

Smart AI Assistance for Multiple Application

Anusha P

Dept. Of Electronics and Communication Engineering
NIE Institution of Technology
Mysuru, India

Shreyas A T

Dept. Of Electronics and Communication Engineering
NIE Institution of Technology
Mysuru, India

Bindushree C M

Dept. Of Electronics and Communication Engineering
NIE Institution of Technology
Mysuru, India

Ashwini D S

Assistant Professor,
Dept of Electronics and Communication Engineering
NIE Institution of Technology
Mysuru, India.

Abstract : Visual impairment is one of the biggest limitations for humanity, especially in this day for blind and age when information is communicated a lot by text messages rather than voice. In this project, we develop a smart AI for image text to speech algorithm that converts an image's text to speech. The basic framework is an digital image processing and free defined image to text data base, which extracts only the region of interest (i.e. region of the image that contains text) and converts that text to speech. It is implemented using a Dynamic Bayesian Networks with free defined commands image file. MATLAB software uses the image read command to convert the image files to speech samples. The captured image undergoes a series of image pre-processing steps to locate only that part of the image that contains the text and removes the background. Two tools are used to convert the new image (which contains only the text) to speech. They are OCR (Optical Character Recognition) software and TTS (Text-to-Speech) engines. The audio output is heard through the laptop inbuilt audio jack using speakers or earphones. The system takes the speech at run time through a microphone and processes the sampled speech to recognize the image text. In the training phase, the text digits are recorded using the OCR and Dynamic Bayesian Networks with free defined commands image file. MATLAB software uses the image read command to convert the image files to speech samples. An OCR (Optic Character Recognition) Model is used for speech recognition, which converts the image text to speech. It uses Dynamic Bayesian Network canny edge detection Kalman filter, morphological operation and windows speech recognition.

Key words: OCR, Segmentation, Text extraction, Templates, TTS, MATLAB

I. INTRODUCTION

Speech recognition is an interdisciplinary subfield of computer science and computational linguistics that develops methodologies and technologies that enable the recognition and translation of spoken language into text by computers. It is also known as automatic speech recognition (ASR), computer speech recognition or speech to text (STT) or text to Speech. It incorporates knowledge and research in the computer science, linguistics and computer engineering fields. Some speech recognition systems require "training" (also called "enrolment") where an individual speaker reads text or isolated vocabulary into the system. The system analyses the person's specific voice and uses it to fine-tune the recognition of that person's speech, resulting in increased accuracy.

Systems that do not use training are called "speaker independent" systems. Systems that use training are called "speaker dependent". Speech recognition applications include voice user interfaces such as voice dialling (e.g. "call home"), call routing (e.g. "I would like to make a collect call"), domestic appliance control, search key words (e.g. find a podcast where particular words were spoken), simple data entry (e.g., entering a credit card number), preparation of structured documents (e.g. a radiology report), determining speaker characteristics, speech-to-text processing (e.g., word processors or emails), and aircraft (usually termed direct voice input). The term voice recognition or speaker identification refers to identifying the speaker, rather than what they are saying. Recognizing the speaker can simplify the task of translating speech in systems that have been trained on a specific person's voice or image speech to text, it can be used to authenticate or verify the identity of a speaker as part of a security process. From the technology perspective, speech recognition has a long history with several waves of major innovations. Most recently, the field has benefited from advances in deep learning and big data. The advances are evidenced not only by the surge of academic papers published in the field, but more importantly by the worldwide industry adoption of a variety of deep learning methods in designing and deploying speech recognition systems.

II. LITERATURE REVIEW

[1] In the paper entitled with "OCR Based Facilitator for the Visually Challenged", Shalini Sonth et al. Here this paper proposes the performance of an ocr which is based on smart book reader mainly for visual impaired people. There is a requirement of Low cost portable smart book reader for visual impaired people.

[2] In the paper entitled with "OCR Based Image Text to Speech Conversion Using MATLAB", Sneha. Madre et al. This paper proposes a uses of an optical Character recognition which is used to convert text character into the audio signal. The

text is preprocessed and then used for recognition by segmenting each character. The text is processed and later it is used for identification of each character using the process

called segmentation. After segmentation extraction of each letter is carried out. This Text file is then converted into the audio signal.

