

CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames

D.E. Maroulis ^{a,*}, D.K. Iakovidis ^a, S.A. Karkanis ^a, D.A. Karras ^b

^a *RTS-image Group, Department of Informatics and Telecommunications, University of Athens, Panepistimiopolis, Ilisia, 15784 Athens, Greece*

^b *Hellenic Aerospace Industry, Schematari, Greece*

Received 11 May 2001; received in revised form 21 January 2002; accepted 24 January 2002

Abstract

In this paper, we present CoLD (colorectal lesions detector) an innovative detection system to support colorectal cancer diagnosis and detection of pre-cancerous polyps, by processing endoscopy images or video frame sequences acquired during colonoscopy. It utilizes second-order statistical features that are calculated on the wavelet transformation of each image to discriminate amongst regions of normal or abnormal tissue. An artificial neural network performs the classification of the features. CoLD integrates the feature extraction and classification algorithms under a graphical user interface, which allows both novice and expert users to utilize effectively all system's functions. It has been developed in close cooperation with gastroenterology specialists and has been tested on various colonoscopy videos. The detection accuracy of the proposed system has been estimated to be more than 95%. As it has been resulted, it can be used as a supplementary diagnostic tool for colorectal lesions.

© 2002 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Medical imaging; Computer aided diagnosis; Endoscopy; Texture; Neural networks; Colorectal lesions

1. Introduction

In the last decades, several systems able to support medical diagnosis have been proposed to augment the natural capabilities of medical specialists in the diagnostic process. Many of these systems utilize imaging techniques and artificial intelligence algorithms to characterize and assess

abnormal biological formations in the human body [1–4]. A system that would assist physicians to diagnose colorectal cancer or detect colorectal cancer precursors would definitely increase the efficiency of the typical screening methodologies.

The risk for colorectal cancer can be reduced through regular screenings, since pre-cancerous polyps can be detected and removed before the cancer even starts. Colonoscopy is the screening procedure, which includes direct visual examination of the colon by means of a fiberoptic endoscope. It is the most accurate screening technique for detecting polyps, which also allows biopsy of

* Corresponding author. Tel.: + 301-727-5307/5317; fax: + 301-727-5333.

E-mail addresses: rtsimage@di.uoa.gr (D.E. Maroulis), dakarras@hol.gr (D.A. Karras).

lesions and resection of most of the polyps [5]. Endoscope's output, i.e. the color images acquired during the procedure, is usually recorded by standard video equipment for reevaluation and archiving purposes. The physician searches for changes in the cellular pattern (pit pattern) of the colon lining, which might be the very earliest sign of polyps. It has been proposed that pit patterns, which appear as textural variations of the tissue, can be used for the qualitative and quantitative diagnosis of colorectal lesions [6,7]. These textural alterations of the colonic mucosal surface signify that this property could also be used for the automatic detection of these lesions.

In this paper, we present CoLD (Colorectal Lesions Detector) an innovative system to support colorectal cancer diagnosis and detection of precancerous polyps, by processing endoscopy images or video frame sequences acquired during colonoscopy. The innovation of CoLD lies on the detection of colorectal lesions from colonoscopy video-frames, using texture segmentation, in conjunction with neural network-based classification techniques.

CoLD provides an integrated environment, which incorporates texture segmentation techniques that are efficiently applied on colonoscopy images [8–13], aiming to the discrimination between image regions that depict normal or abnormal tissue. The proposed methodology is considered in two phases. The first, implements the extraction of image features, where the texture information of the image is represented by a set of descriptive statistical features calculated on the wavelet transformation of the image. In the second phase, an ANN (artificial neural network) undertakes the classification of these features.

ANNs possess interesting function approximation capabilities since they approximate arbitrary nonlinear mappings [14,15] without restrictions of the input space dimensionality [16]. This characteristic makes them useful in biomedical applications, where, usually, a multitude of input features are involved [17,18].

Detection of colorectal lesions is subject to uncertainties due to inaccurate measurements and lack of precise modeling of the lesion's image characteristics, especially for the small ones.

Therefore, the advent of ANNs in computer-aided diagnosis is significant due to their generalization capabilities.

In the next section of this paper we present the computational methods and the design principles that we have considered developing this system. In the third section, we describe CoLD's structure from both the user's and developer's point of view. In the fourth section we describe technical details regarding the implementation and the performance of the proposed system and finally in the last section a set of conclusions is summarized.

2. Methods

2.1. Image features extraction

Texture is an innate property existing in any image and especially in medical images since the tissue itself carries a dominant textural appearance. Although there is no rigorous definition of the texture itself, the significance of the information that it carries has been denoted in the literature for different machine vision applications [19–26]. The classification of image regions within colonoscopy video-frames can be treated as a texture classification problem, by exploiting the textural characteristics of the corresponding regions for the discrimination between the lesions and the normal tissue. The texture of the tissue is considered as a composition of pit patterns. A pit pattern describes the structure of the tissue's surface [6,7]. In the framework that we have developed for the recognition of lesions from the video-frame sequences, the texture is investigated in two stages. In the first stage, each frame is transformed to its wavelet domain and in the next stage the statistical features are estimated.

2.1.1. Texture on the wavelet domain

The characterization of image textures depends on the scale of analysis. The scale relates to the size of the textural elements or neighborhood used for textural analysis. Psycho-visual studies [27–29] indicate that the human visual system processes images in a multi-scale manner. This

approach motivates the use of multi-scale or multi-resolution approaches for texture analysis. Wavelet transform provides a formal approach to multi-resolution representation [30].

In the proposed approach each image frame, is transformed to its wavelet domain using the discrete case of the transformation DWT (Discrete Wavelet Transform). Two functions mutually orthonormal are initially adopted: the scaling function φ and the mother wavelet ψ . Other wavelets are then produced by translations of the scaling function φ and dilations by the mother wavelet ψ , according to the equations:

$$\begin{aligned} \varphi_{j_0,k}(t) &= 2^{j_0/2} \varphi(2^{j_0}t - k) \\ \psi_{j,k}(t) &= 2^{j/2} \psi(2^j t - k), \quad j = j_0, j_0 + 1, \dots, k \in Z \end{aligned} \tag{1}$$

for some $j_0 \in Z$, where Z is the set of integers. Referring to the literature, the scaling function φ defines a kernel function and the mother wavelet ψ results to an oscillation of the input signal [31]. Families of scaling functions can act as suitable bases for $L^2(R)$ or for alternative spaces. The structure of a wavelet basis is deterministic in location and frequency due to translation and dilation, respectively. A function $g \in L^2(R)$ can be represented in a wavelet series, using a given basis, as

$$f(t) = \sum_k c_{j_0 k}(k) \varphi_{j_0 k}(t) + \sum_{j=j_0+1}^{\infty} \sum_{k \in Z} w_{jk} \psi_{jk}(t) \tag{2}$$

where $c_{jk} = \langle g, \varphi_{j_0 k} \rangle$ and $w_{jk} = \langle g, \psi_{jk} \rangle$, where $\langle \cdot, \cdot \rangle$ is the standard L2 inner product of two functions:

$$\langle g_1, g_2 \rangle = \int_R g_1(t) g_2(t) dt \tag{3}$$

DWT, which is defined by (Eq. (2)) can lead to better texture modeling [25], so we can better exploit the well-known local information extraction properties of the wavelet image decomposition [32]. The wavelet bases that are used by the proposed system are those of Haar and Daubechies. The first one is considered simple compared to the other wavelet bases. The Daubechies bases are considered for their orthonormal property, which is important for the

maintenance of the textural structure in the different scales of the transformation [33,34]. One-level wavelet decomposition of the images has been performed resulting in four wavelet bands for further processing (Fig. 1).

