



Journal of Artificial Intelligence

ISSN 1994-5450

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Handwritten Devanagari Character Recognition using Artificial Neural Network

P.B. Khanale and S.D. Chitnis
 D.S.M. College, Parbhani, India

Corresponding Author: P.B. Khanale, D.S.M. College, Parbhani, India

ABSTRACT

The reading skills of computer are still way behind that of human beings. Most character recognition systems cannot read degraded documents and handwritten characters or words. Devanagari, an alphabetic script, is used by over 500 million people all over the world. In this study, we are presenting a Devanagari handwritten character recognition system using Artificial Neural Network. Up to 96% recognition rate is achieved for certain characters.

Key words: Pattern classification, pattern recognition, soft computing

INTRODUCTION

India is multilingual country of more than 1 billion population with 18 constitutional languages and 10 different script. Devanagari, an alphabetic script, is used by a number of Indian languages. It was developed to write Sanskrit but was later adapted to write many other languages such as Marathi, Hindi, Konkani and Nepali. Many other Indian languages use close variant of this script (Masica, 1991). Although, Sanskrit is an ancient language and no longer spoken, written material still exists. Hindi is world's third most commonly used language after English and Chinese and there are approximately 500 million people all over the world that speak and write in Hindi. Devanagari has about 11 vowels and 34 consonants. Figure 1 shows Devanagari characters. It is used as the writing system for over 28 languages including Sanskrit, Hindi, Kashmiri, Marathi and Nepali and is used by more than 500 million people. Thus, Devanagari Handwritten character

			:	ऌ	अ	आ	इ	ई	उ	ऊ	ऋ	ॠ	ऒ	ओ	औ	क
ऒ	ऑ	औ	ओ	औ	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट	
ठ	ड	ढ	ण	त	थ	द	ध	न	न	प	फ	ब	भ	म	य	
र	र	ल	ळ	ळ	व	श	ष	स	ह				ऽ	।	ि	
ी									ि	ी	ो	ी				
ँ									क	ख	ग	ज	ड	ढ	फ	य

Fig. 1: Devnagari characters

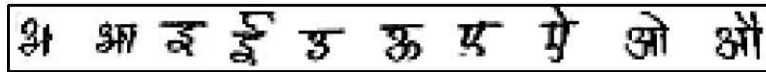


Fig. 2: Selected devanagari characters

recognition is an open field of research which still has large amount of scope for developments. (Bahlmann *et al.*, 2004).

A few models that have been applied for the hand written character recognition system include structure-based models (Aparna *et al.*, 2004; Chan and Yeung, 1998), stochastic models (Li *et al.*, 1998) and learning-based models (Manke and Bodenhausen, 1994). Learning-based models have received wide attention for pattern recognition problems. Neural network models have been reported to achieve better performance than other existing models in many recognition tasks. Support vector machines have also been observed to achieve reasonable generalization accuracy, especially in implementations of handwritten digit recognition (Bin *et al.*, 2000) and character recognition in Roman (Bahlmann *et al.*, 2004), Thai (Sanguansat *et al.*, 2004) and Arabic (Bentounsi and Batouche, 2004) scripts. We have also attempted successfully to recognize Marathi numerals by using Artificial Neural Network (Khanale, 2010a).

As many Indian languages have a similar character set, developing a recognition engine for one Indian language serves as a framework for others as well. Handwritten character recognition for any Indian writing system is rendered complex because of the presence of composite characters.

THE PROBLEM

A recognition system is to be developed for recognition of handwritten devanagari characters by using artificial neural network. The ten selected devanagari characters are given in Fig. 2. Any particular character from data sheet can be selected. The selected character need to be preprocessed and is converted into 5 by 7 matrix of Boolean values. It is further classified into a class based on its unique feature value by the artificial neural network. The input characters may contain noise or they may differ in shape as per the style of writing of a person. We expect the system to classify reasonably well.

