A Majority Voting Scheme for Multiresolution Recognition of Handprinted Numerals

U. Bhattacharya and B. B. Chaudhuri
CVPR Unit, Indian Statistical Institute, Kolkata-108, India
{ujjwal,bbc}@isical.ac.in

Abstract

This paper proposes a simple voting scheme for off-line recognition of handprinted numerals. One of the main features of the proposed scheme is that this is not script dependent. Another interesting feature is that it is sufficiently fast for real-life applications. In contrast to the usual practices, here we studied the efficiency of a majority voting approach when all the classifiers involved are multilayer perceptrons (MLP) of different sizes and respective features are based on wavelet transforms at different resolution levels. The rationale for this approach is to explore how one can improve the recognition performance without adding much to the requirements for computational time and resources. For simplicity and efficiency, in the present work, we considered only three coarse-to-fine resolution levels of wavelet representation. We primarily simulated the proposed technique on a database of off-line handprinted Bangla (a major Indian script) numerals. We achieved 97.16% correct recognition rate on a test set of 5000 Bangla numerals. In this simulation we used two other disjoint sets (one for training and the other for validation purpose) of sizes 6000 and 1000 respectively. We have also tested our approach on MNIST database for handwritten English digits. The result is comparable with state-of-the-art technologies.

1. Introduction

The off-line recognition of hand printed characters, in particular, numerals has been a topic of intensive research during last few years due to its enormous application potential. The application areas include postal code reading, automatic processing of bank cheques, office automation and various other scientific and business applications. On the other hand, the problem is very interesting by nature because each individual writer produces a unique set of characters each time they write.

Many diverse algorithms/schemes for hand printed character recognition [1, 2] exist and each of these have their own merits and demerits. Moreover, they are usually script dependent. A few works include [3] in English, [4] in Chinese, [5] in Korean, [6] in Arabic and [7] in Kanji and so on. Some preliminary works [8, 9, 10] has been done on Bangla, the second-most popular language and script in the Indian subcontinent and the fifth-most popular language in the world.

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 [11]. 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 the present work, we considered multiresolution features based on a wavelet transform and a voting scheme on the responses of a set of multi-layer perceptron (MLP) networks for the classification purpose. The present technique has been tested on two different scripts – Bangla and English and obtained both speed and recognition accuracies comparable to the state-of-the-art techniques. The reason for obtaining high speed recognition rate in the proposed approach is that wavelet, as a feature-extraction tool, fits naturally with digital computer with its basis functions defined by just multiplication and addition operators – there are no derivatives or integrals.

Wavelet based approaches have been proposed previously for recognition of handwritten characters [12, 13] but not in a manner used by us. Since the wavelet transform provides an invariant interpretation of a character image at different resolution levels characterizing different physical structures of the character, this feature extraction scheme intuitively seems to be very effective for handprinted character recognition tasks.

It has been already established that combining the decisions of several classifiers usually result in better classification accuracy. This is perhaps because different classifiers represent different aspects of the input data, while none of them can represent all those together. In the domain of character recognition, significant improvement in recognition performance has been reported in a number of

Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)
0-7695-1960-1/03 $17.00 © 2003 **IEEE**

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

Figure 1. Smooth...smooth components of wavelet decompositions of a numeral image at different resolution levels (a) Original image (b) 32×32 resolution (c) 16×16 resolution (d) 8×8 resolution

occasions [14, 15] by considering the combination strategy and this is true regardless of whether the classifiers are independent or make use of orthogonal features. There exists a variety of methods [16, 17] implementing the combination of classifiers strategy. The simplest such method is the majority voting strategy [18].

In this paper, a set of three MLP classifiers have been considered where features in the form of wavelet coefficient matrices at three coarse-to-fine resolution levels are used. These wavelet coefficient matrices is obtained by considering the Daubechies-4 wavelet transform which is the simplest in the family of wavelet filters discovered by Daubechies [19]. On the other hand, Daubechies wavelet transform is capable of picking up details which may be missed by Haar wavelet. We considered the pyramidal algorithm [20] for obtaining wavelet transforms of an input image hierarchically and the low...low or smooth...smooth components (Figure 1) at three consecutive levels of the hierarchy have been considered as feature sets for the three MLP classifiers.

Three MLP network architectures are trained using training sets at three coarse-to-fine resolution levels. Also a validation sample set is used to determine the termination of training. Majority voting scheme is applied on the set of three responses for an input numeral by the three MLP networks and the final classification is obtained. If each of the three MLPs produce three distinct responses for an input numeral, then it is said to be rejected by the proposed scheme.

The rest of this article is organized as follows. Section 2 provide brief overviews of wavelet transform. We describe the preprocessing, training of the set of MLP networks and the proposed recognition scheme in Section 3. Experimental results are reported in Section 4. Concluding remarks are given in Section 5.

