A Review on Geometrical Analysis in Character Recognition

Arya A[1], Anil Kumar A[2]

[1](Computer Science & Engineering, College of Engineering, Karunagappally/ Cusat, India)
[2](Computer Science & Engineering, College of Engineering, Karunagappally/ Cusat, India)

Abstract: *Character Recognition is a prominent area in the field of image processing. Most of the basic concepts in image processing have been introduced in connection with research carried out in character recognition. Character recognition is still an active field of research and till now it has not acquired sufficient recognition efficiency. Character recognition is used in areas like postal address identification, mail sorting, and bank cheque processing and person identification. In this paper, we have provided a review on existing methods for handwritten character recognition based on geometrical properties of the character. The efficiency of the character recognition is depends on the appropriate preprocessing and feature extraction method. Handwritten character recognition is a difficult task because the handwriting is varying from person to person.*
Keywords: *Character Recognition, feature, geometrical, handwritten, preprocessing*

I. Introduction

Character recognition technique helps in recognition of character written in paper documents and converts it into a digital format. It is an active area of research and it continues to be a challenging research topic. Character Recognition can be classified based upon two major criteria: the data acquisition process and the text type. OCR is the most successful application in automatic pattern recognition. Optical Character Recognition can be classified into Printed Character Recognition (PCR) and Handwritten Character Recognition (HCR). The block diagram of character recognition is shown in Figure 1.
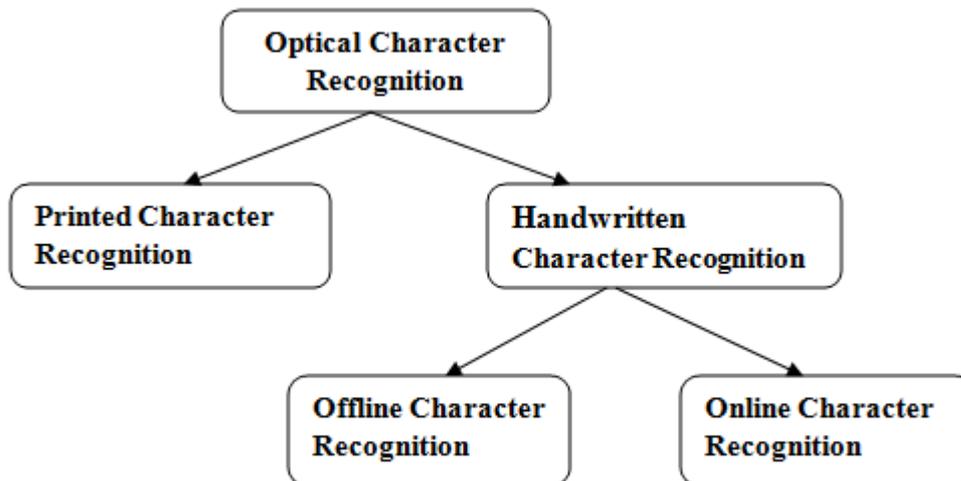
Fig 1: Classification of Character Recognition

In Printed Character Recognition the characters or text are typed from some input devices like any type of keyboard or typewriter etc. These texts are taken from some books, journals, newspapers etc. Handwritten Character Recognition is automation process and can improve the interface between man and machine in a lot of applications. HCR is the ability of computer to receive and interpret intelligible handwritten input from sources such as paper documents. Handwritten Character Recognition is more complicated than machine printed because writing style may vary from one user to another and large amount of noise will be occurs in the handwritten character recognition during the writing of the text and scanning of the document. Based on the data acquisition process HCR is classified into two. Online Character Recognition and offline Character Recognition.

In character recognition system Pre-processing, Segmentation, Feature Extraction and Classification are the different stages.

II.   Phases Of Character Recognition System

Different stages in character recognition are described in Figure 2.

```
┌─────────────────────┐
│  Image Acquisition  │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│    Preprocessing    │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│    Segmentation     │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│  Feature Extraction │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│    Classification   │
└─────────────────────┘
```
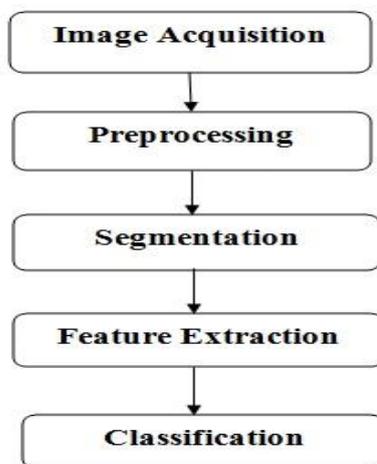
Fig 2: Stages of Character Recognition

A. Image Acquisition:-

The character recognition system acquires an input image through a digital scanner or any other suitable digital input device. The input captured may be in color, gray or binary from scanner or digital camera.

