

Handwritten Devanagari Compound Character Recognition

^[1] Juhee Sachdeva, ^[2] Dr. Shilpa Kulkarni

^[1] Ph.D. Scholar, Jaipur National University, Jaipur, India, ^[2] AP JNU Jaipur India

Abstract:— Compound characters recognition is a difficult task because characters are written in combination of two simple characters. These characters tends to touch each other in different forms and so the segmentation of compound characters is difficult task because of high error rate. In this paper we present a technique for recognition of compound character using MLP technique. The process of pre classification is implemented using MLP that is trained by Back Propagation method. After Preprocessing stage and pre classification stage Diagonal feature extraction is implemented. The proposed work reported a high recognition rate of accuracy.

Keywords:— Handwritten Character Recognition, Compound character, Diagonal Feature Extraction, Multilayer Perceptron Network

I. INTRODUCTION

Handwritten character recognition is gaining popularity due to its potential application areas which would reduce the task of data entry and save the time in case of Form filling, Postal Automation, and Banking etc.[1]. But developing a system for handwritten character recognition poses a challenge to the researchers due to the varying shape of the character that may depend upon the writer, the physical and mental condition of the writer, the acquisition device, pen width, pen ink color and many other factors. Moreover, handwritten Devanagari characters tend to be more complex due to their structure and shape. They include a large character set with more curves, loops and other details in the characters[3]. There are many character pairs which are quite similar in shape[4]. All these issues demand attention and solution with the help of an efficient recognition system[16][17]. In India, there are twenty two Indian official (Indian constitution accepted) languages, namely Assamese, Bangla, Gujarati, Hindi, Konkani, Kannada, Kashmiri, Malayalam, Marathi, Nepali, Oriya, Punjabi, Bodo, Dogri, Maithili, Manipuri, Santhali, Sindhi, Sanskrit, Tamil, Telugu and Urdu[8]. In the proposed work, we aim at developing a comprehensive system for recognition of handwritten Devanagari characters.

II. DEVANAGARI SCRIPT AND COMPOUND CHARACTERS

Devanagari is the most popular script in India. It consist of 11 vowels and 33 consonants which are called as basic characters. Vowels can be written as independent letters, or with the help of a variety of diacritical marks which are connected above, below, before or after the consonant to which they belong to. When vowels are written in this way they are known as modifiers and the characters so formed are

called conjuncts. Sometimes two or more consonants can combine and take new shapes. These new shape clusters are known as compound characters[5]. So these types of basic characters, compound characters and modifiers are present not only in Devanagari script but also in other scripts. Hindi is the national language of India and written in the Devanagari script. Devanagari is also used for to write Marathi, Sanskrit and Nepali. So besides this, Hindi is the third most popular language in the world. All the characters have a horizontal line as the upper part, known as Shirorekha / headline.

क ख क ग घ च छ ज झ ञ ट
ट ड ढ ण त थ द ध न प फ
व भ म य र ल व श ष स ह

Fig:1 Compound Characters

क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट
ठ	ड	ढ	ण	त	थ	द	ध	न	प	फ
ब	भ	म	य	र	ल	व	श	ष	स	ह

Fig:2 Consonants

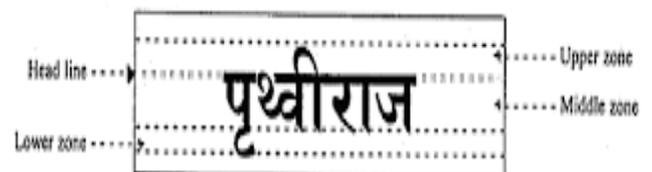


Fig:3 Three zones of a Devanagari word

Devanagari characters can be further classified on the basis of presence of vertical bar

