Research Article CODEN: IJPAKY Impact Factor: 4.226 ISSN: 2319-507X
P. E. Ajmire, IJPRET, 2015; Volume 4 (4): 1-10 IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

HANDWRITTEN DEVANAGARI (MARATHI) CHARACTER RECOGNITION USING SVM AND MLP

P. E. AJMIRE[1], R V DHARASKAR[2], V M THAKARE[3]

1. Associate Professor, Dept. of Computer Science, G. S. Science Arts & Comm. College, Khamgaon. (M.S.), India.
2. Former Director, Disha Education Society, Raipur, (C.G), India.
3. Professor & Head, Dept. Of Comp. Sci., PGTD of CS, S G B Amravati University, Amravati, (M.S.), India.

Accepted Date: 22/11/2015; Published Date: 01/12/2015

Abstract: - Many researchers are working to automate the process of reading, understanding and interpretation of handwritten character. In this paper, we present the new machine learning algorithm for Handwritten Devanagari (Marathi) characters recognition. Offline handwritten character recognition is an important area of Document Analysis and Recognition (DAR). DAR is a mechanism in which the document images are processed to obtain features and its recognition. Each character is stored as an image. The histogram oriented gradient features are extracted. The algorithm is designed to select the extracted features. There is 25% reduction in the feature set. These features are classified using SVM and the obtained average recognition accuracy is on testing is 95.83% and on cross validation is 95.82%. MLP classifiers accuracy is 95.45% on testing and on cross validation is 95.32 % with TanhAxon as Transfer function.

Keywords: Character Recognition, feature extraction, SVM, MLP, Classification.

PAPER-QR CODE

Corresponding Author: MS. P. E. AJMIRE

Access Online On:

www.ijpret.com

How to Cite This Article:

P. E. Ajmire, IJPRET, 2015; Volume 4 (4): 1-10

INTRODUCTION

In the recent years, the use of technology in our day to day life is increased exponentially. Every human being is egger to convey his feelings, thoughts to friends and family members. So it is necessary make the machine as intelligent as possible. The recognition and processing of handwritten character is motivated largely by desire to improve man and machine communication. Many researches are working on this topic but still the task of handwritten Devanagari character recognition remains challenging task to the researchers. In the field of pattern recognition, character recognition has its own importance. There are various scripts and languages used in the world. In this modern era, computers are used widely in communication, education, etc. The task of recognition specially, character recognition looks like very simple for human beings. During the study of methods of handwritten character recognition, it is found that the human brain is the best system to investigate and model than one which is already very successful at such a task. Human beings are studying these things from childhood. For human beings, it is necessary to study the object and stores the characteristics or features of that object. The same task is very complex for machines. In addition to the errors caused by image quality, image features, segmentation, and recognition. The image complexity also has influence on the recognition rate of handwritten character recognition [1, 2]. Devanagari character recognition is complicated task due to the presence of multiple loops, conjuncts, upper and lower modifiers and the number of disconnected and multi-stroke characters, in a word and some writer use header line.

For instance, OCR systems will be more widely used to store, search, and expert from paper-based documents. [3]

The area of character recognition can be listed as,

i. Sorting of Postal Documents using Address

ii. Verifying checks in Bank.

iii. Reading and analyzing documents in Bank and other offices.

iv. Signature recognition.

v. Vehicle Nameplate Recognition.

vi. Analyzing the person

vii. Crime detection using handwriting.

RELATED WORK: India is home of several hundred languages. There are 1,652 different languages in India and 350 are major languages [4, 5]. Most of the languages belong to the Dravidian and Aryan families. Hindi and English are the official languages of the Republic of India. Hindi is the most widespread language of India. There are 22 constitutionally recognized languages in India [6]. Hindi is dominant among them as nearly 41% people speak in Hindi.

