On the Choice of Training Set, Architecture and Combination Rule of Multiple MLP Classifiers for Multiresolution Recognition of Handwritten Characters

U. Bhattacharya[1] S. Vajda[2] A. Mallick[1] B. B. Chaudhuri[1] A. Belaid[2]
[1]CVPR Unit, Indian Statistical Institute, Kolkata, India
[2]READ Group, LORIA Research Center, Nancy, France
Email: ujjwal@isical.ac.in

Abstract

A script independent recognition scheme for handwritten characters using multiple MLP classifiers and wavelet transform-based multiresolution pixel features is presented. We studied four different approaches for combination of multiple MLP classifiers and observed that a weighted majority voting approach provided the best recognition performance. Also, a thumb rule for the selection of network architecture has been obtained and a dynamic strategy for selection of training samples has been studied. The dynamic training set selection approach often makes the training procedure several times faster than the traditional training scheme. In our simulations, 98.04% recognition accuracy has been obtained on a test set of 5000 handwritten Bangla (an Indian script) numerals. Our approach is sufficiently fast for its real life applications and also script independent. The recognition performance of the present approach on the MNIST database for handwritten English digits is comparable to the state-of-the-art technologies.

1. Introduction

Extensive research works on recognition of off-line handwritten characters, in particular numerals, have been carried out since the last few decades because it is a classical problem of pattern recognition and it has enormous application potentials [1]. A wide variety of algorithms / schemes exist for off-line recognition of isolated handwritten characters [2, 3] and each of these has their own merits and demerits.

The existing character recognition schemes differ from each other mainly with respect to feature selection / extraction method and classifier used. Several authors considered neural networks for the classification purpose and many high accuracy character recognition systems are neural network (NN) based [4] mainly because they perform satisfactorily in the presence of incomplete or noisy data and they can learn from examples. LeCun et al. [5] successfully applied constrained backpropagation based network for the numeral recognition task. Knerr et al. [6] showed that only single-layer neural network may be successfully employed for recognition of handwritten digits. The hybrid scheme of [7] successfully used self-organizing feature map for obtaining graph-representation of an input character image and subsequently a sub-grouping was done depending on a few shape features and final classification was performed with the help of MLP classifiers. Shimizu et al. [8] successfully used mirror image learning to improve the recognition accuracy of auto-associative neural networks.

Possibly the most important aspect of a handwriting recognition scheme is the selection of a good feature set which is reasonably invariant with respect to shape variations caused by various writing styles. A large number of feature extraction methods are available in the literature [3]. So instead of proposing another feature extraction method, it seems justified to investigate how an existing feature extraction method(s) can be used along with an intelligent classification strategy to achieve both speed and acceptable recognition accuracies in different scripts.

In this article, we present a detail analysis of a multiresolution recognition approach for isolated handwritten numerals. The present approach does not consider any script specific feature. Smooth. . .Smooth components of the input numeral image at different resolutions of Daubechies' wavelet transform are considered as the features at the respective resolution levels. At each such resolution level a distinct MLP has been used as the classifier. Similar features had been considered before in a cascaded approach [9] providing approximately 93% recognition accuracy on Bangla numeral database. Later, a majority voting scheme [10] improved the recognition accuracy on Bangla numeral database to approximately 97.2%. In the present study, this recognition accuracy has been improved to approximately 98%. This improvement to the recognition accuracy is the result of three modifications over the earlier

Proceedings of the 9th Int'l Workshop on Frontiers in Handwriting Recognition (IWFHR-9 2004)
0-7695-2187-8/04 $20.00 © 2004 IEEE

approach. These modifications consist of (a) using a dynamic selection strategy for training samples; (b) obtaining a thumb rule for the selection of optimal architectures of each MLP at different resolution levels and (c) considering a weighted majority voting scheme for the combination of the outputs of different MLP classifiers. Each of the above three actions contributed to the improvement of the performance. Dynamic training set selection strategy not only significantly reduces the time required for the training of each MLP, additionally it marginally improves the generalization performance. The classification accuracy provided by an individual MLP, largely depends on the number of nodes in its hidden layer(s). During extensive simulation runs, we obtained a thumb rule for determining the hidden layer size providing a near-optimal classification performance. Finally, we considered four different strategies for combining the results of the concerned MLPs and the best result was obtained using a weighted majority voting scheme.