DESIGN OF RECOGNITION SYSTEM

Data collection: A special sheet is designed for data collection. Data is collected from people domain with 10 samples of each character from about 40 persons from different fields and age. Data acquisition is done manually, i.e., data collection for the experiment has been done from the different individuals. Writers were provided with the plain A4 sheet and each writer has asked to write Devanagari characters from अ to औ for one time. The collected documents are scanned using HP-scan jet 5400c at 300 dpi which is usually a low noise and good quality image. The digitized images are stored as binary images in BMP format. A sample of Devanagari handwritten characters from the data set is shown as shown in Fig. 3.

Block diagram of recognition system: Figure 4 shows the basic block diagram of the recognition system.

The hand written devanagari characters are scanned and a digitized document is obtained. Out of the available characters, a particular character is selected by using segmentation. The image of character is cropped and it is resized to fix rows and columns. The result is that each character is

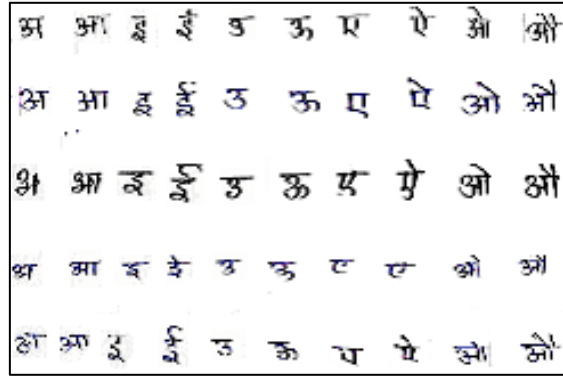


Fig. 3: Data collection

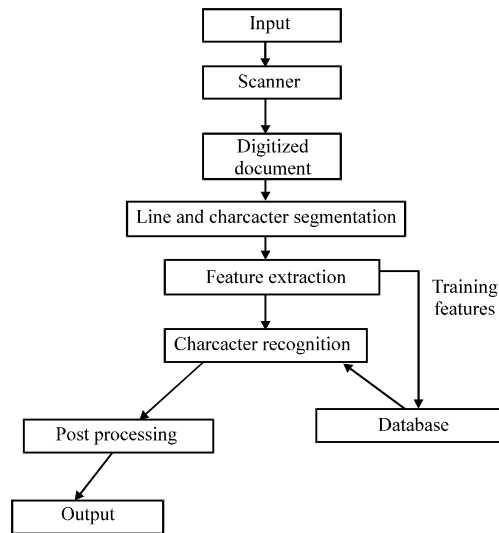


Fig. 4: Basic block diagram of recognition system

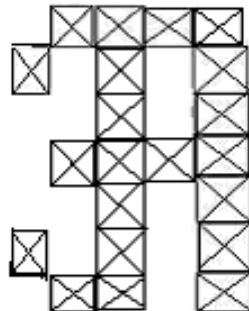


Fig. 5: Representation of character अ

represented by 5 by 7 grid of Boolean values. For example, the character अ is represented as shown in Fig. 5.

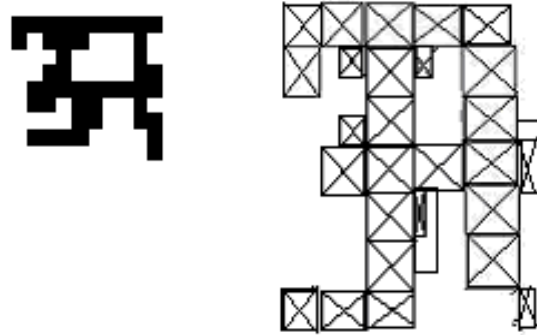


Fig. 6: Image of handwritten character ३ and its representation

However, images of handwritten characters differ from that of ideal one and contain certain different shapes as per style of writing. The image of handwritten ३ and its 5 by 7 representation is as shown in Fig. 6.

For the purpose of recognition a trained artificial neural network is used. The network is trained for a set of ideal and handwritten characters. The network determines a unique feature value for each character. It is further compared with that of ideal one to determine recognition. A reasonably well recognition is expected from the network.

To determine the feature value, character image is decomposed into directional planes. Each directional plane is partitioned into equal sized zones and then sum of pixel values in each zone is taken to have feature value.

