International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 7 (2015) pp. 16923-16930

http://www.ripublication.com

Segmentation Approach for Offline Cursive Handwritten Words

Sonal Bansal[1] and Dr. Rinku Dixit[2]

M.Tech Scholar[1] and Assistant Professor[2]
Department Of Computer Science[1] and Deparment Of Information Technology[2],
Manav Rachna College of Engineering, Faridabad (Haryana) - 121002, India
sonelbansal26@gmail.com , rinkudixit.mrce@mrei.ac.in

Abstract

In this paper we are trying to highlight some of the techniques used for segmentation of offline handwriting and also proposed a novel algorithm for Segmentation of cursive handwritten words in English alphabets (in small letter). The segmentation algorithm analysis will be conducted on 10 cursive handwriting samples taken from an individual. Before doing segmentation, preprocessing of an image will be performed, which will enhance the segmentation process up to some extent. The segmentation algorithm deals in vertical manner. Amongst given samples, only 12 alphabets are getting segmented correctly. By segmentation we also observed that few alphabets do not get segmented when they come in middle of a word instead of in the beginning.

Keywords: Segmentation; Preprocessing; Slant Correction

1. Introduction

Handwriting Recognition is one of the research areas in the field of Pattern Recognition. The main aim of handwriting recognition is to mimic the handwriting of humans. Handwriting Segmentation is one of the very important aspects for recognition of characters. If segmentation is done correctly then it will be helpful for the neural network to recognize the character. So, here we are exploring one of the major steps for effective handwriting recognition.

Handwriting Segmentation is a very complex task as everyone has its own way of forming a character and when it comes to cursive handwriting which itself is complex as there is more variation in the character formation in terms of shape, font and style of writing a character. There are a number of segmentation algorithms that

have been developed according to the analytical perspective of the handwriting with excellent recognition rates. Most of these segmentation algorithms are either fused with neural network to correctly segment characters or work with fixed sized alphabet words for segmentation. The algorithm proposed in this paper has been designed to segment irregular or variable size alphabets and has not been paired with neural networks for segmentation purpose. The segmentation efficiency achieved so far is almost 100% for few alphabets.

Segmentation is also referred to as 'presegmentation' or 'oversegmentation'. Segmentation is broadly classified into three levels: - line segmentation, word segmentation and character segmentation. Line Segmentation approach segments page text into lines, Word Segmentation approach segments line into words and lastly Character Segmentation segments the word into characters which can be fed to neural network for recognition [1]. A number of algorithms have been developed for segmentation of cursive handwriting. One such algorithm is heuristic feature detection algorithm [2] which is used to get the segment points of handwritten words and valid segment points will be trained with neural network. The algorithm is being implemented on printed as well as cursive handwritten words taken from CEDAR database. Amjad and Mohammad [3] proposed a character segmentation algorithm on cursive handwritten words taken from a benchmark database IAM and also they use an Artificial Neural Network to train the valid segmentation points so, that efficiency of an algorithm increases i.e. they integrate the segmentation algorithm and neural network. Nicchiotti and Scagliola [4] proposed a segmentation method for the cursive words based on simple ligature model. The segmentation procedure consist of three steps: - Firstly, detect PSPs (Possible Segmentation Points) which uses a ligature model to detect features such as local minima contour and holes detection, Secondly, determination of cut direction which is used to cut the image either in vertical or slanting manner. It is fruitful and can be recognized correctly in many word images in slanting direction rather than vertical direction and thirdly, Stroke generation, includes the aggregation and rejection. Zimmermann and Bunke [5] proposed the segmentation scheme of offline cursive handwriting lines taken from IAM database. The approach adopted is bounding box to separate words from text line which uses horizontal coordinates by using HMM (Hidden Markov Model) based segmentation. Cheng and Blumenstein [6] develop a fusion of neural networks and segmentation technique known as enhanced neural network based segmentation for improving the segmentation technique of cursive handwritten words. This improvement technique uses a segmentation algorithm and MDF (Modified Direction Feature) for SPV (Segmentation Point Validation), LC (Left character) Extraction and CC (Centre character) extraction to increase overall segmentation results. Blumenstein [7] also proposed the segmentation of offline cursive script using neural network techniques. An Enhanced Heuristic Segmenter (EHS) technique is used to detect the ligature and neural assistant which is a hybrid technique which combines character classifier with the heuristic rules and the neural confidence values will be taken from MDF for SPV, LCV (Left Character Validation) and CCV (Centre Character Validation) and finally fusion of confidence value of Segmentation Area and confidence value of LC and CC take place. A new methodology is used to segment characters and recognition which

