International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 02 Issue: 04 | July-2015 www.irjet.net p-ISSN: 2395-0072

ZONING DENSITY BASED FEATURE EXTRACTION FOR RECOGNITION OF MARATHI HANDWRITTEN CHARACTER

Ratnashil N Khobragade[1], Dr. Nitin A. Koli[2], Mahendra S Makesar[3]

[1] *Assistant Professor, P G Dept of Comp Sci, SGB Amravati University, Amravati, Maharastra, India*

[2] *Head, Computer Center, SGB Amravati University, Amravati, Maharastra, India*

[3] *Assistant Professor PRMCEM Badnera, Maharastra, India*

Abstract- *In this work we have evaluated the performance of the MLPNN based classifiers for classification of Marathi handwritten. For MLP NN, various learning rules and transfer are investigated for different number of hidden layers and processing elements are set. We used Scale Conjugate Gradient algorithm as default-learning rule. We have used our own database 2150 samples divide in training set of 1720, test set of 215, validation set of 215 samples. The 35 features are calculated from 35 zones and applied to MLP. The 35 features vector of 215 samples is used for testing and 35 features vector of 215 samples is used for validation purpose. Recognition performances of this MLP, is observed on the training, testing and the validation sets, are 97% , 83% and 84 %respectively.*

Key Words: Feature Extraction, Handwritten Devanagari character, Marathi character Recognition, zoning, SVM, Artificial Neural Network.

1. INTRODUCTION

Handwritten character recognition is a frontier area of research for the past few decades and there is a large demand for OCR on handwritten documents. Even now no complete hand written text recognition system is available in Indian scenario and it is difficult due to large character set of Indian languages and the presence of vowel modifiers and compound characters in Indian script. Some reports have appeared for isolated handwritten characters and numerals of a few Indian languages. Majority of them was based on Bangla and Devanagari script.

OCR is the process of converting the scanned images of machine printed or handwritten text, symbols, numerals, letters in to the format which can be processed by the computer such as ASCII. OCR can be classified as on-line or off-line based on the data acquisition process. There are two types of characters handwritten and machine printed. Handwritten characters are non-uniform, there size, shape depends on writer and the pen used by the writer. Handwriting of same writer may vary depending on the situation in which he is writing. OCR is one of the most fascinating and challenging areas of pattern recognition with various practical application potentials. It can contribute to the advancement of an automation process in many applications like reading aid for the blind, automatic text entry into the computer for DTP, ledgering, automatic reading for sorting of postal mail, bank cheques and other documents etc.

1.1 Characteristics of Marathi Script

Marathi script is derived from Devanagari. It is an official language of Maharashtra. It is the 4th most spoken language in India and 15th most spoken language in the world. Marathi script consists of 16 vowels and 36 consonants making 52 alphabets. Marathi is written from left to right. It has no upper and lower case characters. Every character has a horizontal line at the top called as the header line. The header line joins the characters in a word. Vowels are combined with consonants with the help of specific characteristic marks. These marks occur in line, at the top, or at the bottom of a character in a word. Marathi also has a complex system of compound characters in which two or more consonants are combined forming a new special symbol. In India huge volumes of historical documents and books (handwritten or printed in Devanagari script) remain to be digitized for better access, sharing, indexing, etc[13]. The objective of this research is to study the handwritten character recognition and explore a multi-feature multi-classifier scheme for handwritten Marathi characters.

2. LITERATURE REVIEW

Recognition of the text like human is still a challenging task for machine. The off-line Handwritten Character Recognition which requires more research to recognize the text. Large number of researchers worked on English language since six decades but for Indian languages it is still a dream. Recognition of handwritten characters has been a popular research area for many years because of its various application potentials. The detail literature review for the development in optical

character recognition with respective to Devanagari Handwritten character and Marathi Manuscript such as image preprocessing, segmentation, feature extraction, neural network classifiers and their implementation etc. have been discussed here.

