursively to words, and those words to text lines and then to paragraphs. Mobile camera based text detection and translation [6] describes the features of software modules which are developed in android smart phones. One of the module can be identified and match the scanned objects to a database of objects. The other two modules are capable of detecting colors and locate the direction which consists of maximum brightness regions in captured scenes.Text detection and recognition in natural scenes and consumer videos[7] proposed an end-to-end system for video text detection and recognition ,system consists of three steps: text localization ,text line aggregation and text line recognition. Here they have used MSER region as candidates and have applied a text/non-text SVM classifier over each candidates. Partial Least Square(PLS) technique is used for large set of features in classification and speeds up the

classification.Effective Text Localization in Natural Scene Images with MSER, Geometry-based Grouping and AdaBoost[8] proposed a novel and effective approach to accurately localize scene texts.Maximally stable extreme regions (MSER) are extracted as letter candidates.Then,after elimination of non-letter candidates by using geometric information, candidate regions are constructed by grouping similar letter candidates using disjoint set. Candidate region features based on horizontal and vertical variances, stroke width, color and geometry are extracted. An AdaBoost classifier is built from these features and text regions are identified. Portable camera-based assistive text and product label reading from hand-held objects for blind persons [9] proposed a reading framework with the help of camera based assistive text reading for helping blind people to read text labels from hand-held objects.A Gaussian based approach where first,the object is recognized ,then region of text is identified and performed different image processing operations.

The text is extracted form the object and isolated[11][12] by motion based object detection.then the region is found where the text is present based on edge properties . At the end the required text is extracted [13].Text Extraction from Video Using Conditional Random Fields[14] describes the conditional random field(CRF) based approach to detect the text lines from video frames. Proposed consists of text block extraction based on edges,CRF labeling for text regions and the text line aggregation, and support vector machine(SVM) prediction.

3. PROPOSED ALGORITHM

The proposed work deals with two parts.

1. Optical Character Recognition.

2. Text to Speech conversion.

1. Optical Character Recognition

The components of OCR system consists of

1.1. Character Image Scanning

In computing, a scanning is a device that optically scans images, printed text, typed text and converts it into a digital image.

1.2. Binarization

Binarization is the process of converting a gray scale image into binary image by thresholding .The binary document image allows the use of first binary arthimatic during processing, and also requires less space to store. Because of the complexity of the OCR operation, the input of the character recognition phase in the most methods is binary images

1.3. Segmentation

Segmentation of text is a process by which the text is partitioned into its coherent parts. The text image contains a number of text lines. Each line again contains a number of words. Each word may contain a number of characters. The segmentation scheme such as line segmentation, word segmentation and character segmentation are proposed where lines are segmented then word and finally characters. These are then put together to the effect of the recognition of individual characters. Segmentation is carried out by detecting the edge. The edges are identified by the pixel values. If the specific pixel value is like 110,we replace it by black pixel and otherwise we replace it by white pixel. By the help of binary pixels the segmentation is carried out.

1.4. Recognition.

Once we get the character by character segmentation we store the character image in a structure. The character has to be compared with the predefined character set. Preliminary data will be stored for all characters for an identified font and size. This data contains the following information. Character ASCII value, character name, character BMP image etc. The recognized character information will be compared with the predefined data which we have stored in the system. As we are using the same font and size for the recognition there will be exact one unique match for the character. This will identify us the name of the character[9]. If the size of the character varies it will be scaled to the known standard and then recognizing process will be done.