NEURAL NETWORK AND ITS TRAINING

The neural network receive 35 Boolean values as a 35 element input vector. It responds to it, by outputting 10 element vector output with 1 in numeral position and 0 elsewhere. Also, the network should recognize handwritten characters, that is, network must make few mistakes with noise and different style of writing of hand written characters.

The selected architecture of the network is a two layer feed forward network with 10 neurons each. The transfer function is log-sigmoid. The training of the network is done with backpropagation algorithm. The network is trained with batch propogation with adaptive learning rate. The performance function used is sum squared error. The goal was set to 0.1. Figure 7 shows training of network. For more details about network, one can refer to our publication on Marathi numerals (Khanale, 2010b).

The network is trained with ideal vectors until it has a 0.1 sum squared error. Then the network is trained with 10 sets of ideal and noisy vectors. For noisy vectors the goal was set to 0.2. Figure 8 shows training of the network with noise. After training, the network is ready to use.

Testing of network: A GUI application is developed where the scanned image of hand written characters can be loaded. Any particular character from the set of characters can be cropped and it is further pre processed to have binary 5 by 7 representation of the character. Further, by using the trained network one can evaluate the feature value of character.

Figure 9 shows feature value of ideal ३. Figure 10 shows feature value of handwritten ३.

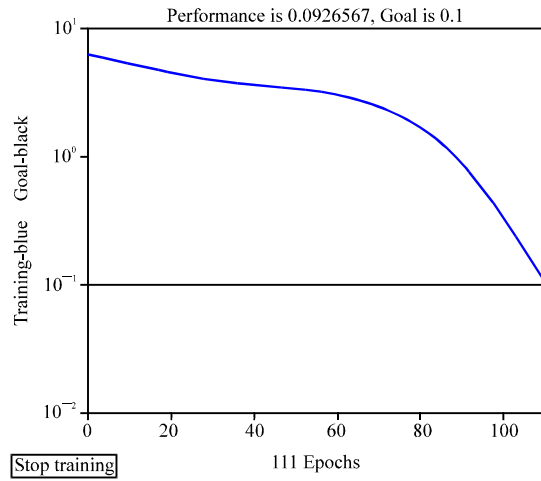


Fig. 7: Training of the network

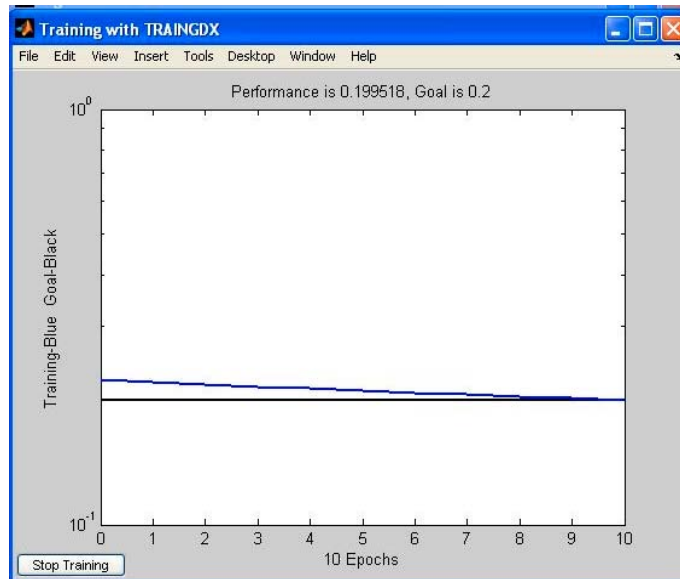


Fig. 8: Training of the network with noise

Table 1: Feature values of printed ideal characters

अ	आ	इ	ई	उ	ऊ	ए	ऐ	ओ	औ
38	53	86	41	93	60	74	30	63	95

RESULTS AND DISCUSSION

The performance of the network is given in Fig. 11. Observe that up to noise level of 0.25 recognition is 100%. Table 1 shows feature values of ideal printed characters. Table 2 shows feature values of handwritten characters. Table 3 gives recognition rate which is based on difference between feature values of ideal and handwritten characters. In Table 3, the last row indicates the