[3] In the paper entitled with “OCR (Optical Character Recognition) Based Reading aid”, Sona.P et al.The paper describes a multilingual voice creation toolkit that supports the user in building voices for the Open source MARY TTS platform, for two state speech synthesis technologies: unit selection and HMM-based synthesis.

[4] In the paper entitled with “Blind Navigation System Using Artificial Intelligence”, Ashwani Kumar et al.The paper describes the conversion of the visual data by image and video processing into an alternate rendering modality that will be appropriate for a blind user.

[5] In the paper entitled with “Optical Character Technique Recognition Algorithms”, Abhishek Verma et al.The paper proposes a uses of an optical character generation system for conversion of images. The parsing consists of three phases: Character extraction, Recognition and post processing. During the recognition phase, the template with maximum 892 correlation is declared as the character present in the image.

[6] In the paper entitled with “Text to speech conversion using OCR”, Jisha gopinath et al.

The paper describes the image is converted into text then the text is converted into speech using MATLAB, LAB VIEW and Android Platform.

III. METHODOLOGY

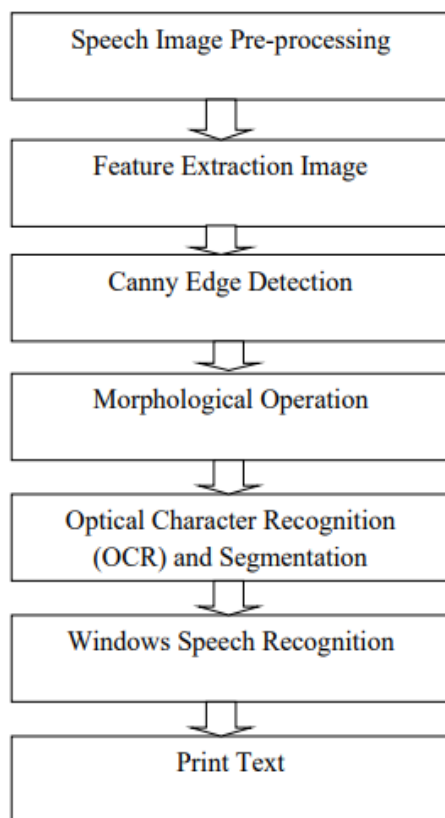


Fig.1. Building blocks of Image to Speech Processing

➤ **Speech Image Pre-Processing:** Preprocessing of speech signals is considered a crucial step in the development of a robust and efficient speech or speaker recognition system. The results get after applying the proposed strategies on some test voice signals are encouraging. It converts the pre-processed image, which is in .JPG form, to a .txt file.

➤ **Feature extraction:** Feature extraction is process of obtaining different features free defined commands image file such as apply today, bell, hell, now, save, heavy metal... etc. Feature extraction is the transformation of original data to a data set with a decreased number of variables, which contains the most discriminatory information. By decreasing the bandwidth of the input data, improved processing speed is achievable

Canny Method or Canny Filter: $BW = \text{edge}(I, 'canny')$ specifies the Canny method. $BW = \text{edge}(I, 'canny', \text{thresh})$ describes sensitivity thresholds for the Canny method. thresh is a two-element vector which the first element is the low threshold, and the second element is the high threshold. If you describes a scalar for thresh , this scalar value is used for the high threshold and $0.4 * \text{thresh}$ is used for the low threshold.

➤ **Morphological:** Morphological image processing is a collection of non-linear operations related to the shape or morphology of features in an image. A morphological operation on a binary image creates a new binary image in which the pixel has a non-zero value only if the test is successful at that location in the input image.

➤ **Optical Character Recognition:** OCR is a reader which recognizes the text character of the computer which may be in printed or written form. the OCR templates of each character are used to recognize the character i.e. scanning process is carried out. After this, the character image is translated in ASCII code which is further used in data processing.

- Tesseract is searching templates in pixels, letters, words and sentences. It utilize two-step approach that calls adaptive recognition. It needs one data stage for character recognition, then the second stage to fulfil any letters, it wasn't insured in, by letters that can match the word or sentence context.

