Detection of Small Bowel Tumors in Endoscopic Capsule Images by Modeling Non-Gaussianity of Texture Descriptors

D. Barbosa, J. Ramos, A. Tavares and C. S. Lima,

Abstract—This paper presents an approach to the automatic detection of small bowel tumors by processing endoscopic capsule images. The most significant texture information is selected by using wavelet processing and captured in the image domain from an appropriate synthesized image. Co-occurrence matrices are used to derive texture descriptors by modeling second order statistics of color image levels. These descriptors are then modeled by using third and fourth order moments in order to cope with distributions that tend to become non-Gaussian especially in some pathological cases. The proposed approach is supported by a classifier based on radial basis functions procedure for the characterization of the image regions along the video frames. The whole methodology has been applied on real data and shows that higher order moments can be effective in modeling capsule endoscopic images regarding tumor detection.

I. Introduction

Conventional endoscopy presents some important limitations in the diagnosis of small bowel problems, since it is limited to the upper gastrointestinal (GI) tract, at the duodenum, and to lower GI tract, at terminal ileum. Therefore, prior to the invention of wireless Capsule Endoscopy (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]. Due to the referred limitations, a very significant part of the small intestine, which has a medium length of six meters, is not seen by these conventional techniques.

The conventional endoscopy diagnosis procedure consists in an exam that uses a flexible endoscope, with a video camera in the distal tip, to acquire intra-corporeal images from the GI tract. The endoscope is introduced through the mouth (upper GI endoscopy) or trough the rectum (lower GI endoscopy), into the body passively by pushing it from outside. The motion of the tip of the endoscope is controlled by the operating physician by manipulating wires in the shaft of the endoscope from outside of the human body. 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 [2].

In 2000, the development of capsule endoscopy 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 introduction of CE also represented the first major technological innovation in GI diagnostic medicine since the flexible endoscope [1]. More recently, a new technique, the double-balloon enteroscopy (DBE), has been introduced into clinical practice [3]. DBE has the potential to examine the entire length of the small bowel with biopsy and therapeutic capability. However it is a time consuming procedure that requires specialist training for the operating physician. Therefore CE can be used as

Manuscript received April 7, 2008.

D. Barbosa, C. S. Lima, and A. Tavares are with the University of Minho, Industrial Electronics Department, Campus de Azurém 4800-058 Gimarães, Portugal phone: +351 253 604706; fax: +351 253 604709; e-mail: clima@ dei.uminho.pt..

J. Ramos is with Capucho´s Hospital, Gastroenterology department, Alameda Sto. António dos Capuchos, Lisbon, Portugal..

a first line diagnosis method, while DBE can be used as a confirmatory or therapeutic modality for lesions first visualized by CE [4].

The first commercially-available wireless video capsule was the M2A^{TM} (by Given Imaging Ltd., Yoqneam, Israel), a pill-like wireless capsule (11mm×26 mm), which contains a miniaturized camera, a light source and a wireless circuit for the acquisition and transmission of signals [5]. The capsule is passively propelled trough the entire GI tract, capturing images at a rate of two frames per second. Image features include a 140° field of view, 1:8 magnification allowing visualization of individual villi, 1–30 mm depth of view, and a minimum size of detection of about 0.1 mm. By the time battery power expires after about 8 hours, the camera will have captured about 55,000 images, which are transmitted to the hard drive in a belt worn by the patient [6]. The capsule is excreted in the patient’s stool, usually within 24–48 h, and not reused [4].

Unlike conventional endoscopy, no drugs are administered to the patient and air insufflation is not neeeded. This might make the test more sensitive, as drug-induced drops in blood pressure or air tamponade (that is, compression of the blood vessels within the bowel lumen by insufflated air), which frequently occur during conventional endoscopy, can make it difficult to visualize small bleeding vessels. Some investigators maintain that use of the capsule camera is a more ‘physiological’ form of endoscopy [6]..

Currently, the M2A^{TM} capsule was replaced by the PillCam^{TM} SB, an endoscopic capsule optimally design for small bowel data acquisition. Given Imaging has also endoscopic capsules best suited to esophagus and colon analysis (PillCam^{TM} ESO and PillCam^{TM} COLON). Olympus Corporation has also launched an endoscopic capsule for the study of the small bowel [1]. The time required to a physician to analyze the resulting video is, on average, 40-60 min[4]. 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. 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.

