Detection of Small Bowel Tumors in Capsule Endoscopy Frames Using Texture Analysis based on the Discrete Wavelet Transform

Daniel J. C. Barbosa, Jaime Ramos, Carlos S. Lima

***Abstract*— Capsule endoscopy is an important tool to diagnosis tumor lesions in the small bowel. The capsule endoscopic images possess vital information expressed by color and texture. This paper presents an approach based in the textural analysis of the different color channels, using the wavelet transform to select the bands with the most significant texture information. A new image is then synthesized from the selected wavelet bands, trough the inverse wavelet transform. The features of each image are based on second-order textural information, and they are used in a classification scheme using a multilayer perceptron neural network. The proposed methodology has been applied in real data taken from capsule endoscopic exams and reached 98.7% sensibility and 96.6% specificity. These results support the feasibility of the proposed algorithm.**

I. Introduction

THE conventional endoscopy suffers from some limitations, since it is limited to the upper gastrointestinal (GI) tract, at the duodenum, and to lower GI tract, at terminal ileum. The medium length of the small bowel is six meters, so a very significant part of the small intestine is not seen by these conventional techniques. In GI tract, great skill and concentration are required for navigating the conventional endoscope because of its flexible structure. Discomfort to the patient and the time required for diagnosis heavily depend on the technical skill of the physician and there is always a possibility of the tip of the endoscope injuring the walls[1]. In 2000, the development of capsule endoscopy (CE) opened a new chapter in small bowel examination, allowing the visualization of the entire GI tract, reaching places where conventional endoscopy is unable to. CE is a simple, non-invasive procedure that is well accepted by the patient and can be performed on an outpatient basis. The PillCamTM SB videocapsule (Given Imaging Ltd., Yoqneam, Israel) is a wireless capsule (11mm×26 mm), which contains a miniaturized camera, a light source and a wireless circuit for the acquisition and transmission of signals[2]. The result is a seven hours video with more than 50.000 frames per exam. Average small bowel transit time is about 90 minutes [2]. The time required to a physician to analyze the resulting video is, on average, 40-60 min[3]. Since this task requires complete concentration from the reader, being, nevertheless, prone to errors, and is time consuming, there is the need to develop computer systems to support the medical diagnosis.

Manuscript received April 7, 2008.

C. S. Lima and Daniel J. C. Barbosa are with the Industrial Electronics Department, Minho University, Portugal.

Jaime Ramos is with the Gastroenterology Department, Hospital dos Capuchos, Portugal.

A small bowel tumor is diagnosed in approximately 2.5–9% of patients submitted to CE, indicating that the frequency of these neoplasms is considerably higher than was previously thought. At least 50% of small intestine tumors identified with CE are malignant [3].

The automatic detection of abnormalities can be based in alterations in the texture of the small intestine mucosa. Maroulis *et al.*[4][5] proposed two different methods based in the analysis of textural descriptors of colonoscopy videos wavelet coefficients. The first uses second-order statistical features that are calculated on the wavelet domain of each image, at the bands 1,2,3 of the wavelet transform. The second is based on the covariance of second-order textural measures in the wavelet domain, namely in the bands 4,5,6. Kodogiannis *et al.*[1] proposed two different schemes to extract features from texture spectra in the chromatic and achromatic domains, namely a structural approach based in the theory of formal languages, where a textured image is considered a sentence in a language, of which the alphabet is a set of texture primitives called textons, constructed in accordance with a certain grammar determining the layout of such texture primitives within a pattern. In Kodogiannis *et al.*[1] work is also proposed a statistical approach, where statistical descriptors are calculated from the histograms of the RGB and HSV color spaces of CE video frames.

The Multi-Layer Perceptron (MLP) networks are commonly used in classification problems, because they have the ability to detect complex non-linear relationships in the data. There is an extensive range of applications of these neural networks, and so, a vast theoretical and practical background in this subject[6].

The proposed methodology focus the feature extraction process from the endoscopic capsule video frames, with a method based in statistical descriptors of cooccurrence matrix calculated for an image reconstructed from the wavelet coefficients of the selected bands, being tested different wavelet bands to evaluate where is the most significant texture information for classification purposes. The proposed approach is performed in the color spaces RGB and HSV. These features are the input of a MLP network, in a classification scheme used to classify real data from Hospital dos Capuchos patients.