End-Bar Compound characters

स्त ख्य ज्य त्व म्य च्य ल्प न्म न्न त्प ल्य स्व त्य

Middle Bar Compound Character

क्फ क्क क्व

No-Bar Compound Character

इड डृ दृ

III. PROPOSED METHODOLOGY

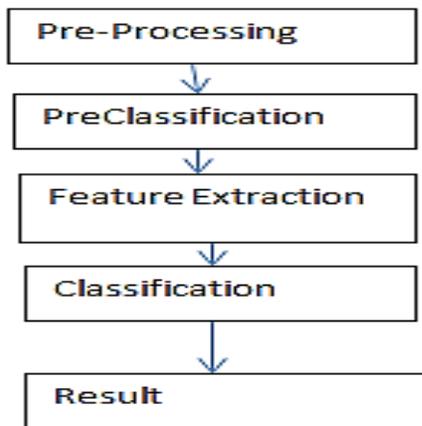


Fig: 4 Offline Character Recognition Systems

To recognized the handwritten data from document, there are number of steps which are involved while recognition, firstly the document is scanned using scanner [1]. This scanned document is converted into image. Then image is preprocessed with set of valuable steps and convert it into a character/script as per the environment. The scanned image undergoes number of valuable preprocessing steps so as to increase the ratio of recognition of the handwritten document. The general steps for Handwritten Character recognition is Image Acquisition, preprocessing, feature extraction, classification and recognition[3].

3.1 Preprocessing:

Preprocessing consist of various operations performed on image. It enhances the image making it suitable for next level of segmentation. It removes noise from image. All work has been done in MATLAB[21]. Preprocessing of compound characters have following steps involved:

Binarization: Upgradation of grey scale image in to binary image is Binarization.

Noise Elimination: Noise can occur at any stage like image capturing, transmission or compression. Noise degrades the quality of image. So different filters and morphological operations are available for removal of image noise

Size Normalization: Normalization is applied to obtain characters of uniform size. It reduce the size of image without getting the structure altered.

Thinning: To remove the selected foreground pixels from images, thinning is used. Image thinning extracts a skeleton of image without the loss of topological properties.



Fig5: Character Image after Preprocessing

3.2 Pre-Classification:

The large number of compound character set with a varying style of writing variations require pre-classification of the characters before the final recognition. Structural features of the characters are like vertical line, horizontal line, end points, junction points are used for classification purpose . The first stage employs classification using global features like presence of vertical line in the character, its position in the character and the presence of enclosed regions in the character. The detection of global features is followed by the detection of the local features like end points and their position in the character. On the basis of global feature, characters are classified on the basis of presence of vertical line i.e. character with vertical bar at the end, character with no end bar and character with vertical bar at middle. After global features, local features are obtained by partitioning the character in four quadrants and detection of end points.



Fig6: Character Partitioning

3.3 Feature Extraction:

Feature extraction can be considered as finding a set of parameters (features) that define the shape of the underlying character as precisely and uniquely as possible[24]. The term feature selection refers to algorithms that select the best subset of the input feature set. Methods that create new features based on transformations, or combination of original features are called feature extraction algorithms[9][10]. We are using diagonal Feature Extraction technique for proposed work.

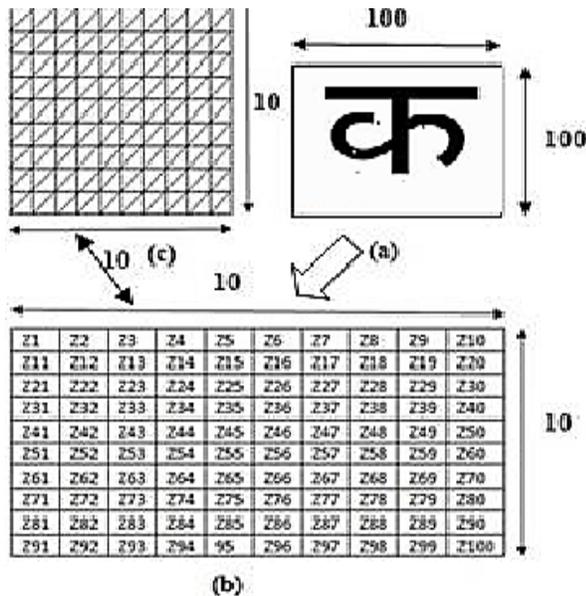


Fig:7 Diagonal Feature Extraction

Diagonal Feature Computation

Every Character Image is divided into 100 equal zones, each of size 10x10 pixels (Fig 8)[12].The Features are extracted from each zone pixels by moving along the diagonals of its respective pixels. Each zone has 19 diagonals,foreground pixels present in each diagonal are summed up so as to get single sub feature and thus 19 sub features are obtained from each zone. These 19 sub features values are averaged to form a single feature value and placed in the corresponding zone. This procedure is sequentially repeated for all the zones. Finally 100 features are extracted for each character[14].

