International Journal of Advance Research in Science and Engineering
Volume No. 10, Issue No. 12, December 2021
www.ijarse.com
IJARSE
ISSN 2319 - 8354

EASY TRANS USING MACHINE LEARNING

Santhiya T, Sanjai KRP, Ramya P

[1)] *Computer Science and Engineering (CSE), Bannari Amman Institute of Technology, India*
[2)] *Computer Science and Engineering (CSE), Bannari Amman Institute of Technology, India*
[3)] *Computer Science and Engineering (CSE), Bannari Amman Institute of Technology, India*

ABSTRACT

The point of our undertaking to mechanize the application to defeat from the language boundary among nations and among states inside the same country. Existing framework, having a different application for every single cycle like camera, Google interpreter and Optical Character Recognition (OCR) text scanner. However, individuals expect the application comprises of the relative multitude of three offices together. Application can see the message present in the image which set aside in system or found using camera and make a translation of the message into the language required and show the understanding outcome back on to the screen of structure. Optical Character Recognition (OCR) is utilized to digitize the words. In the last advance we feed the identified text into our text-to-discourse synthesizer (TTS) for text to discourse change. The above mentioned calculation is applied on images from reports to regular image scenes. Promising results have been represented which show the precision and generosity of the proposed structure and engage its rational execution in evident application.

Keywords - *Text Extraction, Image Processing, Digitization, Optical character recognition, Language translation, Text to speech.*

I. INTRODUCTION

These days in correspondence the language hindrance are make issue for viable correspondence for this we introduced this application. Discourse acknowledgment and text interpretation are fundamentally utilized for changing the discourse over to text and text to discourse for understanding the language which are spoken by client during correspondence, on account of this individual can perceive the discourse are spoken by any other individual. OCR is used for image to text conversion. Optical person acknowledgment (OCR) is accustomed to separating the text from pictures which can be manually written, billboards and so on This picture extraction is accustomed to understanding the language of sentences which into picture.

1.1 Problem Statement

Usage of mobile phones has become a big deal. Numerous Text to discourse, discourse acknowledgment, multilingual interpretation, text extraction applications are produced for versatile clients. In any case, comparative sort of uses isn't created for work area clients. We are making an application that comprise all of the above application in one single application. To help the blind people and illiterates who are facing difficulty in understanding the content they have.

1.2 Objectives

Our objective point is to join all various assignments, for example, discourse acknowledgment, text interpretation, text combination and text extraction from picture all inserted in one with the goal that we get an easy to understand application. Following are the few objectives,

- To improve the speech recognition and text to speech conversion.
- Creates binary image for the image recognition through image processing.
- Reinforcing the sound result.

1.3 Scope

We encourage this application for workspace application. Here we are coordinating the image extraction, language translation and text to speech conversion, all together in a single framework so user doesn't need to download different diverse applications.

II. METHODOLOGY

The proposed framework comprises of three main fundamental modules, text extraction, language translation and text to speech discourse change. The picture handling module catches text in the picture. At first the picture handling module, where OCR changes over .jpg (image) format to .txt (digitalized file) format and during the voice handling module it changes over .txt to .mp3(audio)format. OCR is vital component in this module. OCR or Optical Character Recognition is an innovation that naturally perceives the person through the optical system, this innovation insinuate the capacity of the human feeling of vision. Prior to taking care of the picture to OCR it is changed over to double picture to build the picture acknowledgment exactness. Picture double transformation is finished by Image wizardry programming, which is one more open source device for picture control. The result of OCR is the text, which is put away in a record design (.txt). It needs some supporting condition all together get the negligible deformity. Voice handling module changes the text into discourse and cycles it with explicit qualities so the sound can be perceived and clear.

2.1 Image processing

A pixel is a solitary speck of a specific tone. An image is essentially a variety of pixels. More the pixels in a picture, the higher will be its goal. A PC doesn't comprehend the pictures and images, that is a picture of a sign is actually a sign, it simply realizes that the primary pixel is this tone, the following pixel is that tone, and shows each of its pixels so that you might be able to see.

Stage 1: Pre-Processing the Image

Before text can be pulled, the picture should be kneaded in some ways to make extraction more straightforward and bound to succeed. This is called pre-handling, and distinctive programming arrangements utilize various mixes of methods. The more normal pre-handling methods we are utilizing are:

Binarization

Each and every pixel in the picture is changed over to one or the other dark or white. The objective is to clarify which pixels have a place with text and which pixels have a place with the foundation, which speeds up the genuine OCR process

Deskew

Since archives are seldom filtered with wonderful arrangement, characters might wind up skewed or even topsy turvy. The objective here is to distinguish even text lines and afterward pivot the picture so those lines are really flat.

