Impact Factor (SJIF): 5.301

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444

Volume 6, Issue 5, May-2019

Multilingual Character Recognition And Character Translation For Devnagari Character Over The Image

Miss. Gunjal Harshada Arun

Miss. Rupanavar Divya Bapurao

Miss. Darekar Swamini Shahaji

Prof. Pawar A.H.

Department of Information Technology

S.V.P.M. College of Engineering Malegaon(bk), Baramati

Abstract **—**A Technology used for image translation the user can translate the text on images or pictures taken of printed text (posters, banners, menu list, sign board, document, screenshot etc.). Translation can be done by applying optical character recognition (OCR) technology to an image to extract any text contained in the image, and then have this text translated into a language of their choice, and the applying image processing on the original image to get the translated image with a new language. We primarily focus on images taken with a mobile phone camera. To pre-process images and Translate the character in user selected language. Character translation is the technique which is used for the translate character in the one language to another language like Marathi to Hindi or Hindi to English. The translator is taking the character which is segmented by the character segmentation and then translates it into our specific language. The purpose of this project is to develop such a tool which takes an Image as input and extract characters (alphabets, digits, symbols) from it. This Image can be of handwritten document or Printed document. All users are understanding the contents of an image or the valuable information, there is need of analyzing the text appears in it.

Keywords- Optical Character recognition (OCR), image processing, Translation.

I. INTRODUCTION

Translation is important for people who can travel different places. They can face problem for communication with there local peoples. They don't understand the language or message written on the board. People wish to translate the document in their own specified Indian language. Optical Character Recognition (OCR) is a process by which text characters can be input to a computer by providing the computer with an image.

Optical Character Recognition (OCR) is a process by which text characters can be input to a computer by providing the computer with an image. The computer uses an OCR Engine–a computer program with the specific function of making a guess which letter (recognizable to a computer) an image (recognizable to a human) represents. Message can written on board people are not understood. So that time they have need to character recognition and character translation in the different image.

Character translation is the technique which is used for the translate the character in the one language to another language like Marathi to Hindi or Hindi to English. Translator is take the character which is segmenting by the character segmentation and then translate it in our specific language.

II. EXISTING SYSTEM

In Existing System, robust algorithms for character segmentation and recognition are presented for multilingual Indian document images of Latin and Devanagari scripts. These documents generally suffer from their layout organizations, local skews, and low print quality and contain intermixed texts (machine-printed and handwritten).

1) Google Translate:
   Google Translate is a free multilingual machine translation service developed by Google, to translate text.
   It first translates text to English and then to the target language. Google translate can be used to translate text data in multiple languages. It can also translate text over the image, but only first language used is English and target output is many other languages.
2) Microsoft Translator:
   Microsoft Translator is a free, personal translation app for 60+ languages, to translate text, voice, conversations, camera photos and screenshots. Microsoft Translator only supports 61 languages, one of which is Klingon. Not all features are available for all languages. In this application can translate text in foreign languages.
3) Photron Image Translator:
   Extract and translate text from your images with over 95 percentage accuracy. If you ever gone outside in a foreign country and saw boards, banners or pamphlets in a new language then this app is useful for user. It is an image translation application. To translate text over the image, but it does not translate Devnagari language text over the image.

Disadvantages of Existing System

1. Image text cant not translate.
2. Do not translate Devanagari text over the image.
3. Only used for foreign country.

III. PROPOSED SYSTEM

The country India is a multi-lingual multi-script country and there are twenty-two languages. Eleven scripts are used to write these languages and Marathi, Hindi and English are the most popular script in India. First research report on handwritten characters was published in 1977 but not much research work is done after that. In Proposed system, we are working on handwritten Marathi, Hindi and English characters. Many research reports are available towards character recognition but to the best of our knowledge there are not any researches available about multi lingual character recognition and translation.

