Recognition of Similar Shaped Handwritten Marathi Characters Using Artificial Neural Network

1. Mrs.Archana P.Jane, 2. Prof.Mukesh.A.Pund

*M.E (Scholar), Prof.Ram Meghe Institute of Research & Technology, Badnera
** Associate Prof. Prof.Ram Meghe Institute of Research & Technology, Bandera

Abstract:
         The growing need have handwritten Marathi character recognition in Indian offices such as passport, railways etc has made it vital area of a research. Similar shape characters are more prone to misclassification. In this paper a novel method is provided to recognize handwritten Marathi characters based on their features extraction and adaptive smoothing technique. Feature selections methods avoid unnecessary patterns in an image whereas adaptive smoothing technique form smooth shape of charecters.Combination of both these approaches leads to the better results. Previous study shows that, no one technique achieves 100% accuracy in handwritten character recognition area. This approach of combining both adaptive smoothing & feature extraction gives better results (approximately 75-100) and expected outcomes.

Keywords: character recognition, features Extraction, adaptive Smoothing, Image segmentation, pattern matching, Image Pixel Rating.

1.Introduction:

         In recent years, handwritten Marathi character recognition has grabbed a lot of attention as Marathi being primary official language in Maharashtra has wide application in areas like passport, railways, postal address reading etc.Handwritten character recognition [4] consist of six main steps.

i. Handwritten drawn of character

ii.   Training on handwritten characters.

iii. Testing on handwritten characters.

§ Pixel rating an Image

§ Image smoothing

§ Features extraction.

§ Image segmentation.

§ Patterns matching

§ Result display as per their % of pattern matched.

Earlier, traditional classifiers such as Nearest Neighbor (NN), Hidden Markov Models (HMM) etc. were adopted for character recognition, however they exhibit certain limitations. Machine learning (ML) algorithms [6] provide a promising alternative in character recognition based on the feature set given to them. Each character image sample can be expressed in terms of some 26 Computer Science & Information Technology (CS & IT) quantifiable attributes called features. A variety of features can be extracted such as primitives, profiles etc.Multi Layer (ML) algorithm is then trained with this list of measured features, so that it maps these input features onto a class among certain predefined classes [1, 2]. Then the classifier can be used to determine the class of unknown samples used for testing.

2. Literature Survey:

         Character recognition task has been attempted through many different approaches like template matching, statistical techniques like NN, HMM, Quadratic Discriminant function (QDF) etc.Template matching works effectively for recognition of standard fonts, but gives poor performance with handwritten characters and when the size of dataset grows. It is not an effective technique if there is font discrepancy [4]. HMM models achieved great success in the field of speech recognition in past decades, however developing a 2-D HMM model for character recognition is found difficult and complex [5]. NN is found very computationally expensive in recognition purpose [6]. N. Araki et al. [7] applied Bayesian filters based on Bayes Theorem for handwritten character recognition. Later, discriminative classifiers such as Artificial Neural Network (ANN) and Support Vector Machine (SVM) grabbed a lot of attention. In [3] G. Vamvakas et al. compared the performance of three classifiers: Naive Bayes, K-NN and SVM and attained best performance with SVM. However SVM suffers from limitation of selection of kernel. ANNs can adapt to changes in the data and learn the characteristics of input signal [8].Also, ANNs consume less storage and computation than SVMs [9]. Mostly used classifiers based on ANN are MLP and RBFN. B.K.

