References

[1] Bruce A. Draper. From knowledge bases to Markov models to PCA. In *Proceedings of Workshop on Computer Vision System Control Architectures*, Graz, Austria, 2003.

[2] B. Draper, A. Hanson, and E. Riseman. Knowledge-directed vision: Control, learning and integration. *Proceedings of the IEEE*, 84(11):1625–1637, 1996.

[3] R. Rimey and C. Brown. Control of selective perception using bayes nets and decision theory. *International Journal of Computer Vision*, 12:173–207, 1994.

[4] B. Draper, J. Bins, and K. Baek. ADORE: adaptive object recognition. *Videre*, 1(4):86–99, 2000.

[5] B. Draper, U. Ahlrichs, and D. Paulus. Adapting object recognition across domains: A demonstration. In *Proceedings of International Conference on Vision Systems*, pages 256–267, Vancouver, B.C., 2001.

[6] R.S. Sutton and A.G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2000.

[7] A. G. Barto, S. J. Bradtke, and S. P. Singh. Learning to act using real-time dynamic programming. *Artificial Intelligence*, 72(1):81–138, 1995.

[8] I. Levner. Multi resolution adaptive object recognition system: A step towards autonomous vision systems. Master's thesis, Department of Computer Science, University of Alberta, 2003.

[9] Vadim Bulitko and Ilya Levner. Improving learnability of adaptive image interpretation systems. Technical report, University of Alberta, 2003.

[10] I. Levner, V. Bulitko, G. Lee, L. Li, and R. Greiner. Towards automated creation of image interpretation systems. In *Australian Joint Conference on Artificial Intelligence (To appear)*, 2003.

[11] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Macmillian College Pub. Co., 1994.

[12] S. Agarwal and D. Roth. Learning a sparse representation for object detection. In *7th European Conference on Computer Vision*, volume 4, pages 113–130, Copenhagen, Denmark, 2002.

[13] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York, 1973.

[14] Michael Kirby. *Geometric Data Analysis: An Emprical Approach to Dimensionality Reduction and the Study of Patterns*. John Wiley & Sons, New York, 2001.

[15] P.M.Hall, R.R.Martin, and A.D. Marshall. Merging and splitting eigenspace models. *IEEE Trans. on Pattern Analysis & Machine Intelligence*, 22(9):1042–1050, 2000.

[16] Sunil Arya, David M. Mount, Nathan S. Netanyahu, Ruth Silverman, and Angela Y. Wu. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM*, 45(6):891–923, 1998.

[17] Dan Pelleg and Andrew Moore. Accelerating exact k-means algorithms with geometric reasoning. In Surajit Chaudhuri and David Madigan, editors, *Proceedings of the Fifth International Conference on Knowledge Discovery in Databases*, pages 277–281. AAAI Press, aug 1999.

[18] R. Kohavi, P. Langley, and Y. Yun. The utility of feature weighting in nearest-neighbor algorithms, 1997.

[19] Vadim Bulitko, Greg Lee, and Ilya Levner. Evolutionary algorithms for operator selection in vision. In *Proceedings of the Fifth International Workshop on Frontiers in Evolutionary Algorithms*, 2003.

[20] K. Kira and L. Rendell. The feature selection problem: Traditional methods and a new algorithm. *In Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI-92)*, pages 129–134, 1992.

[21] J. Bins and B. Draper. Feature selection from huge feature sets. In *Proceedings of International Conference on Computer Vision*, volume 2, pages 159–165, 2001.

[22] Ron Kohavi and George H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324, 1997.

Optimal Wavelets and Neural Networks for Pattern Recognition

G. Y. Chen, T. D. Bui and A. Krzyzak
Department of Computer Science, Concordia University
1455 De Maisonneuve West, Montreal, Quebec, Canada H3G 1M8.
$\{guang\_c, bui, krzyzak\}$@cs.concordia.ca

Abstract

We investigate the applications of optimal wavelets and neural networks in the recognition of handwritten numerals. Wavelet transforms have been successfully applied in many applications including pattern recognition. However, which kind of wavelet should be used is still an open problem. We propose to use a combination of optimal wavelets and neural networks in pattern recognition applications. The optimal wavelet filters can be obtained by minimizing the mean square error of the neural network output. The same error objective function is used for training the weights of the neural network. We conducted some experiments on unconstrained handwritten numeral recognition and observed 1.60% increase in the recognition rate compared to the Daubechies-4 wavelet on the Concordia handwritten numeral database.

Keywords: Optimal wavelets, neural networks, handwriting recognition.

