Texture Classification Based on Coevolution Approach in Multiwavelet Feature Space

Jing-Wein Wang

Center of General Studies, National Kaohsiung University of Applied Sciences
415 Chien-Kung Road, Kaohsiung 807, Taiwan, R.O.C.
Tel.: 886-7-3814526 Ext. 3350, Fax.: 886-7-5590462
jwwang@cc.kuas.edu.tw

Abstract. To test the effectiveness of multiwavelets in texture classification with respect to scalar Daubechies wavelets, we study the evolutionary-based algorithm to evaluate the classification performance of each subset of selected feature. The approach creates two populations that have interdependent evolutions corresponding to inter and intra distance measure, respectively. With the proposed fitness function composed of the individuals in competition, the evolution of the distinct populations is performed simultaneously through a coevolutionary process and selects frequency channel features of greater discriminatory power. Consistently better performance of the experiments suggests that the multiwavelet transform features may contain more texture information for classification than the scalar wavelet transform features. Classification performance comparisons using a set of twelve Brodatz textured images and wavelet packet decompositions with the novel packet-tree feature selection support this conclusion.

1 Introduction

Multiwavelets have recently attracted a lot of theoretical attention and provided a good indication of a potential impact on signal processing [1]. In this paper, a novel texture classification scheme using Geronimo-Hardin-Massopust (GHM) discrete multiwavelet transform (DMWT) [2] is proposed. The goal is both to extend the experimentation made in [1], and to test the effectiveness of multiwavelets in texture classification with respect to the scalar Daubechies wavelet [3]. An important problem in wavelet texture analysis is that the numbers of features tend to become huge. Inspiration by Siedlecki and Sklansky [4], Wang *et al.* [5] proposed the Max-Max method based on Genetic Algorithms (GAs) [6] to evaluate the classification performance of each subset of selected features. A feature of GAs is that the chromosomes interact only with the fitness function, but not with each other. This method precludes the evolution of collective solutions to problems, which can be very powerful. The approach proposed here an evolutionary framework in which

successive generations adaptively develop behavior in accordance with their natural needs.

This paper is organized as follows. A brief review of DMWT is presented in Section 2. The coevolutionary-based feature selection algorithm is proposed in Section 3. The texture classification experimental results are presented in Section 4. We summarize the conclusions in Section 5.

2 Discrete Multiwavelet Transforms

For a multiresolution analysis of multiplicity $r > 1$, (MRA), an orthonormal compact support multiwavelet system consists one multiscaling function vector $\Phi(x) = (\phi_1(x), \dots, \phi_r(x))^T$ and one multiwavelet function vector $\Psi(x) = (\psi_1(x), \dots, \psi_r(x))^T$. Both Φ and Ψ satisfy the following two-scale relations:

$$\Phi(x) = \sqrt{2} \sum_{k \in \mathbf{Z}} H_k \Phi(2x - k) \quad (1)$$

$$\Psi(x) = \sqrt{2} \sum_{k \in \mathbf{Z}} G_k \Phi(2x - k). \quad (2)$$

Note that multifilters $\{H_k\}$ and $\{G_k\}$ are finite sequences of $r \times r$ matrices for each integer k . Let $V_j, j \in \mathbf{Z}$, be the closure of the linear span of $\phi_{l,j,k} = 2^{j/2} \phi_l(2^j x - k)$, $l = 1, 2, \dots, r$. By exploiting the properties of the MRA, as in the scalar case, any continuous-time signal $f(x) \in V_0$ can be expanded as

$$\begin{aligned} f(x) &= \sum_{l=1}^r \sum_{k \in \mathbf{Z}} c_{l,0,k} \phi_l(x - k) \\ &= \sum_{l=1}^r \sum_{k \in \mathbf{Z}} c_{l,1,k} 2^{j/2} \phi_l(2^j x - k) + \sum_{l=1}^r \sum_{J \leq j < 0} \sum_{k \in \mathbf{Z}} d_{l,j,k} 2^{j/2} \psi_l(2^j x - k) \end{aligned} \quad (3)$$

where

$$\mathbf{c}_{j,k} = \left(c_{1,j,k}, \dots, c_{r,j,k} \right)^T \quad (4)$$

$$\mathbf{d}_{j,k} = \left(d_{1,j,k}, \dots, d_{r,j,k} \right)^T \quad (5)$$

and

$$c_{l,j,k} = \int f(x) 2^{j/2} \phi_l(2^j x - k) dx \quad (6)$$

$$d_{l,j,k} = \int f(x) 2^{j/2} \psi_l(2^j x - k) dx. \quad (7)$$

For the two-dimensional discrete multiwavelet transform, a 2-D MRA of multiplicity N for $L^2(R^2)$ can be obtained by using the tensor product of two MRA's of multiplicity N of $L^2(R)$. Fig. 1 shows a textured image of size 512×512 and its one-level decomposition with the GHM multiwavelet transform.