2. Text to Speech Synthesizer

A text to speech synthesizer is used to define text to speech as the automatic production of speech, through a grapheme-to-phoneme transcription [16] of the sentences to utter.
The functional block diagram is shown below:

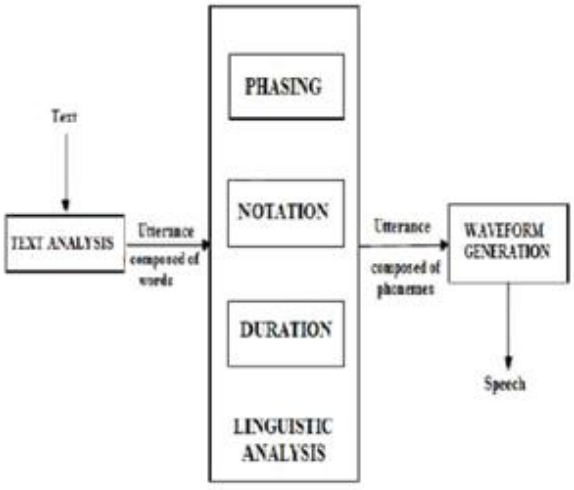

Fig 1: Proposed Method

The functional diagram of a very general Text-To-Speech synthesizer comprises of Natural Language Processing module (NLP) and a Digital Signal Processing module (DSP). NLP is capable of producing a phonetic transcription the text read,

together with the desired intonation and rhythm (prosody).DSP module which transforms the symbolic information it receives into speech.

The below figure shows the Natural Language Processing component and Digital Signal Processing component ,text-to-speech synthesizer.

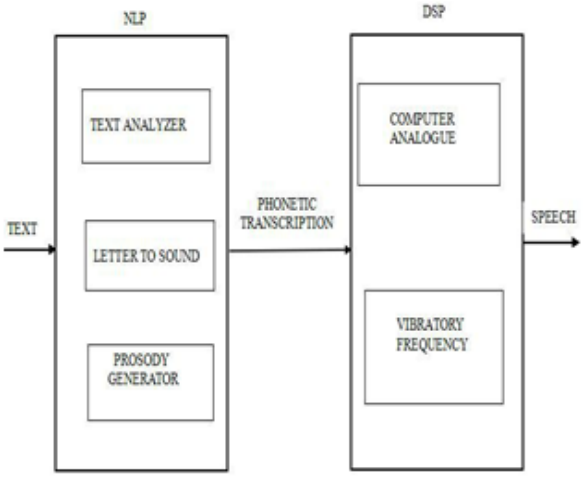

Fig 2: Text-To-Speech Synthesizer

2.1. The Natural Language Processing(NLP) component

The NLP module contains a series of text input and produces a phonetic transcription together with the desired intonation and prosody(rhythm) that is ready to pass on the DSP module. The NLP module is composed of three major components: text analyzer, letter-to-sound (LTS) and prosody generator. Text analysis is a language dependent component in TTS system. It is invoked to analyze the input text. This process is divided into three major steps:pre-processing, morphological analysis and contextual analysis. During pre-processing stage, the main components of the input text are identified[10]. Usually,preprocessing also segments the whole body of text into paragraphs and organizes these paragraphs into sentences. Finally,preprocessing divides the sentences into words. The morphological analysis servers the purpose of generating pronunciations and syntactic information for every word in the text. Morphological analysis determines the root form of every word and allows the dictionary to store just headword entries, rather than all derived forms of a word.Contextual analysis considers words in their context and determines the part-of-speech(POS) for each word in the sentence.Context analysis is essential to solve problems like homographs(words that are spelled the same way but have different pronunciations).Letter-To-Sound (LTS) module is responsible for automatically determining the incoming text's phonetic transcription. Prosodic features consist of pitch,duration, and stress over the time. Prosodic features can be divided into several levels such as syllable , word, or phrase level[15].

2.2. Digital Signal Processing (DSP) Module

Digital Signal Processing (DSP) Module

The operations involved in the DSP module are the computer analogue of dynamically controlling the articulatory muscles and the vibratory frequency of the vocal folds so that the output signal matches the input requirements . This can be basically achieved in two ways:Explicitly,in the form of a series of rules which formally describe the influence of phonemes on one another and Implicitly,by storing examples of phonetic transitions and co-articulations into a speech segment database,and using them just as they are ,as ultimate acoustic units[15].

