DETECTION OF LESIONS IN ENDOSCOPIC VIDEO USING TEXTURAL DESCRIPTORS ON WAVELET DOMAIN SUPPORTED BY ARTIFICIAL NEURAL NETWORK ARCHITECTURES

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

[(1)]Dept. of Informatics, University of Athens, Ilissia,15784 Athens, Greece.
{sk, diakov, dmarou}@di.uoa.gr.
[(2)]Hellenic Aerospace Industry, 2 Rhodou str., A. Ilioupolis, 16342 Athens, Greece, dakarras@hol.gr.

ABSTRACT

Video processing for classification purposes in medical imaging is an area with great importance. In this paper a framework for classification of suspicious lesions using the video produced during an endoscopic session is presented. The proposed approach is based on a feature extraction scheme that uses second order statistical information of the wavelet transformation. These features are used as input to a Multilayer Feedforward Neural Network (MFNN) architecture, which has been trained using features of normal and tumor regions. The system uses a limited number of frames with a rather small population of training vectors. The classification results are promising, since the system has been proven capable to classify and locate regions, that correspond to lesions with a success of 94 up to 99%, in a sequence of the video-frames. The proposed methodology can be used as a valuable diagnostic tool that may assist physicians to identify possible tumor regions or malignant formations.

1. INTRODUCTION

Medical diagnosis is based on information obtained from various sources, such as results of clinical examinations and histological findings, patient's history and other data that physician considers in order to reach a final diagnostic decision. Imaging techniques have been extensively used, in the last decades, as a valuable tool in the hands of an expert for a more accurate judgment of patients' condition. In addition, approaches using medical images for automatic detection of lesions, have been proposed [1-6]. Such methodologies can increase expert's identification ability while decreasing the need for aggressive intervention. Moreover, the shortcomings of biopsies, such as discomfort for the patient, delay in diagnosis, and limited number of tissue samples can be also minimized.

In endoscopic sessions the physician uses video sequences in order to locate lesions suspicious for malignancy. Small image regions may possibly represent such lesions in early stages. The evaluation of small regions is obviously difficult. Lesions at this stage can either be underestimated or even possibly not estimated at all. A system capable to classify image regions to normal or abnormal will act as a second - more detailed - "eye", by processing the endoscopic video. Its exceptional value and contribution in supporting the medical diagnosis procedure is high.

In this paper, we present an approach that aims to the detection of abnormal regions during an endoscopic procedure by processing the video frames of the session. The proposed approach is based on feature extraction techniques using second order statistics of the wavelet transformation of each video-frame. This statistical information is estimated utilizing cooccurrence matrices forming textural signatures of the corresponding regions. Texture is the main property to be evaluated for the discrimination between malignant and benign lesions [7, 8]. A MFNN architecture has been used for classification and characterization purposes of different regions examined. The recognition capability of the proposed approach has been estimated to a percentage, which ranges between 94 and 99% of success.

The paper that follows is organized in 5 sections. In section 2, there is a description of the wavelet transformation and the textural analysis used to define the feature vectors. The recognition system is described in section 3. In section 4, we present the results of experiments performed for different instances of the corresponding endoscopic videos. Conclusions at which we have reached after the application of the proposed methodology to different endoscopic videos are summarized in section 5.

2. FEATURE EXTRACTION SCHEME

As it has already been expressed [7, 8], a major characteristic for the discrimination between normal and possibly abnormal regions is the texture of the tissue along with other features as deformation, color change and bleeding spots. The pit pattern, a pattern that describes the textural structure of the tissue can be primarily used as the information for an accurate decision. According to this, the classification of region in endoscopic video frames can be treated as a texture classification problem. In the following paragraphs the scheme for the extraction of textural features is presented.

2.1. A novel scheme for features extraction based on the DWT

In the proposed approach each image frame, is transformed to its wavelet domain using the discrete case of the transformation (Discrete Wavelet Transform).