2.1.2. Textural feature vectors

Statistical measurements based on second-order statistics can be used as textural features. These statistical descriptors were estimated using the cooccurrence matrices of each region in the image. Cooccurrence matrices [35,36], represent the spatial distribution dependence of the gray levels within an area. We have considered four angles, namely $0^\circ, 45^\circ, 90^\circ, 135^\circ$ as well as a predefined distance of one pixel in the formation of the matrices. Among the 14 statistical measures, originally proposed [35,36], we have considered only four. Namely, angular second moment, correlation, inverse difference moment and entropy. These four measures provide high discrimination accuracy [8–13,19,21], which can be only marginally increased by adding more measures in the feature vector.

Energy – angular second moment

$$f_1 = \sum_i \sum_j^{N_g, N_g} p(i, j)^2$$

Correlation
$$f_2 = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i \cdot j) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

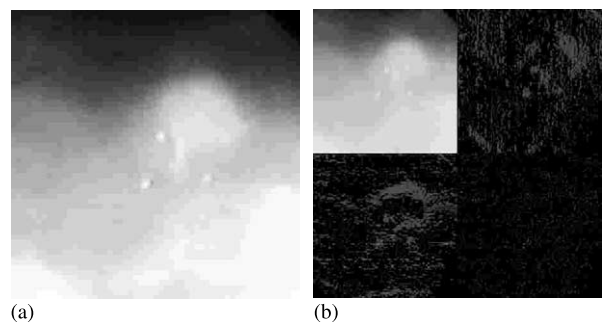


Fig. 1. (a) Original image, (b) One-level wavelet decomposition of (a).

Inverse difference moment

$$f_3 = \sum_i \sum_j \frac{1}{1 + (i-j)^2} p(i, j)$$

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

where $p(i, j)$ is the ij th entry of the normalized cooccurrence matrix, N_g is the number of gray levels of the image, μ_x , μ_y , σ_x , and σ_y are the means and standard deviations of the marginal probability $p_x(i)$ obtained by summing the rows or columns of matrix $p(i, j)$.

2.2. Artificial neural network architecture

The proposed system incorporates MLP (Multi-layer Perceptron) neural networks [37] as a building block of the colon cancer diagnosis system based on video-frames, due to the following facts [37]:

- The capability of MLP neural networks to learn difficult tasks and their quite satisfactory generalization capabilities even in the case of classification problems with multiple input features and classes.
- The capability of MLP neural networks to generalize well, even with noisy data and missing information.
- The MLP neural networks are universal approximators whose generalization capabilities are not affected, at least to the first-order, from the curse of dimensionality.
- MLP training needs a finite only training set of patterns, unlike other ANNs, which need a much larger number of training patterns.

It is well known that MLP is an ANN architecture that can be efficiently trained in large-scale problems by using a supervised procedure, the BP (Back-Propagation) algorithm. This algorithm involves a nonlinear optimization methodology, which attempts to minimize an objective function formulated in the output layer of the MLP architecture. A common formula for this function is:

$$E(n) = \frac{1}{2} \sum_{j=C} e_j^2(n)$$

where $e_j(n) = d_j(n) - y_j(n)$, the error signal, $d_j(n)$ the desired response and $y_j(n)$ the function signal appearing at the output of neuron j at iteration n .

The BP algorithm can be found as a stochastic optimization procedure (on-line version) or as a conventional optimization algorithm (off-line version) [37]. For large scale and difficult real world problems the former method is widely adopted. This on-line BP training procedure is integrated in CoLD using the following formulation.

The weight update of any weight w_{ji} at every pattern presentation is given by:

$$\Delta w_{ji}(n) = \alpha \Delta w_{ji}(n-1) + \eta \delta_j(n) y_i(n)$$

where α and η are the momentum and learning rate, respectively, while δ_j is the local error of neuron j and y_i is the output function of neuron i . The local error δ_j takes on the following forms:

$$\delta_j(n) = e_j(n) \phi'_j(u_j(n))$$

for any neuron j in the output layer and

$$\delta_j(n) = \phi'_j(u_j(n)) \sum_k \delta_k(n) w_{kj}(n)$$

for any neuron j in a hidden layer. In this last form δ_k is the symbol for the local error of neuron k belonging to the upper hidden layer and ϕ_j symbolizes the activation function of neuron j , and $u_j(n)$ is the weighted sum of all synaptic inputs plus bias, of neuron j at the n th iteration.

3. System description

3.1. Structure and functions

The proposed program consists of two major sets of functions organized in two blocks, as shown in Fig. 2, the Image Data Management block (IDMb) and the Classification block (Cb).

The flowchart shown in Fig. 3 presents abstractly the interrelation between the processes of each block. CoLD accepts as input, images that were acquired by an endoscope during colonoscopy. The endoscope's output is recorded using a standard video recording system (VCR), digitized by a frame-grabbing device and finally stored as a sequence of 8-bit non-interlaced grayscale frames.