4. IMPLEMENTATION

The proposed system is implemented using MATLAB.
Operational stages of the system can be summarized as:

- **Image Acquisition**

The images have been acquired by the camera and those are stored in the database. Below figure (3)shows the sample images.

Fig .3: Sample Images

- **Image Pre-processing**

The image which are in RGB format are converted to gray images by gray scale conversion. For this process,we calculate the average value of RGB for each pixel and if the average value is below than the specific value like 110,we replace it by black pixel and otherwise we replace it by white pixel.

- **Image Filtering**

Image filtering consists of removal of noise from the image. Firstly we count the threshold in gray image then according to that threshold we convert it into black and white image. Then we remove all the objects less than 30 pixels by bwareaopen command. bwareaopen morphologically opens binary image (remove small objects).

Fig 4: Sample Image pre-processed and filtered.

- **Optical Character Recognition(OCR)**

Optical Character Recognition consists of character image scanning, Binarization, Segmentation and Recognition.Correlation is calculated between the two variables such as R=corr2(A,B) computes the correlation co-efficient between A and B, where A and B are matrices or vectors of the same size.A and B can be numeric or logical and R is a scalar double.

- **Text-to-speech conversion via mobile**

The text to speech conversion is done by the system.speech.synthesis.speechsynthesizer by the system and it is remotely controlled by the mobile by which visually impaired user can hear the output voice.

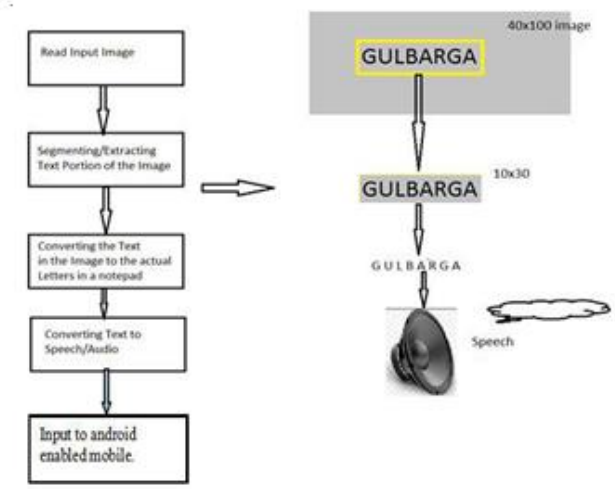

Fig 5: Test Result

There are two types of database created in this work.

1. Captured Text Images.
2. Created Images By Paint.

Following figure shows the implementation of the work.

- **Database/templates formation**

This window shows the captured images.

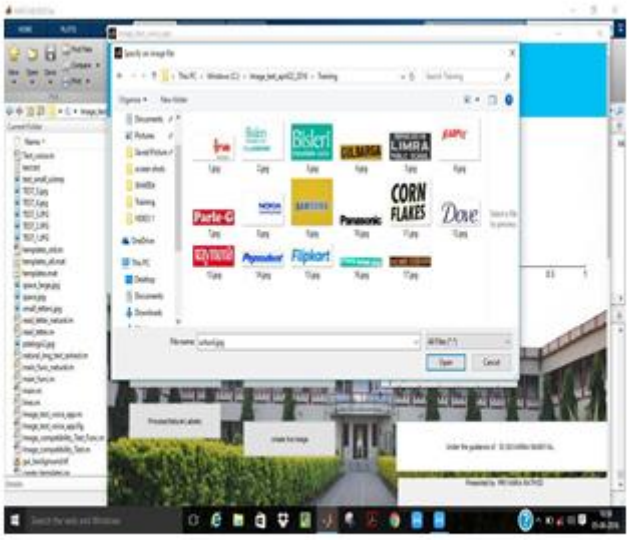

Fig. 6: The Database Creation of Sample Images

- **Image Acquisition**

Image are acquired from a high definition camera and totally 17 images are trained and stored in the database.

