

Evaluation of Texture-based Schemes in Neural Classifiers Training

G.D. MAGOULAS¹, S.A. KARKANIS², D.A. KARRAS³ and M.N. VRAHATIS⁴

⁽¹⁾ Birkbeck College, University of London, WC1E 7HX, United Kingdom.

⁽²⁾ Technological Educational Institution of Lamia, GR-35100 Lamia, Greece.

⁽³⁾ Technological Educational Institution of Halkida, GR-34400 Evoia Island, Greece

⁽⁴⁾ Department of Mathematics, University of Patras, GR-26110 Patras, Greece

Abstract: In this paper different schemes for the extraction of textural features contained in images are examined with respect to their influence on the training and testing performance of feedforward neural networks. Moreover a novel DWT-based scheme, which estimates the features from second-order statistics of the wavelet transform of the image, is comparatively evaluated. It is demonstrated that the new scheme leads to the design and selection of feedforward neural network architectures with the best texture classification accuracy.

Key-words: Feature extraction, texture classification, wavelets, cooccurrence matrices, gray level run length moments, fractal dimension, statistical texture analysis, backpropagation neural networks, generalization.

1 Introduction

Texture plays an important role in numerous applications related to recognition using information from images. The description of the texture is based on a number of measurements evaluated on a transformation of the examined image, or image region. These measurements define the descriptors of the texture, which are further used for defining the feature vector to be used for recognition. This kind of information has been used in different application with great success.

Techniques that are 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, [3][4][8][13].

The emergence of the 2-D wavelet transform[11][14][16] 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 an image providing richer problem specific-information than other methods.

In this paper, several feature extraction techniques for texture classification of images are compared by examining the discrimination abilities of their textural descriptors. 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 development

and use of a novel wavelet-based textural descriptor for a classification scheme based on Feedforward Neural Networks (FNNs) and on the investigation of the effect of different textural descriptors on FNNs learning and generalization capabilities.

The next section presents the different texture-based schemes used in this paper. Then the wavelet-based scheme is described. This is followed by an experimental study of the various textural descriptors. The papers ends with some conclusions and future work.

2 Extracting Textural Information

Texture carries information about the microstructure of the regions and the distribution of the gray levels. A scheme for the recognition of regions based on textural information should be capable of encoding the properties of the texture using a number of parameters, named descriptors. These descriptors are usually represented by sets of statistical measures defining vectors that are used, consequently, for the recognition process.

The approach followed has two major processing stages. The first stage consists of all the processing procedures that will be performed on an image to extract all the identifiable features, which will form the feature vectors. To this end, one usually chooses a family of texture attributes that correspond to the components of the feature vectors and account for the main spatial relations between the gray levels of the texture. The second processing stage, in our case this is realized by a FNN classifier, decides how to incorporate in one body the information obtained from the first stage together with background and prior information, such as temporal data,

relationships about features, etc., in order to draw inferences. Below, three widely known feature extraction methods are briefly described.

2.1 The cooccurrence matrices

Cooccurrence Matrices (CM), [3], represent the spatial distribution and the dependence of the gray levels within a local area. Each $p(i,j)$ entry of the matrices, represents the probability of going from one pixel with a gray level (i) to another with a gray level (j) under a predefined distance and angle. From these matrices several sets of statistical measures, or feature vectors, are computed 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 using the four statistical measures, [3][4], shown in Table 1, where N_g is the number of gray levels, μ_x , μ_y are the marginal mean values of x (along the horizontal pixel axis) and y (along the vertical pixel axis), respectively, and σ_x , σ_y are the corresponding standard deviations. Thus, a set of 16 features for each window is obtained.

2.2 The run-length encoding descriptor

The Run Length Matrix (RLM) 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 gray level (i) continue in the direction q [13]. As with the cooccurrence matrix, $q = 0^\circ, 45^\circ, 90^\circ$ and 135° offer the greatest interest. Five features can be calculated from the run length matrix as shown in Table 1, where N^2 denotes 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 (FD) is a feature that characterizes the roughness of an image [8]. An efficient and accurate method for the evaluation of the fractal dimension in texture classification tasks, which is a variation of the well-known box-counting procedure, has been introduced in [12].