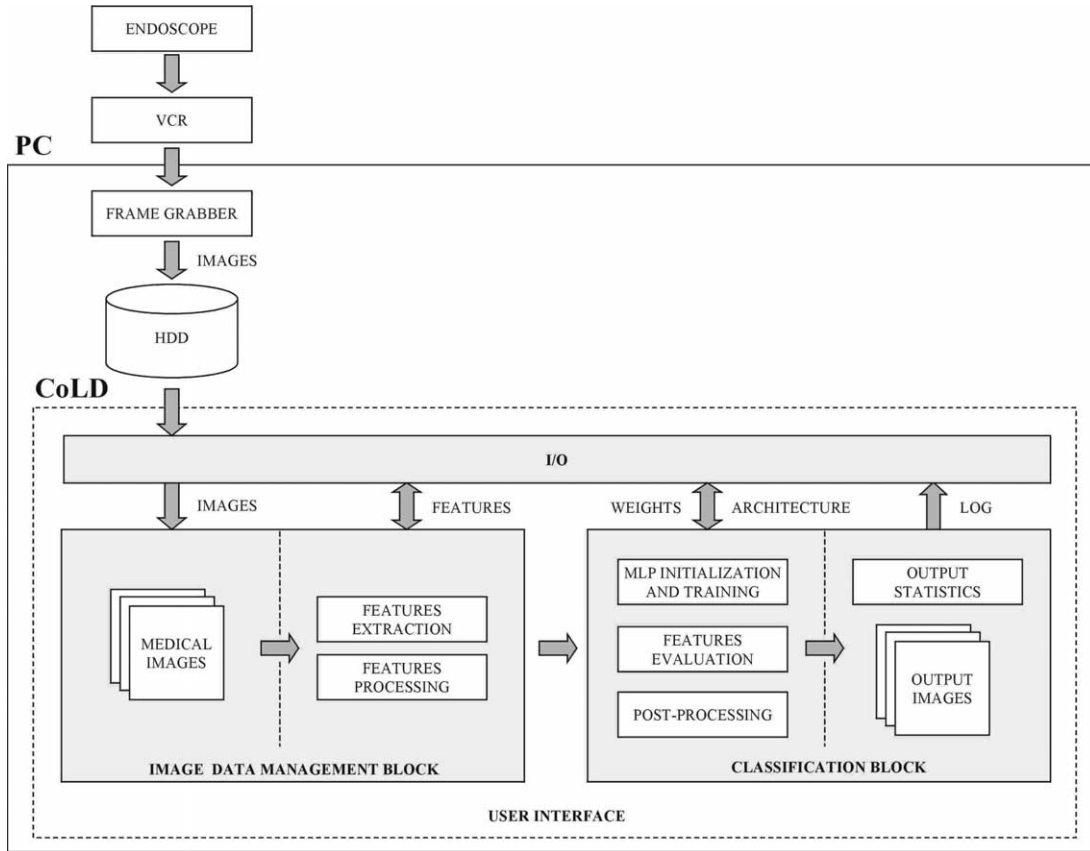


Fig. 2. Block diagram of CoLD.

3.1.1. Image data management block

The acquired frames are forwarded to the (IDMb). This block incorporates functions for:

- *Feature extraction*, which implement the algorithms for the estimation of the statistical measures on the wavelet domain of the images.
- *Feature processing*, which implement the arithmetic processing of the features and the management of the feature vectors, including normalization, shuffling, sorting, counting, and editing.
- *I/O tasks*, regarding the storage and retrieval of features estimated over the input images.

3.1.2. Classification block

Cb realizes a set of functions that mainly aim to the classification of input data, organized in vectors, for the detection of lesions in the

colonoscopy video-frames. The classification of the input vectors is a two-stage procedure. It includes the training stage where the MLP uses a representative set of vectors to adjust its weights and the evaluation stage, which leads to the characterization of regions in the input frames. Detailed reports on the evaluation results are issued and can be displayed as output images depicting the spatial distribution of regions that possibly correspond to abnormal tissue. The functions incorporated in the Cb can be summarized into the following four categories:

- *MLP initialization and training functions* that are used for the configuration of the MLP architecture, its initialization, and the modification of its parameters.
- *Evaluation functions*, which implement the forward pass of the input features through the

MLP, and associate the output vectors with the existent classes.

- *Post-processing functions*, that can be applied at the output of the MLP to enhance the results of the evaluation.
- *Functions for I/O tasks* that can be used for storing and retrieving MLP configurations that

apply in different cases.

Intermediate classification results as occur during the training or the evaluation stage can be logged in plain text file format, for reexamination. Table 1, presents a complete list of functions corresponding to each of the described blocks shown in Fig. 2.

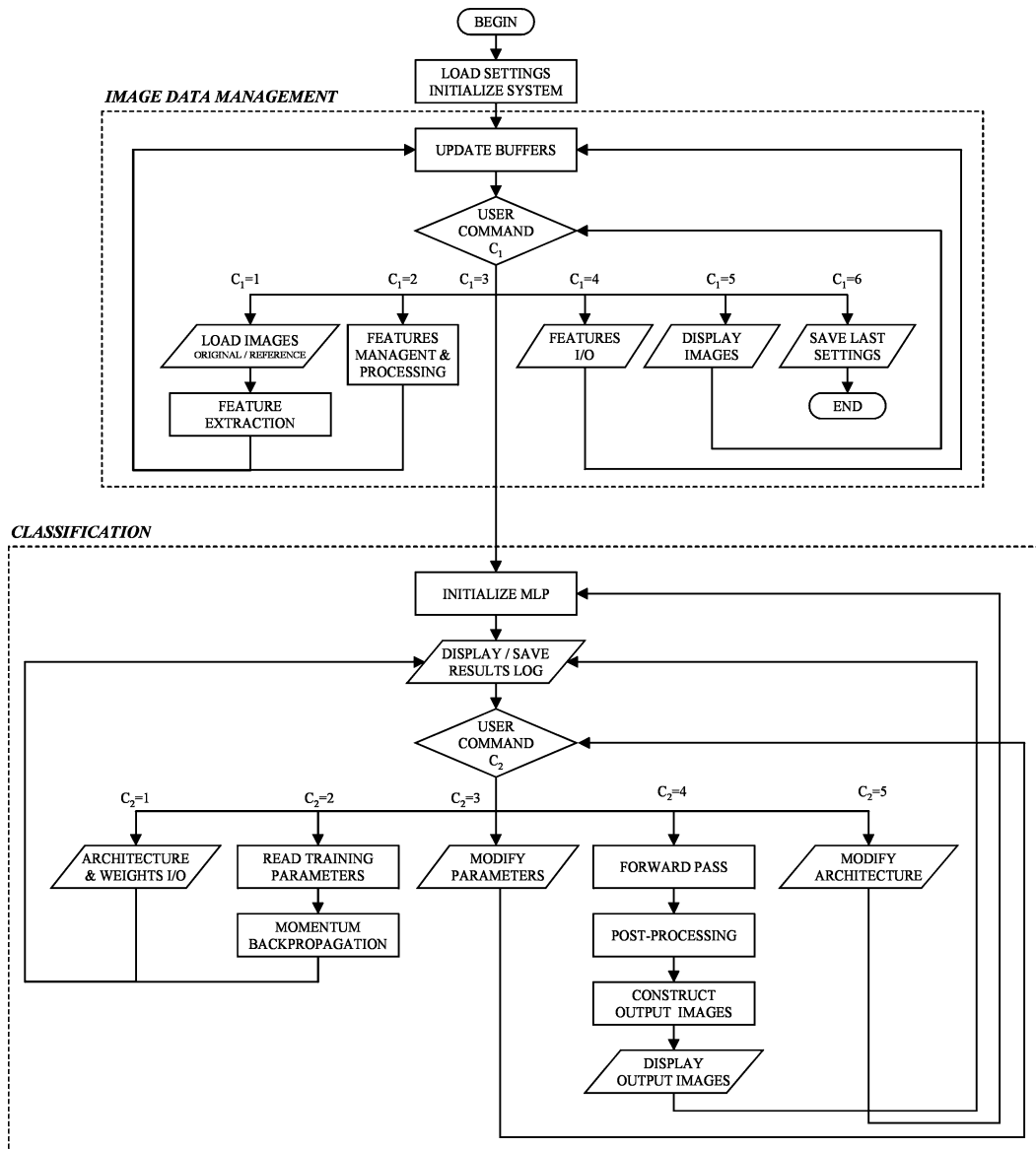


Fig. 3. Flowchart of CoLD.

