FEATURE EXTRACTION AND SELECTION FOR HANDWRITTEN DEVANAGARI VOWELS RECOGNITION

P E Ajmire[1], R V Dharaskar[2], V M Thakare[3]

[1]*G. S. Science, Arts & Commerce College, Khamgaon*
[2]*Director, DES(DMAT-Disha Technical Campus), Raipur*
[3]*HoD, Dept. of Comp.Sci, P.G.T.D, S.G.B.A.U. Amravati (M.S.), India*

Abstract

This paper deals with feature extraction and selection algorithm for Handwritten Devanagari Vowels Recognition (HDVR). The various techniques of feature classifications for Handwritten Devanagari Vowels Recognition (HDVR) are employed for this task. The overall accuracy for support vector machine 92.11% .

Keywords— *Devanagari, Handwritten, Vowel, Features, SVM*

----------***----------

1. INTRODUCTION

Many researchers are working in this area of Image Processing and still handwritten Devanagari character recognition is challenging task. The name Devanagari comes from the Sanskrit words Deva (god), and Nagari (city); together they mean, literally, the script of the "City of the Gods", where this city is the body of the individual [1]. Devanagari script is used for Hindi, which is official language of the Union Government of the Republic of India. Devanagari script consists of 16 vowels and 36 consonants making a total of 52 letters. These vowels are classified in two groups. The first group consists of 12 vowels, which are as follows.

अ आ इ ई उ ऊ ए ऐ ओ औ अं अः

Fig. 1: First Group of 12 Devanagari Vowels.

The second group consists of the 4 vowels:

ऋ, ॠ, A\, ऑ'

Fig.2: Second Group of 4 Devanagari Vowels.

There are 13 Vowels of Devanagari script that are commonly used in Marathi and Hindi. These vowels are rarely found in the word, some of which are tongue-twister. The vowel 'ऋ' still finds use in words like ऋIषI (sage), ऋतU (season) etc. In Hindi and particularly in Marathi the vowels plays important role to design meaning full words. There are some words, which contains only vowels e.g Aa¡. These vowels are used as modifiers. Table I shows some vowels are used as modifiers.

Table 1: VOWELS AS MODIFIERS

A	**e**
Aa	**a**
š	**i**
¡	**I**
£	**u**
¤	**U**
Ao	**e**
AO	**W**
A'	**'**
A"	**"**

2. RELATED WORK

Kaustubh Bhattacharyya at el. [2] proposed ANN-based Innovative Segmentation Method for Handwritten text in Assamese. The ANN trained with different character shapes used as a decision making tool while selecting segmentation boundaries. As the segmentation boundary is fixed dynamically, the system can deal effectively with segmentation problems that suffer due to written induced variations in size and shape of characters without inclinations.

C N Mahender at el [3] proposed a structured based feature extraction of handwritten Marathi word. Authors applied rule based recognition and achieved 85% to 90% recognition rate. Ved Prakash Agnihotri [4] presented a system using neural network. Diagonal based feature extraction is used for extracting features and extracted 54 features of the handwritten Devanagari script. After that these feature of each character image is converted into chromosome bit

string of length 378. More than 1000 sample is used for training and testing purpose in this proposed work. It is attempted to use the power of genetic algorithm to recognize the character. The training set contains 904 characters and testing set contains 204 characters. The precision of offline Devanagari system is 85.78%. Joshi, N at el describe a system for the automatic recognition of isolated handwritten Devanagari characters obtained by linearzing consonant conjuncts. Authors [5] use structural recognition techniques to reduce some characters to others. The residual characters are then classified using the subspace method. The proposed system is evaluated for the writer dependent scenario. Brijmohan Singh at el [6] uses two different methods for extracting features from handwritten Devanagari characters, the Curvelet Transform and two different classifiers, viz., the Support Vector Machine (SVM) with Radial Basis Function (RBF), and the k-Nearest Neighbour (k-NN) classifier are used for recognition of handwritten character. The k-NN classifier performs the best with accuracy 93.8%.

2. FEATURE EXTRACTION

Handwritten database is design for this work. For this task the person from different age, gender, educational qualifications are selected. After pre-processing [7-8], character image is normalised to 40 x 40 pixel size. Histogram oriented gradient features are extracted [9]. These features are reduced to 63 from 81 with selection algorithm.

3. CLASSIFICATION

Support vector machine is a supervised classification; basically it is a statistical classification. It is well known as binary classifier [10]; it can be adopted for multiple class tasks. The support vectors machine is effective on high dimensional data as it supports vectors lie closest to the decision boundary. SVM with a small number of support vectors can have good generalization, even when the dimensionality of the data is high [11]. The data is distributed as 60:20:20 i.e. 60% for Training, 20% for Cross Validation and remaining 20% for Testing Classification Accuracy of SVM for all vowels along with Mean Square Error, Mean Absolute Error, Min and Max absolute Error is computed and it is given in Table II. The confusion Matrix for all vowels is given in Table III

