Computer Vision Based Text Scanners (Handwritten Character Recognition)

Sakthi J
Final year Student at Prathyusha Engineering College, Department of CSE, Thiruvallur, India.
Soniya G
Final year Student at Prathyusha Engineering College, Department of CSE, Thiruvallur, India.
Thamba Meshach W
Assistant Professor at Prathyusha Engineering College, Department of CSE, Thiruvallur, India.

----------ABSTRACT----------

Humans have unique handwritten styles which makes an obstacles to handwritten character algorithms very difficult. Till to date multiple researches are done on handwritten character recognition to identify the handwritten character styles. In most researches ANN (artificial neural network) has been used which gives an high accuracies. By using real time image processing this system can be implemented to apply multiple handwritten data for schools and universities.

Keywords - MATLAB C, OCR TRAINER, SEGMENTATION, CHRACTER RECOGNITION, ANN

I. INTRODUCTION

Handwriting recognition (or HWR) is the ability of a computer to receive and interpret intelligible handwritten input fromsources suchas paper documents, photographs, touch-screens and other devices. The image of the written text may be sensed "off line" from a piece of paper by optical scanning (optical character recognition) or intelligent word recognition. Alternatively, the movements of the pen tip may be sensed "on line", for example by a pen-based computer screen surface, a generally easier task as there are more clues available.Handwriting recognition principally entails optical character recognition. However, a complete handwriting recognition system also handles formatting, performs correct segmentation into characters and finds the most plausible wordsOff-line handwriting recognition involves the automatic conversion of text in an image into letter codes which are usable within computer and text-processing applications. The data obtained by this form is regarded as a static representation of handwriting. Off-line handwriting recognition is comparatively difficult, as different people have different handwriting styles. And, as of today, OCR engines are primarily focused on machine printed text and ICR for hand "printed" (written in capital letters) text.

II. EXISTING SYSTEM

Pattern Recognition provides the solution to various problems from speech recognition, face recognition to classification of handwritten characters and medical diagnosis. Handwritten Character Recognition involves recognition of handwritten numerals alphabets, symbol etc. In last few years, lots of work has been done by the researchers in the field of handwritten character recognition leading to formulation of efficient least time-consuming classifiers. Most of the times, we easily recognize characters despite the presence of inherent variability in size, slant and styles. But when it comes to implement an unconstrained handwritten character recognition system artificially it is not that much easy. By unconstrained we mean that there are inherent variations in style, thickness and size of the written character.

2.1 Disadvantage of Existing System

- Human errors are prone to arise
- Problem of over fitting

III. PROPOSED SYSTEM

The proposed project is to utilize a web camera to capture images of handwritten alphabets and numbers on the sheet, with an algorithm that has the capability to recognize handwritings and computerized numerals. In short, an intelligent neural network has to be developed for robust handwriting recognition of the numbers from the webcam to be able to be input into a database. The neural network method will be deeply explored to obtain the optimal solution with the lowest overshoot and errors.

3.1 Advantages Of Proposed System

- Implemented to apply multiple handwritten
- Proposed methods also solve the problem of over fitting

IV. METHODOLOGY USED

- Preprocessing
- Character Edge Detection
- Future Extraction
- Classifiers

4.1 Preprocessing

Data pre-processing is an important step in the data mining process. The phrase garbage in garbage out is particularly applicable to data mining machine learning projects. Data-gathering methods are often loosely controlled, resulting in out of range values impossible data combinations, missing values, etc. Analysing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis. Often, data pre-processing is the most important phase of a machine learning project, especially in computational biology.

4.2 Character Edge Detection

Edge detection techniques have therefore been used as the base of another segmentation technique. The edges identified by edge detection are often disconnected. To segment an object from an image however, one needs closed region boundaries. The image includes rich information that is very significant for obtaining the image. Characteristicby object recognition. Edge detection refers to the process of identifying and locating sharp discontinuities in an image. By using canny edge algorithm, we can detect out the edges of a input character.

4.3 Future Extraction

In feature extraction stage each character is represented as a feature vector, which becomes its identity. The major goal of feature extraction is to extract a set of features, which maximizes the recognition rate with the least number of elements. o Due to the nature of handwriting with its high degree of variability and imprecision obtaining these features, is a difficult task. Feature extraction methods are based on 3 types of features:

- Statistical
- Structural
- Global transformations and moments

4.4 Classifiers

k-Nearest Neighbour (k-NN), Bayes Classifier, Neural Networks (NN), Hidden Markov Models (HMM), Support Vector Machines (SVM), etc There is no such thing as the "best classifier". The use of classifier depends on many factors, such as available training set, number of free parameters etc.The output of feature extraction is a feature vector obtained from previous phase is assigned as an input to next phase i.e. classification or class label and recognized by means of supervised and unsupervised method. Here the data set is separated into training and test set for every character.

