Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 7, July 2015, pg.351 – 357

RESEARCH ARTICLE

TO ANALYSIS OF A HAND WRITING RECOGNITION USING K-NEAREST NEIGHBOR(KNN), NEURAL NETWORK (NN) AND DECISION TREE CLASSIFIERS

Dhanwada Sujitha

Electronics and Communication Engineering, Godutai Engineering College For Women Gulbarga, India-585104
sujithad448@gmail.com

Abstract— Handwriting recognition is an important application of forensic science. It mainly finds in application in obtaining the identity of the user from handwriting samples. Every writer has a unique handwriting style, however similar two handwritings may appear and they differ in high dimensional feature space. Role of automated handwriting recognition technique is to leverage the advantage of high dimensional feature space and identify the handwriting pattern.

Several handwriting recognition techniques are being proposed in the past. HMM is most popular of these techniques. HMM relies on log likelihood estimation that checks the input features with every class. Therefore the recognition time is high. In a large practical dataset of thousands of handwritings, HMM becomes very slow. Therefore new models are needed that can offer better accuracy in handwriting recognition with faster recognition time.

In this work we propose a unique handwriting recognition framework that recognizes the handwriting pattern by recognizing the writer of independent words being segmented from a handwriting sample and then detect the user based on maximal match. Further we analyze our technique with three classifiers: K-Nearest Neighbor, Neural Network and Decision Tree classifier and prove that decision tree classifier performs better than the rest in the context of handwriting recognition.

I. INTRODUCTION

Handwriting recognition system performs correct segmentation format and finds the most possible words. Logically the brain sends signal along with the muscles to the implementing the handwritings. The identify handwriting sample, to relevant features of the handwritten script, and interact the features. Understanding the pattern when user is giving input is easier and is called online handwriting recognition. The system identifies the

direction vector set and by analysing the movement on the digitizer, the system can learn the writing style of the user.

Therefore to meet the aforementioned challenges with offline handwriting analysis and writer identification we propose a unique handwriting analysis method with KNN, ANN and Decision Tree Classifier methods.

II. METHODOLOGY

Overview of the System

Handwriting is a pattern. A person handwriting defined by several factors. If you analyze the handwriting a single part will have several features for instance what is the spacing in between characters. So if you have significance spacing you know that handwriting much more clear. Some handwriting will tilt right side, some handwritings will tilted left side. If convert the image into binary image if you column vies. The background image is different. Because of some times use ruled paper, plain paper etc...We can recognize the handwriting first find out the inner pattern. That is if have to remove redundant information. It is defined as information presenting image .it is not useful for the processing purpose. Here color of text and background is depending on writing screen. If analyzing two hand writings pattern we didn't check any things. Whether it is written blue, green, red (or) not. Handwritings should starts with eliminating redundant information.

The front end is designed using Mat lab guide interface for graphical user interface (GUI). The code is driven with the code for event being written in each push button's call-back method. The first step is pre-processing. The current state of art in handwriting recognition treats entire handwriting as a single pattern. The method relies on writing pattern and not the written text. Therefore common features that are used for handwriting recognition is the tilt, pressure, line spacing, character spacing, word spacing, character curvature as the patterns recognizing the handwriting. Therefore forging a handwriting pattern is rather easy because the pattern can be synthesized in a more general context. When we collected hand written samples and scan the process of test. Next Step is filtering. In this system remove the background by first using Gaussian blurring for low pass filtering the image.

The next step is segmented process. In this work each and independent word selected and these are forming a directory it is called as segmented directory. While training process each independent word extracted and saved database along user name or user number. Next process continue to using the classifiers Such as k -nearest-neighbour, neural network and decision tree classifiers put in the training and testing methods.

In this project to test the theory presented earlier and perform the handwriting recognition, desired in this recognizer was implemented in mat lab. The implementation was based on presented chapter 4 in which handwritten instances from individual user are used in the training and testing methods.

The front end is designed using Mat lab guide interface for graphical user interface. The code is event driven with the code for event being written in each push button's callback method.

