ISSN No:-2456-2165

Text Extraction from Digital Images with Text to Speech Conversion and Language Translation

Akhil Chandran Miniyadan
Assistant Professor, Department of Information Technology
College of Engineering Trikaripur
Kasaragod, Kerala, India

Nithya G P
Assistant Professor, Department of Information Technology
College of Engineering Thalassery
Kannur, Kerala, India

Abstract:- In this age of digital information, the need for digitalizing anything and everything is a growing need. Today most of the information is available either on paper or in the form of photographs. Tracking and modifying information from images is inconvenient and time-consuming. Thus we need to extract textual images into editable text. We do have technologies to extract text but they are mainly against clean backgrounds and it seemed to generate erroneous results. Thus, there is a need for a system to extract text from general backgrounds with more accuracy. Generating such text can be utilized as an input to a TTS Text-to-Speech where it converts any text into a speech signal and later on translated into the desired language. Our market is beamed with text extraction, TTS and translation products but is often in separate products. Our aim to achieve is combining all these text manipulations in one product and to be able to sync with existing text modifying products.

Keywords:- *Optical Character Recognition, Firebase ML Kit, Binarization, TTS, Discrete Contour Evolution.*

I. INTRODUCTION

Information today has been highly graphical and are stored in the form of images or in videos. Yet recent technology is restricted to how to retrieve those informational texts from the image. That's why text extraction plays a vital role in many applications that include information retrieval, digital library, multimedia systems, and many more. Text can be of immense use if can be converted to audio, especially for reducing visual reliability. Text-to-speech (TTS) is a process of producing spoken word from the text. As the world has grown into a global village, the diversity in native languages shouldn't make anyone from experience life as an outsider. Thus language conversion to the desired language of the user can be of great use, in another lingual atmosphere.

Text manipulations have been always in trend and in need. Text to speech systems was initially developed to assist the visually impaired by offering a computer-generated spoken voice that would read out loud the text. It beneficial, especially for blind people as they will be able to understand what is written by hearing it. Language translation is very helpful, especially to understand signboards.

In this paper, we propose a system which extracts text from a given picture and then convert it into speech and it can also be converted into many other languages. The system tries to combine the needed text manipulations in one single product.

II. EXISTING METHODOLOGY

Now, the old method to perform the text to speech conversion requires a web camera to acquiring an image and converting it into a text document using Optical Character Recognition (OCR). The next stage involves natural language processing and digital signal processing for converting the text into speech using Text to Speech synthesizer (TTS). The following paragraphs discuss various methods in the field.

The method proposed in [1] is a combination of scene text detection and scene text recognition algorithm. An image with text is given as input, preprocessing methods are used to remove noises. Binarization helps in identifying text from an image. Thinning and scaling is performed by connectivity algorithm if any data is lost during preprocessing. The approach in [1] uses a character descriptor to separate text from an image. The detected text is converted using a descriptor and wavelet feature. Sibling of each character is calculated using an adjacent character grouping algorithm. Stroke-related features are extracted using skeletons and character boundaries. There are 3 main steps in the implementation as follows: Given a synthesized patch from a training set, we obtain character boundary and character skeleton by applying discrete contour evolution (DCE) and skeleton pruning on the basis of DCE.

In robust algorithm [2], proposes a new text detection and extraction method that overcomes the weakness of previous approaches. The input image is first transformed into a binary image and edge detection is applied. Instead of performing a simple thresholding method, maximally stable external regions (MSER) are detected. These regions contain the text components and are appointed as white pixels. However, the resulting binary image does not reveal the exact boundaries of the text. For this reason, MSER binary image is enhanced by performing a thresholding operation on each connected component. Edges are then detected and fed into a stroke width detector where strokes, stroke widths, and connected components are found and filtered. In this using a robust algorithm, that proved to be effective on blurred images and noisy images as well.

Image processing done by implementing Tess-Two tesseract library [3] in Android studio and Google Vision API. Different Image processing techniques will be used to enhance the image quality by removing noise from an image, improving contrast and brightness of an image and a cropping image feature will also be provided for increasing the accuracy of detection of text from an image. For translating to Braille python library is used.

In [4] a text which is compatible with an android phone is extracted to speech. The system reads the text from the image and tells the user about it. It detects text area from natural scene image and extracts text information from the detected text regions. In text detection, analysis of color decomposition and horizontal alignment is performed to search for image regions of text strings. This method can distinguish text in the image from its background. An Adaboost learning model is applied to localize text in camera-based images. General methods include first, detecting text from the image and second, converting the image into speech. In TEXT DETECTION adjacent grouping methods and text, character stroke methods configuration is used. In the extracted text to speech, the mobile speaker informs the user of the text in speech or audio.

The mechanical system [5] captures the text from the image of a textbook and extract image from the text. The mechanical set will be consisting of a webcam or maybe an android mobile phone. The captured image is stored in a GUI. The different image processing techniques used are RGB to a grayscale image, contrast adjustment, segmented character recognition. The segmented character (image) is the input to optical character recognizer which converts it into text. The text to speech synthesizer is used to convert text into speech.

III. PROPOSED METHODOLOGY

Our proposed method combines the concept of Optical Character Recognition (OCR), Text To Speech Synthesizer (TTS) and language translator. Basically chunking the whole product into three stages.

