✚IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY
GRAPHICAL USER INTERFACE BASED OPTICAL CHARACTER RECOGNITION USING MATLAB

Jyoti Dalal[*1] & Sumiran Daiya[2]
[*1]M.Tech., ECE, Sat Kabir Institute of Technology and Management, SKITM, Haryana
[2]Assistant Professor, ECE, Sat Kabir Institute of Technology and Management, SKITM, Haryana

ABSTRACT
Optical character recognition (OCR) is becoming a powerful tool in the field of Character Recognition, now a days. In the existing globalized environment, OCR can play a vital role in different application fields. Basically, OCR technique converts images into editable format. This technique converts images in the form of documents such as we can edit, modify and store data more safely for longtime. This paper presents basic of OCR technique with its components such as pre-processing, Feature Extraction, Classification, post-processing etc. There are various techniques have been implemented for the recognition of character. This Review also discusses different ideas implemented earlier for recognition of a character. This paper may act as a supportive material for those who wish to know about OCR.

Keywords: OCR, Boundary Detection & noise removal , Feature Extraction.

I.   INTRODUCTION
Now a days, globalization is reaching to a great level. In this globalized environment, character recognition techniques also getting a valuable demand in number of application areas. OCR is an effective technique which converts image into suitable format such that data can be edit, modify and stored. This technique performs several operations such as, scans the input image, processes over the scanned image thereby image gets converted into portable formats .For instance, the hard copy of old historical books, novels, etc. .cannot be stored safely for a long time. Rather, its safety has limitations. If we apply OCR technique for such cases, the different historical documents can be stored, modified for a longtime. OCR also having variety of applications in almost all fields, including security. OCR implementation helps us to edit, store and process over the scanned data more effectively. User can handle the stored data whenever he wants with the internet support. So we use Optical character recognition is most successful application used in pattern recognition. purposed OCR system consists of the following basic components:
- Input Image
- Pre-processing
- Feature Extraction
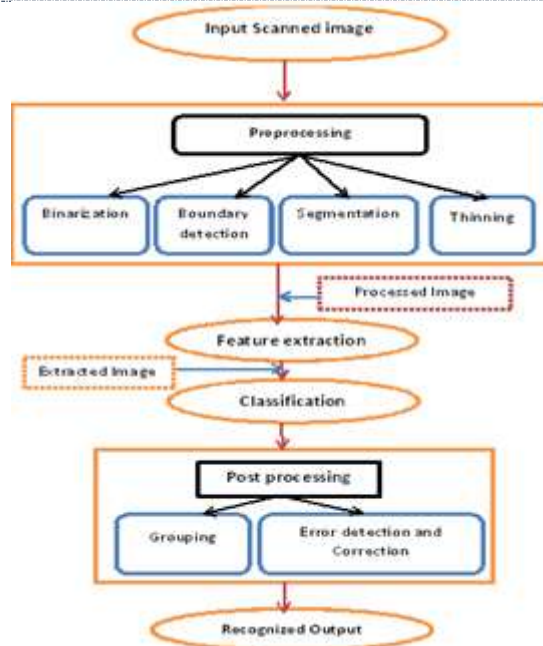- segmentation
- word extraction

Fig- 1.1: Processing Stages of OCR Technique

1.1 Input Image
Firstly, image of input data is optically scanned. The scanned image can be any document of different dimensions. This scanned input image is fed to pre-processing section so as to process over that scanned image.

1.2. Pre-Processing
Pre-processing includes several operations over the scanned image, so that input image becomes suitable and comfortable for applying to further sub sections. Basically the objective of pre-processing is to improve the quality of scanned input image. Noise removal, mathematical operations can also be processed in this Pre-processing section. It includes binarization, boundary detection, segmentation, thinning. It performs the several operations over the scanned input data.

1.3 Binarization
Binarization plays an important role in pre-processing. It is necessary to convert a color image into black and white format. So we can process over that black and white image. Basically separation of background and actual image area referred as foreground of a scanned image is called binarization.

1.4 Boundary Detection & noise removal
The binarized image is now applicable for boundary detection noise removal. In this operation the boundaries of scanned image is detected. It detects all the boundaries of image. It is necessary to detect the boundaries so as to select an individual character.

1.5 Segmentation
This is important operation of OCR as rate of recognition is directly proportional to segmentation. In this process, every individual character is separated. This isolates the different sub-parts of an image. It is used to separate pixels of an image as per the contents in data like words, paragraph etc.

1.6 Feature Extraction
For the accuracy of OCR system, the appropriate Feature Extraction method should be selected. While processing over the image some features should be separated. The typical features are Edges, Corners, Ridges, etc. This method of separation is called as Feature Extraction. The accuracy of an OCR technique depends on selection of proper feature extraction method.

