International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 05 | May-2016 www.irjet.net p-ISSN: 2395-0072

Optical Character Recognition for Cursive Handwriting

Mehak Naz Mangoli[1], Prof. Sujata Desai[2]

[1]*Student, Department of Computer Science and Engineering, BLDEA College of Engineering and Technology, Karnataka, India*

[2]*Professor, Department of Computer Science and Engineering, BLDEA College of Engineering and Technology, Karnataka, India*

Abstract - *Recognition of Handwritten text has been one of the active and challenging areas of research in the field of image processing and pattern recognition. It has numerous applications which include, reading aid for blind, historical document recognition, bank cheques. In this paper we focus on recognition of cursive English alphabet in a given scanned text document with the help of SVM. The first step is image acquisition which acquires the scanned image followed by normalization of scanned image followed by feature extraction, then making image suitable for segmentation where image is decomposed into sub images and finally each character classification using SVM classifier. Feature Extraction improves recognition rate by making all the characters into a common height. A new combination methods have been developed and has been used for this recognition, which is implemented in MATLAB 2013 a.*

Key Words: Offline Handwriting Recognition, Support Vector Machine, Optical Character Recognition.

1. INTRODUCTION

Handwriting recognition has been one of the most fascinating and challenging research areas in field of image processing and pattern recognition in the recent years. It contributes immensely to the advancement of automation process and improves the interface between man and machine in numerous applications. In general, handwriting recognition is classified into two types as off-line and on-line handwriting recognition methods. The on-line methods have been shown to be superior to their off-line counter parts in recognizing handwritten characters due to the temporal information available with the former. However, in the off-line systems, comparably high recognition accuracy level is obtained.

The off-line handwriting recognition (OHR) still remains an active area for research towards exploring the newer techniques that would help improving recognition accuracy. It is because of the fact that several applications including mail sorting, bank processing, document reading and postal address recognition require offline handwriting recognition systems. Character recognition is nothing but Machine simulation of human reading [1], [2]. It is also known as Optical Character Recognition. It contributes immensely to the advancement of an automation process and can improve the interface between man and machine in numerous applications. Several research works have been focusing on new techniques and methods that would reduce the processing time while providing higher recognition accuracy. Study reveals that the methods of Character Recognition have grown up sequentially [3], [4]. The recognition of isolated handwritten character was first investigated [5], but later whole words were addressed.

2. LITERATURE SURVEY

An early notable attempt in the area of character recognition research is by Grimsdale in 1959. The origin of a great deal of research work in the early sixties was based on an approach known as analysis-by-synthesis method suggested by Eden in 1968. The great importance of Eden's work was that he formally proved that all handwritten characters are formed by a finite number of schematic features, a point that was implicitly included in previous works. This notion was later used in all methods in syntactic (structural) approaches of character recognition.

K. Gaurav, Bhatia P. K. [7] *Et al*, this paper deals with the various pre-processing techniques involved in the character recognition with different kind of images ranges from a simple handwritten form based documents and documents containing colored and complex background and varied intensities. In this, different preprocessing techniques like skew detection and correction, image enhancement techniques of contrast stretching, binarization, noise removal techniques, normalization and segmentation, morphological processing techniques are discussed. It was concluded that using a single technique for preprocessing, we can't completely process the image. However, even after applying all the said techniques might not possible to achieve the full accuracy in a preprocessing system.

Salvador España-Boquera *et al* [8], in this paper hybrid Hidden Markov Model (HMM) model is proposed for recognizing unconstrained offline handwritten texts. In this, the structural part of the optical model has been modeled with Markov chains, and a Multilayer Perceptron is used to

estimate the emission probabilities. In this paper, different techniques are applied to remove slope and slant from handwritten text and to normalize the size of text images with supervised learning methods. The key features of this recognition system were to develop a system having high accuracy in preprocessing and recognition, which are both based on ANNs.

In [9], a modified quadratic classifier based scheme to recognize the offline handwritten numerals of six popular Indian scripts is proposed. Multilayer perceptron has been used for recognizing Handwritten English characters [10]. The features are extracted from Boundary tracing and their Fourier Descriptors. The character is identified by analysing its shape and comparing its features that distinguish each character. Also an analysis has been carried out to determine the number of hidden layer nodes to achieve high performance of the back propagation network. Recognition accuracy is less for Handwritten English characters with less training time.

In [11], diagonal feature extraction has been proposed for offline character recognition. It is based on ANN model. Two approaches using 54 features and 69 features are chosen to build this Neural Network recognition system. To compare the recognition efficiency of the proposed diagonal method of feature extraction, the neural network recognition system is trained using the horizontal and vertical feature extraction methods.

A. Brakensiek, J. Rottland, A. Kosmala, J. Rigoll [12] *et al*, in this paper a system for off-line cursive handwriting recognition is described which is based on Hidden Markov Models (HMM) using discrete and hybrid modeling techniques. Handwriting recognition experiments using a discrete and two different hybrid approaches, which consist of discrete and semi-continuous structures, are compared. A segmentation free approach is considered to develop the system. It is found that the recognition rate performance can be improved of a hybrid modelling technique for HMMs.

R. Bajaj, L. Dey, S. Chaudhari *et al* [13], employed three different kinds of features, namely, the density features, moment features and descriptive component features for classification of Devanagari Numerals. They proposed multi classifier connectionist architecture for increasing the recognition reliability and they obtained less accuracy for handwritten Devanagari numerals.

Sandhya Arora in [14], used four feature extraction techniques namely, intersection, shadow feature, chain code histogram and straight line fitting features. Shadow features are computed globally for character image while intersection features, chain code histogram features and line fitting features are computed by dividing the character image into different segments.