Table 1
CoLD's table of functions

Image data management block	Classification block
<i>Feature extraction</i>	<i>MLP Initialization</i>
Extract training features from images	Modify MLP architecture
Extract testing features from images	Initialize MLP synapses
<i>Arithmetic processing and management</i>	<i>Training</i>
Normalize the features	Modify training parameters
Count the patterns of each class	Train MLP with Backpropagation
Sort the feature vectors by class	
Shuffle the feature vectors	<i>Evaluation</i>
Equalize the feature vectors	Evaluate image data
Remove feature vectors	
Manual data editing	<i>Post-processing</i>
Clear the training data	Remove noise after evaluation
Clear the testing data	
Copy the training data as testing data	<i>I/O</i>
Copy the testing data as training data	Log training and evaluation results
Change class labels	Load MLP architecture and weights
<i>I/O</i>	Save MLP architecture and weights
	Save MLP weights in text file
Load training images	Save MLP outputs in text file
Load images for evaluation	Visualize output image sequence
Load training data from disk	
Load testing data from disk	
Save the training data to disk	
Save the testing data to disk	
Display sequence of loaded images	

3.2. The user interface

The development of a user interface should take into account user requirements that can be determined by user needs and the tasks the system is intended to support [38]. Compatibility between the user's understanding of the system and his

skill or knowledge, can be achieved by a user interface design which takes into account the user's level of skill or knowledge, the functionality of user tasks, their important procedural characteristics and the type of information use [39].

CoLD's user interface has been designed for scientists working in the field of medicine, including those working in the field of medical imaging, image analysis and recognition. This aspect introduces the need to incorporate a design strategy, which will allow the accommodation of multiple usage classes, from novice or first-time users, to knowledgeable intermittent or even expert frequent users. Permitting a level-structured approach to learning [40], novices can be taught a minimal subset of objects and actions with which to get started and progressively expand their potential to more complex tasks. On the other hand, users having strong knowledge of the supported tasks and interface concepts can show rapid progress, work faster and soon learn how to take advantage of most options provided by the system.

The functions that are essential for the system's operation are provided either through menu options or icon-buttons. Novice users can directly apply these functions simply through the icon-buttons, and knowledgeable intermittent or expert users can select the additional functions, which are mainly targeted to the arithmetic processing and management of the features, through menu options.

Moreover, on program termination the users are prompted to save the current parameters in a binary file. The latter saved parameters will be automatically loaded when the application is reinitiated, so the user will not have to readjust all program parameters that suit to his/her endoscopy acquisition setup. A set of default parameters, decided upon the experimental results executed in cooperation with the medical experts, is available to be restored at any time by the user. Under this framework the provided environment appears to be less confusing for novices without delimiting the potentials provided to more experienced users.

The user interface is consisted of three panels:

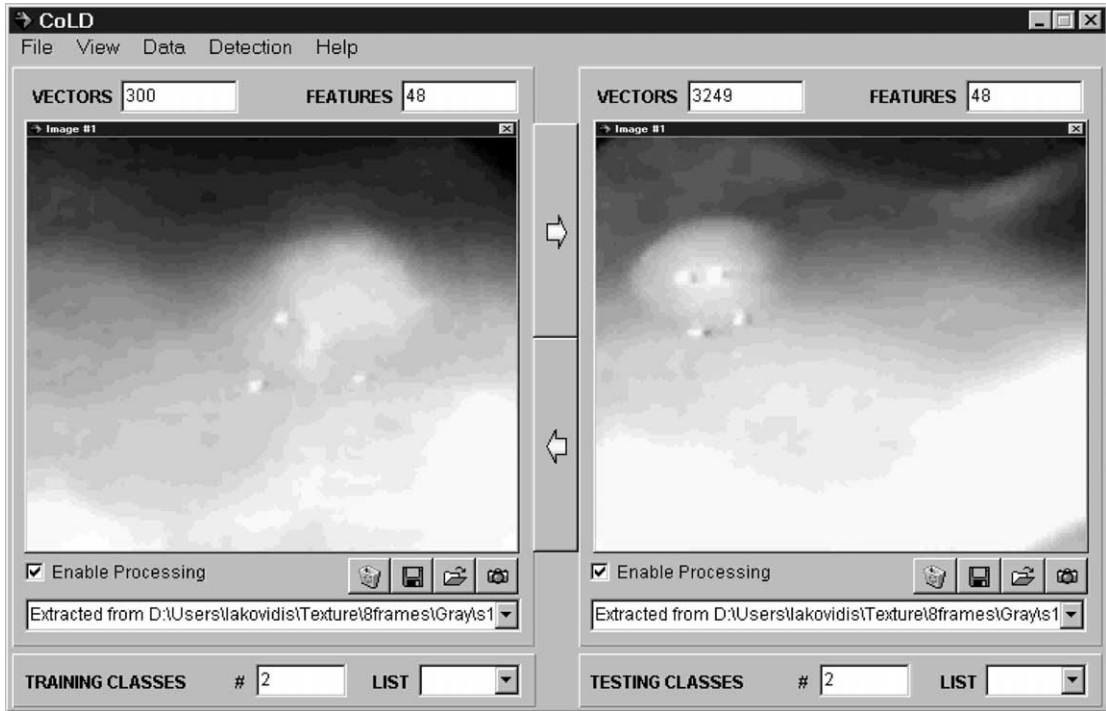


Fig. 4. CoLD's primary panel. Training images are presented on the left and the testing images for evaluation are presented on the right.

- The *primary panel* that is the main window of the application (Fig. 4), controls the arithmetic processing functions of IDMb and handles the management of the input images and features.
- The *feature extraction panel* (Fig. 5), controls the application of the feature extraction algorithms on the input images, that are available in the IDMb.
- The *MLP panel* (Fig. 6), controls the training and evaluation functions included in the Cb.

The program is supported by help options, including extensive online documentation, as well as tips and explanatory legends that pop up over the graphical controls when the mouse pointer passes above.

3.2.1. Primary panel

This panel has been developed aiming to the display of the input frame sequences as well as to the management of the corresponding features. It is organized in two sections. The first section,

which is located on the left, has been assigned to manage and display the training data set. The second is located on the right and has been assigned to manage and display the testing data set,

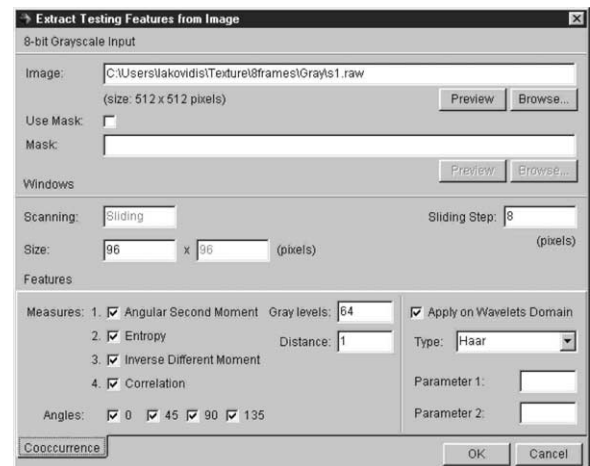


Fig. 5. CoLD's feature extraction panel.