2. Wavelet descriptor for multiresolution feature extraction

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 [20]) 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 [19] 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.)

In the above, principle has been described for obtaining wavelet transform of a single dimensional array. However, simple extension of this 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.

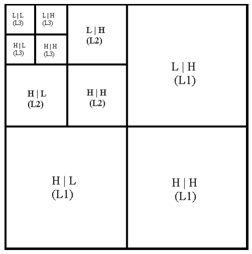

Figure 2. Layout of wavelet decomposition for an image (L $\rightarrow$ low-pass filter, H $\rightarrow$ high-pass filter, L$j \rightarrow j$th level)

3. Recognition scheme

3.1. Preprocessing

As preprocessing, we considered only size normalization. The input grey scale image is first scaled to 64×64 image by using the moment method [21]. No further preprocessing like tilt correction, smoothing etc. are considered.

3.2. Training of MLPs

Two important aspects of the training of MLP networks are

- Designing the training sets and
- Termination of training

Designing the training set. The recognition performance of an MLP network highly depends on the choice of a representative training set. Manual selection of training samples is definitely a good approach to this problem. However, since this approach is extremely tedious, we have chosen the training set randomly from the available data on Bangla numerals. In fact, we performed random selections with respect to three different seed values and experimental results will be provided corresponding to the best of these three. In each of these cases, the size of the training set is 50% of the available data size. On the other hand, the MNIST database of handwritten English digits, available from http://yann.lecun.com/exdb/mnist/ has a training set of 60,000 examples, and a test set of 10,000 examples. In our simulations, we randomly selected 50,000 samples from the training set of MNIST database to form training set of our experiments and the rest 10,000 samples have been used to form the validation set.

Termination of training. There are various termination criteria available in the literature. The recognition performance highly depends on how much training has been given to the network. An effective strategy of judging training adequacy is the use of a validation set. With increased training, the recognition error on the validation set decreases monotonically to a minimum value but then it starts to increase, even if the training error continues to decrease. For better network performance, training is terminated when

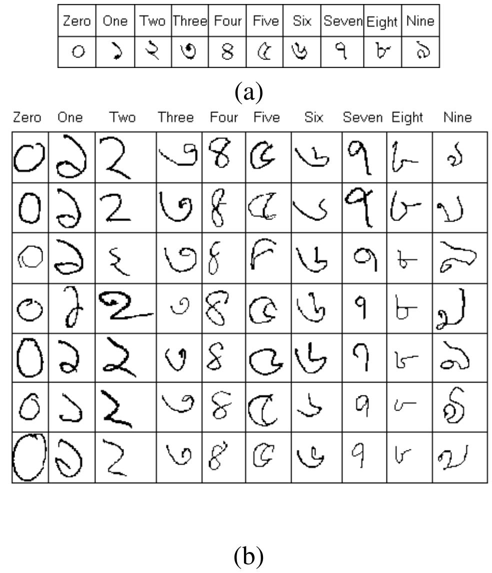

Figure 3. (a) **Ideal samples of Bangla numerals** (b) **A typical sample data subset of handwritten Bangla numerals**

the validation error reaches its minimum. In our simulations , we considered 1000 samples for Bangla numerals and 10,000 samples for English numerals as the two validation sets. In each case, equal representation of all the 10 classes has been taken into consideration.

3.3. Recognition scheme

The proposed recognition scheme has been simulated on the domains of handprinted Bangla numerals and MNIST database of handwritten English numerals. Ideal Bangla numerals and 70 handwritten samples (7 different numeral characters per class) are shown in Figure 3. A sample from the MNIST database may be found in [22].

The bounding box (minimum possible rectangle enclosing the image) of an input image of a numeral is first computed and then this is normalized to the size 64×64 using the moment method [21]. Wavelet decomposition algorithm is applied to this normalized image recursively for three times to obtain 8×8 smooth. . .smooth approximation of the original image as the final decomposition. In this procedure, we also obtain 32×32 and 16×16 smooth. . .smooth approximations of the original image at intermediate stages. Theoretically, this decomposition algorithm could be applied up to the fifth time to obtain 4×4 and 2×2 approximations. However, during our simulations, it is observed that different numerals are not generally distinguishable from these smaller possible approximations.

