International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 5 Issue: 5 1340 – 1343

A Survey on Feature Extraction Techniques for Handwritten Character Recognition

[1]Rama Gaur
[2]Dr. V.S. Chouhan
[1]Ph.D. Scholar (ECE)
Jodhpur National University, Jodhpur, Rajasthan, India
[2]Professor and Head ECE department
MBM Engineering college, Jodhpur, Rajasthan, India

Abstract: - The efficiency and accuracy of the optical character recognition system is very much dependent on the appropriate choice of the feature extraction techniques used. In the field of pattern recognition hand written character recognition is an interesting field of research. If the features calculated or observed for a character are enough and complete to specify it then classification errors can be minimized. So feature extraction plays a vital role in the handwritten character recognition. This paper describes about the different techniques of feature extraction.

*****

I. INTRODUCTION

OCR will improve the communication interface between man and machine. It is able to convert machine printed or hand written document into editable text format. Major Steps in an OCR System are described in fig1.1.

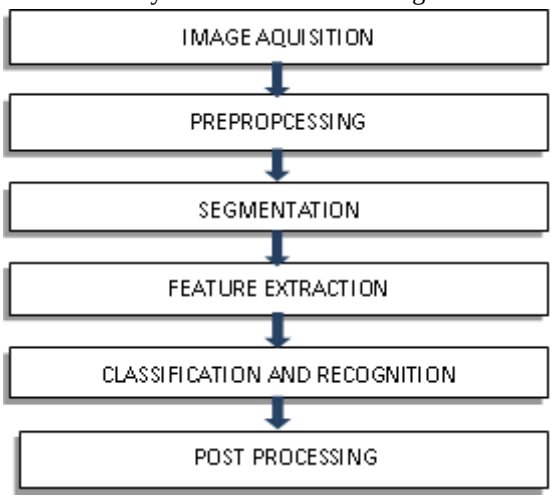

Figure 1: OCR System

II. LITERATURE SURVEY

This paper [1] is about the character recognition techniques used and also brief about type of character recognition technique i.e. online/offline. Further classification of online and offline methods also explained in this paper. In paper [2]it explains the techniques used in the OCR, mainly it is divided as online character recognition and offline character recognition. The offline character recognition technique like the magnetic character recognition and optical character recognition are used. Optical character recognition can be for a handwritten or a printed document for the recognition.

In paper [3], the paper details about the handwritten character recognition and algorithms for the pre-processing, and performance with the practical systems.

Feature Extraction Techniques

Feature extraction is done after the preprocessing phase in character recognition system. The primary task of pattern recognition is to take an input pattern and correctly assign it as one of the possible output classes. This process can be divided into two general stages: Feature selection and Classification. Feature selection iscritical to the whole process since the classifier will not
be able to recognize from poorly selected features.Criteria to choose features given by Lippmanare:"Features should contain information required to distinguish between classes, be insensitive to irrelevantvariability in the input, and also be limited in number, to permit, efficient computation of discriminant functions and to limit the amount of training data required

Feature extraction techniques are classified into three major groups as [4].

Statistical features.
Global transformation and series expansion
Geometric and topological feature

Statistical feature 1.

Projection Method

In the projection method it compares data through a projection. Black pixel counts are taken along parallel lines through the image area to generate distributions. The direction of projection will be horizontal axis, vertical axis,

 ISSN: 2321-8169
1340 – 1343

diagonal axis or all of the above. The character can be divided vertically and horizontally in equal parts and then vertical and horizontal projection found and used as features to recognize the character [1]. Fig. 2 shows horizontal and vertical projection.