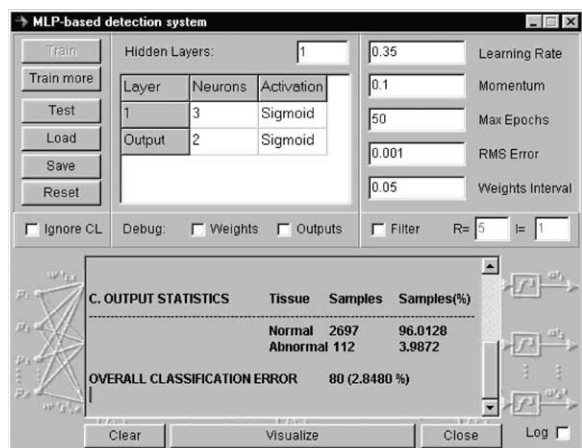


Fig. 6. CoLD's MLP panel.

which is intended for evaluation. Initially only the input images are visible. In case more frames are present in the input, the user is capable of altering the displayed frames in sequence.

The IDMB processing functions are available as menu options and may be applied on both or on each data set independently. Numeric indicators providing information on the contents of each set, including the total number of feature vectors, vectors' dimension, the image-file sources and the class labels of vectors, are displayed on each section of the primary panel.

3.2.2. Feature extraction panel

The image feature extraction process can be controlled through this panel, which provides various options regarding the selection of the feature set. The available parameters that can be set or adjusted through this panel are much more than the parameters a physician would actually use to extract image features. These include, the type of statistical measures to be used, the type of wavelet transformation to be applied, the number of training vectors to be involved, the depth of the image as well as the size of the window sample to be used. This aspect of the program can prove to be very useful for researchers working in the field of medical image analysis. By changing the default settings, which have been optimally decided after extensive experimentation, they can actually experiment on various images and apply this program in many pattern recognition problems.

3.2.3. MLP panel

This panel can be considered as the interface between the user and the Cb through which the resulted frames indicating the evaluation results can be obtained. It provides various options to setup the MLP architecture as well as options to adjust the training parameters. These parameters are related to the number of hidden layers, the number of neurons in each layer and their activations, the learning rate, the momentum, the number of BP training iterations, the value of the expected classification error threshold and the initial weights' interval. The incorporation of these parameters increases the flexibility and the adaptability of the system in different diagnostic environments. Different endoscopes usually produce significantly different images, due to their different output properties. Differences can also be noticed between different VCRs, and frame grabbers. The users are able to setup a fine-tuned recognition system, which will meet their own needs. The trained networks can be retrained repeatedly. Once training is over, the MLP structure can be saved in binary format, and loaded on user's demand.

After the evaluation process, the output images are displayed on a separate window, which facilitates the visual comparison between the input and output images. In addition post-processing options are also available on the same panel.

3.3. Implementation of image features extraction

The extraction of textural features from colonoscopy images or video-frames aims to the construction of textural signatures of image regions. It is based on the estimation of the statistical texture measures on the wavelet domain of the image. The regions are considered as square image windows of $w \times w$ size. Each of these regions is transformed by the corresponding wavelet basis for 1-level decomposition resulting to bands 2, 3 and 4. The cooccurrence matrix is calculated for each band and the 16 statistical measures are evaluated for each of the three bands. The total of 48 textural values constructs a feature vector for the examined region.

The feature extraction algorithm can be applied for the extraction of training or testing feature sets, which will be used for system’s training or the evaluation of the corresponding images, respectively (Fig. 7). The training set, is calculated on a limited set of regions that are randomly

chosen and encountered, while the testing set, is calculated using overlapping windows, raster-scanning the whole image. The distance between two neighboring windows is STEP pixels. Any examined region is assigned to a class, corresponding to the normal or abnormal case, and

DEFINITIONS	
X	Horizontal image size (pixels)
Y	Vertical image size (pixels)
(x, y)	Pixel coordinates within current image
w	Window size (pixels)
STEP	Number of pixels to step
COUNT	Temporary variable used as counter
TOTAL	Total number of training vectors
CALC VECTOR	Calculate feature vector using cooccurrence matrix analysis

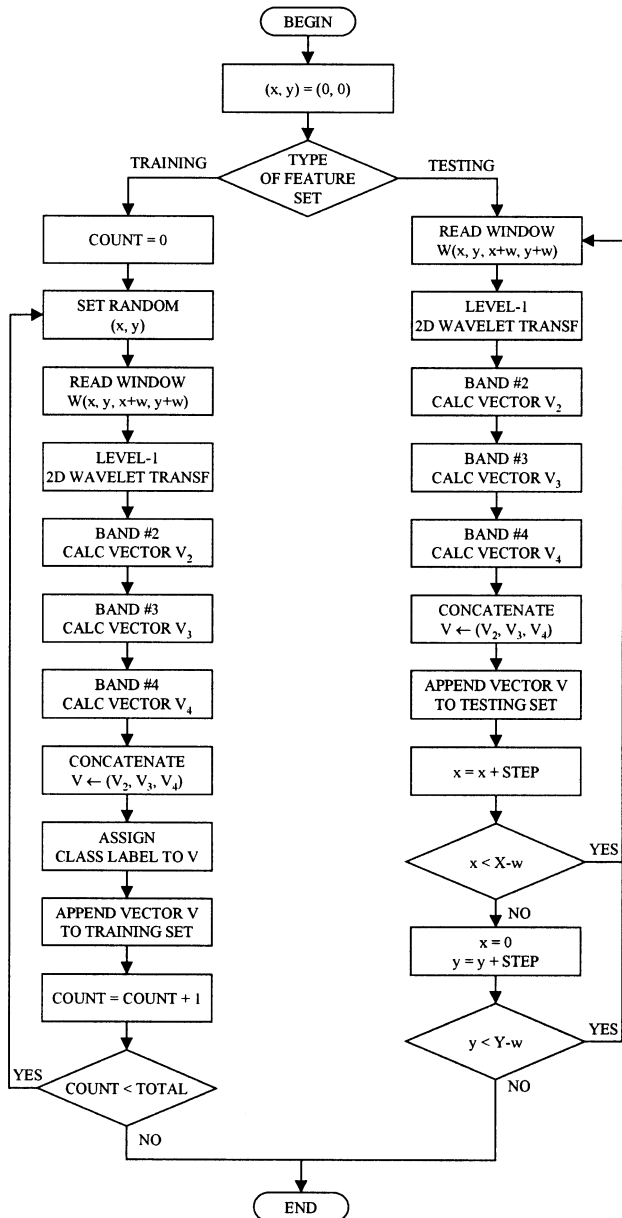


Fig. 7. Flowchart of the feature extraction algorithm.

this information is added as the 49th component of each vector. The class information is obtained by a reference image, which displays the different image regions in different color.