After the introduction of CE, it was discovered that that the prevalence and malignancy rates for small bowel tumors are much higher than previously reported and that the early use of CE can lead to earlier diagnoses and reduced costs, and hopefully prevent cancer [1]. 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 [4]. However, the early diagnosis of small bowel tumor is difficult, because signs are vague and laboratory tests are unhelpful [7]. There are no specific symptoms for benign or malign small bowel tumors and, normally, they are detected in advanced stages, due to mechanical obstruction of the GI tract. However, obscure GI bleed can be an earlier symptom and a key factor for an early diagnosis of these lesions [1]. Small bowel tumors are a significant finding at CE and are often missed by other methods of investigation. Note that, even in malignant lesions, treatment is potentially curative in the absence of metastatic disease.

The automatic detection of abnormalities can be based in alterations in the texture of the small intestine mucosa. Maroulis *et al.* [8][9] 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. In the work of Abyoto *et al.*[10] has been observed that the textural information is localized in the middle frequencies and lower scales of the original signal. 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, 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.*[2] work is also proposed 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 [11][12], are proposed different classification schemes for capsule endoscopic video frames based in statistical measures taken from texture descriptors of co-occurrence matrices, using the discrete wavelet transform to select the bands with the most significant texture information for classification purposes. In previous investigation, it was observed that abnormal capsule endoscopic images tend to present non-Gaussian distributions of the texture descriptors, while normal frames present a normal distribution of these descriptors. Therefore, the Gaussianity of the co-occurrence matrices texture descriptors may be the used in as features in a classification scheme to identify abnormal frames.

The proposed methodology in the present paper focus the feature extraction process from the endoscopic capsule video frames, with a method based in higher order statistic evaluation of texture descriptors taken from co-occurrence matrices, calculated for an image reconstructed from the wavelet coefficients of the selected wavelet bands, which contain the most important texture information for classification purposes. The proposed approach is performed in the HSV color space. These features are the input of a radial basis functions (RBF) neural 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. The texture descriptors are calculated over the co-occurrence matrix calculated from the new image synthesized from the selected wavelet coefficients, for every color channels. These are statistical descriptors that contain second order color level information captured from the co-occurrence matrix, which are mostly related to the human perception and discrimination of textures. With color level information we intend to mean color intensity. Since human perception is a complex pattern identification process, it is proposed the correlation between the descriptors, using higher order statistics to correlate the texture data present in the three color channels, as well as non-Gaussianity identification in the texture descriptors of each color channel.

Texture can be roughly classified as fine, course, grained, smooth, etc and consists of texture primitives or elements, sometimes called textons. However, primitives are hard to define. Primitives for the checkered textile or fabric, can be defined by at least two hierarchical levels, where the first corresponds to textile checks of knitted stripes while the second corresponds to the finer texture of the fabric or individual stitches. Therefore texture description is scale dependent. The wavelet transform is perhaps the most appropriate tool for scale dependent signal analysis.

Texture can be, however, more precisely defined to make machine recognition possible. Tone and structure of a texture are features that help to define more precisely textures [13]. Tone is based mostly in pixel intensity properties in the primitive, while structure is the spatial relationship between primitives. Repeated occurrence of some color level configuration can constitute an interesting texture description, since rapid variations with distance define fine textures while slowly variations define coarse textures. Co-occurrence matrices encode precisely this information and can be used to extract statistical descriptors for texture classification and pattern recognition systems based on textural descriptors. The statistical model is usually built by estimating 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 pixel intensity of colour levels i and *j* at a given direction *θ.* The angular displacement used is the set {0, π/4, π/2, 3π/4}. It is considered only 4 statistical measures among the 14 originally proposed [13]. 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 co-occurrence matrix, defined respectively as:

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

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

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

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

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

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

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

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

where *p(i,j)* is the *ijth* entry of normalized co-occurrence 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).

In the ambit of this paper these features were obtained from pre-processed images, which are synthesized from source images where information not relevant for texture analysis was discarded.