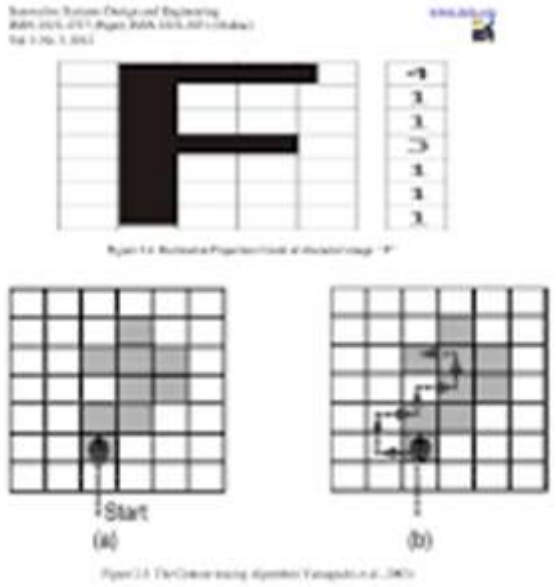

Fig. 2 Projection method

2. Border Transition Technique (BTT)

In border transition technique it assumes that all the characters are oriented vertically. Each character is divided into four equal quadrants. The scanning and calculation of zero-to-one transition in both vertical and horizontal directions in each division take place.

3. Zoning

Zoning widely usedmethod in character recognition tasks. In this method, the character images are divided into zones of predefined sizes and then features are computed for each of these zones. Zoning obtains local characteristics of an image. Here, we have divided the preprocessed character images into 16 zones (4 × 4) as in and then pixel density features were computed for each of the zones (Fig. 3). The average pixel density was calculated by dividing the number of foreground pixels by the total number of pixels in each zone i.

D(i)= Number of foreground pixels in zone i / Total number of pixels in zone i

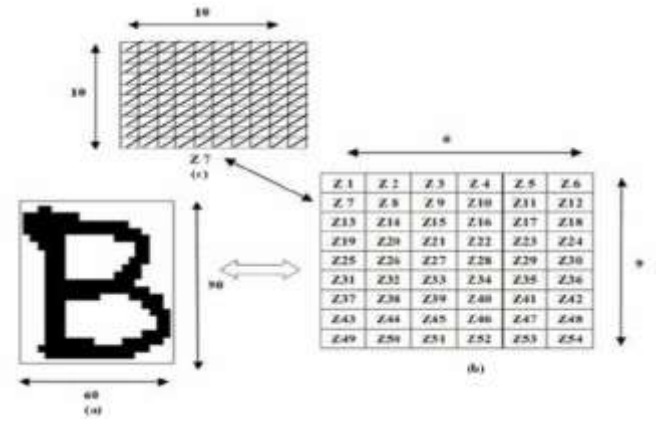

Fig. 3 zoning

4. Graph Matching Method

In a graph matching method it uses structural feature of character. It is useful method to change of font or rotation. In this three features are defined. Here first, an end point is connected only one pixel which has information of position. Than a branch point is connected more than three pixels having feature information which is connected the branch point. The information includes of features like, position and direction A curve point is connected two pixels. However a straight line is also connected two pixels. In order to distinguish between a curve point and a straight line, direction information is used.

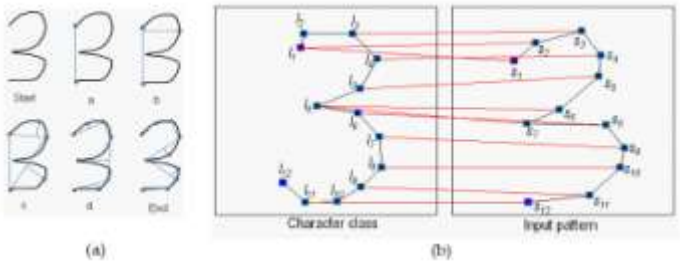

Fig. 4 Graph matching method

Global transformation and series expansion In this various techniques are:

1) Fourier transforms
2) Gabor transforms
3) Fourier Descriptor
4) Wavelets
5) Moments
6) Karhunen-Loeve expansion etc.

1) Fourier Descriptors

Shape feature vector consists of the Fourier descriptors. After boundary the pixel set of an object was computed. In fourier descriptor technique it uses centroid distance to decide shape signature from boundary. Here this centroid distance function is the periodic function we

consider and decompose into fourier series. Fourier transform is used for shape to determine Fourier coefficients, and pixel brightness will find out into computational process of the Fourier coefficients so that shape features can be computed. The Fourier coefficients that we have find are invariant to translation, scaling, rotation and change of start point are used as Fourier descriptors [5].

Note: Fourier series means decomposing a periodic function into sum of set of sine and cosine functions.