3.4 Classification

Classification is the decision making phase of a Handwritten character recognition system that makes use of the features extracted from the previous stage in the process[11][13]. In the proposed system Multilayer perceptron (MLP) is used for classification. In MLP three layers namely Input layer, hidden layer and output layer are used. It is fully connected layer network.

The MLP Algorithm

Step 1: Initialize weights at random, choose a learning rate η
Until network is trained:
For each training example i.e. input pattern and target output(s):
Step 2: Do forward pass through net (with fixed weights) to produce output(s) i.e., in Forward Direction, layer by layer:
Inputs applied Multiplied by weights Summed ‘Squashed’ by sigmoid activation function, Output passed to each neuron in next layer. Repeat above until network output(s) produced

Step 3: Compute error for each output unit δ_k , layer by layer compute error.
Step 4: Next update all the weights Δw_{ij} by Gradient Descent and go to step 2

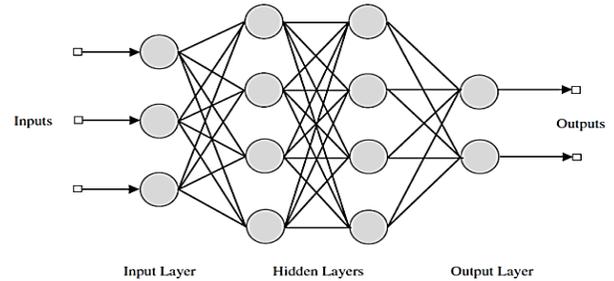


Fig:8 Structural design of Multilayer Perceptron

To train the network to recognize Devanagari Alphabet characters, the corresponding 5x7(6x8) grids are applied to the input of network. Then the weights are calculated using the equations provided. The initial learning rate was experimentally set to 1.5 which is divided by a factor of 2 every 100 iterations and is reset to the initial value after every 300 iterations and the momentum rate is set to 0.95. the characters are considered recognized if all the outputs of the network were no more than 0.01 off their respective desired values.

IV. RESULT AND CONCLUSION

This paper presented a technique for Devanagari compound characters using Pre classification and Multi-layer perceptron. No standard database is available for compound characters so data is collected from people of all age groups. The dataset characters are first preprocessed and then pre-classified based on global and local features. New feature extraction technique i.e. Diagonal Feature Extraction technique for the recognition of handwritten Devanagari characters. Diagonal Feature extraction technique applied with MLP classification technique proves to be powerful tools for recognition system. Feature extraction uses multiple features which gives better recognition accuracy. The MLP algorithm proposed in the paper has the ability to recognize stimulus patterns if a set of input patterns are repeatedly presented to it, it gradually acquires the ability to recognize these patterns. It is not necessary to give any instructions about the categories to which class the pattern belongs. The network is a working design of some neural mechanism of visual pattern recognition. The MLP Neural Network can be used in any kind of recognition system such as cancer cell detection, face recognition system and finger print recognition system. The implementation of fully connected MLP method give reasonable results towards recognizing characters.

