



GURMUKHI CHARACTER RECOGNITION USING NEURO-FUZZY WITH EIGEN FEATURE

GAGANDEEP KAUR¹, MEHAK AGGARWAL²

¹STUDENT IN LLRIET, MOGA

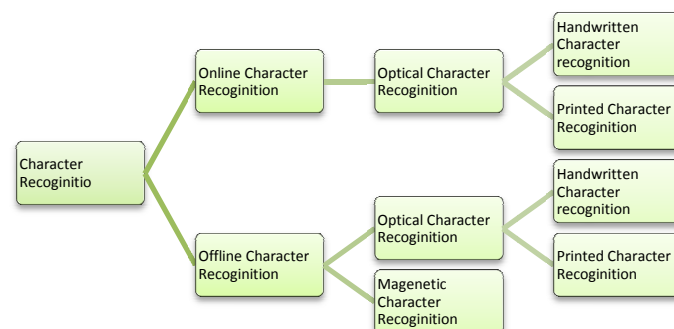
²LECT. AT LLRIET, MOGA

Abstract: Optical character recognition, abbreviated as OCR, is the process of converting the images of handwritten, typewritten or printed text (usually captured by a scanner) into machine editable text or computer process able format, such as ASCII code. Many techniques has been proposed in this field the most of the work done in the field is in the transformation based feature extraction also some of the researcher use the ANN soft computing technique for the classification. In this work Offline character recognition for Gurmukhi lipi has been done using eigen feature. In order to establish correctness 3 different fonts of Gurmukhi lipi has been used. We have done training and testing over multiple samples. In proposed approach a trial has been made to give maximum accuracy in recognition.

Keyword: OCR, ASCII, ANN, PCA

I. INTRODUCTION

Character recognition is a process, which associates a symbolic meaning with objects (letters, symbols and numbers) drawn on an image, *i.e.*, character recognition techniques associate a symbolic identity with the image of a character. Mainly, character recognition machine takes the raw data that further implements the process of preprocessing of any recognition system. On the basis of that data acquisition process, character recognition system can



be classified into following categories

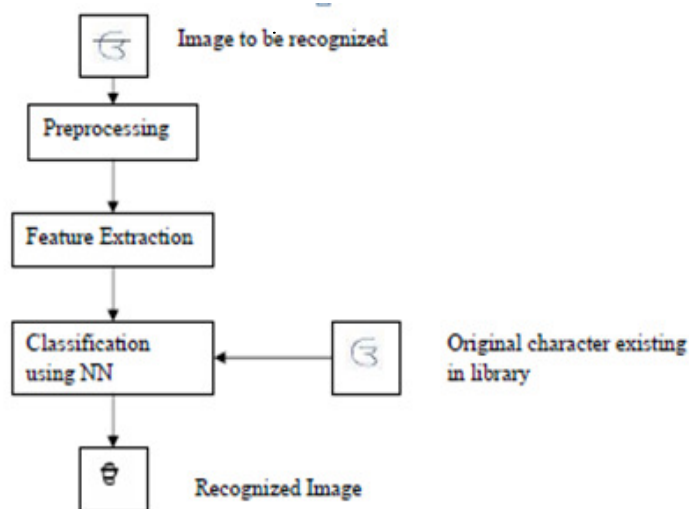
Character recognition can be categorized into following two parts: -

Online Character Recognition

Offline Character Recognition

Off-line handwriting recognition refers to the process of recognizing words that have been scanned from a surface (such as a sheet of paper) and are stored digitally in grey scale format. After being stored, it is conventional to perform further processing to allow superior recognition. In case of

online handwritten character recognition, the handwriting is captured and stored in digital form via different means. Usually, a special pen is used in conjunction with an electronic surface. Gurumukhi script which is mainly used to write Punjabi language, and this consists of 35 characters, in 35 characters 10 vowels, forming 41 consonants including 35 basic characters, 10 modifiers, 6 additional modified consonants in figure 2, 3 sound modifiers (semi-vowels), 3 half characters, which lie at the feet of consonants