In [1] M. Hanmandlu et al. presents the modified exponential membership function fitted to the fuzzy sets for recognition of handwritten Hindi characters based on features consisting of normalized distances obtained using the Box approach. The normalized distance used as a feature is found to be effective. Sandhya Arora et al. suggested combine multiple feature extraction techniques for handwritten devnagari character recognition [2]. Four feature extraction techniques namely, intersection, shadow feature, chain code histogram and straight line fitting are used to extract different features. On experimentation with a dataset of 4900 samples and obtained 92.80% overall recognition rate. Satish Kumar[3] suggested three tier strategy to recognize the hand printed characters of Devanagari script using multiple features and multi-stage classifier. The recognition rate of 94.2% is achieved with this scheme on database consisting of more than 25000 characters belonging to 43 alphabets. Invariant Moments for handwritten devanagari vowels recognition presented by R. J. Ramteke [4], is independent of size, slant, orientation, translation and other variations in handwritten vowels. Ten samples of each vowel from 25 people have been sampled and a database was prepared with normalized image to 40X40 pixel size. The Fuzzy Gaussian Membership function for classification gives 94.56 rate of success. In [5] O V Ramana Murthy, M Hanmandlu considered the character image divided into predefined number of zones and a feature is computed from each of these zones based on the pattern (black) pixels contained in that zone. Some of such features are sum squared distance, histogram average pixel density. In [6] Shailendra Kumar Shrivastava and Pratibha Chaurasia use the energy features of segment characters for the classification of Devanagari character using SVM. The best result obtained in DATASET1 Linear kernel 96%, Quadratic kernel 100%, RBF kernel 97%, and polynomial kernel 100%. Dinesh V. Rojatkar et. al. investigates LRTB feature based classifier using single hidden layer feed-forward neural network with five fold cross validation applied to handwritten Devanagari consonant characters in [7]. They found best network at fold 5 with 80 neurons at trial 3. Networks analyzed on account of confusion matrix, reveals the greater details for individual classes. Average classification accuracy on training, validation, test and combined dataset is 99.40%, 97.38%, 97.05% and 98.98% respectively on the total dataset size of 8224 samples distributed uniformly within 32 classes of typical Devnagari consonants. In [8] Sushama Shelke, Shaila Apte presents a novel approach for recognition of unconstrained handwritten Marathi characters using multistage feature extraction and classification scheme. and achieved 95.40% recognition rate. Zoning based feature extraction is propose by O. V. Ramana Murthy and M. Hanmandlu [9] in which character image is divided into predefined number of zones and a feature is computed from each of these zones. In [10] optimal classifier for the categorization of handwritten Marathi consonant characters using a single hidden layer feed-forward neural network with five fold cross validation is proposed by D V Rojatkar et. al. And obtained Overall, classification accuracy on training, validation, test and combined dataset is 99.58%, 97.88%, 97.62% and 99.05% respectively on the total dataset size of 8224 samples distributed uniformly within 32 classes of typical Devnagari consonants. A detail survey of preprocessing, segmentation, feature extraction, classification and matching techniques for optical character recognition of general scripts presented by Ratnashil N Khobragade et. al. in [11]. In [12] Compound characters are one of the features of Marathi script. Mrs. Snehal S. Golait, Dr. L.G. Malik present a short review on feature extraction for Marathi handwritten compound character recognition.

3. THE PROPOSED APPROACH

The generalized data flow diagram for Marathi Handwritten Character Recognition is shown in figure 5.1.

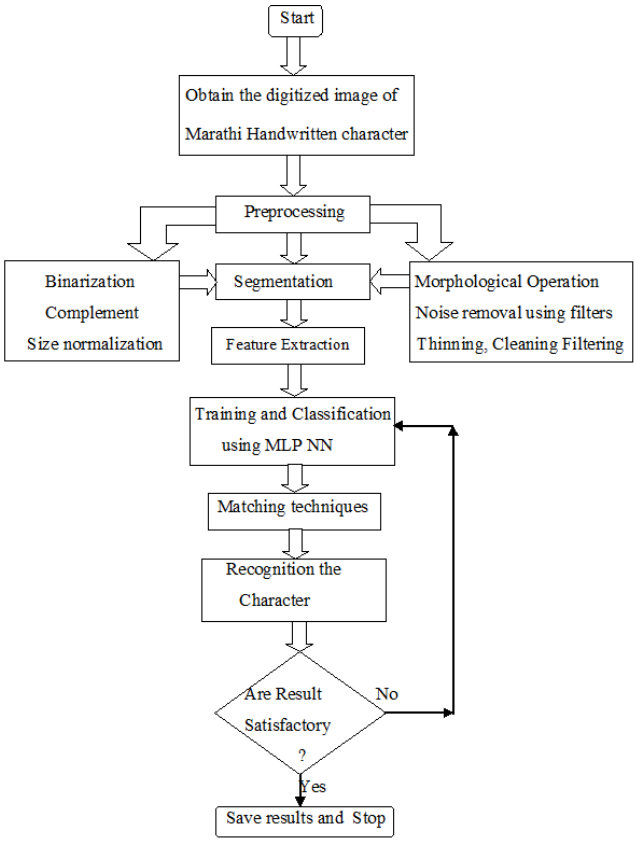


Figure 5.1 Block diagram for Handwritten Marathi character recognition (HMCR)