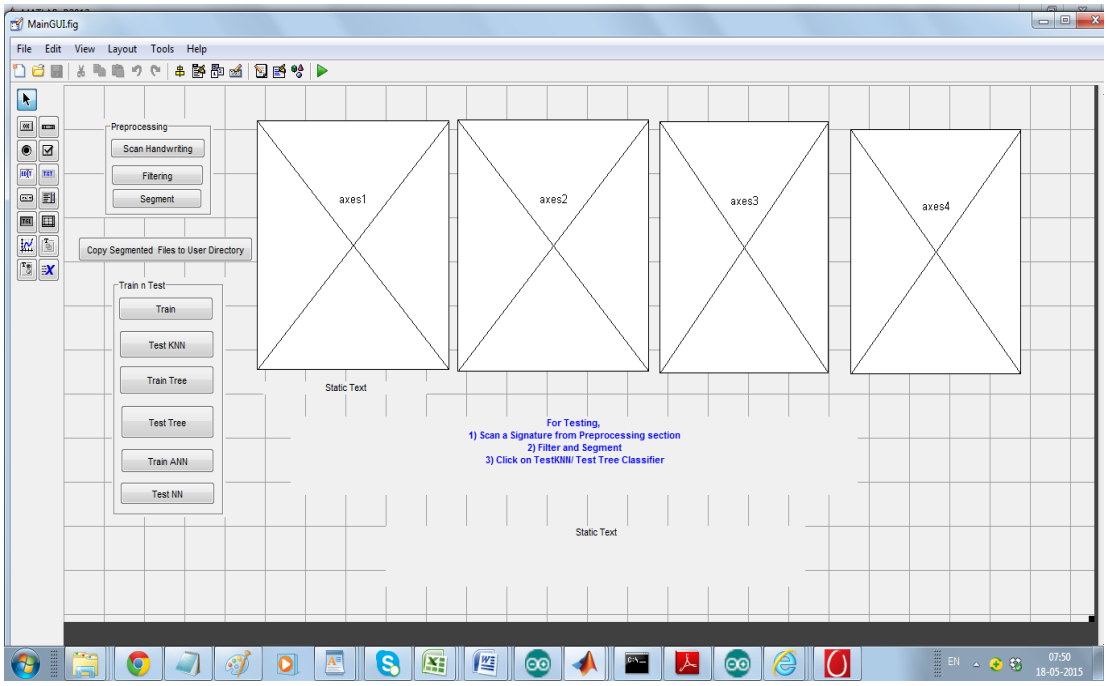


Fig 1 User Interface design

III. DESCRIPTION OF THE WORK

In this work we have taken the hand written samples of 20 users on a plain A4 sheet and scanned it, next step was prepare database for the experiment. so we took the those samples of individual user and scanned then to extract the different lines/instances of that particular user. We have taken 4 or 5 words each line.

In this we have taken only some hand written samples .the first step has scanned handwritten samples each individual user and collect the hand written samples. We create folder directory. in this directory contain 20 users hand written samples. The second next step is filtered and segmentation of each individual users. The third step is training and testing of k nearest neighbour, neural network and decision tree classifiers methods.

Install the GUI .it means graphical user interface method .GUI defined in the previous chapter. then build the Test improve the training and testing methods then finding Accuracy of each individual person using three test methods. This experiment is performed for 1to 20 users with increasing training and testing methods. First the image containing a particular user's scanned document is given as input. Part of user1's handwriting sample is shown in figure 2.

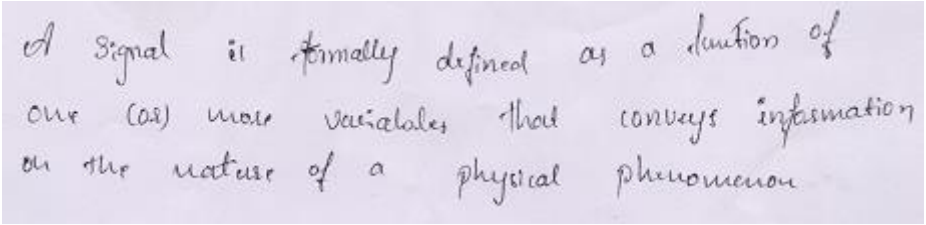


Fig.2 User 1 Handwriting Sample

This is pre-processed using the pre-processing steps defined in the previous chapter. The result of binary filtered document is shown in fig.