1 Introduction

Wavelet transforms have been proven to be very successful in many applications such as image compression, image denoising, signal processing, pattern recognition, and computer graphics, to name only a few. For a given application, which wavelet should we choose to use? It is desirable to choose a wavelet that is best suitable for the problem in hand. We call this wavelet *optimal wavelet*. Optimal wavelets have been used in several areas such as image compression, signal representation, signal denoising, feature extraction, face recognition, image pattern recognition, etc. However, optimal wavelets have not been applied to pattern recognition in combination with neural networks. Here we briefly review some of the optimal wavelet papers. Mallet et al. [1] proposed a new and innovative technique based on adaptive wavelets, which aims to reduce the dimensionality and optimize the discriminatory information. Instead of using standard wavelet bases, they generate the wavelet which optimizes specified discriminant criteria. Wang et al. [2] used Genetic Algorithm (GA) to find an optimal basis derived from a combination of frequencies and orientation angles in the 2-D Gabor wavelet transform. This approach provides a more accurate and efficient projection scheme and therefore a better face classification result. Chapa et al. [3] developed a technique for deriving a bandlimited wavelet directly from the desired signal spectrum in such a way that the mean square error between their spectra is a minimum. The technique includes an algorithm for finding the scaling function from an orthonormal wavelet, and algorithms for finding the optimal wavelet magnitude and phase from a given input signal. A revised version of this paper appeared in *IEEE Transactions on Signal Processing* [4]. Zhuang et al. [5] studied the problem of choosing an image based optimal wavelet basis with compact support for image data compression and provided a general algorithm for computing the optimal wavelet basis. Tewfik et al. [6] studied the problem of choosing a discrete orthogonal wavelet with a support size that is equal to or smaller than a prespecified limit to represent a given finite support signal up to a fixed resolution scale. The techniques are based on optimizing certain cost functions. The optimization with a set of constraints can be converted to an optimization problem without constraints by means of parametrization. Das et al. [7] proposed an algorithm for construction of optimal compactly supported N-tap orthonormal wavelet for signal denoising. Simulated annealing is used for the optimization of the parametrization of the wavelet FIR filter bank coefficients. Golden [8] considered the problem of optimizing at each resolution level the parameters of a two-band quadrature mirror filter analysis bank to achieve maximal decorrelation of the two decimated output sequences.

In this paper, we propose to use optimal wavelets together with neural networks in pattern recognition applications. Since we have to solve an optimization problem in order to find the optimal wavelets, it is suggested to use the optimal wavelets in a pattern recognition descriptor that employs both wavelet features and neural networks. We can obtain the optimal wavelet filters by minimizing the mean square error of the neural network output. The same objective function can be used to train the neural network weights. Experiments show a recognition rate increase for unconstrained handwritten numeral recognition. In order to extract better features, we use different filters for different wavelet decomposition levels.

The paper is organized as follows. Section 2 reviews the parametrization of compactly supported wavelets. Section 3 explains how optimal wavelets are applied to pattern recognition. Section 4 gives some experimental results. Finally, section 5 concludes the paper.

2 Parametrization of Compactly Supported Wavelets

The construction of a wavelet depends on a scaling function $\phi(x)$ that obeys a 2-scale dilation equation:

$$\phi(x) = \sqrt{2}\sum c_k \phi(2x-k)$$

where c_k is a set of real filter. The sequence $\{c_k\}$ must satisfy the following conditions [9], [10]:

$$\sum c_k = \sqrt{2}$$

$$\sum c_k c_{k+2m} = \delta(m)$$

$$\sum(-1)^k k^m c_k = 0, m = 0,1,\cdots(M-1$$

where $M \geq 1$ and $\delta(m)$ denotes a discrete Kronecker delta function. M controls the compact support of the wavelet, and it is equal to the number of vanishing moments of the wavelet $\psi(x)$ corresponding to $\phi(x)$.

The wavelet $\psi(x)$ is defined as

$$\psi(x) = \sqrt{2}\sum d_k \phi(2x-k)$$

where $d_k = ((-1)^k c_{M-1-k}$.