II.1- Image pre-processing

The image pre-processing stage synthesizes an image containing only the most relevant textural information from the source image. The most relevant texture information often appears in the middle frequency channels [14]. Texture is the discrimination information that differentiates normal from abnormal lesions, regarding colorectal diagnosis [8], [9], [15] and [16], hence it is likely to be extrapolated to small bowel diagnosis with similar characteristics.

The wavelet transform allows a spatial/frequency representation by decomposing the image in the corresponding scales. When the composition level decreases in the spatial domain it increases in the frequency domain providing zooming capabilities and local characterization of the image [17]. This spatial/frequency representation, which preserves both global and local information, seems to be adequate for texture characterization.

Color transformations of the original image *I* result in three decomposed color channels:

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

where *i* stands for the color channel.

A two level discrete wavelet frame transformation is applied to each color channel (I^i). This transformation results in a new representation of the original image by a low resolution image and the detail images. Therefore the new representation is defined as:

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

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

Since the textural information is better presented in the middle wavelet detailed channels, then second level detailed coefficients would be considered. However, the relatively low image dimensions (256 X 256) limit the representation of the details, becoming the first level more adequate for texture representation [12]. Thus, the image representation consists of the detail images produced from (6) for the values l=1, 2, 3 as shown in figure 1. This results in a set of 9 subimages:

$$\{D_l^i\} \quad i = 1,2,3 \quad l = 1,2,3. \qquad (7)$$

For the extraction of the second order statistical textural information co-occurrence matrices would be used calculated over the nine different subimages. However, in order to diminish the dimension of the observation vectors the image to process can be synthesized from inverse wavelet transform with the coefficients of the large scales (lower frequencies) discarded. This procedure reduces the dimensionality of the observation vector by a factor of 3 since only three images need to be processed (colour channels) instead of the nine obtained in the wavelet domain. Our results confirmed that the most relevant texture information is maintained through Inverse wavelet transform, which is used to synthesize a new image from the selected wavelet bands. In [12], it was demonstrated that the most significant texture information for classification purposes, in capsule endoscopic frames, is at the lowest wavelet scale. Therefore, let Si be a matrix that has the selected wavelet coefficients at the corresponding positions and zeros in all other positions:

$$S^i = \{D_l^i\}, \quad i = 1,2,3 \quad l = 1,2,3 \quad (8)$$

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

$$N^i = IDWT(S^i), \qquad i = 1,2,3. \quad (9)$$

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

These matrices capture spatial interrelations among the intensities within the synthesized image level. The co-occurrence matrices are estimated in four different directions resulting to 12 matrices:

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

where *i* stands for the color channel and *α* for the direction in the co-occurrence computation.

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

$$\begin{aligned} &F_{m}\left(C_{\alpha}\left(N^{i}\right)\right) \qquad i=1,2,3 \\ &\alpha=0,\frac{\pi}{4},\frac{\pi}{2},3\frac{\pi}{4} \qquad m=1,2,3,4 \end{aligned} \qquad (11)$$

where *m* stands for statistical measure.

Since each feature represents a different property of the synthesized image it is expected that similar textures will have close statistical distributions and consequently they should have similar features. This similarity between features can be statistically modeled in a tri-dimensional space since features can be simultaneously observed in the three channel colors.

While the texture descriptors can be considered statistically independent, their occurrence together in the three color channels is likely to be correlated. The correlation between two descriptors measures their tendency to vary together and constitutes the sufficient statistics when the multivariate density is normally distributed. One of the main properties of the multivariate normal distributions is that the marginal distributions are also normal however the converse (reverse?) is not necessarily true. As example, Figure 4 and 5 show the distribution of F1, for the color channel H, in a set of 100 normal and 92 tumoral images. It is evident that the distribution of F1 is not Gaussian especially in cases of tumor.

In spite of many multivariate statistics used in practice converge in distribution to a multivariate normal, which is acceptable regarding the multivariate central limit theorem, Gaussianity tests can help to model more accurately multivariate distributions. Gaussianity tests can be based on the Mahalanobis distances from the sample mean, computing the squared Mahalanobis distances:

$$m_{i}^{2}=\left(x_{i}-\bar{x}\right)^{T} S^{-1}\left(x_{i}-\bar{x}\right) \qquad (12)$$

for the *n* multivariate observations and plotting them against the Chi 2 (χ2) distribution percentiles:

$$\chi^{2}_{(1-\alpha_{i})p} \ where \ 1-\alpha_{i}=\frac{(i-0.5)}{n} \qquad (13)$$

where in equation (12) *T* stands for transpose and *S* for sample covariance matrix.

