

Handwritten Character Recognition using Artificial Neural Network

Mr. Vaisakh Anil¹ Ms. Twinkle Roy² Ms. Swapna Shakkeer P³ Ms. Stephy Paul⁴ Ms. Sonu Sebastian⁵
^{1,2,3,4,5}UG Student

^{1,2,3,4,5}Department of Computer Science & Engineering
^{1,2,3,4,5}KTU University, India

Abstract— In the present paper, we use the artificial neural network to recognize the character. In this paper it is developed Off-line strategies for the isolated handwritten English character (A TO Z) and (0 to 9) This method improves the character recognition method. Pre-processing of the Character is used Binarization, thresholding and segmentation method. The proposed method is based on the use of feed forward back propagation method to classify the characters. The ANN is trained using the Back Propagation algorithm. In the proposed system, English numerical letter is represented by binary numbers that are used as input then they are fed to an ANN. Artificial Neural network followed by the Back Propagation Algorithm which compromises Training.

Keywords: Neural Network, Back Propagation Method Segmentation, Image Processing Toolbox, Matlab

I. INTRODUCTION

The purpose of this project is to take handwritten English characters as input, then process the character, train the artificial neural network algorithm, to recognize the pattern and modify the character to a beautified version of the input. This project is aimed at developing software which will be useful in recognizing characters of English language. This project is restricted to English characters only. It can also be further developed to recognize the characters of different languages. It engulfs the concept of neural network. One of the primary means by which computers are endowed with human like abilities is through the use of a neural network. Artificial Neural networks are particularly useful for solving problems that cannot be expressed as a series of steps, like recognizing patterns, classifying them into groups, series prediction and data mining.

A artificial neural network trained for classification is designed to take input samples and classify them into groups. These groups can be fuzzy, without clearly defined boundaries. This project concerns detecting freestyle handwritten characters.

II. LITERATURE SURVEY

According to the reference, IJETAE Volume 4, Issue 5, May 2014 this paper explains comparative analysis between Random Transform and Hough Transform, which are applied for error detection and correction. This paper explains implementation of Handwritten Character Recognition in Matlab, compared with current working method of Handwritten Character Recognition. This system achieved recognition rate near about 92%. As per reference, IJSR Publications, Volume 2, Issue 6, June 2012 this paper discusses recognition of offline English character. This explains a model called Hidden Markov Model (HMM) for character recognition. The Novel feature Extraction method is used for implementing HMM. By collecting 13000 samples

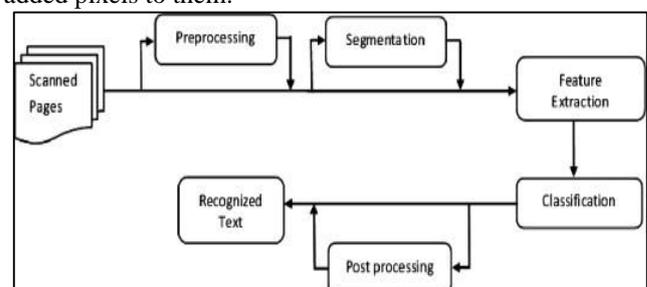
from 100writers they have tested performance of Handwritten Character Recognition technique and got accuracy of near about 94%.

As per reference, IJARECE Volume 2, Issue 5, May 2013 this paper implements the Handwritten Character Recognition technique in Matlab. This paper explains how matlab is more convenient and effective for Handwritten Character Recognition technique. The performance of Handwritten Character Recognition has been tested with samples in this approach.

As per reference, European Academic Research, Volume I, Issue 5/ August 2013 this paper discusses the OCR technique with its components. This achieved a good recognition rate.

III. IMAGE PROCESSING

The first phase in our character recognition process is converting the image to Binary image by thresholding the given character image. Two intensity values are available in binary image. These values are Black and White. We are use zero for Black and one for white. Thus the color of the character is White and the background is black. Preprocessing techniques are needed on color, grey- level or binary document images containing text and/or graphics. In character recognition systems most of the applications use grey or binary images uniform background and/or water marks making it difficult to therefore; the desired result from preprocessing is a binary image containing text only. Thus, to achieve this, several steps are needed, first, some image enhancement techniques to remove noise or correct the contrast in the image, second, thresholding to remove the background containing any scenes, watermarks and/or noise, third, page segmentation to separate graphics from text, fourth, character segmentation to separate characters from each other and, finally, morphological processing to enhance the characters in cases where thresholding and/or other preprocessing techniques eroded parts of the characters or added pixels to them.