Fig. 1. One-level decomposition of the DMWT for the Brodatz texture D15

3 Feature Selection Using Evolutionary Algorithms

3.1 Initialization

In the proposed method that is derived from the principles of the natural species evolution theory [7] individuals are grouped in populations and thereafter referred to as *inter* population P_b and *intra* population P_w are randomly created. In our case, the two populations have interdependent evolutions (coevolution). The term *inter* reflects the reluctance of this individual for the opposite class. This reluctance is quantified by the mean square distance between pattern points that belong to different classes. An individual of the population P_b, I_x , will compete with each individual of the population kernel K_b which is the collection of individuals with best inter distances. The term *Inter* is formulated as follows:

$$[Inter]_{I_x \in P_b} = \sum_m (I_x \Leftrightarrow I_m) \text{ with } I_m \in K_b, m = 1, \dots, M,$$

$$(I_x \Leftrightarrow I_m) = \begin{cases} D_b^x - D_b^m & \text{if } D_b^x > D_b^m, \\ \rho & \text{if } D_b^x \leq D_b^m, \end{cases} \quad (8)$$

where D_b is the Euclidean distance between classes and ρ is a penalty. Conversely, the term *Intra* reflects the attraction of this individual for its own class. An individual of the population P_w, I_x , will compete with each individual of the population kernel K_w which is the collection of individuals with best intra distances.

3.2 Evaluation

A character string whose length is equal to the number of features, called the chromosome, as done in GAs represents an individual of the population. With a direct encoding scheme, a bit of one indicates that the feature is used and zero indicates that the feature is not used. The fitness of an individual is derived from her genotype made up of the genes which guide the discrete wavelet decomposition of each textured image, in accordance with our proposed packet-tree representation described in Section 3.4. The genetic representation is used to evolve potential solutions under a set of twelve Brodatz textured images [8]. A best individual of the population kernel K_b will compete with each of the best individuals of the opposite population kernel K_w . The combined results of these competitions directly provide the fitness function. Following the abovementioned, the fitness function Θ is defined as a number composed of two terms:

$$\Theta = (1 - \xi \cdot \delta / \chi) \cdot ([Inter] - [Intra]), \quad (9)$$

where ξ is the weighting constant greater or equal to one, δ is the number of features selected, χ is the number of training samples. The evaluation process of Θ is randomly combined with the *Inter* individual of the population kernel K_b and the *Intra* individual of the population kernel K_w .

3.3 Genetic Operations

According to the roulette wheel selection strategy [6], the combination of populations P_b and P_w individuals with higher fitness value in equation (9) will survive more at the next generation while ones with lower fitness will be removed from the population. The size of each of the population remains constant during evolution. Reproduction step consists of generating new individuals from the combinative individuals previously selected and is performed with crossover and mutation operations. The crossover operator is performed to create new chromosomes and the mutation operation randomly changes a bit of the string. In our method, the combinative individuals selected in the previous step are used to as the parent individuals and then their chromosomes are combined by the following criterion so as to toward the chromosomes of two offspring individuals.

Combinative Crossover Criterion

If the i -th gene of the *inter* individual and the i -th gene of the *intra* individual are the same, then the i -th gene of the offspring individual is set as either individual. Where i is the index of a gene in an individual. If not, the i -th gene of the offspring individual will be set as either individual at random.

