IJCSC International Journal of Computer Science & Communication Vol. 1, No. 2, July-December 2010, pp. 141-144

Handwritten English Character Recognition Using Neural Network

Anita Pal[1] & Dayashankar Singh[2]

[1,2]Department of Computer Science & Engineering, U.P.Technical University, Lucknow, India
Email: [1]anitapal13@gmail.com, [2]dss_mec@yahoo.co.in

ABSTRACT

In this paper, work has been performed to recognize Handwritten English Character using a multilayer perceptron with one hidden layer. The feature extracted from the handwritten character is Boundary tracing along with Fourier Descriptor. Character is identified by analyzing its shape and comparing its features that distinguishes each character. Also an analysis was carried out to determine the number of hidden layer nodes to achieve high performance of backpropagation network in the recognition of handwritten English characters. The system was trained using 500 samples of handwritings given by both male and female participants of different age groups. Test result was performed on 500 samples other than samples for training that indicates that Fourier Description combined with backpropagation network provide good recognition accuracy of 94% for handwritten English characters with less training time.
Keywords: Handwritten Character Recognition, Feature Extraction, Backpropagation network, Boundary Tracing, Fourier Descriptor and Multilayer Perceptron Network.

1. INTRODUCTION

Neural Networks are recently being used in various kind of pattern recognition. Handwritings of different person are different; therefore it is very difficult to recognize the handwritten characters. Handwritten Character recognition is an area of pattern recognition that has become the subject of research during the last some decades. Neural network is playing an important role in handwritten character recognition. Many reports of character recognition in English have been published but still high recognition accuracy and minimum training time of handwritten English characters using neural network is an open problem. Therefore, it is a great important to develop an automatic handwritten character recognition system for English language [1]. In this paper, efforts have been made to develop automatic handwritten character recognition system for English language with high recognition accuracy and minimum training and classification time. Experimental result shows that the approach used in this paper for English character recognition is giving high recognition accuracy and minimum training time.

2. CHARACTER MODELING

2.1. English Characters

The English language consists of 26 characters (5 vowels, 21 consonants) and is written from left to right. A set of hand written English characters is shown in Figure 1.

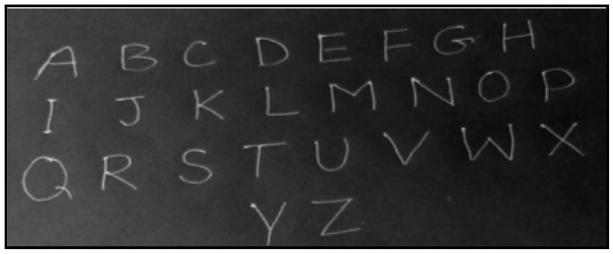

Fig. 1: A Set of Handwritten English Characters

2.2. Scanning and Skeletonization

Handwritten characters are scanned and it has been converted into 1024 (32X32) binary pixels. The skeletonization process will be used to binary pixel image and the extra pixels which are not belonging to the backbone of the character has been deleted and the broad strokes has been reduced to thin lines [1]. Skeletonization is illustrated in Figure 2.

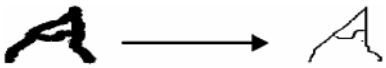

Fig. 2: Skeletonization of an English Character

2.3. Normalization

There are lots of variations in handwritings of different persons. Therefore, after skeletonization process, normalization of characters is performed so that all characters could become in equal dimensions of matrix. In this paper, characters are normalized into 30X30 pixel character and shifted to the top left corner of pixel window.

3. CHARACTER RECOGNITION SYSTEM

The block diagram of Character recognition system is shown in following Figure 3.

Fig. 3: Block Diagram of the Recognition System

The procedure of handwritten English character recognition is as follows:

- Acquire the sample by scanning.
- Skeletionization and Normalization operations are performed.
- Apply Boundary Detection Feature Extraction technique.
- Neural network Classification.
- Recognized Character.