IV. SEGMENTATION

Segmentation of hand written text document into individual character or digit is an important phase in document analysis, character recognition and many other areas. Character segmentation has become a crucial step for mail address

recognition in the automatic post mail sorting system. Also out of available text segmentation methods, we do not have a universal accepted solution. The reason for not achieving satisfactory recognition rates is the difficult nature of cursive handwriting and difficulties in the accurate segmentation and recognition of cursive and touching characters. In order to segment text from a given input document image, it is necessary to detect all the possible text regions. In the case of printed scripts, segmentation is a relatively simple task. In the case of overlapped scripts, broken characters, connected characters, loosely configured characters, and mixed scripts, segmentation is difficult. Overlapped, broken, connected and loosely configured characters are major causes of segmentation errors. Segmentation of Text Image is used to locate each individual character and its boundaries. It involves process of labeling, which assigns the some label to spatially align units i.e., pixel, connected components or characteristic points such that a group of pixels with the similar label share specific visual features.

A. Line Segmentation

Line segmentation is the first and a primary step for text based image segmentation. It includes horizontal scanning of the image, pixel- row by pixel-row from left to right and top to bottom. At each pixel the intensity is tested. Depending on the values of the pixels we group pixels into multiple regions from the entire image. The different region indicates different content in the image file. Subsequently the desired content can be extracted. Due to inaccuracies in the scanning process and writing style, the writing may be slightly tilted or within the image. This can hurt the effectiveness of later algorithms and, therefore, should be detected and corrected.

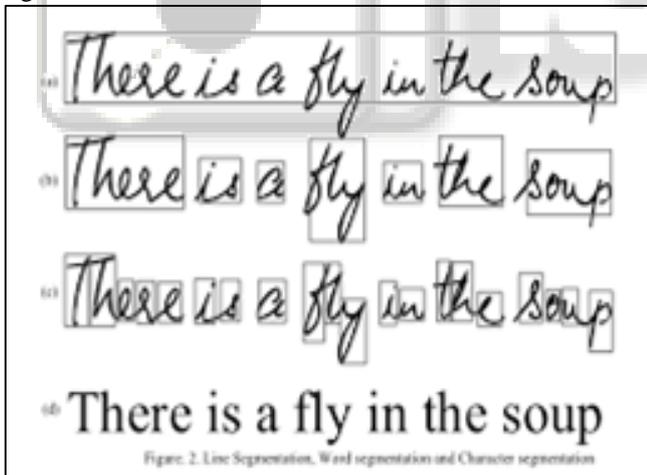
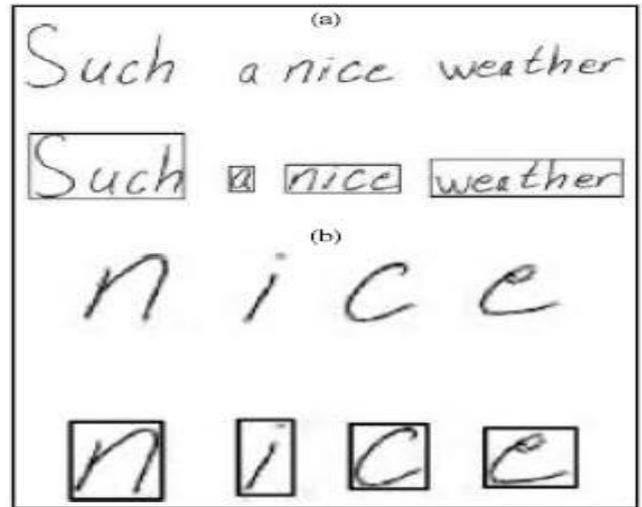