II. Features Extraction

The proposed method relies on a color textural features extraction process based in textural analysis. Since the low-frequency components of the images do not contain major texture information, the most important bands in the wavelet transform are those in which are present medium and high frequency, texture encoding, information. To reduce the final number of features, a new image is synthesized from the selected wavelet coefficients, where the new image contains only the vital texture information from the selected wavelet bands. Thus the synthesized image contains the significant textural information present in the wavelet decomposed sub-images. To evaluate if the most significant textural information for classification purposes is encoded as high frequencies (lowest scale) or as medium frequencies (middle scales), the proposed algorithm was implemented selecting different wavelet bands. The proposed features set is calculated over the cooccurrence matrix calculated from the new image synthesized from the selected wavelet coefficients, for every color channels. These features are statistical descriptors that contain second order color level information captured from the cooccurrence matrix, which are mostly related to the human perception and discrimination of textures. The cooccurrence matrix encodes the synthesized image level (for each color channel) spatial dependence based on the estimation of the second order joint-conditional probability density function *f(i,j,d,θ)*, which is computed by counting all pairs of pixels at distance *d* having wavelet coefficients of color levels *i* and *j* at a given direction *θ*. The angular displacement used is the set {0, π/4, π/2, 3π/4}. In the proposed algorithm are considered only 4 statistical measures among the 14 originally proposed by Haralick [7]. They are angular second moment (F1), correlation (F2), inverse difference moment (F3), and entropy (F4), representing the homogeneity, directional linearity, smoothness and randomness of the cooccurrence matrix, defined respectively as:

$$F1 = \sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j)^2. \qquad (1)$$

$$F2 = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} (i.j)\,p(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y}. \qquad (2)$$

$$\mu_x = \sum_{i=1}^{N} i \sum_{j=1}^{N} p(i,j). \qquad (2a)$$

$$\mu_y = \sum_{j=1}^{N} j \sum_{i=1}^{N} p(i,j). \qquad (2b)$$

$$\sigma_x = \sum_{i=1}^{N} (i-\mu_x)^2 \sum_{j=1}^{N} p(i,j). \qquad (2c)$$

$$\sigma_y = \sum_{j=1}^{N} (j-\mu_y)^2 \sum_{i=1}^{N} p(i,j). \qquad (2d)$$

$$F3 = \sum_{i=1}^{N}\sum_{\substack{j=1\\ i\neq j-1}}^{N} \frac{1}{1+(i-j)}\,p(i,j). \qquad (3)$$

$$F4 = \sum_{i=1}^{N}\sum_{\substack{j=1\\ p(i,j)\neq 0}}^{N} p(i,j)\log_2 p(i,j). \qquad (4)$$

where *p(i,j)* is the *ijth* entry of normalized cooccurrence matrix, *N* the number of levels of the synthesized image and μ_x, μ_y, σ_x, σ_y are the means and standard deviations of the marginal probability $p_x(i)$ obtained by summing up the rows of the matrix p(i,j).

The proposed algorithm can be decomposed in the following categories:

A. Wavelet coefficients selection and new image synthesis

Each capsule endoscopic video frame can be decomposed in the three color channels:

$$I^i, \qquad i = 1,2,3. \qquad (5)$$

where *i* stands for the color channel.

These three color channels are originally in the RGB color space, but can be transformed to other color spaces. In this paper is presented the comparison between RGB and HSV color spaces results to the proposed algorithm. Then a two level discrete wavelet transformation is applied to each color channel (I^i). This transformation results in a new representation (W^i) of the original image by a low resolution image and the detail images. The wavelet bases used were the Daubechies bases. The new representation is defined as:

$$W^i = \{L_n^i, D_l^i\}, \quad i = 1,2,3 \quad l = 1,\ldots,6. \quad (6)$$

where *l* stands for the wavelet band and *n* is the decomposition level.

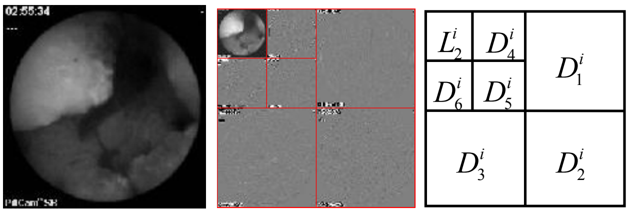

Fig. 1. Example of two level wavelet decomposition scheme of the original image for color channel *i*.

Since is not clear at which wavelet scale is the texture information in capsule endoscopic frames, at this point the algorithm can select different wavelet bands, with the coefficients corresponding to a given scale. The different selected bands were used to estimate in which scale is the most significant textural information for classification of the capsule endoscopic frames. Therefore, let S^i be a matrix that has the selected wavelet coefficients at the corresponding positions and zeros in all other positions:

$$S^i = \{D_l^i\}, \quad i = 1,2,3 \quad l = 1,2,3 \vee l = 4,5,6. \quad (7)$$

where *l* stands for the wavelet band and *i* for the color channel. Note that *l* depends of the selected wavelet scale.

A new image is then synthesized from the selected wavelet bands, trough the inverse wavelet transform. Let N^i be the reconstructed image, for each color channel:

$$N^i = IDWT(S^i), \qquad i = 1,2,3. \quad (8)$$

where i stands for color channel and *IDTW()* is the inverse wavelet transform.