4.5 Software Requirement

- Operating System: Windows 10 and above.
- Tool: image processing
- Language: MATLAB C

4.5 Hardware Requirement

- 120 GB memory space
- 2 GB RAM (minimum)

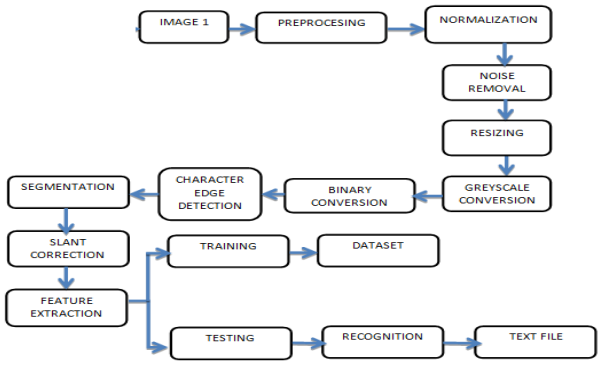

Fig 1: System Architecture

V. MODULE IMPLEMENTATION

Every human has unique handwritten styles which is a difficult task now a days.

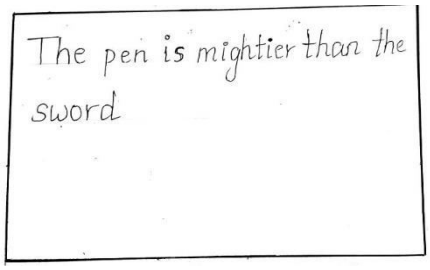

Fig 1: The input should be handwritten where that is scanned by scanner for a high quality of an image.

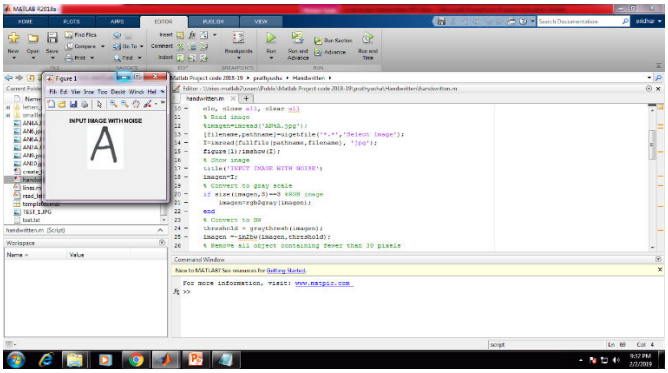

Fig 2: noise removal of a letter A that is trained.

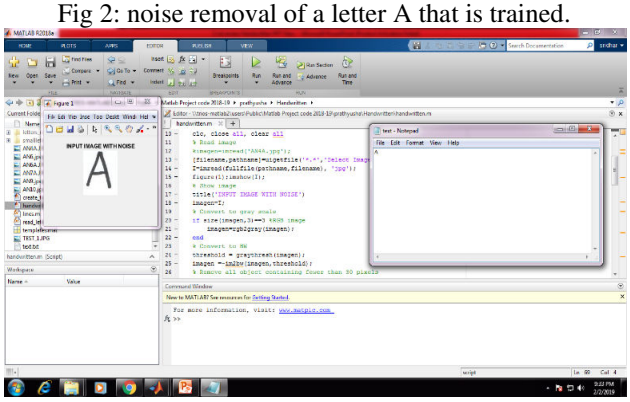

Fig 3: After trained we can run the program to detect the character of handwritten.

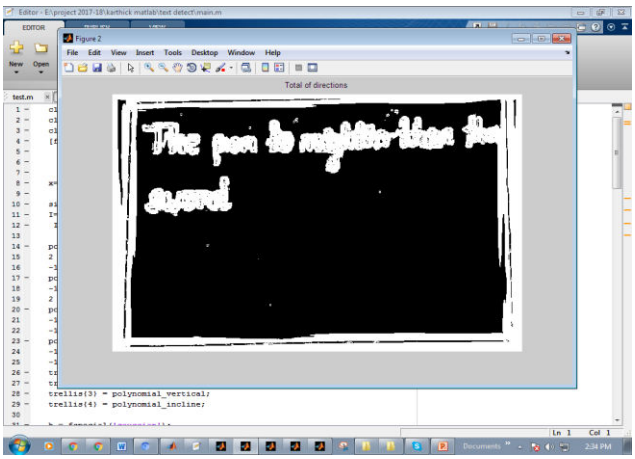

Fig 4: After canny edge detection the input image seems to be above.

