9

Multiscale Texture Descriptors for Automatic Small Bowel Tumors Detection in Capsule Endoscopy

Daniel Barbosa[1], Dalila Roupar[2] and Carlos Lima[2]
[1]*Katholieke Universiteit Leuven*
[2]*Universidade do Minho*
[1]*Belgium*
[2]*Portugal*

1. Introduction

Conventional endoscopic exams do not allow the entire visualization of the gastrointestinal (GI) tract. Push enteroscopy (PE) is an effective diagnostic and therapeutic procedure, although it only allows exploration of the proximal small bowel (Pennazio et al., 1995). Simultaneously, convetional colonoscopy is limited at the terminal ileum. Therefore, prior to the wireless capsule endoscopy era, the small intestine was the conventional endoscopy's last frontier, because it could not be internally visualized directly or in it's entirely by any method (Herrerías & Mascarenhas-Saraiva, 2007). The small intestine accounts for 75% of the total length and 90% of the surface area of the gastrointestinal tract. In adults it measures about 570 cm at *post mortem*, which is substantially longer than conventional video endoscopes (100-180 cm) (Swain & Fritscher-Ravens, 2004). Intraoperative enteroscopy is the most complete but also the most invasive means of examining the small bowel (Gay et al., 1998).

Given the technical and medical improvements introduced on the assessment of the gastrointestinal (GI) tract, Capsule Endoscopy (CE) is considered as the first major technological innovation in GI diagnostic medicine since the flexible endoscope (Kaffes, 2009). More recently, a new technique, the double-balloon enteroscopy (DBE), has been introduced into clinical practice (Yamamoto & Kita, 2006). DBE has the potential to examine the entire length of the small bowel with biopsy and therapeutic capability. Nevertheless, it is a time consuming procedure that requires specialist training for the operating physician. We should note that DBE and CE are complementary tools and not competitive (Chen et al., 2007). Hence, the diagnostic ease of CE can be complemented with a targeted and often therapeutic DBE (Kaffes, 2009). Therefore, CE can be used as a first line diagnosis method, while DBE can be used as a confirmatory or therapeutic modality for lesions first visualized by CE (Pennazio, 2006).

The endoscopic capsule is a pill-like device, with only 11mm x 26 mm, and includes a miniaturized camera, a light source and circuits for the acquisition and wireless transmission of signals (Iddan et al., 2000). 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 (Pennazio, 2006). Capsule endoscopy has evolved in a few short years to become a first-line, noninvasive diagnostic technique for the small bowel. CE is now being utilized worldwide to assess patients for obscure gastrointestinal bleeding, possible Crohn's disease, celiac disease and small bowel tumors (Lee & Eisen, 2010). It is now available in over 4500 practice sites around the world (Munoz-Navas, 2009).

The time required to a physician to analyze the resulting video is, on average, 40-60 min (Pennazio, 2006). The reading time and interpretation of CE exams is very time consuming given that more than 50,000 images have to be reviewed (Delvaux & Gay, 2006; Mergener et al., 2007), which contributes to the high cost of a CE exam (Westerhof et al., 2009). Thus, a computer assisted diagnosis tool to help the physicians to evaluate CE exams faster and more accurately is an important technical challenge and an excellent economical opportunity.

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 (Herrerías & Mascarenhas-Saraiva, 2007).

The application of texture analysis techniques to classify capsule endoscopic frames is feasible and presents promising results. Kodogiannis *et al.* proposed two different schemes to extract features from texture spectra in the chromatic and achromatic domains (Kodogiannis et al., 2007). The first is a structural approach based in the theory of formal languages, while the second is a statistical approach, where statistical texture descriptors are calculated from the histograms of the RGB and HSV color spaces of CE video frames. Recently, Mackiewicz *et al.* proposed an automatic capsule endoscopy segmentation algorithm based in color and texture features to determine the topological division of capsule endoscopic videos (Mackiewicz et al., 2008). Several other CE image processing methodologies were reviewed recently by Karargyris and Bourbakis (Karargyris & Bourbakis, 2010).

In the present chapter, several multiscale texture descriptors are extracted from both wavelet and curvelet domains and their classification performance is assessed. In section 2, we present a brief introduction to multiscale image representation, namely through Discrete Wavelet Transform and Discrete Curvelet Transform. In section 3, the algorithm used to extract the multiscale texture descriptors is described, as well as modelling techniques that allow to extract statistical dependence of textural descriptors taken in different color channels. In section 4, the implementation details of the method are discussed. In section 5, the performance of the proposed methods is assessed and discussed. Section 6 resumes the key findings and presents future research orientation.

2. Multiscale representation of image information

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. 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. However, the directional information 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 (Candès & Donoho, 2000). 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 (Candès et al., 2006). Therefore, this tool can be used as a multi-resolution and multi-directional representation of the information within an image. The Discrete Curvelet Transform (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.