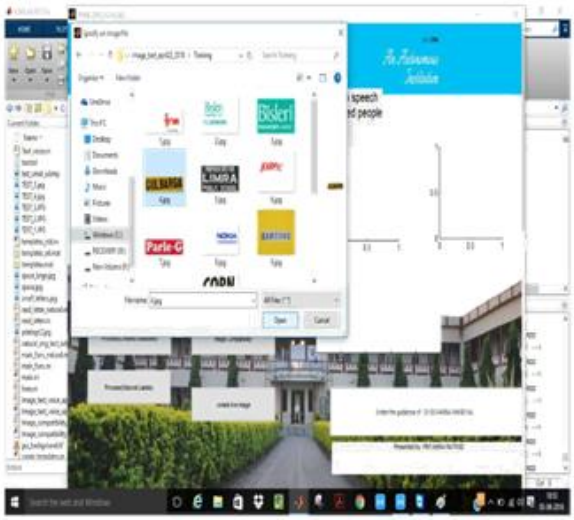

Fig.7: Selecting Trained Images

- **Image Selection**

The trained images are firstly preprocessed, and convert them from RGB to gray scale. The images are filtered by cropping the texted areas. Images are cropped by connected component of detected edges.

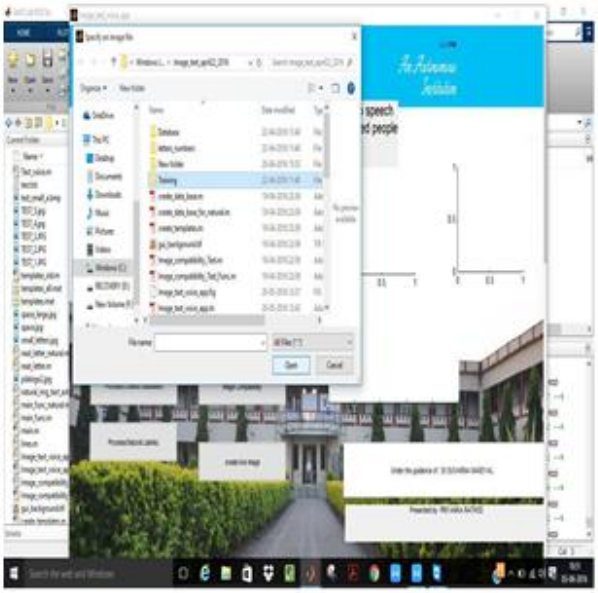

Fig 8: Selecting the Trained Images

1.1 Test Results of Captured Images

1.1.1 Processing the Captured Image

The Captured text images are trained and they are pre-processed.

