Comparison Study of Textural Descriptors for Training Neural Network Classifiers

G.D. MAGOULAS[(1)], S.A. KARKANIS [(1)], D.A. KARRAS[(2)] and M.N. VRAHATIS[(3)]

[(1)] Department of Informatics
University of Athens, GR-157.84 Athens, Greece

[(2)] Department of Business Administration
University of Piraeus, Greece

[(3)]Department of Mathematics,
University of Patras, Patras, Greece

Abstract: - In this paper the use of feedforward neural networks in the characterization of images by texture content is investigated. An in depth experimental study is conducted comparing several well known textural feature extraction techniques along with a novel discrete wavelet transform based methodology. It is demonstrated that the new technique leads to the design and selection of feedforward neural networks architectures with the best texture classification accuracy.

Key-words: - Feature extraction, texture classification, wavelet-based texture descriptors, cooccurrence matrices, gray level run length moments, fractal dimension, statistical texture analysis, feedforward neural networks, training.
IMACS/IEEE CSCC'99 Proceedings, Pages:6221-6226

1 Introduction

The problem of texture-based segmentation and classification of images is of considerable interest in many image-processing applications, like remote sensing, medical imaging, and quality control. Analysis of textures requires the identification of proper descriptors or features that differentiate the textures in the image for further segmentation, classification and recognition. Techniques capable of producing appropriate textural descriptors include simple statistical measures of gray level distribution, measures of local density or other gradient features, the use of run lengths and the computation of second order statistics, like the cooccurrence matrices, [2][3][8][12].

The emergence of the 2-D wavelet transform [7][10][13][14] as a popular tool in image processing offers the ability of robust feature extraction in images. Due to their strong localization properties, wavelets have been proven to be appropriate for the description of the textural information in the image providing a richer problem specific-information. Furthermore second order statistical measures can be evaluated on the output of the 2D wavelet transformation in order to create a more reliable framework for the generation of the textural descriptors [13].

In this paper we propose a novel methodology for texture classification of images by examining the discrimination abilities of their textural properties. Besides neural network classifiers and the 2-D wavelet transform, the tools utilized in this paper are the cooccurrence matrices, the fractal dimension, and the gray level run length methods for textural feature extraction.

The contribution of the paper lies on the use of a novel texture descriptor for a classification scheme based on Feedforward Neural Networks (FNNs) and on the investigation of the effects of different textural descriptors on FNN learning and generalization capabilities. The experimental study conducted in this paper aims precisely at illustrating this latter investigation.

2 Textural Descriptors used

In this section three widely known feature extraction methods are briefly described.

2.1 The cooccurrence matrices

Cooccurrence matrices [2] represent the spatial distribution and the dependence of the gray levels within a local area. Each (i,j)th entry of the matrices, represents the probability of going from one pixel with gray level (i) to another with a gray level (j) under a predefined distance and angle. From these matrices, sets of statistical measures are computed (called feature vectors) for building different texture models.

In our experiments, we have considered four angles, namely 0, 45, 90, 135 as well as a predefined distance of one pixel in the formation of the cooccurrence matrices. Therefore, we have formed four cooccurrence matrices. According to our experiments, the following four statistical measures out of the 14, originally proposed by Haralick [2][3], provide high discrimination accuracy that can be only marginally increased by adding more measures in the feature vector:

- ***Energy - Angular Second Moment***

$$f_1 = \sum_i \sum_j p(i,j)^2$$

- ***Correlation***

$$f_2 = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i * j) p(i,j) - m_x m_y}{s_x s_y}$$

- ***Inverse Difference Moment***

$$f_3 = \sum_i \sum_j \frac{1}{1 + (i - j)} p(i,j)$$

- ***Entropy***

$$f_4 = - \sum_i \sum_j p(i,j) \log(p(i,j)) .$$

Thus, using the above mentioned four cooccurrence matrices we have obtained 16 features describing spatial distribution in each window.