It can be shown that discrete orthonormal wavelets of support less than $2N-1$, where N is an even number, can be parametrized by $N-1$ real parameters each taking values in $[0,2\pi)$ [11]. Let

$$H(z) = \sum c_k z^k$$

$$G(z) = \sum d_k z^k$$

where $z = e^{-jw}$. Then we can have

$$\begin{pmatrix} H(z) \\ z^{2(N-1)}G(z) \end{pmatrix} = E(z^2)\begin{pmatrix} 1 \\ z \end{pmatrix}$$

where $E(z)$ is a polyphase matrix defined by

$$E(z) = V_{N-1}(z)V_{N-2}(z)\cdot\cdot(V_1(z)V_0$$

and

$$V_0 = \begin{pmatrix} cos\theta_0 & -sin\theta_0 \\ sin\theta_0 & cos\theta_0 \end{pmatrix}$$

$$V_k(z) = I + (z-1)v_k v_k^T, 1 \leq k \leq N-1.$$

$$v_k = \begin{pmatrix} cos\theta_k \\ sin\theta_k \end{pmatrix}$$

By looking at the definition of $E(z)$ we know that a sequence c_k of length $2N$ is actually parametrized by N free parameters. Furthermore, the wavelet $\psi(x)$ has at least one vanishing moment if and only if $\theta_0 = 3\pi/4$. We can give some of the parametrized sequences for $N = 2$. The sequences $\{c_k\}_{k=0}^3$ can be listed as follows:

$$c_0 = sin(\theta_1) * sin(\theta_1 - \pi/4)$$

$$c_1 = sin(\theta_1) * sin(\theta_1 + \pi/4)$$

$$c_2 = cos(\theta_1) * sin(\theta_1 + \pi/4)$$

$$c_3 = cos(\theta_1) * sin(\pi/4 - \theta_1)$$

where $\theta_1 \in [0,2\pi)$.

After the parametrization, the optimization problem to find the optimal wavelet for a specific cost function becomes an optimization problem without constraints. This is much easier to solve in practice.

3 Applications to Pattern Recognition

Wavelets and neural networks are very popular and effective in pattern recognition. Here we list three descriptors that use both wavelet features and neural networks:

1. Lee et al. [12] proposed a scheme for multiresolution recognition of unconstrained handwritten numerals using wavelet transform and a simple multilayer cluster neural network. The wavelet features of handwritten numeral at two decomposition levels are fed into the multilayer cluster neural network.

2. Wunsch et al. [13] gave a descriptor by extracting wavelet features from the outer contour of the handwritten characters and feeding the features into neural networks. Their experiments were done on handprinted characters.

3. Chen et al. [14] developed a descriptor by using multiwavelets and neural networks. The multiwavelet features are extracted from the outer contour of the handwritten numerals and fed into neural networks. This descriptor gives higher recognition rate than the one given in [13] for handwritten numeral recognition.

In order to obtain an optimal wavelet for a pattern recognition problem, we need to solve an optimization problem. The objective function should be defined by a cost function that need to be minimized. For neural networks, we can use the distance between the network output and the desired output as the objective function in order to train the weights. For optimal wavelets, we can use the same objective function to find the optimal filters. The training of the neural network weights and optimal wavelet filters can be done interchangeably. That is, we modify the optimal wavelet filter parameters once, then we modify the weights of the neural networks once. We call this as one cycle of training. We repeat the cycle in this way for many times until a prespecified number of cycles is reached or the error of the network is below a threshold. When we update the neural network weights, we treat the wavelet filters as fixed and use back-propagation to modify the weights. When we update the optimal wavelet filter parameters, we fix the neural network weights and modify the optimal wavelet filter parameters by using the *Newton-Raphson Method*:

$$\theta(n+1) = \theta(n) - \frac{E(\theta(n))}{E'(\theta(n))}$$

where $E(\theta)$ is the objective function that is defined as the mean square error of the neural network output, and $E'(\theta)$ is the derivative of $E(\theta)$ with respect to the optimal wavelet parameter θ. We accept the modifications on θ only when E decreases. If E increases because of the modification, we restore the previous value of $\theta(n)$. Similarly, we only accept the modifications to the weights of the neural network when E decreases. Otherwise, we restore the weights and reduce the learning rate of the neural network. For every wavelet decomposition level, we have two optimal wavelet parameters, one for the row and one for the column. It should be mentioned that we do not use Simulated Annealing to find the global minimum for the optimal wavelet filters because there is no meaning to spend too much time on learning the optimal wavelet filters when the neural network weights are not fixed. Figure 1 shows the steps of our optimal wavelet neural network descriptor.

The optimal wavelet filters to be found are different for 1D signals and 2D images. When we use the contour of the numeral, we need to find only one set of optimal wavelet filters for every wavelet decomposition level

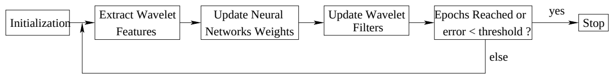

Figure 1: The flow chart of our optimal wavelet neural network descriptor.

because we have 1D input signal. However, for images we need to find two sets of optimal wavelet filters for every wavelet decomposition level: one for the row and one for the column. We carry out experiments on optimal wavelets with filter length equal to four.