The above approximations of the original image are gray-valued images and we apply simple thresholding technique to obtain these as binary images. The present recognition scheme is a majority voting scheme. The binarized versions of 32×32 16×16 and 8×8 approximations of the original image are fed to the input layers of three MLP networks. Different responses at the nodes of the output layers of each of the three MLPs are compared. The output node with maximum value recognizes the input image. Thus, there may occur a case, when an input numeral may be classified into three different categories and in such a situation, the input numeral is said to be rejected by the proposed scheme. On the other hand, if at least two MLP networks classify the input numeral into the same category, then the majority voting scheme recognizes it to belong to this class. During our simulation runs, it has been observed that by extending the proposed classification scheme into further stages cannot improve the classification accuracy.

4. Experimental results

The authors are not aware of the availability of any standard database of handprinted Bangla numerals. So, such a database had been developed with the help of a group of University students. Our database consists of 12,000 isolated handprinted Bangla numerals equally distributed over all classes. These data had been collected from different sections of the population of West Bengal, India and this includes possible variations with respect to age, sex, education, place of origin, income group and profession. Since there appears variation in the writing style of a single individual at different points of time, each individual has been approached on 4 occasions for the sample. However, for simulation results on English we considered MNIST database consisting of a total of 70000 isolated handprinted English numerals. In the following, simulation results on both the scripts are provided – figures within brackets correspond to English numerals.

The whole set of 12,000 (70,000) samples of Bangla (English) numerals, is composed of three subsets called training set, validation set and test set consisting of respectively 6000 (50,000), 1000 (10,000) and 5000 (10,000) samples. In the 32×32, 16×16 and 8×8 resolution levels, the correct classification percentages are respectively 95.98% (96.5%), 96.10% (96.55%) and 93.78% (94.33%) on respective test datasets. After application of majority voting scheme on these three classification results, final true classification is 97.16% (97.57%) and rejection is 0.76% (.45%). Thus we achieved only 2.08% (1.98%) misclassification which is comparable to the existing state-of-the art techniques. Table 1 shows the break-up of the final true classification percentages. The final confusion matrices corresponding to Bangla and English databases are given in Table 2 and Table 3 respectively.

5. Conclusion

Wavelets have been studied thoroughly during the last decade. Its potential in image compression tasks has been already established and its applicability in various other image processing problems are getting explored. In this paper we considered a majority voting approach on wavelet feature based multiresolution recognition of handprinted numerals. In this approach we considered three MLP classifiers corresponding to the three different resolution levels. This strategy of using multiple classifiers of similar type helps to improve recognition accuracy without significant increase in computation of features. However, proposed approach does not restrict further use of multiple classifiers having less (if not zero) correlation. In fact, in the latter situation, the proposed ensemble of MLPs may be considered as a single classifier.

Table 1. Break-up of the result of majority voting scheme on the two datasets (Here T denotes true classification while E denotes misclassification).

Script	32 × 32	16 × 16	8 × 8	Percentage
Bangla	T	T	T	93.84
	E	T	T	0.49
	T	E	T	0.51
	T	T	E	2.32
Total				97.16
English	T	T	T	94.06
	E	T	T	0.51
	T	E	T	0.52
	T	T	E	2.48
Total				97.57

The proposed approach is independent of the script. It improves the recognition accuracy on Bangla and also observed to produce recognition result on English MNIST database comparable to the state-of-the-art techniques. Also, this is fast enough for its implementation in real-life applications and this recognizes more than sixty numerals per second on a Pentium-IV Desktop Computer. Finally, the wavelet based features are also not affected in the presence of moderate noise or discontinuity or small changes in orientation. In Figure 4, we have shown the numeral of Figure $1(a)$ affected by noise and its wavelet-based feature images at different resolution levels.

Table 2. Final Confusion Matrix on Bangla dataset

	০	১	২	৩	৪	৫	৬	৭	৮	৯
০	**97.48**	0.28	0.28	0.45	0.23	0.17	0.15	0.09	0.16	0.67
১	0.20	**97.23**	0.47	0.26	0.27	0.27	0.18	0.39	0.12	0.58
২	1.10	0.25	**97.86**	0.09	0.04	0.14	0.05	0.13	0.03	0.26
৩	1.10	0.34	0.12	**97.47**	0.05	0.24	0.10	0.22	0.04	0.26
৪	0.32	0.55	0.17	0.32	**96.52**	0.04	0.66	0.45	0.75	0.09
৫	0.12	0.15	0.17	0.66	0.21	**96.93**	0.81	0.16	0.58	0.13
৬	0.08	0.11	0.05	1.23	0.10	0.29	**97.89**	0.04	0.05	0.03
৭	0.42	0.20	0.34	0.01	1.43	0.35	0.04	**96.85**	0.23	0.05
৮	0.39	1.39	0.69	0.35	1.03	0.55	0.49	0.32	**94.50**	0.22
৯	0.03	0.08	0.07	0.28	0.16	0.04	0.25	0.07	0.06	**98.87**

Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)
0-7695-1960-1/03 $17.00 © 2003 **IEEE**