Verma [10] presented a system for HCR using MLP and RBFN networks in the task of handwritten Hindi character recognition. The error back propagation algorithm was used to train the MLP networks. J. Sutha et al. in [11] showed the effectiveness of MLP for Tamil HCR using the Fourier descriptor features. R. Gheroie et al. in [12] proposed handwritten Farsi character recognition using MLP trained with error back propagation algorithm. Computer Science & Information Technology (CS & IT) 27 similar shaped characters are difficult to differentiate because of very minor variations in their structures. In [13] T. Wakabayashi et al. proposed an F-Ratio (Fisher Ratio) based feature extraction method to improve results of similar shaped characters. They considered pairs of similar shaped characters of different scripts like English, Arabic/Persian, Devnagri, etc. and used QDF for recognition purpose. QDF suffers from limitation of minimum required size of dataset. F. Yang et al. in [14] proposed a method that combines both structural and statistical features of characters for similar handwritten Chinese character recognition. As it can be seen that various feature extraction methods and classifiers have been used for character recognition by researchers that are suitable for their work, we propose a novel feature set that is expected to perform well for this application. In this paper, the features are extracted on the basis of character geometry, which are then fed to each of the selected ML algorithms for recognition of SSHMC.

3.Proposed Method:

In this paper, we proposed a novel method based on combinations on pixel rating an image, feature extractions and Image pattern matching. Proposed method gives considerable expected outputs than previous proposed character recognition algorithms like HMM, NN, ML etc.Proposed method consists of following phases
3.1 Training
3.2 Testing

3.1 Training

In this proposed work, Training on images consists of listing all handwritten images with respect to its standard Marathi character images. See the training set data in the following figure.

Handwritten Images	Standard Marathi Character Images
अ	अ
आ	आ
इ	इ
ई	ई
उ	उ
स	स
क	क
र	र
न	न
म	म
द	द
प	प

In Training phase, we are having 51 Marathi characters and we Trained 20 handwritten characters with respect to each corresponding standard Marathi character images. The no. of images in Training affects turnaround time of entire process execution.

$$Ti \propto N \text{ -----------------Eq.3.1.1}$$

Where
T_i=Turnaround time. N=No. of Training images.
An accuracy of result depends on the no. of trained handwritten images per standard Marathi character image.

$$Ac \propto \frac{Him}{Sim} \text{-----------------Eq.2.1.2}$$

Where
A_c= Accuracy of character Recognition
H_{im}=No. of handwritten patterns. S_{im}= Standard Marathi image=1

3.2 Testing

Testing is a phase where no. of Marathi handwritten character image is tested against training set handwritten images. Testing consists of following phases:
3.2.1 RGB-to-Binary image conversion

3.2.2 Pixel rate an image
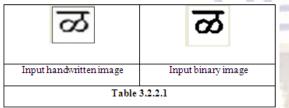
3.2.3 Image edge smoothing.

3.2.4 Image segmentation.

3.2.5 Pattern Matching.

3.2.6 Experiment Result.

3.2.1 RGB-to-Binary image conversion

As we maintained Training images as a binary images, there is need to convert testing image into binary image. Binary image avoids unnecessary image segmentation and features extraction.

Input handwritten image	Input binary image
Table 3.2.2.1	

3.2.2 Pixel rate an image

Input binary image	Pixel Rate Image
Table 3.2.2.2	

Pixel rate an image used to identify an image pixel value either on or off. We set pixels height and width equal to 10 of which we got result shown in above table 3.2.2.2. Height, Width of an pixel rate image should be proportionate to size of an binary image .From the table 3.2.2.2, it is observed that having pixels size 10 x 10 leads to loss of pixels which are represented with an equation

$$Pl = \int_1^n Pnhw(\log_2 n) \text{----------Eq.3.2.1}$$

Where

P_l= No. of Pixels loss

P_{nhw}= Pixels new height, width. n=No. of off pixels in binary image.

3.3 Image edge smoothing

From literature review, it is observed that patterns in handwritten characters have large deviation factor with respect to its standard image pattern . Deviation factor is a mod difference value between no. of pixels patterns in training and testing images and represented with

Where

D_f= Deviation factor.