First research report on handwritten Devanagari characters was published in 1977 by I. K. Sethi and B. Chatterjee [7] but not much research work is done after that. At present researchers are working on handwritten Devanagari characters. Devanagari script is the most popular script. It is phonetic in nature and the writing system for this script maps the sound of the aksharas to specific shapes. Devanagari is used in many Indian languages like Sanskrit, Hindi, Nepali, Marathi, Konkani, Sindhi etc. More than 300 million people around the world use Devanagari script [8]. This script forms the foundation of Indian languages. So Devanagari script plays a very major role in the development of literature and manuscripts. S Arora et al [9] discussed the characteristics of Support vector Machine (SVM) and Artificial Neural Network (ANN) classification Method and achieved 98.16% recognition rate. J.Pradeep et al [10] presented a diagonal based feature extraction method and obtained recognition accuracy of 97.8 % for 54 features and 98.5% for 69 features is recorded. Ashutosh Aggarwal et al [11] use the gradient representation as the basis for extraction of features. The experimental results using Support Vector Machines (SVM) are presented. The result demonstrates high performance of these features with cross validation accuracy of 94%. M M Majgaonker and T J Siddiqui [12] carried out a case study for discovering suffixes of Marathi Language for this purpose they evaluate rule-based and unsupervised Marathi stemmer. The rule based stemmer uses a set of manually extracted suffix stripping rules where as the unsupervised approach learns suffix automatically from a set of words extracted from row Marathi word. R Jayadevan et al [13] proposed the recognition technique which is a combination of two approaches. The first approach is based on gradient, structural and cavity (GSC) features along with binary vector machine (BMV) technique. The second approach is based on vertical projection profile (VPP) feature and dynamic time warping (DTV). Ved Prakash Agnihotri [14] presented a system using neural network. Diagonal based feature extraction is used for handwritten Devanagari script. The precision of offline Devanagari recognition system is 85.78%. V S Tapkir at el [15] proposed algorithm which is based on projection profiles. The profile based methods can only segment non-overlapping lines and characters. The experimental result observed 98% accuracy. Pratibha Singh at el [16] proposed mini-batch stochastic gradient descent based learning applied to multilayer perceptron. This technique reduces the variance in the estimate of the gradient.

PROPOSED SYSTEM: The task begins with the database preparation. Database preparation is basic and important task in this research work, as no standard handwritten Devanagari / Marathi database is available [17]. In handwritten Devanagari character, there is lack of such official benchmark [18]. In our proposed method, we designed the database of 36 following characters.

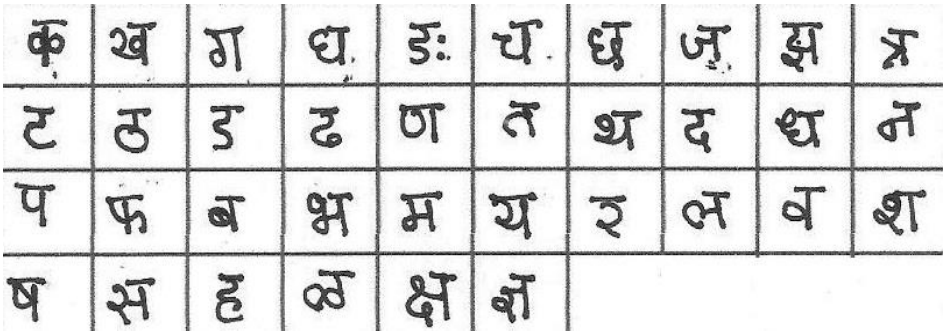

Figure 1 Basic Characters

Following Algorithm is used to developed database.

1. Read Datasheet file (Scan file)
2. Convert to gray
3. Convert gray to binary
4. Take compliment of this data file
5. Crop each Line and then each character
6. Resize each character to 40 x 40 size

After pre-processing the document image each image in the database is of size 40 x 40, features relevant to the classification problem are extracted. The histogram oriented gradient features are extracted. HOG features describe the shape of the image by the distribution of intensity gradients or edge directions. A HOG feature vector represents local shape of an object [19]. HOG features are relatively invariant to local geometric and photometric transformations. The features encode local shape information from regions or from point locations within an image. From these 81 features 63 features are selected by using feature selection technique.

To calculate the gradient, the method is to apply 1-D centered, point discrete derivative mask in one or both horizontal and vertical directions. This method requires filtering intensity data of the image with the following filter kernels:

Dx=[-1,0,1] and .....(1)

$Dy=[-1,0,1]^T$ .....(2)

So, if I is the image, then x and y derivatives using convolution operation is

Ix=I * Dx .....(3)

Iy=I * Dy .....(4)

The magnitude of gradient is $|G|=\sqrt{I^2_x+I^2_y}$ .....(5)

The orientation of the gradient is given by ɵ = arctan (Iy/Ix) .....(6)

RESULT AND DISCUSSION:

The main task of classification is to use the feature vectors provided by the feature extraction algorithm to assign the object to a category. In our work, we used Support Vector Machine and Multilayer perceptron for the classification of Devanagari characters. Support Vector Machine has been used successfully for pattern recognition and regression tasks [20] formulized under the concept of structural risk minimization rule [21]. SVMs are known to generalize well even in high dimensional spaces under small training sample conditions [22]. SVMs have been successfully applied to a number of applications ranging from face detection, verification, and recognition. The results are shown in table 1.