2.1 Wavelet Transform

The scientific breaktrough achieved with the introduction of the Wavelet Transform has inequivocally changed the research direction on the biomedical signal/image processing. Indeed, since the seminal works of Daubechies (Daubechies, 1988) and Mallat (Mallat, 1989), more than 9000 papers and 200 books were published between the late eighties and 2003, with a significant part being focused in biomedical applications (Unser et al., 2003). While not being exhaustive, and in order to provide to the interested reader an idea of some applications of wavelets in medical imaging, the DWT has been applied in several problems such as:

- Image denoising
- Compression of medical images
- Feature extraction and image classification
- Tomographic reconstruction
- Image Enhancement

Since the main goal of this section is to provide to the reader the key concepts regarding the DWT, the mathematical details will be kept to a minimum. The Continuous Wavelet Transform (CWT) is a signal representation in a scale-time space, and the CWT coefficients of a time-varying signal $x(t)$ are given through:

$$X_{\Psi}(\tau, s) = \int_{-\infty}^{+\infty} x(t)\Psi^{*}\left(\frac{t-\tau}{s}\right) dt \tag{1}$$

where Ψ is the mother wavelet function and * stands for the complex conjugate.
Analogously, signal can be recovered from its wavelet coefficients through the Inverse Continuous Wavelet Transform (ICWT):

$$x(t) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} X_{\Psi}(\tau, s)\Psi\left(\frac{t-\tau}{s}\right) d\tau ds \tag{2}$$

By varying both the scale and the translation shift parameters, s and τ respectively, we can obtain a family of daughter wavelets from the mother wavelet function Ψ:

$$\Psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\Psi\left(\frac{t-\tau}{s}\right), \tag{3}$$

Thus, for a fixed value of the scale parameter s, the CWT, which is now a function of the continuous shift parameter τ, can be written as a convolution equation where the filter corresponds to a rescaled and time-reversed version of the wavelet as shown by equation (1) setting t=0. Combining the variation of the scale parameter s and the time shift parameter τ, the CWT provides a complete scale-time representation of a signal. Furthermore, a given CWT coefficient can be simply seen as the inner product:

$$X_{\Psi}(\tau, s) = \langle x(t), \Psi_{s,\tau} \rangle. \tag{4}$$

Each $\langle x(t), \Psi_{s,\tau} \rangle$ can be seen as a quantom of information that is represented as a rectangle in the time-frequency plane. However, it can be shown through the uncertainty principle theorem that this rectangle has a minimum surface that limits the joint time-frequency resolution. The ultimate consequence of this limit is the trade-off between temporal and frequency resolution. Thus, an increase in the frequency resolution would lead to a decrease in the temporal resolution. In limit, we could perfectly identify the frequency content of the signal and simultaneously lose all the localization details of these frequency components, as in the case of the typical Fourier transform. However, and given the relevance of the temporal localization of abnormal frequency content phenomena, there is the need of a better use of the joint time-frequency resolution. A possible solution could be the use of Short Time Fourier Transform (STFT). However, the constant window size implies a uniform tilling of the time-frequency plane. The wavelet can tackle this problem very elegantly, as will be shown ahead.

As can be easily understood, by varying both the scale s and the time shift τ parameters, the CWT representation of the signal $x(t)$ is highly redundant. Indeed, the CWT maps the information within a one-dimensional signal to a two-dimensional time-scale joint representation. This implies a heavy computational burden which reduces its application potential to daily life problems. Nonetheless, this problem can be tackled by limiting the continuous scalability and translatability of the daughter wavelets $\Psi_{s,\tau}$. Using such scheme, the mother wavelet function Ψ generates a smaller family of daughter wavelets:

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \Psi \left(\frac{t - k\tau_0 s_0^j}{s_0^j} \right) \tag{5}$$

where j and k are integers and s_0>1 a fixed dilation step. The discretization of the time-scale plane is usually achieved using a dyadic sampling, where s_0=2 and τ_0=1.

A wavelet atom $\Psi_{j,k}(t)$ is localized around the point $2^j k$ and has a support size proportional to the scale 2^j. Using this approach, the scale index j corresponds to the level of focus from the which the signal is viewed, which is related to the frequency range involved. Indeed, a lower j corresponds to the high frequency contents, which can be easily deducted from the support size.

In the time-frequency plane, the Heisenberg resolution box of $\Psi_{j,k}(t)$ is a dilation by 2^j and translation by $2^j k$ of the Heisenberg box of Ψ. This leads to a perfect tiling of the time-frequency plane illustrated in Fig. 1. The height and width of the Heisenberg boxes in the time-frequency plane represents the resolution on the frequency and time domains respectively. Note that there is an intrinsic trade-off between good temporal localization of high frequency content and good frequency resolution of low frequency content in the signal, as in opposition to the homogeneous time-frequency resolution of the STFT. This is the central advantage of the Wavelet Transform when compared to the Short Time Fourier Transform.