T_r= No. of segments of training image. T_s= No. of segments of testing image.

$$Tr = Ts = \sum_1^i \max_{1 \le hw \le 100} d\left(\frac{s}{dn}\right) \text{--------Eq.3.3.2}$$

Where

S= size of an image.

hw= new height, width of segment. D_n = new segment size

Input binary image	Image After Edge Smoothing

Average searching time for character recognition of test image uses best-fit search approach with training set images.

$$O(t) = O(N) \text{ -------eq.3.3.3}$$

Where
O (t) =Average searching time.

3.4 Image segmentation

To recognize character, segmentation is done based on their patterns of size 2x2 as per equation 3.3.2.

3.5 Pattern Matching

We form fuzzy pattern match of pixels value (Equivalent decimal value) of segmented image based on which we match an object array patterns with following fuzzy rules:

$$Ptm = \sum_{i=1}^{s} Tst \; v \left(Trt \; v Trt + i \; v \; ...... \; v \; \frac{Trt}{i}\right) \text{----eq3.5.1}$$

Where

P_{tm}= fuzzy pattern match.

S= size of training image=size of testing image.

3.2.6 Experiment result

Sr No	Test Image	Recognize Image 1	Recognize Image 2	Recognize Image 3	Recognize Image 4	Recognize Image 5	Recognize Image 6	Recognize Image 7	Recognize Image 8	Recognize Image 9
1	प	प	व	ब	फ	म	क	ष	भ	य
Std.marathi char,		प	व	ब	फ	म	क	ष	भ	य
Deviation Factor		0.0000	0.1923	0.2345	0.2378	0.3124	0.3564	0.4219	0.5638	0.5698
Recognize Character						प				

Sr No	Test Image	Recognize Image 1	Recognize Image 2	Recognize Image 3	Recognize Image 4	Recognize Image 5	Recognize Image 6	Recognize Image 7	Recognize Image 8	Recognize Image 9
2	म	म	प	भ	फ	ष	स	य	क	न
Std.marathi char,		म	प	भ	फ	ष	स	य	क	न
Deviation Factor		0.0000	0.1276	0.2234	0.2567	0.3015	0.3567	0.4216	0.5216	0.5767
Recognize Character						म				

Sr no	Test Image	Recognize Image 1	Recognize Image 2	Recognize Image 3	Recognize Image 4	Recognize Image 5	Recognize Image 6	Recognize Image 7	Recognize Image 8	Recognize Image 9
3	फ	फ	प	व	क	अ	भ	ष	थ	घ
Std.marathi char,		फ	प	व	क	अ	स	ष	थ	घ
Deviation Factor		0.0000	0.1435	0.24545	0.2535	0.25467	0.258965	0.46556	0.525464	0.58967
Recognize Character						फ				

Sr.no.	No. of Training Images / standard Marathi characters	Type of Images	% Loss of Pixels/Ima ge in Pixel rating Image	Average % Accuracy
1	10	JPG	33%	70%-80%
2	20	JPG	33%	80%-85%
3	30	JPG	33%	85%-90%
4	50	JPG	33%	90%-98%

Experiment Result 2

Conclusion:

From the study of Literature survey and proposed method ,we conclude that ,proposed method gives considerable and expected accuracy than previous character recognitions techniques like HMM,ML,NBP etc.Experiment results shows that ,proposed method achieved an accuracy nearer to 98% provided no. of training samples per standard Marathi images should be maximum as possible as.

Future Work:

In the process of recognizing handwritten character, human brains may fails that's why to keep an expectations to achieve 100% accuracy is not expectable. A future work is needed to correctly analyze segments patterns and fuzzy rules mentioned in an equation 3.5.1 to achieve better accuracy which should be independent of no. of training set images.

References

[1] T.M. Mitchell, "Machine Learning", McGraw-Hill Education (ISE Editions), December 1997.

