ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (1135-1139), Month: April - June 2015, Available at: www.researchpublish.com

Segmentation of Kannada Handwritten Characters and Recognition Using Twelve Directional Feature Extraction Techniques

[1] Lohitha B.J, [2]Y.C Kiran

[1]M.Tech. Student Dept. of ISE, Dayananda Sagar College of Engineering, Bangalore
[2]Research Scholar, CSE Dayananda Sagar College of Engineering, Jain University, Bangalore, Associate Professor, Dept. of ISE Dayananda Sagar College of Engineering, Bangalore

Abstract: **The OCR system provides an efficient way to translate human readable characters to machine readable characters. It is to identify and analyze a document image by dividing the page into line, words, and then characters. Handwriting character recognition is a challenging and interesting task in the field of pattern recognition. In this paper we are segmenting the Kannada handwritten characters. For the feature extraction we are using twelve directional feature extraction techniques and for the recognition back propagation neural network is used. Experimental results show the good recognition rate towards segmented handwritten Kannada characters.**

Keywords: **Handwritten Kannada Character Recognition Feed forward neural network, Recognition accuracy rate.**

1. INTRODUCTION

Nowadays, recognition systems are used in many fields that have different nature. The optical character recognition (OCR) was started from the recognition of machine printed digits and characters and then it was developed to the recognition of machine printed words. Gradually, handwritten digit, character and word recognition were introduced into this domain. Several research works have been focusing towards evolving the newer techniques that would reduce the pre-processing time and to provide higher recognition accuracy. The applications of OCR is in bank cheque processing, form data entry, vehicle plate recognition, postal address block detection and recognition, Automatic insurance documents key information extraction, Extracting business card information into a contact list.

The handwritten character recognition system is classified into two types. They are off-line system and on-line system. In off-line system the writing is usually captured optically by scanner and it is available as image. It involves automatic conversion of text in an image into letters code that is usable within a computer. The probability of recognizing handwriting recorded with digitizer as a time sequence of pen co-ordinates is known as on-line character recognition. The handwritten character recognition system has following steps

1. Image acquisition
2. Segmentation
3. Pre-processing
4. Feature extraction
5. Classification

In this paper the features of directions of pixels of the characters with admiration to their neighbouring pixels are extracted with the help of gradient values. The direction has been divided into 12 regions with each region covering angle of 30 degree, hence direction value of any pixel may have only 12 values assigned from 1 to 12. This approach increases the information content and gives better recognition rate with reduced recognition time.

2. LITERATURE REVIEW

M. Amrouch, Y. es-saady et al [1] proposed a global approach to the recognition of handwritten Amazigh characters using directional features and Hidden Markov Model. Feature vector is extracted from an image of a character using sliding window technique based on Hough transform and translated into sequence of observations. Finally the class of the character is determined by the classifier.

K.S Prasanna Kumar [3] introduces a novel way of feature extraction for Optical Character Recognition (OCR) customized for Kannada characters. The algorithm described here relies on breaking the character into four equal parts and using one of the quarters for extraction. The algorithm is deliberately kept away from all the complexities and the number of features to be extracted is also minimized so as to increase the efficiency and speed of recognition. The algorithm also describes a conflict resolution technique helpful in effectively utilizing the algorithm.

Bindu S. Moni, G. Raju [2] implemented a fixed meshing for the off-line recognition of handwritten isolated Malayalam characters. The 12-directional features are extracted to form a feature vector. Classification has been carried out by implementing the Quadratic discriminant Function and Modified Quadratic discriminant Function.

Cheng-Lin Liu [4] represented selected feature extraction methods in off-line handwritten Chinese character recognition. The experimental results showed, among the 8-directional, 12-directional and 16-directional feature extraction techniques, 12-directional method has better trade off between accuracy and complexity.

3. PRE-PROCESSING

The raw data of handwritten characters no matter how it is acquired will be subjected to a number of preprocessing steps to make it useable. First convert image input to gray scale image and then convert gray scale image to binary image. The preprocessing phase aims to extract the relevant textual parts and prepares them for segmentation and recognition. Noise presented in the image is removed by using filtering and morphological operations. Normalization is used to converts the random character into the standard or specific size.