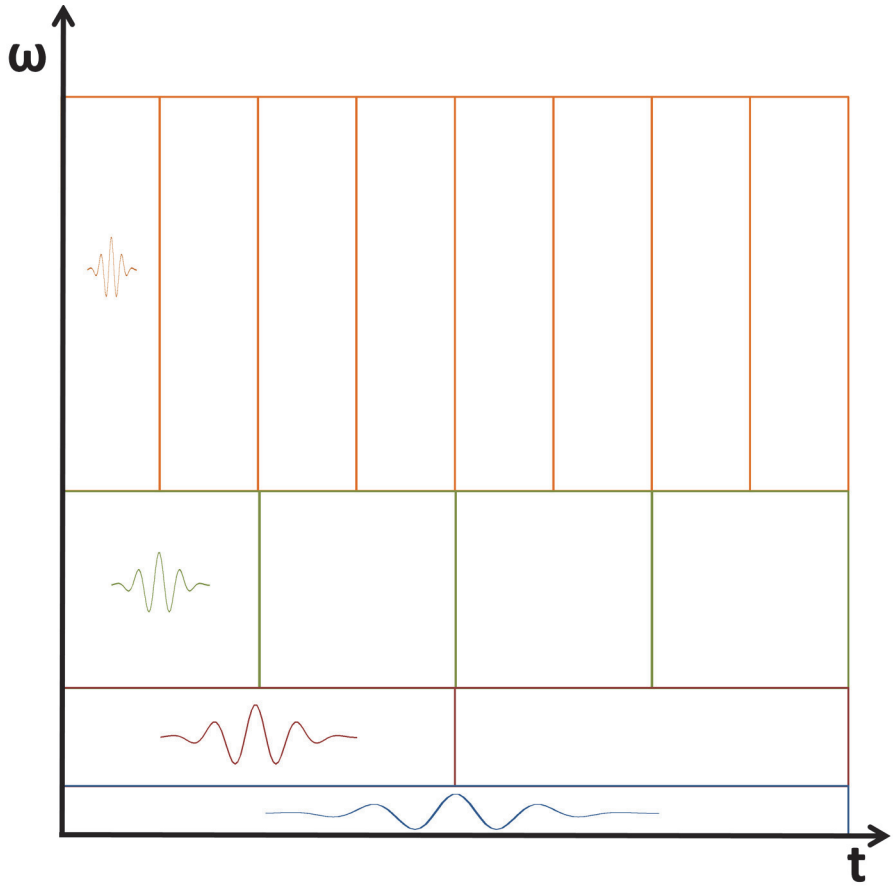

Fig. 1. Inhomogeneous dyadic tilling of the time-frequency plane with wavelet atoms.

From the previous figure, another property of the wavelet atoms $\Psi_{j,k}(t)$ becomes evident: their frequency spectrum has a band-pass nature. Indeed, from Fourier theory, we know that the dilatation operation in the time domain corresponds to a compression in the frequency domain and a shift towards the zero frequency. However, this would imply that an infinite number of scales would be needed to cover the entire frequency spectrum, as can be seen in Fig. 2. Since this would not be a plausible solution, Mallat (Mallat, 1989) introduced the scalling function to cover the spectrum not spanned by the wavelet atoms. Indeed, when analyzing a signal using a combination of a scalling function and wavelets, the scalling function will represent the signal information in the spectrum covered by all the wavelet atoms up to a scale j, while the remaining spectrum is analyzed by wavelets. This can be observed in Fig. 3.
When regarding the schematic representation of the Wavelet Transform in Fig. 3, one can easily observe that the tilling of the spectrum with scaling and wavelet functions is similar to consecutive low-pass and band-pass filter operations. Due to its low-pass nature, the scalling function will allow to extract coefficients that express an approximate version of the signal

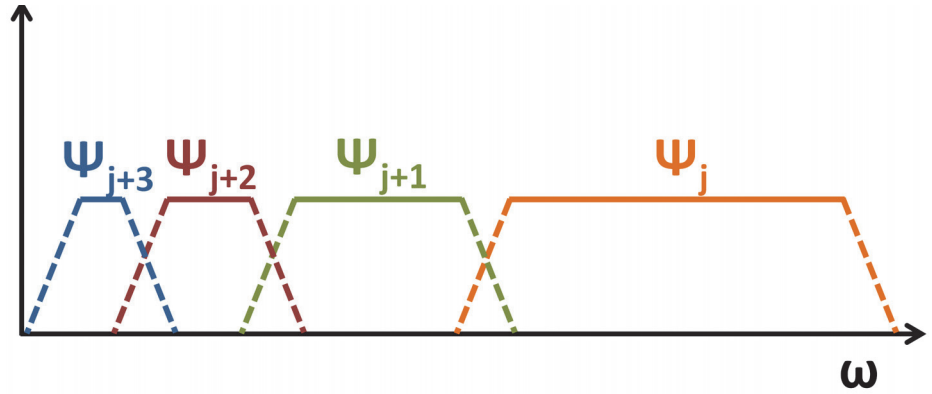

Fig. 2. Wavelet atoms spectrum resulting from dyadic scaling of the mother wavelet Ψ.