Despeckle

Whether or not the picture has been binarized, there might be commotion that can meddle with the ID of characters. Despeckling disposes of that commotion and attempts to streamline the picture.

Line Removal

Recognizes all lines and markings that probably aren't characters, then, at that point, eliminates them so the real OCR process doesn't get confounded. It's particularly significant when checking reports with tables and boxes.

Drafting

Isolates the picture into particular lumps of text, for example, distinguishing sections in multi-segment archives.

Stage 2: Processing the Image

First of all, the OCR interaction attempts to build up the gauge for each line of text in the picture (or on the other hand on the off chance that it was drafted in pre-handling, it will deal with each zone each in turn). Each distinguished line of characters is taken care of individually. For each line of characters, the OCR programming distinguishes the dividing between characters by searching for vertical lines of non-text pixels (which ought to be clear with legitimate binarization). Each piece of

pixels between these non-message lines is set apart as a "token" that addresses one person. Consequently, this progression is called tokenization. When each of the possible characters in the picture are tokenized, the OCR programming can utilize two distinct strategies to recognize which characters those tokens really are: The tokens and glyphs should be of comparative size or, in all likelihood not a solitary one of them will coordinate.

2.2 Pattern Recognition

Every token is contrasted pixel-with pixel against a whole arrangement of known glyphs—including numbers, accentuation, and other uncommon images—and the nearest match is picked. This method is known as lattice coordinating(Pattern Recognition). There are few disadvantages in this. First Second, the tokens should be in a comparative text style as the glyphs, which precludes penmanship. However, in the event that the symbolic's text style is known, design acknowledgment can be quick and exact.

2.3 Feature Extraction

Every token is looked at contrary to changed guidelines that depict what sort of character it very well may be. For. model, two equivalent stature vertical lines associated by a solitary flat line are probably going to be a capital H

Stage 3: Post-Processing the Image Lexical Restriction

All words are thought about against a vocabulary of endorsed words, and any that don't coordinate are supplanted with the nearest fitting word. A word reference is one illustration of a dictionary. This can help right words with mixed up characters, like "thorn" rather than "thorn".

2.4 Voice handling

In this strategy our fundamental plan to change got text over to discourse with the assistance of coding in raspberry pi utilizing text to discourse (TTS) synthesizer. The text to discourse synthesizer is introduced in raspberry pi. The result can be tuned in through sound speakers.

III. Description

Image processing: In the initial step the gadget is moved over the printed page and the inbuilt camera catches the pictures of the text. The nature of the picture caught will be so high to have quick and clear acknowledgment because of the great goal camera. Letters will be separated and changed over into computerized structure. It comprises of three stages: Skew Correction, Linearization and Noise Removal. The caught picture is checked for slanting. There are potential outcomes of picture getting slanted with one or the other left or right direction. Here the picture is first lit up and binarized. The limit with respect to incline area checks for a state of bearing and accepting that recognized then a clear picture turn is finished till the lines coordinate with the even midpoint, which conveyed an

inclination amended picture. The commotion presented during catching or of succession of characters into sub picture of individual image (because of low quality of the page must be cleared for additional handling. Division: This activity tries to decay a picture characters). The binarized picture is checked for interline spaces. On the off chance that entomb line spaces are identified then the picture is portioned into sets of sections across the interline hole.

The lines in the sections are checked for even space convergence concerning the foundation. Histogram of the picture gave to distinguish the width of the even lines. Then, at that point, the lines are filtered upward for vertical space convergence. Here histograms are utilized to identify the width of the words. Then, at that point, the words are disintegrated into character width calculation. Include Extraction: In this stage we accumulate the fundamental highlights of the picture called highlight maps. One such technique is to recognize the edge in the picture, as they will contain the necessary text. Tesseract: This product is utilized to change over the picture record to message document by separating the messages from the picture and putting away it in the record with.txt expansion.

