



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 2, February 2017

A Survey on Off-line Application for Handwritten Text Recognition and Modification using Back-propagation and LAMSTAR

Pallavi Baviskar, Pooja Bankar, Gayatri Kavathekar, Pooja Deshmukh

Asst. Professor, Dept. of Computer Engineering, P.E.S. Modern College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India.

BE Student, Dept. of Computer Engineering, P.E.S. Modern College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India.

BE Student, Dept. of Computer Engineering, P.E.S. Modern College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India.

BE Student, Dept. of Computer Engineering, P.E.S. Modern College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India.

ABSTRACT: This paper mainly discuss about handwritten text recognition with modification. The proposed system can be applicable for all offline handwritten text. It will smoothly work on digits as well as alphabets. This application will help to recognise and modify publicly available dataset.

This paper enlighten various steps useful for recognition and modification. In that Image Enhancement, Image Binarization, Image segmentation and neural network are discuss for offline handwritten recognition. Using back propagation and LAMSTAR are methods of neural network are choose for maximum recognition rate. This paper is also focus on related work for on line and offline recognition techniques. Using those methodologies and techniques comparative analysis is discuss in depth.

KEYWORDS: Image Enhancement, Image Binarization, Region-based segmentation, Edge Detection, Feature Extraction, Back Propagation Neural Network, LAMSTAR Neural Network.

I. INTRODUCTION

Lot of a historical important data is available in written format. This historical data needs to be stored for future use. Due to huge demand for storing the information present in these documents, digitizing this data has become a need now a days. One simple way to store information in computer system is to first scan the documents and then store them as Images. But to reuse this information it is very difficult to read the individual document line by line and search the desired contents from these documents. So digitizing this data becomes important.

The proposed application allows user to convert an image of a handwritten document into printed text format. In extend of usefulness it also allows user to change the content of the document, if required. Main advantage of this application is that it works without Internet connection.

Proposed application prompts user to upload an image of the document which user wants to convert into printed format. Once the image is selected, the image will be processed as per the requirement for further processing. Then the image will be provided as input to the neural network. After processing, the user will get the text in printed format. If user wants to change the content of the document, the application will provide him the option for editing. Once the editing is done, user can save the document as PDF format.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 2, February 2017

II. RELATED WORK

Different applications are present for Optical Character recognition. In paper “ARTIFICIAL NEURAL NETWORK BASED OPTICAL CHARACTER RECOGNITION” the authors provides a methodology for OCR in recognition and classification of character from an image. It makes use of Feature Vector which helps in defining character uniquely by means of Artificial Neural Network. But these applications are unable to work with handwritten text.

There are different methods available for implementing handwritten text recognition like, Clustering, Feature Extraction, Pattern Matching, ANN, k-NN Classifier, and Direction Based [1]. The paper concluded that implementing neural network for recognition purpose not only gives faster and most accurate results, but is also able to recognize the pattern that it has encountered for the very first time [7]. In research paper “Handwritten Character Recognition in English: A Survey” authors studies various methods for implementing handwritten text recognition. This study shows that neural network provides higher accuracy giving better result. Thus, based on the comparative study of both papers we concluded that neural network provides more flexibility in recognition process as compared to other techniques like SVM, Statistical techniques, Clustering [3].

Back-propagation NN is the most commonly used techniques for pattern recognition. But this method has some drawbacks, which can be improved with the help of LAMSTAR-NN. The study shows when Back-propagation neural network is implemented with LAMSTAR-NN gives higher rate of accuracy [4].

Paper “Neural Networks for Handwritten English Alphabet Recognition” demonstrates the use of neural networks for developing a system that can recognize hand-written English alphabets, where each English alphabet is represented by binary values that are used as input to a simple feature extraction system, whose output is fed to our neural network system.

The main feature of this application is to provide user an off-line service for recognition and modification purpose in a single application.