Table 2: Performance of All Vowels

Output	Mean Squared Error	Mean Abs Error	Min Abs Error	Max Abs Error	Linear Corr. Coeff.	**%age of Accuracy**
A	0.0573	0.1644	0.0025	1.0233	0.4942	66.67
A'	0.0428	0.1475	0.0003	0.6786	0.6858	83.33
¤	0.0459	0.1595	0.0042	0.6414	0.7243	85.71
Ao	0.0439	0.1545	0.0031	0.7013	0.8049	88.89
Ee	0.0378	0.1514	0.0009	0.5763	0.7634	100.00
A"	0.0321	0.1385	0.0027	0.4810	0.6629	100.00
£	0.0609	0.1825	0.0060	1.0179	0.7251	88.89
Aa	0.0345	0.1434	0.0012	0.5347	0.7247	100.00
©	0.0461	0.1678	0.0007	0.6199	0.8351	100.00
AO	0.0394	0.1517	0.0053	0.7272	0.7381	100.00
E	0.0421	0.1582	0.0018	0.5715	0.7043	100.00
š	0.0331	0.1449	0.0058	0.4730	0.7587	83.33
¡	0.0691	0.2066	0.0019	0.7871	0.5498	80.00
Avg.	0.0450	0.1593	0.0028	0.6795	0.7055	90.53

Table 3: Confusion Matrix

Output	A	A'	¤	Ao	Ee	A"	£	Aa	©	AO	E	š	¡
A	**4**	0	0	0	0	0	0	0	0	0	0	0	0
A'	1	**5**	0	0	0	0	0	0	0	0	0	0	0
¤	0	0	**6**	0	0	0	1	0	0	0	0	0	0
Ao	0	0	0	**8**	0	0	0	0	0	0	0	0	0
Ee	0	0	0	0	**4**	0	0	0	0	0	0	0	0
A"	1	1	0	0	0	**4**	0	0	0	0	0	0	0
£	0	0	1	0	0	0	**8**	0	0	0	0	0	0
Aa	0	0	0	0	0	0	0	**4**	0	0	0	0	0
©	0	0	0	0	0	0	0	0	**9**	0	0	0	0
AO	0	0	0	1	0	0	0	0	0	**4**	0	0	0
E	0	0	0	0	0	0	0	0	0	0	**5**	0	0
š	0	0	0	0	0	0	0	0	0	0	0	**5**	1
¡	0	0	0	0	0	0	0	0	0	0	0	1	**4**

4. CONCLUSION

The histogram oriented gradient is used to detect the object in three dimensional environment (3-D). Here it is attempt to use these features for 2-D image of character for handwritten Devanagari vowels recognition. The statistical classifier SVM is used for the classification. The overall accuracy of recognition of handwritten Devanagari vowels is given in Table IV.

Table 4: Overall Accuracy of Recognition

SVM Classifier	
On	Accuracy
Training	100
Cross Validation	85.82
Testing	90.52
Overall Accuracy	**92.11**

REFERENCES

[1] Online encyclopedia, “Omniglot”, of writing systems and languages.

[2] K. Bhattacharyya and K. K. Sarma, "ANN-based Innovative Segmentation Method for Handwritten text in Assamese", International Journal of Computer Science Issues, Volume 5, pp.9-16, October 2009.

[3] C N Mahender and K V Kale, “A Structured based Feature Extraction of Handwritten Marathi Word”, International Journal of Computer Applications, Vol. 16, No.6, February 2011.

[4] Ved Prakash Agnihotri, “ Offline Handwritten Devanagari Script Recognition”, Information Technology and Computer Science, Vol. 8, pp. 37-42, 2012.

[5] Joshi, N, Sita G, Ramakrishnan A.G and Deepu V and Madhvanath S, " Machine recognition of online handwritten Devanagari characters", Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference , Vol. 2, pp. 1156-1160, IEEE 2005.

[6] Brijmohan Singh, Ankush Mittal and Debashis Ghosh, "An Evaluation of Different Feature Extractors and Classifiers for Offline Handwritten Devnagari Character Recognition", Journal of Pattern Recognition Research, Vol 2pp. 269-277, 2011.

[7] H K Anasuya Devi, “Thinning: A preprocessing technique of OCR for the Brahmi Script”, Ancient Asia Vol. 1, 2006.

[8] Dipti Deodhare, NNR Ranga and Suri R. Amit, “Preprocessing and Image Enhancement Algorithms for a Form-based Intelligent Character Recognition System”, International Journal of Computer Science & Applications, Vol.II, No.II, pp.131-144, 2005.

[9] P E Ajmire, R V Dharaskar and V M Thakare, " Handwritten Devanagari (Marathi) Character Recognition using MLP and SVM", International journal of Pure and Applied Research in Engineering and Technology, Vol 4(4), pp. 1-10, 2015.

[10] Chih-Wei Hsu and Chih-Jen Lin, “A Comparison of Methods for Multi-class Support Vector Machines”, Department of Computer Science and Information Engineering National Taiwan University Taipei 106, Taiwan.

[11] Koby Crammer and Yoram Singer, “On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines”, Journal of Machine Learning Research 2, pp.265-292, 2001.