[2] M. Zafar, D. Mohamad, M.M. Anwar, "Recognition of Online Isolated Handwritten Characters by Back propagation Neural Nets Using Sub-Character Primitive Features", IEEE Multitopic Conference(INMIC), 2006, pp. 157 - 162

[3] G. Vamvakas , B. Gatos, S. Petridis, N. Stamatopoulos , "An Efficient Feature Extraction and Dimensionality Reduction Scheme for Isolated Greek Handwritten Character Recognition", IEEE Ninth International Conference on Document Analysis and Recognition(ICDAR), 2007, vol. 2,pp. 1073 – 1077 Figure 2. Misclassification Rate of Bayesian Network and C4.5 with FULL,CFS and CON Features Sets 34 Computer Science & Information Technology (CS & IT)

[4] J.R. Prasad, U.V. Kulkarni, R.S. Prasad, "Offline handwritten character recognition of Gujrati script using pattern matching", IEEE 3rd International Conference on Anti-counterfeiting, Security,and Identification in Communication, 2009, pp. 611-615.

[5] H.S. Park, S.W. Lee, "An HMMRF-Based Statistical Approach for Off-line Handwritten Character Recognition", IEEE Proceedings of the 13th International Conference on Pattern Recognition, 1996, vol. 2, pp. 320 – 324.

[6] C.L. Liu, H. Sako, H. Fujisawa, "Performance evaluation of pattern classifiers for handwritten character recognition", International Journal on Document Analysis and Recognition (IJDAR), 2002, vol. 4, pp. 191–204.

[7] N. Araki, M. Okuzaki, Y. Konishi , H. Ishigaki , "A Statistical Approach for Handwritten Character Recognition Using Bayesian Filter", IEEE 3rd International Conference on Innovative Computing

Information and Control, 2008, pp. 194 – 194.

[8]   N. Arica, F.T. Yarman-Vural, "An Overview Of Character Recognition Focused On Off-line Handwriting", IEEE Transactions on Systems, Man, and Cybernetics, 2001, vol. 31, pp. 216 –233.

[9]   F. Kahraman, A. Capar, A. Ayvaci, H. Demirel, M. Gokmen, "Comparison of SVM and ANN performance for handwritten character classification", Proceedings of the IEEE12th Signal Processing and Communications Applications Conference, 2004, pp. 615 – 618.

[10]   B.K. Verma, "Handwritten Hindi Character recognition Using Multilayer Perceptron and Radial Basis Function Neural Networks," Proceedings of IEEE International conference on Neural Networks, 1995, vol. 4, pp. 2111-2115. [11] J. Sutha, N. Ramaraj, "Neural Network Based Offline Tamil Handwritten Character Recognition System", IEEE International Conference on Computational Intelligence and Multimedia Applications, 2007, vol. 2, pp. 446 – 450.

[12]   R. Gharoie, M. Farajpoor, "Handwritten Farsi Character Recognition using Artificial Neural Network", International Journal of Computer Science and Information Security, 2009, vol. 4.

[13]   T. Wakabayashi, U. Pal, F. Kimura, Y. Miyake, "F-ratio Based Weighted Feature Extraction for Similar Shape Character Recognition" IEEE 10th International Conference on Document Analysis and Recognition (ICDAR), 2009, pp. 196-200.

[14]   F. Yang, X.D. Tian, X. Zhang, X.B. Jia, "An Improved Method for Similar Handwritten Chinese Character Recognition", IEEE Third International Symposium on Intelligent Information Technology and Security Informatics (IITSI), 2010, pp. 419 – 422.

[15]   N.J. Nilsson, "Introduction to Machine learning", An early draft of a proposed textbook, Robotics Lab, Deptt of Computer Science, Stanford University, Stanford, CA 94305, 1998.[16] WEKA 3: Data Mining With Open Source MachineLearningSoftware in JAVA URL: http://www.cs.waikato.ac.nz/ml/weka/.

[17]   N. Friedman, D. Geiger, M. Goldszmidt, "Bayesian Network Classifiers, Machine learning, 1997,pp. 131-163.