4 Experimental Results

In this paper, we only test the feasibility of optimal wavelets in handwritten numeral recognition. In fact, optimal wavelets can be used in any other pattern recognition applications as long as wavelet features and neural networks are used. We use a simple multilayer feed-forward neural network in our experiments. The original numeral image is first normalized to a 16×16 matrix. The centroid of the handwritten numeral is moved to the center of the image. We also scale the handwritten numeral so that the maximum distance from the outer contour to the centroid is half the length of the box containing the numeral image. We perform wavelet transform on the image for two successive levels. We obtain four 8×8 images after the first level wavelet decomposition, and four 4×4 images after the second level wavelet decomposition. These two levels of wavelet features are fed into the neural network. One hidden layer with 80 neurons is used in the neural network. The output layer has ten neurons for the ten digits. The network target is defined as a vector with 10 real values. Only one value is 0.9 representing the numeral category by its location in the vector. All other values in the vector are set to 0.1. The optimal wavelet filters can be learned by using Newton-Raphson method. The objective function is defined as the mean square error of the neural network output. This objective function is used for training both the weights of the neural network and the optimal wavelet filters. We consider different wavelet filters for different wavelet decomposition levels for both rows and columns. The initial value of the optimal wavelet filter parameter θ is set to $\pi/12$. This corresponds to the Daubechies-4 wavelet. The neural network weights are initialized as random values at the beginning. The training epoches are set to 4800. The Concordia University *CENPARMI* handwritten numeral database is used for the training and testing. This database contains 6000 unconstrained handwritten numerals originally collected from dead letter envelopes by the U.S. postal service at different locations. The numerals in the database are stored in bi-level format. We use 4000 numerals for training and 2000 for testing. Some of the numerals are very difficult to recognize even with human eyes. For comparison, we compare the recognition rates of the same algorithm with standard wavelets and with optimal wavelets. We get a recognition rate of 91.70% by using optimal wavelets, whereas 90.10% for the same network with Daubechies-4 wavelet. Note that we only used simple wavelet features in our experiments and proved that higher recognition rate can be obtained by employing optimal wavelets. If we use more complicated invariant wavelet features, then much higher recognition rate is expected for our handwritten numeral recognition task. We list the optimal wavelet filters we obtained by using the *CENPARMI* handwritten numeral database as follows:

$$
\begin{aligned}
f_1 &= (-0.1159, \quad 0.1814, \quad 0.8230, \quad 0.5257) \\
g_1 &= (-0.1163, \quad 0.1825, \quad 0.8234, \quad 0.5246) \\
f_2 &= (-0.1424, \quad 0.2897, \quad 0.8495, \quad 0.4174) \\
g_2 &= (-0.0722, \quad 0.0914, \quad 0.7793, \quad 0.6157)
\end{aligned}
$$

Note that f_1 and f_2 are the optimal wavelet filters for row wavelet decomposition level 1 and 2, respectively. $g1$ and $g2$ are the optimal wavelet filters for column wavelet decomposition level 1 and 2, respectively. Figure 2 and Figure 3 illustrate the shapes of these scaling functions and mother wavelets. Optimal wavelets depend on the training dataset and the neural network used, so different optimal wavelet filters should be obtained for different datasets and descriptors.

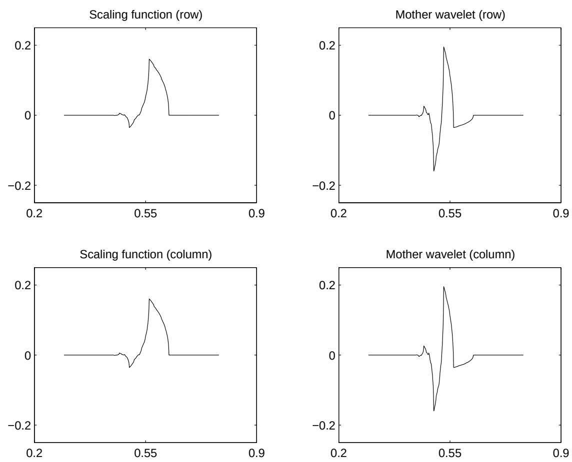

Figure 2: The scaling functions and mother wavelets for the first level wavelet decomposition for row and column, respectively.

