

Real-time OCR for Regional Languages

Problem statement

OCR for regional languages is a quite innovative way that enables end users to hear the contents of text images instead of reading through them. It combines the concept of Optical Character Recognition (OCR) and Text to Speech Synthesizer (TTS). Text-to-Speech conversion is a method that scans and reads English alphabets and numbers that are in the form of image using OCR technique and converts them into voice.

OCR for English languages has been able to achieve a high percentage of accuracy in conversion but the OCR for Regional Languages is lacking in terms of accuracy. Moreover, text extraction from colour images is a challenging task in computer vision.

Background

Feature extraction and classification are the heart of OCR. Many of the researchers have found out a number of possible solutions Text to Speech Conversion using different speech synthesis. The existing TTS system consists of two phases. The first is text analysis, where the input text is transcribed into a phonetic or some other linguistic representation. The second one is the generation of speech waveforms. For numerical text to speech system, domain specific synthesis is applied. Typically, off-line OCR systems are designed for a particular script or language. However, the ideal approach to multilingual or regional language OCR would likely be to develop a system that can, with the use of language-specific training data, be re-targeted to process different languages with minimal modifications. This is still an open area of research with plenty of challenges. This is particularly true for regional languages handwriting recognition due to the added complexity of variations in writing styles even within the same scripts. There are lot of challenges for multilingual or regional languages OCR in pre-processing, feature extraction, script identification and recognition modelling.

Methodology

Optical character Recognition (OCR) is a process that converts scanned or printed text image, handwritten text into editable text for further processing. After text extraction it is converted to speech.

- Firstly, acquire the character image and read it.
- At second step i.e., pre-processing step, the colour image is converted into grayscale, then this grayscale image is converted into binary image by performing the threshold operation
- Character is extracted and resized in this step. Letters are resized according to templates size
- Then the load templates and 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.

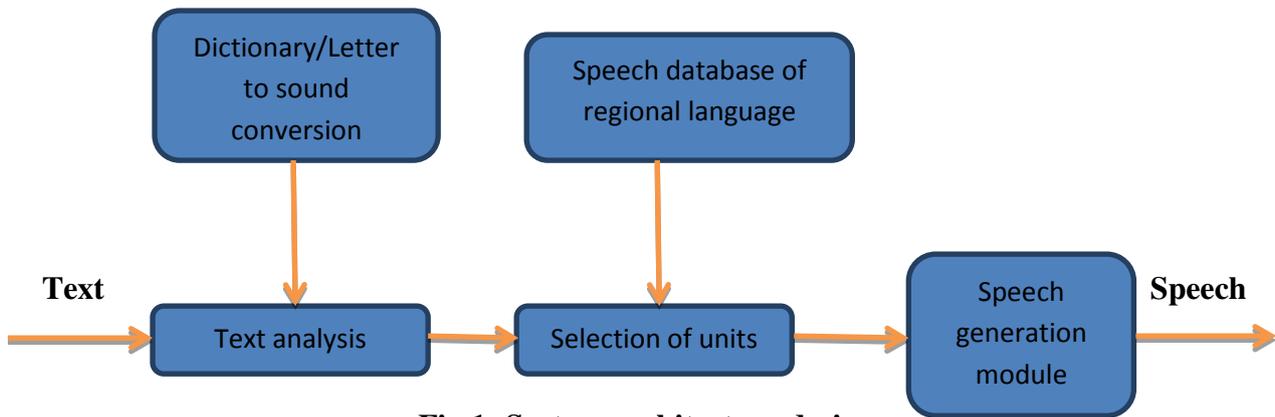


Fig 1: System architecture design

This system which is converting text to speech, is helpful for the blind people and people with low eyesight to read books and documents. It is also helpful for the dumb people to write the text which is then converted to speech. The speech is produced from the text by Text to speech synthesizer, by means of a grapheme-to-phoneme converting sentences to text content. Text Normalization is the transformation of text into the form of pronounceable. A complete system architecture design of the real-time OCR for regional languages is given in Fig. 1.

Experimental design

Dataset

The database containing words or sentences gives superior quality of outputs. Furthermore, a synthesizer can also include human voice characteristics to generate a synthetic but somewhat normal voice. Speech synthesizer is evaluated on the basis of the similarity of output speech obtained with that of the natural speech. Speech database is used for regional language, which performs formant synthesis.

Evaluation measures

Measures such as Accuracy, precision, and recall will be computed.

Hardware and software requirements

It includes camera, Ethernet and speaker. The input text image is captured by camera and the processed by image processing techniques and audio signal is given out on speakers. The Ethernet is used to access the e-dictionary website. Python based Deep Learning libraries will be exploited for the development and experimentation of the project. Training of the data will be conducted on NVIDIA GPUs. This software is used to convert the image file to text file by extracting the texts from the image and storing it in the file with .txt extension.