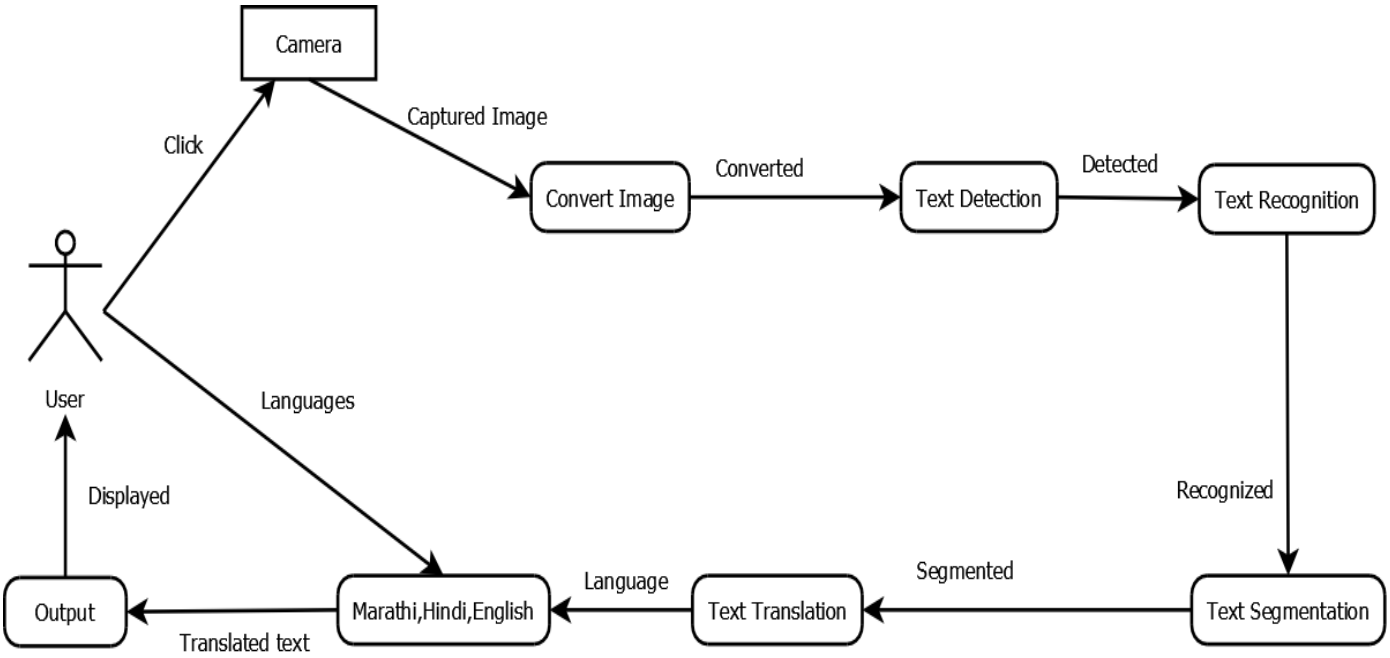
The conversion for the image text to the multiple language that is the people understandable format various person can't understand the advertised image language that's why we convert that text to the multiple language using the translator. Our system is used for every user which can not to understand the image text in this purpose user simply add the image and convert that image text in multiple languages. Admin can add the different data set with multiple language text.

Advantages of Proposed System:

1. Implement translate text over the image from Devnagari languages (Marathi, Hindi) to English or Devnagri languages.
2. To translates written words or text to another language.

3. It can identifies text in a picture taken by the users and translates text on the screen instantly.
4. To understand the content or message in an image.

IV. SYSTEM ARCHITECTURE



In the above figure System Architecture, Shows that user can click on the camera for the capture different images that images on the form of Devanagari Script which is converted into other specific languages (Marathi, Hindi, English).Images can converted into binary scale or gray scale.

System can be take a capture image and then process on that images start from the text detection. After the image conversion we can processed on this character that is the character Detection. Character Detection is the technique which is detect the character over the images.

Character Recognition is a process by which text or characters can be input to a computer by providing the computer with an image. The system uses an OCR Engine--a computer program with the specific function of making a guess which letter (recognizable to a computer) an image (recognizable to a human) represents.