4. SEGMENTATION

Segmentation is an important task of any Optical Character Recognition (OCR) system. It separates the image text documents into lines, words and characters. In this paper segmentation stage consists of two stages. In the first stage, mathematical morphology technique is used for constructing bridge between the components .In the next stage, projection technique is proposed for the segmentation of the text into line, words and characters.

4.1 Morphology:

Initially, all the connected components in a document image are detected and removed from the binary image using connected component analysis algorithm. For a component, if the number of on pixels is very small compared to a preset threshold then we remove that component. After this process, the proposed method uses morphology operation that is by using appropriate size of structure element, erosion and dilation will be applied to the binary image. In erosion the last zero value pixel present at the boundary of the image is converted into 1 and in dilation last one value pixel present at the boundary is converted to zero. After dilation, the dilated image is inverted and then the content present in the image is cropped by identifying the rows. The rows are identified by finding the minimum and maximum positions of the zero valued pixels.

4.2 Projection technique:

After the completion of first stage, the next stage is to extract individual character present in the image. In order to extract individual character, a technique based on projection is used. A projection profile is a histogram giving the number of ON pixels accumulated along parallel lines. A vertical projection profile is a one-dimensional array where each element

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (1135-1139), Month: April - June 2015, Available at: www.researchpublish.com

denotes the number of ON pixels along a column in the image. The horizontal projection profile is a one-dimensional array where each element denotes the number of ON pixels along a row in the image. To segment each text word into several characters, we use the valleys of the vertical projection of each text word obtained by computing the column-wise sum of black pixels. The position between two consecutive vertical projections where the histogram height is least denotes one boundary line. Using these boundary lines, every text word is segmented into several characters.

5. FEATURE EXTRACTION

This is also called as data extraction & gives data from perspective areas. Features are a set of numbers that capture the salient characteristics of the segmented image. The statistical features are derived from the statistical distributions of pixels, such as zoning, moments, projection histograms or direction histograms. Structural features are based on the topological and geometrical properties of the character, like strokes and their directions, end-points or intersection of segments and loops. Mathematically, the gradient of a two-variable function, here the image intensity function is at each image point, a 2D vector with the components given by the derivatives in the horizontal and vertical directions. At each image point, the gradient vector points in the direction of largest possible intensity increase, and the length of the gradient vector corresponds to the rate of change in that direction. This implies that the result of the Sobel operator at an image point which is in a region of constant image intensity is a zero vector and at a point on an edge is a vector which points across the edge, from darker to brighter values. To extract gradient feature, 3×3 Sobel operators are used. It uses two templates to compute the gradient components in horizontal and vertical directions, respectively.

-1	0	+1
-2	0	+2
-1	0	+1

Horizontal

+1	+2	+1
0	0	0
-1	-2	-1

Vertical

The two gradient components at location (i, j) are calculated by:

$$g_v(i,j) = f(i-1,j+1) + 2f(i,j+1) + f(i+1,j+1) - f(i-1,j-1) - 2f(i,j-1) - f(i+1,j-1)$$

$$g_h(i,j) = f(i-1,j-1) + 2f(i-1,j) + f(i-1,j+1) - f(i+1,j-1) - 2f(i+1,j) - f(i+1,j+1)$$

Grad = arctan $[g_h(i,j)/g_v(i,j)]$

After computing the gradient of each pixel of the character, the gradient values are mapped onto 12 direction values to the angle span of 30 degree between any two adjacent direction values. The orientations of these 12 directional values are shown in Figure 1.