5 Conclusion

In this paper, we propose to use optimal wavelets in the wavelet neural network for pattern recognition. Experimental results show that by adaptively finding the optimal wavelets we get higher recognition rate for unconstrained handwritten numeral recognition. Note that in addition to handwritten numeral recognition, we can also use optimal wavelets in other pattern recognition applications as long as wavelet features and neural networks are used.

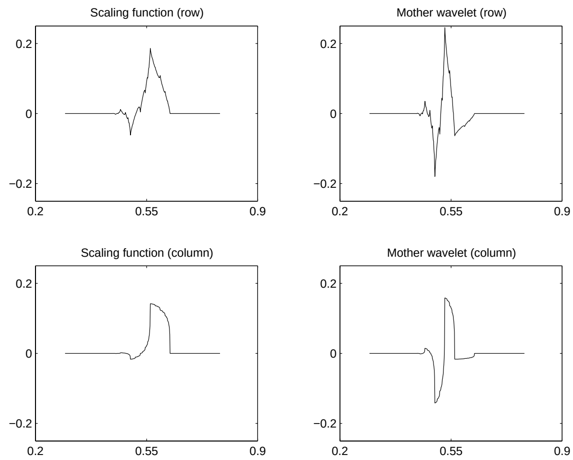

Figure 3: The scaling functions and mother wavelets for the second level wavelet decomposition for row and column, respectively.

Acknowledgments

This work was supported by research grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and by the Fonds Quebecois de la Recherche sur la Nature et les Technologies (FQRNT) of Quebec.

References

[1] Yvette Mallet, Danny Coomans, Jerry Kautsky, Olivier De Vel, "Classification Using Adaptive Wavelets for Feature Extraction," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 10, pp. 1058-1066, 1997

[2] Xiaoling Wang and Hairong Qi, "Face Recognition Using Optimal Non-Orthogonal Wavelet Basis Evaluated by Information Complexity," *16 th International Conference on Pattern Recognition (ICPR'02)*, Volume 1, pp. 10164, August 11 - 15, 2002, Quebec City, QC, Canada.

[3] Joseph O. Chapa and Mysore R. Raghuveer, "Optimal matched wavelet construction and its application to image pattern recognition," *SPIE Proceedings* Vol. 2491, Wavelet Applications II, 1995, Orlando, FL, USA, 1995.

[4] Joseph O. Chapa and Mysore R. Raghuveer, "Algorithms for Designing Wavelets to Match a Specified Signal," *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3395-3406, 2000.

[5] Yan Zhuang and J. S. Baras, "Constructing Optimal Wavelet Basis for Image Compression," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2351-2354, 1996.

[6] Ahmed H. Tewfik, Deepen Sinha and Paul Jorgensen, “On the Optimal Choice of a Wavelet for Signal Representation,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 747-765, 1992.

[7] A. Das, U. B. Desai and P. P. Vaidya, “Search for Optimal Basis for signal Denoising in the Space of n-tap Wavelets,” *International Symposium on Signal Processing and its Applications (ISSPA)*, Kuala Lumpur, Malaysia, pp. 96-99, 2001.

[8] S. A. Golden, “Identifying Multiscale Statistical Models using the Wavelet Transform,” M.S. Thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, USA, 1991.

[9] I. Daubechies, “Orthonormal bases of compactly supported wavelets,” *Comm. on Pure and Applied Mathematics* **41**, pp. 909-996, 1988.

[10] G. Strang, “Wavelets and Dilation Equations: A Brief Introduction.” *SIAM Review*, vol. 31, no. 4, pp. 614-627, 1989.

[11] H. Zou and A. H. Tewfik, “Parametrization of Compactly Supported Orthonormal Wavelets,” *IEEE Transactions on Signal Processing*, vol. 41, no. 3, pp. 1428-1531, 1993.

[12] Seong-Whan Lee, Chang-Hun Kim, Hong Ma and Yuan Y. Tang, “Multiresolution Recognition of Unconstrained Handwritten Numerals with Wavelet Transform and Multilayer Cluster Neural Network,” *Pattern Recognition*, vol. 29, no. 12, pp. 1953-1961, 1996.

[13] Patrick Wunsch and Andrew F. Laine, “Wavelet descriptors for multiresolution recognition of handprinted characters,” *Pattern Recognition*, vol. 28, pp. 1237-1249, 1995.

[14] G. Y. Chen, T. D. Bui and A. Krzyzak, “Contour-Based Handwritten Numeral Recognition using Multiwavelets and Neural Networks,” *Pattern Recognition*, Vol.36, No.7, pp.1597-1604, 2003.