REFERENCES

1. Optical Character Recognition by Rice, Stephen V., Nagy, George, Nartker, Thomas A Springer Science & Business Media, 31-May-1999.
2. Gonzalez, Rafael C., and Richard E. Woods. "Digital image processing." (2002).
3. Cheriet, Mohamed, et al. Character recognition systems: a guide for students and practitioners. John Wiley & Sons, 2007.
4. H.Bunke P.S.P Wang-"Handbook of character recognition and document image analysis." (1997).
5. Connell, Scott D., R. M. K. Sinha, and Anil K. Jain. "Recognition of unconstrained online Devanagari characters." Pattern Recognition, 2000. Proceedings. 15th International Conference on. Vol. 2. IEEE, 2000.
6. Bansal, Veena, and R. M. K. Sinha. "On how to describe shapes of Devanagari characters and use them for recognition." Document Analysis and Recognition, 1999. ICDAR'99. Proceedings of the Fifth International Conference on. IEEE, 1999.
7. Bansal, Veena, and R. M. K. Sinha. "Segmentation of touching and fused Devanagari characters." Pattern recognition 35.4 (2002): 875-893.
8. Pal, U., and B. B. Chaudhuri. "Indian script character recognition: a survey." pattern Recognition 37.9 (2004): 1887-1899.
9. Nixon, Mark, Mark S. Nixon, and Alberto S. Aguado. Feature extraction & image processing for computer vision. Academic Press, 2012.
10. Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A. (Eds.). (2008). Feature extraction: foundations and applications (Vol. 207). Springer.
11. Liu, Cheng-Lin, and Hiromichi Fujisawa. "Classification and learning methods for character recognition: Advances and remaining problems." Machine learning in document analysis and recognition. Springer Berlin Heidelberg, 2008. 139-161.
12. J.Pradeep, E.Srinivasan and S.Himavathi, " Diagonal Based Feature Extraction For Handwritten Alphabets Recognition System Using Neural Network", (IJCSIT), Vol 3, No 1, Feb 2011.
13. Dinesh Acharya U., N. V. Subbareddy & Krishnamoorthi, " Combined Classifiers In Recognition Of Handwritten Kannada Numerals: A Hybrid Approach", International Journal of Information Technology and Knowledge Management July-December 2009, Volume 2, No. 2, pp. 305-311.
14. Anita Jindal, Renu Dhir and Rajneesh Rani, " Diagonal Features and SVM Classifier for Handwritten Gurumukhi Character Recognition ", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 5, May 2012 ISSN: 2277 128X .
15. Hadar I. Avi-Itzhak, Thanh A. Diep, and Harry Garland, "High Accuracy Optical Character Recognition Using Neural Networks with Centroid Dithering ", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 17, No. 2, February 1995.
16. Veena Bansal and R.M.K. Sinha, "Integrating Knowledge Sources in Devanagari Text Recognition System", Technical Report, TRCS-97-248, I.I.T. Kanpur, India, 1997.
17. R. M. K. Sinha, "A Journey from Indian Scripts Processing to Indian Language Processing", IEEE Annals of the History of Computing, pp8-31, Jan-Mar 2009.
18. Lee, Yuchun. "Handwritten digit recognition using k nearest-neighbor, radial-basis function, and backpropagation neural networks." Neural computation 3.3 (1991): 440-449.
19. Plamondon, Réjean, and Sargur N. Srihari. "Online and off-line handwriting recognition: a comprehensive survey." IEEE Transactions on pattern analysis and machine intelligence 22.1 (2000): 63-84.
20. MAJIDA ALI ABED HAMID ALI ABED ALASADI "Simplifying Handwritten Characters Recognition Using a Particle Swarm Optimization Approach" EUROPEAN ACADEMIC

International Journal of Science, Engineering and Management (IJSEM)
Vol 3, Issue 1, January 2018

RESEARCH, VOL. I, ISSUE 5/ AUGUST 2013
ISSN 2286-4822.

21. Kauleshwar Prasad, Devvrat C. Nigam, Ashmika Lakhotiya: Character Recognition Using Matlab's Neural Network Toolbox, International Journal of u-and e- Service, Science and Technology Vol.6 No. 1, February, 2013.
22. Anita Pal & Dayashankar Singh, "Handwritten English Character Recognition Using Neural Network" International Journal of Computer Science & Communication Vol. 1, No. 2, July-December 2010, pp. 141-144.
23. Gunjan Singh, Sushma Lehri "Recognition of Handwritten Hindi Characters using Backpropagation Neural Network", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (4) , 2012, 4892-4895.
24. Sandhya Arora, "Combining Multiple Feature Extraction Techniques for Handwritten Devanagari Character Recognition", IEEE Region 10 Colloquium and the Third ICIIS, Kharagpur, INDIA, December 2008.
25. Yegnanarayana, B. Artificial neural networks. PHI Learning Pvt. Ltd., 2009.