- Image to text conversion: The above diagram shows the flow of Image-To-Speech. The first block is the image pre-processing modules and the optic character recognition. It converts the pre-processed image, which is in .JPG form, to a .txt file. We are using the Tesseract OCR.

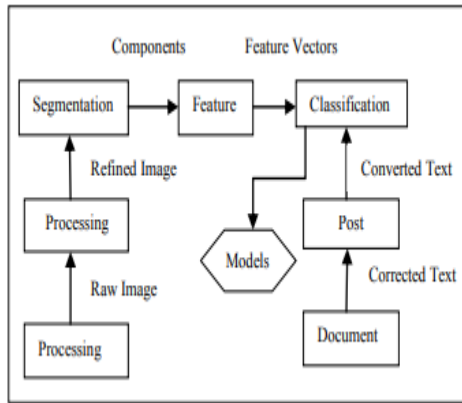


Fig.2. OCR Frame work

➤ **Text to speech conversion:** The second block is the voice processing module. It converts the image file to an audio output. Here, the image is converted to speech using a speech synthesizer called windows.

IV. FLOW CHART

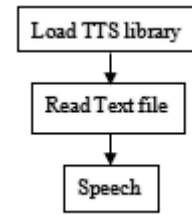
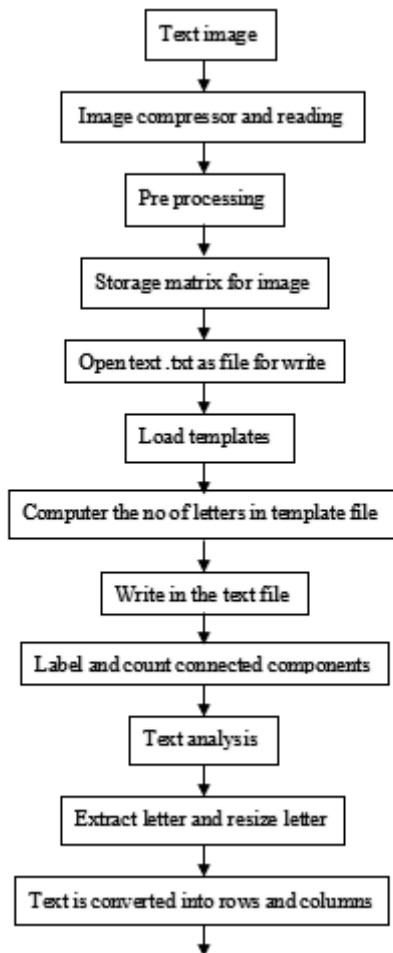


Fig.3.Flowchart

1. Step: The image is captured using a webcam and it is stored in the form of (.jpg) file format. Next step image is read/display using the imread command. It will read the image from stored file.

2. Step: The pre-processing conversion of the original RGB image into a grayscale is done by rgb2gray command. In this, as discussed above the pixels are made set and unset. The commands Rgb2gray it convert RGB image or color map to grayscale image. Rgb2gary conversion RGB image to grayscale by removing the hue and the saturation information which is Re-training the luminance.

The Command fid = open (filename) open the file using filename holds the name of the file to be opened. And characters are stored in an empty matrix.

3. Step: The image filtering. Filtering the unwanted noise is removed which smoothing the image for further processing.

4-10. Step: The process performs segmentation of each and every line, every letter. A correlation process in which the file is loaded so as to match the letter into stored templates.

11. Step: The next step is the conversion of text to speech. The first analysis, text and it will convert into speech using MATLAB.