Higher Order Statistics (HOS) leads to a lower number of model parameters, therefore it is perhaps the most appropriate choice, especially in applications requiring a significant amount of computational effort, such as the case of massively processing of capsule endoscopic images based on co-occurrence matrices. Additionally one constraint always associated with Gaussian mixture modeling is the choice of the number of Gaussian components, which optimum value depends greatly on the application and is usually not a priori known.

II.2- Modeling non-Gaussianity of texture statistical measures

Second order statistics is a well established theory that is completely adequate to represent random vectors. Nevertheless, it is limited by the assumptions of Gaussianity, linearity, stationarity, etc. HOS characterized by higher order moments are adequate to model non-Gaussian distributions under the assumption that all the moments are finite and so their knowledge is in practice equivalent to the knowledge of their probability function [18].

Third and fourth order moments have precisely meaning in separating Gaussian from other distributions. The third central moment:

$$\mu_3 = E\left\{\left(x-\bar{x}\right)^3\right\} \tag{14}$$

gives a measure of assymmetricity of the probability density function around their mean (skewness), while the fourth central moment gives a measure of the peaky structure of the distribution when compared with the Gaussian. Higher than fourth order moments are used seldom in practice, hence not tried in the ambit of this paper. Therefore second, third, and fourth order moments, were used in the ambit of this paper:

$$\begin{gathered} E\left\{F_m^i, F_m^j\right\} \\ E\left\{F_m^i, F_m^j, F_m^k\right\} \\ E\left\{F_m^i, F_m^j, F_m^k, F_m^l\right\} \end{gathered} \tag{15}$$

About the fourth order moment only the fourth central moment was used in order to model subgaussianity and supergaussianity of the marginal distributions for all the four statistical texture descriptors.

The second order moments or correlation of the same statistical measure between different color channels is computed as

$$\phi_{F_m^i, F_m^j} = \sum_\alpha F_m\left(C_\alpha\left(I_s^i\right)\right) X F_m\left(C_\alpha\left(I_s^j\right)\right) \tag{16}$$

which results in the computation of 24 coefficients, six different correlations computed for each descriptor. The third order moments are computed as :

$$E\left\{\left(F_m^i\right)^3\right\} = \sum_\alpha \left(F_m\left(C_\alpha\left(I_s^i\right)\right)\right)^3 \tag{17}$$

$$\begin{gathered} E\left\{F_m^i, F_m^j, F_m^k\right\} = \sum_\alpha F_m\left(C_\alpha\left(I_s^i\right)\right) X \\ F_m\left(C_\alpha\left(I_s^j\right)\right) X F_m\left(C_\alpha\left(I_s^k\right)\right) \quad i \neq j \neq k \end{gathered} \tag{18}$$

which results in the computation of 16 coefficients, four different third order moments computed for each descriptor. The fourth order moments are computed as:

$$E\left\{\left(F_m^i\right)^4\right\} = \sum_\alpha \left(F_m\left(C_\alpha\left(I_s^i\right)\right)\right)^4 \tag{19}$$

which results in the computation of more 12 coefficients.

Summing up 28 higher order moments to the second order moments, each frame is characterized by a set of 52 components in the observation vector. These components constitute the input of the radial basis function.

III. Radial Basis Functions

Radial basis functions (RBF) are hidden activation functions embedded in a two layer neural network. RBF's have their roots entrenched in much older pattern recognition techniques as for example clustering and mixture models. The input into an RBF network is nonlinear while the output is linear. Due to their nonlinear approximation properties, RBF networks are able to model complex mappings, which perceptron neural networks can only model by means of multiple intermediary layers. RBF's are the neural networks for excellence used in statistical applications since the mappings are based on the similarity of the underlying statistics between the training set and the pattern to be tested. In RBF's for pattern recognition applications the most used activation function is the Gaussian, however Gaussian mixtures have been considered in various fields. The Gaussian activation function for RBF networks is given by:

$$\Phi_j(\mathbf{X}) = \exp\left[-\left(\mathbf{X} - \boldsymbol{\mu}_j\right)^T \sum_j^{-1}\left(\mathbf{X} - \boldsymbol{\mu}_j\right)\right] \quad (20)$$

for j=1 …L, where **X** is the input feature vector, L is the number of hidden units and **μ**j and ∑j are the mean and covariance matrix of the *j*th Gaussian function.