II. LITERATURE REVIEW

Madan J. et al, “Development of a generic structural feature extraction method for printed Gurumukhi and similar scripts”,. The problem is to develop a generalized set of features that can be easily applied on the input script like Gurumukhi or similar scripts. An approach had been devised in which the same feature set identified can be applied on various input scripts, rather than adding a new language each time. In this paper, generalized set of structural features had been developed that can be applied on Gurumukhi and similar input scripts. The results of proposed methodology gives better results on characters. The experimental results demonstrate that fonts with reasonably thick stroke and which are made up of clear lines or curves.

Rani S. et al, “Recognition based classification of Gurmukhi manuscripts”, In this paper a recognition based classifier for the Gurmukhi manuscripts into various categories has been proposed. Literature reveals that different degradation styles gives different kind of results. So it becomes difficult to identify such characters. This paper was extremely useful for researchers engaged in recognizing the manuscript documents in any script, because same kinds of problems can be found in manuscript document of any script of the world.

Jindal P. et al, “Line and Word Segmentation of handwritten text documents written in Gurmukhi Script using midpoint detection technique”, In this paper a method for line and word segmentation has been proposed. Proposed work has been implemented on handwritten script. The main aim of this paper was Line and Word Segmentation of text document written in script using mid-point detection technique.. The proposed method was evaluated in terms of accuracy and compared with the existing methods. Two different handwritten scanned images had been taken as input images for line and word segmentation. The results of the proposed method were quite promising.

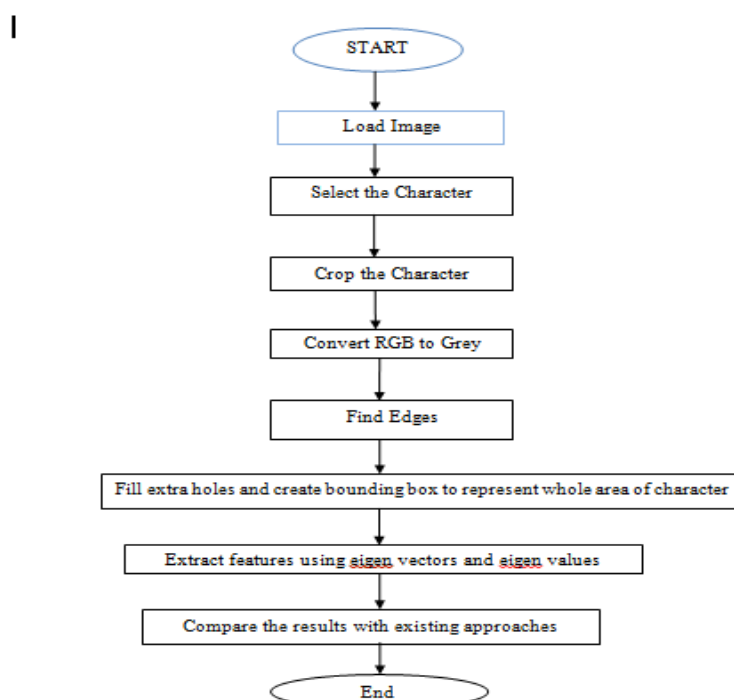
Aggarwal A. et al, “Handwritten Gurmukhi character recognition”, This paper deals with the offline recognition of handwritten Gurmukhi characters. In this two features of character i.e gradient and curvature of character image were computed. The extracted features were then fused together to form a composite feature vector containing both gradient and curvature information. Dimensionality of the generated composite feature vectors was set to 400. The efficiency of these feature sets was tested on a handwritten database of Gurmukhi characters containing 7000 sample character images.

Kaur S. et al, “Gurmukhi Printed Character Recognition using Hierarchical Centroid Method and SVM”, proposed the system for the recognition of printed Gurmukhi character. Hierarchical centroid method is has been used for extracting the feature from images of printed characters. The main advantage of using this method was that it had provided different feature vector and could play important role for manuscript recognition. The dataset used in this study consists of 29 different font styles of the printed characters. The classification has been done by using Support Vector Machine.