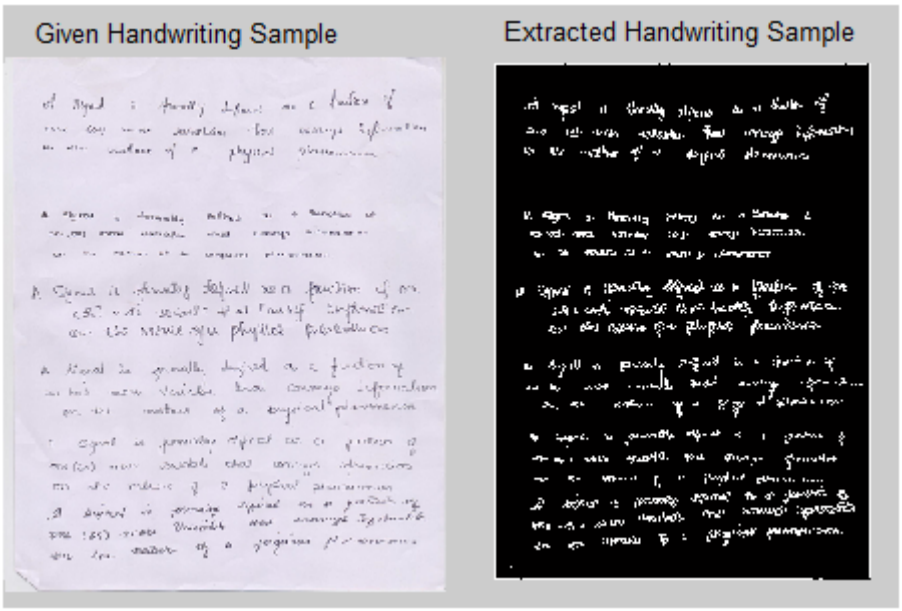


Fig 3 Filtering and Binarization of User 1 Sample

Once the document was binary, pixels with ones, i.e. the written part could be conceived as a group of pixels being linked through the neighbourhood context. As such a bounding box can be obtained across all bounded regions which essentially defines the written words. Bounding boxes are the rectangles which defines the diagonal coordinates of the bounded region. Words can be segmented through these bounded coordinates.

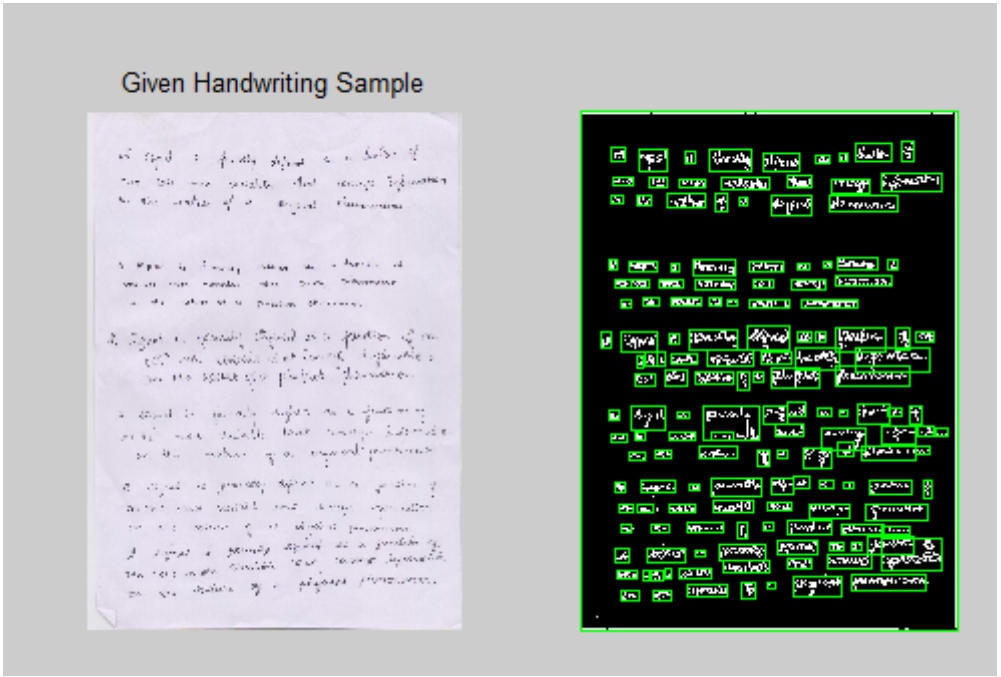


Fig 4 Bounding box across all the words

In training process statistical features from all the images present in dataset folders are extracted. Each user's words are segmented in the folder with names "1","2" for first, second writer respectively. Therefore in training process, the system scans every word stored in 1 directory while training user1's handwriting pattern