The Gaussian activation function can model more accurately groups of features that have tendency to vary together and so more likely to represent similar patterns. This modeling is optimal regarding statistical classification purposes especially when second order statistics need to be accurately modeled. Given the central limit theorem, this characteristic perhaps justifies the fact that usually RBF's present a better degree of generalization that feed forward neural networks, in spite of sometimes performing poorer in the training set. This conclusion can be found in several scientific papers [19]. Besides modeling probability density functions, RBF's networks have been shown to implement the Bayesian rule [20].

The classification scheme described in this paper used a standard radial basis function, which basic scheme is shown in figure 2, with 52 input units, 35 RBF units and 1 output neuron. The training algorithm was the well known hybrid learning process where the centers were computed by clustering, the spreads of the Gaussians chosen by normalization and the Least Mean Square algorithm for computing the network weights. The output neuron was used to classify the data into 2 classes normal and abnormal.

IV. Experimental Results

The experimental training set consisted of 400 frames from several endoscopic exams taken at the Capucho's Hospital in Lisbon by Doctor Jaime Ramos. Half of these frames do not present any abnormality, while the rest were selected as representing tumor pathology pattern. The selection was made by the physician. The test set contains 600 normal capsule endoscopic frames and 250 frames with tumor evidences. The training and testing data are from different patients. Figures 7 and 8 shows some frames belonging to the training set.

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 [21]. These two measures can be calculated as:

$$Sensitivity = \frac{d}{c+d}.100\ (\%) \quad (21)$$

$$Specificity = \left(100 - \frac{b}{a+b}.100\right)(\%) \quad (22)$$

Where *a* is the number of true negative patterns, *b* is the number of false positive patterns, *c* is the number of false negative patterns and *d* is the number of true positive patterns.

The classification performance is high when both Sensitivity and Specificity are high, in a way that their tradeoff favors true positive or false positive rate depending on the application.

In order to compute the co-occurrence matrix for the new image, synthesized from the wavelet coefficients from the selected bands, a new algorithm was implemented, to avoid computing co-occurrences in the image corners where no image information exists. The co-occurrence computation was done considering d=1 since we intend to capture fine texture details in an image with relatively poor spatial resolution. 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.[22], 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.

The used colour space was HSV and the obtained pair (Sensitivity, Specificity) was for the described data set (91 ±0.4%, 92±0.1%). Table 1 resumes the most relevant results. The results are given in statistical terms, and, to test the importance of the higher order statistics, the classification vector for each frame had the second order moments, given by (16) or the second and third order moments, given by (16), (17) and (18), or the second, third and forth order moments, given by (16), (17), (18) and (19).

Classification vector	Specificity (%)	Sensitivity (%)
2nd order moments	90.1±0.5	90.5±0.6
2nd and 3rd order moments	91.2±0.3	92.4±0.4
2nd, 3rd, and 4th order moments	91.3±0.4	92.2±0.5

Table 1 - Classification performance of the proposed algorithm

From the analysis of the classification performance for the proposed algorithm, it is obvious that modeling the non-Gaussianity of the texture descriptors leads to better classification results. However, the addition of fourth order moments do not clearly improves the classification performance. Note also that to correctly estimate higher order moments, larger amounts of data is needed, and so the classification improvement with the addition of higher order moments will be more evident in larger datasets.

V. DISCUSSION AND FUTURE WORK

The results of this paper show that higher order statistics for texture descriptors can be used as classification parameters in order to classify capsule endoscopic video frames. It was also shown, that higher order statistics lead to superior classification performance. However, the improvement from modeling non-Gaussianity should be clearer in larger datasets, since it is well known that higher order moments need much more data to be accurately estimated than second order moments. Therefore, future work will include the augment of the available tumor frames, in order to better identify the importance of higher order statistics. Different classification schemes will also be subject of future investigation.

