ISSN 2395-1621

Image to Text and Text to Speech Conversion for Efficient Literacy Education

Pranali Newaskar, Aishwarya Phalke, Sonali Satpute, Trupti Padmwar,
Mrs. Snehal Shinde

pranalinewaskar2@ gmail.com
aishwaryaphalke2322@gmail.com
sonalisatpute155@ gmail.com
snehalshindejspmntc@gmail.com
truptipad17@gmail.com

JSPM Narhe Technical Campus

ABSTRACT

Optical Character Recognition has been used widely in various sectors over years where computer can recognize the characters from the image. OCR generally have three steps which are pre-processing, recognition and post-processing or verification process. Text Recognition is one of the difficult undertakings of PC vision with considerable viable interest. Optical character acknowledgment (OCR) empowers various applications for robotization. This task centers around word identification and acknowledgment in regular pictures. In contrast with perusing text in examined reports, the designated issue is essentially really testing. The utilization case in center works with the likelihood to identify the content region in normal scenes with more prominent exactness in view of the accessibility of pictures under imperatives. This is accomplished utilizing a camera mounted on a truck catching moreover pictures nonstop. The distinguished content region is then perceived utilizing Tesseract OCR method.

General Terms: Optical character recognition, image to text recognition, Text Recognition, Feature extraction.

Keywords: OCR, Pre-Processing, Classification, Recognition.

ARTICLE INFO

Article History

Received: 17th July 2021

Received in revised form :

18th July 2021

Accepted: 20th July 2021

Published online :

27th July 2021

I. INTRODUCTION

Images are becoming increasingly important and most of the information is presented as a combination of words and images. In today’s schools, many text-based learning materials are provided for students in a digital format.

An application used to select images and accordingly translate the data from the image into selected languages to increase the readability, understand ability which can be helpful for students as well as tourism. It is important that students not only have ability to derive literal meaning from texts but also develop an understanding of how the texts are produced.

Several methods for image-to-text multilingual translators are reviewed in detail. By disabling the gaps, which are identified by thorough review of the literature, an improved methodology is proposed. As a result, the development of application goes through five major phases including: Image selection, extraction, recognition, translation and speech conversion. Furthermore, Optical Character Recognition algorithm is particularly used for extraction and recognition of character with high accuracy under different environmental circumstances.

It translates text just by selecting an image from media storage and translation instantly appears on the user's screen in language selected by the user. The proposed solution may mainly be helpful in literacy education, for learning different languages and possibly it can work as visitors’ assistant. The application is also useful for the students with certain disabilities like dyslexia through the facility of text to speech using various methodologies.

II. RELATED WORK

In a proposed system, we are going to overcome existing drawbacks and provide real time features based image processing domain system by using open-cv python. Image processing done through algorithm and methods. Image to text and text to speech estimation done through inception modules created in tessaract in python.

We are going to develop following modules:

Image Acquisition:- Using open computer vision library. We are going to capture real time images of user. After getting faces we are forwarding these images for feature extraction and image processing.

Image processing:- After getting images by using filters and further used to process or deform.

Feature Extraction:- Feature extraction is the step of getting component features like color, shape, texture etc from real time images. Feature extraction is very much important for the initialization of processing techniques and finally characters recognition. Among all these features, character localization and detection is essential, from which locations of all other features are identified.

Image To Text:- Our main aim behind to develop this application is to give multi factor system. Which overcomes existing security problem by our recognition based tessaract. An un- authenticated things recognized then predicted data will send to our system.

Tools and Technologies Used

Image Processing:

Every image is formation of RGB colors. Each and every captured image has some noise, un- wanted background. Thus there is necessity of process those captured image before assign to our recognition module. Pre-processing unit made up of noise removal, grey image conversion, binary image conversion of input images after that feature extraction done on those samples. In future extraction five steps are applied in which finding the unconventionality. Next elongations of images are evaluated by calculating pixel segmentation as well as rotation of input images.

Recurrent Neural Networks:-

Brain-inspired systems are to replicate how humans learn. Consist of input, invisible and output layers that transform the input into something that the output layer can use. It is Excellent for finding patterns which are complex for humans to extract and teach the machine to recognize. RNN gathers their knowledge by detecting the patterns and relationships in data and learns (or is trained) through experience, not from programming. RNN takes in processed images as input.

III. IMPLEMENTION DETAILS

Algorithm:

1] Tessaract

This paper proposes a text recognition pipeline for translating text in the ordinary scenes. The goal is achieved in two steps. Initially, the text area is noticed which plays an important part in the overall performance of the OCR engine. Secondly, the noticed text area is translated using tesseract V5. Tesseract is an precise and open-sourced OCR engine from Google. A detailed explanation of the steps involved is discussed in the following sections.

Data Collection

The images are taken using a USB camera module mounted at the back end of a logistic truck. Images are captured when the swap body is selected up, transported and set down at the target location. Moreover, all images are stored in the cloud using Microsoft Azure. The images obtained contains noise and it is needed to pre-process the images. Also, images are captured round- the-clock and hence there is a weighty variation in the brightness of images.