Character Segmentation is the method which is used for the divide the character in multiple form. This method is also used for the character in the number of multiple sub parts for the translation.

Character translation is the technique which is used for the translate the character in the one language to another language like Marathi to Hindi or Hindi to English. Translator is take the character which is segmenting by the character segmentation and then translate it in our specific language.

V. OUTCOME OF THE SYSTEM

Input Image:

आज सोमवार आहे

```
ImageName /home/harshada/Documents/images/4.png
Please select your language eng/ mar/ hin for OCR mar
Language selected is  mar
==============OPTION 1============================
PIL-IMAGE-COLOR
 अ◌ज स◌ोम◌वा◌र आ◌हे
==========================================
================OPTION 2==========================
IMAGE-GRAY
 अ◌ज स◌ोम◌वा◌र आ◌हे
==========================================
=================OPTION 3==========================
PIL-IMAGE-GRAY
 अ◌ज स◌ोम◌वा◌र आ◌हे
==========================================
================OPTION 4==========================
IMAGE-BLUR
 अ◌ज स◌ोम◌वा◌र आ◌हे
==========================================
=================OPTION 5==========================
PIL-IMAGE-BLUR
 अ◌ज स◌ोम◌वा◌र आ◌हे
==========================================
=================OPTION 6==========================
IMAGE-TH
 अ◌ज स◌ोम◌वा◌र आ◌हे
==========================================
=================OPTION 7==========================
PIL-IMAGE-TH
 अ◌ज स◌ोम◌वा◌र आ◌हे
==========================================
Please select your option 1 -7 7
Option selected is  7
Please select your translate language en-(English), mr-(Marathi), hi-(Hindi)en
Source Language: mar Destination language:  en
Translated text
 Today is Monday
```

VI. CONCLUSION

The purpose of this system is to develop such a tool which takes an Image as input and extract characters (alphabets, digits, symbols) from it. This Image can be of handwritten document or Printed document. That image can be used as a form of data entry from printed records. And to translate the recognized character into specified Indian language.

Optical character recognition (OCR) technology converts images of text into movable type. Recommendation engines, powered by machine learning, suggest what movies or television shows to watch next based on user preferences.

VII. REFERENCES

[1] Parul Sahare and Sanjay B. Dhok, Multilingual Character Segmentation and Recognition Schemes for Indian Document Images. DOI 10.1109/ACCESS.2018.2795104, IEEE Access

[2] Rehman, and T. Saba, Performance analysis of character segmentation approach sfor cursive script recognition on benchmark database, Digit. Signal Process., vol. 21, no. 3, pp. 486-490, May 2011.

[3] C.Z. Shi, S. Gao, M.T. Liu, C.Z. Qi, C.H. Wang, and B.H. Xiao, Strokedetector and structure based models for character recognition: A comparative study, IEEE Trans. Image Process., vol. 24, no. 12, pp.49524964, Aug. 2015.

[4] M.K. Sharma, and V.P. Dhaka, Segmentation of English o_ine handwritten cursive scripts using afeedforward neural network, Neural Comput. and Applic., vol. 27, no. 5, pp. 13691379, Jul. 2016.

[5] S. Nomura, K. Yamanaka, O. Katai, H. Kawakami, and T. Shiose, A novel adaptive morphological approach for degraded character image segmentation, Pattern. Recogn., vol. 38, no. 11, pp. 19611975, Nov.2005.

[6] V.N.M. Aradhya, G.H. Kumar, and S. Noushath, Multilingual OCR system for South Indian scripts and English documents: An approach based on Fourier transform and principal component analysis, Eng. Appl. Artif. Intel., vol. 21, no. 4, pp. 658668, Jun. 2008.

[7] K.C. Santosh, and L. Wendling, Character recognition based on non-linear multi-projection pro_les measure, Front. Comput. Sci., vol. 9, no. 5, pp. 678690, Oct. 2015.