Wireless capsule endoscopic frame classification scheme based on higher order statistics of multi-scale texture descriptors

D. Barbosa[1], J. Ramos[2] and C. Lima[1]

[1] Industrial Electronics Department, University of Minho, Braga, Portugal
[2] Gastroenterology Department, Hospital dos Capuchos, Lisboa, Portugal

Abstract— **The gastrointestinal (GI) tract is a long tube, prone to all kind of lesions. The traditional endoscopic methods do not reach the entire GI tract. Wireless capsule endoscopy is a diagnostic procedure that allows the visualization of the whole GI tract, acquiring video frames, at a rate of two frames per second, while travels through the GI tract, propelled by peristalsis. These frames possess rich information about the condition of the stomach and intestine mucosa, expressed by color and texture in these images. These vital characteristics of each frame can be extracted by color texture analysis. Since texture information is present as middle and high frequency content in the original image, two new images are synthesized from the discrete wavelet coefficients at the lowest and middle scale of a two level Discrete Wavelet Transform of the original frame. These new synthesized images contain essential texture information, at different scales, which can be extracted from statistical descriptors of the coocurrence matrices, which are second-order representations of the synthesized images that encode color and spatial relationships within the pixels of these new images. Since the human perception of texture is complex, a multi-scale and multicolor process based in the analysis of the spatial color variations relationships, is proposed, as classification features. The multicolor texture information is modeled by the third order moments of the texture descriptors sampled at the different color channels. HSV color space is more related to the perceptive human characteristics, therefore it was used in the ambit of this paper. The multi-scale texture information is modeled by covariance of the texture descriptors within the same color channel of the two synthesized images, which contain texture information at different scales. The features are used in a classification scheme using a multilayer perceptron neural network. The proposed method has been applied in real data taken from several capsule endoscopic exams and reaches 94.6% of sensitivity and 93.7% specificity. These results support the feasibility of the proposed algorithm.**

Keywords— **Discrete Wavelet Transform, Texture Analysis, Capsule Endoscopy, Computer Aided Diagnosis**

I. Introduction

Until the introduction of wireless capsule endoscopy, it was not possible to see the gastrointestinal (GI) tract in its entire length, since conventional endoscopy is limited, in upper GI tract endoscopy, at duodenum and at terminal jejunum, in lower GI tract endoscopy. Therefore, the vast majority of the small bowel, which has a medium length of six meters, is not seen by these conventional techniques. Consequently, prior to the invention of CE, the small intestine was the conventional endoscopy's last frontier, because it could not be internally visualized directly or in entirely by any method [1]. Furthermore the conventional endoscopic procedures are uncomfortable to the patient and require advanced technical skills from the operating physician, in order to correctly navigate the flexible endoscope. Note also that there is the risk of injuring the GI tract walls with the tip of the endoscope [2].

The introduction of wireless Capsule Endoscopy (CE) in the clinical practice provided a simple and effective diagnostic tool to observe GI mucosa abnormalities, until then not easily seen by traditional imaging techniques, since GI tract could not be internally visualized directly or in its entire length by any conventional method [1]. The endoscopic capsule is a pill-like device, with only 11mm×26 mm, and includes a miniaturized camera, a light source and a wireless circuit for the acquisition and transmission of signals [3]. The acquired video frames are wireless transmitted to a receiver, which stores them in a hard disk drive. The camera captures images at a rate of two frames per second, for about eight hours, resulting in more than 50.000 video frames per exam [4]. The average time taken by a physician to analyze a capsule endoscopic is between 40-60 min [5]. During the exam analysis, it is necessary complete concentration by the doctor, since an abnormal frame can be in the middle of a normal frames video segment. So the analysis of a capsule endoscopic video is a time consuming task, prone to errors, and so it claims for computational help. Note also that having an expert physician analyzing, for a long period, a capsule endoscopic exam is also very costly, and, therefore, exists an important economic opportunity to develop a computer assisted diagnosis tool to this task.

The detection of abnormalities based in texture alterations of intestine mucosa has been previously reported. In the Maroulis et al. work [6][7], different classifications schemes, based in textural features extracted from the Discrete Wavelet domain, were proposed to classify colonoscopy videos. Kodogiannis et al.[2] 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 , and a statistical approach, where statistical texture descriptors are calculated from the histograms of the RGB and HSV color spaces of CE video frames. In authors' previous work [8][9], are proposed different algorithms to classify capsule endoscopic video frames, based in textural descriptors taken from coocurrence matrices, using the discrete wavelet transform to select the bands with the most significant texture information for classification purposes. In the present work is proposed an algorithm based on higher order statistics of multi-scale texture descriptors, in order to model the complex process of human perception of texture. The texture features are the input of a multilayer perceptron neural network, a well known classifier in pattern recognition problems.