2.2 The run-length encoding descriptor

The run length matrix P with elements $p(i,j)$, where the (i)th dimension corresponds to the gray level and has a length equal to the maximum gray level n, while the (j)th dimension corresponds to the run length and has length equal to the maximum run length l, represents the frequency that (j) points with a gray level (i) continue in the direction q [12]. As with the cooccurrence matrix, q = 0ï , 45ï , 90ï and 135ï offer the greatest interest. Five features can be calculated from the run length matrix as shown in the equations below:

- ***Long Run Emphasis***

$$r_1 = \frac{\sum_i \sum_j j^2 p(i,j)}{\sum_i \sum_j p(i,j)} ,$$

- ***Short Run Emphasis***

$$r_2 = \frac{\sum_i \sum_j \frac{p(i,j)}{j^2}}{\sum_i \sum_j p(i,j)} ,$$

- ***Gray Level Nonuniformity***

$$r_3 = \frac{\sum_i \left[\sum_j p(i,j) \right]^2}{\sum_i \sum_j p(i,j)} ,$$

- ***Run Length Nonuniformity***

$$r_4 = \frac{\sum_j \left[\sum_i p(i,j) \right]^2}{\sum_i \sum_j p(i,j)} ,$$

- ***Run Percentage***

$$r_5 = \frac{\sum_i \sum_j j^2 p(i,j)}{N^2} ,$$

where N^2 is the number of points in the image.

The run lengths are expected large for coarse textures, especially structural textures, but can be quite small for fine textures. The nonuniformity features are small, when the gray levels or the run lengths are similar throughout the matrix, while the long run length is large when there is high intensity clustering in the texture.

2.3 The fractal dimension

The fractal dimension is a feature that characterizes the roughness of an image [8]. A well-known method for evaluating the fractal dimension is a variation of the well-known box-counting procedure, which is efficient and accurate for texture classification tasks [11].

Following this approach, the grey-level image is considered as a 3-dimensional space (x,y,z), with (x,y) denoting a 2-dimensional location, and (z) denoting the grey level. This 3-dimensional space is partitioned into cubes of size $r \times r \times r$. The position of the columns of the cubes, vertical to the (x,y) pixel plane is assigned as (i,j), where

$$(i,j) = (x/r, y/r),$$

and the boxes are enumerated from bottom to top. In every column (i,j) the cubes k and l which contain the minimum and maximum grey levels of the column, respectively, are found.

The fractal dimension D is estimated through the least mean square linear fit of $\log(N_r)$ against $\log(1/r)$ for different values of r:

$$D = \frac{\log(N_r)}{\log(1/r)}.$$

The number N_r is computed as

$$N_r = \sum_{i,j} n_r(i,j),$$

where $n_r(i,j) = l - k + 1$.

It is possible that two images of different texture and different optical appearance have the same fractal dimension. Thus, the discrimination capability of the fractal dimension, in some cases, is problematic. In order to alleviate this problem, the fractal dimension has been computed in the original subimage, as well as in the first two lower resolution versions of the original subimage and the first two sets of detail subimages, containing higher horizontal and vertical frequency spectral information. Decomposing the original image through the dyadic wavelet transform [6] has produced the subimages.

The above feature extraction procedure is originally proposed in [4]. Following this technique, seven-dimensional training patterns can be created from each image region.

3 A Novel DWT Textural Descriptor

The problem of texture discrimination, aiming at labeling image areas, is considered in the wavelet domain, since it has been demonstrated that discrete wavelet transform (DWT) can lead to better texture modeling [13]. We use the popular 2-D DWT schemes [6][14] performing a one-level wavelet decomposition of the image regions, thus resulting in four wavelet channels. Concerning the wavelet decomposition of the image regions, among the one approximate and the three detail wavelet channels 2, 3, 4 (frequency index) we have selected for further processing only the three detail channels, whose variances are the largest, since they might carry more information than the approximate one.

A more sophisticated approach is proposed by applying cooccurrence analysis to the three detail wavelet channels and extracting $3 \times 16 = 48$ relevant measures [13][14].

3.1 The wavelet transform

Wavelets offer a general mathematical approach for hierarchical function decomposition. According to this transformation, a function, which can be a function representing an image, a curve, signal etc., can be described in terms of a coarse level in addition with details that range from broad to narrow scales.