In this report, we present simulation results on two databases of different scripts – (i) MNIST database [11] for handwritten English numerals and (ii) ISI database [12] for handwritten Bangla (the second most popular script in the Indian subcontinent) numerals. Samples from each of these two English and Bangla databases are shown in Fig. 1(a) and Fig. 1(b) respectively. On (i) we obtained 98.546% correct classification and 0.694% rejection on a test set of 10000 English numerals and on (ii) we obtained 98.04% correct classification and 0.74% rejection on a test set of 5000 Bangla numerals. The speed of recognition is more than 60 numerals per second on a PIV desktop computer.

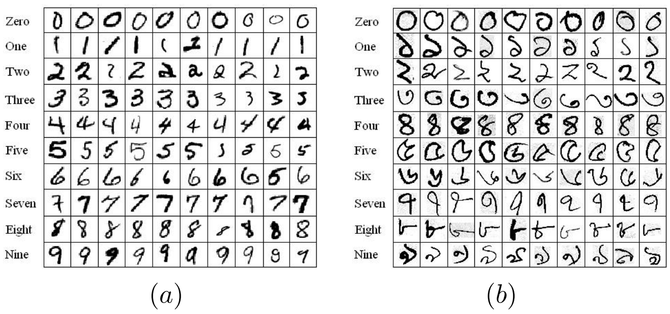

Figure 1. Typical sample sets of handwritten numerals (a) English (b) Bangla

The rest of this article is organized as follows. In Section 2 we describe wavelet transform-based multiresolution features. MLP classifiers and their combination strategies are described in Section 3. The multiresolution recognition scheme and its various issues improving the performance are described in Section 4. Simulation results are reported in Section 5. Section 6 concludes the present article.

2. Wavelet descriptor for multiresolution representation of pixel image

A wavelet transform is orthogonal and operates on an input vector whose length is an integral power of two. This is a fast linear operation. It generates a vector which is of the same length but numerically different from the input vector. Wavelet transform can be viewed as a rotation in function space, from the input domain to a different domain. The basis functions of the wavelet domain are called wavelets. Wavelets are quite localized both in space and in frequency.

There exist infinitely many possible sets of wavelets and different sets of wavelets make different trade-offs between how compactly they are localized in space and how smooth they are. A wavelet transform is usually implemented by a binary tree of filters. The art of finding good wavelets lies in the design of these set of filters which achieve the above trade-offs and also make the perfect reconstruction of the original signal possible.

The working principle of a wavelet transform is as follows. An input signal x is split into a lowpass or smooth component x_0 and a highpass or detail component x_1 respectively by a lowpass filter L and a highpass filter H. Both of these two components are down-sampled in the ratio 2:1. The lowpass component x_0 is then split further into x_{00} and x_{01} by using the above filters for the second time and these are again down-sampled in the ratio 2:1. This process (pyramidal algorithm [13]) of splitting and down-sampling is continued as far as required or a trivial size of the smooth...smooth component (usually 2) is reached.

The first and simplest possible orthogonal wavelet system is the Haar wavelet (Thesis of A. Haar, 1909). However, Daubechies [14] constructed a set of orthonormal wavelet basis function that are perhaps the most elegant. These wavelets are compactly supported in the time-domain and have good frequency domain decay. This describes the reason behind our choice of Daubechies wavelet transform. The simplest member of this family of wavelets is the Daubechies-4 wavelet which has only four coefficients

$$l_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}, l_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}, l_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}, l_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}$$

The above coefficients form the lowpass or smoothing filter L and another set of four coefficients

$$h_0 = l_3, h_1 = -l_2, h_2 = l_1 \text{ and } h_3 = -l_0$$

form the highpass filter H. (In signal processing contexts L and H are called quadrature mirror filters.)

A simple extension of the above principle to multidimensional arrays is possible. A wavelet transform of an image, a 2-dimensional array, is easily obtained by transforming the array on its row index (for all values of its column indices), then on its column. Each transformation corresponds to multiplication by an orthogonal matrix and by matrix associativity, the result is independent of the order in which the row or column are transformed.

The layout of application of wavelet transform recursively on an image is shown in Figure 2. The successive application of the transform produces an increasingly smoother version of the original image.

