Handwritten Character Detection Using Optical Character Recognition Method

Ms. Shubhangee S. Bhagat[1], Ms. Ankita R. Joshi[2], Ms. Vaishnavi S. Gajbhiye[3], Ms. Sweety R. Nandanwar[4], Prof. M.M.Ingle[5]

[1, 2, 3, 4, 5]*Electronics and Telecommunication Department, G. H. Raison I College of Engg. Nagpur, R.T.M.N.U., Nagpur, India.*

Abstract: The handwritten character recognition using MATLAB converts the handwritten notes, printed documents, author books, papers etc into the editable word format. Using OCR(Optical Character Recognition) technique this can be possible with high speed and accuracy within a short interval of time. Now-a-days this technique serves a key factor and create the paperless environment. This paper presents a details of Handwritten Character Recognition.
Keywords : Handwritten Character Recognition, Optical Character Recognition.

I. INTRODUCTION

Character Recognition is the process of detection and recognition of the character. Today, many researchers have developed algorithms to recognize printed as well as handwritten characters. But the problem of interchanging data between human beings and computing machines is a challenging one. In reality, it is very difficult to achieve 100% accuracy. Converting image into editable text requires identifying characters from input image acquired by digitizing paper using scanner and is also known as offline character recognition. Offline handwritten character recognition is the process of finding the letters and words present into the digitized image. In this MATLAB software is used. MATLAB is high performance language for technical computing, it integrates computation and programming in easier way which resolves the problems and the solutions are expressed in mathematical form. It is used in including matrix calculations, developing and running algorithms etc but here we are using this software for creating the GUI(Graphical User Interface) and data visualization. In this paper, the section II gives the description of the basic working principle of character recognition with a detailed literature survey. The other section consist of methodology used and conclusion .

II. LITERATURE REVIEW

Hassiba Nemmour[6] gave stress on handwritten Arabic word recognition. In that they explained two approaches for word recognition such as analytic and holistic approaches. Ridgelet transform is used for feature extraction. Advantage of using Ridgelet is to highlight the line singularities in the handwritten words. SVM is used as the classifier. It's accuracy was 84% efficiency

Bindu S Moni[7] presented Malayalam character recognition based on modified quadratic classifier and directional features. The image was converted in to 72 X 72 by using the size normalization . To decompose the character image in to blocks or zones meshing technique is introduced. Sobel Mask is used for normalization of image which is taken for feature extraction and directional features are obtained by applying sobel mask. Recognition efficiency of 95.32% is obtained.

[2] Nusaibath C in offline handwritten character recognition the input image is acquired by digital camera and scanner. Binarization and skeletonization is used for Preprocessing for correcting the input image. Line segmentation, word segmentation and character segmentation are also used to segment the characters. Gabor filtering is good method for most relevant features identification . Additional features taken are aspect ratio, ratio of horizontal and vertical grid values. Neural network is used for classification. Recognition efficiency obtained is 96.80%.

[4]Pritpal Sigh presented handwritten character recognition for Gurumukhi numerals. Here the stages of character recognition include preprocessing, feature extraction and classification. The filtered image binarised and normalized to 32 X 64. Features are taken as the wavelet coefficients. In addition to this wavelet feature zonal features are also consider here. Recognition rate of 88.8% is obtained.

G Y Chen suggested invariant pattern recognition. Invariant means features must be independent. Features that are invariant under translation, rotation, scaling is obtained by Fourier transform. For palm print classification Contourlet transform extract the features and invariant feature is obtained by taking Fourier of the Contourlet coefficients.

Mamatha H R is the recognition of Kannada numerals. Kannada is the official language of state Karnataka. It is derived from the southern Bramhi lipi. To preprocess the image binarisation and thinning done. To overcome the limitations of wavelet a new

approach is introduced called curvelet transform. So the features are curvelet coefficients and standard deviation is the dimension reduction technique. For the classification of characters KNN used. This paper gives accuracy of 90.5%.

Angshul Majumdar focus on the recognition of Bangali characters. Curvelet coefficients are taken as the feature values. Here two thinned and thickened version of the image is considered. The fundamental concept is that if the character can't recognize with original image it will be recognized with morphologically altered variations. For testing and training five variations of input image is taken and classified using KNN. Overall accuracy of 96.80% is obtained.