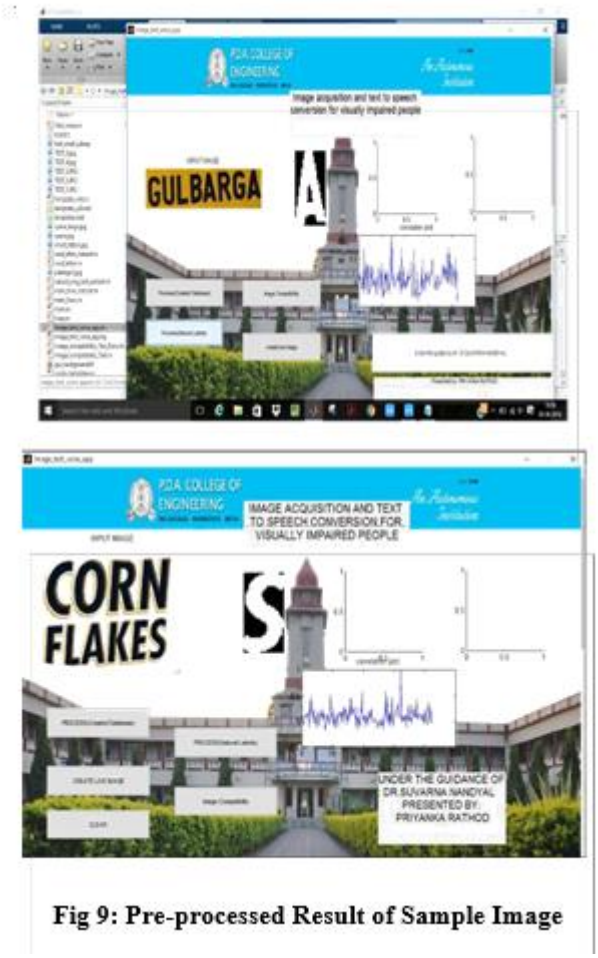

Fig 9: Pre-processed Result of Sample Image

1.1.2 Converting Image to Text

The images which are captured naturally are converted from image to text by the optical character recognition, which performs correlation by comparing two matrices variable.

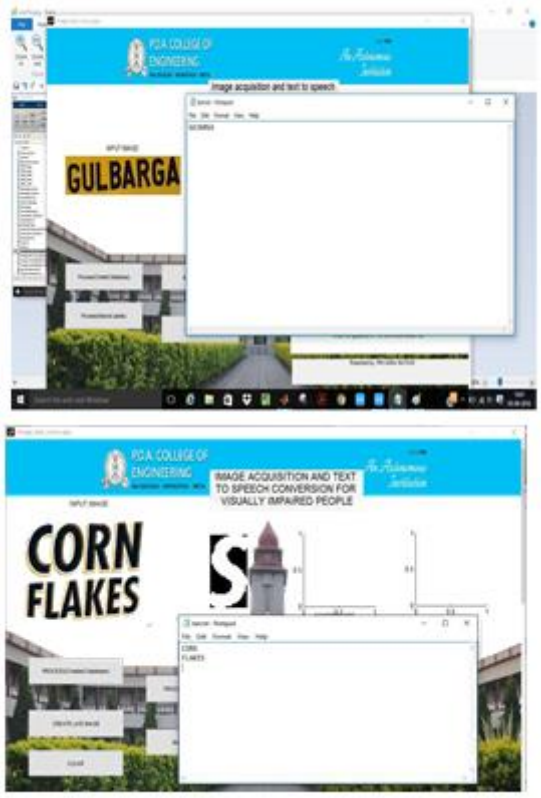

Fig 10: Image is converted to Text

1.2 Test Images Created By Paint

1.2.1 Sample Image Created By Paint

We can create a Live Image by using Paint and that image can be converted to text to speech using MATLAB.

Fig 11: Sample Image Created By Paint

1.2.2 Process Image Created By Paint

The created live image is converted to text to speech, text is audible via mobile through remote control software called as team viewer.