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

where $j_0 \in Z$, Z is the set of integers, φ is the scaling function, ψ is the mother wavelet, $c_{jok} = <g, \varphi_{jok}>$ and $d_{jk} = <g, \psi_{jk}>$, where $<., .>$ is the standard inner product of two functions. The wavelet bases used for the transformation were those of Daubechies, using all the wavelet coefficients d_j. These bases were considered due to their orthonormal property, an important issue that maintains the textural structure along the different scales of the transformation [9]. One-level wavelet decomposition of the images has been performed resulting in four wavelet bands for further processing.

2.2. Textural feature vectors

Statistical measurements based on second order statistics used as textural features. These statistical descriptors were estimated over the cooccurrence matrices of each region in the image. Cooccurrence matrices [10], 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. From these matrices, sets of statistical measures are computed (called feature vectors) for building different texture models. We have considered four angles, 0^o, 45^o, 90^o, 135^o, as well as a predefined distance of one pixel in the formation of the cooccurrence matrices. Among the 14 statistical measures, originally proposed [10], we have considered four. Namely, angular second moment, correlation, inverse difference moment and entropy. These measures provide high discrimination accuracy [10], which can be only marginally increased by adding more components in the feature vector.

2.3. The algorithm

The proposed methodology is based on the processing of image windows from each video frame according to the following steps:

1. One-level DWT is applied on the window according to base chosen.
2. A set of 16 statistical features is estimated for each of the three wavelet bands of the transformed window.
3. These 48-component feature vectors form the input vector to the neural classifier.

3. RECOGNITION PROCESS USING ARTIFICIAL NEURAL NETWORKS

Scientific interest in models of neuronal networks or artificial neural networks (ANNs) mainly arises from their potential ability to perform interesting recognition tasks. Advances in ANNs may contribute to the design and development of new computational tools to analyze multidimensional and multimodal medical images. This holds also in the case of sequences of images obtained through minimally invasive imaging procedures, e-specially when therapy is guided by these images (video-surgery, interventional radiology, guided radiotherapy, etc.).

The most popular ANN is the so-called MFNN. In a MFNN, whose l-th layer contains M_l neurons, where l = 1,...,L. Artificial neurons operate according to the following equations:

$$net_j^l = \sum_{i=1}^{M_{l-1}} w_{ij}^{l-1,l} y_i^{l-1} \ ,$$

$$y_j^l = f(net_j^l) \ ,$$

where net^l_j is, for the j-th neuron in the l-th layer (j = 1,.. M_l), the sum of its weighted inputs. The weights for connections from the i-th neuron at the (l-1) layer to the j-th neuron at the l-th layer are denoted by $w^{l-1,l}{}_{ij}$, y^l_j is the output of the j-th neuron that belongs to the l-th layer, and the logistic function $f(net^l_j) = (1 + \exp(-net^l_j))^{-1}$ is the j-th's neuron non-linear activation function.

Training a MFNN to recognize abnormalities in image regions is typically realized by adjusting the network weights through a gradient descent method following an error correction strategy. In a MFNN this operation corresponds to the minimization of network's learning error. In the presented experiments training is performed with the on-line version of the momentum back-propagation algorithm [11]. The incorporation of momentum represents a minor modification to the weight update, yet it may have some beneficial effects on the learning behavior of the algorithm [11]. After training, the MFNN is able to discriminate between normal and abnormal texture regions by forming hyperplane decision boundaries in the pattern space.

Optimal generalized learning can be achieved by a MFNN structure that has the minimal number of neurons necessary to map the exemplar input instances for K classes into the associated K target identifiers (outputs)

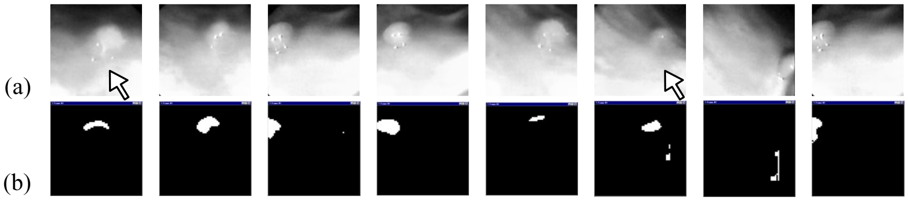

Figure 1. (a) Original video-frames, (b) classification results after post-processing. The arrows indicate the frames used for training.

and satisfy the validation and verification tests [12]. Additionally, minimum number of neurons reduces the computational complexity, which becomes a crucial factor when the system is to be incorporated in an integrated application environment.