B. Preprocessing:-

Preprocessing is the image Enhancement Technique. Preprocessing is important because it converts the image into a form suitable for subsequent processing and feature extraction. Major preprocessing steps are:

1.   Binarization
2.   Noise Removal
3.   Normalization
4.   Morphological Operations

In character recognition systems most of the applications use gray or binary images since processing color images is computationally high. Image Binarization converts the image into bi-level form. The method used for binarization is known as thresholding. Different thresholding techniques are global thresholding, local thresholding and ostu's thresholding. Noise is a random variation of image Intensity and visible as grains in the image. It may produce at the time of capturing or image transmission. Common manifestation of noise in binary images takes the form of isolated pixels, salt-and-pepper noise. The major objective of noise removal is to remove any unwanted bit-patterns, which do not have any significance in the output. Major noise reduction technique is filtering. Mean filter, median filter and min-max filters are used for noise reduction. Median filter is widely used for the reduction of salt and pepper noise. Normalization changes the range of pixel intensity values. Size Normalization is the process of converting the random sized image into standard sized image. This size normalization avoids inter class variation among characters. Morphological operations which increase or decrease objects in size. Common morphological operations are erosion and dilation. Erosion shrinks the character image and dilation adds the pixels to the character boundary. Skeletonization or thinning is a morphological operation in which a single pixel wide representation of an image is obtained without changing its connectivity.

C. Segmentation:-

Segmentation is an important task of any character recognition system. It separates the image text documents into lines, words and characters. The accuracy of character recognition system mainly depends on the segmentation algorithm being used. First lines are segmented from the document and then each word is segmented from these lines and finally individual characters are extracted.

D. Feature Extraction:
      Feature extraction is the process of extracting the relevant features in the character image. The main aim of the feature extraction is to increase the recognition rate. Efficiency of the character recognition system depends on the feature extraction step. Feature extraction is of three types [11]:
1.   Statistical features.
2.   Global transformation and moments
3.   Structural features.

      Statistical features are based on how the data is collected and selected. It helps to create a hypothesis about the data. Statistical feature extraction is based on probability theory and hypothesis. Partitioning into regular or irregular regions, profiles and projections, distances and crossings are the statistical feature extraction methods. Global transformation and moments includes various techniques like Fourier transform, Gabor transforms, Wavelets, Moments and curvelets. Structural features are based on geometrical and topological properties of the character. Loops, curves, lines, T-point, cross, aspect ratio, strokes and their directions, inflection between two points, chain code etc. are the structural features.

E. Classification:-
      Classification is the final phase of character recognition system. It is the process which to find the class labels of objects whose class label is unknown. This phase uses the feature vectors that we obtained from the feature extraction stage. Classification aims to classify features according to its properties. Training and testing is done at the classification phase. Most commonly seen classifiers are Artificial Neural Network, SVM, K-Nearest Neighbor and Nearest Neighbor classifier. Now a day Genetic Algorithm has been successfully applied for classification. In addition to these classifiers MQDF, Adaboost , Multilayer perceptron and Decision tree are used for the classification in character recognition system.

III.   Literature Review
      Character recognition is an active area of research and many researchers have been done in the field of character recognition. This section provides an extensive review of existing methods of character recognition based on geometrical feature extraction:

      Abdul Rahiman M et.al. [1] proposed a handwritten character recognition system based on vertical and horizontal line positional analyzer algorithm. In this paper the preprocessing technique used is noise removal. Median Filter is used for the reduction of noise. Then for segmentation character separation and line separation are used, which result in isolated character. In this paper feature extraction is based on Horizontal and vertical line count and position. After feature extraction decision tree classifier is used for classification of characters. In this work they mentioned it gives an accuracy of 91%.