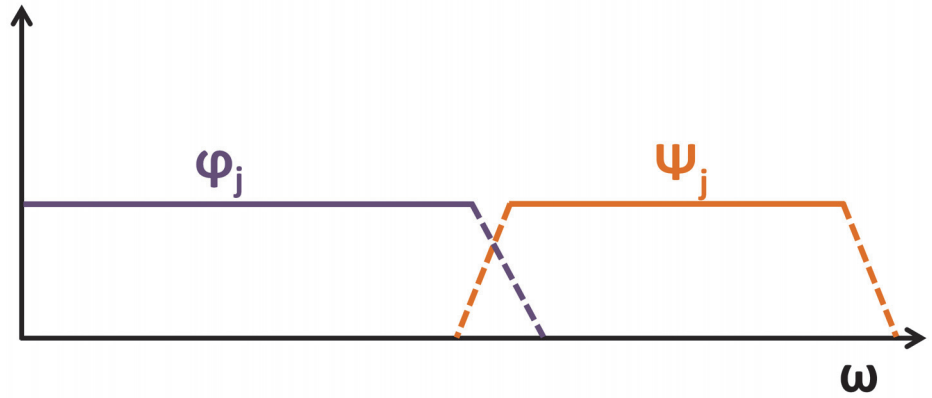

Fig. 3. Scaling function (ϕ) and wavelet atoms (Ψ) spectra. Note how the scaling function (ϕ_j) covers the spectral region spanned by the wavelet atoms until $j+1$.

content, whereas the wavelet coefficients correspond to detail information at different levels of focus. Although this concept, commonly designated as subband coding, is not unique from the Wavelet Transform, it greatly helps its practical implementation.

2.1.1 Discrete Wavelet Transform

In several practical problems, the signal under analysis is of discrete nature. Thus, there is the need to discretize the Wavelet Transform in order to use it in sampled discrete signals. This can be efficiently done using recursive filtering in iterated filter-bank as show in Fig. 4. Indeed, it can be shown that both scaling and wavelet coeficcients at a given scale can be calculated from the coefficients at the previous scale using appropriate filters. Furthermore, half of the samples after applying the filter can be eliminated according to the Nyquist's rule, since the signal now has only half of the bandwith.

There is an important relationship between these low and high-pass used recursively in the iterated filter-bank. Indeed, these filters are not independent from each other and satisfy the

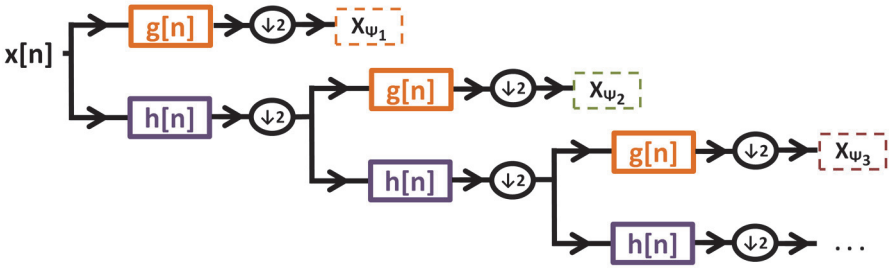

Fig. 4. Iterated filter-bank for practical implementation of the Discrete Wavelet Transform.

condition of Quadrature Mirror Filters (QMF):

$$g[L-1-n] = (-1)^n \cdot h[n], \tag{6}$$

where $g[n]$ is the high-pass filter, $h[n]$ is the low-pass filter and L is the filter lenght. The consecutive filter and subsampling operation at each scale can be expressed by:

$$y_{approx} = \sum_n x[k] \cdot h[-n+2k], \tag{7}$$

$$y_{detail} = \sum_n x[k] \cdot g[-n+2k], \tag{8}$$

Note that, given the finite time resolution of sampled signals, there is an upper limit in the scale at which the signal is analysed, which is of course dependent on the low-pass filter length. However, it is common that the DWT analysis is not done until this limit, being thus the output of the low-pass filtering corresponding to the scalling function coefficients. However, and for sake of readability, we shall globally denominate the coefficients arising from the application of these iterated filter-banks simply as DWT coefficients.
As for the continuous case, the original signal can be recovered from its DWT coefficients applying Inverse Discrete Wavelet Transform (IDWT). In order to reconstruct the signal, the iterated filter-banks has to be reversed. Thus, at each scale, the DWT coefficients are upsampled, by simply introducing a zero between every two samples, and then filter the signal with synthesis filters, closely related with the analysis filters used in the direct DWT. However, and in order to achieve perfect reconstruction, the analysis and synthesis filters must fulfill the following conditions (Daubechies & Sweldens, 1998):

$$\tilde{h}(z)h(z^{-1}) + \tilde{g}(z)g(z^{-1}) = 2, \tag{9}$$

$$\tilde{h}(z)h(-z^{-1}) + \tilde{g}(z)g(-z^{-1}) = 0, \tag{10}$$

where g and h are the analysis filters used in the direct DWT and $\tilde{h}$ and $\tilde{g}$ are the synthesis filters used in the inverse DWT.

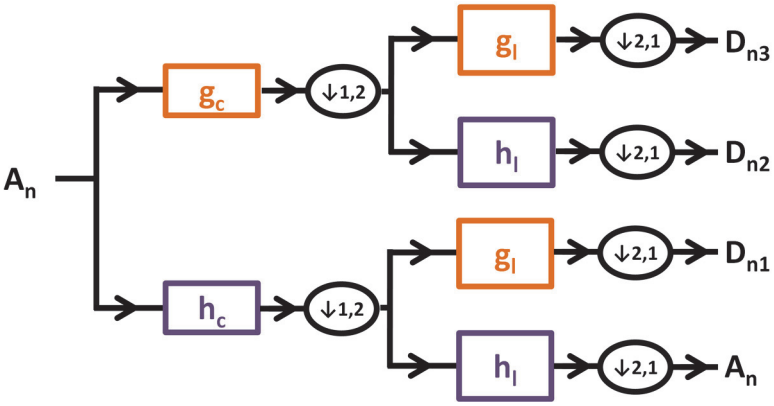

Fig. 5. Iterated filter-bank for 2D Discrete Wavelet Transform in images. Note that A will keep an approximation of the original image, whereas D will carry detail information.

2.1.2 2D Discrete Wavelet Transform

Given the one-dimensional nature of the signals under analysis in the Wavelet Transform framework, the direct application of this mathematical tool to higher dimensional signals is not directly possible. However, a N-D signal can still be analyzed along each of its N dimensions. Using this reasoning, Mallat also introduced in his seminal paper a very elegant extension of the concepts of multi-resolution decomposition in order to use the DWT in image processing problems. The key idea is to expand the application of 1D filterbanks to the 2D in a straightforward manner, by simply applying the designed QMF filters to the columns and rows separately. Thus, the DWT applied to an image can be implemented using an iterated filter-bank, illustrated in Fig. 5, which can be described as consecutive filtering and subsampling operations:

$$A_n = [h_l \star [h_c \star A_{n-1}]_{\downarrow 1,2}]_{\downarrow 2,1}, \tag{11}$$

$$D_{n1} = [g_l \star [h_c \star A_{n-1}]_{\downarrow 1,2}]_{\downarrow 2,1}, \tag{12}$$

$$D_{n2} = [h_l \star [g_c \star A_{n-1}]_{\downarrow 1,2}]_{\downarrow 2,1}, \tag{13}$$

$$D_{n3} = [g_l \star [g_c \star A_{n-1}]_{\downarrow 1,2}]_{\downarrow 2,1}, \tag{14}$$

where $(i, j) \in \mathbb{R}^2$, $\star$ denotes the convolution operator, $\downarrow 1,2$ ($\downarrow 2,1$) is the sub-sampling operation applied to the columns (lines) and A_0 is the original image I. The low and high-pass filter h and g are the QMF filters referred in the previous section. As shown nicely in Fig. 6, A_n arises from a low-pass filtering operation, being thus an approximated version of the image at scale n. On the other hand, D_{n1} and D_{n2} arise from high-pass filter in a specific direction, namely horizontal and vertical, while low-pass filtering is applied in the other direction. Thus, these coefficients held detail information along a specific direction in the image, at scale n. Lastly, D_{n3} arises from high-pass filtering in both horizontal and vertical directions, possessing thus information regarding details in the diagonal direction. This behaviour can be easily observed when applying the DWT transform to a square binary image, as shown in Fig. 6.

From the previous image, another important characteristic of DWT is evident: it provides a sparse representation of the information present in image. This has been a feature widely used in compression schemes, being perhaps the JPEG2000 the best know example.