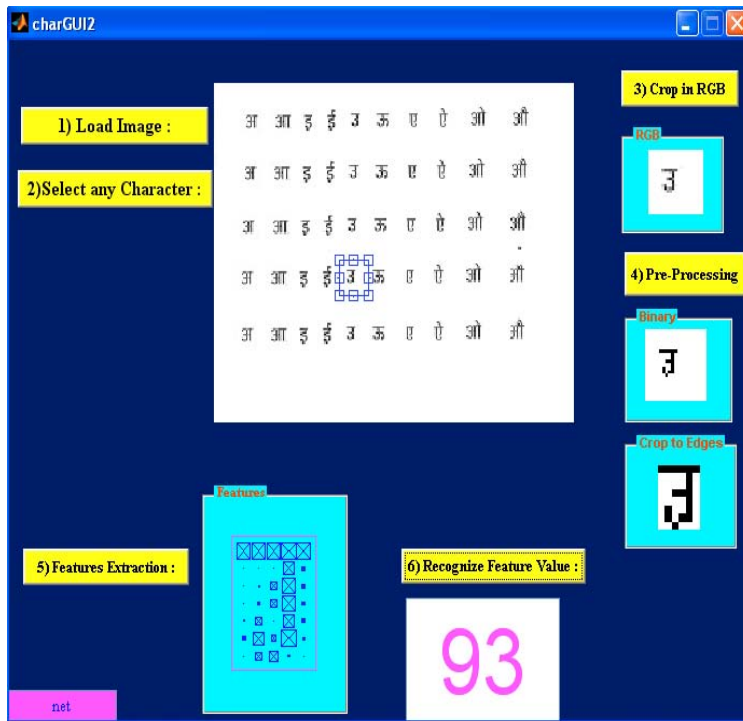


Fig. 9: Feature value of ideal character उ

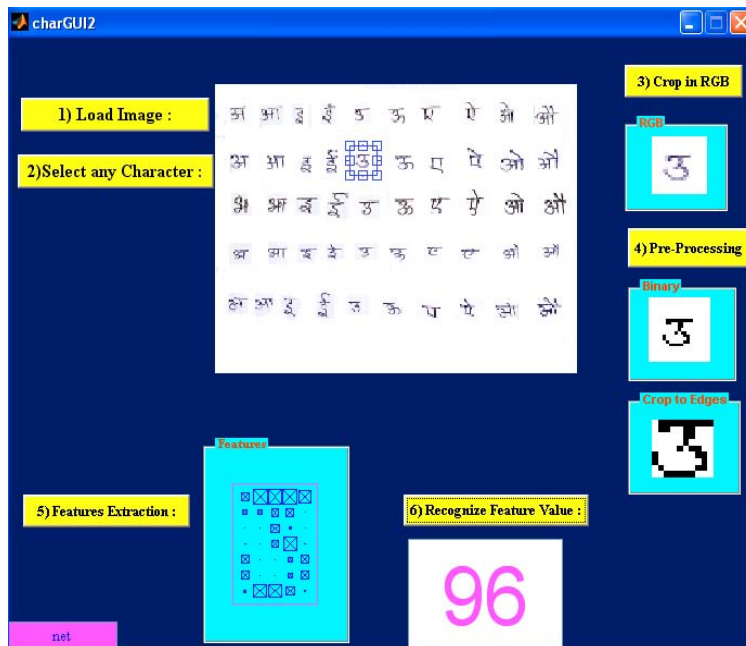


Fig. 10: Feature value of handwritten character उ

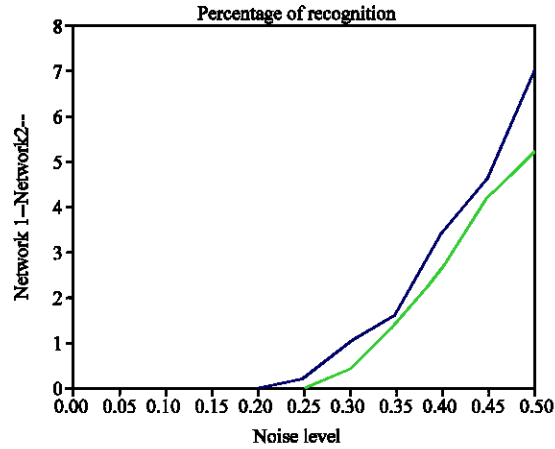


Fig. 11: Performance of the network

Table 2: Feature values of handwritten characters

अ	आ	इ	ई	उ	ऊ	ए	ऐ	ओ	औ
54	81	90	39	99	70	62	24	57	70
34	61	88	42	95	43	65	31	44	59
76	75	88	45	96	65	63	31	49	91
26	52	89	68	96	53	76	26	62	95
41	79	93	40	95	52	53	90	21	55

Table 3: Recognition rate of handwritten characters

अ	आ	इ	ई	उ	ऊ	ए	ऐ	ओ	औ
84.0	72	96.0	98	94	90.0	88	94.0	94.0	75
96.0	92	98.0	99	98	83.0	91	99.0	81.0	64
62.0	78	98.0	96	98	95.0	89	99.0	86.0	96
88.0	99	97.0	73	97	93.0	98	96.0	99.0	100
97.0	74	93.0	99	98	92.0	79	40.0	58.0	60
85.4	83	96.4	93	97	90.6	89	85.6	83.6	79

average of recognition rate. Observe that up to 96% recognition is achieved for handwritten devanagari characters. The neural network method described here can also be extended to other Indian scripts.

REFERENCES

- Aparna, K.H., V. Subramanian, M. Kasirajan, G. Vijay Prakash, V.S. Chakravarthy and S. Madhvanath, 2004. Online handwriting recognition for Tamil. Proceedings of the 9th International Workshop on Frontiers in Handwriting Recognition, (IWFHR'04), Tokyo, Japan, pp: 438-443.
- Bahlmann, C., B. Haasdonk and H. Burkhardt, 2004. Online handwriting recognition with support vector machines - a kernel approach. IEEE Transactions Pattern Analysis Machine Intelligence, 26: 299-310.
- Bentounsi, H. and M. Batouche, 2004. Incremental support vector machines for handwritten Arabic character recognition. Proceedings of the International Conference on Information and Communication Technologies. pp: 1764-1767.