will make use of every best characteristic of the gray scale image. In this approach, character segment region is being determined using projection and topographic features in gray scale image and then these segmented characters find a nonlinear character boundaries using multistage graph algorithm and finally character recognition is being adopted using recognition based segmentation approach [8]. Singh and Dhaka [9] proposed two vertical segmentation techniques based on average height of word i.e. ½ and 2/3 height of word. The segmentation of 2/3 height is better as compare to ½ height of the word. Saba, Rehman and Sulong [10] proposed a character segmentation of handwritten words based on characters geometric feature i.e. segmentation of closed characters, segmentation of open characters and finally ANN (Artificial Neural Network) is being used to trained the valid segment points of cursive words. The training with ANN provides high speed and accuracy. Ranadhir and Moumita [11] describes the rule based segmentation where input data is English cursive handwriting words. The segmentation has been performed on cursive words. Three steps have been followed for segmentation: firstly, is to find the baselines i.e. upper baseline, Lower baseline, Middle baseline, Ascender line and Descender Line. Secondly, over segmentation algorithm is used to assign a CSP that is Candidate Segmentation Point which will be used in next step to validate rule based validation. Third step is to apply rule based validation and also they propose four rules to remove the wrong segmentation points.

This paper focuses on one of the major steps for recognizing a character i.e., the Segmentation approach for offline cursive handwritten words. In this paper, we are dealing with segmenting a word into an isolated character commonly known as Character Segmentation. The rest of the paper is organized in following sections: Section 2 describes the proposed methodology of our work, Section 3 specifies the result obtained by the segmentation algorithm and Conclusions are specified in Section 4 followed by references in the last.

2. Proposed Methodology

The proposed work revolves around the segmentation of cursive handwriting. The approach consists of three main steps: - Image Acquisition, Preprocessing and Feature Extraction. MATLAB 7.11.0 is used for implementing the algorithms. The cursive handwritten word samples have been taken from an individual and segmentation algorithm will be performed on it. These samples contain every alphabet so, to test the segmentation algorithm on each alphabet. The steps required for segmentation of handwritten words are described as follows in Fig. 1: -

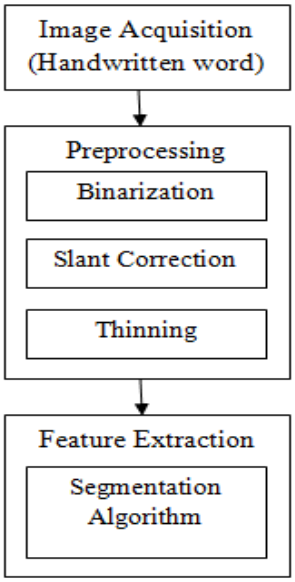

Fig.1 :-Segmentation Steps

Image Acquisition

The first step in any handwritten recognition system is to collect the data. The sample set used for current research are the scanned handwritten words of an individual on a plain white paper. The handwritten word image can be of variable size i.e. no fixed dimensions or pixel values for the word image. This scanned image is used as an input and further this will be preprocessed.