4. THE EXPERIMENTAL SETUP AND RESULT

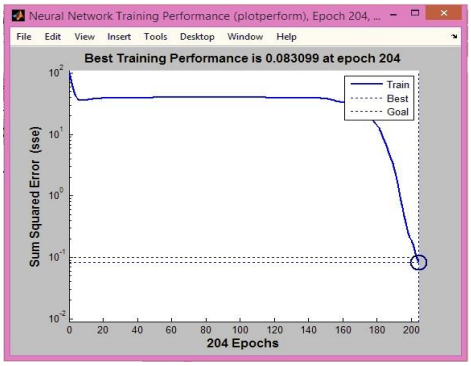
It is important to design and develop a feature extraction technique for the successful recognition of handwritten Marathi characters and to design a suitable NN based classifier. Here we describe results in detail based on zoning density based feature extraction technique developed for Marathi Handwritten Character Recognition. Performance are evaluated and tested. The experimentation of proposed approach is carried out by using Matlab11.0 with core i3 processor and other required minimum configuration.

4.1 Data collection

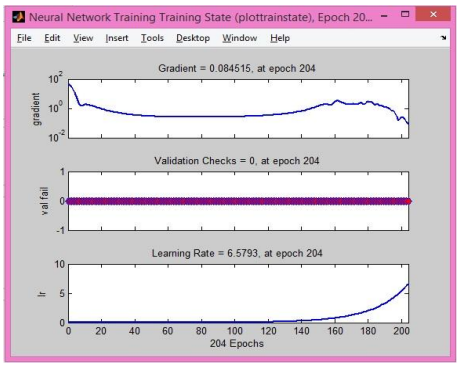
We have used our own database and some other database for the experimentation. For the preparation of the training, testing and validation sets, a database of 2150 samples of 43consonents symbols of Marathi character is formed from each of 50 people of different age groups and sexes. It is optically scanned by scanner by 300 dpi. A training set of 1720 samples, test set of 215 samples and validation set of 215 samples are then formed through random selection of character samples of each class from the initial database. All these samples are scaled to 7x5 pixel images then converted to gray image and then binary images through thresholding.

Zoning based feature extraction is one of the most popular methods in character recognition. The character image is divided into predefined number of zones and a feature is computed from each of these zones. This feature is based on the pattern of black pixels contained in that zone. The character image is divided into 7x5 zones. We sum up all the pixels in one zone called it pixel density in that zone this become a feature. Total 35 features are calculated from 7x5 i.e. 35 zones.

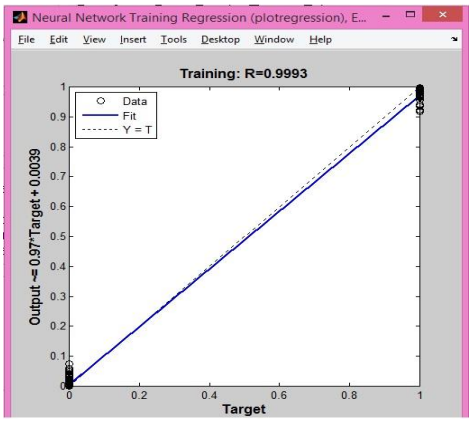
In this work, MLP with one hidden layer is chosen. This is mainly to keep the minimum computational requirement without affecting its function approximation capability. The 35 features vector of 1720 samples is used for training the MLP. The 35 features vector of 215 samples is used for testing and 35 features vector of 215 samples is used for validation purpose. Extensive experiments are performed on the dataset. The result shows the significant improvement over the other method of character recognition.



a)



b)



c)

Figure 1 a) Training performance verses Sum Squared Error graph, b) Neural Network Training State Performance Graph, c) Neural Network Training Regression Graph.

Recognition performances of this MLP, as observed on the training, testing and the validation sets, are 97% , 83% and 84 %respectively. Some samples of experimental results are shown in Figure 1 a, b and c.

5. CONCLUSION AND FUTURE SCOPE

In this work we have evaluated the performance of the MLPNN based classifiers for classification of Marathi handwritten characters and results are found to be satisfactory. For MLP NN, various learning rules and transfer are investigated for different number of hidden layers and processing elements are set. We used Scale Conjugate Gradient algorithm as default-learning rule. We have used our own database 2150 samples divide in training set of 1720, test set of 215, validation set of 215 samples. The 35 features are calculated from 35 zones and applied to MLP. The 35 features vector of 215 samples is used for testing and 35 features vector of 215 samples is used for validation purpose. Recognition performances of this MLP, is observed on the training, testing and the validation sets, are 97% , 83% and 84 %respectively.