Text Detection

Pre-processing Steps

Pre-processing is a key module and undoubtedly affects the results in successive steps. Images kept in the cloud contains a significant amount of noise. The camera is mounted at a somewhat inclined angle with respect to the text required for recognition. The following methods are used to accomplish the pre-processing stage which has proved critical in text extraction. Apart from the low brightness of night images, there is a reverse flashlight falling exactly on the checksum number which was a challenge for obtaining better results. Fixing the brightness of the image is an important requirement for the Tesseract OCR engine to work efficiently.

The images attained are constrained and the text area of interest is always on the lower right side of the image. The constraints and individuality of input images helped in fixing a cropping area so that the unwanted areas from the image were eliminated.

Brightness check: After cropping the images, the following problem was the varying brightness levels of the image. The night images also faced the difficulty of an additional flash-light which markedly changed the illumination effects on the image. Systematically to differentiate the images, the brightness of the images were calculated using root mean square (RMS) pixel brightness method. The images are then categorized into different types followed by gamma correction. This method demonstrated success when the gamma values of the images are varied depending on the brightness levels of the is. A lookup table (LUT) was form mapping the input pixel values to the output gamma-corrected values. It caused in quicker gamma improvement using Open CV Skew correction: As already discussed above, the images are captured at an angle and to deal with the geometric distortion caused by the camera place, a skew correction was performed to get an aligned text area. In adding, deskewing and dewarping the image are seen important for efficient text recognition in later stages. After finding the edges by a canny edge detector, the strong lines in the image are detected using hough line transformation. Progressive probabilistic Hough line conversion is used which reduces the computation effort. There are two main type of arguments: minimum length of line and the maximum allowed gap between line segments which together helped to detect the strong lengthy lines in the image. Finally, geometric change such as perspective transformation is used to achieve necessary skew correction. The transformation matrix for perspective transformation is completed from above-mentioned Hough line detection.

Finding the Text Area In order to decrease the noise, blurring is applied to the image. Using a blackhat operator, dark regions in the images are discovered. The text in the images is dark and in contradiction of a light background which reinforced the blackhat operation. As a next step, using a Scharr operative, the gradient magnitude illustration of the blackhat image is computed. Scharr operator optimizes the rotation symmetry by minimizing the biased meansquared angular error. The gaps among the characters are closed using a morphological final operation. Finally, the Otsu thresholding way is applied to the image.

Text recognition

Text recognition is done using Tesseract V5 from Google. Tesseract is deep learning-based text recognition model which not only possess great accuracy levels but also provisions a wide variety of Languages. However, Tesseract OCR contains of some important assumptions for text recognition. It performs accurately with documented text and the precision is generally limited to controlled conditions. Hence, pre-processing of images and extraction of the text area are done keeping these factors in mind.

This project used Tesseract 5 with LSTM as the OCR engine mode. Moreover, fine-tuning the page segmentation modes (PSM) resulted in significant enhancements in the overall accuracy of the pipeline. After cautious tests and comparisons, two different page segmentation approaches were used. For the first part of recognition the image was preserved as a single character and for the second part, the automatic page segmentation method was used. A detailed explanation of the steps involved is explained in the next section. The detected text is then assessed using text post-processing. The checksum value is distinctly calculated each time based on the first 10 characters after recognition. The last digit in square brackets is checksum number. In mandate to verify the recognized text, the calculated checksum number is matched with the detected checksum number. If they are equal, then the extracted text is proved and stored as true detection. Excitingly, later tests with a larger dataset resulted in false-positive cases which are discussed in the following section. After careful assessment of some of the false-positive cases, the clue was to cross-verify by breaking the detected ROI and separately extract the text. A two step assessment process was performed. The noticed ROI is manually split into two parts: the registration number and the checksum number. Later, the text recognition process was done separately on each part and cross verified with the checksum number. PSM 3 resembles to automatic page segmentation method and the presence of square boxes around the checksum number was a reason to go with this technique. Furthermore, the character was inadequate to only digits using tesseract arguments. The cell recalls values over arbitrary time intermissions, and the three gates adjust the flow of information into and out of the cell. The cell of the model is accountable for keeping track of the dependencies among the elements in the input sequence. The input gate controls the range to which a new value flows into the cell, the forget gate panels the extent to which a value remains in the cell, and the output gate controls the extent to which the value in the cell is used to calculate the output activation of the LSTM unit. Though, there are certain variants of the LSTM model such as Gated Recurrent Units (GRUs) that do not have the output gate. LSTM Networks are generally used on time-series data for classification, processing, and making predictions. The cause for its popularity in time- series application is that there can be several lags of unknown duration between important events in a time series. To overwhelmed the drawback of traditional RNNs, 3 gates are added into the cell of the network to simplify the notion of memory:

LSTMs has four gate: 1) forget (f), 2) input (i), 3) memory (c) and 4) output gate (o).

Given an old memory Ct1, the new cell memory Ct is computed.

