EVALUATION OF TEXTURAL FEATURE EXTRACTION SCHEMES FOR NEURAL NETWORK-BASED INTERPRETATION OF REGIONS IN MEDICAL IMAGES

S.A. Karkanis[(1)], G.D. Magoulas[(2)], D.K. Iakovidis[(1)], D.A. Karras[(3)] and D.E. Maroulis[(1)]

[(1)]University of Athens, Department of Informatics, Panepistimiopolis 15784 Athens, Greece
{sk,diakov,dmarou}@di.uoa.gr
[(2)]Department of Information Systems and Computing, Brunel University, UB8 3PH, United Kingdom
[(3)]Hellenic Aerospace Industry, 2 Rhodou, A. Ilioupolis, 16342, Athens, Greece.

ABSTRACT

A few approaches have been presented in the literature towards the discrimination of texture in medical images. Recently, medical experts proposed that the more valuable information for discriminating among normal and suspicious for cancer regions in endoscopic images is the texture of the examined tissue. Texture can be encoded by a number of mathematical descriptors. Three well-known textural descriptors, as well as a new wavelet-based one are used in this paper for an accurate study and evaluation of the methodologies encountered. Experiments conducted include tests with various images from the Brodatz album, as well as interpretation of tissue regions in endoscopic image. In all cases the recognition task is supported by multilayer perceptron type neural network architectures.

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. The recognition capability of a system is estimated by applying the proposed approach in texture reference images as well as in images related with the pertinent application [1]. In the proposed study the feature extraction approaches used are evaluated in recognition tasks in medical images and more specifically on images that come from endoscopic sessions. The final result which is also a requirement of the system is a highly rate in success of the recognition building block.

The processing of medical images is a difficult task. The formation of different tissues is variable among different people, the illumination conditions are not the same during different sessions (e.g. endoscopic ones), and the final decision of the doctors is based on subjective judgements (according to his/her experience of similar diagnostic cases). The use of textural descriptors for a final accurate treatment decision can offer valuable information. As it has already been accepted by the literature an important discrimination factor between normal and abnormal tissues is the appearance of the texture on their surface [2].

In this paper we evaluate three well-known textural descriptors, as well as a novel wavelet-based one in recognizing different texture types found in textural reference images from the Brodatz album, and in discriminating regions in endoscopic images. The results shown that the use of the novel feature extraction technique gave the best result, achieving high recognition.

The paper is organized as follows. In the next section the textural descriptors used in our experiments are briefly described. In Section 3, the experiments and discussion of the results are presented. The conclusions follow in the fourth section.

2. FEATURE EXTRACTION

Three different statistical descriptors were used in this paper, [3-5], i.e. cooccurrence matrices, run length encoding, fractal dimension and a new Discrete Wavelet Transform-based (DWT) descriptor [6]. A common issue in the above mentioned techniques, is the scheme of processing of the input image. The image is divided into regions of square non-overlapping windows of equal dimensions. Features are estimated for each image window and the set of these constitute the components of the corresponding feature vector. Multilayer Perceptron (MLP) type neural networks have been involved in the

image classification and labeling approach of each window [7].

2.1. Cooccurrence matrices

Cooccurrence matrices [3] represent the spatial distribution and the dependence of the gray levels within a local area. Based on these matrices, sets comprised of four of statistical measures are computed for four different angles, namely 0^o, 45^o, 90^o, 135^o. The measures are: energy, correlation, inverse different moment and entropy. Thus we obtain 16 features describing spatial distribution in each image-window.

2.2. Run-length encoding

The run length matrix P with elements $p(i,j)$, where i-th dimension corresponds to the gray level and has a length equal to the maximum gray level n, while 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. Five features, like long run emphasis, short run emphasis, gray level nonuniformity, run length non uniformity, run percentage, can be calculated from the run length matrix [4]. 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 if the gray 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. Fractal dimension

The fractal dimension is a feature that characterizes the roughness of an image [5]. We have used a variation of the well-known box-counting procedure for evaluating the fractal dimension, which is efficient and accurate for texture classification tasks. 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 *rxrxr*. The position of the columns of the cubes, vertical to the *(x, y)* pixel plane is assigned as *(i, j)*, where *(i, j) = (x/y, 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 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.
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, decomposed using the dyadic wavelet transform [6]. Seven features are calculated producing a seven-dimensional space created from each image region.

2.4. DWT Textural Descriptor

It has been demonstrated that discrete wavelet transform (DWT) can lead to better texture modeling [6]. DWT schemes performing a one-level wavelet decomposition of the image resulting in the four wavelet bands offers a transformation that maintains the textural properties in different scales of transformation. The three wavelet channels 2, 3, 4 (frequency index) we have selected for further processing. On each of this channel cooccurrence analysis is applied extracting by this way 3x16 = 48 relevant measures, the components of the feature vectors. The image is scanned along with a window of predefined dimensions, it is transformed by the DWT and the statistical measurements are calculated for each window creating this way the corresponding vector. The total set of the vectors is used as input to the neural network that will result to the classification of each corresponding region into one of the two classes: normal region or abnormal (possibly suspicious) region of the image.

