MVA '98 IAPR Workshop on Machine Vision Applications, Nov. 17-19, 1998, Makuhari, Chiba, Japan

8—5 2-D Non-separable Wavelet Bases for Texture Classification with Genetic Feature Selection

B. Y. Liao♱, J. S. Pan♱, J. W. Wang† and L. Hong♱

♱National Kaohsiung Institute of Technology, Taiwan, R.O.C.
† National Cheng Kung University, Taiwan, R.O.C.

Abstract

In this paper, the performances of texture classification based on pyramidal and uniform decomposition are comparatively studied with and without feature selection. This comparison using the subband variance as feature explores the dependence among features. It is shown that the main problem when employing 2-D non-separable wavelet transforms for texture classification is the determination of the suitable features and that yields the best classification results. A Max-Max algorithm which is a novel evaluation function based on genetic algorithms is presented to evaluate the classification performance of each subset of selected features. Experimental results have shown the selectivity of the proposed approach and do capture the texture characteristics.

1 Introduction

The wavelet transform is a multiresolution technique, which can be implemented as a pyramidal or uniform (tree-structured) decomposition. Several contributions have proposed pyramidal and tree-structured wavelet transforms as alternatives for texture feature extraction and classification [5-7]. Although pyramidal and tree structured wavelet transforms can map the useful information content into a lower dimensional feature space, yet they achieves no dimensionality reduction in the space of the original features.

The feature selection is applied for texture classification in this paper. The main goal of feature selection is to select a subset of q features from the given set of Q features, $q < Q$, without significantly degrading the performance of the recognition system. Achieving this goal requires a capability for evaluating the effectiveness of a feature subset and an effective strategy for searching for the best q features from the given Q features. Recently several authors proposed randomized population-based heuristic search techniques that simulates natural process in biology to select feature subsets for use with decision tree or nearest neighbor classifiers [1-2]. This technique is known as genetic algorithms (GA) and is based on the assumption that large domains of data are organized and evolve in a manner similar to processes occurring in nature.

In this paper, a novel evaluation function defined as the inter-class distance minus the intra-class distance is proposed in a genetic algorithm to explore the importance of individual feature in the optimal classification. Our approach is to use the Euclidean distance as the distance measure in the proposed evaluation function. The inter-class distance is defined as the mean square distance between pattern points that belong to different classes. The intra-class distance is the mean square distance between pattern points of the same class. In our algorithm, optimum feature selection is dictated by the maximization of the evaluation function. After energy-based features are extracted from pyramidal or uniform decomposition via 2-D non-separable wavelet transforms, the evaluation function of the genetic algorithm for feature selection is combined with the simplified Mahalanobis classifier to optimize classification by searching for near-optimal feature subsets.

This paper is organized as follows. 2-D non-separable wavelets are described in Section 2. In Section 3, the classification algorithms is presented. Section 4 describes a novel evaluation function for the feature selection problem. Section 5 presents experimental results of classification and some discussions to explore the importance of individual feature in optimal classification.

2 2-D Non-Separable Wavelets

In most of the previous works on two dimensional texture processing the sampling rate changes used are separable and can be performed alone one dimension at a time. However, for two dimensional textures, a non-separable wavelet basis [3] opens a possibility of having schemes better adapted to the human visual system and thus will be used in this work to decompose the texture images. If it is desirable to have features invariant to rotation of the texture image, a 2-D non-separable wavelet basis can be used to cope with the problem.

3 Classification algorithm

The classification algorithm can be separated into learning phase and classification phase as follows:

Learning Phase

(1) Decompose a given texture image using the 2-D non-separable wavelet basis into image subbands.
(2) Calculate the sampled energy of each subband. If the decomposed image is $f(r, s)$ with $0 \le r \le R$-1 and $0 \le s \le S$-1, the energy e is calculated as follows:

$$mean = \frac{1}{RS}\sum_{r=0}^{R-1}\sum_{s=0}^{S-1} f(r, s), \tag{1}$$

$$e = \frac{1}{RS}\sum_{r=0}^{R-1}\sum_{s=0}^{S-1}[f(r, s) - mean]^2. \tag{2}$$

(3) Use the decomposed images as the next input, increase the level by 1. Note, it decomposes the image with a multiresolution scale factor of $\sqrt{2}$.
(4) Repeat the steps 1 to 3 and terminate until the desired resolution is reached.
(5) After the desired resolution, we have the feature set given by

$$\{ e_l^j \mid j = 0, 1, \ldots, J, \text{ and } l = 0, 1, \ldots, L-1 \},$$

where e_l^j denotes the energy from the lth subband at level j. For the case of pyramidal decomposition, l = 0 and 1 ($L = 2^1$) indicating the lowpass and highpass bands respectively. For the case of uniform decomposition, $l = 0, \ldots, 2^J - 1$ ($L = 2^J$) indicating the uniform bands.
(6) Repeat the steps 1 to 5 for each sample image from the same texture.
(7) Generate the mean and variance of energy for each texture.
(8) Repeat the process for all textures.