IV. FLOWCHART

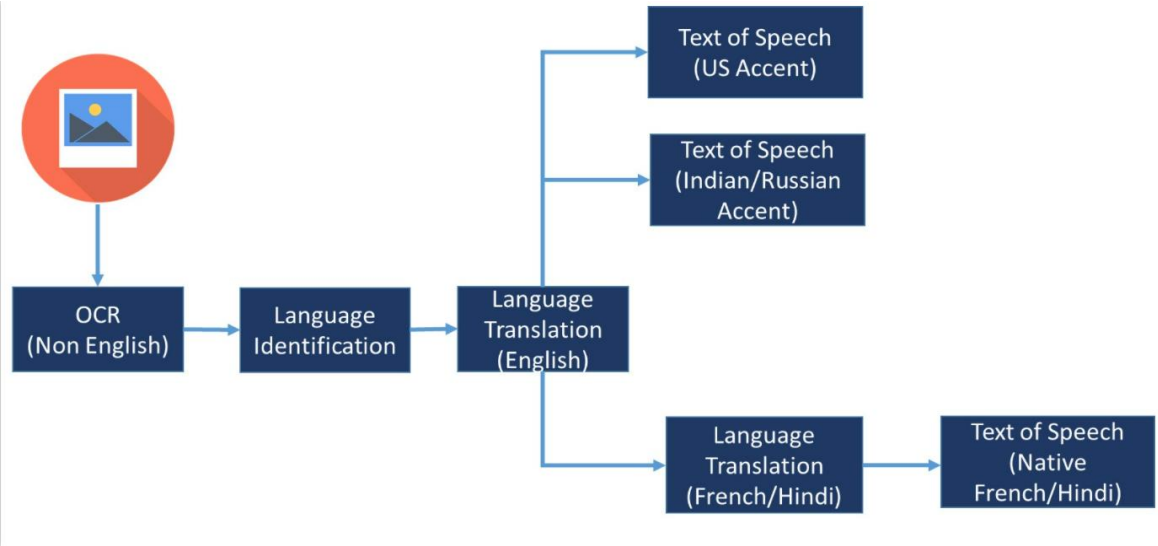

Fig. 1

V. IMPLEMENTATION

5.1 Optical Character Recognition

Optical character recognition, abbreviated as OCR, is the mechanical or electronic change of filtered pictures of written by hand, typewritten or printed text into machine encoded text. It is generally utilized as a type of information section from a type of unique paper information source, regardless of whether archives, deals receipts, mail, or quite a few printed records. It is essential to the

computerization of printed messages so they can be electronically looked, showed on-line and utilized in machine cycles, for example, machine interpretation, message to-discourse and message mining.

5.2 Text translation using Google Translator:

Google translator is a free multilingual measurable and neural machine interpretation administration created by Google, to decipher text and sites from one language into another. Google Translator can progressively interpret text between large number of language sets. Interpretation lets sites and projects automatically coordinate with the interpretation administration. Google interpretation utilizes a making an interpretation of technique to a framework called neural machine interpretation. It utilizes profound learning strategies to decipher entire sentences all at once. Neural machine interpretation: Google Neural Machine Translation (GNMT) is a neural machine interpretation (NMT) framework created by Google and presented in November 2016, that utilizes a fake neural organization to expand familiarity and exactness in Google. Google Translate is a neural machine interpretation framework utilizes a huge start to finish fake neural organization that attempts to perform profound learning, specifically, long momentary memory organizations. GNMT works on the nature of interpretation over SMT in certain occurrences since it utilizes a model based machine interpretation (EBMT) technique in which the framework "gains from a large number of models". It interprets entire sentences all at once, instead of simply piecing by piece. It utilizes this more extensive setting to assist it with sorting out the most important interpretation, which it then, at that point, revamps and acclimates to be more similar to a human talking with legitimate language structure. The GNMT organization can embrace Interlingua machine interpretation by encoding the semantics of the sentence, rather than by remembering phrase-to-state interpretations. By utilizing a huge number of models, GNMT works on the nature of interpretation, utilizing more extensive setting to conclude the most significant interpretation. The outcome is then revised and adjusted to move toward linguistically based human language. Right when Google Translate produces an understanding recommendation, it looks for plans in a colossal number of reports to help choose the best understanding. By recognizing designs in archives that have as of now been interpreted by human interpreters, Google Translate makes informed conjectures (AI) with regards to what a proper interpretation ought to be. Presently, discourse is orchestrated for the interpreted message utilizing Google Text-to-discourse.

5.3 Text to speech using Text-to-Speech Speech Synthesis:

Google Text-to-Speech changes over text into human-like discourse in excess of 180 voices across 30+ dialects and variations. It applies historic examination in discourse union (WaveNet) and Google's amazing neural organizations to convey high-loyalty sound. It creates discourse that copies human voices and sounds more normal, lessening the hole with human execution by 70%. Cloud

Text-to-Speech is fueled by WaveNet, programming made by Google's UK-based AI auxiliary DeepMind. WaveNet utilizes AI to create discourse. It then waveforms from a data set of human discourse and yet again makes them at a pace of 24,000 examples each second.