3.4 Feature Selection

After computation of the fitness function for all the combination of the two kernel individuals, a feature selection step is activated for choosing the individuals allowed

reproducing at the next generation. The strategy of feature selection involves selecting the best subset A_q ,

$$A_q = \{\alpha_u \mid u = 1, \dots, q; \alpha_u \in B\} \quad (10)$$

from an original feature set B ,

$$B = \{\beta_v \mid v = 1, \dots, Q\}, Q > q. \quad (11)$$

In other words, the combination of q features from A_q will maximize equation (9) with respect to any other combination of q features taken from Q , respectively. The new feature β_v is chosen as the $(\lambda+1)$ st feature if it yields

$$\text{Max}_{\forall \beta_v} \text{Max}_{\forall \alpha_u} \Delta[\text{Inter}](\alpha_u, \beta_v), \quad (12)$$

where $\alpha_u \in A_\lambda$, $\beta_v \in B - A_\lambda$, and $\Delta[\text{Inter}](\alpha_u, \beta_v) = [\text{Inter}](\alpha_u, \beta_v) - [\text{Inter}](\alpha_u)$. $[\text{Inter}](\alpha_u)$ is the evaluation value of equation (8) while the feature α_u is selected and $[\text{Inter}](\alpha_u, \beta_v)$ is the evaluation value of equation (8) while the candidate β_v is added to the already selected feature α_u . In a similar way, the feature selection mechanism minimizes intra measure and helps to facilitate classification by removing redundant features that may impede recognition. The proposed schemes consider both the accuracy of classification and the cost of performing classification.

To speed up such a selection process, we consider a packet-tree scheme that is based on fitness value of equation (9) to locate dominant wavelet subbands. The idea is given as below.

- 1) For a predetermined number of levels, given textured subbands that were decomposed with wavelet packet transforms into.
- 2) Initially select subbands that can be viewed as the parent and children nodes in a tree at random.
- 3) After initialization, the subbands at the current level will be selected only if the predecessor at the previous level was selected. Otherwise, we just skip these successors and consider the next subbands.
- 4) Generate a representative tree of selected features for each texture by averaging the selected feature vectors over all the sample images.
- 5) Repeat the process for all textures.

4 Teyxture Classification Experiments

4.1 Experiment Design

In this section, we present the experimental results on twelve 512×512 images with 256 gray levels found in the Brodatz's album. The reported results for each classification task have the following parameter settings: population size = 20, number of generation = 1000, and the probability of crossover = 0.5. A mutation probability value starts with a value of 0.9 and then varied as a step function of the number of

iterations until it reaches a value of 0.01. For the scalar wavelet we randomly chose one hundred 256×256 overlapping subimages as training samples for each obtained original textured image and used one thousand samples in the multiwavelet. Due to the curse of dimensionality, this arrangement considers the subband structure resulting from passing a signal through a multifilter bank is different from a scalar wavelet. We tackled the classification problem also with the D4 wavelet transform and the GHM multiwavelet transform. Textural features are given by the extrema number of wavelet coefficients [5], which can be used as a measure of coarseness of the texture at multiple resolutions. Then, texture classifications without and with feature selection were performed using the simplified Mahalanobis distance measure [9] to discriminate textures and to optimize classification by searching for near-optimal feature subsets. The mean and variance of the decomposed subbands are calculated with the leave-one-out algorithm [9] in classification.

The training phase of texture classification composed of the evaluation, selection, and reproduction steps correspond to a generation is repeated until the fitness function does not progress, or when a maximum number of generations is reached. Then, the best string with the highest fitness value in equation (9) at the last generation is preserved for further use in the classification phase. During the classification phase, the unknown texture (one of the twelve Brodatz textures) is matched against the database and the best match is taken as the classification result.

4.2 Experimental Results and Discussions

4.2.1 Texture Classification without Feature Selection

The performance of the classifier was evaluated with three different randomly chosen training and test sets. Algorithms based on the two types of wavelets have been shown to work well in texture discrimination. At the levels 1-3 of Table 1, the percentage of correct classification rate without feature selection improves as the number of levels increases. This observation is expected since there is no problem of curse of dimensionality. However, the classification performance is not monotonically dependent on the feature space dimension, but decreases after having reached a local maximum, as has been shown in Table 1. At the level 4, the inclusion of further feature results in performance degradation. In a similar way, the classification rate is even down to 95.30% at the level 3 of Table 2. Theoretically, multiwavelets should perform even better because scalar wavelets cannot simultaneously possess all of the desired properties, such as orthogonality and symmetry. However, average classification accuracies in the latter have produced performance capable of surpassing the results of the former. This owes to the fact that the number of features should be reduced more to alleviate the serious Hughes phenomenon [9] when the training samples are limited.