Wavelets offer an novel technique for computing the levels of detail present, under a framework that is based on a chain of *approximation* vector spaces $\{V_j \subset L^2(\Re^2),\ j \in Z\}$ and a *scaling* function f such that the set of functions $\{2^{-j/2} f(2^{-j}t-k) : k \in Z\}$ form an orthonormal basis for V_j. These two components introduce a mathematical framework presented by Mallat [6] and called multiresolution analysis.

A MultiResolution Analysis (MRA) scheme of $L^2(\Re^2)$ can be defined as a sequence of closed subspaces $\{V_j \subset L^2(\Re^2),\ j \in Z\}$ satisfying the following properties :

Containment: $V_j \subset V_{j-1} \subset L^2$; for all $j \in Z$.

Decrease: $\lim_{j\to\infty} V_j = 0$, i.e. $\bigcap_{j>N} V_j = \varnothing$, for all N.

Increase: $\lim_{j\to-\infty} V_j = L^2$, i.e. $\bigcup_{j<N} V_j = L^2$, for all N.

Dilation: $u(2t) \in V_{(j-1)} \Leftrightarrow u(t) \in V_j$.

Generator: There is a function $f \in V_0$ whose translation $\{f(t-k : k \in Z)\}$ forms a basis for V_0 .

By defining *complementary subspaces* $W_j = V_{j-1} - V_j$, so that $V_{j-1} = V_j + W_j$ then we can write, according to the "*increase*" property that

$$L^2(\Re^2) = \sum_{j \in Z} W_j \qquad (1)$$

The subspaces W_j are called *wavelet subspaces* and contain the difference in signal information between the two spaces V_j and V_{j-1}. These sets contribute to a *wavelet decomposition* of L^2 according to Eq.(1).

3.2 Selecting feature descriptors in the wavelet domain

The problem of texture classification, aiming at discriminating among various texture classes, is considered in both the time and the wavelet domain, since it has been demonstrated that discrete wavelet transform (DWT) can lead to better texture modelling [7]. In this way, we can better exploit the well known local information extraction properties of the wavelet signal decomposition as well as the features of the wavelet denoising procedures [10]. It is expected that this kind of information, considered in the wavelet domain, should be smooth due to the time-frequency localization properties of the wavelet transform. It is interesting, that only the 2-D Haar wavelet transform, which is considered as a simple one compared with the other wavelet bases, has exhibited the expected and desired properties. We have performed a one-level wavelet decomposition of the images, thus resulting in four wavelet channels.

As already mentioned, among the three channels 2, 3, 4 (frequency index) the one whose histogram presents the maximum variance, which is the channel that represents the most clear appearance of the changes between the different textures, has been selected for further processing.

The subsequent step in the proposed methodology is to obtain image windows from the selected wavelet channel and the original image of dimensions M×M and 2M×2M respectively. Feature extraction is conducted by using the information that comes from the cooccurrence matrices [2]. Among the 14 statistical measures, originally proposed by Haralick [3], that are derived from each cooccurrence matrix we have considered only four of them: angular second moment, correlation, inverse difference moment and entropy. These measures, as experiments indicated, provide high discrimination accuracy that can be only marginally increased by adding more measures in the feature vector. Using the above mentioned cooccurrence matrices 16 features describing spatial distribution in each window in the wavelet domain have been obtained. For each window in the image of the selected wavelet channel, a feature vector containing 16 features that uniquely characterizes it in the wavelet domain has been formed. For each such window a set of four features has been obtained by calculating the above four mentioned statistical measures. Finally, these 48 dimensional feature vectors form the input vector of the neural classifier.

4 Comparative Experimental Study

A total of 12 Brodatz texture images [1] of size 512×512 has been used. They are shown in Figure 1. From each texture image 10 subimages of size 256×256, with 256 gray levels depth, were randomly selected, and the feature extraction methods described in Sections 2 and 3 have been applied. For each feature extraction method 30 simulation runs have been performed using FNNs with 5 to 60 neurons in the hidden layer in order to find the architecture with the best average generalization capability. The best available architecture for each case is exhibited in Table 1.