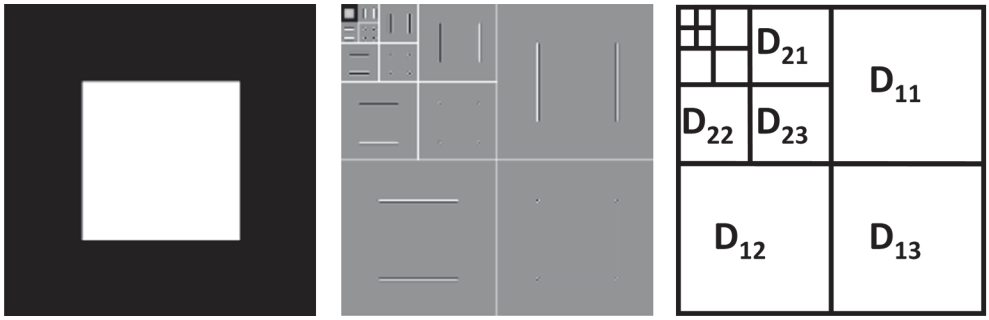

Fig. 6. Example on the directional sensitivity of the 2D Discrete Wavelet Transform (right: original image, center: DWT coefficients up to j=4, left: correspondence between the DWT coefficients and the equations (11))-(14)).

Note that, as for the 1D case, the original image can still be recovered from its DWT coefficients, trough inverse transform. Therefore, the DWT can be manipulated in order to enhance features, which can be then synthesized to a new image using IDWT .

2.2 Curvelet transform

The multi-resolution capability of the DWT has been vastly explored in several fields of signal and image processing, as seen in the last section. The ability of dealing with singularities is another important advantage of the DWT, since wavelets provide an optimal representation for one-dimensional piecewise smooth signal (Do & Vetterli, 2005). However, and as seen in the previous section, the application of DWT to multidimensional data is done in a separate way along each dimension. As can be easily understood, natural images are not simply stacks of 1-D piecewise smooth scan-lines, and therefore singularities points are usually located along smooth curves rather than having an independent location. Indeed, these intermediate dimensional structures like discontinuities along curves often provided structures of interest and relevant information within the image. However, and being the DWT directional sensitivity limited to three directions, there is the need to develop mathematical tools to overcome this limitation. Thus, the limitations of the DWT triggered the quest for new concepts capable of overcome these limits. The alternatives proposed until now make use of a combination between concepts from traditional multi-scale analysis and ideas taken from geometry.

Given the focus of the present book being the DWT and its applications, the description of multi-directional and multi-scale transforms will be kept brief and mostly conceptual. The first attempts to extract multi-scale information at different orientations date from the beginning of the nineties, with the introduction of steerable pyramids. This was the first approach to this problem, being a practical, data-friendly strategy to extract information at different scales and angles. More recently, more advance techniques have been proposed, such as the Curvelet Transform (Candès & Donoho, 2000) and the Contourlet Transform (Do & Vetterli, 2005). These new and promising image analysis tools are already starting to prove its usefulness.

The Continuous Curvelet Transform has been introduced by Candès and Donoho and has as key concept the anisotropic notion of scale and high directional sensitivity in multiple directions (Candès & Donoho, 2000). Thus, and contrarily to DWT, the orientational sensitivity is not limited to the horizontal, vertical and diagonal directions but rather span

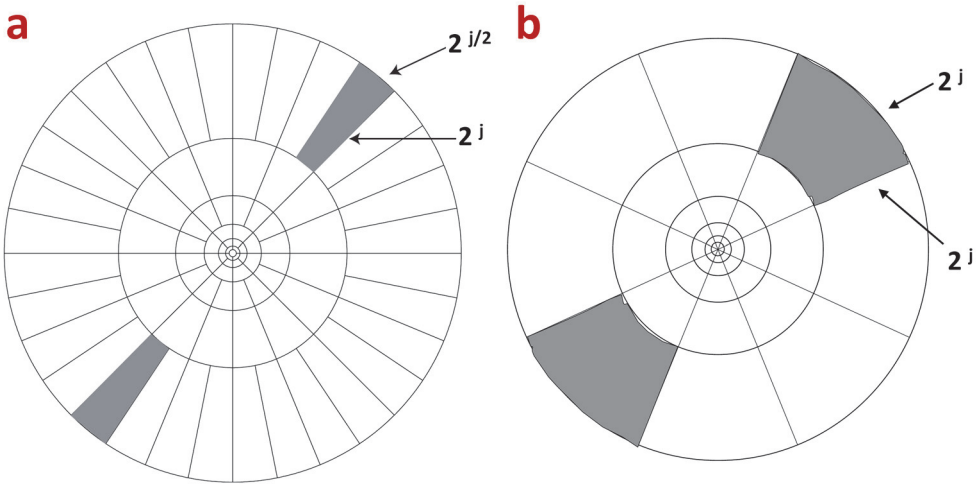

Fig. 7. Tilling of the 2D Fourier polar plane for the Curvelet Transform (a) and the 2D Discrete Wavelet Transform (b). Note that the radius stands for the frequency, where higher radius will encode higher frequency information, and the angle for the orientation.