For each capsule endoscopic frame a new image is calculated, reconstructed from the selected wavelet bands, with the essential textural information of the original image.

B. Cooccurrence matrix and statistical descriptors

Statistical measurements based on second-order statistics can be used as textural features. These statistical descriptors were estimated using the cooccurrence matrices of each reconstructed image. For the extraction of the second order statistical textural information, cooccurrence matrices were calculated for the different color channels. These matrices capture spatial interrelations among the intensities within the reconstructed image levels and represent the spatial distribution dependence of the gray levels within an image, determining how often different combinations of pixel brightness values occur in an image. The cooccurrence matrices are estimated in four different directions α:

$$C_\alpha\left(N^i\right) \quad i = 1,2,3 \quad \alpha = 0, \frac{\pi}{4}, \frac{\pi}{2}, 3\frac{\pi}{4}. \quad (9)$$

where i stands for the color channel and α for the direction in the cooccurrence computation.

Four statistical measures given by equations (1), (2), (3), (4) are estimated for each matrix C resulting in 48 statistical descriptors:

$$F_m\left(C_\alpha\left(N^i\right)\right) \quad i=1,2,3 \quad \alpha = 0, \frac{\pi}{4}, \frac{\pi}{2}, 3\frac{\pi}{4} \quad m=1,2,3,4. \quad (10)$$

where m stands for statistical measure.

Then the mean and variance for each F_m is calculated over α, for every color channels, resulting in a set of 24 components per frame, which constitute the input of the MLP network.

III. Multilayer Perceptron Neural Network

The classification scheme described in this paper used a standard MLP network, with 24 input neurons, 2 output neurons (normal and tumor) and a variable number of neurons in the hidden layer (6, 8 and 12). The training algorithm was the well known back propagation learning process, in which the values of each connection are adjusted in order to reduce the value of the error function. The two output neurons were used to classify the data into 2 classes, normal and tumor. The choice of a simple classification scheme was done to make the results more representative of the effectiveness of the proposed algorithm than of the classifier itself.

IV. Implementation and Results

The experimental dataset consisted of capsule endoscopic video segments of different patients' exams, taken at the Hospital dos Capuchos in Lisbon by Doctor Jaime Ramos. The training set was constructed with images from normal segments of capsule endoscopic videos, some of them taken from exams with pathological cases. The tumor images were taken from capsule endoscopy exams with this pathology. The final training dataset was composed by 100 normal images and 104 tumor images. The final test dataset was composed by 100 normal images and 92 tumor images.

Fig. 2. Example of a normal intestinal tissue frames

Fig. 3. Example of a tumor intestinal tissue frames

In order to compute the cooccurrence matrix for the new image, synthesized from the wavelet coefficients from the selected bands, a new algorithm was implemented, to avoid computing cooccurrences in the image corners where no image information exists. The cooccurrence computation was done considering d=1. A similar algorithm was also developed to calculate the histograms of each frame.

A 3.2 GHz Pentium Dual Core processor-based with 1 GB of RAM was used with MATLAB to run the proposed algorithm. The average processing time per frame is about 1 minute, which is unacceptable to real world applications. However, in the work of Arvis *et al.*[8], there is the reference that the reduction of the gradation levels of each color channel from 256 levels to 32 levels does not compromise the texture analysis process. Therefore the processing time per frame drops considerably, to about 1 second per frame, without significant loss of performance. However the vast majority of the pixels in the reconstructed image have a level very close to zero, so the most of the information is included in a few, very close, levels, which will lead to a loss of texture information, as very close levels in the 256 levels image are converted to the same level in the 32 levels image. To overcome this limitation, we have to disperse the pixel values to all available range with a simple multiplication by a constant. Therefore the textural information will be present in all the 256 gray levels, and consequently in all the 32 gray levels, after the conversion.

In the HSV color space, the Hue range of the analyzed endoscopic capsules frames is very restricted, as we can see in figure 4. So, and for this particular method, where the Hue range is very limited, we propose the dispersion of the H color channels pixel values (*H=H* x *k*) to all the available range. With this simple operation, we will be able to distinguish small variations in H, without loss of information, since there is an appropriated dispersion of the H information to all the available levels. The practical validation of this principle can be observed in table 2.

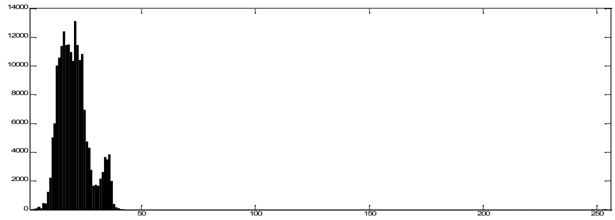

Fig. 4. H values range in the dataset.

