IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.6, Issue.10, Oct. 2018 E-ISSN: 2347-2693

Hand Written English Character Recognition using Pattern Sampling Recognition Technique (PSRT)

Rakesh Kumar Mandal

Dept. of Computer Science & Application, University of North Bengal, India

*Corresponding Author: rakesh_it2002@yahoo.com

Available online at: www.ijcseonline.org

Accepted: 26/Sept/2018, Published: 31/Oct/2018

***Abstract* –** This is the era of intelligent computing devices. Efforts are going on in all over the world to develop machines and programs which can solve the problems that human beings can solve with ease. One such field is recognition of handwritten characters by computers. In this paper the neural network is first trained using perceptron-learning algorithm. The target pattern is a collection of distinct patterns set for each character. While testing the target pattern is sampled and the distortion in the sampled pattern was compared with the original one. 30% or less of such distortion was considered for the identification of a particular character. The results showed that such methods produce accuracies of at least 90% and more for the hand written upper case English alphabets.

***Keywords*-** Character recognition, Sampling, Perceptron, Learning Algorithm, Neural Network

I. INTRODUCTION

In this fast corporate world it has become cumbersome to waste time in getting the documents typed from a typist. Hand written character recognition software recognizes the characters written by an executive or somebody else. But it is very difficult to design software which is capable of identifying the characters with great accuracy. K-nearest neighbor methods can be used to recognize the patterns [1]. In K-nearest neighbor method the pattern is obtained by looking into k number of nearest patterns having the least Euclidean distance with that of the pattern.

Artificial Neural Nets are used to train the nets and later using the nets identifying the characters [2]. But obtaining 100% accuracy is still a challenge for many such nets.

Variation in handwriting leads to great difficulty in identifying the character patterns. Different writing styles lead to the distortion in patterns giving false results. A strong generalization method was required to identify the distorted patterns.

Multiscale Training Technique (MST) is used in many places to solve the generalization problem [3,4,5]. Methods become very complex to solve this kind of problem. A simple approach can be used to solve the problem. An approach was made to solve the problem by using single layer net.

Work has been done on identification of the characters in Devnagri script by combining multiple feature extraction techniques like intersection, shadow feature, chain code histogram and straight line fitting, [6].

Another approach towards feature extraction technique was to calculate only twelve directional feature inputs depending upon the gradients, where the features of the hand written characters were the directions of the pixels with respect to their neighboring pixels, [7].

Hybrid methods were also applied to recognize the hand written characters. One such method was a prototype learning/matching method that can be combined with support vector machines (SVM) in pattern recognition, [8].

In this paper a new concept of recognizing hand written character pattern is developed and implemented called Pattern Sampling Recognition Technique (PSRT). Distinct character target patterns are taken for different characters. Each distinct target pattern (say x) is further taken as a single vector, which is a group of 10, such patterns (xxxx…10 times).

In general the overall recognition technique is divided into two main sections, namely training and testing. Training requires training the net with PSRT target patterns. Testing requires the net with different handwritten patterns and checks the accuracy with which the net identifies the patterns.

II. METHODOLOGY

The characters were written on a piece of A4 size paper with uniform square sized boxes Figure 1. Each character is written in one box. Blue/Black ballpoint pen is used to write the characters on the paper. 26 uppercase letters of English alphabets were taken for the training purpose. Using either a scanner or a good resolution digital camera captured these characters. The captured images were then used to generate input vectors, which were fed to the perceptron neural network for the training. Pattern Sampling Recognition Technique (PSRT) was then used for setting the targets. Weights were calculated by using the above-mentioned technique (PSRT). Various patterns written by many people were taken to test the network. MATLAB was used to test the results and the accuracy was measured which was found 95%.

III. PREPARATION OF INPUT PATTERNS

The characters prepared as explained in section 2, are scanned using a scanner or a good resolution digital camera. The following example explains this:

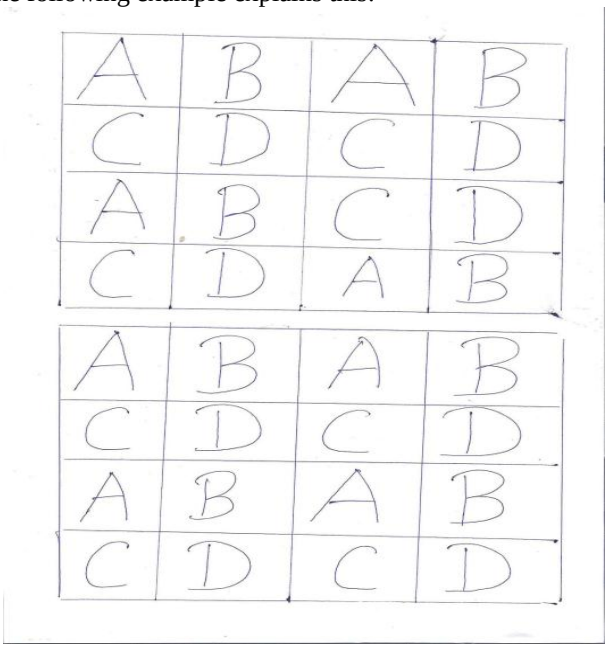

Figure 1. Training character set

The captured or scanned images are in RGB scale. These images have to be converted into grayscale format before further processing can be done. Each character was en-caged in a square of size 79 x 79 pixels by using MATLAB programming logic. Finally, 26 different vectors of size (79 x 79) were produced in the binary format. Later on all the 0s of the vectors were replaced by –1s for the better calculation. Perceptron learning rule [2] has been implemented in each sample vector.

IV. PATTERN SAMPLING RECOGNITION TECHNIQUE (PSRT)

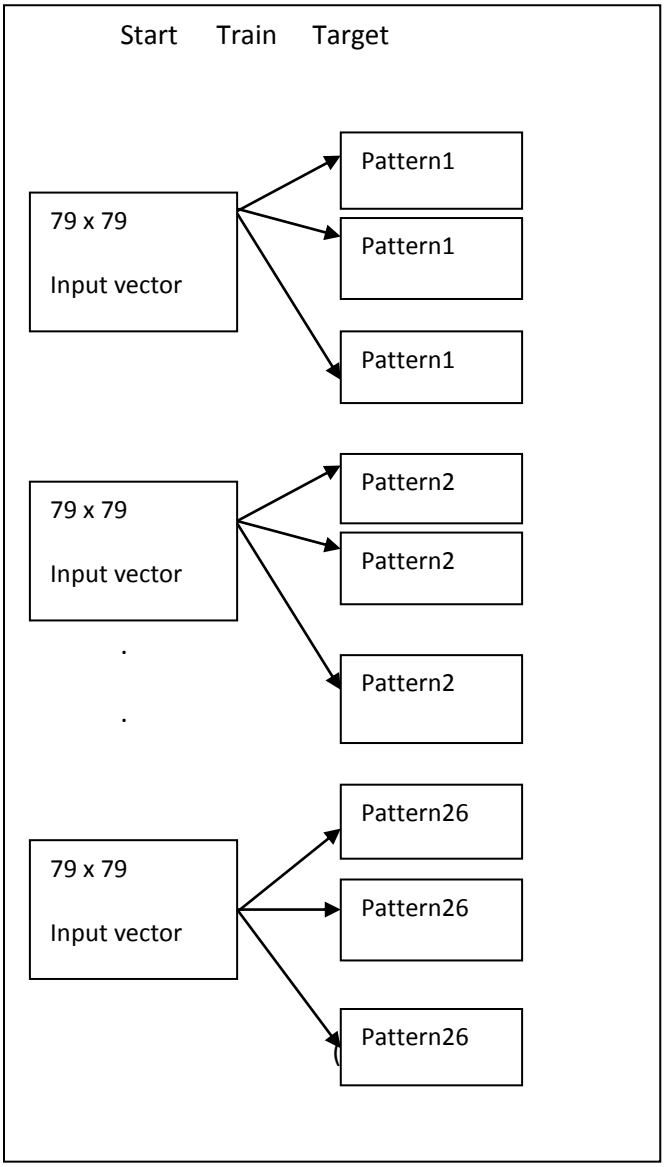

Figure 2. Conceptual diagram of PSRT training

The training started with 79 x 79 pixel input vector Figure 2. These input vectors were fed into the neural network for training. After being trained for a few epochs, the neural network adjusts its weight to produce the targets. The targets were designed in the following way Figure 3:

Pattern1	Pattern1	…	Pattern1
Pattern2	Pattern2	…	Pattern2
……..	…….	…	……..
Pattern26	Pattern26	…	Pattern26

Figure 3. Target vector format (79 x 79 x 10)

Then the sample characters with distortions were fed to the neural network for testing Figure 4. Slight distortions give deviated target outputs. The extent of distortions was checked to produce the results Figure 5.