III. PROPOSED ALGORITHM

Description of the Proposed Algorithm:

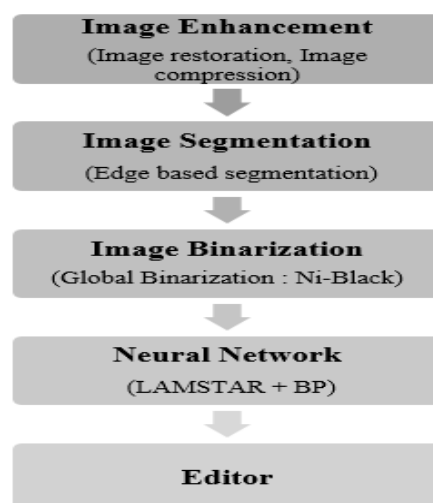


Figure. Phases for Handwritten recognition and modification

The image uploaded by the user will be enhanced for better use. It will be restored, if in case the image is damaged, and will be compressed for removing redundancy of the image data in order to be able to store in an efficient form. The stored image will be further subject to edge based segmentation. Using Ni-Black technique of Global Binarization this segmented

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 2, February 2017

image will be binarized[5]. After the image is binarized and segmented, features of the image will be extracted in order to recognize the pattern. This extracted will be stored for further processing. Once the pre-processing of the image is done, it will be given as input to the neural network for pattern recognition. Back-propagation and LAMSTAR neural networks are used for pattern recognition.

The input to our application will be a scanned image of any format. Once the image is uploaded on the application, it will be processed further as shown in the architecture diagram. The image will be restored, compressed, segmented and binarized. The features of the image will be stored, compared with the patterns stored in three different databases, i.e., database for Upper Case Letters, for Lower Case Letters and for digits. A pattern will be searched in all three databases, and outputs will be produced. These outputs will then be compared with each other, using methods of neural network like back-propagation and LAMSTAR. Depending on the results of these methods, final output will be generated. This output will then be stored into a text file. Once all the characters present in the image are transferred in the text file, this text file will be given to the editor for modification purpose. Editor will allow user to change the content, insert a picture, facility of Bold, underline and italic font and to see the preview of the PDF file. After modification, the text file will be saved as a PDF file.

