

# Automatic detection of small bowel tumors in Capsule Endoscopy based in Color Curvelet Covariance statistical texture descriptors

Daniel J. C. Barbosa, Jaime Ramos, José Higinio Correia and Carlos S. Lima

**Abstract**— Traditional endoscopic methods do not allow the visualization of the entire Gastrointestinal (GI) tract. Wireless Capsule Endoscopy (CE) is a diagnostic procedure that overcomes this limitation of the traditional endoscopic methods. The CE video frames possess rich information about the condition of the stomach and intestine mucosa, encoded as color and texture patterns. It is known for a long time that human perception of texture is based in a multi-scale analysis of patterns, which can be modeled by multi-resolution approaches. Furthermore, modeling the covariance of textural descriptors has been successfully used in classification of colonoscopy videos. Therefore, in the present paper it is proposed a frame classification scheme based in statistical textural descriptors taken from the Discrete Curvelet Transform domain, a recent multi-resolution mathematical tool, which is based in an anisotropic notion of scale and high directional sensitivity in multiple directions, being therefore suited to characterization of complex patterns as texture. The covariance of texture descriptors taken at a given detail level, at different scales, is used as classification feature, in a scheme designated as Color Curvelet Covariance. The classification step is performed by a multilayer perceptron neural network. The proposed method has been applied in real data taken from several capsule endoscopic exams and reaches 97.2% of sensitivity and 97.4% specificity. These promising results support the feasibility of the proposed method.

## I. INTRODUCTION

CONVENTIONAL endoscopic exams do not allow the entire visualization of the gastrointestinal (GI) tract. The conventional upper GI endoscopy only reaches the duodenum, while lower GI endoscopy is limited to the terminal ileum, which means that the vast majority of the small bowel, which has an average length of six meters, is not visible with these conventional techniques. Other important disadvantages of conventional endoscopy methods are the discomfort caused to the patients and the risk of injuring the GI walls with the tip of the endoscope, since the correct navigation of the endoscope requires a highly trained physician.

Capsule Endoscopy (CE) was introduced in 2000, being considered the most important technological innovation in GI diagnostic medicine since the flexible endoscope [1]. The capsule endoscopic exam is simple and non-invasive procedure that is well accepted by the patient and can be performed on an outpatient basis. The endoscopic capsule is

a pill-like device, with only 11mm×26 mm, and includes a miniaturized camera, a light source and circuits for the acquisition and wireless transmission of signals [2]. As the capsule travels through the GI tract, it acquires video frames at a rate of two per second and sends them to a hard disk receiver that is worn in the belt of the patient, in a wireless communication scheme. At the end of the battery life, usually around eight hours, more than 50000 images will have been acquired. If no complications arise, the capsule should be in the patient's stool, usually within 24–48 h, and not reused [3]. The analysis of this huge amount of data is done in a workstation, with proprietary software that allows the visualization of the video, by an expert physician and takes, in average, 40-60 min [3]. This task requires a highly focused viewer, since any distraction of the physician may lead to misevaluation of exams. Nevertheless, the human evaluation is prone to errors. Furthermore, having an expert physician analyzing the exam for a long time is an important parcel in the total cost of the exam, so there is an important economic opportunity to develop a computer assisted diagnosis tool to this task.

After the introduction of CE, it was discovered 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, reduced costs and, hopefully, prevent cancer [1].

The use of statistic textural descriptors has been successfully applied in classification schemes for identification of abnormalities in colonoscopy videos in the work of Karkanis *et al* [4]. Kodogiannis *et al.* 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 based in statistical texture descriptors extracted from the different color channels' histograms [5]. From authors' previous work [6,7], the application of texture analysis techniques to classify capsule endoscopic frames is feasible and presents promising results.

The classification scheme described in this paper uses statistical texture descriptors taken from the Discrete Curvelet Transform, a powerful multi-resolution mathematical tool, and a standard MLP network trained through the well known back-propagation learning process. The choice of a simple classification scheme was done to make the results more representative of the choice of the features. The proposed algorithm has been used to classify real data from Hospital dos Capuchos' patients.

Manuscript received April 22, 2009.

José Higinio Correia, 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.