Using the interface described in Fig. 5, the user can adjust a number of parameters for experimentation as well as for a fine-tuning of the system. Such parameters are: the components of the feature vectors, the wavelet type, the angles and the distance for which the cooccurrence matrix will be calculated, the number of training feature vectors, the size of the scanning window, the step of the scanning procedure and the number of gray levels, which also determines the dimension of the cooccurrence matrix.

3.4. Training and evaluation

Cb provides a configurable MLP architecture, where the number of hidden layers and the number of neurons in each hidden layer can be defined by the user. Moreover, the type of activation function of each neuron and the type of connectivity among layers as being full or partial are user defined. The MLP implementation in CoLD is based on Intel's RPL (Recognition Primitives Library), which is optimized for Intel Pentium processors. Four highly compact coded routines are used in this implementation accomplishing fast training and testing results (for initialization of the MLP structure, training, evaluation of inputs, and finally the deallocation of the MLP structure) [41]. The number of inputs comprising the input layer is equal to the number of features of each feature vector. The input features are normalized within the range of $[0, 1]$. For each feature vector x_j presented in the input layer, an output vector is generated in the output layer. The number of output neurons is equal to the number of classes defined in the training data set. The k th of the M possible classes to which the input vector can belong is denoted by C_k . The network is trained with binary target values:

$$d_{kj} = \begin{cases} 1 & \text{when input belongs to class } C_k \\ 0 & \text{when input does not belong to class } C_k \end{cases}$$

Based on this notation, class C_k , is represented by the M -dimensional target vector $[0, \dots, 1, \dots, 0]^T$, where 1 is located at the k th position.

The MLP can be trained and retrained repeatedly. During training, the training parameters can be changed. The MLP structure that has given the best success rate can be saved in binary format, and loaded on user's demand.

Testing features corresponding to an image or a sequence of images to be evaluated are presented in the input of a trained MLP. Each output vector is classified upon the index of its maximum component. The maximum component is considered as 1 and the rest as 0, according to the representation of C_k .

The output images are constructed by overlapping windows corresponding to the windows that were used for the extraction of the testing feature set of feature vectors. Each of these windows is colored upon its class label. Windows corresponding to image regions illustrating abnormal tissue are colored white and the rest are colored black. These images are then laid over the original images. White areas over the original image designate regions of the image corresponding to abnormal tissue. The output results can be further optimized following a post-processing stage.

Fig. 8, illustrates input and output images, respectively, after training the system with the image designated with the arrow. Image dimensions are 512×512 pixels. The features have been calculated with a window size of 96×96 pixels, with values quantized at 64 gray levels. For the 2D-DWT the Haar type of wavelet have been used. The training set was consisted of only 150 48-dimensional feature vectors. The number of testing feature vectors that have been extracted for each testing image was 2809, for $STEP = 8$ pixels ($96 - 8 = 88$ pixels overlap between two neighboring windows). Thus, the total number of testing feature vectors is $8 \times 2809 = 22472$. It is worth noting that the training set is approximately 0.7% of the testing set. The classification of these vectors was performed by a 48-3-2 MLP. For each neuron, the logistic activation function:

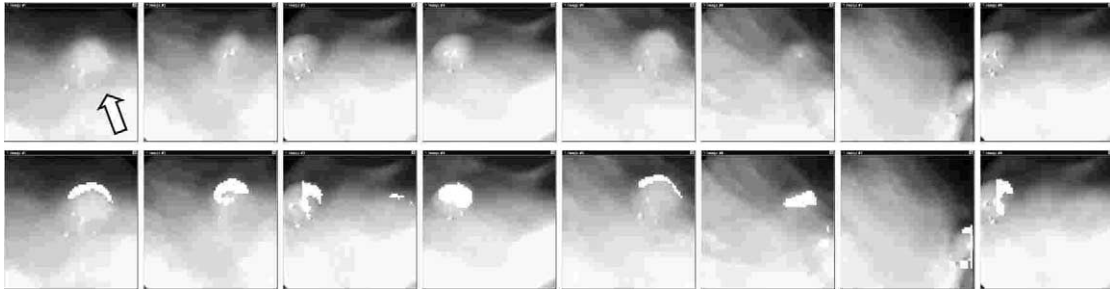


Fig. 8. Original (upper row) and output (lower row) frames indicating the results of detection.

$$\phi_j(u_j(n)) = \frac{1}{1 + \exp(-\alpha u_j(n))},$$

$$\alpha > 0 \text{ and } -\infty < u_j(n) < \infty$$

was used, where $u_j(n)$ is the weighted sum of all synaptic inputs plus bias, of neuron j at the n th iteration and α is the momentum constant. The learning rate, the momentum and the initial weights interval that have been applied were set $\eta = 0.35$, $\alpha = 0.1$ and 0.05 , respectively. The MLP was trained after 48 training epochs; i.e. complete presentations of the entire training set to the MLP.

The overall average classification error for all 8 frames of Fig. 8 is 4.02% with a standard deviation of 2.26%. Table 2, presents the confusion matrix of CoLD's average classification results for these frames, comparing to the physician's characterization. As it can be derived by this table, the percentage of the image regions corresponding to normal tissue that are characterized as abnormal by the system, is relatively higher than the percentage of vectors corresponding to abnormal tissue that are characterized as normal by the system. This fact indicates that the probability of false alarms is higher than the probability of neglecting regions corresponding to suspicious lesions. Despite the classification rates, it is worth noting that the system has spotted accurately the location of the abnormal regions on each image, regardless the classification percentage, which actually quantifies the matching between the resulted images and the reference images that have been used to assess the system's classification performance.

3.5. Post-processing

Isolated misclassified vectors or groups of vectors may appear as spots or small oblong formations within the output image. A post-processing stage following the classification process performed by the MLP, appears to be necessary in most cases, to minimize this type of noise in the output images.

Experimentation on noisy output images has shown that the application of a majority voting decision criterion produces significantly clearer output. The output image overlay is constructed by overlapping windows of intensity determined by the class labels assigned by the MLP. This property can be used to strengthen the characterization of these image regions.

