Neural Network based textural labeling of images in multimedia applications

S.A. Karkanis[+], G.D. Magoulas[+], and D.A. Karras[++]

[+] University of Athens, Dept. of Informatics, Typa Build., Panepistimiopolis,

GR 15784, Athens Greece, {sk|magoulas}@di.uoa.gr

[++] University of Piraeus, Dept. of Business Administration, dakarras@www.hotmail.com

Abstract

In this contribution is investigated the use of multilayer perceptron type neural networks in the characterization of images by texture content. The paper is focused on the effects of textural feature extraction methods on the network architecture, training performance and generalization capability when applied in indexing of images contained within multimedia image databases. 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 proposed technique leads to the design and selection of multilayer perceptron architectures with the best texture classification accuracy.

Keywords: *multimedia image databases, texture description, features extraction, neural networks.*

1. INTRODUCTION

An important problem in the development of multimedia systems is the design of complex image content seeking query mechanisms on large image databases. Understanding image characteristics and describing these images involving such characteristics requires significant effort in order to depict an image in terms of a symbolic representation that best matches its information content in the multimedia database. Such a symbolic representation, i.e. a string, may become the corresponding image index in a complex query mechanism.

This paper deals with the design of image indices by labeling the corresponding regions in terms of their second order characteristics and, more specifically, texture. The proposed index design scheme is simple: an image is divided in rectangular regions of predefined dimensions and each one is labeled according to its textural content. Textural classification may play a significant role in the solution of the image-indexing problem in multimedia applications.

A texture-based image indexing method is usually composed of three stages. The first stage aims at the description of the texture and at the extraction of efficient textural descriptors. The second stage is devoted to the division of the image in regions. Dividing an image into overlapping or non-overlapping square regions of equal dimensions is the simplest technique attaining reasonable results. The last stage consists of classification and labeling of these regions in terms of their textural content using statistical pattern recognition techniques.

This paper proposes a new descriptor for texture classification based on measures obtained from the detail coefficients of the Discrete Wavelet Transform (DWT). An image division scheme into regions of square non-overlapping windows of equal dimensions has been adopted. Multilayer Perceptron (MLP) type neural networks have been involved in the image indexing approach to classify and label the textural content of each window.

The contribution of this paper lies not only on the use of a novel texture descriptor and the application of MLPs in this image indexing task for multimedia applications, but mainly on investigating the effects of different textural descriptors on MLP learning and generalization capabilities regarding such a task. The experimental study conducted in this paper aims, precisely, at illustrating this latter investigation.

2. TEXTURAL FEATURE EXTRACTION TECHNIQUES INVOLVED

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

2.1 Cooccurrence analysis based texture descriptor

For each image region-window, previously described, we use the information that comes from the cooccurrence matrices [1]. These matrices represent the spatial distribution dependence of the gray levels within an 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. More matrices are formed for specific spatial distances and predefined angles. From these matrices, sets of statistical measures are computed (called feature vectors) for building different texture models. We have considered four angles, namely 0^o, 45^o, 90^o, 135^o as well as a predefined distance of one pixel in the formation of the cooccurrence matrices. Therefore, we have formed four cooccurrence matrices. Among the 14 statistical measures, originally proposed by Haralick [1,2], that are derived from each cooccurrence matrix we have considered only four. Namely, angular second moment, correlation, inverse difference moment and entropy.

- ***Energy - Angular Second Moment*** $$f1 = \sum_i \sum_j p(i,j)^2$$
- ***Correlation*** $$f2 = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i * j) p(i,j) - m_x m_y}{s_x s_y}$$
- ***Inverse Difference Moment*** $$f3 = \sum_i \sum_j \frac{1}{1+(i-j)} p(i,j)$$
- ***Entropy*** $$f4 = -\sum_i \sum_j p(i,j) \log(p(i,j))$$

We have experimentally found, that these measures, provide high discrimination accuracy which can be only marginally increased by adding more measures in the feature vector. Thus, using the above mentioned four cooccurrence matrices we have obtained 16 features describing spatial distribution in each window corresponding to a region in which an original image is divided in order to apply the proposed image indexing scheme.