III. WORKING PRINCIPLE

The methodology is described as Handwritten recognition is divided into six steps which are as follows:

Image acquisition, pre-processing, segmentation, feature extraction, classification and post processing. The block diagram of the basic character recognition is shown in fig1.
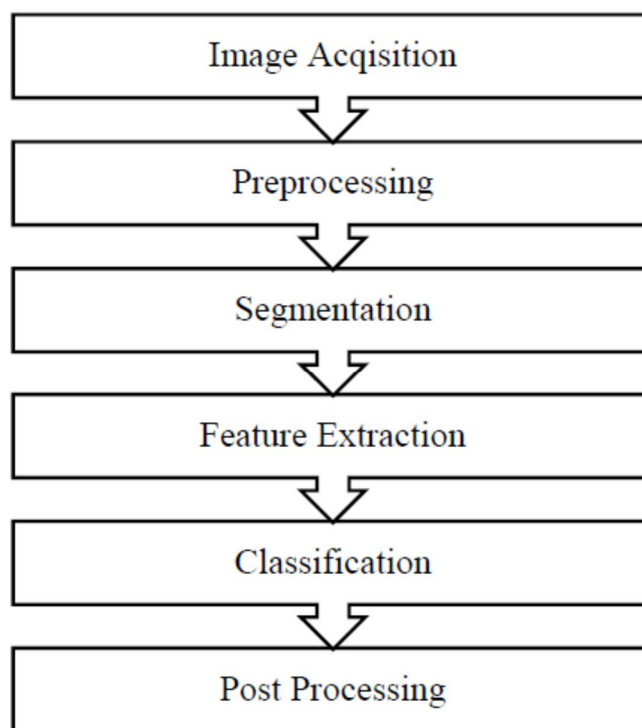

Figure 1: Block Diagram of Character Recognition

off-line handwritten character recognition involves the automatic conversion of text in an image into letter codes which are usable within computer and text processing application. The data obtained by this form is regarded as a static representation of handwritting. Off-line character recognition often involves scanning a form or document written sometime in the past. This means the the individual characters contained in the scanned image will need to be extracted. following is the process for converting handwrtten notes into the text format. Firstly, the image acqisition is carried out in which the image is to be taken for the preprocessing. In preprocessing the pixels of image are scanned for the text conversion and they are determined for if the image is scannable or not. Then next is the segmentation is carried out in which the location of the image is segmented for the feature extraction. the feature exctraction is the important feature in which our data base is compaired with the input image and the features into the data base are scanned. Then the classification of the our data base image and the acqisition image samples are classified. Then these samples are send for the post processing. In the post processing the scanned image is converted into the text format.

IV. CONCLUSION

There is clear need for optical character recognition in order to provide a fast and accurate method to search both existing images as well as large achieves of existing paper documents. However existing character recognition programs suffer from a flawed tradeoff between speed and accuracy, making it less attractive for large quantities of documents. for instance, binarization and tinning reduce each letter to the minimum amount of information necessary to still be able to recognize the letter.
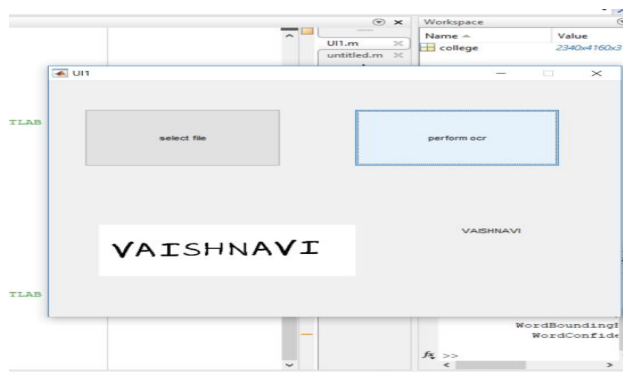
Fig.2. Data base of different A alphabet

Fig.3 Matlab GUI output

REFERENCES

[1] Aji George and Faibin Gafoor, "Contourlet Transform Based Feature Extraction For Handwritten Malayalam Character   Recognition Using Neural Network", IRF International Conference Chennai, pp: 107-110, 2014.

[2] Nusaibath C and Ameera Mol P M,"Off-line Handwritten Malayalam Character Recognition using Gabor Filters", International Journal of Computer Trends and Technology, pp: 2476-2479, 2013

[3] Mamatha H. R, Sucharitha S and Srikanta Murthy K,"Handwritten Kannada Numeral Recognition based on the Curvelets and Standard deviation", International Journal of Processing Systems, pp: 74-78, 2013.

[4] Pritpal Singh and Sumit Budhiraja,"Offline Handwritten Gurmukhi Numeral Recognition using Wavelet Transforms", I. J Modern Education and Computer Science, pp: 34-39, 201

[5] G. Y. Chen and B. Kegl,"Invarient Pattern Recognition using Contourlets and Adaboost", Pattern Recognition Society Elsevier, pp: 1-13, 2012.

[6] Hassiba Nemmour and Youcef Chibani,"Handwritten Arabic Word Recognition based on Ridgelet Transform and support Vector Machines", IEEE, pp: 357-361, 2011.

[7] Bindu S Moni and G Raju,"Modified Quadratic Classifier and Directional Features for Handwritten Malayalam Character Recognition", IJCA Special Issue on"Computer Science-New Dimensions and Perspectives", pp: 30-34, 2011.

[8] Angshul Majumdar,"Bangala Basic Character Recognition Using Digital Curvelet Transform", Journal of Pattern Recognitio Research, pp: 17-26, 2007.

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089   (24*7 Support on Whatsapp)