Class labels of each window from the upper left to the lower right corner of the image are sequentially put under evaluation. The class label of each window is determined by the class label of its neighbors Fig. 9. The new class label (i), is assigned to be equal to the class label of the majority of its neighbors n_{rk} within a radius of N , i.e. $r \leq N$, $k = 1, 2, \dots, 8r$. The system uses default

Table 2
CoLD's characterization results compared to physician's characterization results

Physician	CoLD	
	Normal (%)	Abnormal (%)
Normal	93.99	6.00
Abnormal	2.79	97.21

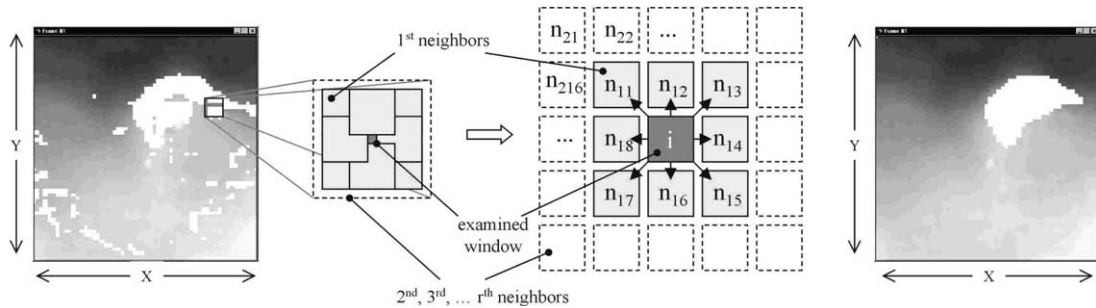


Fig. 9. CoLD's output image before (left) and after (right) post-processing procedure (middle).

values of $N = 2$, but can be also adjusted on user's demand, depending on window size and on the STEP value.

4. Implementation, requirements and performance

CoLD's system prototype was primarily developed in MATLAB command line interface. Promising results of the applied methodology on still images and in sequences of video-images [8–13], led to the integration of this experimental setup into a standalone application, which is now able to process more images in less time.

The latest version of CoLD that is presented in this paper, was developed in Borland C++ Builder 3, exploiting the graphical objects it provides for the development of the user interface. Key computational tasks, were implemented using the functions provided by Intel's performance libraries [42], as described previously in this paper. These libraries are dynamically linked to the main program (as DLLs) and they provide functions that are optimized for Intel Pentium processors. The display performance, which is important for the playback of the input or the output video-frames, were implemented using DirectX functions which exploit the capabilities of the display hardware [43].

The minimum hardware requirements are mainly imposed by the requirements of Intel's libraries. An Intel Pentium processor PC with MMX support, 32MB RAM and 10MB of hard disk space can be considered as the least hardware specifications for CoLD's application. The operat-

ing system required is a Windows 98/ME or Windows NT4/2000 with DirectX support.

CoLD has been tested on an Intel Pentium III 550 MHz processor PC, with 256MB RAM. The video-frames that have been used for CoLD's performance evaluation are shown in Fig. 8. Training and testing feature sets are considered for a single frame. The parameters that have been used for training and evaluation are described in the preceding section. The time measurements corresponding to the feature extraction processes are presented in Table 3.

Training and evaluation times were measured for the 48-X-2 MLP, where $X \in \{3, 5, 10, 15, 20, 25, 30, 35, 40\}$ and for 100 training epochs. The performance results in milliseconds versus the number of hidden neurons X used are presented in Fig. 10. As it can be observed the computational complexity for both training and evaluation processes is linear in the number of hidden units that are used, since the experimental study presents an MLP architecture with 1-hidden layer [37]. All measurements do not include any I/O operations.

Table 3
CoLD's performance in feature extraction

Training/Testing set	One frame time measurements (s)
150 vectors	1.94
2809 vectors	34.65

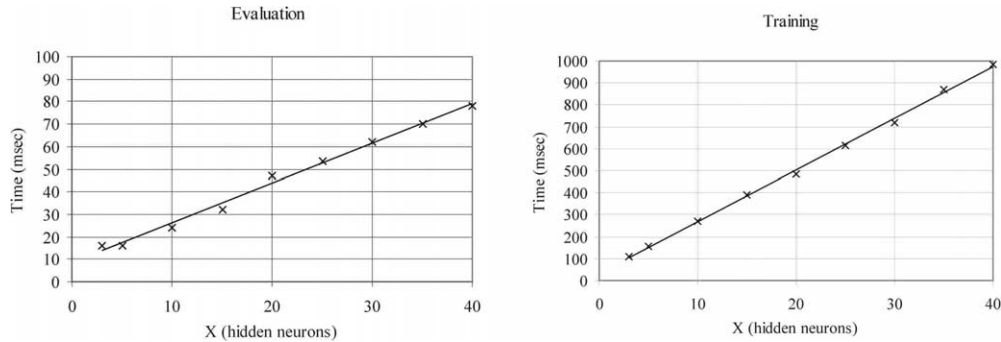


Fig. 10. CoLD's performance for training and evaluation with a single video-frame.

5. Conclusion and prospects

This paper describes CoLD, a system that supports medical diagnosis of colorectal lesions using images or video-frames acquired during colonoscopy. The major contribution of CoLD in the process of medical diagnosis, is that it can provide additional information to physicians on the characterization of the colonic mucosal surface, by exploiting its textural characteristics, which are consequently used for the classification of the corresponding image regions as normal or abnormal. All the functions included can be easily applied through the options provided by the system's user interface. CoLD offers to novice users the capability to use it directly with predefined parameters but also offers to expert users options to tune all the relative parameters of the algorithms applied, for research needs. Two weeks examination data have been used for extensive testing of CoLD on single images as well as on colonoscopy video frame sequences. The average success rate in the detection of lesions exceeds 95%, without retraining needed for the same types of lesions used for training, even when tested on different patients. Possible future extensions of CoLD can be summarized as:

(1) Study for optimization and parallelization of the time consuming functions, which will be executed on a multiprocessing environment. Such functions are mainly those that are related to the extraction of the textural features, which can be identically applied on different parts of the images

or even on different frames by the processors concurrently. Such an implementation will add to CoLD characteristics for a near to real time application.

(2) Incorporation of different feature extraction and classification algorithms that could provide higher detection success rates.

(3) Integration of CoLD with standard endoscopy equipment that is widely used by the specialists or even with similar equipment that will be built in the future.

6. Availability of the software

A demonstration version of the presented software is available for downloading from our web site <http://rtsimage.di.uoa.gr/download.htm>.

Acknowledgements

This work was partially funded by National and Kapodistrian University of Athens, Special Account of Research Grants. We would like to acknowledge Prof. M. Tzivras M.D., University of Athens, Medical School, Department of Pathophysiology, Gastroenterology Section, for the provision of endoscopy videos. In addition, we are most grateful to Dr M.O. Schurr M.D., University of Tübingen Section for MIC, also for the provision of endoscopy videos and their effort on the evaluation of CoLD.