## II. DISCRETE CURVELET TRANSFORM

It is known for a long time that human perception of texture is based in a multi-scale analysis of patterns [8], which can be modeled by multi-resolution approaches. In fact, the multi-resolution ability of the Discrete Wavelet Transform (DWT) has been vastly explored in several fields of image processing such as compression, denoising and classification [9]. However, the directional of the DWT is limited, which might not be enough to capture all the complex texture patterns within an image.

Introduced in 2000, the Continuous Curvelet Transform (CCT) is based in an anisotropic notion of scale and high directional sensitivity in multiple directions [10]. While wavelets are certainly suitable for dealing with objects where the interesting phenomena, e.g., singularities, are associated with exceptional points, they are ill-suited for detecting, organizing, or providing a compact representation of intermediate dimensional structures. Given the significance of such intermediate dimensional phenomena, there has been a vigorous effort to provide better adapted alternatives by combining ideas from geometry with ideas from traditional multi-scale analysis [11]. Therefore, this tool can be used as a multi-resolution and multi-directional representation of the information within an image.

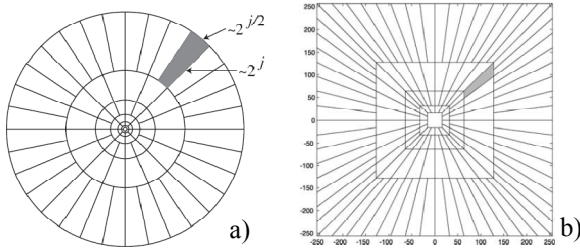


Fig. 2. CCT tiling of the frequency domain (a) and basic tiling of the digital coronization process (b)

The CCT is based in the tiling of the 2D Fourier space in different concentric coronae, one of each divided in a given number of angles, accordingly with a fixed relation. Now, to each of this polar “wedges” will be associated a frequency window  $U_j$  (figure 2-a) that will correspond to the Fourier transform of a curvelet  $\varphi_j(x)$  function, which can be thought of as a “mother” curvelet, since all the curvelets at scale  $2^j$  may be obtained by rotations and translations of  $\varphi_j(x)$ . So the CCT can be defined by a pair of windows  $W(r)$ , a radial window, and  $V(t)$ , an angular window. These are both smooth, nonnegative, and real-valued, with  $W$  taking positive real arguments and supported on  $r \in (1/2, 2)$  and  $V$  taking real arguments and supported on  $t \in [-1, 1]$ . These windows will always obey the admissibility conditions:

$$\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1, \quad r \in (3/4, 3/2). \quad (1)$$

$$\sum_{l=-\infty}^{\infty} V^2(t-l) = 1, \quad t \in (-1/2, 1/2). \quad (2)$$

Now, for each  $j \geq j_0$ , it is introduced the frequency window  $U_j$  defined in the Fourier domain by:

$$U_j(r, \theta) = 2^{-3j/4} W(2^{-j} r) V\left(\frac{2^{\lfloor j/2 \rfloor} \theta}{2\pi}\right). \quad (3)$$

where  $\lfloor j/2 \rfloor$  is the integer part of  $j/2$  and  $U_j$  corresponds to a polar “wedge” as seen in figure 2-a. The frequency window  $U_j$  will correspond to the Fourier transform of a curvelet  $\varphi_j(x)$  function. Consider an equispaced sequence of rotation angles  $\theta_l = 2\pi \cdot 2^{l-2j} \cdot l$ , with  $l=0,1,\dots$  such that  $0 \leq \theta_l \leq 2\pi$ , whose spacing is scale dependent, and the sequence of translation parameters  $k = (k_1, k_2) \in \mathbb{Z}^2$ . With these notations, curvelets are defined (as function of  $x=(x_1, x_2)$ ) at scale  $2^j$ , orientation  $\theta_l$  and position  $x_k^{(j,l)} = R_{\theta_l}^{-1}(k_1 \cdot 2^j, k_2 \cdot 2^j)$  by:

$$\varphi_{j,l,k}(x) = \varphi_j\left(R_{\theta_l}\left(x - x_k^{(j,l)}\right)\right). \quad (4)$$

where  $R_{\theta}$  is the rotation by  $\theta$  radians and  $R_{\theta}^{-1}$  its inverse (and also its transpose). A curvelet coefficient  $c(j,l,k)$  is then simply the inner product between an element  $f \in L^2(\mathbb{R}^2)$  and a curvelet  $\varphi_{j,l,k}$ :

$$c(j,l,k) = \langle f, \varphi_{j,l,k} \rangle = \int_{\mathbb{R}^2} f(x) \overline{\varphi_{j,l,k}(x)} dx. \quad (5)$$