III. PROBLEM FORMULATION

Automatic character recognition has been an active research area, which is intended with translating images of characters into a standard encoding scheme representing them. If an image could be converted into editable form, the modification and manipulation of the data for improvement in readability become easier. Many techniques has been proposed in this field the most of the work done in the field is in the transformation based feature extraction also some of the researcher use the ANN soft computing technique for the classification purpose but in the case of the transformation based recognition problem mainly come into work is that the transformed feature value are very less differentiable and cannot be recognizes easily and Also the ANN’s problem is that it only classify well in case of the data already given in training module if any out of range data is there it is not efficient enough to recognize it so need to provide a system which will give better feature extraction and more refined classification.

Proposed methodology works as described in the flow chart below.



IV. RESULTS AND DISCUSSIONS

Result obtained after applying the process are as follows. The image given below shows a sample of inputs feed to the proposed system.

ੳ ਅ ਏ ਸ ਹ ਕ ਖ ਗ ਘ
 ਙ ਚ ਛ ਜ ਝ ਵ ਟ ਠ ਡ
 ਢ ਣ ਤ ਥ ਦ ਧ ਨ ਪ ਫ
 ਬ ਭ ਮ ਯ ਰ ਲ ਵ ਙ

Fig: Set of input Characters

Figure shown below shows the result of selected character with eigen feature

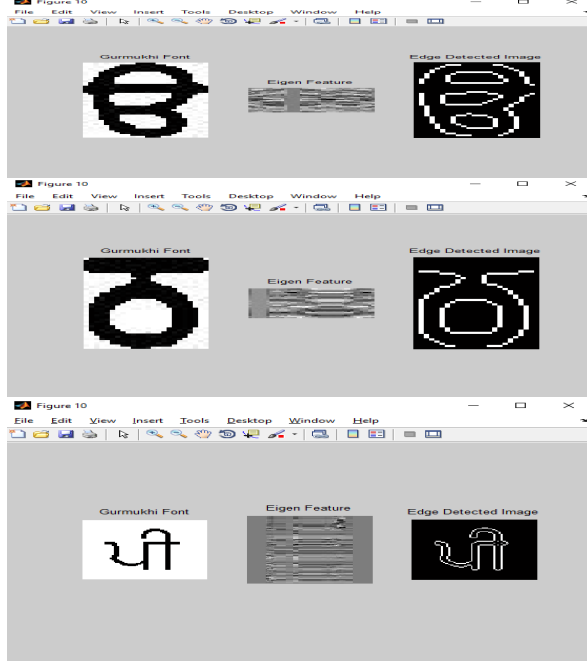
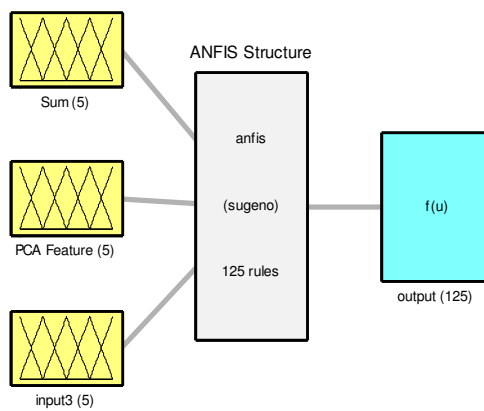


Fig : Extraction of feature for some characters

The figure shown below gives the inference system for the inputs. This is a combined inference system with 125 samples.