1.7 Classification
The feature extracted data must have gone through the process of Classification. This process classifies the extracted individual character in proper way.

1.8 Post-Processing
This is the last and an important phase of OCR technique. It includes different operations like Grouping, Error detection and correction. Whatever the data being operated through different operations such as, binarization, segmentation, Feature

extraction, Classification etc. is fed to post-processing. That means different features of input scanned image are extracted. That feature extracted data is an individual character. It is unable to get detailed information from that individual character. So, it is necessary to collect individual character in appropriate and sequential manner. The process of collecting individual characters of the same contents to form a string is termed as Grouping. By using error detecting and correcting algorithms, errors can also be eliminated.
Finally, we get the recognized output character.

II.   LITERATURE REVIEW ON OPTICAL CHARACTER RECOGNITION
As per reference, IJETAE Volume 4, Issue 5, May 2016**[1]** this paper explains comparative analysis between Random Transform and Hough Transform, which are applied for error detection and correction. This paper explains implementation of OCR in Matlab, compared with current working method of OCR. This system achieved recognition rate near about 92%.

As per reference, IJSR Publications, Volume 2, Issue 6, June 2015 **[2]** this paper discusses recognition of off-line English character. This explains a new model Hidden Markov Model (HMM) for character recognition. The Novel feature Extraction method is used for implementing HMM. By collecting 13000 samples from 100 writers they have tested performance of OCR technique and got accuracy of near about 94%.

As per reference, IJARECE Volume 2, Issue 5, May 2013 **[3]** this paper implements the OCR technique in Matlab. This paper explains how matlab is more convenient and effective for OCR technique. The performance of OCR has been tested with samples in this approach.

As per reference, European Academic Research, Volume I, Issue 5/ August 2013 **[4]** this paper discusses the OCR technique with its components. This achieved a good recognition rate by implementing Particle Swarm Optimization Approach.
 Optimization (BFO). In this proposal PSO and BFO are used to achieve most advantageous harmonic compensation. This paper also discusses the efficiency of both approaches PSO and BFO by comparing them.
As per reference , IJERA Volume 1, Issue 4, pp. 1736-1739 [7] this paper presents an overview of the various O.C.R. systems for Gurumukhi which are developed for handwritten isolated Gurumukhi text. This paper discusses details of different feature extraction methods with its comparative analysis.

As per reference, International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 1, January 2014[8] this explains OCR technique for both handwritten and printed Guajarati script. For this implementation, linear recognition technique has been used. This paper explains how linear recognition technique is efficient in OCR for error detection and correction.

As per reference, International Journal of Computer Application, volume 23, no.1, pp. 21-24, 2011.[9] This paper not only explain OCR for different font size and style, but also tests the performance of proposed OCR system with four groups of different font size and style. This proposed system achieved recognition rate near about 96%.

III.   RESULT OUTPUT

Finally, we get the recognized output character
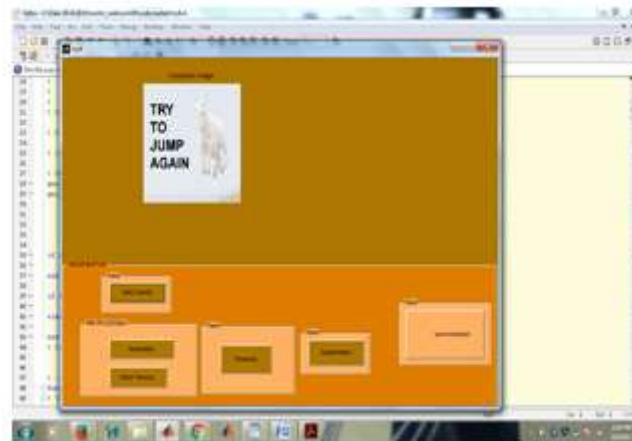

