Detecting Abnormalities in Colonoscopic Images
by Textural Description and Neural Networks

S.A. Karkanis[1], G.D. Magoulas[1], M. Grigoriadou[1], and M. Schurr[2]

[1]*Department of Informatics, University of Athens, GR-15784 Athens, Greece.*
`{sk,magoulas,gregor}@di.uoa.gr`

[2]*Eberhard-Karls-University of Tuebingen, Section for Minimal Invasive Surgery,
Waldhoernlestrasse 22, D-72072 Tuebingen, Germany.*
`{marc.schurr}@uni-tuebingen.de`

ABSTRACT

In this paper the performance of a simple scheme for the discrimination of different texture regions in colonoscopic images is investigated. The proposed scheme uses textural descriptors based on second order gray level statistics and employs a multilayer feedforward neural network to discriminate among normal and cancer regions. Preliminary results indicate that this scheme is capable of detecting abnormalities within the same image with high accuracy. It can be also successfully applied on different images to detect abnormalities that belong to different cancer types.

INTRODUCTION

In diagnostic colonoscopy, the medical expert based on a distributed percept of local changes interprets the physical surface properties of the tissue such as the roughness or the smoothness, the regularity and the shape to detect abnormalities. Adjacent surfaces showing different surface properties are distinguished on the basis of the texture differences. It is well known that the performance of human perception on this complex task is outstanding. The medical expert has the ability to add or remove components from an image so as to give meaning to what he sees. He can also adapt to changes so much so that even a distorted image can be recognized. No machine available today can achieve the equivalent perceptive power of a human being.

The idea of using machine learning methodologies to aid interpretation of colonoscopic images is meritorious. However, it is important for one to realize the vast difficulties in physical attributes of the colon. No two colons are similar. Even within the same colon, one section may have very different characteristics from another. This fact introduces severe limitations in the use of computer-assisted endoscopy for interpreting colonoscopic images. Given a colonoscopic image, the "true" features that are associated with it are not exactly known to us. Usually, one or more feature extraction models are used to provide values for each feature's parameters. The findings are then used to infer the results. In (Kwoh, 1995) multiple probabilistic networks were used for subsets of findings to improve the colonoscopy procedure. This approach resembles the way different expert contributions are combined.

In this work, we use texture information to train an artificial neural network for the detection of cancer regions in colonoscopic images. In the next section the proposed approach is described. Then, some preliminary simulation results are reported. The paper ends with some conclusions and future work.

TEXTURE DISCRIMINATION FOR ABNORMALITIES DETECTION

Texture plays an important role for the characterization of regions from digital images. It carries information about the micro-structure of the regions and the distribution of the gray levels. A scheme for the classification of such regions based on the texture information should be capable of encoding the properties of the texture using a number of descriptors. These descriptors are usually represented by sets of statistical measures defining by this way the vectors to be used, consequently, for the recognition.

The approach followed has two processing stages. The first stage consists of all the processing that will be performed on an image to extract all the identifiable features. The information is purely dependent on what can be extracted from the "raw" image. The second stage decides how to incorporate the information from the first stage together with background and prior information such as temporal data, relationships about features, etc., to draw inferences. This process can be considered analogous to an expert who will consolidate all the facts and verify the truth of the hypotheses.

Textural Description Using Second Order Gray Level Statistics

In this work we use texture information for the detection of cancer regions in colonoscopic images by finding some quantitative description of a texture. Usually one chooses a family of texture attributes which account for the main spatial relations between the gray levels of texture. The texture models which underlie the attributes belong to one of three categories: structural, statistical and random process. In the first case, a texture is characterized by a family of primitives, and by the way according to which they are spatially organized (e.g. Gabor filters, Daugman, 1988). The second category involves the use of statistical tools and inference: gray level co-occurrence matrices (Haralick, 1979; Gotlieb and Kreyszig, 1990), gray level run length statistics (Galloway, 1975; Siew *et al.*, 1988), gray level difference (Weszka *et al.*, 1976), second-order moments (Chen, 1982). According to the third category textured images are considered as realizations or samples of spatial random fields. Texture features are extracted by fitting random fields to image data. This encloses auto-binomial Markov fields modeling (Cross and Jain, 1983), autoregressive Gaussian models and the Gaussian Markov models (Chellappa and Chatterjee, 1985; Chellappa *et al.*, 1985; Cohen *et al.*, 1991; Kashyap and Chellappa, 1983).