4.2.2 Texture Classification with CGA Feature Selection

By making a comparison with the previous results, the classification errors in Tables 3 and 4 mostly decrease when the used features are selectively removed from all the features at the level 4 of Table 1 and level 3 of Table 2, respectively. This decrease is due to the fact that less parameter used in place of the true value of the class

conditional probability density functions need to be estimated from the same number of samples. The smaller the number of the parameters that need to be estimated, the less severe the Hughes effect can become. In the meanwhile, we also noticed that the multiwavelet outperforms the scalar wavelet with the packet-tree selection. This is because the extracted features in the former are more discriminative than the latter and, therefore, the selection of a subband for discrimination is not only dependent on the wavelet bases, wavelet decompositions, and decomposed levels but also the fitness function.

Table 1. Classification results (correct rate in %) of the pyramidal decomposition using wavelet transforms without feature selection

Level Sample set	1	2	3	4
1	95.17	98.42	97.83	97.33
2	95.17	98.92	98.75	97.25
3	94.83	97.75	99.00	96.92
Average	95.00	98.36	98.53	97.17

Table 2. Classification results (correct rate in %) of the pyramidal decomposition using multiwavelet transforms without feature selection

Level Sample Set	1	2	3
1	95.42	96.17	95.25
2	95.00	96.25	95.91
3	94.42	96.67	94.75
Average	94.95	96.36	95.30

Table 3. Classification results (correct rate in %) using the wavelet packet decomposition with coevolutionary GA feature selection

Sample Set	$\xi = 1$	$\xi = 2$	$\xi = 3$	$\xi = 4$	$\xi = 5$
1	98.47	98.48	98.29	98.49	98.43
2	98.41	98.38	98.62	98.29	98.49
3	98.41	98.43	98.38	98.51	98.49
Average	98.43	98.43	98.43	98.43	98.47

Table 4. Classification results (correct rate in %) using the multiwavelet packet decomposition with coevolutionary GA feature selection

Sample Set	$\xi = 1$	$\xi = 2$	$\xi = 3$	$\xi = 4$	$\xi = 5$
1	98.83	98.80	98.54	98.82	98.58
2	98.85	98.72	98.80	98.71	98.66
3	98.96	98.73	98.90	98.79	98.73
Average	98.88	98.75	98.75	98.77	98.79

5 Conclusions

This paper introduces a promising evolutionary algorithm approach for solving the texture classification problem. Further work is to explore the feasibility of our new methods by incorporating with recent preprocessing techniques of multiwavelets.

Acknowledgements

The author would like to acknowledge the support received from NSC through project number NSC 90-2213-E-151-010.

References

1. Strela, V., Heller, N., Strang, G., Topiwala, P., and Heil, C.: The Application of Multiwavelet Filter Banks to Image Processing. *IEEE Trans. Image Process.*, 8 (1999) 548-563
2. Xia, X. G., Geronimo, J. S., Hardin, D. P., and Suter, B. W.: Design of Prefilters for Discrete Multiwavelet Transforms. *IEEE Trans. Signal Process.*, 44 (1996) 25-35
3. Daubechies, I. (ed.): *Ten Lectures on Wavelets*. SIAM, Philadelphia, Penn. (1992)
4. Siedlecki, W. and Sklansky, J.: A Note on Genetic Algorithm for Large-Scale Feature Selection. *Pattern Recognition Letters*, 10 (1989) 335-347
5. Wang, J. W., Chen, C. H., Chien, W. M., and Tsai, C. M.: Texture Classification using Non-Separable Two-Dimensional Wavelets. *Pattern Recognition Letters*, 19 (1998) 1225-1234
6. Goldberg, D. E. (ed.): *Genetic Algorithms in Search, Optimization, and Machine Learning*. MA: Addison-Wesley (1989)
7. Bäck, T.: *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*, Oxford University Press, New York (1996)
8. Brodatz, P. (ed.): *Textures: A Photographic Album for Artists and Designers*. NY: Dover (1966)
9. Devijver, P. A. and Kittler, J. (ed.): *Pattern Recognition: A Statistical Approach*. Prentice-Hall, Englewood Cliffs, NJ (1982)