

Optical Character Recognition for Handwritten Cursive English characters

Aparna.A , Prof.I.Muthumani

*Department of ECE,
Alagappa Chettiar College of Engineering and Technology, Karaikudi, India*

ABSTRACT-Optical Character Recognition (OCR) is the technique which enables a machine to automatically recognize the characters or scripts written in the users' language. Optical Character Recognition (OCR) has become one of the most successful applications of technology in the field of pattern recognition and artificial intelligence. In this project a scanned image is translated into machine editable text by means of using Optical Character Recognition. Here a hand written English cursive word is scanned and this image is fed into the computer in which it is recognized using Hidden Markov Model, and converted into the same word in equivalent printed characters . A new combination algorithm is developed and has been used for this recognition and generation work, which is implemented in MATLAB 2010 a. Among the heavy competition in this application area, this project is being developed to achieve better accuracy and overcome all the draw backs found in the other available OCR algorithms.

KEYWORDS-Optical Character Recognition (OCR), Cursive Handwriting, Hidden Markov Model (HMM)

I. INTRODUCTION

Machine replication of human functions, like reading, is an ancient dream. However, over the last five decades, machine reading has grown from a dream to reality. Optical character recognition has become one of the most successful applications of technology in the field of pattern recognition and artificial intelligence. Many commercial systems for performing OCR exist for a variety of applications, although the machines are still not able to compete with human reading capabilities. OCR is a process to translate the human readable hand-written characters to machine readable characters. Text

detection and character recognition, which is known as Optical Character Recognition (OCR) has become one of the most successful applications of technology in the field of pattern recognition and artificial intelligence. The origin of character recognition can actually be found back in 1870. During the first decades of the 19th century several attempts were made to develop devices to aid the blind through experiments with OCR. However, the modern version of OCR did not appear until the middle of the 1940's with the development of digital computer. The motivation for development from then on, was the possible applications within the business world.

Several algorithms are available using statistical, structural, pattern and neural network systems for the recognition of handwritten characters. Major disadvantages of character recognition using neural networks are large training time and computational cost. To extract the characters, filters are commonly used. To recognize the characters, there exists a method using Gabor filter [1]. The major drawback of Gabor filtering is that they are computationally intensive. The median filter overcomes the limitations of the linear filters [2]. It effectively removes the errors presented in the segmented characters. Character segmentation is one of the most important and difficult stage in the recognition process. Segmentation is carried out by means of stoke based and projection profile methods [3] [4].

In this paper, we propose an effective statistical based recognition method to recognize the handwritten cursive English characters.

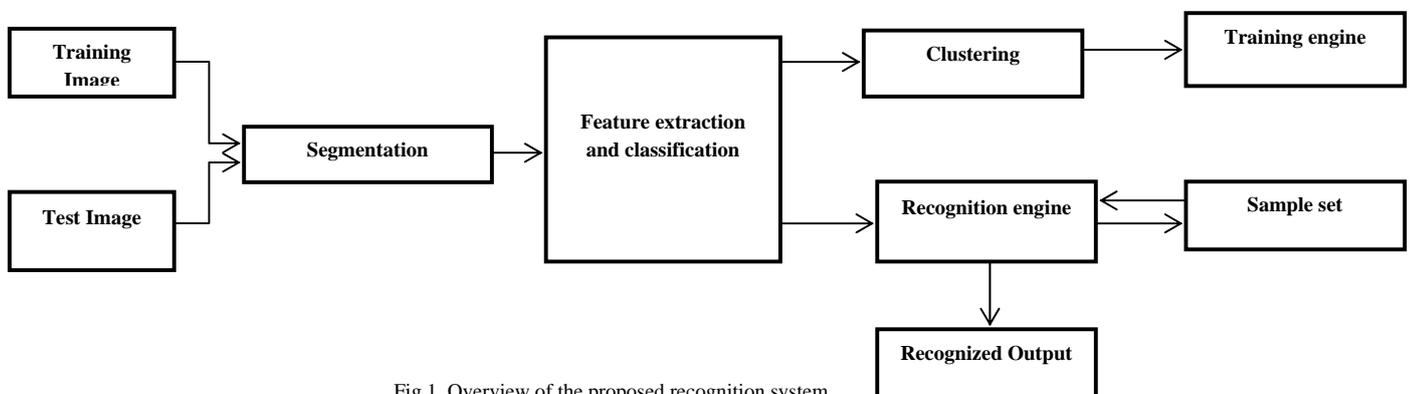


Fig 1. Overview of the proposed recognition system

II. METHOD

This method consists of four steps. They are: image acquisition, preprocessing of an image, segmentation of characters, feature extraction of characters and character recognition.