Cooccurrence matrices measures		Run-length encoding measures	
Energy-Angular Second Moment:	$f_1 = \sum_i \sum_j p(i, j)^2$	Long Run Emphasis:	$r_1 = \frac{\sum_i \sum_j j^2 p(i, j)}{\sum_i \sum_j p(i, j)}$
Correlation:	$f_2 = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i * j) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$	Short Run Emphasis:	$r_2 = \frac{\sum_i \sum_j \frac{p(i, j)}{j^2}}{\sum_i \sum_j p(i, j)}$
Inverse Difference Moment:	$f_3 = \sum_i \sum_j \frac{1}{1 + (i - j)} p(i, j)$	Run Length Nonuniformity:	$r_3 = \frac{\sum_i \left[\sum_j p(i, j) \right]^2}{\sum_i \sum_j p(i, j)}$
Entropy:	$f_4 = - \sum_i \sum_j p(i, j) \log(p(i, j))$	Gray Level Nonuniformity:	$r_4 = \frac{\sum_j \left[\sum_i p(i, j) \right]^2}{\sum_i \sum_j p(i, j)}$
		Run Percentage:	$r_5 = \frac{\sum_i \sum_j j^2 p(i, j)}{N^2}$

Table 1. Cooccurrence matrices and run-length descriptors.

Following this approach, the gray-level image is considered as a 3-dimensional space (x, y, z) , with (x, y) denoting a 2-dimensional location and (z) denoting the gray 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 that contain the minimum and maximum gray 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 , $D = \log(N_r)/\log(1/r)$.

The number N_r is computed as $N_r = \sum_{i,j} n_r(i, j)$, where $n_r(i, j) = l - k + 1$. Note that in practice, it is possible that two images of different texture and different optical appearance may have the same fractal dimension. Thus, the discrimination capability of the fractal dimension is problematic in some cases. To deal with this problem the feature extraction procedure proposed in [5] has been used: 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, which contain higher horizontal and vertical frequency spectral information. The subimages have been produced by decomposing the original image through the dyadic wavelet transform [7]. This technique results in seven-dimensional training patterns for each image window.

3 A Novel Wavelet-based Textural Descriptor

The wavelet transform is a general mathematical approach for hierarchical function decomposition. A function representing an image, curve, or signal, can be described by means of this transformation in terms of a coarse level and levels of details, which range from broad to narrow scales.

More specifically, wavelets offer a novel framework for computing the levels of detail present in an application context, which is called MultiResolution Analysis (MRA) [7]. This framework is based on a chain of *approximation* vector spaces $\{V_j \subset L^2(\mathbb{R}^2), j \in \mathbb{Z}\}$, where \mathbb{Z} is the set of integers, and a *scaling* function ϕ such that the set of functions $\{2^{-j/2}\phi(2^{-j}t - k): k \in \mathbb{Z}\}$ form an orthonormal basis for V_j . An MRA scheme of $L^2(\mathbb{R}^2)$ can be defined as a sequence of closed subspaces $\{V_j \subset L^2(\mathbb{R}^2), j \in \mathbb{Z}\}$ satisfying the following properties:

1. *Containment*: $V_j \subset V_{j-1} \subset L^2$; for all $j \in \mathbb{Z}$.
2. *Decrease*: $\lim_{j \rightarrow \infty} V_j = 0$, i.e. $\bigcap_{j > N} V_j = \emptyset$, for all N , where N is the set of natural numbers.

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

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

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

By defining *complementary subspaces* $W_j = V_{j-1} - V_j$, so that $V_{j-1} = V_j + W_j$, we can write using the “*increase*” property that

$$L^2(\mathbb{R}^2) = \sum_{j \in \mathbb{Z}} W_j. \tag{1}$$

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 Relation (1).

The special case of the Discrete Wavelet Transform (DWT) has been suggested by various researchers as a suitable approach to perform efficient texture modeling [14]. By considering the problem of texture classification in the wavelet domain, 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 procedure [11]. It is expected that this kind of information, considered in the wavelet domain, should be smooth enough due to the time-frequency localization properties of the wavelet transform.

In our experiments, we have used the popular 2-D DWT schemes [7][16] performing a one-level wavelet decomposition of the image regions. It is worth noting that 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. Thus, in the first stage of the proposed methodology a one-level wavelet decomposition of the images has been performed resulting in four wavelet channels: the *approximate* one and the three *detail* wavelet channels 2, 3, 4 (frequency index).