- Bin, Z., L. Yong and X. Shao-Wei, 2000. Support vector machine and its application in handwritten numeral recognition. Proceedings of the 15th International Conference on Pattern Recognition, Sept. 3-8, Barcelona, Spain pp: 720-723.
- Chan, K.F. and D.Y. Yeung, 1998. Elastic structural mapping for online handwritten alphanumeric character recognition. Proceedings of 14th International Conference on Pattern Recognition, Brisbane, (ICPRB'98), Australia, pp: 1508-1511.
- Khanale, P.B., 2010a. Recognition of marathi numerals using artificial neural network. *J. Artificial Intell.*, 3: 135-140.
- Khanale, P.B., 2010b. Recognition recognition system based on the properties and architectures of the human motor system. Proceedings of the International Workshop on Frontiers in Handwriting Recognition, (FHRtures of the human motor system. Proceedings of the International Workshop on Frontiers in Handwriting Recognition, (FHR'90), Montreal, pp: 195-211.
- Li, X., R. Plamondon and M. Parizeau, 1998. Model-based online handwritten digit recognition. Proceedings of 14th International Conference On Pattern Recognition, Aug. 16-20, Brisbane, Australia, pp: 1134-1136.
- Manke, S. and U. Bodenhausen, 1994. A connectionist recognizer for online cursive handwriting recognition. Proceedings of International Conference on Acoustics, Speech and Signal Processing, ICASSP 94, April 12-19, University of Karlsruhe, pp: 433-436.
- Masica, C.P., 1991. *The Indo-Aryan Languages*. Cambridge University Press, Cambridge, ISBN: 9780521299442.
- Sanguansat, P., W. Asdornwised and S. Jitapunkul, 2004. Online Thai handwritten character recognition using hidden Markov models and support vector machines. Proceedings of the International Symposium on Communications and Information Technologies, 2004, Oct. 26-29, Japan, pp: 492-497.