(a) (b) (c) (d) (e)

Figure 2. (a) **Layout of wavelet decomposition (L $\rightarrow$ low-pass filter, H $\rightarrow$ high-pass filter, $j \rightarrow$ level no.);** (b) **Original image of a Bangla handwritten numeral;** (c) **The size normalized(32×32) image;** (d) **and** (e) **Smooth...smooth components of wavelet decompositions at resolution levels** 16×16 **and** 8×8 **respectively.**

3. MLP classifiers

3.1. BP training of MLP classifiers

To train the MLP network we have used backpropagation algorithm (BP) [15]. This algorithm performs a gradient descent in the connection weight space on an error surface defined by

$$E = \frac{1}{P}\sum_{p} E_p, \quad \text{where,} \tag{1}$$

$$E_p = \frac{1}{2}\sum_{k}(t_{pk} - y_{pk})^2 \tag{2}$$

Here P is the total number of patterns in the training set and $\{t_{pk}\}$, $\{y_{pk}\}$ are respectively, the target and output vectors corresponding to the p-th input pattern. The quantity E is called the system error. In BP algorithm, weight updation rules are given by

$$w_{jk}(t+1) = w_{jk}(t) - \eta\frac{\partial E_p(t)}{\partial w_{jk}(t)} + \alpha\Delta w_{jk}(t-1), \tag{3}$$

$$w_{ij}(t+1) = w_{ij}(t) - \eta\frac{\partial E_p(t)}{\partial w_{ij}(t)} + \alpha\Delta w_{ij}(t-1) \tag{4}$$

where $w_{jk}(t)$ is the weight connecting a hidden node j with an output node k while $w_{ij}(t)$ is the weight connecting an input node i with a hidden node j at time t. $\Delta w_{jk}(t-1)$ is the modification amount to w_{jk} at time $t-1$. $\eta(>0)$ and $\alpha(0 < \alpha < 1)$ are respectively called the learning rate and momentum factor.

3.1.1 Dynamic selection of training samples

An important issue of BP training is that a trained MLP network must have sufficient generalization capability. Traditionally, MLP networks are trained using the whole available training database – a larger training database should result in better training of the concerned MLP. On the other hand, more than sufficient number of presentations of the same set of samples during its training result in overtraining of the concerned MLP network and thus the generalization performance of the MLP gets affected.

It may often happen that a large training database consists samples all of which are not sufficiently different (differing only by a few pixels) from each other. In such situations, the MLP network may get overtrained by those patterns which are almost similar. Such a situation was observed during our study. Similar situations also arise in signal processing or speech processing jobs where it is usual to consider an active learning scheme which is capable of selecting a concise subset from the available large training database and thus reduces both the necessary training time and the chance of overtraining. Such an approach for dynamic selection of training samples is found in [16].

For the backpropagation training of our MLP networks, we used concise training sets selected dynamically. The algorithm for the selection of such a training set is given below.

Algorithm:
Step 1: Initialize current training subset(T_s) by randomly selecting a small fraction of the whole training database(T)
Step 2: Do
  Train the MLP network with the current subset T_s;
  Check its current generalization performance
  using the validation set.
 while{generalization performance gets improved}
Step 3: Present the samples from $(T - T_s)$ to the MLP;
 Select the next fraction of samples from $(T - T_s)$
 for which the current MLP's response are the worst.
Step 4: Obtain modified T_s by adding these samples.
Step 5: Repeat Steps 2-4 till generalization gets improved.
Step 6: Stop.

During our simulations, it was observed that the above approach can efficiently select considerably small number of samples from a large training database providing equivalent or slightly better generalization capability and also reduces the training time. During our simulations a 2.3GHz computer took more than 75 hours CPU time to train a

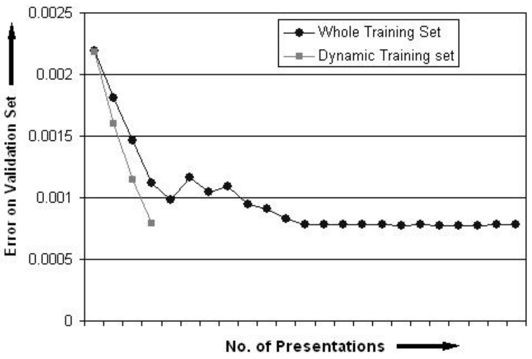

Figure 3. System errors during training using all the training samples and the dynamically selected training samples as a function of the number of presentations of training samples to the MLP

$1024 \times 100 \times 10$ MLP network using the whole training set of the MNIST database. On the other hand, the above dynamic training set selection strategy reduces the training time to less than 5 hours CPU time. Thus the time complexity of training MLP networks using a large training database gets reduced substantially. Related results have been summarized in Table 1. Also, the rates at which learning progress in the two cases are shown in Fig. 3.