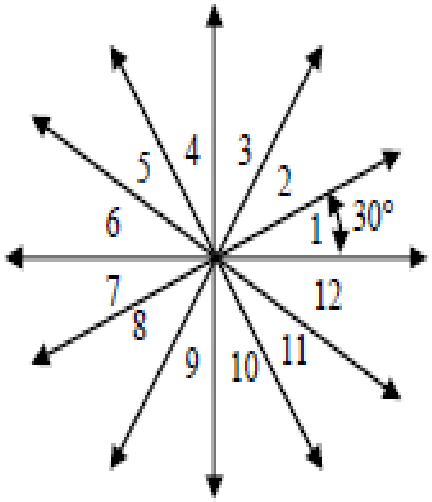


Figure 1: Orientation of 12 directions

Gradient values	Direction
g = -1	0
$0 \leq g \leq \Pi/6$	1
$\Pi/6 < g \leq \Pi/3$	2
$\Pi/3 < g \leq \Pi/2$	3
$\Pi/2 < g \leq 2\Pi/3$	4
$2\Pi/3 < g \leq 5\Pi/6$	5
$5\Pi/6 < g \leq \Pi$	6
$\Pi < g \leq 7\Pi/6$	7
$7\Pi/6 < g \leq 4\Pi/3$	8
$4\Pi/3 < g \leq 3\Pi/2$	9
$3\Pi/2 < g \leq 5\Pi/3$	10
$5\Pi/3 < g \leq 11\Pi/6$	11
$11\Pi/6 < g \leq 2\Pi$	12

Figure 2: mapping of gradients on 12 directions

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (1135-1139), Month: April - June 2015, Available at: www.researchpublish.com

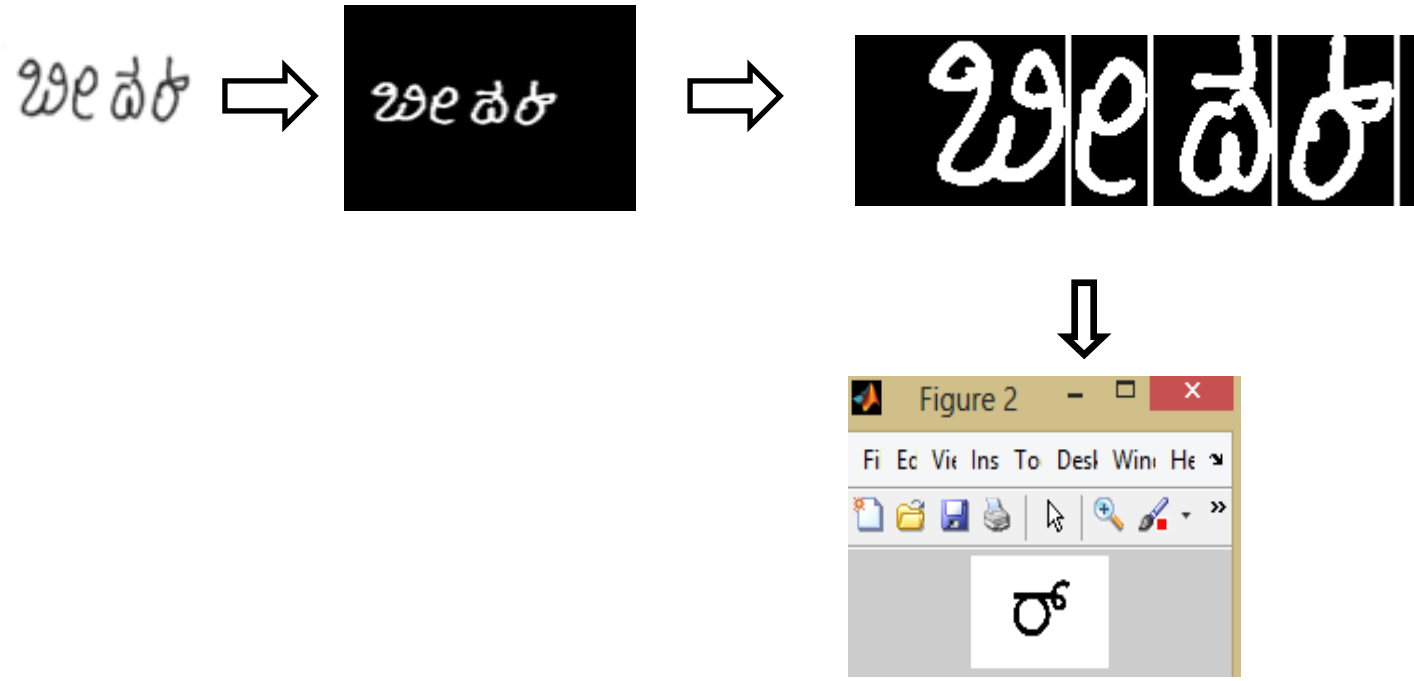
To calculate directional distribution values of background pixels for each foreground pixel, we have used the masks for each direction. The pixel at centre is foreground pixel under consideration to calculate directional distribution values of background. The weight for each direction is computed by using specific mask in particular direction. The mapping of gradient values on 12 directional values can be given as in Figure 2.