Accuracy of SVM on 60-20-20 % of Data for Training-Testing-Cross Validation

Sr.No.	Char	Testing	Training	CV
1	K	95	100	100
2	Kha	94.73	100	95.83
3	G	100	100	100
4	Gh	86.36	100	100
5	Dn	91.66	100	88.88
6	Ch	100	100	95.83
7	Chha	100	100	95
8	J	100	100	95.65

9	Z	89.47	100	100
10	Tra	100	100	100
11	Ta	100	100	95.65
12	Th	100	100	94.44
13	D	90.9	100	100
14	Dh	90.9	100	80.95
15	Na	100	100	100
16	T	96.15	100	100
17	Tha	100	100	95
18	Da	93.33	100	95.42
19	Dha	95.23	100	85
20	N	100	100	94.73
21	P	87.5	100	95.23
22	F	93.75	100	100
23	B	91.66	100	90.9
24	Bh	100	100	95
25	M	100	100	85
26	Y	100	100	100
27	R	100	100	100
28	L	94.73	100	92
29	V	86.36	100	100
30	Sh	88	100	100
31	Sha	94.73	100	100
32	S	94.44	100	91.66
33	H	95	100	100
34	La	100	100	100
35	Ksh	100	100	100
36	Dnya	100	100	87.5
Average %age of Accuracy		95.8306	100	95.8242

Table 1: Result of SVM Classifier

MLP classifier is employed for recognition of characters using the same feature set. The MLP is a special kind of Artificial Neural Network (ANN). MLP has been chosen because of its well-known learning and generalization abilities, which is necessary for dealing with imprecision in input patterns. Architecturally, an MLP is a feed-forward layered network of artificial neurons.

Each artificial neuron in the MLP computes a TanhAxon and Sigmoid function of the weighted sum of all its inputs. An MLP consists of one input layer, one output layer and one hidden or every neuron in a layer of the MLP is connected to all inputs of each neuron in the immediate next layer of the same. Neurons in the input layer of the MLP are all basically dummy neurons as they are used simply to pass on the input to the next layer just by computing an identity function each. The numbers of neurons in the input and the output layers of an MLP are chosen depending on the problem to be solved. The number of neurons in other layers and the number of layers in the MLP are all determined by a trial and error method at the time of its training. An ANN requires training to learn an unknown input-output relationship to solve a problem. For the present work, we use Momentum and Step as a learning rule. The results are shown in table 2.

MLP with T.F-TanhAxon and L.R-Momentum				
Sr.No.	Char	Testing	Training	CV
1	K	100	100	100
2	Kha	92.3	100	100
3	G	100	100	100
4	Gh	94.1176	100	78.57143
5	Dn	95.2381	100	100
6	Ch	100	100	96
7	Chha	96.7742	100	85.71429
8	J	95	100	100
9	Z	100	100	89.47
10	Tra	100	100	100
11	Ta	94.4444	100	86.36364
12	Th	100	98.3871	90.47619
13	D	100	96.7213	100
14	Dh	90	98.2456	95.65217
15	Na	100	100	100
16	T	90.9091	100	94.11765
17	Tha	97.561	100	97.5
18	Da	90	97	96.5
19	Dha	89.4737	100	89.47368
20	N	90.4762	91.3793	95.2381
21	P	100	100	88.88889
22	F	100	100	100

23	B	83.3333	100	91.30435
24	Bh	86.6667	100	95.65217
25	M	90.9091	91.2281	85.71429
26	Y	91.3043	100	100
27	R	100	100	100
28	L	100	100	96
29	V	93.75	100	94.44444
30	Sh	95	100	95.45455
31	Sha	100	100	100
32	S	94.1176	100	93.33333
33	H	100	100	100
34	La	95.6522	100	100
35	Ksh	100	100	100
36	Dnya	89.4737	100	95.83333
Average %age of Accuracy		95.4584	99.2489	95.32507

Table 2: Accuracy obtained by MLP

CONCLUSION: An attempt is made to apply new technique for feature extraction. The method is promising one. The recognition rate is increased by reducing the features by nearly 23% and hence it is remarkable. Some researchers used these features for object detection. In this work, as per the information available, this is first time these features are used for character recognition. The obtained average result is on testing is 95.83% and on cross validation is 95.82%. MLP classifiers accuracy is 95.45% on testing and 95.32 on cross validation with Tanhaxon as Transfer function. The handwritten character recognition accuracy is better by using SVM classifier but the results of MLP classifiers are also comparable.