The features extraction could be further change and analyze to reduce the dimensionally and computational complexity. The work can be extended for recognition of words and complete sentences for Marathi manuscript. The method could be implemented for the other scripts.

REFERENCES

[1] M. Hanmandlu, O.V. Ramana Murthy Vamsi Krishna Madasu, "Fuzzy Model based recognition of handwritten Hindi characters", Digital Image Computing Techniques and Applications, 0-7695-3067-2/07 $25.00 IEEE 2007.

[2] Sandhya Arora et.al., "Combining Multiple Feature Extraction Techniques for Handwritten Devnagari Character Recognition", Region 10 Colloquium and the Third ICIIS, Kharagpur, INDIA December 8-10, IEEE 2008.

[3] Satish Kumar, "Devanagari Hand-printed Character Recognition using Multiple Features and Multi-stage Classifier", International Journal of Computer Information Systems and Industrial Management Applications IJCISIM ISSN: 2150-7988 Vol.2, pp. 039-055, 2010.

[4] R. J. Ramteke, " Invariant Moments Based Feature Extraction for Handwritten Devanagari Vowels Recognition", International Journal of Computer Applications 0975 - 8887 Volume 1 – No. 18, 2010.

[5] O V Ramana Murthy, M Hanmandlu, "Zoning based Devanagari Character Recognition", International Journal of Computer Applications 0975 – 8887 Volume 27– No.4, August 2011.

[6] Shailendra Kumar Shrivastava, Pratibha Chaurasia,"Handwritten Devanagari Lipi using Support Vector Machine", International Journal of Computer Applications (0975 – 8887) Volume 43– No.20, April 2012.

[7] Dinesh V. Rojatkar, Krushna D. Chinchkhede, G.G. Sarate , "Design and Analysis of LRTB feature based Classifier applied to Handwritten Devnagari Characters : A Neural Network Approach", International Conference on Advances in Computing, Communications and Informatics ICACCI pp. 97-101, 2013.

[8] Sushama Shelke, Shaila Apte, "A Novel Multi-feature Multi-Classifier Scheme for Unconstrained Handwritten Devanagari Character Recognition", 12th International Conference on Frontiers in Handwriting Recognition 2010.

[9] O. V. Ramana Murthy, M. Hanmandlu, "Zoning based Devanagari Character Recognition", International Journal of Computer Applications 0975 –8887 Volume 27–No.4, pp 21-25, August 2011.

[10] Dinesh V. Rojatkar, Krushna D. Chinchkhede, G.G. Sarate, "Handwritten Devnagari Consonants Recognition using MLPNN with Five Fold Cross Validation" International Conference on Circuits Power and Computing Technologies ICCPCT pp. 1222-1226, IEEE 2013.

[11] Ratnashil N Khobragade, Dr. Nitin A. Koli, Mahendra S Makesar, "A Survey on Recognition of Devnagari Script", International Journal of Computer Applications & Information Technology, Vol. II, Issue I, IJCAIT January 2013.

[12] Mrs. Snehal S. Golait, Dr. L.G. Malik," Review on Feature Extraction Technique for Handwritten Marathi Compound Character Recognition", 2013 Sixth International Conference on Emerging Trends in Engineering and Technology 978-1-4799-2560-5/13, IEEE 2013.

[13] Ratnashil N Khobragade, Dr. Nitin A. Koli, Mahendra S Makesar, "Analysis of Methods for Recognition of Devnagari Script", International Journal of Pure and Applied Research in Engineering and Technology, Volume 2 (8) pp. 27-38, IJPRET, 01-04-2014.

BIOGRAPHIES

Ratnashil N Khobragade is Assistant Professor in PG Dept of Computer Science, SGB Amravati university, Amravati. He is pursing Ph.D. in Computer Science/Engg., completed M.E. in year 2009 and B.E. in 2002. He has more than 12 years of experience in teaching and research. Throughout his teaching career he has taught more than 20 subjects at various UG and PG level courses. At UG and PG level he has guided more than 30 projects/dissertation. He has published 18 papers in International & National level Journals and also International Conferences and National level Conferences.

Mahendra S Makesar has done M.E. in Computer science and engineering and persuing his Ph.D from Sant Gadge Baba Amravati University, Amravati. His areas of interest are Image processing, multimedia system and DBMS. He is research scholar at SGBAU, Amravati.

Dr. Nitn A Koli is a Head, Computer Centre, Sant Gadge Baba Amravati University Amravati. He has published many research papers in national and international journals. His areas of interest are image processing, networking and DBMS.