Classification Phase

(1) Decomposes an unknown input texture using the steps 1 to 5 of the *Learning Phase*.
(2) Calculate the distance from the input texture to each texture in the database.
(3) Assigns the unknown input texture to texture i if $D_i < D_{i'}$ for all $i' \neq i$.

4 Feature selection using genetic algorithms

The evaluation function has to consider accuracy of classification and the cost of performing classification. The strategy of feature selection involves selecting the best subset A_q,

$$\mathrm{A}_q = \{\alpha_u \mid u = 1, \ldots, q;\ \alpha_u \in \mathrm{B}\} \tag{3}$$

from an original feature set B,

$$\mathrm{B} = \{\beta_v \mid v = 1, \ldots, Q\}, Q > q. \tag{4}$$

In other words, the combination of q features from A_q will maximize an evaluation function $J(\cdot)$ with respect to any other combination of q features taken from Q. In the Max-Max method, the new feature β_v is chosen as the (k+1)st feature if it yields

$$\operatorname*{Max}_{\forall \beta_v} \operatorname*{Max}_{\forall \alpha_u} \Delta J(\alpha_u, \beta_v) \tag{5}$$

where $\alpha_u \in A_k$, $\beta_v \in B - A_k$, and $\Delta J(\alpha_u, \beta_v) = J(\alpha_u, \beta_v) - J(\alpha_u)$. $J(\alpha_u)$ is the value of the evaluation function while the feature α_u is selected and $J(\alpha_u, \beta_v)$ is the value of the evaluation function while the candidate β_v is added to the already selected feature α_u.

In this paper, we define the evaluation function J as

$$J = (1 - \xi \times \delta/\chi) \times (D_1 - D_2) \tag{6}$$

where ξ is the constant greater or equal to one, δ is the number of selected features, χ is the number of training samples, D_1 is the Euclidean distance between classes, and D_2 is the Euclidean distance within class. The following constraints of Eqs. (7) and (8) are used to bias the search process so as to provide an improvement in the population's average fitness:

$$\text{the latest } (1 - \xi \times \delta/\chi) \times D_1 > \text{the previous } (1 - \xi \times \delta/\chi) \times D_1, \tag{7}$$

and

$$\text{the latest } (1 - \xi \times \delta/\chi) \times D_2$$

< the previous $(1 - \xi \times \delta/\chi) \times D_2$. (8)

In this research, we consider the standard GA by randomly creating an initial population of size P. A direct encoding scheme is used to construct binary GA parent strings to produce offspring that contain some parts of both parent's genetic material. Mutation is an operator that introduces variations into the string and applied to each generation. In order to obtain the optimal string, the population diversity is needed. At early iterations the mutation probability P_m is set to high and as the optimal string is being approached few changes in the present strings are necessary. The process of crossover, evaluation, and selection is repeated for a predetermined number of generations and the best string obtained is taken as the optimal one.

5 Experimental results and discussions

Classification experiments were conducted using twelve standard 512 × 512 Brodatz textures with 256 gray levels obtained from a public archive. One hundred 256 × 256 overlapping subimages are randomly chosen from the original image and used in the training and the classification phases. The mean and variance of the decomposed subbands are calculated with the leave-one-out algorithm. During the classification phase, the unknown texture is matched against the database and the best match is taken as the classification result. The reported results for each classification task have the following parameter settings: population size $P = 200$, number of generation = 20, and the probability of crossover $P_c = 0.5$. A mutation probability value starts with $P_m = 0.9$. The P_m value is then varied as a step function of the number of iterations until it reaches a value of 0.1.

In the following, we summarize the classification results obtained from the twelve classes of textures. Table 1 shows the performances of the pyramidal decomposition using features from all last levels with and without feature selection. The band ordering of the "Feature Selection Vector" in the last column is from high to low, for example, they are 0H, 1H, 2H, 3H, 4H, 5H, 6H, 7H, and 7L for levels 0 - 7. Table 2 summarizes the classification results of the uniform decomposition using features from the last level. The band ordering of the "Feature Selection Vector" in the last column is from high to low, for example, they are 30H, 31H, ..., 37H, 30L, 31L, ..., and 37L for level 3.