2.2 The Run-length encoding texture descriptor

The run length matrix, p(*i,j*), is a method of statistical analysis, that represents the frequency that j points with a grey level *i* continue in the direction q [3]. The (*i*)th dimension of the matrix corresponds to the grey level and has a length equal to the maximum grey level, n, while the (*j*)th corresponds to the run length and has length equal to the maximum run length, l. As with the co-occurrence 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, where *A* is the area of the image

$$\text{Run Percentage} = \sum_{i=1}^{n}\sum_{j=1}^{l} p(i,j) \Big/ A$$

$$\text{Long Runs Emphasis} = \sum_{i=1}^{n}\sum_{j=1}^{l} j^2 p(i,j) \Big/ \sum_{i=1}^{n}\sum_{j=1}^{l} p(i,j)$$

$$\text{Run Length Nonuniformity} = \sum_{j=1}^{l}\left\{\sum_{i=1}^{n} p(i,j)\right\}^2 \Big/ \sum_{i=1}^{n}\sum_{j=1}^{l} p(i,j)$$

$$\text{Grey Level Nonuniformity} = \sum_{i=1}^{n}\left\{\sum_{j=1}^{l} p(i,j)\right\}^2 \Big/ \sum_{i=1}^{n}\sum_{j=1}^{l} p(i,j)$$

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

2.3 The Fractal dimension based texture descriptor

The fractal dimension is an image feature that characterizes the roughness of an image [4]. However, it is possible that two images of different texture and different optical appearance have the same fractal dimension. Thus, its discrimination capability, in some cases is problematic.

In order to alleviate this problem, the fractal dimension was 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. The subimages were produced by decomposing the original image through the dyadic wavelet transform [5]. The aforementioned feature extraction procedure is originally proposed in [6]. Following this procedure, seven-dimensional training patterns can be created from each image region.

3. A NOVEL DWT DISTRIBUTION BASED 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 [7]. We use the popular 2-D DWT schemes [8,9]. We have performed 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 select for further processing only the three detailed channels, whose variances are the largest, since they might carry more information than the approximate one. With respect to this, we should say at this point that Unser [10] has pointed out local variance of the wavelet coefficients as an appropriate measure for classifying texture. A more sophisticated approach is proposed by applying cooccurrence analysis to the three detail wavelet channels and extracting $3 \times 16 = 48$ relevant measures.

4. COMPARATIVE EXPERIMENTAL STUDY

The proposed image-labeling scheme for multimedia applications heavily depends on the accuracy of the texture classification stage. The experimental study below outlined is a preliminary evaluation of the performance of the image indexing system components associated with its labeling phase.

A total of 12 Brodatz texture images [11]: 3, 5, 9, 12, 15, 20, 51, 68, 77, 78, 79, 93 (see Figure 1) of size 512×512 has been used. From each texture image 10 subimages of size 256×256, with 256 gray levels depth, were randomly selected, and the above mentioned feature extraction techniques have been applied. The MLP generalization capability has been tested using patterns from 20 subimages of the same size randomly selected from each image.