12. Step: Finally get the speech for given image.

V. RESULT

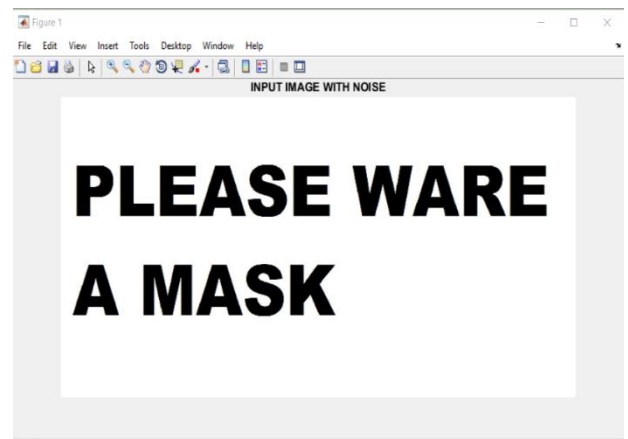


Fig.4. Input Text image

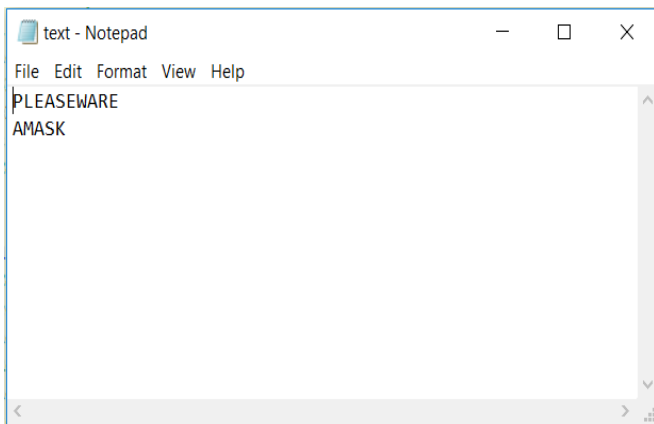


Fig.5. Output Text image

- [6] Jisha Gopinath S1, Arvind S2, Pooja chandranS3, Saranya S4, "Text to speech conversion using OCR", IJETAE, Vol.05, 2015.
- [7] Van Erp J.B.F and van Veen, H.A.A.C, "A mobile cloud pedestrian crossing for the blind," IEEE, pp 405-408,vol.212,2017.
- [8] Team Advisor: Joe McManus "wireless Navigation System for Visually Impaired," Engineering and Environmental Psychology," Capstone Research Project April 25, 2015.
- [9] Anusha A. Pingale 1, Devshree D. Mistry 2, Shital R. Kotkar 3, "smart bus alert system for navigation of blind persons," IEEE, pp.1123, 2017.
- [10] Saima Aman1 and Stan Szpakowicz1 "Identifying Expression of Emotion In Text," IEEE, pp. 196–205, vol. 4629, 2007.
- [11] Rishabh Gulati, "GPS Based Voice Alert System for the Blind", International Journal of Scientific & Engineering Research, Volume 2, Issue 1, January-2011.

VI. SCOPE AND FUTURE WORK

The scope of this project is promising. Optical Character Recognition finds its applications in Medicine, Online Retailers, Education, and more. The proposed system works on both English and Hindi scripts. Our aim for future work is to extend the same functionality for some more Indian regional languages such as Kannada, Tamil, Telegu, etc

VII. CONCLUSION

This paper proposes an implementation of a smart OCR based reader for the visually impaired. The project is implemented using the OCR and Dynamic Bayesian Networks with free defined commands image file is used. MATLAB software uses the image read command to convert the image files to speech samples. An OCR (Optic Character Recognition) Model is used for speech recognition, which converts the image text to speech. It uses Dynamic Bayesian Network canny edge detection Kalman filter, morphological operation and windows speech recognition techniques are used. The major goal of this project is to provide an affordable hand held device to the under-represented sections of the society, i.e., the blind and the visually impaired.

BIBLIOGRAPHY

- [1] Shalini Sonth et al, "OCR Based Facilitator for the Visually Challenged" International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques IEEE conference record (ICEECCOT) 2017.
- [2] Sneha. Madre et al, "OCR Based Image Text to Speech Conversion Using MATLAB", Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018) IEEE
- [3] Sona p et al, "OCR (Optical Character Recognition) Based Reading aid" International Research Journal of Engineering and Technology 2016
- [4] Ashwani Kumar, Ankush Chourasia, et al, "Blind Navigation System Using Artificial Intelligence", March 2018
- [5] Abhishek Verma et al, OCR- Optical Character Recognition, IJARSE, Vol. 5, Issue No 09, September 2016, pp 181-191.