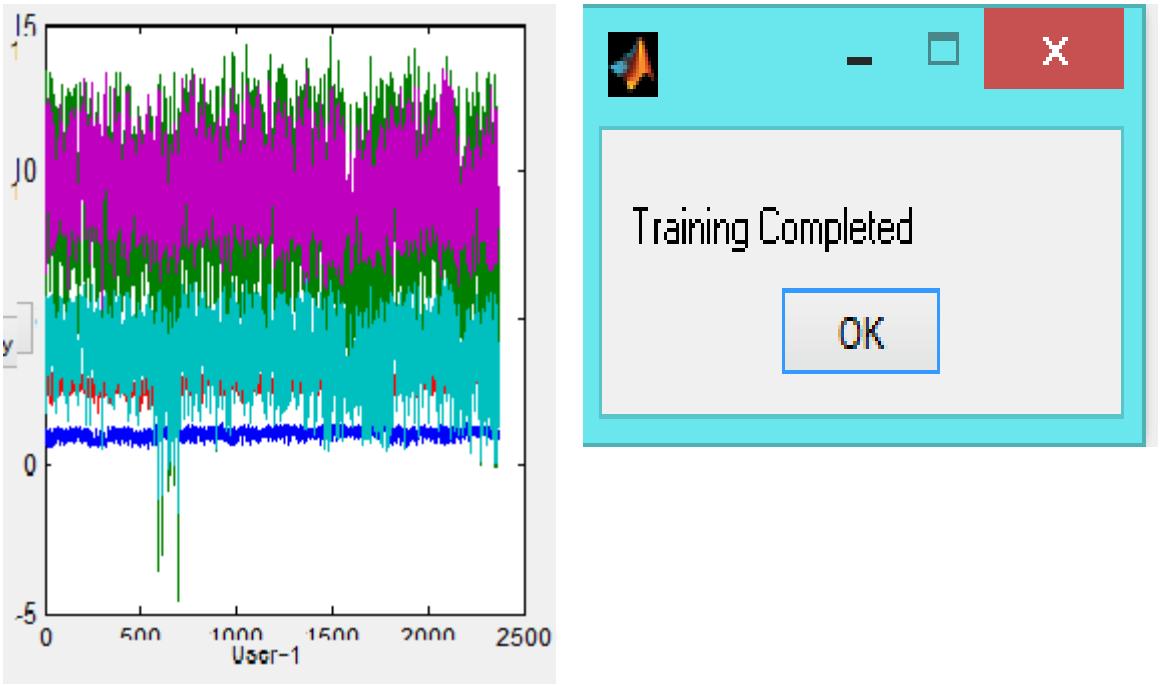


Fig 5 Completion of Training Process

The accuracy is defined as the number of words being successfully classified as being part of a specific user. Therefore if the system scans 50 words written by user one and 40 of these words are classified as being written by user one, then the accuracy for user one will be 80%.

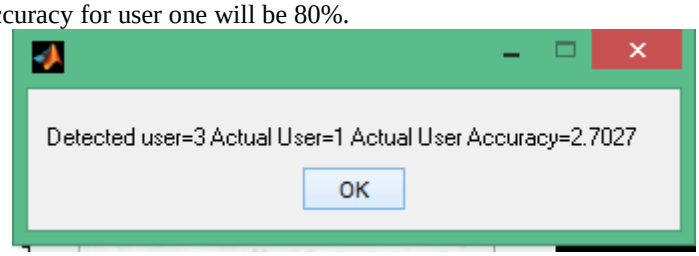
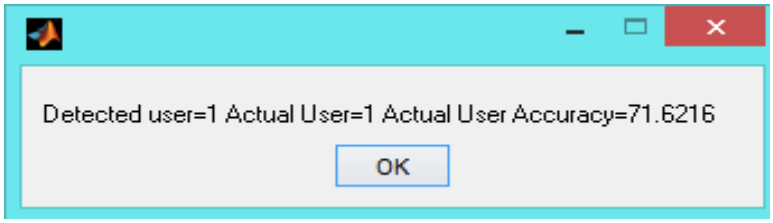
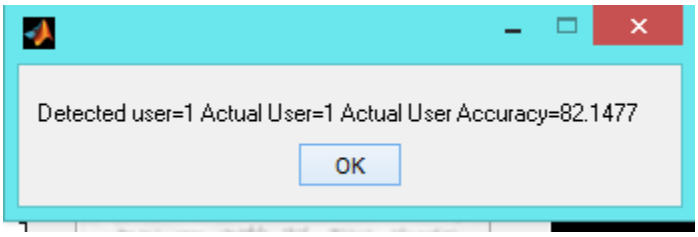