4. FEATURE EXTRACTION

In this paper, to extract the information of the boundary of a handwritten character, the eight-neighbor adjacent method has been adopted. This scans the binary image until it finds the boundary. The searching follows according to the clockwise direction. For any foreground pixel p, the set of all foreground pixels connected to it is called connected component containing p. The pixel p and its 8-neighbors are shown in Figure 4. Once a white pixel is detected, it checks another new white pixel and so on. The tracing follows the boundary automatically. When the first pixel is found, the program will be assigned the coordinates of that position to indicate that this is an origin of the boundary. The new found pixel will be assigned as a new reference point and starts the eight-neighbor searching. In this way, the coordinates of the initial point are varied according to the position. As the tracer moves along the boundary of the image, the corresponding coordinates will be stored in an array for the computation of Fourier Descriptors. During the boundary tracing process, the program will always check the condition whether the first coordinates of the boundary are equal to the last coordinates. Once it is obtained; means the whole boundary has been traced and boundary tracing process completes [2].

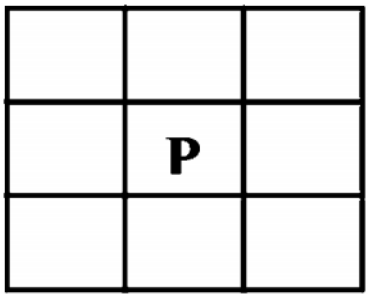

Fig. 4: Pixel P and Its 8-neighbor

Fig. 5: Boundary Tracing of a Character

4.1. Fourier Descriptors

Fourier Descriptors are involved in finding the Discrete Fourier coefficients a[k] and b[k] for $0 \le k \le L-1$,

$$a[k] = 1/L \sum x[m]e^{-jk(2\pi/L)m} \quad (1)$$

$$b[k] = 1/L \sum y[m]e^{-jk(2\pi/L)m} \quad (2)$$

Fourier coefficients derived according to equations (1) and (2) are not rotational or shift invariant but Fourier Descriptors that have the invariant property with respect to rotation and shift,the following operations are defined.For each n compute a set of invariant descriptors $r(n)$

$$r(n) = [\,|a(n)|^2 + |b(n)|^2]^{1/2} \quad (3)$$

Computing a new set of descriptors $s(n)$ by eliminating the size of character from $r(n)$

$$s(n) = r(n)/r(1) \quad (4)$$

$a(n)$, $b(n)$ and invariant descriptors $s(n)$, $n = 1,2,\ldots..(L-1)$ were derived for all of the characters.

5. NEURAL NETWORK

5.1. Recognition

Recognition of handwritten characters is a very complex problem. The characters could be written in different size, orientation, thickness, format and dimension. This will give infinite variations. The capability of neural network to generalize and insensitive to the missing data would be very beneficial in recognizing handwritten characters. In this paper, for English handwritten character recognition in neural Feed Forward Multi-Layer Perceptron network (MLPN) with one hidden layer has been used. For training, back-propagation algorithm has been implemented [1, 5].

5.2. Multilayer Perceptron Network

The multilayer perceptron neural networks with the EBP algorithm have been applied to the wide variety of problems. In this paper, two-layer perceptron i.e., one hidden layer and one output layer has been used [5]. Structure of MLP network for English character recognition is shown in Figure 6.

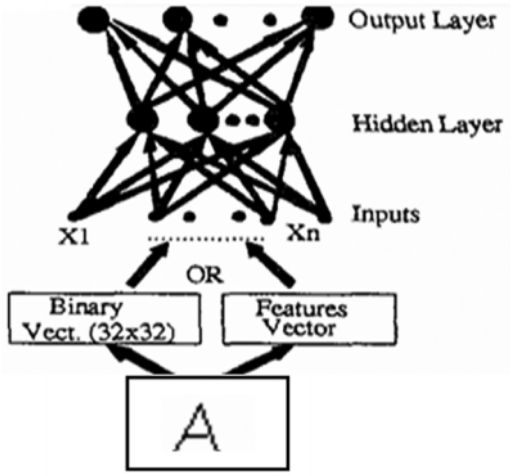

Fig. 6: Multilayer perceptron Network (MLPN)