2) Wavelets

In the wavelets Transform it represents a mathematical way used to study non-stationary signals. Therefore, its usefulness has been increasingly adapted over the last 10 years. It was employed in different fields such as communication technology, geophysics and image processing. The wavelet transform provides an appropriate basis for image handling because of its useful features. The assets of the wavelet transform are: The ability to compact most of the signal's energy into a few transformation coefficients, which is called "energy compaction" and the capability to acquire and characterizeefficiently low frequency components (image Backgrounds) and also high frequency transients (image edges). The features are extracted fromWavelet transform coefficients. Neural network classifies feature vectors for analysis and processing. wavelet analysis helps in, obtaining matrix of data, where time and frequency domain information is present. Another waveform is compressed or stretched to obtain wavelets of different scales that are used along time comparing them with the original signal [6].

3) Moments

It is an effective technique to estimate shape of characters. As is known that this feature extraction technique work better if operations like normalization of character size and geometric operations are performed correctly by floating point arithmetic. The features drawn by invariants moment technique are used to calculate seven different parameter of a character. Moment invariants remain invariant under scaling,translation,reflection, rotation. These are the measures of the pixel distribution around the center of gravity of the character. The orthogonal moments are the very frequently used in many applications, because of their property of orthogonality [7].

Geometric and topological features

This method extracts the geometric features of the character. Those features are based on the basic line types that form the character skeletons. This system gives a feature vector as its output. The various steps involved in geometric method are:

i) In preprocessing (binarization, skeletonization) is done on the input image.

ii) The features extracted from the image of character includes the positions of various line segments in the character images.

iii) After the step (ii) the image will be divided into equal size of windows, and the feature is done on individual windows.

iv) To extracting the different segments of line in a particular zone, the entire skeleton in that zone will be traversed. So that fixed pixels in the character skeleton were defined as starters, intersections.

v) After that the line of each segment is determined, based on this information feature vector is formed and each of the zone has a feature vector corresponding to it.

The contents of each zone feature vector are

- No. of horizontal lines
- No. of vertical lines
- No. of Right diagonal lines
- No. of Left diagonal lines
- Normalized Length of all horizontal lines.
- Normalized Length of all vertical lines.
- Normalized Length of right all diagonal lines

III. CONCLUSION

This paper represents a review of the feature extraction techniques available for the optical character recognition. As the efficiency is based on the technique we use for the extraction of the features all the techniques will give different attributes of the character. This study will be help full to the researchers to make the choice of feature extraction technique for character recognition.

Reference

[1] Suruchi G. Dedgaonkar, Anjali A. Chandavale, Ashok M. Sapkal , "Survey of Methods for Character Recognition",International Journal of Engineering and Innovative Technology (IJEIT) Volume 1, Issue 5, May 2012.

[2] Er. Neetu Bhatia, "Optical Character Recognition Techniques: A Review"International Journal of Advanced Research in Computer Science and Software Engineering ,Volume 4, Issue 5, May 2014 ISSN: 2277 128X.

[3] ReÂjeanPlamondon, "On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey" Fellow, IEEE, and Sargur N. Srihari, Fellow, IEEE, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, Vol. 22, no. 1, January 2000

[4] AshlinDeepa R.N, R.RajeswaraRao,"Feature Extraction Techniques for Recognition of Malayalam Handwritten Characters: Review"International Journal of Advanced Trends in Computer Science and Engineering, Vol. 3, No.1, Pages: 481– 485 (2014)

[5] M. Bokser, \Omni document technologies," Proceedings of the IEEE, vol. 80, pp. 1066-1078, July 1992

[6] Gang Zhang ; Coll. of Inf. Sci. & Eng., Northeastern Univ., Shenyang ; Ma, Z.M. ; Qiang Tong ; Ying, " The shapefeature extraction usingfourierdesciptors with Brightness", presented at the, IIHMSP '08 International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2008, Page(s):71 - 74 Print ISBN:978-0-7695-3278-3

[7] R. J. Ramteke, "Invariant Moments Based Feature Extraction for Handwritten Devanagari Vowels Recognition", 2010 InternationalJournal of Computer Applications (0975 - 8887) Volume 1 – No. 18.

.