a wider set of orientations, which stems from the anisotropic notion of scale (parabolic rescaling). In fact, the CCT is based in the tilling of the 2D Fourier space (which has both scale and orientation information) in different concentric coronae, one of each divided in a given number of angles, accordingly with a fixed relation shown in Fig. 7. As it can be observed, there is still a division of the scale information (the center of the 2D Fourier plane corresponds to low frequencies, while the outer regions to higher frequency components), being the directional sensitivity dependent on the scale.

Each of these different regions corresponds to a different curvelet. Indeed, these polar wedges can be defined by the superposition of a radial window and an angular window. To each of this polar wedges, a tight frequency window or coronae can be associated in the 2D Fourier space. This frequency window will then correspond to the Fourier transform of a curvelet function $\psi_{j,\theta}$ function. In fact, and for a single scale, all $\psi_{j,\theta}$ may be obtained by rotations and translations of a mother curvelet ψ_j. As for the wavelet case, the curvelets coefficients will then simply arise from the inner product between the image and the rotation/translation of the mother curvelet.

In the seminal work of Candès and Donoho a discretization scheme was also proposed. However, its complexity led to further research, from which arose the conceptually simpler, faster and less redundant second generation of curvelets, proposed in 2006 (Candès et al., 2006). However, it must be noted that the Curvelet Transform is continuous in its nature and will always have to be redesigned in order to be applied to discrete image data. Keeping this in mind, Do and Vetterli have introduced the Contourlet Transform, a discrete mathematical tool having similar multi-resolution and directional sensitivity characteristics (Do & Vetterli, 2005). However, the Contourlet Transform is formulated from a double filter bank approach, as shown in Fig. 8. The first stage corresponds to Laplacian pyramid which separates the image content into different detail levels. The second step is a a directional filter bank which links

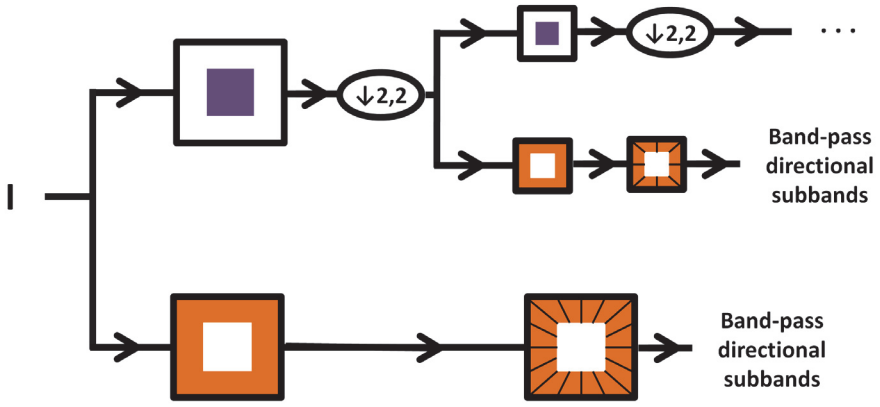

Fig. 8. Two-stage iterated filter-bank for implementation of Contourlet Transform.

point discontinuities into linear structures. Thus, there is multi-scale and multi-directional decomposition of the image content.
Although sharing similar fundamental concepts, Curvelet and Contourlet transforms are not equivalent. Indeed, the theoretical support of Curvelet Transform outranks the Contourlet Transform, in the sense that the curvelet elements have a sharply defined location in the frequency plane, materialized as a polar wedge (in contrast to other approaches with discrete nature). On the practical side, this means that contourlets lack smoothness along the ridge in the spatial domain and exhibit spurious oscillations which may be the source of numerous problems, especially if one wants to use these transforms for scientific computing. On the other hand, the Contourlet Transform is directly designed for discrete applications, whereas the discretization scheme of the curvelet transform faces some intrinsic challenges in the sampling of the Fourier plane in the outermost coronae. Furthermore, the Contourlet Transform presents less redundancy and has a lower computational burden.
As a final note, Curvelet coefficients are, as in the DWT, an 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.

3. Feature extraction algorithm

3.1 Multiscale image pre-processing

In order to extract different scale information from capsule endoscopic frames, both Discrete Wavelet Transform and Discrete Curvelet Transform (DCT) were applied to CE data. 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 both domains. Note that in the coarsest detail level there is no directional information for both transforms. On the other hand higher detail levels possess directional information. The DWT has 3 sub-bands of different directional information, while the DCT has a number of subbands dependent on the parameters used. In the present case, there were 8 and 16 sub-bands of different directional information in the medium and finest detail scales of the

Fig. 9. Example of a capsule endoscopic frame (center) and its DWT coefficients (left) and DCT coefficients (right). Note that the number of coefficients in the DCT image is not equivalent to the other two images.