The first stage involves extracting text from general scenes with accuracy using the camera and converting it into editable text using Optical Character Recognition (OCR). The second stage involves natural language processing where the detected text is converted into speech using Text to Speech synthesizer (TTS). Finally, the detected text is converted to different languages as per use.

A. Steps of Our Proposed Methodology

Basically the method deals with capturing a textual image, processing the image to extract text, converting the text to speech and optionally converting the text to other languages.

1) Image Capturing: This step involves to extract text from general scenes with accuracy, mainly from digital images by means of digital image processing techniques. This is done by using phone camera. Focused images captured on-the-go or those stored in the memory space can be utilized in this stage.

2) Image Processing: The captured image, in this stage, is processed. For this, we implement efficient text recognizing functions in firebase ML Kit's text recognition API. The process is efficient and tends to retrieve standard and uniform texts from the image. The textual image is scanned to identify elements of recognition like block, symbol, word and paragraph by the region bounded by coordinates of the area. After recognizing the elements of text, they are compared with supported languages available in Google Cloud platforms. This extracted text can be the input of the next module of TTS. Firebase ML provides machine learning capabilities both online and offline, thus being flexible as required.

3) Text-To-Speech(TTS): Text-to-speech (TTS) is a process of producing spoken word from by converting a given text to voice. Text-to-Speech function is a speech synthesizer that vocalizes text in real time in a natural way. Android TTS library helps to achieve this stage where the extracted text gets an audio output in a computer-generated voice. The text received from the text recognizer is analyzed against the database for TTS engine and spoken out in an audio format.

4) Language Translation: Language translation can be a great relief when in a different language speaking place. Several languages converting APIs are available in the market and can be opted for the purpose. The extracted text can undergo those operations to get desired language alterations for better understanding. For language translation, a self- learning statistical machine translation service can be used.

In this system, based on the analysis of translated texts a dictionary is constructed which works as the database of words and are compared along when a translation is required.

Fig. 1 portrays the proposed methodology. The text extraction, text to speech and language translation is done using best available methods.

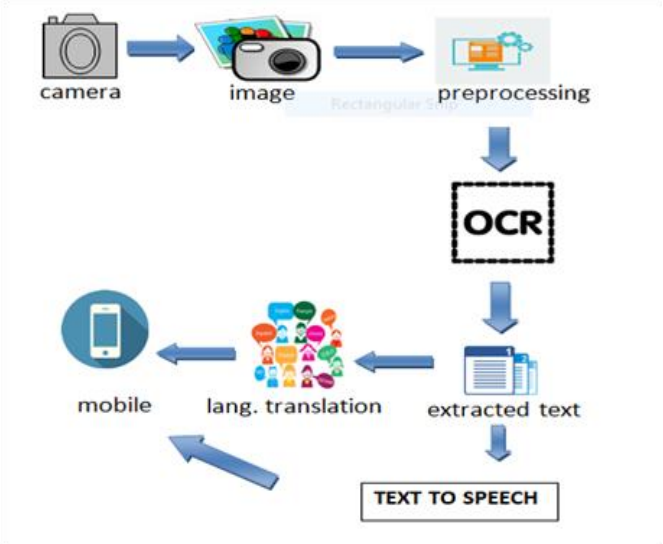

Fig. 1 Proposed Architecture

IV. CHALLENGES

Text recognition accuracy is directly proportional to the clarity and focus of the image. Thus image focus should be of desirable standards, for suitable results. Also, the image captured should have its texts of primary coverage and avoid unnecessary background elements in the image which only acts as noise for the process.

The way an idea is conveyed via words is different in a different context and that too different in a different language. Language translation challenges include this change in con-textual meaning, as it only translates to the general meaning.

V. CONCLUSION

Textual images and graphical information are in the scene today. So does their manipulative applications. Text extraction and manipulation have been in technology for some years now. Still, no product has gained any noticeable user satisfaction. By combining the much needed textual image manipulative operations, our method is an effort to fill that gap.

REFERENCES

[1]. M. Prabaharan and K. Radha, “Text extraction from natural scene images and conversion to audio in smart phone applications,” International Jour- nal of Innovative Research in Computer and Communication Engineering, vol. 3, pp. 19–23, 2015.
[2]. J. Yuan, Y. Zhang, K. K. Tan, and T. H. Lee, “Text extraction from images captured via mobile and digital devices,” in 2009 IEEE/ASME International Conference on Advanced Intelligent Mechatronics. IEEE, 2009, pp. 566–571.
[3]. K. Kalaivani, R. Praveena, V. Anjalipriya, and R. Srimeena, “Real time implementation of image recognition and text to speech conversion,” Int. J. Adv. Res. Technol, vol. 2, pp. 171–175, 2014.
[4]. V. Yadav and N. Ragot, “Text extraction in document images: highlight on using corner points,” in 2016 12th IAPR Workshop on Document Analysis Systems (DAS). IEEE, 2016, pp. 281–286.
[5]. N.-M. Chidiac, P. Damien, and C. Yaacoub, “A robust algorithm for text extraction from images,” in 2016 39th International Conference on Telecommunications and Signal Processing (TSP). IEEE, 2016, pp. 493–497.