      Sumedha B. Hallale et.al. [2] have proposed a Twelve Directional Feature Extraction. In this paper image preprocessing step include noise removing, skeltonization and normalization. The median filter is used for noise removing. Skeleton of the image can be obtained by thinning algorithm. In this paper Size Normalization is used for changing the random sized pixel intensity values to standarized image. Feature Extraction is based on structural features. This paper is based on Directional Feature Extraction. This work done in English Alphabets. In this paper the experimental result shows 82.87% recognition rates. Dayashankar Singh et.al. [3] proposes same method of feature extraction.

      Prachi Mukherji et. al. [4] proposed a shape feature extraction of Handwritten Devanagiri characters. In this paper filtering technique can be applied for noise removal. Noise removal and image smoothing is done using the Gaussian Filter. Then the binarization is achieved using Otsu's thresholding. Then extract the skeleton of the character. In this paper sum of neighbors are used for separating an image in its constituent strokes. This paper uses an average compressed Direction coding algorithm for stroke feature extraction. In this paper [4] classification is based on the decision tree classifier. Advantage of this method is that simple and works efficiently to detect curves and turning points in Devnagari script. This work gives an overall recognition accuracy of 86.4%.

      Parikh Nirav Tushar et. al. [5] proposed a Chain Code Based Handwritten Cursive Character Recognition. In this the work is done in English alphabets. In this paper the preprocessing steps used are Binarization and slant correction. For binarization thresholding is applied. Here Global thresholding is used. In feature Extraction stage chain codes are constructed and which form the feature of the character. In this paper [5] Artificial Neural Networks are used for classification. This method gives recognition of 80%.

      Ved Prakash Agnihotri [6] proposed an Offline Handwritten Devanagiri script recognition. In which the preprocessing steps include binarization and noise reduction. In this paper ostu's thresholding is used for binarization. Filtering technique is used noise reduction. In this paper no specific filtering technique is

explained, suitable filtering is used. In feature extraction stage features are extracted by zone based approach. In this work for classification they use genetic algorithm. This system gives a recognition accuracy of 85.78%

E. K. Vellingiriraj[7] proposed Recognition of Ancient Tamil Handwritten Characters in Palm manuscripts. The preprocessing stage consists of size normalization, binarization image thicken. For binarization thresholding is used. Dilation can be applied to increase the thickness of the Tamil character. The feature extraction module extracts the basic components of Tamil characters, such as loops, straight lines, zigzag lines and the position of the connection between the loop and the straight line. For classification genetic algorithm is used.

J.Pradeep et.al. [8] proposed a Recognition of English offline handwritten character using Neural Network. Noise Reduction is done using median filter and binarization is using Otsu's global technique. Sobel filter can be applied for the detection of edges. Dilation and filling are the other preprocessing methods. Diagonal feature extraction is done in for feature extraction step. Character image is divided into 54 equal zones, 54 features are obtained from the character image. For classification, feed forward back propagation neural network is used. Compared to the other methods of feature extraction the diagonal features provide good recognition rate. This system yields a recognition rate of 98%.

S.V. Rajashekararadhya et.al. [9] proposed an Efficient Zone Based Feature Extraction Algorithm For Handwritten Numeral Recognition. In this paper the preprocessing techniques used are normalization and thinning. Thinning gives the skeleton of the character image. Zone based method can be applied for feature extraction. Three algorithms were used in this work. These algorithms gives the zone based distance metric feature extraction, zone centroid and zone based distance metric feature extraction and hybrid feature extraction. Feed Forward Back propagation Neural Network can be used for the classification stage. This system gives a recognition accuracy of 97.5%

Amritha Sampath et.al.[10] proposed Freeman Code Based Online Handwritten Character Recognition For Malayalam. The preprocessing stage consist of noise reduction and thresholding is used for the noise reduction. In feature extraction stage low level and high level features of the character is extracted. Low level features consist of width, height, aspect ratio etc of the character and the high level features are number and position of loops, straight lines, headlines, curves etc. The directional feature is extracted using freeman chain code. For classification Artificial Neural Network is used.

Comparisons between these literatures are summarized in the Table1.

IV. Conclusion

A detailed study on various aspects of character recognition has been carried out by laying more emphasis on geometrical based analysis. Geometrical based analysis seems to be difficult compared to other methods of character recognition. It has also been analyzed that the recognition efficiency of the geometrical based image analysis depends upon the appropriate selection of preprocessing methods .Improvement can be made in the existing geometrical character recognition techniques by finding new techniques for preprocessing.