II. PROPOSED ALGORITHM

It is known that the human texture identification is a complex multi-scale process. In order to model this complex pattern recognition task, it is proposed a method based in higher order statistics of textural descriptors taken from the coocurrence matrix of images synthesized with the most relevant texture of an original capsule endoscopic video frame. It is also included, in the classification features extracted from each frame, the covariance between texture descriptors taken from coocurrence matrices of images synthesized with different scales of the Discrete Wavelet Transform (DWT) of the original frame, analyzing the variations of the information present at different scales in order to model the human multi-scale process of texture identification. Texture information is encoded as medium and high frequency, since the low-frequency components of the images do not contain major texture information, and, therefore, the lowest scales of the DWT of an image will present the most relevant texture information. To reduce the final number of features per frame, new images are synthesized from the selected wavelet scales, where the new image contains only the vital texture information present in the selected wavelet scale. In the present work, and in order to model the multi-scale aspects of the human texture identification, are synthesized two new images for each video frame, one containing the lowest DWT scale coefficients information (DWT bands 1,2 and 3) and the other containing the second scale DWT scale coefficients information (DWT bands 4,5 and 6). Therefore these two synthesized images contain texture information at different detail level, which allows the implementation of a multi-scale approach to this classification problem. The proposed algorithm can be decomposed in the following steps:

A. Wavelet coefficients selection and new image synthesis

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

$$I^i, \qquad i=1,2,3. \tag{1}$$

where *i* stands for the color channel.

These three color channels are originally in the RGB color space, but are transformed to the HSV color space. 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:

$$W^i = \left\{L_n^i, D_l^i\right\}, \quad i=1,2,3 \quad l=1,...,6 \tag{2}$$

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

Since we want to evaluate the relevant patterns at different scales, it is necessary to select the desired wavelet coefficients, at different DWT scales. In the present work were selected the lowest DWT scale (DWT bands 1,2 and 3) and the second scale DWT scale (DWT bands 4,5 and 6). Therefore, let S^i be matrices that have the selected wavelet coefficients at the corresponding positions and zeros in all other positions:

$$S_x^{\ i} = \left\{D_l^i\right\}, \quad i=1,2,3 \quad l=1,2,3 \vee 4,5,6 \quad x=1 \vee 2 \tag{3}$$

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

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

$$N_x^{\ i} = IDWT(S_x^{\ i}), \quad i=1,2,3 \quad x=1 \vee 2 \tag{4}$$

where *i* stands for color channel, *x* for the selected DWT scale and *IDTW()* is the inverse wavelet transform.

For each capsule endoscopic frame two images are calculated, reconstructed from the selected DWT scales, containing the essential textural patterns of the original image, at different detail level.

B. Cooccurrence matrix and texture descriptors

In spite of existing diverse methods to extract the texture information within an image, in the present work it was chosen an approach based in cooccurrence matrices and texture statistical descriptors, originally proposed by Haralick [10]. 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 $f(i,j,d,\theta)$, 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. 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 the image. From these matrices, statistical descriptors can be calculated in order to extract texture information from the synthesized images. In the proposed algorithm only 4 statistical measures are considered among the 14 originally proposed by Haralick [10], namely angular second moment (F1), correlation (F2), inverse difference moment (F3), and entropy (F4).

There are two synthesized images and for each them are calculated four coocurrence matrices for each color channel, which results in twelve coocurrence matrices for each image. For each coocurrence matrix, four statistical measures are calculated, resulting in a total of 96 texture descriptors, 48 for each image:

$$F_m\left(C_\alpha\left(N_x^i\right)\right) \quad i=1,2,3 \quad x=1\vee 2$$
$$\alpha=0,\frac{\pi}{4},\frac{\pi}{2},3\frac{\pi}{4} \quad m=1,2,3,4 \tag{5}$$

where i stands for color channel, x for DWT scale, m for statistical measure and α for the direction in the cooccurrence computation.

C. Higher-order statistics of the texture descriptors and multi-scale textures covariance

The mean and variance for each F_m are calculated over α, for every color channel, resulting in a set of 24 components per synthesized image. In authors' previous work, it is stated that higher order statistics can be used to model deviations to the Gaussian distribution, which are more accentuated in pathological cases for almost all the texture descriptors. Note that this shift to the normal Gaussian distribution does not affect preferentially any descriptor also. In the present work, and to reduce the size of the final feature set, it was only used the third centered moment to model the non-Gaussianity, calculated for each F_m over α, as:

$$\gamma^3_{x,m,i}=\frac{1}{N}\sum_\alpha\left[\begin{pmatrix}F_m\left(C_\alpha\left(N_x^i\right)\right)\\-E\left\{F_m\left(C_\alpha\left(N_x^i\right)\right)\right\}\end{pmatrix}^3\right] \tag{6}$$

where i stands for color channel, x for DWT scale, m for statistical measure and α for the direction in the cooccurrence computation.