Fig.6 User 3 Handwriting Detection by Neural Network

The same method is followed by both decision tree and KNN classifiers to classify the words to users and then aggregating the detection accuracy. For every user the process is similar. Overall accuracy is measured as the aggregated accuracy for all the users being tested. One more important aspect that needs to be understood is that while testing 50 words of say user 1, if a classifier falsely detects most of the words as user 2 (False recognition rate is greater than correct recognition rate) the user will be falsely recognized as user 2. Once such false recognition produced by neural network is presented in figure 4.6 where we can clearly see that the recognition rate is extremely low. This happens when the network cannot isolate the feature vectors in the model and cannot distinguish between the patterns of two different classes.



(a)



(b)

Fig.7 Testing of User1's Handwriting by a) KNN b) Decision Tree

Comprehensive analysis of user wise recognition is presented in detail in result chapter.

IV. RESULTS AND ANALYSIS

In this project to test the theory presented earlier and perform the handwriting recognition, desired in this recognizer was implemented in mat lab. The implementation was based on presented chapter 4 using this implementation was based on equation presented chapter4using this implementation the hand writings of different users, in which handwritten instances from individual user are used in the training and testing methods.

Strategy 1: user1 state is designed then performs the experiment for different trained instances where different features calculated and improve the feature extraction and passed to user1. The training is completed after that user1 is tested. The performance of the system found to be accuracy 71.62%,84.56%,10.81% for respective trained and tested as shown table 4.1

Table 4.1: User 1 Test

USERS	CLASSIFIERS	DETECTED USER	ACTUAL USER	ACCURACY IN%
1	K-NEAREST NEIGHBOUR	1	1	71.26
1	METHOD(KNN)	1	1	72.48
1	DECISION TREE	5	1	4.05
	NEURAL NETWORK			

V. CONCLUSIONS

In this project we have developed a system for word based handwriting recognition system. Here user's handwriting is first segmented and then trained. We use three different classifiers to test the handwriting of a given word. We detect the writer by selecting the class which is being recognized for most of the user for a given training sample.This is a novel contribution towards handwriting recognition as all past works attempts to recognize handwriting based on a single document. Our handwriting recognition technique adopts voting based classification to select the user with maximum match. A result shows that probabilistic methods suits better for handwriting recognition. Therefore Decision tree classifier outperforms all other classifiers. Performance of neural network was found to be poor as the features are failed to converge because many user's writing patterns are similar. Overall accuracy of Neural Network was found to be 38.6% whereas Decision tree was 88.64% and

KNN was 82.8%. Thus it can be claimed that decision tree classifier is a better suited model for handwriting recognition system. Proposed system can be further improved by combining the classifiers and adding independent sample recognition along with independent word based recognition system.

References

[1] P. Baldi and G. Pollastri. The principled design of large-scale recursive neural network architectures–dagrnns and the protein structure prediction problem. J. Mach. Learn. Res., 4:575–602, 2003.

[2] J. S. Bridle. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In F. Fogleman-Soulie and J.Herault, editors, Neurocomputing: Algorithms,Architectures and Applications, pages 227–236. Springer-Verlag, 1990.

[3] F. Gers, N. Schraudolph, and J. Schmidhuber. Learning precise timing with LSTM recurrent networks.Journal of Machine Learning Research, 3:115–143, 2002.

[4] A. Graves. Supervised Sequence Labelling with Recurrent Neural Networks. PhD thesis.

[5] A. Graves, S. Fern ,Andez, F. Gomez, and J. Schmidhuber. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In Proceedings of the International Conference on Machine Learning, ICML 2006, Pittsburgh, USA, 2006.