Figure-1.2 Project Main Home Page

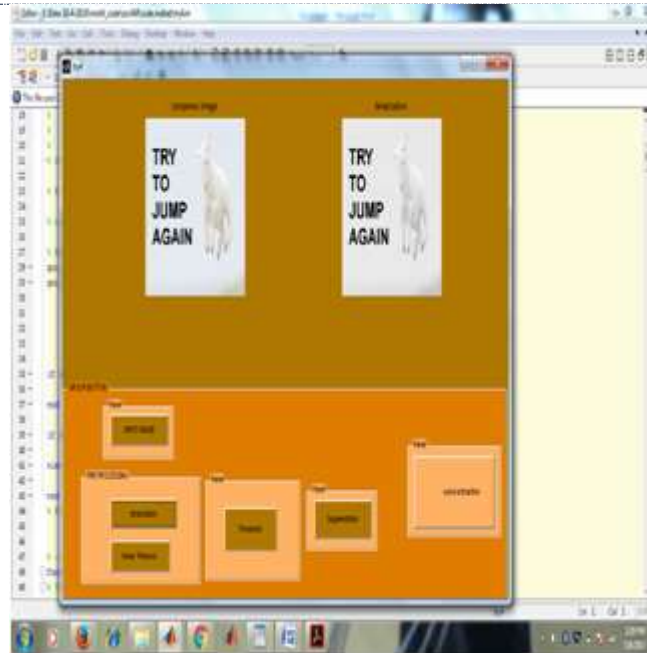*Figure-1.3 Input Image In Gui(Graphical User Interface)*
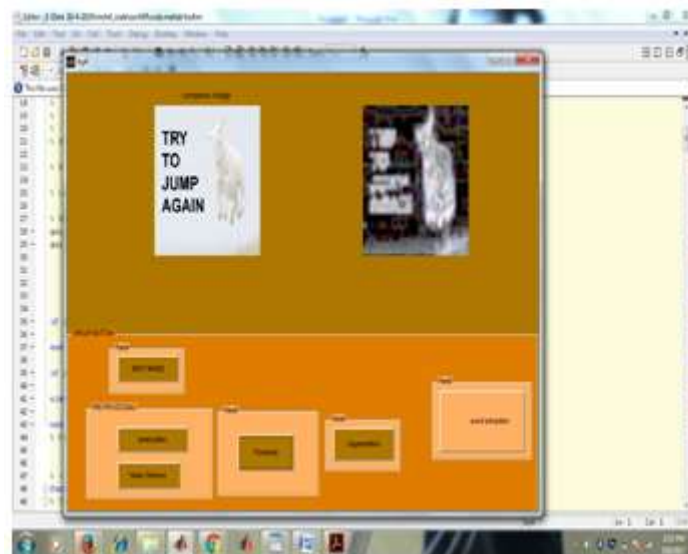
Figure-1.4 Noise Removal Boundary Detection

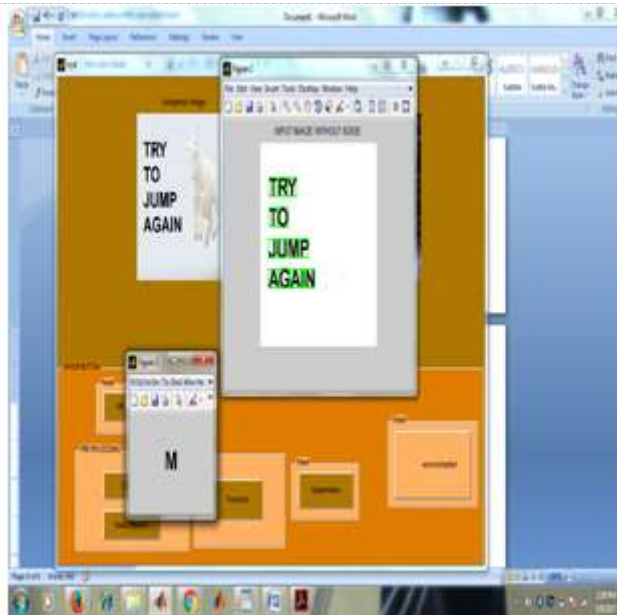*Figure-1.5 images Thresholding*
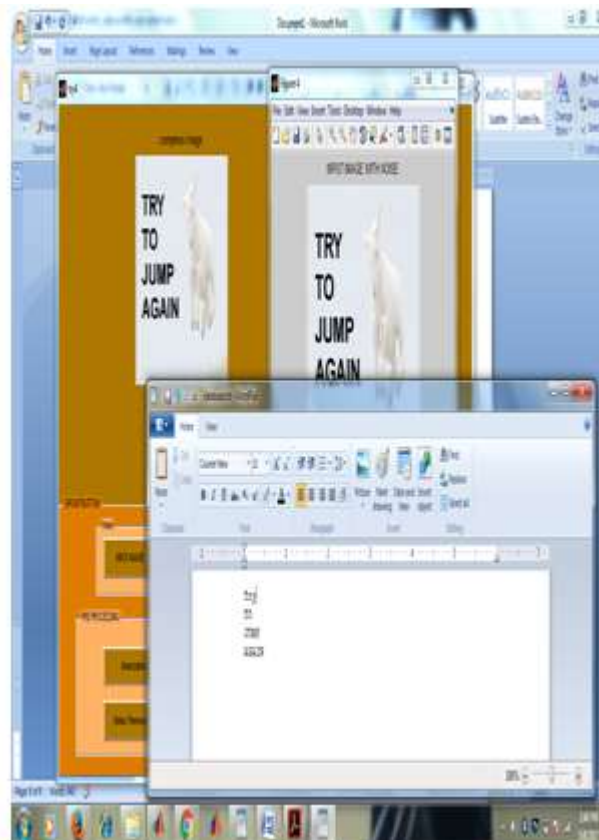
Figure-1.6 Image Segmentation

Figure-1.7 Multiple Charter Recognize

IV.   CONCLUSION

This paper presents nitty gritty portrayal of OCR framework. It incorporates exchange of different sub-parts of OCR method, for example, pre-handling, division, Feature Extraction. The distinctive papers having new calculations and methodologies in order to perceive a character precisely have been examined in this survey.