Table 3. Final Confusion Matrix on English database

	0	1	2	3	4	5	6	7	8	9
0	**96.21**	0.00	0.23	0.15	0.25	0.27	0.25	0.19	0.96	1.00
1	0.20	**97.39**	0.47	0.26	0.27	0.00	0.18	0.00	0.12	0.58
2	1.10	0.25	**97.86**	0.09	0.04	0.14	0.05	0.13	0.03	0.26
3	1.10	0.34	0.12	**97.47**	0.05	0.00	0.10	0.00	0.04	0.26
4	0.32	0.00	0.00	0.32	**98.35**	0.00	0.66	0.00	0.45	0.00
5	0.12	0.15	0.37	0.00	0.21	**96.93**	0.81	0.16	0.58	0.00
6	0.04	0.11	0.05	1.23	0.10	0.29	**97.89**	0.04	0.05	0.03
7	0.00	0.11	0.14	0.00	0.13	0.00	0.00	**99.77**	0.00	0.05
8	1.39	1.39	0.55	0.00	0.00	0.55	0.49	0.32	**94.96**	0.22
9	0.03	0.08	0.07	0.28	0.16	0.04	0.25	0.07	0.06	**98.87**

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

Figure 4. Smooth ... smooth components of wavelet decomposition of a noisy image of a Bangla numeral at different resolution levels (a) **Original noisy image** (b) 32×32 **resolution** (c) 16×16 **resolution** (d) 8×8 **resolution**

References

[1] R. Plamondon, S. N. Srihari, "On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey", *IEEE Trans. Patt. Anal. and Mach. Intell.*, vol. 22(1), 2000, pp. 63-84.

[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] S. N. Srihari, E. Cohen, J. J. Hull and L. Kuan, "A System to Locate and Recognize ZIP Codes in Handwritten Addresses", *IJRE*, vol. 1, 1989, pp. 37-45.

[4] J. Tsukumo and H. Tanaka, "Classification of Handprinted Chinese Characters Using Nonlinear Normalization Methods", *Proc. 9th. Int. Conf. on Patt. Recog.*, 1988, pp. 168-171.

[5] S. W. Lee and J. S. Park, "Nonlinear Shape Normalization Methods for the Recognition of Large-set Handwritten Characters", *Pattern Recognition*, vol. 27, 1994, pp. 895-902.

[6] A. Amin and H. B. Al-Sadoun, "Hand Printed Arabic Character Recognition System", *12th. Int. Conf. on Pattern Recognition*, 1994, pp. 536-539.

[7] H. Yamada, K. Yamamoto and T. Saito, "A Non-linear Normalization Method for Handprinted Kanji Character Recognition – Line Density Equalization", *Pattern Recognition*, vol. 23, 1990, pp. 1023-1029.

[8] U. Pal and B. B. Chaudhuri, "Automatic Recognition of Unconstrained Off-line Bangla Hand-written Numerals", *Advances in Multimodal Interfaces, Springer Verlag Lecture Notes on Computer Science (LNCS-1948)*, Eds. T. Tan, Y. Shi and W. Gao., 2000, pp. 371-378.

[9] A. F. R. Rahman, R. Rahman and M. C. Fairhurst, Recognition of Handwritten Bengali Characters: A Novel Multistage Approach, Pattern Recognition, vol. 35, 2002, pp. 997-1006.

[10] 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.

[11] 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.

[12] P. Wunsch and A. F. Laine, "Wavelet Descriptors for Multiresolution Recognition of Handprinted Characters", *Pattern Recognition*, vol. 28(8), 1995, pp. 1237-1249.

[13] S. W. Lee, C.H. Kim, H. Ma and Y. Y. Tang, "Multiresolution Recognition of Unconstrained Handwritten Numerals with Wavelet Transform and Multilayer Cluster Network", *Pattern Recognition*, vol. 29(12), 1996, pp. 1953 - 1961.

[14] L. Xu, A. Krzyzak and C. Y. suen, "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition", *IEEE Trans. on Syst. Man, Cybern.*, vol. 22, 1992, pp. 418-435.

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

[16] 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.

[17] 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.

[18] L. Lam and C. Y. Suen, "Application of Majority Voting to Pattern Recognition: An Analysis of its Behaviour and Performance", *IEEE Trans. on Syst. Man and Cyebrn. - Part A: Systems and Humans*, vol. 27, 1997, pp. 553 - 568.

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

[20] 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.

[21] R. G. Casey, "Moment Normalization of Handprinted Characters", *IBM J. Res. Develop.*, vol. 14, 1970, pp. 548-557.

[22] 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.

Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)
0-7695-1960-1/03 $17.00 © 2003 **IEEE**