Offline Handwritten English Numerals Recognition using Correlation Method

Isha Vats
M.E (CSE) – IV Sem
Chandigarh University
Gharuan, India

Shamandeep Singh
(Assistant Professor) M.E (CSE)
Chandigarh University
Gharuan, India

Abstract - **In this paper, entire system is based on recognition of offline handwriting digits. This offline handwriting number recognition system has mainly five stages: Preprocessing, Segmentation, Feature Extraction, Classification and Post-processing. The main aim of the proposed work is to efficiently recognize the offline handwritten digits with a higher accuracy than previous works done in this field. . But a difficult problem in this field is the recognition of partially or completely touching handwritten digits and the existing tools are not yet for such conditions. Also previous handwritten number recognition systems are based on only recognizing single digits and they are not capable of recognizing multiple numbers at one time. So the major concern in the proposed work is focused on efficiently performing segmentation for isolating the digits so that multiple number images can be recognized in one step and improving the accuracy of the system and to create a better handwritten number recognition system.**

Keywords - Handwritten English digits ; Pattern recognition; Segmentation; Templates matching

I. INTRODUCTION

The computer was created many years ago, and since then many new technologies and software are invented which have decreased the work load of human. But today most computers still receive input in old fashion from keyboards or a mouse. But it is very difficult to provide high degrees of inputs to computer in less time as such processes are a lot of time consuming. A few years ago, research is introduced in the field of optical character recognition. It is one of the most interesting and challenging areas of pattern recognition. It improves an automation process and proves to be an interface between human and machine.

Handwriting recognition is the ability of a machine to receive and interpret handwritten input from multiple sources like paper documents, photographs, touch screen devices etc. The other name for handwriting recognition system is optical character recognition system. In handwriting recognition system the image of the written text from various devices like scanned paper documents, pictures sensed offline by optical scanning the movements of the pen tip strokes in touch screen phones or tablets sensed online. The aim of such systems is to convert machine printed, hand printed or hand written document files into editable text format.

II. BRIEF LITERATURE SURVEY

Handwriting recognition is of two types: offline handwriting recognition and online handwriting recognition. Offline handwriting recognition is discussed in [4] Miroslav Nohaj, Rudolf Jaskaln, in which preprocessing technique of offline character recognition is discussed and feed forward neural networks technique is used for recognition process. In [9] F. Kimura, M. Shridhar, various classification methods: statistical and structural classifiers are discussed in detail with their algorithms and it includes statistical classifiers (K-algorithm) and Structural classifier (S-algorithm) for recognition of offline characters. In [10]S.V. Rajashekararadhya, Dr P. Vanaja Ranjan, the methods used is zone centroid and image centroid based distance metric feature extraction system to te recognize offline handwritten numerals of four popular south Indian scripts.

In [15] Toru Wakahara, YukihikoYamashita, the k-NN classification method for handwritten numerals recognition with GAT correlation technique is discussed. Correlation method is also called template matching method. GAT correlation method is proposed for feature extraction of handwritten numerals for recognition process. Recognition experiments are performed on handwritten numerical database and achieved recognition rates of 97.50%.

III. PROPOSED WORK

In proposed work, there are ten digits to represent the number system which are consisting of 0 to 9 as shown in Figure 1. Each digit differs from another in terms of shape.
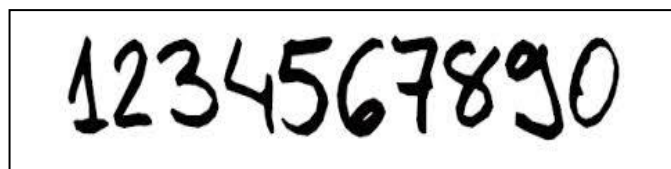

Fig 1. Ten English numerals [8]

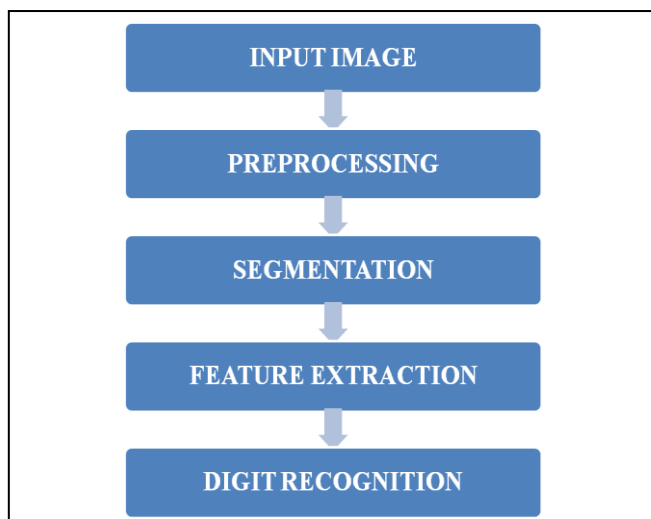The entire work is divided into following major steps:

Fig 2. Basic steps in offline handwritten digits recognition

A. *Image Aquisition*

Initially the image is input to the system, for recognition process. This process is also called as image acquisition. Various methods of image acquisition or image input to the machine are optical scanning of images of various handwritten documents, by clicking photographs through cameras etc in offline handwriting recognition process.