5.4 Process Flow:

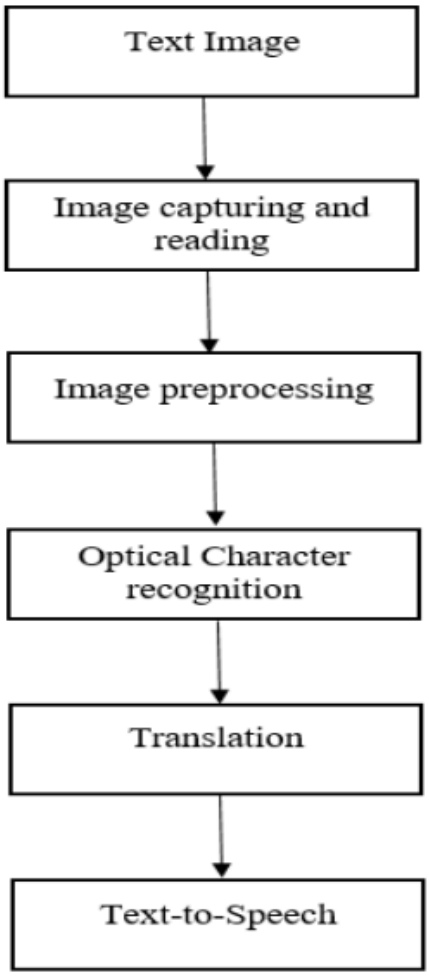

Fig. 2

VI. APPLICATIONS AND ADVANTAGES

6.1 Advantages

- Further developed word acknowledgment abilities, familiarity and exactness.
- Overall market entrance.
- It is really reliable and simple to utilize.
- Upgraded representative execution.

- Broaden the compass of your substance.
- People with different learning styles.
- It is exceptionally helpful for unskilled people.

6.2 Applications

- Changes over text into spoken voice yield.
- It is mostly intended for blind individuals.
- It helps little sorts with additional creating word affirmation, recall information while scrutinizing.
- We can give assorted language access for the people who can't examine and make.
- Impact of Text to Speech on Tourism.
- It very well may be utilized as declaring gadget in schools/universities and in different spots.

VII. EXPERIMENTAL RESULTS:

The entire procedure consists of three stages. First we capture the image of the required text which is then pre-processed to get better results. The input image is shown below

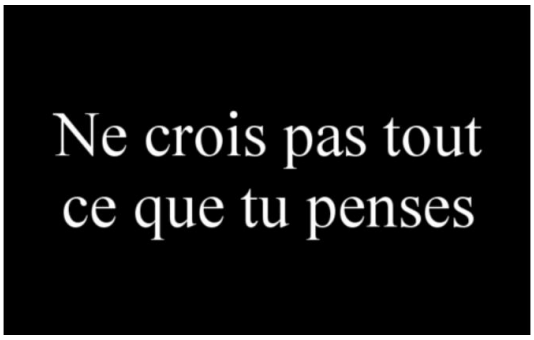

Fig. 3

The image is tokenized and text is extracted using pattern recognition.

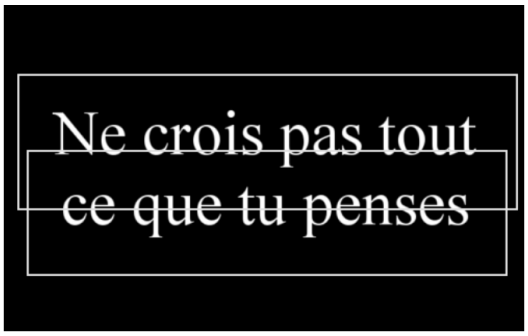

Fig. 4

The text from this image which is required to be translated is recognized using Google Vision and this result is shown below

```
'Ne crois pas tout ce que tu penses'
```

Fig. 5

Now this detected text is translated into user required language using Google Translate. Here it is translated into Hindi language. The result is shown below

```
Do not believe everything you think
```

Fig. 6

Using Text to speech the translated text speech is synthesized simultaneously and voice is heard through speaker.

VIII. CONCLUSION

The proposed strategy effectively identifies the text areas in the majority of the pictures and is very precise in separating the text from the distinguished areas. Based on the experimental analysis that we performed we found out that the proposed method can accurately detect the text regions from images which have different text sizes, styles and color. Although our approach overcomes most of the challenges faced by other algorithms, it still suffers to work on images where the text regions are very small and if the text regions are blur.

REFERENCE

[1] NavyaPant, AyushiTrivedi, Supriya Agrawal, Pinal Shah and SimranSonik, “Speech to text and text to speech recognition systems” , IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727, Vol-20, Issue-2, Ver-I, March.April-2019.

[2] S. Fong, A. Elfaki, M. bin Md Johar and K. Aik, 'Mobile language translator', 2011 Malaysian Conference in Software Engineering, 2019.

[3] Archana A, Shinde D. Text pre-processing and text segmentation for OCR. International Journal of Computer Science Engineering and Technology. 2017:810–12

[4] Gordana Laštovička-Medin, Itana Bubanja, “Hardware approach of text-to-speech in embedded applications” Kaladharan N, An english text to speech conversion system, International Journal of Advanced Research in Computer Science and Software Engineering, Vol-5, Issue-10, October-2020