REFERENCES

1. Amalia Rusu and Venu Govindaraju, “The Influence of Image Complexity on Handwriting Recognition”, University at Buffalo, The State University of New York Buffalo, NY, USA.

2. P. E. Ajmire, R V Dharaskar, V M Thakare, “Study of Complexity of Devanagari Script for Handwritten Character Recognition”, Proceedings of Twelveth IRF International Conference, Chennai, India, pp.109-111, 31st August 2014.

3. Simone Marinai, “Introduction to Document Analysis and Recognition”, Springer-Verlag Berlin Heidelberg 2008.

4. Nandita Chaudhary, "Listening To Culture: Constructing Reality from Everyday Talk", SAGE Publications India.

5. E-Book "Discovering the Life Span", by Pearson Education India.

6. R. Mahesh K. Sinha, "A Journey from Indian Scripts Processing to Indian Language Processing", IEEE Annals of the History of Computing, IEEE Computer Society, 2009.

7. I.K. Sethi, B. Chatterjee, "Machine recognition of constrained hand printed Devanagari", Pattern Recognition, The Journal of the Pattern Recognition Society, Vol. 9, Issue 2, pp 69-75, July 1977.

8. Latesh Malik and P. S. Deshpande, "Recognition of printed Devnagari characters with regular expression in finite state models", Proceedings of the International Workshop on Machine Intelligence Research (MIR Day, GHRCE- Nagpur), 2009.

9. Sandhya Arora, D. Bhattacharjee, M. Nasipuri, D. K. Basu and M. Kundra, "Complementary Features Combined in a MLP-based System to Recognize Handwritten Devanagari Characters" Journal of Information Hiding and Multimedia Signal Processing Vo. 2,No.1 Jan 2011.

10. J. Pradeep, E. Srinivasan and S. Himavathi "Diagonal Based Feature Extraction for Handwritten Alphabets Recognition System Using Neural Network", International Journal of Computer Science & Information Technology, Vol. 3, No 1, Feb 2011.

11. Ashutosh Aggarwal, Rajneesh Rani, Renu Dhir "Handwritten Devanagari Character Recognition Using Gradient Features", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 5, May 2012.

12. M M Majgaonker and T J Siddiqui "Discovering suffixes: A case study for Marathi Language" International Journal on Computer Science and Engineering, Vol.02, No. 08, pp 2716-2720, 2010.

13. R Jayadevan, S R Kolhe, P M Patil and U Pal "Database Development and Recognition of Handwritten Devanagari Legal Amount words", International Conference on Document Analysis and Recognition, IEEE, 2011.

14. Ved Prakash Agnihotri, "Offline Handwritten Devanagari Script Recognition", Information Technology and Computer Science, Vol. 8, pp. 37-42, 2012.

15. Mrs. V S Tapkir and Mrs.S D Shelke, "OCR for handwritten Marathi Script", International Journal of Scientific & Engineering Research Vol. 3, Issue 8, August 2012.

16. Pratibha Singh, Ajay Verma and Narendra S Chaudhari, "Handwritten Devanagari Digit Recognition using Fusion of Global and Local Features", International Journal of Computer Application, Vol.89, Number-1, 2014.

17. Sandhya Arora, D. Bhattacharjee, M. Nasipuri, D. K. Basu and M. Kundra, "Complementary Features Combined in a MLP-based System to Recognize Handwritten Devanagari Characters" Journal of Information Hiding and Multimedia Signal Processing Vo. 2,No.1 Jan 2011.

18. Vikas J Dongre and Vijay H Mankar, "Development of Comprehensive Devanagari Numeral and Character Database for Offline Handwritten Character Recognition", Applied Computational Intelligence and Soft Computing, 2012.

19. Takuya Kobayashi, Akinori Hidaka, and Takio Kurita, "Selection of Histograms of Oriented Gradients Features for Pedestrian Detection", International Conference on Neural Information Processing (ICONIP), LNCS Vol. 4985, pp. 598–607, 2008.

20. C. Burges, "A tutorial on support vector machines for pattern recognition", Data Mining and Knowledge Discovery vol.2 pp. 1-43, 1998.

21. V. Vapnik, "An overview of statistical learning theory", IEEE Transaction on Neural Network vol.10 pp. 989-999, 1999.

22. K. Jonsson, J. Kittler, and Y.P. Matas, "Support vector machines for face authentication", Journal of Image and Vision Computing, Vol. 20. pp. 369-375, 2002.