A Review On Handwritten Character Recognition Techniques

Parul Singla[1], Sargam Munjal[2]
[1]M.Tech(CSE) Scholar, Dept. of CST,
[2]Assistant Professor, Dept. of CST,
Manav Rachna University(Formerly MRCE), Faridabad, India

Abstract- **Pattern Recognition is one of the very important and actively searched trait or branch of artificial intelligence. It is the science that tries to make machines as intelligent as humans to recognize patterns and classify them into desired categories in a simple and reliable way. Nowadays,pattern recognition plays a vital role in hand written character recognition. Different people have different writing styles. So, for recognizing their handwriting, we use pattern recognition system. In this paper, we have described the steps used in the handwritten character recognition system. The steps are image acquisition,preprocessing, segmentation, feature extraction,classification and recognition. The techniques used in the steps of handwritten character recognition system are also described in this paper.**

Index Terms— **Handwritten Character Recognition, Preprocessing, Segmentation, Feature Extraction, Classification.**

I. INTRODUCTION

Pattern Recognition is the association of an observation to past experience or knowledge. With the recent advances in the technology, many pattern recognition tasks have become automated. These include tasks naturally performed by humans, such as speech and handwritten character recognition.

Handwritten Character Recognition is an important area in image processing and pattern recognition field. It is a wide field that covers all sort of character recognition via machine in various application domains. The goal of this area of pattern recognition is to translate human readable characters to machine readable characters. Today, we have automatic character recognizers that help humans in variety of practical and commercial applications. A lot of research work has been done in this area, but still

there is some space in the state of the art. Handwritten characters are non-uniform in nature, as a particular character can be written in different styles and sizes by different writers and even the same writer can write the same character in different styles at different times. Handwritten characters are also vague in nature as there may not be smooth curves or perfectly straight lines all the time.

In this paper, Section2 explains the handwritten character recognition system. Section3 explains the preprocessing and its techniques. Section4 consists segmentation and its types.Section 5 discuss about feature extraction and its techniques. Section 6 is about classification. Section 7 tells the area where handwritten character system is used. Last section concludes the paper..

II. HAND WRITTEN CHARACTER RECOGNITION SYSTEM

Handwritten Character Recognition system consists of many phases such as Preprocessing, Segmentation, Feature Extraction, Classification and Recognition.
The output of one step is the input of next step.
The task of preprocessing relates to the removal of noise and variation in handwritten word patterns. Preprocessing may itself be broken down into smaller tasks such as noise removal, binarization, thinning, edge detection, resizing etc to enhance the quality of images and to correct distortion.

Segmentation is an operation that seeks to decompose an image of a sequence of charactersinto sub images of individual symbols. Its decision, that a pattern isolated from the image is that of a character

(or some other identifiable unit), can be right or wrong. It is wrong sufficiently often to make a major contribution to the error rate of the system.
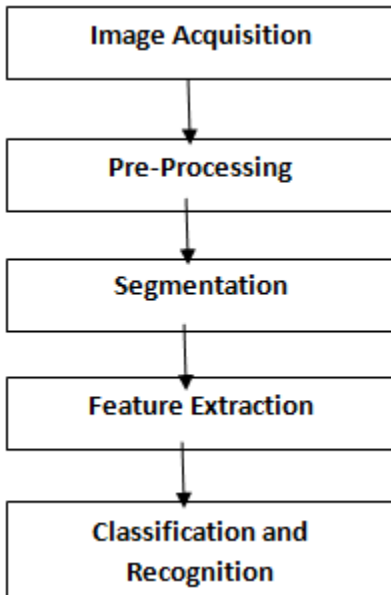

Fig. 1.Block Diagram of handwritten Character Recognition System

Feature Extraction is the process to retrieve the most important data from the raw data. The most important data means that on the basis of that's the character can be represented accurately. The major goal of feature extraction is to extract a set of features, which maximizes the recognition rate with the least amount of elements.

Classification helps compare the rate of accuracy, training, testing time and classification time of new feature extraction technique with some of the existing techniques.

III. PREPROCESSING

Variability that is present in the off line handwritten character is removed by preprocessing.Preprocessing techniques are needed on color, grey-level or binary document images.Such images may also contain non-uniform background and/or watermarks making it difficult to extract the document text from the image without performing some kind of preprocessing, therefore, the desired result from preprocessing is a binary image containing text only.

Thus, to achieve this, several preprocessing techniques are needed.

A. Noise Removal

When the document is scanned, the scanned images might be contaminated by additive noise and these low quality images will affect the next step of document processing. Therefore, a pre processing step is required to improve the quality of images before sending them to subsequent stages of document processing. Due to the noise there can be the disconnected line segment , large gaps between the lines etc. so it is very essential to remove all of these errors so that's the information can be retrieved in the best way.

B. Binarization

Binarization is the process of converting a grey scale image (0 to 255 pixel values) into binary image (0 and 1 pixel values) by selecting a threshold value in between 0 to 1.

C. Edge Detection

Edges characterize object boundaries and are therefore useful for segmentation, registration, and identification of objects. Edge detecting an image significantly reduces the amount of data and filters out useless information, while preserving the important structural properties in an image.