IV. RESULT

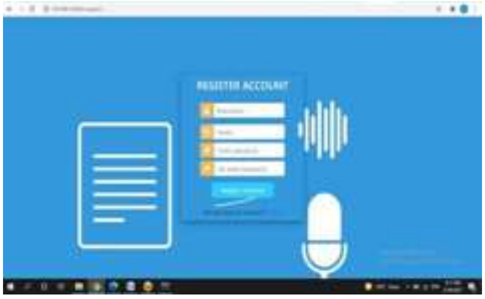


Fig. Registration Page

Above screen represent registration page. Before doing anything user must have to register account.

During registration user enter the values in fields like username, email and password. After that successful registration is done.

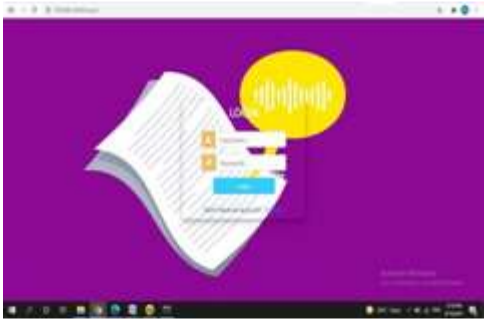


Fig. Login Page

Above screen represent login page. After registration user can log in to the application using username and password.

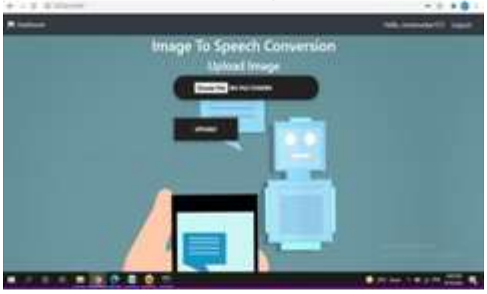
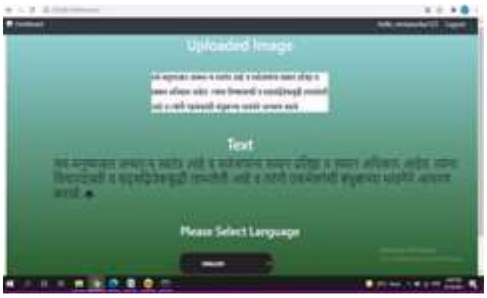


Fig. Upload Image

Above screen represents Image to text conversion. There are two buttons are provided one is for choosing file from media and second is for uploading that image.

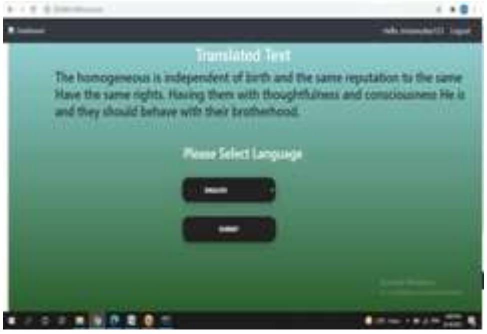


[illegible]

[illegible]

Fig. Image to Text

Above screen represents uploaded image and output of that image that is text output which is extracted from that input image. At bottom there is a option of selecting language for translation.



The homogeneous is independent of birth and the same reputation to the same Have the same rights. Having them with thoughtfulness and consciousness He is and they should behave with their brotherhood.

Fig. Translated Into English

Above screen represents output of translated text. After that there is a option of selecting language option for speech output.

Fig. Translated text Into Speech

Above screen represents Speech output which is converted from the text. This is the final output.

V. CONCLUSION

Based on the proposed methodology, an application has been developed by studying the pervious developments. This app aims to help the people travelling around the globe. We have used OCR algorithm for image extraction and recognition. After capturing an image, it will extract the text from the image and then recognize the characters. When this process cycle is completed it looks up for the text in its database for translation. If it finds the text in its database, it gets translated if not the application will automatically search online translation for the text. The converted text is taken as an input for conversion of that text in to speech.

Future work

In future Android Application will be developed which will be useful for students, tourist etc.

Applications

The proposed project image to text and text to speech is used in public sector.

VI. REFERENCE

1] Quang Anh BUI Salvatore Tabb one “Selecting automatically pre-processing methods to improve OCR performances” IEEE 2017.

2] Gupta Mehul, Patel Ankita, Dave Namrata, Goradia Rahul, Saurin Sheth “Text-Based Image Segmentation Methodology” ICIAME 2014.

3] Mr. Pratik Madhukar Manwalkar, Mr. Shashank H. Yadav “Text Recognation from image” IEEE 2015.

4] Pradeep Kumar Bhatia “A Detailed Review Feature Extraction in Image Processing Systems” IEEE 2014.

5] Rohit Verma, Dr. Jahid Ali “A Survey of Feature Extraction and Classification Techniques in OCR Systems” IJCAIT 2012.

6] Nidhi sawant, “Devanagari Printed Text to Speech Conversion using OCR” ,I-SMAC 2018.

7] Muhammad Ajmal, “Image to Multilingual text conversion for education literacy”, 17th international conference 2018.

8] Minal Acharya,”Scan.it – Text Recognition, Translation and Conversion”.

9] M.S. Akopyan,” Text recognition on images from social media”,2019 IVMEM.