A multi-scale comparison of texture descriptors extracted from the Wavelet and Curvelet domains for small bowel tumor detection in Capsule Endoscopy exams

Daniel J. C. Barbosa, José Higino Correia. and Carlos S. Lima

Abstract— **Traditional endoscopic methods do not reach the entire Gastrointestinal (GI) tract. Wireless Capsule Endoscopy (CE) 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, resulting in huge amounts of data per exam. These 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. Therefore, in the present paper it is proposed a frame classification scheme, based in different combinations of texture descriptors taken at different detail levels of the Discrete Wavelet Transform and Discrete Curvelet Transform domains, in order to compare the classification performance of these multi-resolution representations of the information within the CE frames. 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 91.7% of sensitivity and 89.4% specificity for features extracted from the DWT domain and 94.1% of sensitivity and 92.4% specificity for features extracted from the DCT domain. 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 duodenum, while lower GI endoscopy is limited at the terminal ileum, which means that the vast majority of the small bowel is not visible with these conventional techniques. Furthermore, these medical procedures present other important drawbacks, since they are very uncomfortable to the patients, the navigation of the flexible endoscope is very dependent on the operating physician technical skills, who must be highly trained in order to correctly maneuver the endoscope, and there is always the risk of injuring the GI walls with the tip of the endoscope.

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. Therefore, and given the technical and medical improvements introduced on the assessment of the GI, Capsule Endoscopy (CE) is considered as the first major technological innovation in GI diagnostic medicine since the flexible endoscope [1]. 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 moves through GI tract, propelled exclusively by peristalsis, it acquires images 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. The acquisition of images is limited by the battery life of the device, usually around eight hours, which imply that in a single CE exam more than 50000 images are 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]. Beyond being a boring task, it is prone to errors, as any distraction of the physician may lead to misevaluation of exams. 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 identification of abnormalities in the GI mucosa with texture analysis has been previously reported and from authors' previous work [4,5], the application of texture analysis techniques to classify capsule endoscopic frames is feasible and presents promising results. The proposed method explores the extraction of relevant feature sets at different detail levels of the Discrete Wavelet Transform and the Discrete Curvelet Transform of CE video frames and has been tested in clinical data acquired at the Hospital dos Capuchos.

The classification scheme described in this paper uses 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 (wavelet/curvelet, detail level, descriptors, etc).

Manuscript received April 7, 2009.
José Higino Correia, C. S. Lima and Daniel J. C. Barbosa are with the Industrial Electronics Department, Minho University, Portugal

II. Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) is a mathematical tool that allows a spatial/frequency representation by decomposing the image in different scales with different frequency content, therefore being a multi-resolution representation of the information within the image. It is known for a long time that human perception of texture is based in a multi-scale analysis of patterns [6], which can be modeled by multi-resolution approaches. In fact, the multi-resolution ability of the DWT has been vastly explored in several fields of image processing such as compression, denoising and classification [7]. Furthermore, the spatial/frequency representation of the image content through the DWT, which preserves both global and local information, seems to be an adequate approach for texture characterization. In practice, the 2D DWT can be easily computed applying a separable filter bank to the image:

$$L_n(b) = [H_x * [H_y * L_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2}. \quad (1)$$

$$D_{n1}(b) = [H_x * [G_y * L_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2}. \quad (2)$$

$$D_{n2}(b) = [G_x * [H_y * L_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2}. \quad (3)$$

$$D_{n3}(b) = [G_x * [G_y * L_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2}. \quad (4)$$

where $b \in R^2$, * denotes the convolution operator, $\downarrow 2,1$ ($\downarrow 1,2$) sub-sampling along the rows (columns) and $L_0 = I(b)$ is the original image. H and G are low and band pass filters, respectively. L_n is obtained by low pass filtering, so it is called low resolution image at scale n. The D_{ni} are obtained by band pass filtering in a specific direction. Thus these parameters contain directional detail information at scale n. An example of a two level DWT decomposition can be observed in figure 1. Note that this decomposition allows the separation of image's content in sub-bands of different frequency detail and directional content. However, these three linear directions might not be enough to capture all the complex texture patterns within an image.

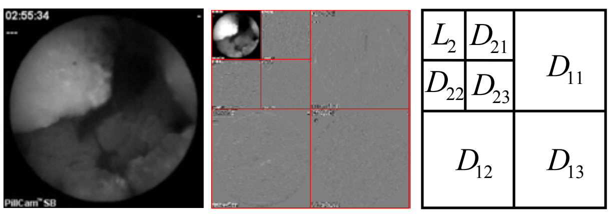


Fig. 1. Two level DWT decomposition of a CE frame

III. Discrete Curvelet Transform

Originally introduced in 2000 by Candes and Donoho, the Continuous Curvelet Transform (CCT) is based in an anisotropic notion of scale and high directional sensitivity in multiple directions [8]. In signal processing, for example, one has to deal with the fact that interesting phenomena occur along curves or sheets, e.g., edges in a two-dimensional (2D) image. 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 research effort to provide better adapted alternatives by combining ideas from geometry with ideas from traditional multi-scale analysis [9]. Therefore, this mathematical tool can be used as a multi-resolution and multi-directional representation of the information within an image.