D. Skew Detection and Correction

While scanning the image, if the source document is not aligned properly, it may cause the components to be tilted. So, skew detection and correction method has been used, which detect and remove the skew from the image and later the boundaries of particular images are adjusted so that image looks like an original image.
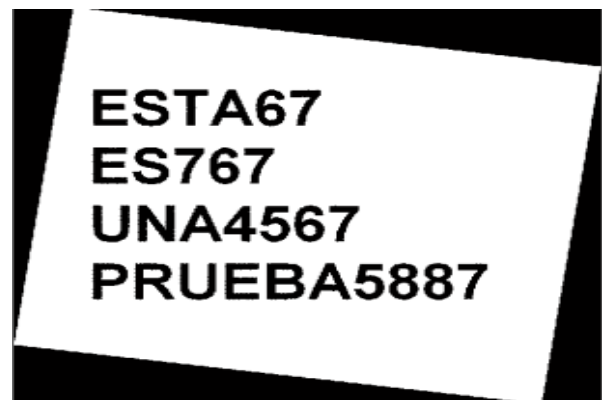

Fig.2.Scanned Image

ESTA67
ES767
UNA4567
PRUEBA5887

Fig.3.Skewed Image

IV.   SEGMENTATION

It is an operation that decomposes an image of sequence of characters into sub images of individual character.

A.   Line Segmentation [2]

In line segmentation, the lines of a text blocks are detected by scanning the input image horizontally. Frequency of black pixels in each row is counted in order to construct a row histogram. When frequency of black pixels in a row is zero, it denotes a boundary between two white pixels consecutive lines.

B.   Character Segmentation

In Character segmentationa word is separated into characters, each individual character and composite character is separated for further identification.

character ⇒ character

Fig.4.Character Segmentation

V.   FEATURE EXTRACTION
VI.

Feature Extraction is the process of getting information about an object or a group of object in order to facilitate classification.
In feature extraction stage each character is represented as a feature vector, which becomes its identity. Due to the nature of handwriting with its high degree of variability and imprecision obtaining these features, is a difficult task.

Steps for Feature Extraction are-

1)Capture the scan character.
2)Perform the normalization process.
3)Perform Binarization.
4) Apply feature extraction technique.
5)Implement neural network classifier.
6)Get the recognition character.

Feature extraction methods falls among these categories. Any method can be a combination of these categories also:

A.   Statistical

It is based on the probability theory and hypothesis. Statistical distribution of pixels of an image takes care of variations in writing styles.

B.   Structural

Structural feature space is extracted such that each value contains some information about structure of the image. Feature values are calculated from the structural and geometrical properties of the character. Examples of structural features are number of horizontal lines or vertical lines, aspect ratio, number of cross points, number of loops, number of branch points, number of strokes, horizontal curves at top or bottom, etc.

C.   Global Transformation and moments

The Fourier Transformation of the outline of the image is extracted.

VII.   CLASSIFICATION[1]

After getting feature space form the binary character image, an efficient classifier is used to classify the class of a character. The most traditional classifier used for handwritten character
Recognition is Neural Network. Other than neural network there are a lot of classifiers for classification problem, for example-Bayesian theory, support vector machine, K-nearest neighbor and artificial Immune system etc.

VIII.   APPLICATIONS

Handwritten Character Recognition Software has simplified the process of extracting data from the handwritten documents and storing it in electronic formats. Banking sectors, Health care industries and many such organizations where handwritten documents are dealt with regularly find Handwritten Character Recognition Software extremely useful.

IX.   CONCLUSION

Immense work and research has been done in handwritten character recognition but so far 100% accuracy has not been achieved which gives scope of further research in this direction. Individual character gives good accuracy but recognition of a word is affected by different handwriting of different individuals.

REFERENCES

[1]Rajbala Tokas1,Aruna Bhadu, A COMPARATIVE ANALYSIS OF FEATURE EXTRACTION TECHNIQUES FOR HANDWRITTEN CHARACTER RECOGNITION, International Journal of Advanced Technology & Engineering Research (IJATER), Volume 2, Issue 4, July 2012.

[2] Er.Naunita,Segmentation of Handwritten Text Document- A Review, *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 3, March 2013.*

[3] Amit Choudhary, Rahul Rishi, Savita Ahlawat, Off-Line Handwritten Character Recognition using Features Extracted from Binarization Technique, 2013 AASRI Conference on Intelligent Systems and Control

[4]JagrutiChandarana, MayankKapadia, Optical Character Recognition, International Journal of Emerging Technology and Advanced Engineering, Volume 4, Issue 5, May 2014.

[5]Jie Liu, Jigui Sun, Shengsheng Wang, Pattern Recognition: An overview, IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.6, June 2006.

[6]Vijay Laxmi Sahu1, Babita Kubde, Offline Handwritten Character Recognition Techniques using Neural Network: A Review, International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064, Volume 2 Issue 1, January 2013.

[7] Yasser Alginahi, Preprocessing Techniques in Character Recognition.