Table 1. Details of the advantages of Dynamic Selection of Training Samples

Hidden nodes	Dynamically selected samples	E on validation set	
		Dynamic	Traditional
50	1200	.00238	.00241
100	1800	.00201	.00216
150	1800	.00168	.00172
200	2400	.00112	.00116
250	2400	.00077	.00079

3.1.2 Termination of training

In the literature, there exist various termination criteria for backpropagation learning. In the present approach, we used a validation set for determining the termination of training. Since usually validation sets are not exclusively provided with the available standard databases, we randomly selected 10,000 samples from the training set of MNIST database and 1000 samples from that of Bangla numeral database to form the respective validation sets. In each case, equal representation of all the 10 classes has been taken into consideration. Training were terminated when the error on the respective validation sets reached the minimum for the first three consecutive instances.

3.1.3 Input layer size

Since for the wavelet transform, it is required to have image sizes in powers of 2 only, each such training/test sample image has been normalized to the size 32×32. Daubechies wavelet transform has been used to obtain smooth..smooth components of these normalized images at the resolution levels 16×16 and 8×8. Thus each numeral sample gives rise to 3 images – 32×32, 16×16 and 8×8. Three MLPs have been trained by presenting samples at the three resolution levels. So, the input layer sizes of these three MLPs are 1024 (32×32), 256 (16×16) and 64 (8×8) respectively.

Table 2. Training results of different MLPs (with different hidden layer sizes) using MNIST database; (a) input layer size is 1024 (32×32), (b) input layer size is 256 (16×16) and (c) input layer size is 64 (8×8)

(a)

Hidden nodes	E (System Error) on validation set	Accuracy %	
		Test Set	Training Set
100	0.001682	96.42	97.89
300	0.001443	96.61	97.95
500	0.001046	97.44	98.31
700	0.000946	97.68	98.78
900	0.000781	97.72	98.79
1100	0.000771	97.80	98.82
1300	0.000774	97.77	98.86

(b)

Hidden nodes	E (System Error) on validation set	Accuracy %	
		Test Set	Training Set
50	0.001508	96.02	97.64
100	0.001119	96.76	97.71
150	0.001021	97.22	98.58
200	0.001007	97.48	98.59
250	0.000969	97.60	98.59
300	0.000978	97.50	98.61
350	0.001003	96.49	98.63

(c)

Hidden nodes	E (System Error) on validation set	Recognition Accuracy (%)	
		Test Set	Training Set
14	0.002169	91.51	97.24
24	0.001704	92.60	97.53
34	0.001585	93.49	98.08
44	0.001414	94.70	98.18
54	0.001293	95.27	98.25
64	0.001278	95.60	98.38
74	0.001289	95.29	98.59

3.1.4 Hidden layer size

The training and the recognition performance of an MLP largely depends on the selection of the number of nodes in its hidden layer(s). During our simulations, it was observed that in the present problem, an optimal performance is achieved when only one hidden layer is considered and its size is approximately equal to the input layer size.

Results given in Tables 2 $(a) - (c)$ justify the above conjecture that an optimal recognition performance can be achieved when the hidden layer size is approximately equal to the size of the input layer.

3.2. Combination of multiple MLP classifiers

In the literature, there exists a variety of methods [17, 18] for the combination of results obtained from multiple classifiers. Different classifiers usually represent different aspects of the input data, while none of them can represent all those together. This is true regardless of whether the classifiers are independent or make use of orthogonal features. In a few handwritten character recognition approaches such as [19] multiple classifiers were used for improved recognition accuracies.

In the present article, we considered three MLP classifiers. For combining the recognition results by these three classifiers we simulated four different rules – sum rule, product rule, majority voting and weighted majority voting.

4. Recognition scheme

An input gray-valued image of a character is first binarized. The bounding box (minimum possible rectangle enclosing the character shape) of the binary image is normalized to the size 32×32. For binarization we considered Otsu's global thresholding technique [20] and for normalization we use a linear size normalization method as described in [21]. No other preprocessing like tilt correction, smoothing etc. are considered.

Wavelet decomposition algorithm is applied to this normalized image recursively for two times to obtain 16×16 and 8×8 smooth…smooth approximations of the original image. These approximations of the original image are gray-valued images and the same thresholding technique is applied to obtain respective binary images.

Above three (32×32, 16×16 and 8×8 approximations) binarized images, corresponding to an input numeral, are fed to the input layers of three MLP networks. Responses at the output nodes of the three MLPs are combined using sum, product, majority voting and weighted majority voting rules. In the sum and product rule, if the maximum response is not significantly more than the second maximum response, then the input numeral is considered to be rejected. On the other hand, in the majority and weighted majority voting approaches, if the maximum votes are obtained by more than one class, then rejection occurs.