REFERENCES

[1] Herrerías, J. M., Mascarenhas, M., 2007, *Atlas of Capsule Endoscopy,* Sulime Diseño de Soluciones, Sevilla.

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

[3] Yamamoto H, Kita H., 2006, *Double-balloon endoscopy: from concept to Reality*, Gastrointestinal Endoscopy Clinics of North America, 16, 347–361.

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

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

[6] Qureshi, W. A., 2004, *Current and future applications of capsule endoscopy*, Nature Reviews Drug Discovery, 3, 447-450.

[7] Franchis, R. de, Rondonotti, E., Abbiati, C., Beccari, G., and Signorelli, C., 2004, Small bowel malignancy, Gastrointestinal Endoscopy Clinics of North America, 14, 139-148.

[8] 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.

[9] 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.

[10] Abyoto, R., Wirdjosoedirdjo, S., and Watanable, R., 1998 *Unsupervised texture segmentation using multiresolution analysis for feature extraction*, J Tokyo Univ. Inform. Sci., 2, 9, 49-61.

*[11]*Lima, C., Barbosa, C., Ramos, J., Tavares, A., Monteiro, L., and Carvalho, L., 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.

[12] Barbosa, C., 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.

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

[14] Van de Wouwer, G. , Scheunders, P., and Van Dyck, D., 1999, *Statistical texture characterization from discrete wavelet representations*, IEEE Trans. Image Processing, 8, 592-598.

[15] Nagata, S., Tanaka, S., Haruma, K. , Yoshihara, M., Summi, K., Kajiyama, G. and Shimamoto, F., 2000, *Pit pattern diagnosis of early colorectal carcinoma by magnifying colonoscopy: Clinical and histological implications*. Int. J. Oncol., 16, 927-934.

[16] Kudo, S., Kashida, H., Tamura, T. , Kogure, E., Imai, Y., Yamano, H. and Hart, A. R., 2000, *Colonoscopic diagnosis and management of nonpolypoid early colorectal cance,*. World J. Surgery, 24, 1081-1090.

[17] Mallat, S. 1998, *A wavelet tour of signal processing*, Academic Press.

[18] Nandi, A., 1999, *Blind Estimation Using Higher-Order Statistics*, Kluwer.

[19] Kovacevic, D., and Loncaric, S., 1997, *Radial basis function-based image segmentation using a receptive field,* Tenth IEEE Symposium on computer-based Medical Systems, pp. 126-130.

[20] Bors, A.G., and Gabbonj, G., 1994, *Minimal topology for a radial basis function neural network for pattern classification*, Digital Signal Processing: a review journal, 4, 3, 173-188.

[21] Swets, J. A., Dawes, R. M., and Monahan, J. , 2000, *,* Psych. Sci Public Interest, 1, 1-26.

[22] 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.

Figures

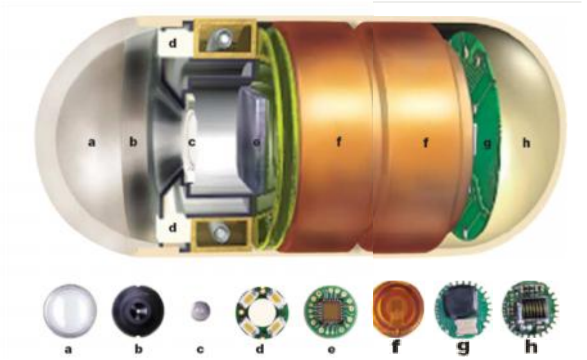

Fig. 1. Given Imaging M2A endoscopic capsule. Optical dome (**a**), lens holder (**b**), short focal-length lens (**c**), six white-light-emitting diode illumination sources (**d**), complementary metal oxide silicon (CMOS) chip camera (**e**), two silver oxide batteries (**f**), UHF band radio telemetry transmitter (**g**), antenna (**h**).