In the second stage feature extraction has been conducted by using the information that comes from the cooccurrence matrices [3], i.e. angular second moment, correlation, inverse difference moment and entropy, on the detail channels in order to create a more reliable framework for the generation of the textural descriptors. Hence, a set of four features has been obtained for each window by calculating these statistical measures as described in Section 2.1. These four measures provide high discrimination accuracy. In this way, a feature vector containing 16 features that uniquely characterizes in the wavelet domain each image window of the selected wavelet channel has been formed. This procedure applied to

the three detail channels has resulted in the extraction of $3 \times 16 = 48$ relevant measures. A mathematical description of the novel DWT-based feature extraction procedure is given below:

Step 1: After splitting the image $f(x,y)$ by the sliding window split process (this is common to all methodologies herein compared), an original image window $f_w(x,y)$ is transformed into its four wavelet bands, denoted as ws_1, ws_2, ws_3, ws_4 , by performing a one-level decomposition through the use of the 2-D DWT. The well known formula of the 1-D DWT decomposition of a signal $f(t)$

$$f(t) = \sum_k c_{j_0}(k) \phi_{j_0,k}(t) + \sum_k \sum_{j=j_0}^J d_j(k) \psi_{j,k}(t),$$

can be reformulated as

$$f(t) = \sum_k c_1(k) \phi_{1,k}(t) + \sum_k d_1(k) \psi_{1,k}(t),$$

where the choice of j_0 sets the coarsest scale whose space is spanned by the wavelet basis $\phi_{j_0,k}(t)$, $J-j_0$ defines the decomposition level of the transform as desired by having $J-j_0$ stages, $\psi_{j,k}(t)$ is the wavelet basis that provides the high resolution details of the signal at scale j , and, finally, $c_{j_0}(k)$ and $d_j(k)$ are the wavelet coefficients (coarse and detail ones) that are calculated using the formulae

$$c_{j_0}(k) = \langle f(t), \phi_{j_0,k}(t) \rangle \text{ and } d_j(k) = \langle f(t), \psi_{j,k}(t) \rangle.$$

We, then, proceed defining the one-level 2-D DWT of an image window $f_w(x,y)$ into its four wavelet bands ws_1, ws_2, ws_3, ws_4 , which will be denoted as $ws_{1,1}, ws_{2,1}, ws_{3,1}, ws_{4,1}$, respectively. But first, let us assume that a 2-D separable scaling function $\Phi^{(1)}(x,y) = \Phi(x)\Phi(y)$ is involved as kernel function in the decomposition. Then we have

$$\begin{aligned} f_w(x,y) = & \sum_{k,m} ws_{1,1}(k,m) \phi_{1,k}(x) \phi_{1,m}(y) \\ & + \sum_{k,m} ws_{2,1}(k,m) \phi_{1,k}(x) \psi_{1,m}(y) \\ & + \sum_{k,m} ws_{3,1}(k,m) \psi_{1,k}(x) \phi_{1,m}(y) \\ & + \sum_{k,m} ws_{4,1}(k,m) \psi_{1,k}(x) \psi_{1,m}(y), \end{aligned}$$

where k and m are indices that run over the elements of the matrices $ws_{q,1}$, $q = 1, 2, 3, 4$. As in the 1-D case, the following formulae are used to obtain the wavelet coefficients $ws_1(k,m), ws_2(k,m), ws_3(k,m)$ and $ws_4(k,m)$

$$\begin{aligned} ws_{1,1}(k,m) &= \langle f_w(x,y), \phi_{1,k}(x), \phi_{1,m}(y) \rangle, \\ ws_{2,1}(k,m) &= \langle f_w(x,y), \phi_{1,k}(x), \psi_{1,m}(y) \rangle, \\ ws_{3,1}(k,m) &= \langle f_w(x,y), \psi_{1,k}(x), \phi_{1,m}(y) \rangle, \\ ws_{4,1}(k,m) &= \langle f_w(x,y), \psi_{1,k}(x), \psi_{1,m}(y) \rangle. \end{aligned}$$

Step 2: Transform the detail wavelet coefficients $ws_2(k,m), ws_3(k,m)$ and $ws_4(k,m)$ resulted from Step 1 into integers using the following formula:

$$intws_q = N_g - \delta * \sqrt{|ws_q|}, \quad q = 2, 3, 4,$$

where N_g is the number of gray levels and δ is a heuristic parameter, [2]. In the experiments presented in the next section we have used the values $N_g = 256$ and $\delta = 3.8$. Thus, this step results in transforming the bands ws_2, ws_3 , and ws_4 into $intws_2, intws_3$, and $intws_4$, respectively.