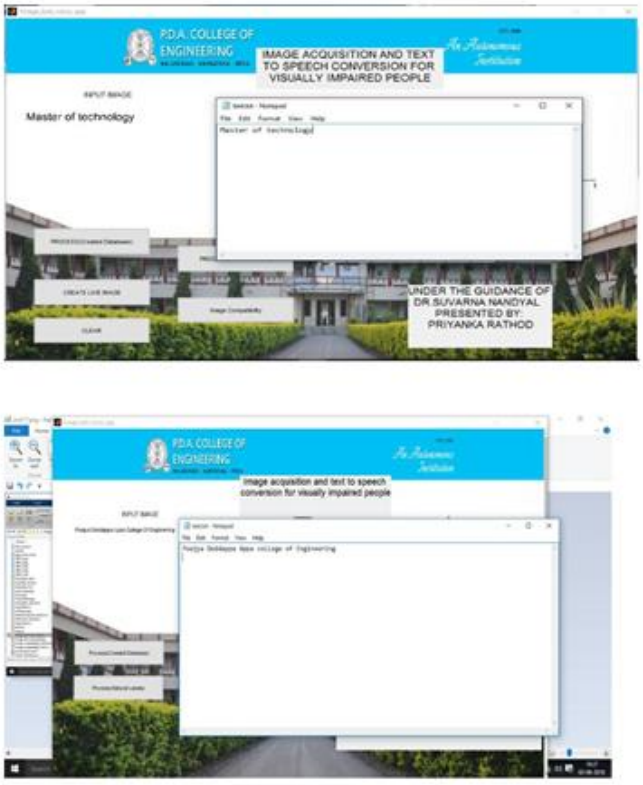

Fig 12: Process Created Image by Paint

1.3 Sample Images from Android Mobile:

Android mobile is used in the work as it is user friendly, fast and efficient to use. The improved android user interface ensures quick and easy remote control sessions.

Fig 13: Sample figures from the Android Mobile

5. CONCLUSION AND FUTURE WORK

The proposed work is an effort to suggest an approach for image to text and text to speech conversion using optical character recognition and text-to-speech technology. The application developed is user friendly cost effective and applicable in the real time. The camera plays a crucial role in the working of the system, hence the image quality and performance of the camera in real-time scenario must be tested thoroughly before actual implementation. This allows us to listen to our printed files, captured images or edit it in a word-processing program, through the Android mobile.

Future work includes navigation issue which may help visually impaired people to identify the shop while shopping ,with the help of maps and assist in geographical information to give the event message for driving people. The work can be implement in every android mobile without using remote control application. Further work can help to reduce the hardware part in the system.

REFERENCES

Baranski Przemyslaw, Polanczyk Macie , "*Mobile travel aid for the blind*", in the proceedings of the 4th European DSP and research conference,2013

Neha Gupta and V.K. Banga," *Image Segmentation for Text Extraction*", in proceeding of 2nd International Conference on Electrical, Electronics and Civil Engineering (ICEECE'2012), Singapore April 28-29, 2012.

R. Chandrasekaran and RM. Chandrasekaran , "*Morphology based Text Extraction in Images"*, in proceeding of International Journal of Computer Science and Technology(IJCST) Vol 2, Issue 4, Oct-Dec. 201

D. Wang and S. N. Srihari, "*Classification of newspaper image blocks using texture analysis*," Computer Vision Graphics, and Image Processing, vol. 47, pp. 327–352, 1989.

T. Pavlidis and J. Zhou, "*Page segmentation and classification,"* Computer Vision Graphics, and Image Processing, vol. 56, no. 6, pp. 484–496, 1992.

K.Matusiak, "*Mobile Camera Based Text Detection and Translation"*, Derek Ma Department of Electrical Engineering Stanford University,IEEE,2013.

Arpit Jain, Pradeep Natarajan, *"Text detection and recognition in natural scenes and consumer videos"*, IEEE International conference on Acoustic, speech and signal processing, 2014.

Hassan ,*"Morphological Text Extraction from Images"*, Arizona State University,Tempe,USA.

Chucai.Yi, "*Portable camera-based assistive text and product label reading from hand-held objects for blind persons"* ,IEEE,2013.

Xujun Peng, Huaigu Cao, Rohit Prasad and Premkumar Natarajan, *"Text Extraction from Video Using Conditional Random Fields"*,International conference on document analysis and recognition, 2011

Malik , *"Extraction of Text in Images"*,New Jersey Institute of Technology, NJ,USA.

Grover, Arora k, Mitra S K, "*Text Exraction from Document Image Using Edge Information"*, S.Inst .of technol., Rourkela.

Aparna.A, I.Muthumani, "*Optical Character Recognition for Handwritten Cursive English characters"*, Intenational Journal of Computer Science and Information Technologies, Vol.5 (1), 2014.

D.Sasirekha, E.Chandra, *"Text To Speech: A Simple Tutorial"*, International Journal of Soft Computing and Engineering, Vol.2, Issue-1, 2012.

Y. Sagisaga, *"Speech Synthesis from Text,"* 1998.

Dutoit, Thierry *"An Introduction to Text-To-Speech Synthesis,"* Boston: Kluwer Academic Publishers, 1997.