Mohammed Z. Khedher, Gheith A. Abandah, and Ahmed M. Al Khawaldeh [15] *et al*, this paper describes that Recognition of characters greatly depends upon the features used. Several features of the handwritten Arabic characters are selected and discussed. An off-line recognition system based on the selected features was built. The system was trained and tested with realistic samples of handwritten Arabic characters. Evaluation of the importance and accuracy of the selected features is made. The recognition based on the selected features give average accuracies of 88% and 70% for the numbers and letters, respectively. Further improvements are achieved by using feature weights based on insights gained from the accuracies of individual features.

3. PROPOSED MODEL

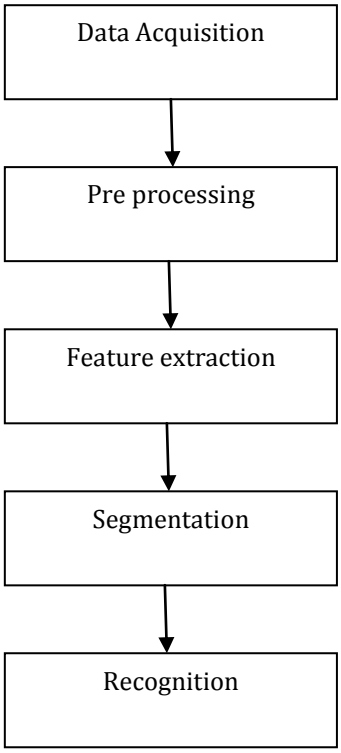

Figure 3: Flow Chart of Proposed Model

Figure 3 shows the flow chart of the proposed method. The proposed system is designed as follows. The handwritten character to be recognized is acquired by using an optical scanner.

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 can be 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.

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 hills on a white surface. After extracting the segmentation regions, characters are segmented.

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.

We use Support Vector Machines to recognize our segmented characters due to the fact it gives in the best results in pattern recognition and is easy to implement. SVM is a set of methods used in machine learning for classification and regression from a set of learning data. SVMs are among the best "off-the-shelf" supervised learning algorithm.

4. CONCLUSION AND FUTURE SCOPE

This work deals with the recognition of cursive handwritten characters using SVM. The samples used are of high quality to reduce the complexities in the recognition process. This work successfully gives individual characters of the input word image. This work gives better results than previous works and is easy to implement due to the use of SVM.

Future work can be done by improving the recognition accuracy and speed in much more better way. It can be improved further to get accurate result in noisy environment. Document retrieving can also be done as future work.

REFERENCES

[1] U. Bhattacharya, and B. B. Chaudhury, "Handwritten numeral databases of Indian scripts and multistage recognition of mixed numerals", IEEE Trans. Pattern analysis and machine intelligence, vol. 31, No. 3, pp. 444-457, 2009.

[2] U. Pal, T. Wakabayashi and F. Kimura, "Handwritten numeral recognition of six popular scripts", Ninth International Conference on Document Analysis and Recognition, ICDAR07, Vol.2, pp.749-753, 2007.

[3] V. K. Govindan and A. P. Shivaprasad,"Character Recognition A review", Pattern recognition, vol.23, no.7, pp.671-683, 1990

[4] J. Pradeep, E. Srinivasan and S. Himavathi, "Diagonal Based Feature Extraction for Handwritten Alphabets Recognition System Using Neural Network", International Journal of Computer Science and Information Technology (IJCSIT),, vol. 3, no. 1, pp. 27-38, Feb 2011.

[5] C. Suen, C. Nadal, R. Legault, T. Mai, and L. Lam, "Computer recognition of unconstrained handwritten numerals", *Proc. IEEE* , 80(7):1162-80.

[6] Aparna.A ,.I.Muthumani, "Optical Character Recognition for Handwritten Cursive English characters", International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 847-848.

[7] K. Gaurav and Bhatia P. K., "Analytical Review of Preprocessing Techniques for Offline Handwritten Character Recognition", 2nd International Conference on Emerging Trends in Engineering & Management, ICETEM, 2013.

[8] Salvador España-Boquera, Maria J. C. B., Jorge G. M. and Francisco Z. M., "Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 4, April 2011.

[9] U. Pal, T. Wakabayashi and F. Kimura, "Handwritten numeral recognition of six popular scripts," Ninth International conference on Document Analysis and Recognition ICDAR 07, Vol.2, pp.749-753, 2007.

[10] Anita Pal & Dayashankar Singh, "Handwritten English Character Recognition Using Neural," Network International Journal of Computer Science & Communication. Vol. 1, No. 2, July-December 2010, pp. 141-144.

[11] J. Pradeep, E. Srinivasan and S. Himavathi, "Diagonal Based Feature Extraction For Handwritten Alphabets Recognition System Using Neural Network", International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 1, Feb 2011.

[12] A. Brakensiek, J. Rottland, A. Kosmala and J. Rigoll, "Offline Handwriting Recognition using various Hybrid Modeling Techniques & Character N-Grams".

[13] Reena Bajaj, Lipika Dey, and S. Chaudhury, "Devnagari numeral recognition by combining decision of multiple connectionist classifiers", Sadhana, Vol.27, part. 1, pp.-59-72, 2002.

[14] Sandhya Arora, "Combining Multiple Feature Extraction Techniques for Handwritten Devnagari Character Recognition", IEEE Region 10 Colloquium and the Third ICIIS, Kharagpur, INDIA, December 2008.

[15] Mohammed Z. Khedher, Gheith A. Abandah, and Ahmed M. Al- Khawaldeh, "Optimizing Feature Selection for Recognizing Handwritten Arabic Characters", proceedings of World Academy of Science Engineering and Technology, vol. 4, February 2005 ISSN 1307-6884.