In this paper we focus on normal/abnormal tissue samples discrimination based upon the analysis of the gray level cooccurrence matrices. This method evaluates a series of matrices that describe the spatial variation of gray level values within a local area. In our simulations, four cooccurrence matrices have been computed for each sample area, with a displacement of one pixel and angles of 0, 45, 90, 135 degrees. The following four features have been computed on each matrix to produce a 16-dimensional feature vector describing each tissue sample:

1) Energy - Angular Second Moment
$$f_1 = \sum_i \sum_j p(i,j)^2$$

2) Correlation
$$f_2 = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i*j)p(i,j) - \boldsymbol{m}_x \boldsymbol{m}}{\boldsymbol{s}_x \boldsymbol{s}_y}$$

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

4) Entropy
$$f_4 = -\sum_i \sum_j p(i,j)\log\big(p(i,j)\big)$$

In the above features, (*i,j*) corresponds to the (*i,j*)th entry of the matrices and 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. In our experiments, these four statistical measures out of the 14, originally proposed by Haralick (Haralick, 1979; Haralick *et al.* 1973) provided high discrimination accuracy that was only marginally increased by adding more measures in the feature vector. A similar situation arises in clinical practice: experts in interpreting colonoscopic images usually utilize only a few "important" features for inference.

Neural Network-based Abnormalities Detection from Texture Samples

Advances in neurocomputing have opened the way for the establishment of medical image interpretation systems which are able to learn complex associations by example. It is acknowledged that the appropriate use of the neural network-based methodologies in medical problem solving could be very effective to improve the efficiency and the quality of medical care. In contrast to the traditional techniques, neural network-based methods have the following advantages:

- They can capture by learning arbitrary functional relationships, not explicitly prescribed, between input and output data.
- They do not need any assumptions established on the statistical distributions underlying the input patterns.
- Their have good generalization ability and fault tolerance.
- They exhibit robustness for imprecision and uncertainty in unconstrained information environment.

Neural network-based methodologies strongly depend on the technology that provides computers with high computing performance for processing large amount of information in reasonable time. The most popular neural network architecture is the multilayer feedforward neural network (MFNN). Consider a MFNN whose *l*-th layer contains N_l neurons, $l = 1,...,M$. The neurons operate according to the following equations:

$$net_j^l = \sum_{i=1}^{N_{l-1}} w_{ij}^{l-1,l} y_i^{l-1},$$
$$y_j^l = f(net_j^l),$$

where net_j^l is, for the j-th neuron in the l-th layer ($j = 1,...,N_l$), the sum of its weighted inputs. The weights from the i-th neuron at the (l-1) layer to the j-th neuron at the l-th layer are denoted by $w_{ij}^{l-1,l}$, y_j^l is the output of the j-th neuron that belongs to the l-th layer, and the logistic function $f(net_j^l) = (1 + \exp(-net_j^l))^{-1}$ is the j-th's neuron nonlinear activation function. Training the MFNN to discriminate among normal and abnormal tissue samples is realised by adjusting the network weights through a gradient descent method. This corresponds to minimizing the error function E:

$$E = \frac{1}{2} \sum_{p=1}^{P} \sum_{j=1}^{N_M} (y_{j,p}^M - t_{j,p})^2 \ ,$$

where $(y_{j,p}^M - t_{j,p})^2$ is the squared difference between the actual output value at the j-th output layer neuron for sample p and the target output value; p is an index over input-output pairs. An enhanced training algorithm is used, which is called Back-Propagation with Variable Stepsize-BPVS (Magoulas *et al.*, 1997). After training, the network is able to discriminate between normal and abnormal texture samples by forming hyperplane decision boundaries in the pattern space.

EXPERIMENTS

In the experiments reported a 16-21-2 MFNN 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. Preliminary results have indicated that this scheme is capable of detecting abnormalities within the same colonoscopic image with 99% of success.

Below, we present results on detecting abnormalities, which belong to two different types, in colonoscopic images taken from two different colons. Image 1 (see Figure 1) is macroscopically a Type IIIs lesion according to (Kudo, 1996). Histologically it is a *low grade cancer*. Image 2 is macroscopically a Type V lesion according to (Kudo, 1996). Histologically it is a *moderately differentiated carcinoma*. The images have been digitized to a size of 2K×2K pixels with 64 gray levels depth per pixel. Textures from 10 normal and 10 abnormal tissue samples, corresponding to a 256×256 image window, have been randomly chosen from each image and used for training the MFNN. Two different conditions on the Classification Error (CE) have been used to terminate the training phase: (*i*) CE< 0.15, and (*ii*) CE< 0.2. In both cases the performance of the trained MFNNs has been tested on a set of 80 texture samples (40 normal and 40 abnormal) from the two images.
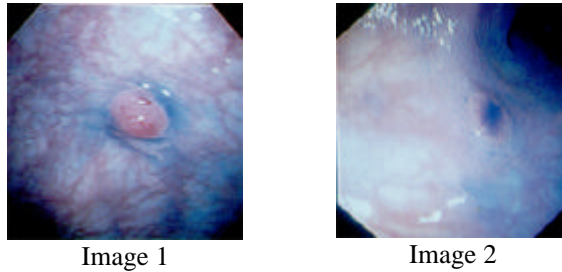

| Image 1 | Image 2 |

Figure 1. Colonoscopic images used in the experiments.