In MLPN with Backpropagation training algorithm, the procedure and calculations as follows:

$$f_j(x) = 1/(1 + e^{-\text{net}}) \text{ and net} = \Sigma w_{ij} o_i$$

where o_i is the output of unit i, w_{ij} is the weight from unit i to unit j. The generalized delta rule algorithm was used to update the weights of the neural network in order to minimize the cost function:

$$E = \tfrac{1}{2}(\Sigma(D_{pk} - O_{pk}))^2$$

where D_{pk} and O_{pk} are the desired and actual values respectively, of the output unit k and training pair p.Convergence is achieved by updating the weights using following formulas:

$$W_{ij}(n + 1) = W_{ij}(n) + \Delta W_{ij}(n) \quad (1)$$

$$\Delta W_{ij}(n) = \eta \delta X_j + \alpha(W_{ij}(N) - W_{ij}(n - 1)) \quad (2)$$

where η is the learning rate, α is the momentum, $W_{ij}(n)$ is the weight from hidden node i or from an input to node j at nth iteration, X_i is either the output of unit i or is an input, and δ_j is an error term for unit j [1]. If unit j is an output unit, then

$$\delta_j = O_j(1 - O_j)(D_j - O_j)$$

If unit j is an internal hidden unit, then

$$\delta_j = O_j(1 - O_j)\ \Sigma \delta_k W_{kj}$$

6. EXPERIMENTAL RESULTS

6.1. Character Database

Five hundred samples were collected from 10 person,50 samples each, out of which 250 samples were used for training (training data) and 250 samples were used for testing the data (test data).

6.2. Procedure and Flowchart

A complete procedure of handwritten English character recognition is given below

- Capture the scanned characters.
- Perform the Normalization process.
- Perform Binarization.
- Apply Feature Extraction Techniques (Boundary tracing technique).
- Implement the Neural Network Classifier.
- Get the recognized character.

A complete flowchart of handwritten English character recognition is given below in Figure 7

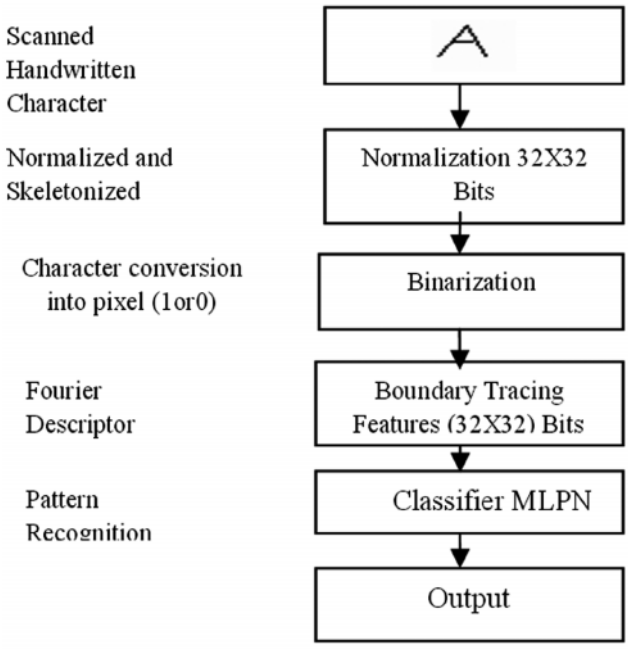

Fig. 7: A System for English Character Recognition

6.3. Results

An analysis of experimental result has been performed and shown in table 1.