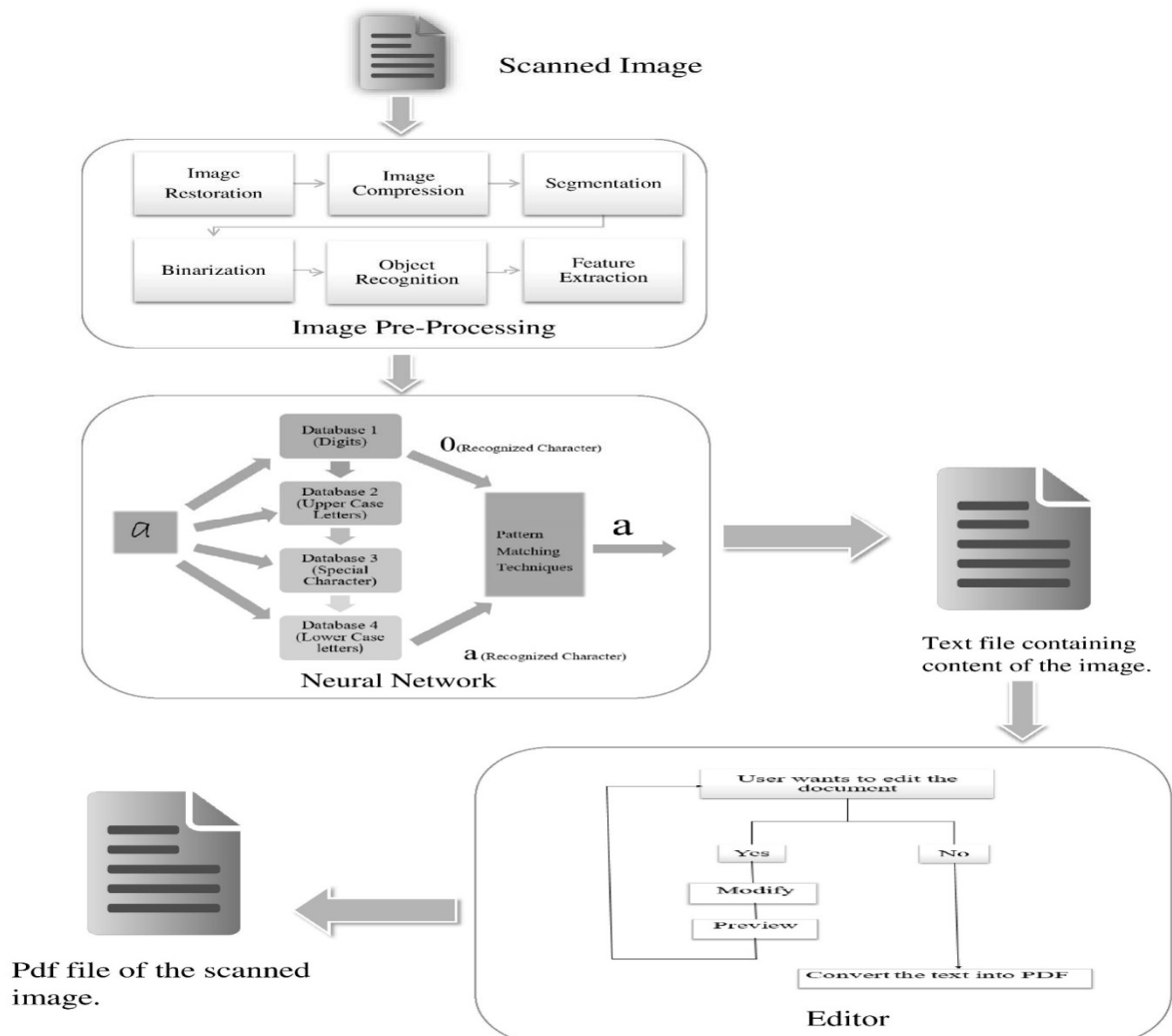


Fig. Architecture diagram for the application.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

IV. CONCLUSION

Handwritten text recognition is challenge in machine learning area. Handwritten text includes various characters and numbers in different formats. By analyzing various methods for recognition handwritten text, we observed that back propagation and LAMSTAR techniques of artificial neural network are effective not only for recognition but modifying text if required. We are aiming to develop an application which will be user friendly at the same time it will give off-line access to the user so. An application provides facility to modify the contents of a document within itself, saving the overhead of searching a different application to convert the document and edit its contents. This application will be a further step towards natural language processing and machine learning. In extension of future scope of this application will provide support for multilingual texts.

REFERENCES

1. Suruchi G. Dedgaonkar, Anjali A. Chandavale, Ashok M. Sapkal, "Survey of Methods for Character Recognition", IJEIT Volume 1, Issue 5, 180-189, May 2012,.
2. Monica Patel, Shital P. Thakkar, " Handwritten Character Recognition in English: A Survey", IJARCCCE, Vol. 4, Issue 2, 345-350, February 2015,.
3. Cheng-Lin Liu, Hiromichi Fujisawa, "Classification and Learning for Character Recognition: Comparison of Methods and Remaining Problems".
4. Majed Valad Beigi, "Handwritten Character Recognition Using BP NN, LAMSTAR NN and SVM" SPRING 2015.
5. Puneet, Naresh Kumar Garg , "Binarization Techniques used for Grey Scale Images", IJCA, Volume 71– No.1, 8-11, June 2013.
6. M. Egmont-Petersen, D. de Ridder, H. Handels , "Image processing with neural networks—a review", Elsevier, Volume 35, Issue 10, October 2002.
7. Daniel Graupe, "THE LAMSTAR NEURAL NETWORK: A BRIEF REVIEW".
8. Yusuf Perwej, Ashish Chaturvedi, "Neural Networks for Handwritten English Alphabet Recognition", IJCA , Volume 20– No.7, April 2011.
9. Vivek Shrivastava and Navdeep Sharma, "ARTIFICIAL NEURAL NETWORK BASED OPTICAL CHARACTER RECOGNITION", SIPIJ, Vol.3, No.5, October 2012.
10. Mark A. Kon and Leszek Plaskota, "Complexity of Predictive Neural Networks", International Conference on Complex Systems, May, 2000.
11. Dr. C P Sumathi and S Karpagavalli, "TECHNIQUES AND METHODOLOGIES FOR RECOGNITION OF TAMIL TYPEWRITTEN AND HANDWRITTEN CHARACTERS: A SURVEY", International Journal of Computer Science & Engineering Survey (IJCSES) Vol.3, No.6, 23- 35, December 2012.
12. Sandhya Arora. Debotosh Bhattacharjee, Mita Nasipuri, L. Malik , M. Kundu and D. K. Basu, "Performance Comparison of SVM and ANN for Handwritten Devnagari Character Recognition" , IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 3, May 2010,