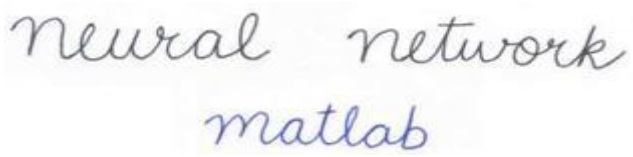

Fig.2 :-Scanned handwritten word image

Preprocessing

This step is to enhance the characters and the text, i.e. if there is any distortion or noise, reduction of the same in the image acquired will be removed in this step. Following preprocessing tasks will be performed on an acquired image as depicts in Table 1,

TABLE I. PREPROCESSING STAGES

Preprocessing Tasks	**Description**
Binarization	This will convert gray scale image to binary image, i.e. in 0 and 1 form.
Slant Correction	This step is very important aspect while dealing with cursive handwriting, as there is a little bit of slanting in handwriting. Proper slanting of character will help in segmenting character accurately as if this step is not performed then certain overlapping of characters with one another will occur which lays down the accuracy of characters.
Thinning	Thinning algorithm will be applied to the word.

Preprocessing results are shown in Fig. 3,

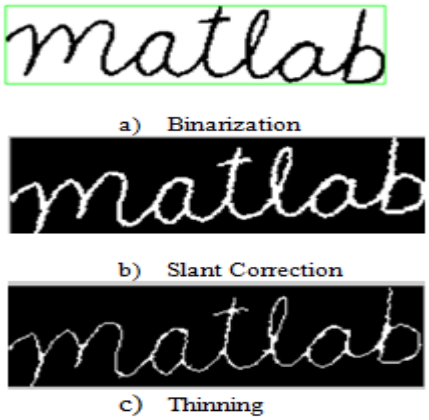

Fig.3:- Preprocessing stages

Feature Extraction

In feature extraction phase the characters will be extracted as patterns. For feature extraction we will use proposed segmentation algorithm. The segmentation algorithm will split the word in vertical manner i.e. the algorithm works on the 2/3 * height of the word image [9]. The proposed segmentation segment the preprocessed image which will remove all extra foreground pixels from the word image and is ready for feature extraction. Many problems arise when we are proposing a segmentation the problems are as follows: there is no uniformity in handwriting of an individual, Characters in words are touching each other or sometimes overlapping also occur and there is no fixed ascender or descender line as samples have been taken on white papers which will be written in irregular form.

The proposed segmentation algorithm is described in Algorithm 1,

Algorithm 1. Segmentation Algorithm
{
For each column matrix
For each row matrix
Check condition for image to contain specific pixel values, (rownumber> twothirdheight) and also check for consecutive pixel values to compute segment points
Store these segment points to cut the image
}
The segmentation result of word 'matlab' obtained depicts in Figure 4,

Fig.4:- Segmentation Result

3. Experimental Results

The proposed segmentation approach has been applied to 10 handwritten sample of an individual which will contain lowercase handwritten words. The sample word contains all alphabets from a –z. The segmentation algorithm will segment 12 alphabets correctly out of 26.

The segmentation results are shown in Table 2,

TABLE II. SEGMENTATION RESULTS

Segmented alphabets	Alphabets	Percentages
Correctly	a, d, e, f, h, k, l , m, n, r, s, t	100%
Incorrectly	b, c, g, i, j, o, p, q, u, v, w, x, y, z	-

Our proposed segmentation will correctly segment alphabets a, d, e, f, h, k, l, m, n, r, s, t whether these alphabets occur in between the words or the starter. The Segmentation rate for these alphabets in any word is 100% whereas some alphabets cannot be segmented correctly.

Discussion of Results

Many researchers have proposed several segmentation techniques but each of them have their different methodology for segmenting a character some of them uses only segmentation algorithm whereas others use a fusion of segmentation with neural

network which is also one of the accurate method to increase the segmentation rate.