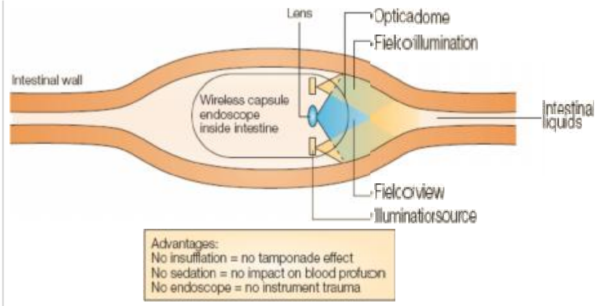
Lens
Opticadome
Fielcoillumination
Intestinal wall
Wireless capsule endoscope inside intestine
Intestinal liquids
Fielcoview
Illuminatiorsource
Advantages:
No insufflation = no tamponade effect
No sedation = no impact on blood profuion
No endoscope = no instrument trauma

Fig. 2. Physiological advantages o f capsule endoscopy

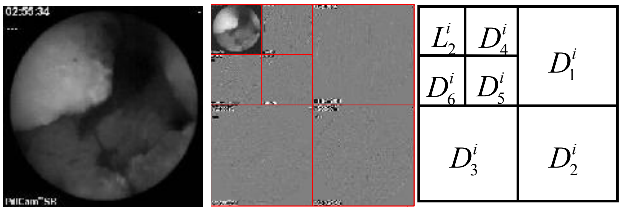

Fig. 3. Example of two level wavelet decomposition scheme of the original image for color channel *i*.

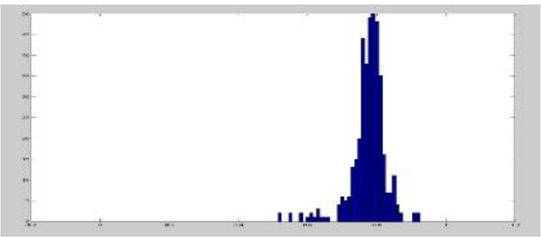

Fig. 4. Distribution of F1 texture descriptors for a set of 100 normal capsule endoscopic frames.

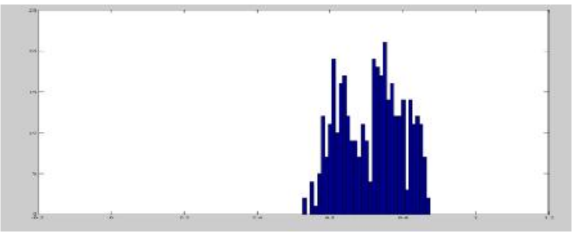

Fig. 5. Distribution of F1 texture descriptors for a set of 92 abnormal (small bowel tumor) capsule endoscopic frames.

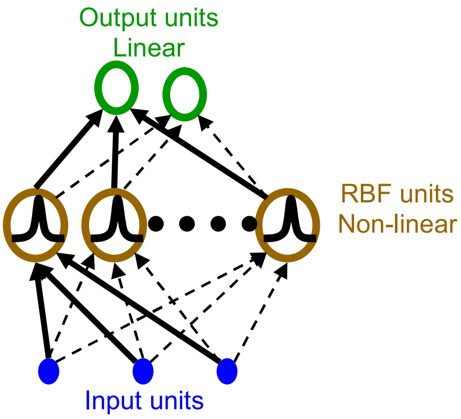

Figure 6. Scheme of a radial basis function with three input and two output neurons.

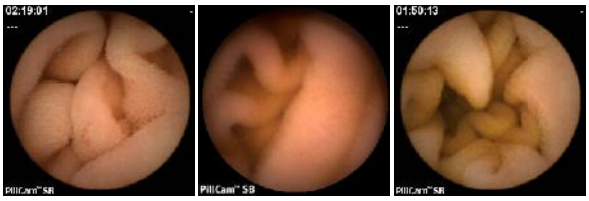

Fig. 7. Example of a normal intestinal tissue frames

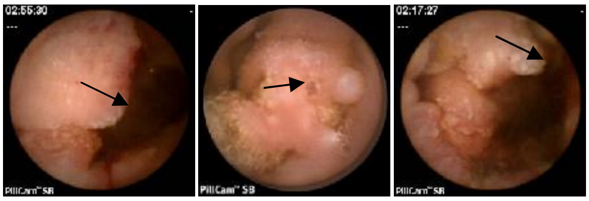

Fig. 8. Example of a tumor intestinal tissue frames