Step 3: For each one of the transformed wavelet band $intws_2, intws_3, intws_4$ consider its corresponding cooccurrence matrices $M_k(intws_q)$, for $k = 1, 2, 3, 4$ and $q = 2, 3, 4$, and extract the corresponding features.

4 Experimental Study

In the first experiment, a total of 12 Brodatz texture images, [1], of size 512×512 has been used, as shown in Figure 2. From each texture image 10 subimages of size 256×256 , with 256 gray levels depth, have been randomly selected, and the feature extraction methods described in Sections 2 and 3 have been applied.

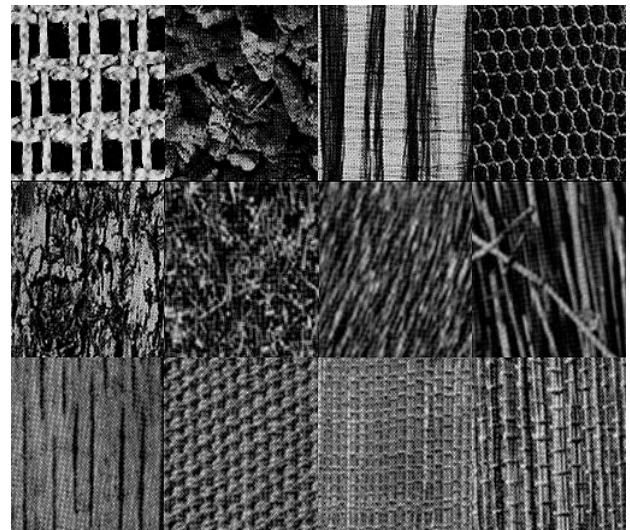


Figure 2. The twelve Brodatz textures.

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 2. For example, an FNN with 48 input neurons, 30 hidden and 12 output neurons with sigmoid activation functions and biases exhibited the best performance for the DWT-based method.

Textural Descriptor	FNN
DWT distribution estimation	48-30-12
Fractal dimension-FD	7-7-12
Cooccurrence matrices-CM	16-40-12
Run length moments-RLM	5-18-12

Table 2. The best available FNN architecture.

Simulations have been performed using five batch training algorithms: the standard Back-Propagation (BP) [10], the Momentum BP (MBP) [10], the Adaptive BP (ABP) [15], the Rprop algorithm [9] and the BP with Variable Stepsize (BPVS) [6]. The termination condition was a classification error less than 3%.

Details on the average performance of the algorithms over the 30 best runs are presented in Table 3 in terms of the average number of gradient and error function evaluations needed to reach the termination condition. Note 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 all algorithms in the training phase.

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 3 the five training algorithms exhibited the best generalization performance when trained with DWT extracted patterns. Especially, BP, MBP, ABP and BPVS reached an average performance of 99%.

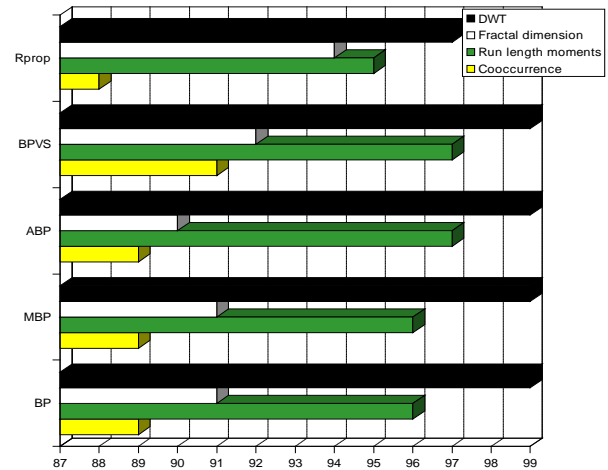
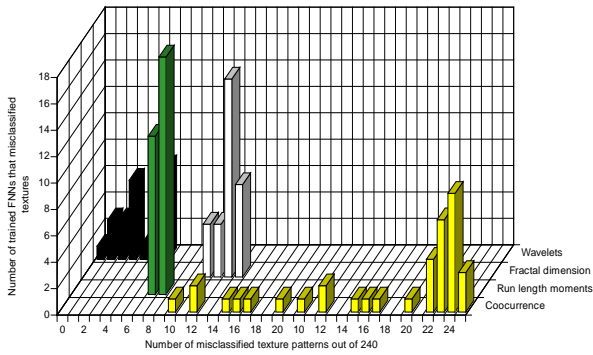


Figure 3. Average performance in the generalization.