B. *Preprocessing*

Many times while inputting the image to the system for handwriting recognition process the image rotates itself, many times it gets ruptured with noise, blurring of image takes place, size of the text in image can be smaller or larger in different cases and many problems come across in this stage. So in preprocessing stage, basically the cleaning of input image is done so that the resulting image after preprocessing can become suitable for further stages of recognition process.[7] Following steps are performed for carrying preprocessing in proposed work are:

- Initially, the colored image is converted into gray scale image using in-built function "rgb2gray".

- Then filtration is done in which the noise is removed from the input image. For this "Weiner" filter is used as it filters binary image that has been corrupted with constant power. It is a low pass filter.

- After this, the input image is converted into "BW". The output BW image replace all pixels in image with a value greater than a specified threshold value with value 1 (white) and all other pixels are replaced with 0 (black).

- In the end, normalization is done in which image text is normalized to a specified range. In present work that specified value is "42*24" pixels.

- Also, all those objects will be removed which contains fewer than 30 pixels. For this "bewareopen" function is used.

C. *Segmentation*

In handwriting recognition techniques, the segmentation is the most important process. Segmentation is done to make the separation between characters in an image [14]. Different ways in which characters can interfere include: overlapping, touching, connected and intersecting pairs etc. For segmenting handwritten digits, lines-words-digit segmentation method is proposed.

- In this method, handwritten text is first segmented into lines, such that text in an image can be written in multiple lines.

- So initially each line having text is counted.

- After that, each line is evaluated for further segmentation process.

- In each line, words are segmented. This is called word segmentation.

- After that, character segmentation takes place in which each word is segmented in single character or letter.

- This happens in a sequential way that is initially lines are segmented, then first line is evaluated in which word segmentation is done and in that line character segmentation is done.

- After the complete evaluation of first line second line is evaluated for segmentation process.

D. *Feature Extraction and Digits Recognition using Template Matching*

In this stage, the unique features of the digits that can classify them at recognition stage are extracted. This is an important step. Its effective functioning improves the recognition rate of the system and reduces the misclassification of the handwritten text while recognition process [15]. Feature extraction is finding the set of unique features which can define the shape of a character precisely.

In the proposed work, classical technique Template matching is used. This technique is also called as Correlation template matching technique. Following steps are performed in this stage:

- In this method, templates of digits are created for training the system such that system remembers those templates.

- These templates are stored in the database for future use.

- When the recognition process starts the input digits for testing are matched with those pre-defined templates in the database.

- The testing image which matches the most with the pre-defined templates in the database are classified in one class.

In this way, recognition process is done.

IV. EXPERIMENTAL RESULTS

MATLAB software is used for implementation. The experiments were performed on many test images of English numerals. The preprocessed input image is segmented line by line and then the segmentation of words is done and finally digit segmentation is done. The scanned image is resized to 42x24 pixels. In the end correlation matching is used for feature extraction and recognition purpose. Finally the accuracy of recognized numerals is checked and efficiency is calculated. Figure 5 shows the output of the scanned input image.
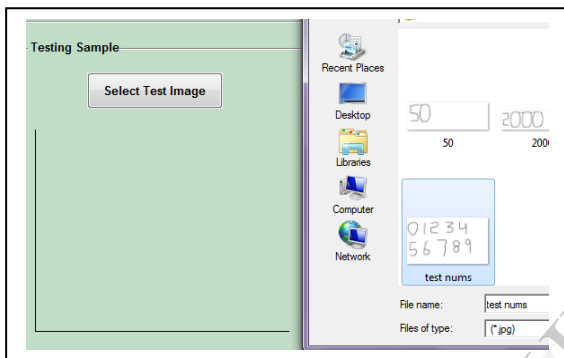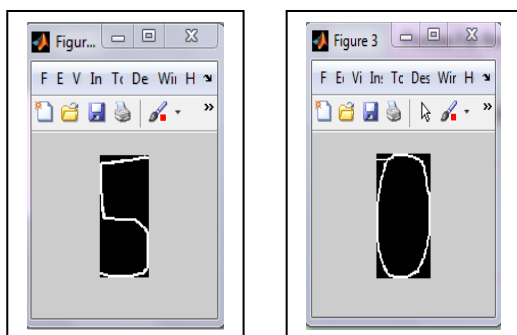