According to [12], the average number of neurons for the hidden layer, can be obtained by:

$$M_{ave} = [1.7095 \log_2(2K)].$$

In order to keep the number of hidden layer neurons as small as possible, the experiments for each training set were conducted starting from:

$$M_0 = [\log_2(K)] + 1$$

hidden neurons and increasing by one until the MFNN learns properly. The problem becomes more complex if we take into account that the training and the testing sets has to be selected out of multiple data sets corresponding to different video-frames.

Assuming that the number of classes presented in each frame is K = 2, the number of output neurons should also be 2. The vectors corresponding to image areas presenting abnormal tissue comprise the first class, and vectors corresponding to image areas presenting normal tissue comprise the second class. Two classes are enough to signify the existence of lesions within the image. The dimension of the feature space designates the number of nodes of the input layer. Thus, in the presented experiments that follow in the next paragraph, utilize fully connected 48-X-2 MFNN architectures with X ≥ 2 hidden neurons.

4. RESULTS AND DISCUSSION

In our experiments, eight video-frames of a small polyp of the colon were used (Fig. 1a.). The video was recorded in standard VHS videotape, were digitized at a size of 512 x 512 pixels and 256 gray levels depth. The resolution of the images can be considered as low to moderate.

The classification capability of the system between the different classes was examined using:

- training sets of different sizes. The number of training samples used were 1% and 0.5% of the total number of testing samples. The total number of testing samples was varying between 22472 and 32768, depending on the window size used. The window sizes that have been tested were 8x8, 16x16, 32x32, 64x64, 96x96 pixels, respectively.
- one frame for training, i.e. the first one (Fig.1a),
- two frames for training, i.e. the first and the sixth frame (Fig.1a, indicated by arrows).

The tests and overall classification rate was computed for the sequence of all 8 video-frames (Fig.1a). The mean classification error, compared to window size and training population is illustrated in Fig.2. As it can be seen from Table 1, the mean classification error (MCE) varies between 4% and 7%. It can also be noticed that MCE is reduced as the window size increases. Small size windows are possible to bring noise to the output of the network. In these cases post-processing for the reduction of this noise, has been implemented. This post-processing uses a voting scheme in the neighborhood of each window aiming to the elimination of isolated regions. In Fig.1b, the results of this post-processing procedure are presented.

Another point that should be noted is that the training samples were limited comparing to the total testing samples, without influencing MCE (Table 1). As it can be observed from Fig.1, the illumination conditions are different amongst different frames. This leads us to the conclusion that the proposed feature extraction scheme is not sensitive to the variances of illumination.

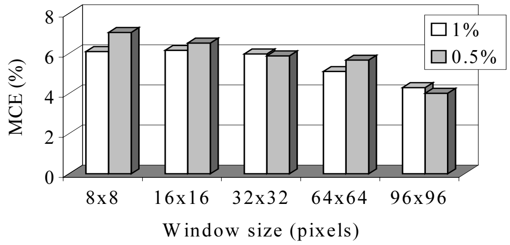

Figure 2. Mean overall classification error (MCE) for various window sizes and training samples.

Overall Classification Error Statistics

Training Set Size	1% of Testing Set			0.5% of Testing Set		
Window size (pixels)	Hidden neurons	Mean (%)	St. Dev. (%)	Hidden neurons	Mean (%)	St. Dev. (%)
96x96	5	4.29	2.36	3	4.02	2.26
64x64	6	5.10	1.80	3	5.67	2.09
32x32	7	5.98	1.90	4	5.88	1.64
16x16	8	6.16	1.49	4	6.52	1.64
8x8	8	6.10	1.25	4	7.05	1.63

Table 1. Overall mean classification error (MCE) and standard deviation for various window sizes and training samples.