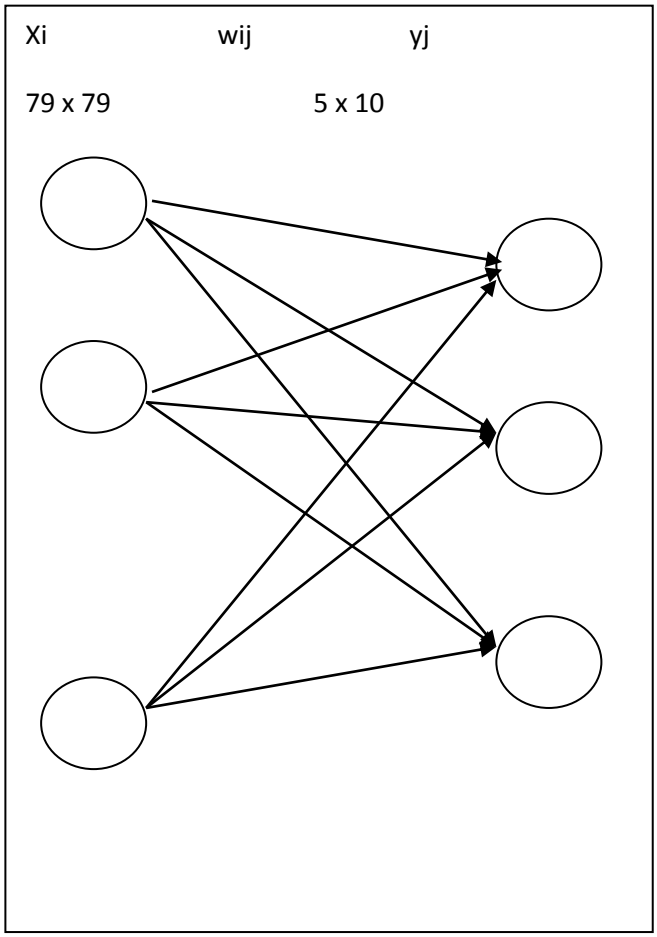

Figure 4. Neural network used for PSRT

In this paper a new concept of sampling distance was developed. Sampling distance is the difference between the number of similar patterns in the target vector for a particular character and the resultant vector found after testing.

$$sd_i = pattern(T_i) - pattern(R_i) \quad (1)$$

Where sd_i is the sampling distance, $pattern(T_i)$ is the sample pattern in the target vector and $pattern(R_i)$ is the sampled pattern in the resultant vector.