Figure 2. Line Segmentation, Word segmentation and Character segmentation

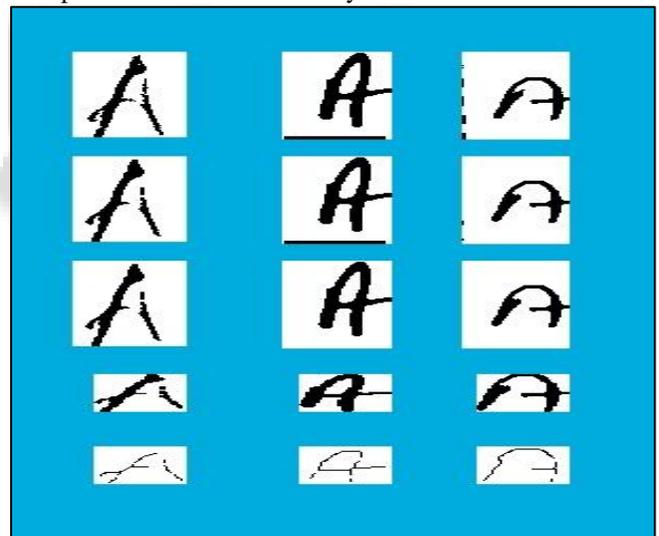
B. Word Segmentation

Word segmentation is the next level of segmentation. It includes vertical scanning of the image, pixel-row by pixel-row from left to right and top to bottom. At each pixel the intensity is tested. Depending on the values of the pixels we group pixels into multiple regions from the entire image. The different region indicates different content in the image file. Subsequently the desired content can be extracted. Slant angle estimation is used to perform skew correction for the extracted word in heavy noise. The skew correction can be performed by determining the angle and rotating the image in the opposite direction.



C. Character Segmentation

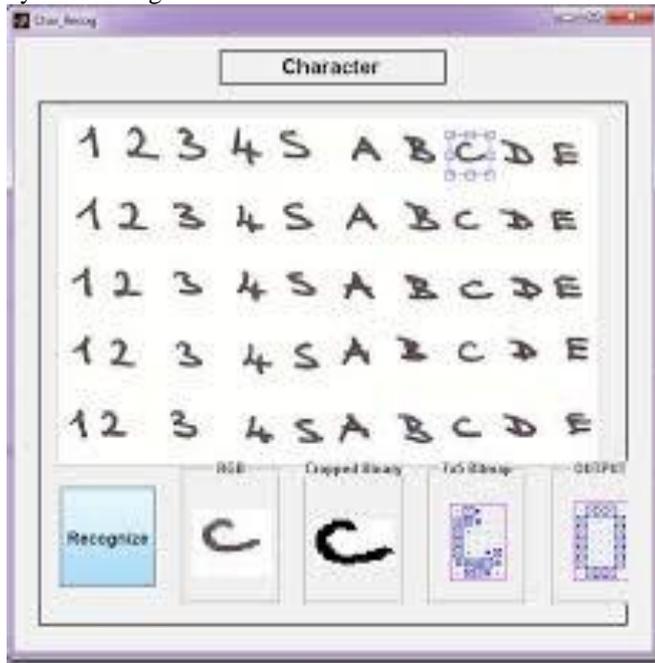
Character segmentation is the final level for text based image segmentation. It is similar to in operations as word segmentation. A few precautions should be followed while performing character segmentation. Figure 2 shows one such problem. The segments as shown in Figure 3c is not accurate, as “h” is extracted as “l” and “i”. Such errors are undesirable. Another precaution is of ligatures. If the text image contains a cursive type font then while segmenting the ligature should be separated for better efficiency.



V. IDENTIFICATION

After the training and testing of the network, the pixelated 7 by 5 sized image of ‘A’ is fed to the network as input. Then the out we get is the resultant 2D matrix plot same as the character ‘A’ from the ideal data-set which was fed to the network as training data-set. Differentiation of characters is done by cropping the boxed characters of the pre-processed image. At first the sub-images are cropped label by label in the sample image, then the character array of image is resized to form a 7X5 matrix pixelated image. This is done because an image array can only be defined with all images of fixed size. Also the size of the character image should be maintained to a size of 7X5 due to the ideal character set is defined as a set of images with 7X5sized 2D matrix with

binary values. For this to be achieved first the images are reshaped to a 7 by 5 aspect ratio image then resized into a 7 by 5 size image.