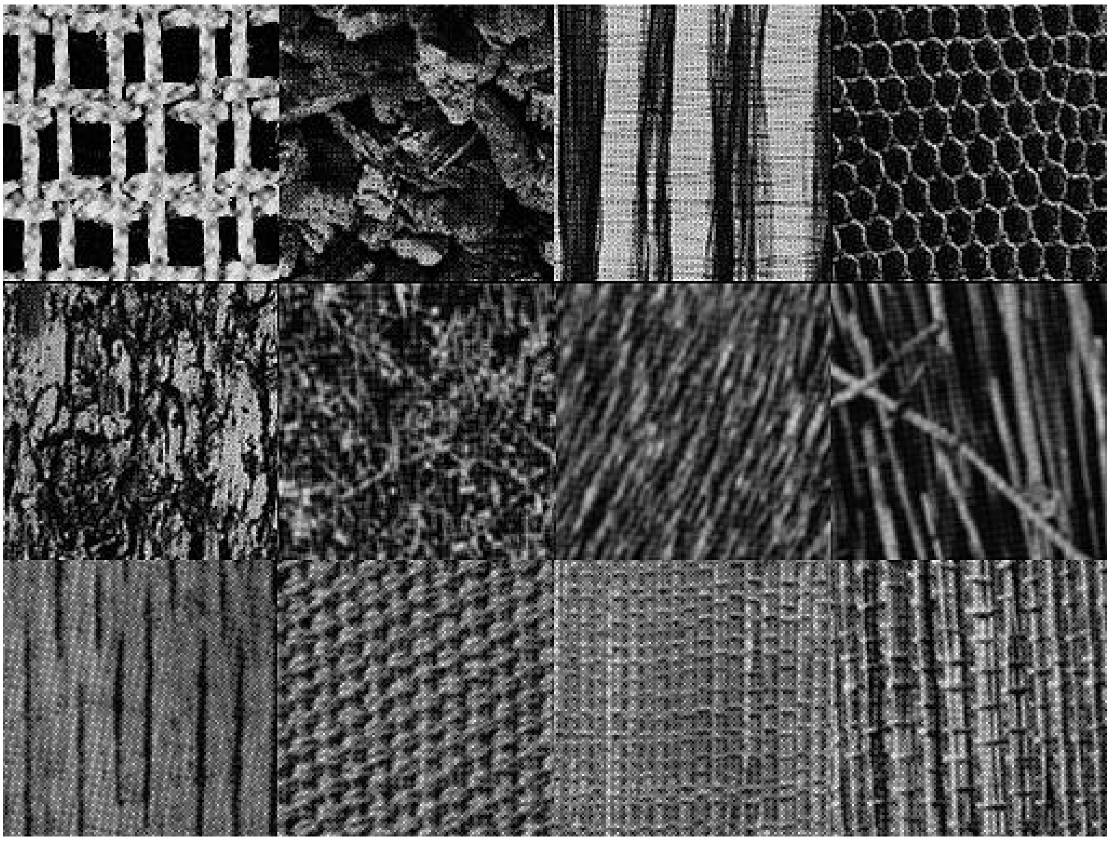

Figure 1. Twelve texture patterns obtained from digitizing images found in the "Brodatz Album". Textures: 20, 5, 51, 3, 12, 9, 93, 15, 68, 77, 78, 79.

A recently proposed learning algorithm, named BPVS [12], has been used to train the MLPs. This algorithm provided better generalization capability than other popular training algorithms when tested on texture classification problems (see [12] for relevant experiments). For each feature extraction method thirty simulation runs have been performed using MLPs with 5 to 50 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. For example, a MPL with 48 input neurons, 30 hidden and 12 output neurons with biases exhibited the best performance for the DWT distribution estimation method.

Feature extraction method	MLP architecture
DWT distribution estimation	48-30-12
Fractal dimension	7-30-12
Cooccurrence analysis	16-40-12
Gray level run length moments	5-10-12

Table 1. The best available MLP architectures.

The average generalization performance of the 30 MLPs that have been trained using DWT features was the best and reached a 99.1%. The number of misclassified test patterns out of 240 for each method is presented in Figure 2. As shown in Figure 2, the MLPs that have been trained using the DWT distribution estimation patterns had significantly better generalization capability than all the others. For example, 13 MLPs trained with DWT distribution estimation patterns misclassified only 3 test patterns out of 240. On the other hand, 15 MLPs trained with Fractal dimension patterns misclassified 13 test patterns out of 240. Note that one MLP trained with DWT distribution estimation patterns achieved 100% classification success, i.e. it exhibited 0 misclassifications.