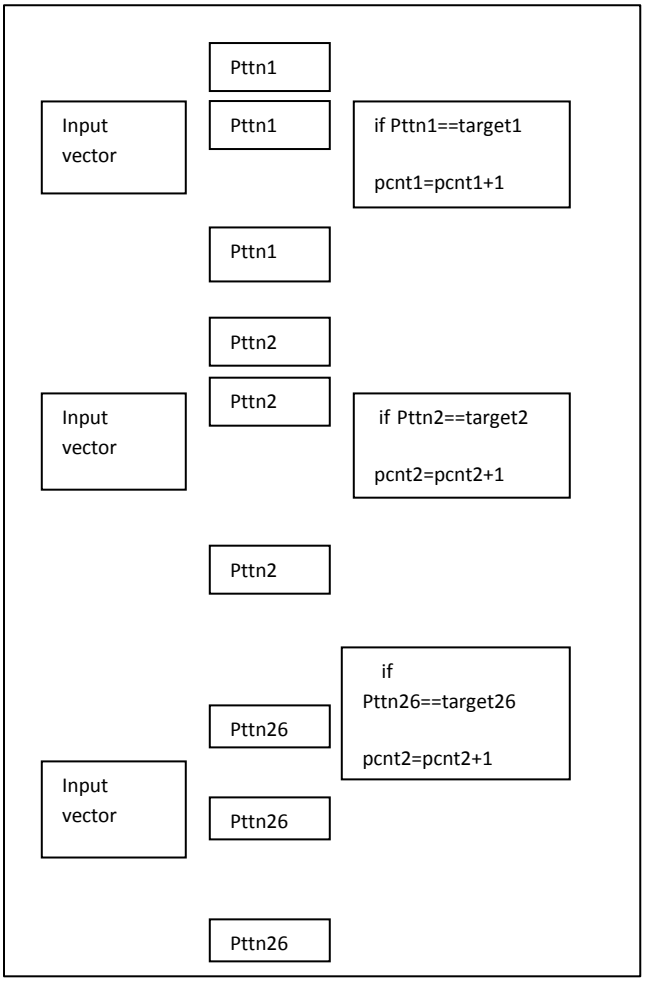

Figure 5. Conceptual diagram for PSRT testing

V. TRAINING AND TESTING METHODOLOGY

The characters were written on a plain A4 size paper. The paper was scanned using a scanner. The image file was read and stored in a matrix. The image file was converted into grayscale. The grayscale image was converted into binary matrix. The binary matrix was divided into sub matrices holding individual characters which were converted into single row vectors. The row vectors were fed to the net one by one for training. The training algorithm used was perceptron learning algorithm. The sample characters were read using the same method and fed to the net for testing. The target patterns were set in such a way that a portion of wrong output will not affect the result.

VI. RESULT ANALYSIS

The method was tested by MATLAB programming. Few characters were taken initially. The accuracy of the method was measured by testing its capability of identifying the characters against the percentage of distortion of the sample character.

Neural Network parameters used in the experiment
Number of neurons in the input unit=1521
Number of neurons in the output unit=50
Number of Epoch=3
Training Algorithm used=Perceptron

Table 1. shows at what percentage of distortion of the sample characters the net is capable of identifying the characters. For characters A, B, C & D were taken initially for training and testing.

Table 1. Identification of characters to the extent of distortion

Identified Characters	Percentage of distortion				
	10%	20%	30%	40%	50%
A	YES	YES	YES	NO	NO
B	YES	YES	YES	NO	NO
C	YES	YES	YES	YES	NO
D	YES	YES	YES	YES	NO

VII. DISCUSSION

The results show that PSRT training allows very fast convergence. The number of epochs required for the training is very less. The accuracy of identifying the characters is also very good.

However, the method was tested for some sample characters which were totally deviated from their positions. The method failed to produce correct outputs for the same as the total deviations from the position produces completely different training pattern and hence a completely different target pattern.

VIII. CONCLUSION

PSRT gives very good accuracy, if the characters were written in boxed sheets. In this paper the method applied used the logic of encaging the characters without using the boxed sheets but the logic provides static encaging. Problems in identifying the characters arises when the characters gets fully deviated from their positions on the sheet. Efficient algorithm is still to be explored to en-cage the characters written on any position of the paper. Some techniques were already being developed. But work is going on to develop some simple technique to solve the problem. Another, problem is to find out some generalization among variations in sizes of the characters using very simple methods.

REFERENCES

[1] L van der Maaten-2009, A New Benchmark Data Set for Handwritten Character Recognition, Available: www.tilburguniversity.nl/faculties/humanities/ticc/.../TR2009002.pdf, (Accessed: 2010, July 19th)

[2] Brown, E.W. (1993), Applying Neural Networks to Character Recognition, Available: http://www.ccs.neu.edu/home/feneric/charrecnn.html (Accessed: 2010, July 19th).

[3] Robinson, G. (1995), The Multiscale Technique, Available: http://www.netlib.org/utk/Isi/pcwLSI/text/node123.html

[4] Handwritten Character Recognition, Available: http://tcts.fpms.ac.be/rdf/hcrinuk.htm

[5] Velappa Ganapathy, and kok Leong Liew, Handwritten Character Recognition Using Multiscale Neural Network Training Technique, World Academy of Science, Engineering and Technology 39 2008

[6] Sandhya Arora, Debotosh Bhattacharjee, Mita Nasipuri, Dipak kumar Basu and Mahantapas Kundu, Combining Multiple Feature Extraction Techniques for Handwritten Devnagri Character Recognition, Available: http://arxiv.org/ftp/arxiv/papers/1005/1005.4032.pdf, (Accessed: 26, July, 2010).

[7] Dayashankar Singh, Sanjay Kr. Singh and Dr. (Mrs.) Maitreyee Dutta, Hand Written Character Recognition Using Twelve Directional Feature Input and Neural Network, Available: http://www.ijcaonline.org/journal/number3/pxc387173.pdf, (Accessed: 26, July, 2010).

[8] Fu Chang, Chin-Chin Lin and Chun-Jen Chen, Applying A Hybrid Method To Handwritten Character Recognition, www.iis.sinica.edu.tw/~fchang/paper/931_chang_F.pdf, (Accessed: 26, July, 2010).