5. Experimental results

We obtained the recognition performances of each of the four combination rules on the test set of 10000 samples of the MNIST database. The misclassification percentages on this set of handwritten English numerals are 1.39%, 1.40%, 1.30% and 0.76% respectively using sum, product, majority voting and weighted majority voting approaches for the combination purpose. The rejections in these four situations are 0.22%, 0.24%, 0.32% and 0.694% respectively. Thus the best performance corresponds to the weighted majority voting scheme and the above misclassification/rejection figures have been obtained by considering 1.8, 0.6 and 0.6 as the different weights for the outputs of three MLPs corresponding to the three fine-to-coarse resolution levels. During our extensive simulations, we considered quite a few such sets of weights and the above set of weights was found to be the best. Consideration of the maximum weight in favour of the finest resolution level is justified by the fact that it carries the maximum information and also the recognition performance of the concerned MLP is the highest.

Similar observations were made on the Bangla numeral database. In this case also, the weighted majority voting approach provided best recognition performance (1.22% misclassification and 0.74% rejection on a test set of 5000 samples) among the four alternative strategies.

In Table 3 and Table 4, the confusion matrices using the weighted majority voting approach on the test sets of respectively English and Bangla numerals are presented.

Table 3. Confusion matrix on the test set of MNIST database using the weighted majority voting approach

	0	1	2	3	4	5	6	7	8	9
0	**99.082**	0	0	0.102	0	0	0.204	0.102	0.102	0
1	0	**98.59**	0.264	0	0.088	0	0.088	0	0.088	0.088
2	0.097	0.191	**98.768**	0.194	0.09	0	0	0.274	0.115	0
3	0	0	0.099	**98.218**	0	0.198	0	0.099	0.099	0
4	0	0	0	0	**99.371**	0	0.324	0	0.102	0.1
5	0.204	0	0	0.134	0	**98.587**	0.2	0.112	0.1	0.102
6	0.191	0.208	0.03	0.101	0	0.104	**98.843**	0	0	0
7	0	0	0.179	0	0.215	0	0	**98.543**	0.185	0.198
8	0.21	0.05	0.105	0.08	0.1	0.104	0	0.2	**98.228**	0.101
9	0	0.029	0	0.099	0.196	0.274	0	0.491	0.194	**97.23**

Table 4. Confusion matrix on the test set of ISI database using the weighted majority voting approach

	০	১	২	৩	৪	৫	৬	৭	৮	৯
০	99.052	0.030	0.000	0.102	0.000	0.000	0.204	0.102	0.102	0.000
১	0.000	98.890	0.164	0.100	0.138	0.100	0.088	0.000	0.088	0.088
২	0.087	0.291	97.868	0.204	0.194	0.000	0.000	0.291	0.194	0.000
৩	0.000	0.000	0.099	98.018	0.120	0.198	0.000	0.099	0.099	0.000
৪	0.000	0.000	0.000	0.000	98.571	0.000	0.400	0.089	0.102	0.205
৫	0.200	0.000	0.000	0.897	0.000	97.508	0.224	0.214	0.192	0.202
৬	0.209	0.209	0.104	0.104	0.000	0.104	98.643	0.000	0.104	0.000
৭	0.000	0.000	0.681	0.000	0.195	0.000	0.000	98.054	0.195	0.195
৮	0.411	0.220	0.205	0.323	0.105	0.205	0.002	0.205	97.222	0.410
৯	0.000	0.099	1.330	0.099	0.496	0.297	0.000	0.496	0.396	96.630

0-7695-2187-8/04 $20.00 © 2004 IEEE

6. Conclusion

The present multiresolution recognition approach for handwritten numerals is independent of the script. The weighted majority voting approach improves the previously reported recognition accuracy on Bangla numeral database. Its performance on English MNIST database is also comparable to the existing state-of-the-art techniques. Moreover, this is fast enough for its implementations in real-life applications and it can recognize more than sixty numerals per second on a Pentium-IV Desktop Computer.

Finally, such a multiresolution pixel image based approach can perform satisfactorily even in the presence of moderate noise or discontinuity or small changes in orientation.

References

[1] C. Y.Suen, M. Berthod and S. Mori, Automatic recognition of handprinted characters – the state the art. *Proceedings of the IEEE*, **68(4)** (1980), pp. 469-487.