System anfis: 3 inputs, 1 outputs, 125 rules

Fig: Inference system of inputs

The graph given below shows the accuracy graph of proposed system

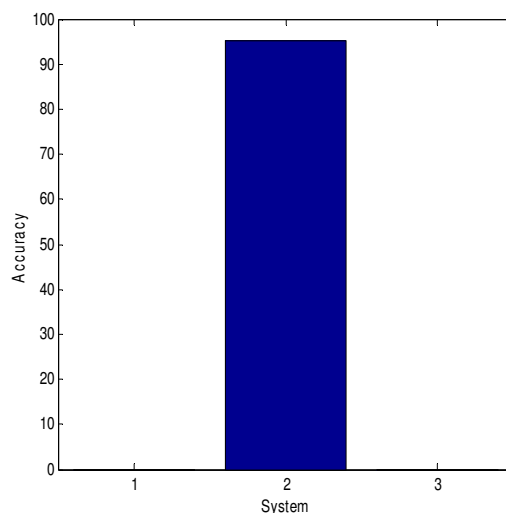


Fig: Accuracy graph of proposed methodology

V. CONCLUSION AND FUTURE SCOPE

Printed Gurumukhi character recognition using neural networks is discussed here. It has been found that recognition of n Gurumukhi characters is a very difficult task. Following are main reasons for difficulty in recognition of Gurumukhi characters :- Some Gurumukhi characters are similar in shape. Characters can be written in different fonts. These facts are justified by the work done here. A small set of all Gurumukhi characters using back propagation neural network is trained, then testing was performed on other character set. The accuracy of network was very low. Then, some other character images in the old character set are added and trained the network using new sets. Then again testing was performed on some new image sets written by different people, and it was found that accuracy of the network increases slightly in some cases. Again some new character images into old character set are added (on which network was trained) and trained the network using this new set. The network is presented new character images and it has been seen that recognition increases, although at a slow rate. The result of the last training by 140 character set and testing with the all characters are presented. It can be concluded that as the network is trained with more number of sets, the accuracy of recognition of characters will increase definitely. Over the past three decades, many different methods have been explored by a large number of scientists to recognize characters. A variety of approaches have been proposed and tested by researchers in different parts of the world, including statistical methods, structural and syntactic methods and neural networks. No OCR in this world is 100% accurate till date. The recognition accuracy of the neural networks proposed here can be further improved. The number of character set used for training is reasonably low and the accuracy of the network can be increased by taking more training character sets. This approach of recognition is used for recognition of Gurumukhi characters only. In future work, this can be implemented for recognition of Gurumukhi words.

REFERENCES

- Singh, G., & Sachan, M. (2015, September). Offline Gurmukhi script recognition using knowledge based approach & Multi-Layered Perceptron neural network. In *Signal Processing, Computing and Control (ISPPCC), 2015 International Conference on* (pp. 266-271). IEEE.
- Grover, S., & Verma, A. (2016, August). Design for emotion detection of punjabi text using hybrid approach. In *Inventive Computation Technologies (ICICT), International Conference on* (Vol. 2, pp. 1-6). IEEE..