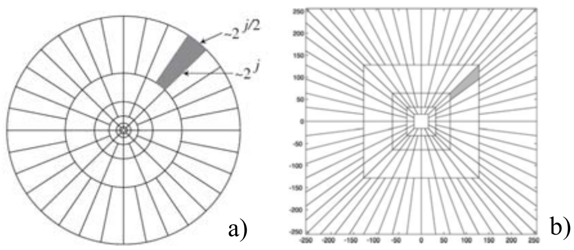


Fig. 2. CCT tiling of the frequency domain (a) and basic tiling of the digital coronization process (b)

The CCT is based in the tilling 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 *Uj* (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:

$$\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1, \qquad r \in (3/4, \; 3/2). \quad (5)$$

$$\sum_{l=-\infty}^{\infty} V^2(t-l) = 1, \qquad t \in (-1/2, \; 1/2). \quad (6)$$

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

$$U_j(r,\theta) = 2^{-3j/4} W(2^{-j} r) V\left(\frac{2^{\lfloor j/2 \rfloor}\theta}{2\pi} \right). \quad (7)$$

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

$$\varphi_{j,l,k}(x) = \varphi_j\left(R_{\theta_l}(x - x_k^{(j,l)}) \right). \quad (8)$$

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

$$c\ (j,l,k) = \langle f, \varphi_{j,l,k} \rangle = \int_{R^2} f(x) \overline{\varphi_{j,l,k}(x)}\ dx. \quad (9)$$

Reference [9] 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, as in the DWT, 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 DWT. 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.

IV. Statistical Texture Descriptors

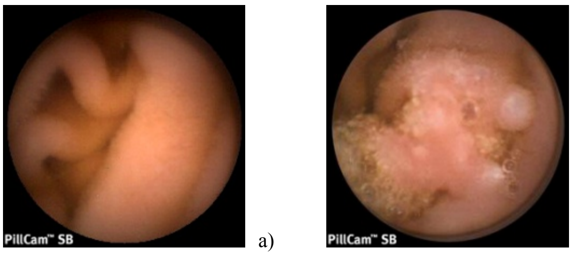
There are several statistical features that can be extracted from the wavelet/curvelet domain as texture descriptors, being the most common the mean, the standard deviation, the energy and the entropy of each sub-band [10]. In the present work, it was decided to start the comparison between the different DWT and DCT detail levels only with the mean and variance as statistical descriptors, in order to better compare the two different multi-resolution domains, being the remaining features added to the feature set, in order to verify if they positively contribute to the classification performance.

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. Therefore, it is vital to only consider the pixels inside this area, since the information regarding to the CE exam is contained in this part.

Since the low frequency components of the images do not contain major texture information, the most important bands in the wavelet/curvelet transform are those in which are present medium and high frequency, texture encoding, information [5]. 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 and DWT are not directional, and consequently do not possess directional sensitivity.

V. 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.



a) b)

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 depends heavily on the selected transform and the number of features (from 0.3s per frame for one feature extracted from the DWT domain to 2s per frame for all the considered features extracted from the DCT domain). Note however that these times can still be greatly improved. The DCT calculation was done with the routines implemented in the tool CurveLab (available for research purposes at www.curvelet.org). A two level DWT and a three scales (including the coarsest) DCT were computed for each color channel of the CE video frames, leading to coarsest, medium and finest detail coefficients for each domain. The selected color space was the HSV, since is more similar to the physiological perception of human eye [12], and therefore more adequated than the standard RGB colorspace.

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.

Table 1 show that the most relevant information for classification purposes is encoded as high frequency content in the DWT and DCT finest detail coefficients. These results correspond to a feature set of mean and variance (feature set A). 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
CLASSIFICATION PERFORMANCE OF DWT VS DCT

Transform Detail Level	DWT M	DWT F	DCT M	DCT F
Specificity (μ±σ%)	84±3.1	89.4±2.1	85.0±1.1	92.4±1.2
Sensibility (μ±σ%))	77.1±4.1	91.7±0.5	80.4±2.2	94.1±0.6

Since it is clear that the finest detail coefficients of both DWT and DCT contain the most relevant information, the inclusion of the entropy and energy descriptors in the feature set was evaluated only at those scales. The feature set B corresponds to the extraction of the mean, variance, entropy and energy from the finest detail scales. The results of this comparison are presented in table 2, where is evident that the increase in classification performance with the addition of entropy and energy to the feature set is not significant.

TABLE II
CLASSIFICATION PERFORMANCE OF DWT VS DCT

Transform Feature Set	DWT A	DWT B	DCT A	DCT B
Specificity (μ±σ%)	89.4±2.1	89.2±1.9	92.4±1.2	91.9±1.5
Sensibility (μ±σ%))	91.7±0.5	92.1±1.1	94.1±0.6	94.3±0.7

VI. CONCLUSION AND FUTURE WORK

The more significant information content for classification purposes is encoded as high frequency information, at the DWT/DCT scales that correspond to the finest detail coefficients. The results clearly show that the descriptors from the DCT domain present a higher discriminative power when compared with DWT descriptors, being the feature set with better classification performance composed by the mean and variance extracted from each sub-band of the considered domains.

Future work will include further investigation of different features extraction schemes from the DCT domain, taking advantage of the high directional sensitivity and multi-resolution information encoding of this mathematical tool.

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] 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 P*roceedings of the IEEE-EMBC2008*, Vancouver, Canada, 2008, pp 1242-1245.

[5] 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.

[6] 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.

[7] L. Baopu and M. Meng, "Texture analysis for ulcer detection in capsule endoscopy images," Image and Vision Computing, In Press.

[8] 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.

[9] E. Candes, L. Demanet, D. Donoho and L. Ying, "Fast discrete curvelet transforms", *SIAM Multiscale Modeling Simul*., 2006.

[10] 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.

[11] J. Starck, N. Aghanim and O. Forni, "Detection and discrimination of cosmological non-Gaussian signatures by multi-scale methods," *Astron. Astrophys,* vol. 416, pp 9-17, 2004.

[12] 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