Every procedure has its own particular uniqueness and level of exactness, yet at the same time a few adjustments must be accomplished for characters of various size and textual styles. A review of highlight extraction and grouping systems for optical character acknowledgment is examined. A great deal of research has been done in this field. Still the work is going ahead to enhance the exactness of highlight extraction and arrangement procedures. Because of algorithmic effortlessness and higher level of adaptability, format coordinating and Correlation technique is anything but difficult to actualize with the change of acknowledgment target classes. Its acknowledgment is most grounded on monotype and distinctive sorts of text styles considering the specimen input pictures for instance written by hand picture and it requires shorter investment and does not require test preparing but rather one format is just fit for perceiving characters of a similar size. The OCR calculation which is actualized in MATLAB (R2010.a/64-bit) gives ideal exactness on a normal as 91.16% when contrasted and existing technique and furthermore the Radon change connected for skew location and remedy gives better outcomes as contrasted and Hough change. These procedures are unique in relation to the others in that no components are really separated. The separation between the example and every model is processed, and the class of the model giving the best match is relegated to the example [8]. The system is straightforward and simple to actualize in equipment and has been utilized as a part of numerous business OCR machines. content lines are practically of same tallness, gave that the script is composed in a particular text dimension. On the off chance that the script is made by a sort machine, without a doubt the text dimension will be uniform all over the place. Between two content lines, there is a limited level band with either no pixel or not very many pixels. Subsequently, by checking break-focuses through them and putting away them will be valuable for recognizing the valleys in it, content line groups can be recovered. It is an operation that tries to break down a picture of arrangement of characters into sub pictures of individual images. Character division is a key necessity that decides the utility of traditional Character Recognition frameworks. It incorporates line, word and character division. Diverse techniques utilized can be grouped in light of the kind of content and system being taken after like acknowledgment based division. The calculation approaches two techniques for Skew recognition and adjustment for the pictures and the correlation for % of precision for two strategies is appeared in table I from which it is watched that Radon Transform gives ideal arrangement with 100% exactness. As a general perspective of the framework model, it could be infer that this framework model has been created by utilizing the strategy that has specified and expounded which is the Template Matching way to deal with perceive the character picture

V. REFERENCES

[1] Jagruti Chandarana, Mayank Kapadia Optical Character Recognition International Journal of Emerging Technology and Advanced Engineering Volume 4, Issue 5, May 2016

[2] Binod Kumar Prasad, GoutamSanyal A model Approach to Off-line English Character Recognition International Journal of Scientific and Research Publications, Volume 2, Issue 6, June 2015

[3] Sandeep Tiwari, Shivangi Mishra, Priyank Bhatia, Praveen Km. Yadav[May 2013] Optical Character Recognition using MATLAB International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)Volume 2, Issue 5, May 2013

[4] Majida Ali Abed Hamid Ali Abed Alasadi Simplifying Handwritten Characters Recognition Using a Particle Swarm Optimization Approach European Academic Research, Volume I, Issue 5/ August 2013

[5] Mahesh Goyani, Harsh Dani, Chahna Dixit Handwritten Character Recognition – A Comprehensive Review International Journal of Research in Computer and Communication Technology, Volume 2, Issue 9, September -2013

[6] Sushree Sangita Patnaik and Anup Kumar Panda Particle Swarm Optimization and Bacterial Foraging Optimization Techniques for Optimal Current Harmonic Mitigation by Employing Active Power Filter Applied Computational Intelligence and Soft Computing Volume 2012, Article ID 897127.

[7] Pritpal Singh, Sumit Budhiraja Feature Extraction and Classification Techniques in O.C.R. Systems for Handwritten Gurumukhi Script – A Survey International Journal of Engineering Research and Applications (IJERA) Volume 1, Issue 4, pp. 1736-1739.

[8] Lipi Shah, Ripal Patel, Shreyal Patel, Jay Maniar Skew Detection and Correction for Gujarati Printed and Handwritten Character using Linear Regression International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 1, January 2014

[9] Ritesh Kapoor, Sonia Gupta, and C.M. Sharma, "Multi-font/size character recognition and document scanning," Int. J. of Computer Application, volume 23, no.1, pp. 21-24, 2011.

CITE AN ARTICLE