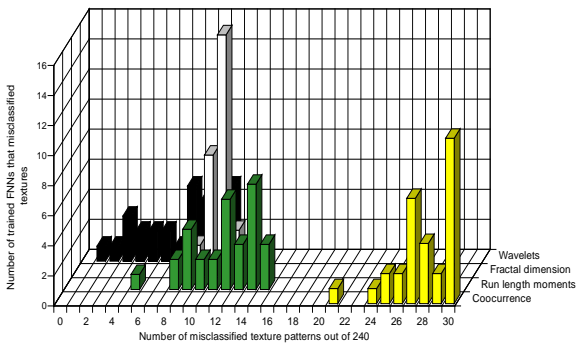
Detailed generalization results for the BPVS trained FNNs are exhibited in Figure 4 (left). As shown by the number of misclassified test patterns out of a set of 240 patterns from each feature extraction method, the FNNs that have been trained with DWT-based patterns had better generalization capability than all the others. For example, 7 FNNs trained with DWT-based 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-based patterns achieved 100% classification success, i.e. it exhibited 0 misclassifications. Relative results are exhibited in Figure 4 (right) for the Rprop trained FNNs. In both cases, DWT patterns provided the better generalization capability than all other extraction methods tested.

	DWT (GRD/EFE)	FD (GRD/EFE)	CM (GRD/EFE)	RLM (GRD/EFE)
BP	1054/1054	688889/688889	924/924	4906/4906
MBP	1067/1067	596430/596430	9236/9236	5006/5006
ABP	327/327	407500/407500	531/531	390/390
Rprop	72/72	469297/469297	172/172	180/180
BPVS	262/386	23597/37678	677/1008	265/388

Table 3. Average number of gradient (GRD)/error function evaluations (EFE) for each method tested.



(a)



(b)

Figure 4. (a) BPVS and (b) Rprop trained FNNs with respect to their number of misclassifications

In order to investigate the effect of the FNN architecture on the effectiveness of the image recognition scheme, we conducted a second set of experiments using 6 different FNN architectures with 5, 10, 15, 20, 25 and 30 hidden nodes, respectively, which were trained using the BPVS from 100 different initial weight sets.

In this case we have used wavelet-extracted pattern, since FNNs trained with wavelet patterns have exhibited the best performance in our tests with texture images. Colonoscopic images from two different colons were used (see Figure 5), and 200 patterns have been used for training and 400 for testing. The results are illustrated in Figures 6-8.

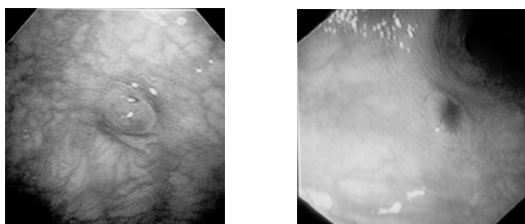


Figure 5. Colonoscopic images: Image 1 on the left, and Image 2, on the right.

In Figure 6, the generalization performance of the trained networks is shown with respect to the number of hidden nodes. FNNs with 15 hidden nodes exhibit the best average generalization capability, denoted by a triangle in the figure. Also, the network that exhibits the best performance overall (96.5%) is based on 15 hidden nodes.

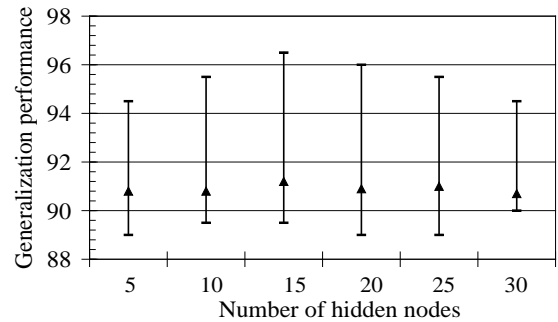


Figure 6. Generalization performance vs. number of hidden nodes.