The goal of feature selection is to find the smallest or least costly subset of features for which the classifer's performance does not deteriorate below a certain specified level. In order to examine which of the strings whose length is equal to the number of features. A bit of one indicates that the feature is used; 0 indicates that the feature is not used. Parent selection emulates the survival-of-the-fittest mechanism in nature. Single-point crossover exchange information between two available features are most important for the discrimination and which are not good representatives of the training samples, the constant ξ is used to tune the number of the selected features of the evaluation function, and its value is determined on a problem-specific basis. In other words, the value of ξ could be different from these levels to the next levels and therefore it is suggested as 2 and 3 in the experiments so as to minimize the number of used features.

In table 1, the percentage of correct classification rate with and without feature selection improves as the number of levels increases. This observation is expected since there is no problem of curse of dimensionality. At the levels 0, 1, and 2, the results of table 2 without feature selection is as good as the results of table 1. At the level 3 of table 2, the performance is increased less than 1% although the number of used features is a double of the level 2 of table 2. The classification rate is even down to 86.21% at the level 4 in our experiments. This decreases classification performance compared to the pyramidal decomposition. The effect of decreasing classification performance with increasing feature space dimension is related to the Hughes effect [4] and has two reasons. First, the cause of it is that only a finite number of training samples was used in these experiments for a given texture. Second, the reason for worse discrimination performance is that the additional features at higher levels of the uniform decomposition may contain noisy information and negates the presence of the features with discriminatory information. This shows that the discriminatory characteristics of texture spread more in low-pass bands and the features extracted from the pyramidal decomposition are more representative for texture images in the experiments. On the other hand, the classification error decreases when the used features are selectively removed from all features at the level 3 of table 2. This decrease is due to the fact that less parameters 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.

The Max-Max evaluation function of doing feature selection and warping of the feature space to optimize classification has been demonstrated to be very powerful in dealing with texture images. Simulations of the selected feature subsets have shown the selectivity of the proposed approach and do capture the texture characteristics. The tabu search approach is also a very powerful optimization technique which can be used for the feature selection which is called tabu feature selection and the experiments are processing now.

References

1 W. Siedlecki and J. Sklansky, "A note on genetic algorithm for large-scale feature selection," Pattern Recognition Letters, 10, (5), pp. 335-347, 1989.

2 M. Richeldi and P. Lanzi, "Performing effective feature selection by investigating the deep structure of the data," Proc. of the Second Int. Conf. on Knowledge Discovery and Data Mining, AAAI Press, pp. 379-383, 1996.

3 J. Kovacevic and M. Vetterli, "Non-separable two- and three- dimensional wavelets," IEEE Trans. Image. Process., 43, pp. 1269-1273, 1995.

4 G. F Hughes, "On the mean accuracy of statistical pattern recognizers,' IEEE Trans. Inform. Theo. IT-14, pp. 55-63, 1968.

5 T. Chang and C. C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," IEEE Trans. Image Process., 2, pp. 429-441, 1993.

6 A. Laine and J. Fan, "Texture classification by wavelet packet signatures," IEEE Trans. Pattern Anal. Machine Intell., 15, pp. 1186-1191, 1993.

7 M. Unser, "Texture classification and segmentation using wavelet frames," IEEE Trans. Image Process., 4, pp. 1549-1560, 1995.

Level	All Features	GA-Selected Subset	Feature Selection Vector
0	83.16	83.16	11
0 - 1	87.29	87.29	111
0 - 2	87.29	87.29	1111
0 - 3	88.80	88.47	10111
0 - 4	91.08	90.24	000111
0 - 5	94.19	93.27	0001101
0 - 6	96.04	94.02	00011101
0 - 7	98.15	97.39	001110001

Table 1 Classification results of the pyramidal decomposition using 2-D non-separable wavelet transforms with and without GA feature selection using features from all last levels (correct rate in %).

Level	All Features	GA-Selected Subset	Feature Selection Vector
0	83.16	83.16	11
1	87.29	87.29	1111
2	87.79	87.54	00001111
3	88.55	89.73	0110011011111011

Table 2 Classification results of the uniform decomposition using 2-D non-separable wavelet transforms with and without GA feature selection using features from the last level (correct rate in %).