References

- [1] L.J. Mango, Computer-assisted cervical cancer screening using neural networks, *Cancer Lett.* 77 (1994) 155–162.
- [2] G.D. Tourassi, C.E. Floyd, H.D. Sostman, R.E. Coleman, Acute pulmonary embolism: artificial neural network approach for diagnosis, *Radiology* 189 (1993) 555–558.
- [3] O. Tsujii, M.T. Freedman, S.K. Mun, Classification of microcalcifications in digital mammograms using trend-oriented radial basis function neural network, *Pattern Recogn.* 32 (1999) 891–903.
- [4] 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* 198 (1996) 671–678.
- [5] D. Rex, R. Weddle, D. Pound, K. O'Connor, R. Hawes, R. Dittus, J. Lappas, L. Lumeng, Flexible sigmoidoscopy plus air contrast barium enema versus colonoscopy for suspected lower gastrointestinal bleeding, *Gastroenterology* 98 (1990) 855–861.
- [6] S. Nagata, S. Tanaka, K. Haruma, M. Yoshihara, K. Sumii, G. Kajiyama, F. Shimamoto, Pit pattern diagnosis of early colorectal carcinoma by magnifying colonoscopy: clinical and histological implications, *Int. J. Oncol.* 16 (2000) 927–934.
- [7] S. Kudo, H. Kashida, T. Tamura, E. Kogure, Y. Imai, H. Yamano, A.R. Hart, Colonoscopic diagnosis and management of nonpolypoid early colorectal cancer, *World J. Surg.* 24 (2000) 1081–1090.
- [8] 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* 9 (2000) 225–230.
- [9] D.A. Karras, S.A. Karkanis, V. Mertzios, Information systems based on neural network and wavelet methods with application to decision making, modeling and prediction tasks, *Kybernetes* 27 (1998) 224–236.
- [10] S.A. Karkanis, D.K. Iakovidis, D.A. Karras, D.E. Maroulis, Detection of lesions in endoscopic video using textural descriptors on wavelet domain supported by artificial neural network architectures, accepted to IEEE International Conference in Image Processing (ICIP)-Proceedings, Paper Number 4815, Thessaloniki, Greece, 2001.
- [11] S.A. Karkanis, G.D. Magoulas, D.K. Iakovidis, D.A. Karras, D.E. Maroulis, Evaluation of textural feature extraction schemes for neural network based interpretation of regions in medical images, accepted to IEEE International Conference in Image Processing (ICIP)—Proceedings, Paper Number 4673, Thessaloniki, Greece, 2001.
- [12] 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, 2000, pp. 423–429.
- [13] S. Karkanis, G.D. Magoulas, M. Grigoriadou, M. Schurr, Detecting abnormalities in colonoscopic images by textural description and neural networks. Advance Course in Artificial Intelligence (ACAI)—Proceedings, Workshop on Machine Learning in Medical Applications, Chania, Greece, 1999, pp. 59–62.
- [14] K.M. Hornik, M. Stinchcombe, H. White, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, *Neural Networks* 3 (1990) 551–560.
- [15] K.M. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, *Neural Networks* 2 (1989) 359–366.
- [16] A.R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, *IEEE Trans. Informat. Theory* 39 (1993) 930–945.
- [17] C.P. Lim, R.F. Harrison, R.L. Kennedy, Application of autonomous neural network systems to medical pattern classification tasks, *Artificial Intelligence Med.* 11 (1997) 215–239.
- [18] A.S. Miller, B.H. Blott, T.K. Hames, Review of neural network applications in medical imaging and signal processing, *Med. Biological Eng. Comput.* 30 (1992) 449–464.
- [19] A. Latif-Amet, A. Ertuzun, A. Ercil, An efficient method for texture defect detection: sub-band domain co-occurrence matrices, *Image Vision Comput.* 18 (2000) 543–553.
- [20] A.A. Kassim, M.A. Mannan, M. Jing, Machine tool condition monitoring using workpiece surface texture analysis, *Machine Vision Applicat.* 11 (2000) 257–263.
- [21] C. Chiu, A novel approach based on computerized image analysis for traditional Chinese medical diagnosis of the tongue, *Comp. Meth. Prog. Biomed.* 61 (2000) 77–89.
- [22] A. Bodnarove, M. Bennamoun, K.K. Kubik, Suitability analysis of techniques for flaw detection in textiles using texture analysis, *Pattern Analysis and Applications* 3 (2000) 254–266.
- [23] D.G. Vince, K.J. Dixon, R.M. Cothren, J.F. Cornhill, Comparison of texture analysis methods for the characterization of coronary plaques in intravascular ultrasound images, *Computerized Medical Imaging and Graphics* 24 (2000) 221–229.
- [24] T. Sziranyi, M. Csapodi, Texture classification by cellular neural networks using genetic learning, *Computer Vision and Image Understanding* 71 (1998) 255–270.
- [25] M. Unser, Texture classification and segmentation using wavelet frames, *IEEE Trans. Image Processing* 4 (1995) 1549–1560.
- [26] M. Hauta-Kasari, J. Parkkinen, T. Jaaskelainen, R. Lenz, Multi-spectral texture segmentation based on the spectral cooccurrence matrix, *Pattern Analysis and Applications* 2 (1999) 275–284.
- [27] B. Julesz, Texton gradients: the texton theory revisited, *Biol. Cybern.* 54 (1986) 245–251.
- [28] A. Rao, G.L. Lohse, Identifying high level features of texture perception, *CVGIP: Graphical Models and Image Processing* 55 (1993) 218–233.

- [29] M. Vanrell, J. Vitria, X. Roca, A multidimensional scaling approach to explore the behavior of a texture perception algorithm, *Machine Vision and Applications* 9 (1997) 262–271.
- [30] A. Cohen, R.D. Ryan, *Wavelets and Multiscale Signal Processing*, Chapman & Hall, London, 1995.
- [31] F. Abramovich, T. Bailey, T. Sapatinas, Wavelet analysis and its statistical applications, *The Statistician* 49 (2000) 1–29.
- [32] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Anal. Machine Intelligence* 11 (1989) 674–693.
- [33] Y. Meyer, *Wavelets: Algorithms and Applications*, SIAM, Philadelphia, PA, 1993.
- [34] I. Daubechies, *Ten lectures on wavelets*. Philadelphia Society for Industrial and Applied Mathematics, 1992.
- [35] R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, *IEEE Trans. Syst., Man Cybernet.* 3 (1973) 610–621.
- [36] R.M. Haralick, Statistical and structural approaches to texture, *IEEE Proc.* 67 (1979) 786–804.
- [37] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd edn, Prentice Hall, New Jersey, 1996.
- [38] B. Myers, S.E. Hudson, R. Pausch, Past, present and future of user interface software tools, *ACM Trans. Human Computer Interact.* 7 (2000) 3–28.
- [39] A. Berrais, Knowledge-based expert systems: user interface implications, *Adv. Eng. Software* 28 (1997) 31–41.
- [40] B. Shneiderman, *Designing the User Interface: Strategies for Effective Human–Computer Interaction*, third edn, Addison-Wesley, Massachusetts, 1998.
- [41] *Intel Recognition Primitives Library Reference Manual*, Intel Corporation, USA, 1999.
- [42] Intel Performance Libraries (WWW page), <http://developer.intel.com/software/products/perflib/>, Accessed: May, 2001.
- [43] Microsoft DirectX Home (WWW page), <http://www.microsoft.com/directx/>, Accessed: May, 2001.