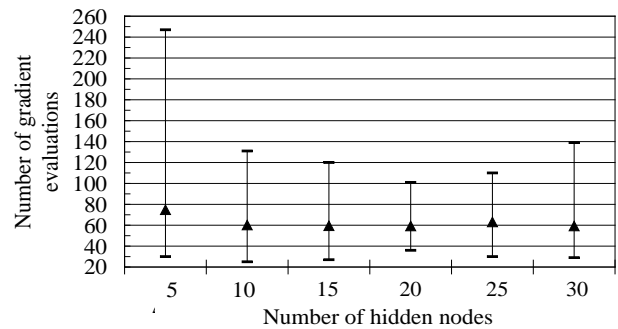


Figure 7. Number of gradient evaluations vs. number of hidden nodes.

In Figure 7, the number of gradient evaluations with respect to the number of hidden nodes is illustrated. Networks with 10 to 20 hidden nodes exhibit the best average number of gradient evaluations (60 gradient evaluations). While the worse case performance was observed with a 5 hidden node FNN (247 gradient evaluations).

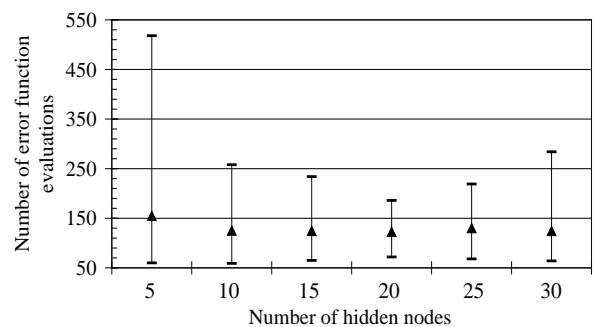


Figure 8. Number of error function evaluations vs. number of hidden nodes.

In Figure 8, networks with 10 to 20 hidden nodes exhibit almost the same average performance with respect to the average number of error function evaluations to train the FNNs. The best performance (59 error function evaluations) was obtained with a 10 hidden nodes architecture.

5 Conclusions

A novel DWT-based technique has been suggested for texture description. The proposed scheme estimates the features from second-order statistics of the wavelet transform of the image. This method, along with three other well known feature extraction techniques have been tested in terms of their effects on the training and generalization performance of the neural network component of a texture classification scheme. The recognition task was executed using various neural network architectures in order to determine the most suitable one for similar applications. The preliminary results indicate that the proposed approach is considerably reliable for demanding applications. Future research will be directed to the examination of the recognition performance of the proposed scheme on medical images acquired by different sources.

References

- [1] Brodatz P., *Textures-A photographic album for artists and designer*, Dover, 1966.
- [2] Buckheit J.B., Chen S., Donoho D.L., Johnstone I.M., Scargle J.D., *Wavelab Reference Manual*, 1995. Available at <http://www-stat.stanford.edu/~wavelab/ftp/WaveLabRef.ps>
- [3] Haralick R.M., Shanmugam K. and Dinstein I., Textural Features for Image Classification, *IEEE Trans. Systems, Man and Cybernetics*, 3, 6, 1973, 610-621.
- [4] Haralick R.M., Statistical and structural approaches to texture, *IEEE Proc.*, 67, 1979, 786-804.
- [5] 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.
- [6] Magoulas G.D., Vrahatis M.N. and Androulakis G.S., Backpropagation training with variable stepsize, *Neural Networks*, 10, 1997, 69-82.
- [7] Mallat S. and Zhong S., Characterization of signals from multiscale edges, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14, 1992, 710-732.
- [8] Pentland A.P., Fractal-based description of natural scenes, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6, 1984, 661-674.
- [9] Riedmiller M., and Braun H., A direct adaptive method for faster backpropagation learning: the Rprop algorithm, in: *Proc. IEEE Int. Conf. Neural Net.*, San Francisco, 586-591, 1993.
- [10] 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.
- [11] 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.
- [12] Sarkar N., Chaudhuri B.B., An Efficient Approach to Estimate Fractal Dimension of Textural Images, *Pattern Recognition*, 25, 9, 1992, 1035-1041.
- [13] 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.
- [14] Unser M., Texture Classification and Segmentation Using Wavelet Frames, *IEEE Trans. Image Processing*, 4, 11, 1995, 1549-1560.
- [15] Vogl T., Mangis J., Rigler J., Zink W., and Alkon D., Accelerating the convergence of the backpropagation method, *Biological Cybernetics*, 59, 1988, 257-263.
- [16] Wickerhauser M.V., *Adapted Wavelet Analysis from Theory to Software*, IEEE Press, 1994.