A. Image Acquisition

The handwritten character to be recognized is acquired by using an optical scanner.

B. Image Preprocessing

The acquired image is not suitable for recognition process. Hence, the acquired image is required to undergo a preprocessing step to convert it to a usable form for further stages of recognition process. The preprocessing stage includes skew correction and normalization.

In optical character recognition (OCR), the text lines in a document must be scanned properly. While scanning through an optical scanner few degrees of skew is unavoidable. Skew refers to a tilt in the scanned image. Skew detection and correction are the important preprocessing steps in character recognition process. This skew in an image is estimated by means of using thinning algorithm along with Hough transform. And the estimated skew is corrected by means of using coordinate transformation method.

Generally handwriting fluctuation occurs between people. Not only the people vary in their writing styles, but also vary in geometric features such as slant. Normalization method is used to take out slant from the handwritten characters. It also refers to changing the range of pixel intensity values. It adjusts the pixel values to a standard range.

C. Segmentation

Segmentation is the process of separating characters in a word and it is the most difficult part in the cursive handwritten recognition process. Here the segmentation regions are identified from the peaks of the vertical projection profile. Vertical projection of a binary image looks like a set of black hills on a white surface. After extracting the segmentation regions, characters are segmented.

D. Feature Extraction

Feature extraction stage employs the extraction of the texture features of the handwritten characters. For this purpose median filter is employed. All the segmented characters are scaled into common height using image resizing technique. Unwanted portions and noise in the segmented characters are removed using median filter.

E. Recognition

Hidden Markov Model (HMM) is used to recognize the segmented characters. A HMM is a doubly stochastic process that is not observable (it is hidden). It is a statistical model. After feature extraction process, the extracted features are recognized using HMM training and ranking process which in turn contains training and test image as input.

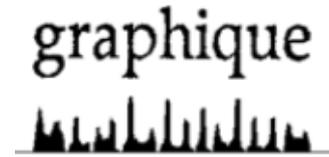


Fig 2. Example for Vertical projection profile

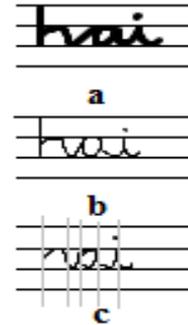


Fig 3. (a) Original input image, (b) Image after preprocessing, (c) Image with detected segmented regions

III. CONCLUSION

This paper deals with the recognition of cursive handwritten English characters using Hidden Markov Model (HMM). A new algorithm is developed, which avoids errors caused by noise in the scanned image by applying median filter. The samples used are of high quality to reduce the complexities in the recognition process. This approach shows a better result in terms of speed and accuracy.

REFERENCES

- [1] Ramanathan.R, Ponmathavan.S, Valliappan.N, Thaneshwaran.L, "Optical Character Recognition for English and Tamil using Support Vector Machine", International conference on Advances in Computing, Control, & Telecommunication Technologies, 2009, pp.610-612
- [2] Kanika Bansal, Rajiv Kumar, "k-algorithm: A modified technique for noise removal in handwritten documents", International Journal of Information Sciences and Techniques, May 2013, volume 3, No 3
- [3] Nafiz Arica, Fatos T.Yarman-Vural, "Optical Character Recognition for cursive handwriting", IEEE transactions on Pattern analysis and Machine Intelligence, volume 24, Issue 6
- [4] Naazia Makkar, Sukhjit Singh, "A brief tour to various skew detection and correction techniques", International Journal for Science and Emerging Technologies with Latest Trends, 2012, pp.54-58
- [5] Bhagirath Kumar, Niraj Kumar, Charulata Palai, Pradeep Kumar Jena, Subhagata Chattopadhyay, "Optical Character Recognition using Ant Miner Algorithm: A Case Study on Oriya Character Recognition", International Journal of Computer Applications, 2013, volume 61-No.3
- [6] Muthumani.I, Uma Kumari C.R, "Online Character Recognition of Handwritten Cursive Script", International Journal of Computer Science Issues, Vol. 9, Issue 3, No 2, May 2012
- [7] Tripathi R.C, Kumar.V, "Character Recognition: A Neural Network Approach" Proceedings published in International Journal of Computer Applications, pp.17-20, 2012
- [8] Ming Ma, Dong-Won Park, Soo Kyun Kim and Syungog An, "Online Recognition of Handwritten Korean and English Characters", Journal of Information Processing Systems, Volume 8, No 4, December 2012