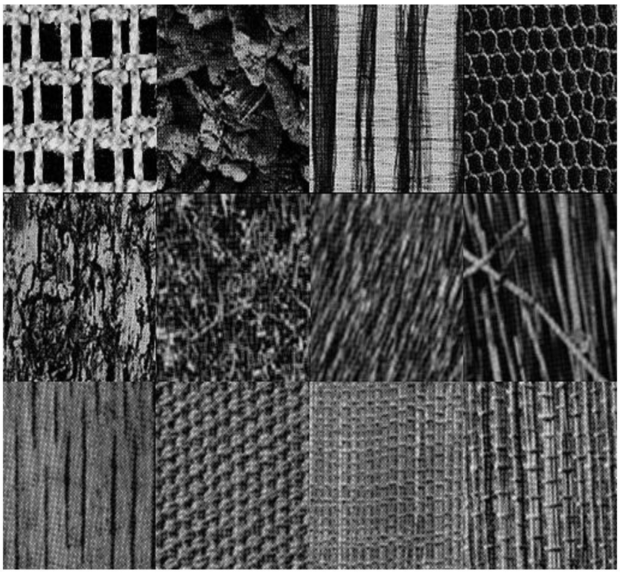

Figure 1. Twelve texture patterns obtained by digitizing images found in the "Brodatz Album". Textures: 20, 5, 51, 3, 12, 9, 93, 15, 68, 77, 78, 79.

For example, an FNN with 48 input neurons, 30 hidden and 12 output neurons with sigmoid activation function and bias exhibited the best performance for the DWT distribution estimation method.

Textural Descriptor	**FNN**
DWT distribution estimation	48-30-12
Fractal dimension	7-7-12
Cooccurrence analysis	16-40-12
Run length moments	5-18-12

Table 1. The best available FNN architectures.

The simulations have been performed using three batch training algorithms: the standard back-propagation (BP) [9], the momentum back-propagation (MBP) [9] and the back-propagation with variable stepsize (BPVS) [5]. Training terminated when the classification error was less than 3%.

Details on the performance of the algorithms are presented in Table 2. The first line in every row contains the average number of gradient evaluations and the second one the average number of error function evaluations needed to reach the termination condition. It is worth noting that in practice the computational cost of a gradient evaluation is considered at least three times more than the cost of an error function evaluation. 100% of success has been achieved for the three algorithms in the training phase.

	DWT	**Fractal dimension**	**Cooccurrence**	**Run length**
BP	1054 1054	688889 688889	924 924	4906 4906
MBP	1067 1067	596430 596430	9236 9236	5006 5006
BPVS	262 386	23597 37678	677 1008	265 388

Table 2. Average of gradient (first line in every row) and error function evaluations (second line in every row).

The generalization capability of the 30 FNNs has been tested using patterns from 20 subimages of size 256×256 randomly selected from each image. As shown in Figure 2 the three training algorithms exhibited the best generalization performance when trained with DWT extracted patterns. In this case, an average performance of 99% has been reached. Note that the average performance of BPVS trained FNNs in all cases was the best.

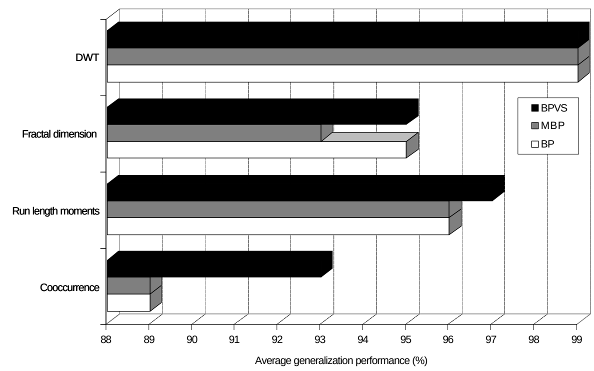

Figure 2. Average generalization performance of the trained FNNs.