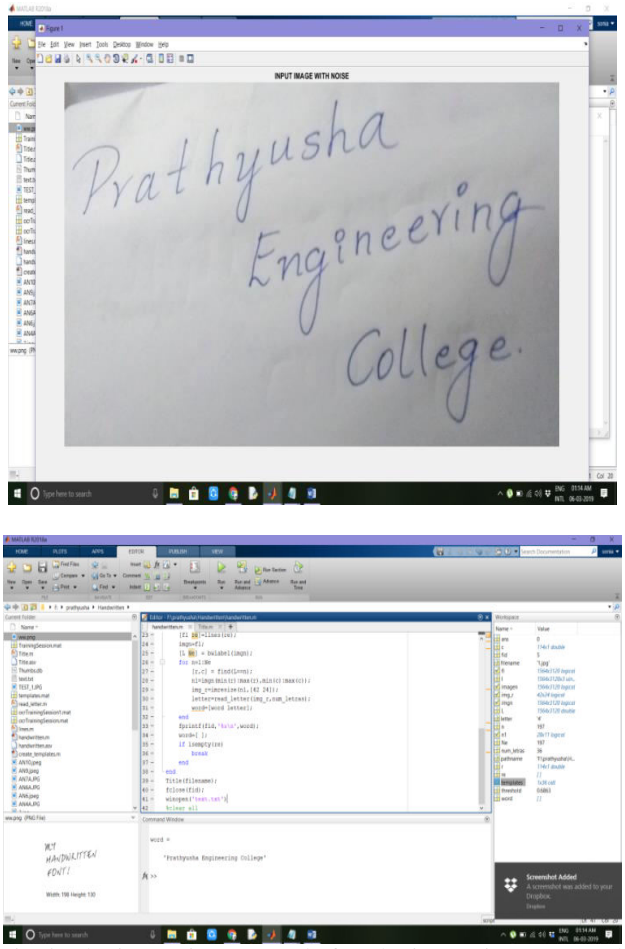

Fig 5: The image is extracted after the image is segmented and trained.

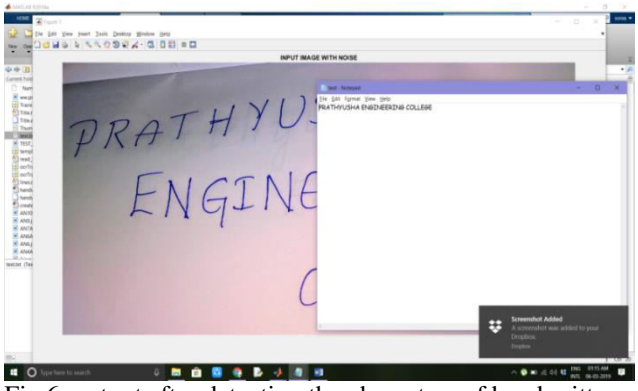

Fig 6: output after detecting the characters of handwritten.

VI. CONCLUSION

Different image preprocessing operations like thresholding, noise removal, and slant correction, resizing and thinning are applied on numeral patterns before presenting them to training network. After preprocessing operations feature extraction is done, by converting image into bit patterns. We presented our data set repeatedly to the network until we achieved the desired efficiency level, and it has been observed that as we repeatedly presented data set to the network, network performance increases. An optimum selection of number of neurons in hidden layer guarantees a good performance. Excessive number of hidden layer neurons leads to overtraining, and an optimum amount helps in better generalization capability of neural network.

REFERENCES

[1] Seema Asht,RajeshwarDass, Dharmendar, "Pattern Recognition Techniques", *International journal of science, engineering and computer technology,* vol. 2, pp. 87-88,March 2012.

[2] C.Y. Suen, Christine Nadal et al.,"*Computer Recognition of Unconstrained Handwritten Numerals," Proc. IEEE,* vol. 80,pp. 1162-1180,July 1992.

[3] Neves,R.F.P. et al.,"A New Technique to threshold the Courtesy Amount of Brazilian Bank Checks," *In Proc. Of 15th IWSSIP, IEEE Press,June 2008.*

[4] AljosaDudarin, ZdenkoKovacic,"Alphanumerical Character Recognition Based on Morphological Analysis," *IEEE 2010.*

[5] C.Y. Suen, M. Berthod, and S. Mori, "*Automatic Recognition of Handprinted Characters –The State of the Art," Proc. IEEE,* vol. 68, pp.469-483, April 1980.

[6] H. Genchiet *al.,"Microprogram Controlled Pattern Processing in a Handwritten Mail Reader-Sorter," Pattern Recognition,* vol. 2,pp. 175-185,1970.

[7] O.D.Trier et al., "*Feature Extraction Methods for Character Recognition-A Survey," IEEE 2008.*

[8] Yanjie Wang et al., "Statistical Modeling and Learning for Recognition-based Handwritten Numeral String Segmentation," Int. conference on document analysis and classification, 2009.

[9] El-Sayed M. El-Alfy," *A hierarchical GMDH-Based Polynomial Neural Network for Handwritten Numeral Recognition Using Topological Features," IEEE 2010*

[10] Amit Choudhary et al., *"Performance Analysis of Feed Forward MLP with Various Activation Functions for Handwritten Numerals Recognition", IEEE Trans.,* vol. 5,pp. 852-856,2010.