Reference [11] proposes two different schemes for the discretization of the CCT, namely the USFFT algorithm and the wrapping algorithm. Both rely in the transformation of the frequency coronae of the CCT of figure 2 in a “Cartesian coronae”, which are based on concentric squares (instead of circles) and shears, in a process designated by digital coronization. The motivation for this digital coronization is the fact that coronae and rotations are not especially adapted to Cartesian arrays, which difficult their computation. Since it is stated that the wrapping algorithm may be simpler to understand and implement, this approach was chosen to calculate the Discrete Curvelet Transform (DCT) in the present work. Further details about the CCT and its discretization can be found in [9].

Therefore, the DCT coefficients are accurate representations of the original image with different detail, given by the different frequency content in each scale, but also with different detail in multiple directions, overcoming the directional limitations of the Discrete Wavelet Transform. This might be well suited for the analysis of complex spatio-frequency patterns as texture.

In the proposed approach, the CE frames were processed with the wrapping algorithm for three scales, with one, eight and sixteen angles respectively.

## III. STATISTICAL TEXTURE DESCRIPTORS

There are several statistical features that can be extracted from the curvelet domain as texture descriptors, being the most common the mean, the standard deviation, the energy and the entropy of each DCT sub-band [12].

Note that capsule endoscopic video frames are usually square images of 256x256 but the information is restricted to a circular area in the middle of the image, as it is observable in figure 3. Given the anisotropy of scale of the DCT, the computation of the textural descriptors will have to be done only for the coefficients within the area correspondent to the circular region of the CE frames,  $N$ . The proposed texture descriptors can be calculated as:

$$\mu = E\{P(i, j)\} = \frac{1}{N} \sum_i \sum_j P(i, j). \quad (6)$$

$$\sigma = \sqrt{E\{(P(i, j) - \mu)^2\}} = \sqrt{\frac{1}{N} \sum_i \sum_j (P(i, j) - \mu)^2}. \quad (7)$$

$$Energy = \frac{1}{N} \sum_i \sum_j P(i, j)^2. \quad (8)$$

$$Entropy = \frac{1}{N} \sum_g p(g) \cdot \log(p(g)). \quad (9)$$

where  $P(i, j)$  corresponds to the pixel at position  $(i, j)$  and  $p(g)$  the probability of a pixel having the gray level  $g$ , ie, the  $g^{th}$  entry of the image's normalized histogram.

Since the low frequency components of the images do not contain major texture information [7], the most important scales in the DCT will be those in which are present medium and high frequency, texture encoding, information. Therefore, no texture descriptors were computed for the scales whose coefficients correspond to low frequency content (coarsest scale coefficients). Furthermore, the coarsest scale coefficients of the DCT are not directional, and consequently do not possess directional sensitivity. From authors' previous work, it is highly expectable that the most relevant information for classification purposes is encoded as high frequency content in the scale correspondent to the highest detail level. Note however, that the computation of the proposed texture descriptors at every DCT angle will result in high dimensionality feature vectors.

The statistical dependence of textural descriptors taken in different color channels is useful to distinguish normal from abnormal texture patterns, as stated in authors' previous work. Furthermore, the same finding was previously reported in [4], for colonoscopy videos. Therefore, in the present work it is proposed a similar framework, designated as Color Curvelet Covariance, where the covariance of textural descriptors in the different color channels will be used as a classification feature. Note that in the present framework, the high directional sensitivity of the DCT will be likely to lead to more robust descriptors than a similar scheme used in textural descriptors taken from Discret Wavelet Transform coefficients. The Color Curvelet Covariance of a texture descriptor can be calculated as:

$$CCC(a, b, s, m) = \sum_{\alpha} (F_m(a, s, \alpha) - E\{F_m(a, s, \alpha)\}) \times (F_m(b, s, \alpha) - E\{F_m(b, s, \alpha)\}). \quad (10)$$

where  $a, b$  represent the different color channels in the covariance calculation,  $F_m$  is the statistic textural descriptor,  $\alpha$  is the considered angle of the DCT coefficients,  $s$  the considered DCT scale and  $E\{F_m(a, s, \alpha)\}$  the average of the statistical textural descriptor  $F_m$  over the different angles  $\alpha$ , in the color channel  $a$ . Note that if  $i=j$ , the previous equation is nothing more than the variance of the textural descriptor  $F_m$  with the DCT angle  $\alpha$ . Conversely,  $CCC(i, j, s, m)$ , for  $i \neq j$ , will give a measure of the similarity of the variation of the textural descriptor  $F_m$  taken from the DCT coefficients at different angles, between two color channels.

Given that the value of the statistical descriptor  $F_m$  already possesses useful information for the texture characterization process, and given that the proposed  $CCC(i, j, s, m)$  features only account for the covariance of  $F_m$  between color channels, it is proposed to also include in the feature set the average value  $E\{F_m\}$ . Note that  $E\{F_m\}$  is only the mathematical expectancy for the value of  $F_m$  and therefore does not possess as much information as the sequence of the different  $F_m$  values taken from the DCT coefficients at different angles, but the inclusion of the full sequence would lead to feature vectors with very high dimensionality, which would compromise the training of the classifier and the speed of the classification step. Furthermore, if the same texture pattern was rotated, the sequence of the different  $F_m$  would be affected, which could decrease the classification performance.

#### IV. IMPLEMENTATION AND RESULTS

The experimental dataset 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 final dataset consisted in 400 normal frames, which were equally divided in two sets, for the MLP network training and testing, and 200 abnormal frames, which were also equally divided in two sets. Examples of the dataset frames can be observed in the figure 3.

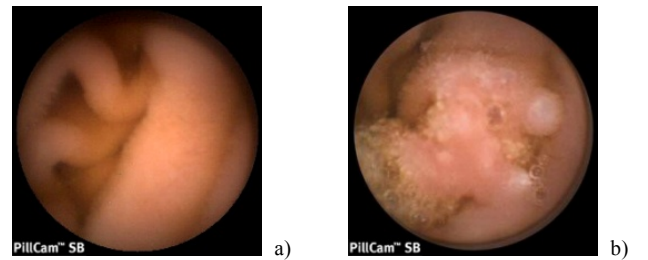


Fig. 3. Example of a normal (a) and an intestinal tumor (b) CE frames

A 2.4 GHz Pentium Core Duo processor-based, with 3 GB of RAM, was used with MATLAB to run the proposed algorithm and the average processing time is approximately 1.5 s. Note however that the implementation of the proposed algorithm was not optimized for speed, so the processing can still be improved. The DCT calculation was done with the routines implemented in the tool CurveLab (available for

research purposes at [www.curvelet.org](http://www.curvelet.org)). A three scales (including the coarsest) DCT was computed for each color channel of the CE video frames, leading to coarsest, medium and finest detail coefficients in the DCT domain. The selected color space was the HSV, since is more similar to the physiological perception of human eye [13], and therefore more adequate than the standard RGB color space.

Given the present framework, different feature sets can be formulated, for comparative purposes. The feature set A was constructed with all the considered  $F_m$  values, the feature set B was constructed with the average values  $E\{F_m\}$ , the feature set C was constructed with the Color Curvelet Covariance values and the feature set D corresponds to features in the feature set B and C. The number of statistical descriptors  $F_m$  used in each dataset was also evaluated. For a clear comprehension, the feature  $A(\mu, \sigma, E, Ent)$  corresponds to values of all the considered statistical textural descriptors of a given CE frame, while the feature set  $B(\mu, \sigma)$  corresponds to the Color Curvelet Covariance values for the statistical textural descriptors  $\mu, \sigma$ .

Table 1 demonstrates that the most relevant information for classification purposes is encoded as high frequency content in the DCT finest detail coefficients. Note that M and F stands for medium and finest detail scales. It was tested also the classification performance of a feature set containing the medium and finest detail, but there was no significant improvement.