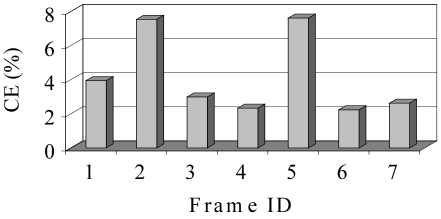

Figure 3. Overall classification error (CE) using one training frame.

The experimental combinations of the above-mentioned experimental parameters generated a large amount of results. The classification error presented in Fig.3, varies between 2% and 7%. Table 2, also describes the classification performance of the proposed system compared to physician's characterization.

		MLP	
Physician	**Regions**	Normal	Abnormal
	Normal	93.99%	6.00%
	Abnormal	2.79%	97.21%

Table 2. Confusion matrix presenting the percentage classification of the MLP's performance in characterizing image regions, using one training frame.

Finally, indicative results using two frames for training are shown in Fig.4. The classification error varies between 1% and 6%.

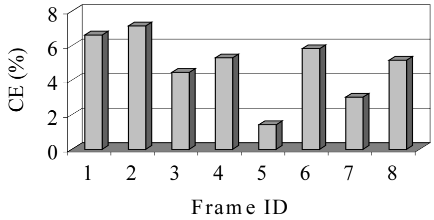

Figure 4. Overall classification error (CE) using two training frames.

5. CONCLUSIONS

In this paper, we presented a framework for the classification of normal and abnormal tissue during endoscopic session. The proposed technique applies a feature extraction algorithm, on sequences of endoscopic video-frames. The estimated features, which characterize the textural properties of the image, are considered on the wavelet domain. These features are used as input to a MFNN architecture for classification. After extensive experimentation on various endoscopic sequences, the results were promising giving a percentage up to 99% of success. Our major future prospects include the implementation of the proposed system for near to real-time processing of the video frames.

6. ACKNOWLEDGMENTS

This work is partially funded by National and Kapodistrian University of Athens, Special Account of Research Grants.

7. 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. Karkanis, G.D. Magoulas, N. Theofanous, "Image recognition and neuronal networks: intelligent systems for the improvement of imaging information," *Minimal Invasive Therapy and Allied Technologies*, vol. 9, pp. 225-230, 2000.

[3] Y. Jiang, R.M. Nishikawa, D.E. Wolverton, C.E. Metz, M.L. Giger, R.A. Schmidt, C.J. Vyborny, K. Doi, "Malignant and benign clustered microcalcification: automated feature analysis and classification," *Radiology*, vol. 198, pp. 671-678, 1996.

[4] A.S. Miller, B.H. Blott, and T.K. Hames, "Review Of Neural Network Applications In Medical Imaging And Signal Processing," *Medical and Biological Engineering and Computing,*" vol. 30, pp. 449-464, 1992.

[5] C. Chiu, "A Novel Approach Based On Computerized Image Analysis For Traditional Chinese Medical Diagnosis Of The Tongue," Computer Methods and Programs in Biomedicine, vol. 61, pp. 77-89, 2000.

[6] J.D. Lee and Y.L. Hsiao, "Extraction Of Tumor Region In Color Images Using Wavelets," *Computers and Mathematics with Applications,* vol. 40, pp. 793-803, 2000.

[7] 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.

[8] S. Nagata, S. Tanaka, K. Haruma, M. Yoshihara, K. Sumii, G. Kajiyama, F. Shimamoto, "Pit Pattern Diagnosis Of Colorectal Carcinoma By Magnifying Colonoscopy: Clinical And Histological Implications," Int. J. Oncol., vol. 16, no. 5, pp. 927-934, 2000.

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

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

[11] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2^{nd} ed., Prentice Hall, New Jersey, 1996.

[12] G.C. Looney, *Pattern Recognition Using Neural Networks: Theory And Algorithms For Engineers And Scientists*, Oxford University Press, New York, pp. 316-319, 1997.