The covariance between the same texture descriptors at different DWT scales is also calculated, in order to model the detail effect in the patterns within the original image, since the two synthesized images contain different detail level information. Let *MSC* be the multi-scale covariance of a texture descriptor in the two synthesized images:

$$MSC_{i,m}=\sum_\alpha\left[F_m\left(C_\alpha\left(N_1^i\right)\right)-E\left\{F_m\left(C_\alpha\left(N_1^i\right)\right)\right\}\right]X$$
$$\left[F_m\left(C_\alpha\left(N_2^j\right)\right)-E\left\{F_m\left(C_\alpha\left(N_2^j\right)\right)\right\}\right] \tag{7}$$

where i stands for color channel, x for DWT scale, m for statistical measure and α for the direction in the cooccurrence computation.

Therefore, each capsule endoscopic video frame will be characterized by 84 features, which will be the input of the MLP network. 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.

III. IMPLEMENTATION AND RESULTS

The experimental dataset for the evaluation for the proposed method was constructed with frames from 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 final dataset consisted in 500 normal frames, which were equally divided in two sets, for the MLP network training and testing, and 150 abnormal frames, which were also equally divided in two sets. Examples of the dataset frames can be observed in the figure 1.

a)


b)

Fig. 1. Examples of: a) normal intestine frame;
b) abnormal intestine frame;

From previous work [8][9], it was concluded that the reduction of the gradation levels of each color channel from 256 levels to 32 levels does not compromise the texture analysis process and leads to the improvement of the processing time per frame, which is about 2 seconds in MATLAB running in a 3.2 GHz Pentium Dual Core processor-based with 1 GB of RAM. Note also that this gradation levels reduction must be followed by a proper dispersion of the pixel values to all the available range, in order to minimize the loss of classification performance.

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:

$$Sensitivity = \frac{d}{c+d}.100\,(\%). \quad (8)$$

$$Specificity = \left(100 - \frac{b}{a+b}.100\right)(\%). \quad (9)$$

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.

Table 1 Classification performance of the proposed algorithm

Feature Set	1	2	3	4
Specificity (μ±σ%)	89.2±1.4	91.4±2.6	92.9±1.7	93.7±2.0
Sensibility (μ±σ%))	90.1±1.8	93.4±3.2	91.9±1.3	94.6±1.6

To test the performance of the proposed algorithm, it was evaluated the classification improvement when the third centered moment and the multi-scale covariance were added to the features extracted for each frame. Therefore, for each video frame, feature set 1 was composed by the mean and variance for each F_m calculated over α, feature set 2 was composed by the elements of feature set 1 plus the third order centered moment, calculated trough (6), feature set 3 was composed by the elements of feature set 1 plus the multi-scale covariance parameters, calculated trough (6), and feature set 4 was composed by all the 84 features.

IV. DISCUSSION AND FUTURE WORK

From the presented results, it is clear that the proposed algorithm has potential to be used in a automatic classification tool to reduce the time spent by the physician in the analysis of a capsule endoscopy exam, namely as a selection process that only shows to the physician the most suspect frames. However, and to assure a robust application, this method has to be tested with a larger dataset, so the future work will include the increase in the available dataset. Also different classification schemes will be evaluated, to optimize the classification performance of the process. Dimensionality studies will be done to the proposed feature set, but different feature sets will be also considered. The main goal of the present research is development of an automatic abnormalities detection system in capsule endoscopy videos, for the most common CE detectable diseases.

REFERENCES

1. Herrerías J, Mascarenhas M (2007) Atlas of Capsule Endoscopy. Sulime Diseño de Soluciones, Sevilla
2. Kodogiannis V, Boulougoura M, Wadge E and Lygouras J (2007) The usage of soft-computing methodologies in interpreting capsule endoscopy. Engineering Applications of Artificial Intelligence 20: 539–553
3. Idden G, Meron G, Glukhovsky A and Swain P (2000) Wireless capsule endoscopy. Nature 415-417
4. Qureshi WA (2004) Current and future applications of capsule endoscopy. Nature Reviews Drug Discovery 3:447-450
5. Pennazio M (2006) Capsule endoscopy: Where are we after 6 years of clinical use?, Digestive and Liver Disease 38:867–878
6. Maroulis D, Iakovidis D, Karkanis A and Karras D (2003) CoLD: a versatile detection system for colorectal lesions in endoscopy video frames. Computer Methods and Programs in Biomedicine 70:151-166.
7. Karkanis S, Iakovidis D, Maroulis D, Karras D, and Tzivras M (2003) Computer-Aided Tumor Detection in Endoscopic Video Using Color Wavelet Features. IEE Trans. On Information Technology in Biomedicine 7:3:141-152.
8. Lima C, Barbosa D et al. (2008) Classification of Endoscopic Capsule Images by Using Color Wavelet Features, Higher Order Statistics and Radial Basis Functions, Proceedings of IEEE-EMBC2008, to be published.
9. Barbosa D, Ramos J, and Lima C (2008) Detection of Small Bowel Tumors in Capsule Endoscopy Frames Using Texture Analysis based on the Discrete Wavelet Transform, Proceedings of IEEE-EMBC2008, to be published.
10. Haralick RM (1979) Statistical and structural approaches to texture. Proc. IEEE 67:786–804