Fig 3. Input image to the system

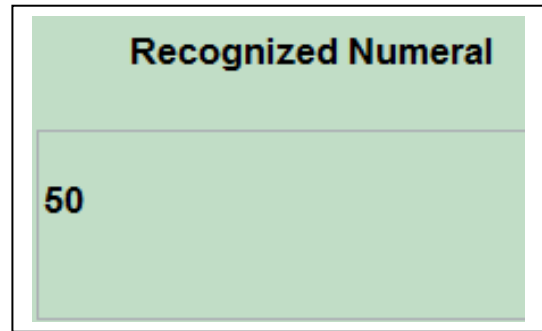Fig 4. Segmentation of input image
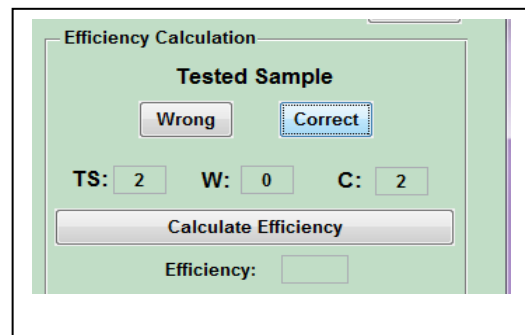

Figure 5. Recognition of input image

Fig 6. Efficiency calculation of recognized digits

V. CONCLUSION

In this proposed work, an approach for recognizing handwritten digit is described. The recognition accuracy of the implementation is promising. In this proposed work, preprocessing work is done in which normalization, filtration is performed using Weiner filter so that input image should be noise free and become appropriate for further steps for recognition. Here used digit segmentation method which can handle a larger variety of touching and overlapped digits, which occur fairly often in images obtained from handwritten material. Finally correlation matrix is used for feature extraction step and features are extracted and used for number recognition process. The efficiency in of present efficiency rate of 97% is achieved.

REFERENCES

[1] Nisha Sharma et al, "Recognition for handwritten English letters: A Review", International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 7, January 2013

[2] Vijay Laxmi Sahu et al, "Offline handwritten character recognition techniques using neural network: A Review", International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064 ,2013

[3] Gergely Timár et al, "Analogic preprocessing and segmentation algorithms for off-line handwriting recognition", Analogic and Neural Computing Systems Laboratory Computer and Automation Institute of the Hungarian Academy of Sciences, 2003

[4] Miroslav Nohaj, Rudolf Jaska, "Image preprocessing for optical character recognition using neural networks", Journal of Patter Recognition Research, 2011

[5] Kanika Bansla, Rajiv Kumar, et al, "K-algorithm: A modified technique for noise removal in handwritten documents", International Journal of Information Science And Techniques (IJIST) Vol.3, 2013

[6] R. Kumar and A. Singh "Algorithm to detect and segment Gurmukhi handwritten text into lines, words and characters" IACSIT International Journal of Engineering and Technology, vol. 3, no.4., 2011

[7] Om Prakash Sharma et al, "Recent trends and tools for feature extraction in OCR technology", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-6, 2013

[8] J.Pradeep et al, "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

[9] F. Kimura, M. Shridhar, "Handwritten numerical recognition based on multiple algorithms", Pattern Recognition, Vol. 24, No. 10. pp. 969-983, 1991

[10] S.V. Rajashekararadhya, Dr P. Vanaja Ranjan, "Efficient zone based feature extraction algorithm for handwritten numeral recognition of four popular south Indian scripts" , Journal of Theoretical and Applied Information Technology, 2008

[11] Chen Huang, Sargur N. Srihari , "Word segmentation of off-line handwritten documents", center of excellence for document analysis and recognition (CEDAR), 2008

[12] Stuti Asthana et al, "Handwritten multiscript numeral recognition using artificial neural networks", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-1, March 2011

[13] Alexander J. Faaborg NY, "Using neural networks to create an adaptive character recognition system", Cornell University, Ithaca, May 2002

[14] Sunith Bandaru, "Handwritten character recognition using neural networks", 2010

[15] Toru Wakahara, YukihikoYamashita, "k-NN classification of handwritten characters via accelerated GAT correlation",Elsevier PatternRecognition Vol. 47, 2014, Pg. 994–1001, 2013

[16] Shelpuk S, "Automated hand recognition as a human computer interface'', MIT Review,pp.1-5, 2012

[17] Lee Y, "Implementation of an interactive interview system using hand gesture recognition'',Neurocomputing,vol.116,pp.272-279, 2013

[18] Kauleshwar Prasad, Devvrat C. Nigam, Ashmika Lakhotiya And Dheeren Umre, " Character Recognition Using Matlab's Neural Network Toolbox'', International Journal of u- and e- Service, Science and Technology Vol. 6, No. 1, February, 2013.

[19] Simmi Dutta, Shruti, Haneet Kaur, Chandani Bhagat, "Digital Document Archiving System with Optical Character Recognition'', International journal of innovative research in electrical, electronics, instrumentation and control engineering Vol. 1, Issue 2, May 2013.

[20] Deborah L. Paul, P. Bryan Heidorn "Augmenting Optical Character Recognition (OCR) for Improved Digitization: Strategies to Access Scientific Data in Natural History Collections'', MIT Review,pp.1-5, 2013

[21] Mansi Shah And Gordhan B Jethava, "A Literature Review On Hand Written Character Recognition", Indian Streams Research Journal Vol - 3, ISSUE −2, March 2013

[22] M. Blumenstein, C. K. Cheng and X. Y. Liu,"New preprocessing techniques for handwritten word recognition ",School of Information Technology, Griffith University, 2002