To evaluate at which wavelet scale is the most significant textural information to classification purposes, different versions of the proposed algorithm were tested, reconstructing the new image from different wavelet bands, as stated in (7). An assessment to the performance of the proposed method at different color spaces was considered to the RGB and HSV color spaces, as seen in table 1. Note that in this table the H dispersion is not considered.

Instead of measuring the rate of successful recognized patterns, more reliable measures for the evaluation of the classification performance can be achieved by using the sensitivity (true positive rate) and the specificity (100-false positive rate) measures. These two measures can be calculated as:

$$Sensitivity = \frac{d}{c+d}.100\ (\%). \qquad (11)$$

$$Specificity = \left(100 - \frac{b}{a+b}.100\right)(\%). \qquad (12)$$

where *a* are the true negative patterns, *b* are the false positive patterns, *c* are the false negative patterns and *d* are the true positive patterns.

To evaluate the performance of the proposed algorithm, they were considered different architectures in the MLP, varying the number of neurons in the hidden layer. The results in the following tables were achieved for 25 runs (training + simulation for the available dataset) for each MLP architecture, for each space color and selected bands.

TABLE I
CLASSIFICATION PERFORMANCE OF THE PROPOSED ALGORITHM

Color Space Wavelet Bands	RGB 1,2,3	RGB 4,5,6	HSV 1,2,3	HSV 4,5,6
Specificity (μ±σ%)	78.0±2.4	77.2±5.6	94.9±1.1	87.8±4.0
Sensibility (μ±σ%))	90.0±2.8	77.4±3.2	96.9±1.2	86.8±3.6

TABLE II
CLASSIFICATION PERFORMANCE FOR DIFFERENT H DISPERSION(k)

Color Space - HSV Wavelet Bands – 1,2,3	k=1	k=2	k=5	k=10
Specificity (μ±σ%)	94.9±1.1	95.1±1.5	96.6±1.1	96.5±0.9
Sensibility (μ±σ%)	96.9±1.2	97.3±0.7	98.7±0.5	98.1±1.5

V. DISCUSSION AND FUTURE WORK

From the presented results, the most important observation is that the most significant textural information for classification purposes is in the wavelet bands 1,2,3, corresponding to higher frequencies, at the lowest wavelet scale, which leads to the conclusion that the key textural information for classification of small bowel tumors is encoded as high frequency information. In Maroulis *et al.* work [6], are used the wavelet bands 4,5,6, but the endoscopic images have 1024x1024 pixels, so in endoscopic capsule frames (256x256 pixels), similar textures will be encoded as higher frequencies than in conventional endoscopy. In the analyzed dataset, was verified that H values of the pixels of the different frames were restricted to a very limited range, so classification performance increase with H values dispersion, because it lead to greater sensibility to detect small variations in H. Note that in different systems to detect abnormalities in endoscopic capsule videos, the H dispersion may not be valid, since the H range may not be so restricted. Trough the testing process, the color space HSV revealed better performance than the RGB, as theoretically predicted. Future work will include the increase in the available dataset and performance evaluation of the proposed algorithm for different color spaces, with different classification schemes and the development of an automatic abnormalities detection system in capsule endoscopy videos, for the most common CE detectable diseases.

REFERENCES

[1] V. S. Kodogiannis, M. Boulougourab, E. Wadge and J.N. Lygourasc, "The usage of soft-computing methodologies in interpreting capsule endoscopy," *Engineering Applications of Artificial Intelligence,* vol. 20, pp. 539–553, 2007.

[2] G. Idden, G. Meron, A. Glukhovsky and P. Swain, "Wireless capsule endoscopy," *Nature*, pp.415-417, 2000.

[3] M. Pennazio, "Capsule endoscopy: Where are we after 6 years of clinical use?," *Digestive and Liver Disease,* vol. 38, pp. 867–878, 2006.

[4] D. Maroulis, D. Iakovidis, A. Karkanis and D. Karras, "CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames," *Computer Methods and Programs in Biomedicine*, vol. 70, pp. 151-166, 2003.

[5] S. Karkanis, D. Iakovidis, D. Maroulis, D. Karras and M. Tzivras, "Computer-Aided Tumor Detection in Endoscopic Video Using Color Wavelet Features," *IEE Trans. On Information Technology in Biomedicine*, vol. 7, no. 3, Sep. 2003.

[6] S. Haykin, *Neural Networks – A comprehensive foundation*. New York: Mcmillan College Publishing Company, 1994.

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

[8] V. Arvis, C.Debain, M. Berducat and A. Benassi, "Generalization of the Cooccurrence Matrix for color images: Application to colour texture classification," *Image Anal Stereol*, vol.23, pp. 63-72, 2004.