6. CLASSIFICATION

A feed forward back propagation neural network is used in this work for classifying and recognizing the Kannada handwritten characters. The neural classifier consists of a hidden layer besides an input layer and an output layer. The hidden layer use tansig activation function and the output layer is a competitive layer as one of the characters is required to be identified at any point in time. Once the neural network has been trained successfully it is then required to identify the same corresponding character. In addition, the network should also be able to handle noise. In practice, the network does not receive a perfect Kannada character as an input.

7. EXPERIMENTAL RESULTS

The database is created by collecting data from 25 different writers so that recognition engine could be trained with different styles of handwriting. The method has been implemented in MATLAB 13b on Dual Core 3 GHz with 4 GB RAM. For experimental purpose, we have considered 500 handwritten words collected from different individuals of various professions.



8. CONCLUSION

andwritten character recognition is the complex task in document image processing. Several reasons for this are quality of the scanner, degraded image, thickness of the pen, handwriting style. Compare to English like languages identifying the Kannada handwritten characters is the most challenging task. This is because of the huge number of characters and shapes in the Kannada language. Although several works has been done in this Kannada character recognition field, still it has got open problem for the researchers. Lack of availability of standard dataset is more difficult to start with the Kannada handwritten character recognition process. Lot of time is required to collect the handwritten words, scanning, cropping and resizing. Also subscripts are can't be identified correctly. Handwritten characters recognition involves data acquisition, preprocessing, segmentation, feature extraction and classification stages. Good combination of data set, feature extraction method and classifiers gives the better accuracy rate. Experimental result shows the good recognition rate towards handwritten Kannada characters.

REFERENCES

[1] M. Amrouch, Y. es-saady, "Handwritten Amazigh character recognition system based on continuous HMMs and directional features", IJMER, vol. 2, issue 2, 2012, pp-436-411.

[2] Bindu S Moni,G Raju "Modified quadratic classifier and directional features for handwritten Malayalam character recognition", Computational Science-New Dimensions and Perspectives, NCCSE, 2011.

[3] K.S. Prasanna Kumar "Algorithm to Identify Kannada Vowels using Minimum Features Extraction method" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-2, Issue-2, January 2013

[4] Cheng-Lin Liu "Handwritten Chinese character recognition: effects of shape normalization and feature extraction", inria 00120408, Dec 2006.

[5] Mamatha H.R, Srikantamurthy K." Morphological Operations and Projection Profiles based Segmentation of Handwritten Kannada Document", International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA Vol. 4– No.5, and October 2012

[6] Shraddha S. Gundal & P. P. Narwade, "Handwritten character recognition using neural Network with four, eight & twelve directional feature extraction techniques", International Journal of Electronics, Communication & Instrumentation Engineering Research and Development (IJECIERD) ISSN(P): 2249-684X; ISSN(E):2249-7951Vol. 4, Issue 2, Apr 2014, 5-12.

[7] Sangame S.K., Ramteke R.J., Rajkumar Benne," Recognition of Isolated Handwritten Kannada Vowels", Advances in Computational Research, ISSN: 0975–3273, Vol. 1, Issue 2, 2009, pp-52-55.

[8] Sumedha B. Hallale, Geeta D. Salunke "Twelve Directional Feature Extraction for Handwritten English Character Recognition" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277- 3878, Volume-2, Issue-2, May 2013.