Table 1
Result of Handwritten English Character using MLPN

No. of Hidden nodes (neurons)	*Learning Rate*	*Momentum Factor*	*No. of Epochs*	*Recognition %*	
				Training Set	*Test Set*
12	0.2	0.8	50	100	89
24	0.2	0.8	100	100	94
36	0.2	0.8	200	100	94

7. CONCLUSION

In this paper, a system for recognizing handwritten English characters has been deveoped. An experimental result shows that Fourier descriptors with back propagation networkyields good recognition accuracy of 94%. The skeletonized and normalized binary pixels of English characters were used as the inputs of the MLP

network. The results of structure analysis shows that if the number of hidden nodes increases the number of epoches taken to recognize the handwritten character is also increases. A lot of efforts have been made to get higher accuracy but stil there are tremendous scope of improving recognition accuracy by developing new feature extraction techniques or modifying the existing feature extraction techniques.

REFERENCES

[1] Verma B.K, "Handwritten Hindi Character Recognition Using Multilayer Perceptron and Radial Basis Function Neural Network", *IEEE International Conference on Neural Network*, **4**, pp. 2111-2115, 1995.

[2] Sutha.J, Ramraj.N, "Neural Network Based Offline Tamil Handwritten Character Recognition System", *IEEE International Conference on Computational Intelligence and Multimedia Application*, 2007, **2**, 13-15, Dec.2007, Page(s): 446-450, 2007.

[3] Yuelong Li Jinping Li Li Meng, "Character Recognition Based on Hierarchical RBF Neural Networks" *Intelligent Systems Design and Applications*, 2006. ISDA '06. Sixth International Conference, **1**, On Page(s): 127-132, 2006.

[4] Dayashankar Singh, Maitreyee Dutta and Sarvpal Harpal Singh, "Comparative Analysis of Handwritten Hindi Character Recognition Technique", *IEEE International Advanced Computing Conference (IACC'09)*, March 6-7, 2009, Thaper University, Patiala.

[5] Dayashankar Singh, Maitreyee Dutta and Sarvpal H. Singh, "Neural Network Based Handwritten Hindi Character Recognition", *ACM International Conference (Compute 09)*, Jan. 9-10, 2009, Bangalore.

[6] F.Kimura, T.Wakabayashi, S.Tsuruoka, and Y.Miyake,"Improvement of Handwritten Japanese Character Recognition using Weighted Direction Code Histogram," *Pattern Recognition*, **30**, No.8, pp. 1329-1337, 1997.

[7] N.Kato, M.Suzuki, and S.Omachi, "A Handwritten Character Recognition System Using Directional Element Feature and Asymmetric Mahalanobis Distance", *IEEE Trans. on PatternAnalysis and Machine Intelligence*, **21**, No.3, pp. 258-262, 1999.

[8] C.C. Tappert, C.J. Suen and T. Wakahara, "The State of the Art in Outline Handwriting Recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **PAMI-12**, No.8, pp.707-808, 1990.

[9] D.S. Yeung, "A Neural Network Recognition System for Handwritten Chinese Character Using Structure Approach," *Proceeding of the World Congress on Computational Intelligence*, **7**, pp. 4353-4358, Orlando, USA, June 1994.

[10] D.Y. Lee, "Handwritten Digit Recognition Using K Nearest-neighbor, Radial Basis Function and Backpropagation Neural Networks, "*IEEE Neural Computation*, **3**, Page(s) 440- 449.

[11] Y.Y. Chung, and M.T. Wong, "Handwritten Character Recognition by Fourier Descriptors and Neural Network", *Proceedings of IEEE TENCON – Speech and Image Technologies for C*.

[12] H. Almualim and S. Yamaguchi, "A Method for Recognition of Arabic Cursive Handwriting," *IEEE Trans. on Pattern and Machine Intelligence*, **9**, No 5, pp.715-722, Sept. 1987.

[13] I.S.I. Abuhaiba and S.A. Mahmoud, "Recognition of Handwritten Cursive Arabic Characters," *IEEE Transaction on PA&MI*, **16**, No 6, pp. 664-672, June 1994.