Detailed generalization results for the BPVS trained FNNs are exhibited in Figure 3. As shown by the number of misclassified test patterns out of a set of 240 patterns from each feature extraction method, the FNNs that have been trained with DWT distribution estimation patterns had better generalization capability than all the others.

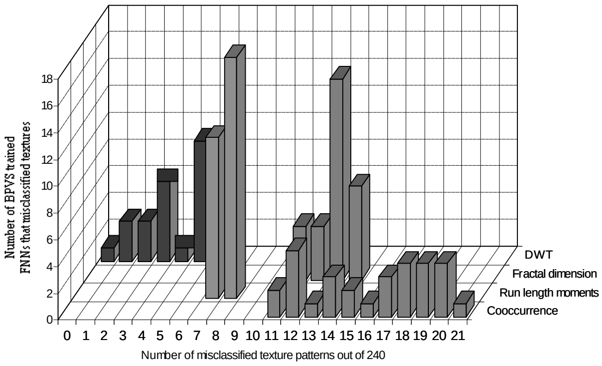

Figure 3. Number of BPVS trained FNNs with respect to their corresponding number of misclassified test patterns for the four texture extraction methods.

For example, 7 FNNs trained with DWT distribution estimation patterns misclassified only 6 test patterns out of 240. On the other hand, 18 FNNs trained with Run length patterns misclassified 8 test patterns out of 240. Note that one FNN trained with DWT distribution estimation patterns achieved 100% classification success, i.e. it exhibited 0 misclassifications.

5. Conclusions

A novel DWT distribution estimation technique has been suggested for texture description. This method, along with three other well known feature extraction techniques, have been comparatively investigated in terms of their effects on the training and generalization performance of the neural network component of a texture classification scheme. The preliminary results indicate that the proposed approach is considerably reliable for demanding applications.

References:

[1] Brodatz P., *Textures-a photographic album for artists and designer*, Dover, 1966.

[2] Haralick R.M., Shanmugam K. and Dinstein I., Textural Features for Image Classification, *IEEE Trans. Systems, Man and Cybernetics*, 3, 6, 1973, 610-621.

[3] Haralick R.M., Statistical and structural approaches to texture, *IEEE Proc.*, 67, 1979, 786-804.

[4] Karayiannis Y.A. and Stouraitis T., Texture classification using the fractal dimension as computed in a wavelet decomposed image, in *Proc. IEEE Work. Nonlin. Sign. & Image Proc.*, Greece, 186-189, 1995.

[5] Magoulas G.D., Vrahatis M.N. and Androulakis G.S., Backpropagation training with variable stepsize, *Neural Networks*, 10, 1997, 69-82.

[6] Mallat S. and Zhong S., Characterization of signals from multiscale edges, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14, 1992, 710-732.

[7] Meyer Y., *Wavelets: Algorithms and Applications*, Philadelphia: SIAM, 1993.

[8] Pentland A.P., Fractal-based description of natural scenes, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6, 1984, 661-674.

[9] Rumelhart D.E., Hinton G.E. and Williams R.J., Learning Internal Representations by Error Propagation, in D. E. Rumelhart, and J. L. McClelland (eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, 1, 318-362. MIT Press, Cambridge, MA, 1986.

[10] Ryan T.W., Sanders D., Fisher, H.D. and Iverson, A.E., Image Compression by Texture Modeling in the Wavelet Domain, *IEEE Trans. Image Processing*, 5, 1, 1996, 26-36.

[11] Sarkar N., Chaudhuri B.B., An Efficient Approach to Estimate Fractal Dimension of Textural Images, *Pattern Recognition*, 25, 9, 1992, 1035-1041.

[12] Siew, L.H., Hodgson R.M. and Wood E.J., Texture measures for carpet wear assessment, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 10, 1988, 92-105.

[13] Unser M., Texture Classification and Segmentation Using Wavelet Frames, *IEEE Trans. Image Processing*, 4, 11, 1995, 1549-1560.

[14] Wickerhauser M.V., *Adapted Wavelet Analysis from Theory to Software*, IEEE Press, 1994.