- I. Madan, J. S., Sidhu, R., & Sharma, D. V. (2016, March). Development of a generic structural feature extraction method for printed Gurmukhi and similar scripts. In *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on* (pp. 2285-2290). IEEE.
- II. Rani, S., & Lehal, G. S. (2016, March). Recognition based classification of Gurmukhi manuscripts. In *Colossal Data Analysis and Networking (CDAN), Symposium on* (pp. 1-5). IEEE.
- III. Jindal, P., & Jindal, B. (2015, December). Line and Word Segmentation of handwritten text documents written in Gurmukhi Script using mid point detection technique. In *Recent Advances in Engineering & Computational Sciences (RAECS), 2015 2nd International Conference on* (pp. 1-6). IEEE.
- IV. Aggarwal, A., & Singh, K. (2015, September). Handwritten Gurmukhi character recognition. In *Computer, Communication and Control (IC4), 2015 International Conference on* (pp. 1-5). IEEE.
- V. Singh, G., & Sachan, M. K. (2015, December). Data capturing process for online Gurmukhi script recognition system. In *Computational Intelligence and Computing Research (ICCR), 2015 IEEE International Conference on* (pp. 1-4). IEEE.
- VI. Aggarwal, A., Singh, K., & Singh, K. (2015). Use of Gradient Technique for Extracting Features from Handwritten Gurmukhi Characters and Numerals. *Procedia Computer Science*, 46, 1716-1723.
- VII. Siddharth, K. S., Jangid, M., Dhir, R., & Rani, R. (2011). Handwritten Gurmukhi Character Recognition Using Statistical and Background Directional Distribution. *International Journal on Computer Science and Engineering (IJCSE)*, 3(06), 2332-2345.
- VIII. Kaur, S., & Bhatia, R. (2016). Gurmukhi Printed Character Recognition using Hierarchical Centroid Method and SVM. *International Journal of Computer Applications*, 149(3).
- IX. Jindal, A., Dhir, R., & Rani, R. (2012). Diagonal Features and SVM Classifier for Handwritten Gurmukhi Character Recognition. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(5).
- X. Aggarwal, A., Singh, K., & Singh, K. (2015). Use of Gradient Technique for Extracting Features from Handwritten Gurmukhi Characters and Numerals. *Procedia Computer Science*, 46, 1716-1723.
- XI. Lehal, G. S., & Singh, C. (2000). A Gurmukhi script recognition system. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on* (Vol. 2, pp. 557-560). IEEE.
- XII. Kumar, M., Jindal, M. K., & Sharma, R. K. (2011, November). k-nearest neighbor based offline handwritten Gurmukhi character recognition. In *Image Information Processing (ICIIP), 2011 International Conference on* (pp. 1-4). IEEE.
- XIII. Mahto, M. K., Bhatia, K., & Sharma, R. K. (2015, March). Combined horizontal and vertical projection feature extraction technique for Gurmukhi handwritten character recognition. In *Computer Engineering and Applications (ICACEA), 2015 International Conference on Advances in* (pp. 59-65). IEEE.
- XIV. Singh, S., Aggarwal, A., & Dhir, R. (2012). Use of Gabor Filters for recognition of Handwritten Gurmukhi character. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(5).
- XV. Siddharth, K. S., Dhir, R., & Rani, R. (2011). Comparative Recognition of Handwritten Gurmukhi Numerals Using Different Feature Sets and Classifiers. In *Proceedings of International Conference on Image Information Processing (ICIIP 2011)*.
- XVI. Lehal, G. S., & Singh, C. (2000). A Gurmukhi script recognition system. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on* (Vol. 2, pp. 557-560). IEEE.
- XVII. Davessar, N. M., Madan, S., & Singh, H. (2003, October). A hybrid approach to character segmentation of Gurmukhi script characters. In *Applied Imagery Pattern Recognition Workshop, 2003. Proceedings. 32nd* (pp. 169-173). IEEE.
- XVIII. Lehal, G. S., Singh, C., & Lehal, R. (2001). A shape based post processor for Gurmukhi OCR. In *Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on* (pp. 1105-1109). IEEE.
- XIX. Singh, P., & Budhiraja, S. (2012). Handwritten Gurmukhi Character Recognition Using Wavelet transforms. *International Journal of Electronics, Communication & Instrumentation Engineering Research and Development (IJEIERD)*, 2(3), 27-37.
- XX. Kumar, N., & Gupta, S. (2016). Offline Handwritten Gurmukhi Character Recognition: A Review. *International Journal of Software Engineering and Its Applications*, 10(5), 77-86.
- XXI. Kumar, M., Jindal, M. K., & Sharma, R. K. (2011, November). k-nearest neighbor based offline handwritten Gurmukhi character recognition. In *Image Information Processing (ICIIP), 2011 International Conference on* (pp. 1-4). IEEE.
- XXII. Sharma, D., & Jhaji, P. (2010). Recognition of isolated handwritten characters in Gurmukhi script. *International Journal of Computer Applications*, 4(8), 9-17.
- XXIII. Singh, S., Aggarwal, A., & Dhir, R. (2012). Use of Gabor Filters for recognition of Handwritten Gurmukhi character. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(5). Vol. 2, No. 5, Pp. 234-240, May 2012