Table 1: Comparison of Geometrical Techniques in CR

Author	Preprocessing	Segmentation	Feature Extraction	Classification	Recognition Accuracy
Abdul Rahiman M[1]	Noise Removal	Line & Character separation	Horizontal and Vertical Line count and position	Decision Tree	91%
Sumedha B. Hallale[2]	Noise Removal, Skeltonization, Normalization.	-	Directional Feature Extraction	ANN	82.87%
Dayashankar Singh[3]	Noise Removal, Skeltonization,	-	Directional Feature Extraction	ANN	-
Pranchi Mukherji [4]	Noise Removal, Binarization, Skeletonization.	-	Average Compressed Direction Coding Algorithm	Decision Tree	86.4%
Parikh Nirav Tushar[5]	Binarization Noise Removal, Skeletonization.	-	Freeman code	ANN	80%
Ved Prakash Agnihotri [6]	Binarization , Morphological operation	-	Zone based- Diagonal Feature	Genetic Algorithm	85.78%

E. K. Vellingiriraj[7]	Binarization, Morphological Operations	-	Stroke Detection , Loop Detection	Genetic Algorithm	-
J.Pradeep[8]	Noise Removal, Binarization, Edge detection, Dilation and filling.	-	Diagonal feature	Feed Forward Back propagation Neural Network.	98%
S.V.Rajashekararadhya[9]	Normalization, Thinning.	-	Image centroid zone, Zone centroid zone, Hybrid centroid zone	Feed Forward Back propagation Neural Network	97.5%
Amritha Sampath[10]	Noise Removal,	-	Freeman chain code	ANN	-

References

[1].   Abdul Rahiman M and M.S. Rajasree, Recognition of Handwritten Malayalam Character using Vertical and Horizontal Line Positional Analyzer Algorithm, 3rd International Conference onMachine Learning and Computing, 2013.

[2].   Sumedha B. Hallale and Geeta .D. Salunke, Twelve Directional Feature Extraction for Handwritten English Character Recognition, International Journal of Recent Technology and Engineering, vol. 2, May.2013

[3].   Dayashankar Singh, Sanjay Kr. Singh and Maitreyee Dutta, Handwritten Character Recognition Using Twelve Directional Feature Input and Neural Network, International Journal of Computer Applications, vol. 1,no. 4, 2010.

[4].   Prachi Mukherji and Priti .P. Rege, Shape Feature and Fuzzy Logic Based Offine Devnagari Handwritten Optical Character Recognition, Journal of Pattern Recognition Research 4, June. 2009.

[5].   Parikh Nirav Tushar and Saurabh Upadhyay, Chain Code Based Handwritten Cursive Character Recognition System with Better Segmentation Using Neural Network, International Journal of Computational Engineering Research, vol.3, May. 2013.

[6].   Ved Prakash Agnihotri, Offline Handwritten Devanagari Script Recognition , I.J.Information Technology and Computer Science, pp. 37-48, July. 2012.

[7].   E. K. Vellingiriraj and P. Balasubramanie, Recognition of Ancient Tamil Handwritten Characters in Palm Manuscripts Using Genetic Algorithm, International Journal of Scientific Engineering and Technology, vol. 2, pp. 342-346, May. 2013.

[8].   J.Pradeep, E.Srinivasan and S.Himavathi, Diagonal Feature Extraction Based Handwritten Character System Using Neural Network, International Journal of Computer Applications (0975-8887), vol.8, no.9, pp.17-22, October 2010.

[9].   S.V. Rajashekararadhya and P. Vanaja Ranjan, Efficient Zone Based Feature Extraction Algorithm for Handwritten Numeral Recognition of Four Popular South Indian Scripts, Journal of Theoretical and Applied Information Technology, pp. 1171-1181, 2005 - 2008.

[10].   Amritha Sampath, Tripti.C and Govindaru.V, Freeman Code Based Online Handwritten Techiniques for Handwritten Character Recognition for Malayalam using Backpropagation Neural Networks, ACIJ, vol.3, no.4, Jul.2012.

[11].   Rajbala Tokas and Aruna Bhadu2, A Comparitive Analysis of Feature Extraction Techniques for Handwritten Character Recognition, International Journal of Advanced Technology & Engineering Research, vol .2, July .2012