[2] N. Arica and F. Yarman-Vural, "An Overview of Character Recognition Focused on Off-line Handwriting", *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 31(2), 2001, pp. 216 - 233.

[3] O. D. Trier, A. K. Jain and T. Taxt, "Feature Extraction Methods for Character Recognition - A Survey", *Pattern Recognition*, vol. 29, 1996, pp. 641 - 662.

[4] M. D. Garris, C. L. Wilson and J. L. Blue, Neural Network-Based Systems for Handprint OCR Applications. *IEEE Transactions on Image Processing,* **7** (1999) pp. 1097-1112.

[5] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, Backpropagation applied to handwritten zip code recognition. *Neural Computation*, (1) 4 (1989) pp. 541-551.

[6] S. Knerr, L. Personnaz, G. Dreyfus, Handwritten Digit Recognition by Neural Networks with single-Layer Training. *IEEE Transactions on Neural Networks*,**3** (1992) pp. 303-314.

[7] U. Bhattacharya, T. K. Das, A. Datta, S. K. Parui and B. B. Chaudhuri, A Hybrid Scheme for Handprinted Numeral Recognition Based On a Self-Organizing Network and MLP Classifiers, International Journal for Pattern Recognition and Artificial Intelligence, 16(7), 2002, pp. 845-864.

[8] S. Shimizu, W, Ohyama, T. Wakabayashi and F. Kimura, Mirror Image Learning for Autoassociative Neural Networks. *Proc. of ICDAR 2003,* August 03 - 06, 2003, Edinburgh, Scotland, **3** pp. 804-808.

[9] U. Bhattacharya, T. K. Das and B. B. Chaudhuri, A cascaded scheme for recognition of handprinted numerals, Proceedings of the third Indian Conference on Computer Vision, Graphics and Image Processing, Ahmedabad, India, 2002, pp. 137 - 142.

[10] U. Bhattacharya and B. B. Chaudhuri, "A majority voting scheme for multiresolution recognition of handprinted numerals", Proc. of the 7th Int. Conf. On Document Analysis and Recognition (ICDAR), Edinburgh, Scotland, vol. 1, 2003, pp. 16-20.

[11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", Proceedings of the IEEE, vol. 86(11), 1998, pp.2278-2324.

[12] U. Bhattacharya, B.B. Chaudhuri, "Databases for research on recognition of handwritten numerals of Indian scripts", In Proc. Of National Workshop on Computer Vision Graphics and Image Processing (WCVGIP-2004), Gwalior (M.P.), India, pp. 18-21, 21-22 Feb., 2004.

[13] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition : The Wavelet Representation", *IEEE Trans. on Pattern Anal. and Machine Int.*, vol. 11(7), 1989, pp 674 -693.

[14] I. Daubechies, "The Wavelet Transform, Time-frequency Localization and Signal Analysis", *IEEE Trans. on Information Theory*, vol. 36(5), 1990, pp. 961-1005.

[15] D. E. Rumelhart, G. E. Hinton and R. J. Williams, Learning internal representations by error propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*. Eds. D. E. Rumelhart and J. L. McClelland, MA: The MIT Press: Cambridge (1986), pp. 318-362.

[16] A. Robel, "The dynamic pattern selection algorithm: Effective training and controlled generalization of backpropagation neural networks", Technical Report, Technische Universitat Berlin, Berlin, 1993.

[17] C. Y. Suen, C. Nadal, T. A. Mai, R. Regault and L. Lam, "Computer Recognition of Unconstrained Handwritten Numerals", *Proc. IEEE*, vol. 80, 1992, pp. 1162-1180.

[18] T. K. Ho, J. J. Hull and S. N. Srihari, "Decision Combination in Multiple Classifier Systems", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 16, 1994, pp. 66-75.

[19] F. Kimura and M. Sridhar, "Handwritten Numeral Recognition Based on Multiple Algorithms", *Pattern Recognition*, vol. 24, 1991.

[20] N. Otsu, A Threshold Selection Method from Grey-Level Histograms, IEEE Trans. Systems, Man, and Cybernetics, vol. 9, 1979, pp. 377-393.

[21] C-L. Liu, K. Nakashima, H. Sako and H. Fuzisawa, "Handwritten digit recognition: investigation of normalization and feature extraction techniques", Pattern Recognition, 37, 2004, 265-279.

0-7695-2187-8/04 $20.00 © 2004 **IEEE**