Segmentation results based on the literature survey is presented in Table 3,

TABLE III. ANALYSIS OF SEGMENTATION RESULT

Author	**Segmentation Approach Used**	**Database Used**	**Segmentation Result**
Zimmermann and Bunke [5]	HMM based segmentation	IAM database	98%
Blumenstein and Verma [2]	Heuristic Segmentation + ANN	CEDAR database	75.06% and 76.52%
Khan and Mohammad [3]	Segmentation algorithm + ANN	IAM database	91.21%
Nicchiotti and Scagliola [4]	Rule based	CEDAR database	86.9%
Cheng and Blumenstein [6]	Feature extraction + ANN	CEDAR	Not mentioned only segmentation error rates are there
Blumenstein [7]	Heuristic algorithm + ANN	CEDAR	-
Singh and Dhaka [9]	Vertical Segmentation algorithm	250 words of three letters (English alphabets)	½ word - 71% 2/3 word– 83%
Saba, Rehman and Sulong [10]	Geometric Feature analysis + ANN	CEDAR	86.44%

4. Conclusion

As we have seen many segmentation algorithms each with its own drawbacks as well as advantages. Segmentation is the most crucial step in any handwriting recognition problem as every hand-writing has its own way of forming characters. In our case, we saw even a single individual cannot write a specific letter in the same manner repeatedly. Letters vary in both shape and size. Hence a new segmentation algorithm has been proposed. Our segmentation algorithm correctly segments only 12 letters so far whereas in remaining letters there are certain segmentation problems and the work is on to overcome those as well.

References

[1] B. Yanikoglu and P. A. Sandon, 1998,"Segmentation of Off-Line Cursive Handwriting using Linear Programming," Pattern Recognition, vol. 31, pp. 1825-1833.

[2] M. Blumenstein and B.Verma, "A New Segmentation Algorithm for Handwritten Word Recognition," Neural Networks, 1999.IJCNN'99.International Joint Conference, vol. 4, pp.2893-2898.

[3] Amjad Rehman Khan and Dr. Zulkifli Mohammad , 2008, "A Simple Segmentation Approach for Unconstrained Cursive Handwritten Words in Conjunction with the Neural Network," International Journal of Image Processing, vol. 2, No, 3, pp. 29-35.

[4] G. Nicchiotti and C. Scagliola, 2000, "A Simple and Effective Cursive Word Segmentation Method," Proceedings of the 7th International Workshop on the Frontiers of Handwriting Recognition(IWFHR – 7), pp. 499-504.

[5] Zimmermann M and Bunke, 2000, "Automatic Segmentation of the IAM Off-line Database for Handwritten English Text," In Proc. Of the 16th Int. Conf. on Pattern Recognition, vol. 4, pp. 35 - 39.

[6] C. K. Cheng and M. Blumenstein, 2005, "Improving the Segmentation of Cursive Handwritten Words using Ligature Detection and Neural Validation," In Processinds of the 4th Asia Pacific International Symposium on Information Technology (APIS 2005), Gold Coast, Australia, pp. 56 – 59.

[7] Michael Blumenstein, 2008, " Cursive Character Segmenattion using Neural Network Techniques," Springer, pp. 259 - 275.

[8] Seong–Whan Lee, Dong-June Lee and Hee-Seon Park, 1996, " A New Methodology for Gray-Scale Character Segmenattion and Recognition," IEEE Transactions on Pattern analysis and Machine Intelligence, vol. 18, no. 10, October, pp. 1045 - 1050.

[9] M. P. Singh and V. S. Dhaka, 2009, " Handwritten Character Recognition using Modified Gradient Descent Technique of Neural Networks and Representation of Conjugate Descent for Training Patterns," International Journal of Engineering Transactions A: Basics, vol. 22, no. 2, pp. 188 - 193.

[10] Tanzila Saba, Amjad Rehman and Ghazali Sulong, 2011, " CURSIVE SCRIPT SEGMENTATION WITH NEURAL CONFIDENCE," International Journal of Innovative Computing, Information and Control, vol. 7, no. 8, August, pp. 4955 - 4964.

[11] R. Ghosh and M. Ghosh, 2005," An Intelligent Offline Handwriting Recognition System Using Evolutionary Neural Learning Algorithm and Rule Based Over Segmented Data Points," Journal of Research and Practice in Information Technology, vol. 37, no. 1, February, pp. 73 - 88.