3. EXPERIMENTS AND DISCUSSION

3.1. Texture Images

A total of 12 Brodatz [8] texture images of size 512×512 has been used (see Fig. 1). From each 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.

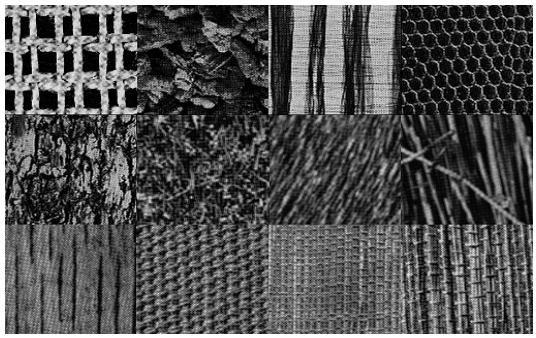

Figure 1. Twelve textures found in the "Brodatz Album".

An adaptive learning rate algorithm has been used to train the MLPs using 30 different initial weight sets, and the best available architecture for each case is exhibited in Table 1.

Feature extraction method	MLP
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. Note that one MLP trained with DWT distribution estimation patterns achieved 100% classification success, i.e. it exhibited 0 misclassifications.

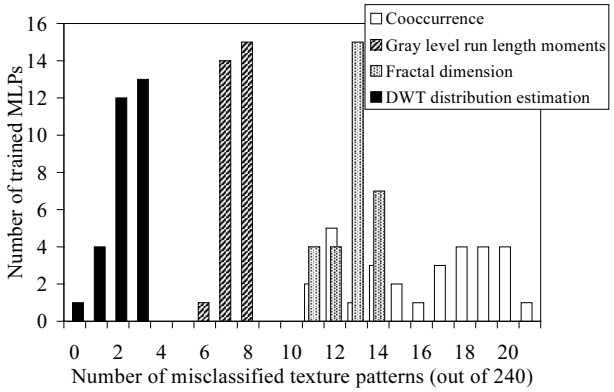

Figure 2. Number of trained MLPs with respect to corresponding misclassified test patterns.

3.2. Endoscopic Images

We conducted two sets of experiments using medical images. In the first set a fixed 16-21-2 MLP architecture (i.e. 16 linear input neurons, 21 nonlinear hidden neurons and 2 nonlinear output neurons) has been used and the cooccurrence matrices for the textural description of tissue samples have been applied.

Below, we present results on interpreting image regions in colonoscopic images from two different colons. Note that the malignant regions in these images belong to two different types: Image 1 (Fig. 3, left) is macroscopically a Type IIIs lesion; histologically it is a *low grade cancer*. Image 2 is macroscopically a Type V lesion; histologically it is a *moderately differentiated carcinoma* [2]. In both cases the performance of the trained MLPs has been tested on a set of 80 texture samples (40 normal and 40 malignant) from the two images.

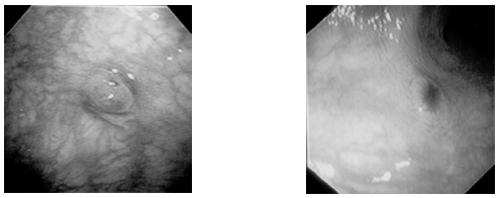

Figure 3. Colonoscopic images: Image 1 on the left, and Image 2, on the right.

The average performance of 10 MLPs than have been trained on this task is summarized in Table 2, where CE is classification error used for training; μ_{CS} is the average generalization performance; μ_{GRD} is the average number of gradient evaluations; μ_{EFE} is the average number of error function evaluations. Although the value of μ_{CS} is very similar for the two cases, the computational cost is quite different. Note also that the CE termination criterion was set high enough to avoid problems with noisy data.

CE (%)	μ_{CS} (%)	μ_{GRD}	μ_{EFE}
15	92.1	87	165
20	91.4	9	18

Table 2. Average performance of the 10 trained MFNNs.

The MLP that exhibited the best generalization performance (95%) was trained with the condition CE<0.15. It misclassified 3 out of the 80 test samples. The best performance using the CE<0.20 termination criterion was 93.75%. Details are shown in Figure 4

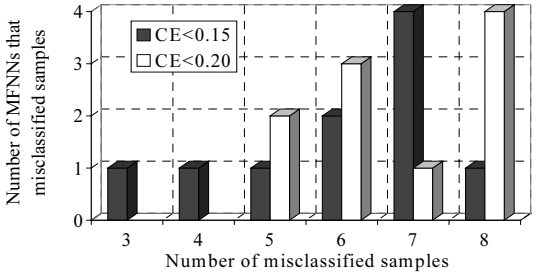

Figure 4. Number of trained MLPs with respect to corresponding misclassified samples.