The average performance of 10 MFNNs than have been trained on this task is summarized in Table 1. The reported parameters are: CE is classification error in the training phase, m_{CS} is the average percentage of success in the testing phase and the computational cost required to train the networks is exhibited in terms of the average number of gradient (m_{GRD}) and of error function evaluations (m_{EFE}). Note that the value of the m_{CS} is very similar for the two cases, however the computational cost is quite different.

CE	m_{CS}	m_{GRD}	m_{EFE}
0.15	92.1	87	165
0.20	91.4	9	18

Table 1. Average performance of the 10 trained MFNNs.

Details on the generalization performance of the 10 networks are exhibited in Figure 2. Note, that MFNNs trained according to the first termination condition (CE<0.15) exhibit better performance than networks trained using the second condition. For example, one network trained using the condition CE<0.15 misclassified 3 out of the 80 test samples (i.e. success 95%), while two networks trained according to the condition CE<0.20 misclassified 5 out of the 80 test samples (i.e. success 93.75%). In conclusion, the neural network-based methodology seems a promising tool for detecting normal and cancer regions in colonoscopic images. We currently investigate several well known textural feature extraction techniques along with a novel discrete wavelet transform based methodology to study the effects of the different textural descriptors on MFNN training and generalization capabilities.
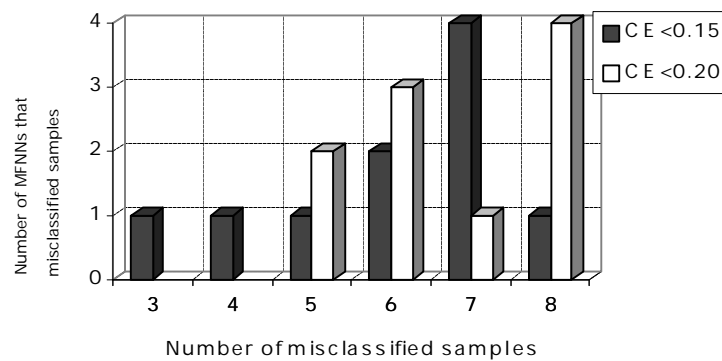

Figure 2. Number of trained MFNNs with respect to their corresponding number of misclassified test samples.

CONCLUSIONS AND FUTURE WORK

In this paper, a scheme for the detection of cancer regions with high accuracy has been presented. The scheme applies a neural network-based methodology and uses texture features extracted by the cooccurrence matrices. A next step will be the application of this approach to a sequence of frames taken from a colonoscopy video in order to construct a system that will be capable to discriminate abnormal regions in real conditions and possibly on-line during operation.

REFERENCES

Chellappa, R. and Chatterjee, S. (1985). Classification of textures using gaussian markov random fields, *IEEE Trans. Acoustic, Speech and Signal Processing*, 33, 959-963.

Chellappa, R., Chatterjee, S., and Bagdazian, R. (1985). Texture synthesis and compression using gaussian-markov random field models, *IEEE Trans. System, Man and Cybernetics*, 15, 298-303.

Chen, C.H. (1982). A study of texture classification using spectral features, *Proc. Int. Conf. Pattern Recognition*, Munich, Germany, 1074-1077.

Cohen, F.S., Fan, Z., and Patel, M.A. (1991). Classification of rotated and scaled textured images using gaussian-markov random field models, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13, 192-202.

Cross, G.R. and Jain, A.K. (1983). Markov random field texture models, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 5, 1, 25-39.

Daugman, J.G. (1988). Complete discrete 2D Gabor transforms by neural networks for image analysis and compression, *IEEE Trans. Acoustic, Speech and Signal Processing*, 36, 1169-1179.

Galloway, M.M. (1975). Texture analysis using gray level run lengths, *Computer Graphics and Image Processing*, 4, 172-179.

Gotlieb, C.C. and Kreyszig (1990). Texture descriptors based on cooccurrence matrices, *Computer Vision, Graphics and Image Processing*, 51, 70-86.

Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification, *IEEE Trans. Systems, Man and Cybernetics*, 3, 6, 610-621.

Haralick R.M. (1979). Statistical and structural approaches to texture, *IEEE Proceedings*, 67, 786-804.

Kashyap, R.L. and Chellappa, R. (1983). Estimation and choice of neighbors in spatial interaction models of images, *IEEE Trans. Information Theory*, 29, 60-72.

Kudo (1996). *Early Colorectal Cancer*, Igaku-Shoin publischers, Tokyo.

Kwoh, C. K. (1995). Probabilistic reasoning from correlated objective data. Ph.D. Thesis. Imperial College, London.

Magoulas, G.D., Vrahatis, M.N., and Androulakis, G.S. (1997). Effective back-propagation training with variable stepsize, *Neural Networks*, 10, 69-82.

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

Weszka, J.S., Dyer, C.R., and Rosenfeld, A. (1976). A comparative study of texture measures for terrain classification, *IEEE Trans. Systems, Man and Cybernetics*, 6, 269-285.