TABLE I  
INFLUENCE OF DETAIL LEVEL IN THE CLASSIFICATION PERFORMANCE

Detail Level Feature set	Medium $A(\mu, \sigma)$	Finest $A(\mu, \sigma)$
Specificity (%)	85.0±1.1	96.7±1.2
Sensitivity (%)	80.4±2.2	97.0±0.6

Given the results in table 1, the following tests only considered the finest detail scale DCT coefficients. In table 2 it is possible to observe the results for the different proposed feature sets. Note that the proposed feature set D has a classification performance similar to the feature set A, with considerably smaller dimensionality of the feature vector.

TABLE II  
CLASSIFICATION PERFORMANCE OF DIFFERENT FEATURE SETS

Feature Set	$A(\mu, \sigma)$	$B(\mu, \sigma)$	$C(\mu, \sigma)$	$D(\mu, \sigma)$
Specificity (%)	96.7±1.2	98.2±0.7	95.9±0.5	97.4±0.4
Sensitivity (%)	97.5±0.6	92.6±0.8	95.9±0.9	97.2±1.1

The inclusion of the energy and entropy statistical descriptors in the proposed algorithm was also tested, although no significant improvement was observed in the classification performance comparing with the results achieved with only the mean and standard deviation descriptors, presented in table 2, therefore not being effective the inclusion of these descriptors, since similar results are achieved for a feature vector of twice the size.

TABLE III  
CLASSIFICATION PERFORMANCE OF DIFFERENT TEXTURE DESCRIPTORS

Feature Set	$A(\mu, \sigma, E, Ent)$	$D(\mu, \sigma, E, Ent)$
Specificity (%)	96.6±0.5	97.5±0.3
Sensitivity (%)	98±1.1	97.4±0.6

## V. CONCLUSION AND FUTURE WORK

The more significant information content for classification purposes is encoded as high frequency information, at the DCT scales that correspond to the finest detail coefficients. The results clearly show that the Color Curvelet Covariance features can be successfully used for CE frames classification purposes, being the performance optimal when the average of the statistical textural descriptors is also included in the feature set. The best statistical textural descriptors in the present framework are the mean and the standard deviation of the DCT coefficients, given that the inclusion of the energy and entropy descriptors does not lead to a significant increase in the classification performance.

## REFERENCES

- [1] J. Herrerías and M. Mascarenhas, *Atlas of Capsule Endoscopy*. Sulime Diseño de Soluciones, 2007, Sevilla.
- [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] 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.
- [5] 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.
- [6] C. Lima C, D. Barbosa et al, "Classification of Endoscopic Capsule Images by Using Color Wavelet Features, Higher Order Statistics and Radial Basis Functions," in *Proceedings of the IEEE-EMBC2008*, Vancouver, Canada, 2008, pp 1242-1245.
- [7] D. Barbosa, J. Ramos and C. Lima, "Detection of Small Bowel Tumors in Capsule Endoscopy Frames Using Texture Analysis based on the Discrete Wavelet Transform," in *Proceedings of the IEEE-EMBC2008*, Vancouver, Canada, 2008, pp 3012-3015.
- [8] J. Beck, A. Sutter and R. Ivry, "Spatial frequency channels and perceptual grouping in texture segregation," *Comput. Vision Graph. Image Process*, 37, pp 299-325, 1987.
- [9] L. Baopu and M. Meng, "Texture analysis for ulcer detection in capsule endoscopy images," *Image and Vision Computing*, In Press.
- [10] E. Candes and D. Donoho, "Curvelets, multi-resolution representation, and scaling laws", *Wavelet Applications in Signal and Image Processing VIII*, SPIE vol. 4119, pp 1-12, 2000.
- [11] E. Candes, L. Demanet, D. Donoho and L. Ying, "Fast discrete curvelet transforms", *SIAM Multiscale Modeling Simul.*, 2006.
- [12] L. Dettori and L. Semler, "A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography," *Computers in Biology and Medicine*, vol. 37, no. 2, pp 486-498, 2007.
- [13] L. Baopu and M. Meng, "Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments," *Computers in Biology and Medicine*, vol. 37, no. 4, pp 141-147, 2009