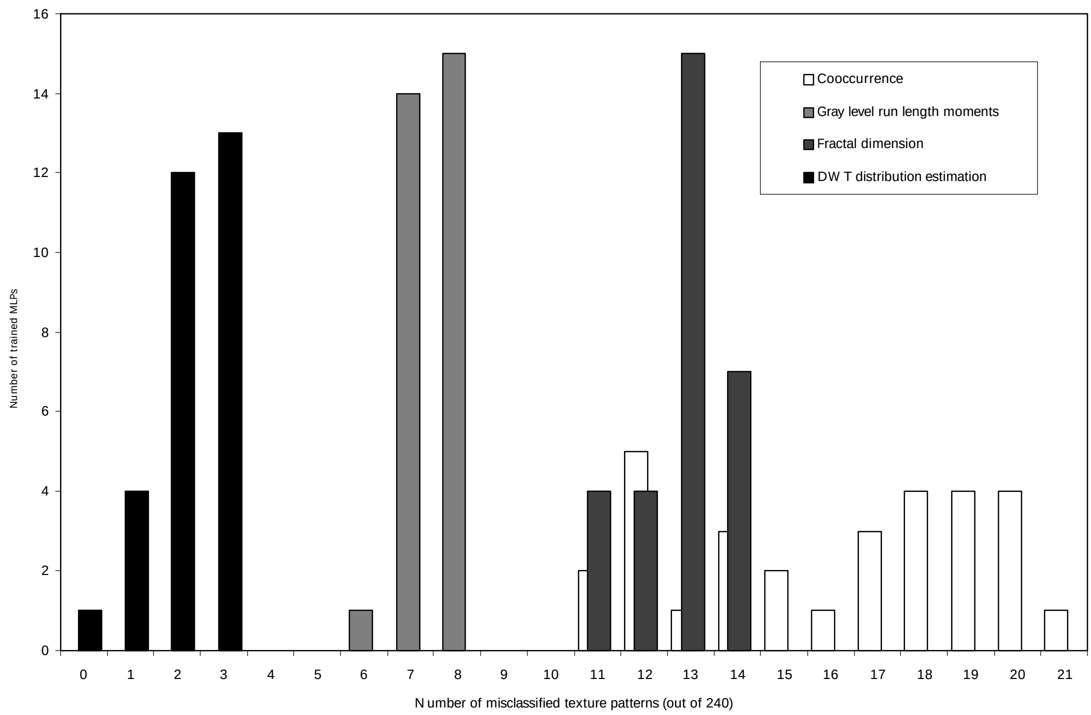

Figure 2. Number of trained MLPs with respect to their corresponding number of misclassified test patterns.

5. CONCLUSIONS

An image-indexing scheme for multimedia applications based on image textural content has been proposed and preliminary evaluated. Regarding its components, a novel DWT distribution estimation technique has been suggested for the texture description stage. This method, along with three other well known feature extraction techniques, have been comparatively investigated in terms of their effects on the generalization performance of the labeling component of the indexing system. The preliminary results indicate that the proposed approach is considerably reliable for demanding applications. Integration aspects of this research effort are under thorough investigation by the authors.

REFERENCES

1. Haralick, R. M., Shanmugam, K. and Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Trans. Systems, Man and Cybernetics*, 3, 6, 610-621.
2. Haralick, R. M. (1979). Statistical and structural approaches to texture. *IEEE Proceedings*, 67, 786-804.
3. Siew, L. H., Hodgson, R. M. & Wood, E. J. (1988). Texture Measures for Carpet Wear Assessment. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 10, 92-105.
4. Pentland, A. P. (1984). Fractal-based description of natural scenes. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 6, 661-674.
5. Mallat, S., & Zhong S. (1992). Characterization of signals from multiscale edges. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 14, 710-732.
6. Karayiannis, Y. A., & Stouraitis T. (1995). Texture classification using the fractal dimension as computed in a wavelet decomposed image. In *Proceedings of the IEEE Workshop on Nonlinear Signal and Image Processing*, Halkidiki, Greece, 186-189.
7. Ryan, T. W., Sanders, D., Fisher, H. D. and Iverson, A. E. (1996). Image Compression by Texture Modeling in the Wavelet Domain. *IEEE Trans. Image Processing*, 5, 1, 26-36.
8. Meyer, Y. (1993). *Wavelets: Algorithms and Applications*, Philadelphia: SIAM.
9. Wickerhauser, M.V. (1994). *Adapted Wavelet Analysis from Theory to Software.* IEEE Press.
10. Unser, M. (1995). Texture Classification and Segmentation Using Wavelet Frames. *IEEE trans. Image Processing*, 4, 11, 1549-1560.
11. Brodatz, P. (1966). *Textures-a photographic album for artists and designer*, Dover.
12. Magoulas G. D., Vrahatis M. N. and Androulakis G. S. (1997). Backpropagation training with variable stepsize, *Neural Networks*, 10, 69-82.