In order to investigate the effect of the MLP architecture on the effectiveness of the image recognition scheme, we conducted a second set of experiments using 6 different MLP architectures with 5, 10, 15, 20, 25 and 30 hidden nodes, respectively, which were trained using 100 different initial weight sets. In this case we have used wavelet-extracted pattern, since MLPs trained with wavelet patterns have exhibited the best performance in our tests with texture images (200 pattern have been used for training and 400 for testing). The results are illustrated in Figures 5-7.

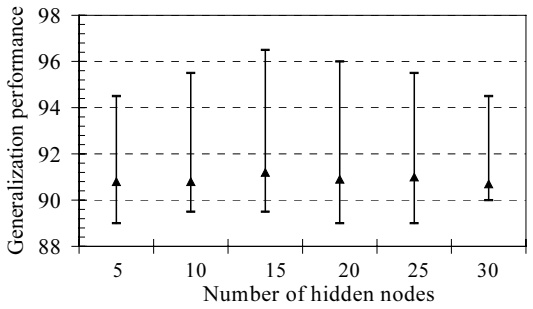

Figure 5. Generalization performance vs. number of hidden nodes.

In Figure 5, the generalization performance of the trained networks is shown with respect to the number of hidden nodes. Networks 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.

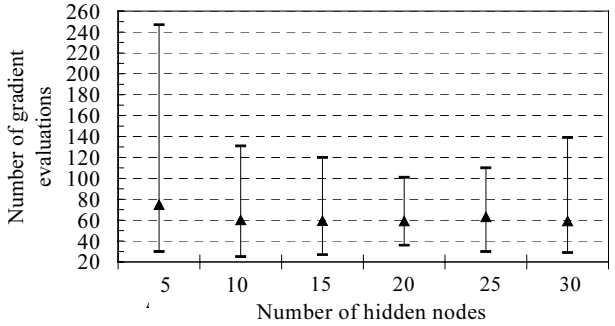

Figure 6. Number of gradient evaluations vs. number of hidden nodes.

In Figure 6, 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 MLP architecture (247 gradient evaluations).

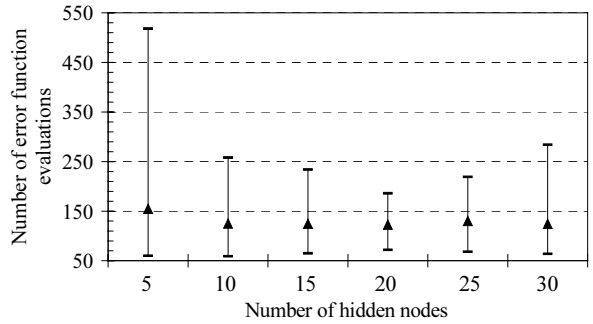

Figure 7. Number of error function evaluations vs. number of hidden nodes.

In Figure 7, 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 MLPs. The best performance (59 error function evaluations) was obtained with a 10 hidden nodes architecture.

4. CONCLUSIONS

Different schemes/methodologies for the extraction of textural features contained in images are examined for their performance in discriminating capability in medical endoscopic images. Moreover a novel DWT-based methodology was proposed which estimates the features from second-order statistics of the wavelet transform of the image. The recognition task was executed using various neural network architectures in order to determine the most suitable one for similar applications. The DWT distribution methodology was the one with which the best results were achieved. Future research will be directed to the examination of the recognition performance of the proposed scheme on medical images acquired by different sources.

5. REFERENCES

[1] S.A. Karkanis, G.D. Magoulas, D.K. Iakovidis, D.E. Maroulis, N. Theofanous, "Tumor recognition in endoscopic video images," 26th EUROMICRO Conference, Maastricht, Netherlands, pp.423-429, 2000.

[2] S.E. Kudo, H. Kashida, et.al. "Colonoscopic Diagnosis And Management Of Nonpolypoid Early Colorectal Cancer," *World Journal of Surgery,* vol. 24, no. 9, pp.1081-1090, 2000.

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

[4] C.H. Chen, "A study of texture classification using spectral features," Proc. Int. Conf. Pattern Recognition, Munich, Germany, pp.1074-1077, 1982.

[5] Y.A. Karayiannis, and T. Stouraitis, "Texture classification using the fractal dimension as computed in a wavelet decomposed image," Proceedings of the IEEE Workshop on Nonlinear Signal and Image Processing, Halkidiki, Greece, pp. 186-189, 1995.

[6] M. Unser, "Texture Classification and Segmentation Using Wavelet Frames," IEEE trans. Image Processing, vol. 4, no. 11, 1549-1560, 1995.

[7] C.P. Lim, R.F. Harrison, R.L. Kennedy, "Application of autonomous neural network systems to medical pattern classification tasks," *Artificial Intelligence in Medicine*, vol. 11, pp. 215-239, 1997.

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