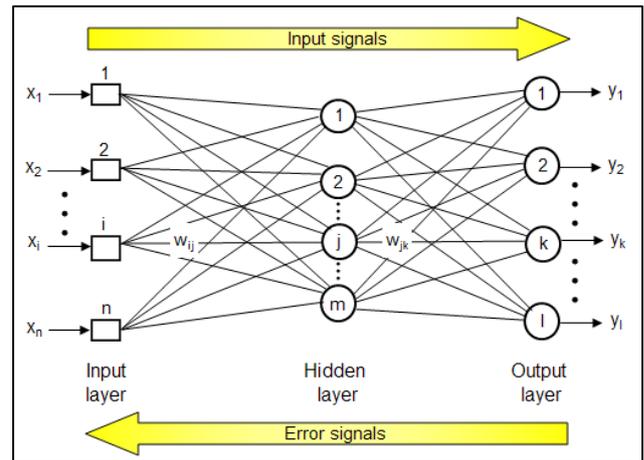
VI. REGENERATION

The input image captured by digital camera is a colored image. Before features are extracted from an image, it may be useful to pre-process the image to reduce irrelevant information or noise and to enhance the image properties that will make feature measurement easier and reliable. There may be random noise that is generated due to different factors such as dirt, dust particles, etc. It can cause significant degradation in the feature extraction process which in turn may lead to higher error rates in the classification process. This noise removal is therefore essential for the system.

VII. TRAINING AND TESTING NETWORK

Next, we create a training vector for the neural network in order to match the input accepted by the neural network function. The steps performed in creating and training the artificial neural network has been illustrated below:

Type tool in Matlab. A dialog box appears are required to Import the Inputs and Targets from the MATLAB workspace. After importing, the created network appears together with the network list. Open the network and select training tab. Here, we can choose the training parameters and data, and finally click on Train option to train the network. I used feed forward back propagation neural network. In other words, implementation based on Multi- Layer Perceptron Network (MLPN) trained with back-propagation was done. The other complex training methods employing Error Back Propagation Algorithm have been used earlier. There were 2 hidden layers used with TANSIG (tan- sigmoid) function. Further, experimental results in the next section illustrated the steps performed.



VIII. NEURAL NETWORK

As the name suggests, Neural is related to neurons, which are an important part of biological nervous system. Like human nervous system processes the information it receives from nerves, in a similarly this Artificial Network processes information to solve specific problems. Every artificial neural network comprises of interconnected neurons which is trained or configured for a specific application. This is widely used in various field of study like pattern recognition, data classification and so forth to analysis a problem and adjust its parameters accordingly. The need of artificial neural network can be realized by comparing it with the use of conventional computers which requires an algorithm to solve a specific problem. Unlike computers, artificial neural networks follow parallel processing architecture thereby resulting in maximum efficiency. Moreover, there are multiple network types like Perceptron, feed forward, feedback networks which present a variable ways to associate input with the output. Artificial Neural Network is not just confined to MATLAB but also suitable for real time systems. It contributes to research in medicine such as neurology to study brain mechanism in detail. The scope of neural network is not just limited to be used alone. It is used in solving Zip Code Recognition problem. It be integrated with other related subjects like Fuzzy logic and Artificial Intelligence for faster response and computations.

IX. CONCLUSION

As presented in the paper, handwritten character recognition using ANN is categorized into three main divisions: Image Acquisition, Image processing and neural network training of the data set. The experimental results illustrate how an input image leads us to character extraction and further usage of artificial neural network recognizes the handwritten patterns, accordingly. The reason for choosing artificial neural networks as a part of the research to perform character recognition is because of their high noise tolerance. The designed systems have the ability to yield the accurate results, provided the correct data set is available at the time of training the network. The current stage of research says that the software does perform well both in terms of speed or accuracy. But the character location is not efficient because of the size of every block varies. This can be taken care of by

initializing the weights during training of data set. There is a scope of improvement the current system.

Hence, a simple yet effective approach for recognition of handwritten characters using artificial neural networks has been described.

REFERENCES

- [1] Character Recognition Using Neural Network Ankit Sharma, Dipti R Chaudhary
- [2] Handwritten Digit Recognition with a Back Propagation Network Y. Le Cun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel
- [3] Character Recognition Using Matlab's Neural Network Kauleshwar Prasad, Devvrat C. Nigam, Ashmika Lakhotiya and Dheeren Umre
- [4] Hand Written Character Recognition Using Artificial Neural Network Vinita Dutt, Sunil Dutt
- [5] Hand-written character recognition using Artificial Neural Network A thesis submitted in partial fulfillment of under graduate program of Bachelor of Technology in Electronics & Instrumentation Engineering
- [6] Feature extraction methods for character recognition-A survey Anil k jain ,torfinn taxt