DCT respectively. Nonetheless, a note should be made regarding the number of coefficients arising from these two transforms. While the DWT leads to a compact representation, with a number of coefficients equal to the number of pixels in the image under analysis, the DCT results in a higher number of coefficients. For illustrative purposes, the DWT and the DCT coefficients of a CE frame are show in Fig.9. Note that the DCT is far less compact.
Color transformations of the original image I result in three decomposed color channels:

$$I^i, \quad i = 1, 2, 3, \tag{15}$$

where i stands for the color channel.
A two level DWT is applied to each color channel, I^i, as shown in Fig. 9. This transformation results in a new representation of the original image by a low resolution image and the detail images, possessing higher frequency content in three orientations. Therefore the new representation is defined as:

$$\gamma^i_{DWT} = DWT(I^i) = \{A^i_n, D^i_{s,d}\}, \quad n=2,\ s=1,2,\ d=1,2,3 \tag{16}$$

where d stands for the wavelet directional sub-band, s for scale and n is the decomposition level.
A three scale DCT is applied to each color channel, I^i, as shown in Fig. 9. This transformation results in a new representation of the original image by a low resolution image and the detail images, possessing higher frequency content in multiple orientations. Therefore the new representation is defined as:

$$\gamma^i_{DCT} = DCT(I^i) = \{A^i_n, D^i_{s,d}\}, \quad n=3,\ s=1,2,\ d=1,\cdots,8 \ \vee \ d=1,\cdots,16 \tag{17}$$

where d stands for the curvelet directional sub-band, s for scale and n is the decomposition level. Note that, contrarily to DWT, the number of directional sub-bands at each scale depends on the scale itself. Thus, for the first scale there are 16 directions, while for the second there are 8 directions.
After the application of DWT and DCT transform to the image data, its content is divided in different multiscale and multi-directional information. It should be stressed that the

textural information is usually better presented in the middle scale coefficients channels. Thus, mid scale coefficients should be considered in theory. However, the relatively low image dimensions (256×256) limit the representation of the details, becoming the first level more adequate, than previously expected, for texture representation Barbosa et al. (2008). Nevertheless an image resolution normalization might be required for different image resolutions.

3.2 Statistical texture features

There are several statistical features that can be extracted from the wavelet and curvelet domains as texture descriptors, being the most common the mean, the standard deviation, the energy and the entropy of each DWT/DCT sub-band (μ, σ, E, Ent) (Dettori & Semler, 2007). Nonetheless, in authors' previous work it was observed that the introduction of energy and entropy in the feature set did not significantly contributed to an increase in the classification performance (Barbosa, Correia, Ramos & Lima, 2009; Barbosa, Ramos, Correia & Lima, 2009). Thus, only the mean and standard deviation in each DWT/DCT sub-band were computed. The proposed texture descriptors can be calculated as:

$$\mu = E\{P(i,j)\} = \frac{1}{N}\sum_i\sum_j P(i,j) \tag{18}$$

$$\sigma = \sqrt{E\{(P(i,j)-\mu)^2\}} = \sqrt{\frac{1}{N}\sum_i\sum_j (P(i,j)-\mu)^2} \tag{19}$$

where $P(i,j)$ corresponds to the pixel value at position (i,j). 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 Fig.9. 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. In order to cope with this constraint, the summation limits of the equations (18)-(19) were set in order to correspond to this area.

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. Since the low frequency components of the images do not contain major texture information, the most important scales in the DWT and DCT will be those in which are present medium and high frequency, texture encoding, information. Furthermore, the coarsest scale coefficients of the DCT and DWT are not directional, and consequently do not possess directional sensitivity. Therefore, each of the afore mentioned statistical features is computed at each sub-band of the DWT/DCT coefficients, for all the color channels, at medium and finest detail level.

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 (Karkanis et al., 2003), for colonoscopy videos. Therefore, 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 DWT coefficients. This was recently suggested by the authors and shown in (Barbosa, Ramos, Correia & Lima, 2009). In the present chapter, this methodology is followed and the scope of the comparison of this recently proposed algorithm and other existing features is expanded. In the Color Curvelet Covariance

framework the covariance of textural descriptors in the different color channels is used as a classification feature. The Color Wavelet/Curvelet Covariance of a texture descriptor can be calculated as:

$$C*C(a,b,s,m) = \sum_{\alpha} ((F_m(a,s,\alpha) - E\{F_m(a,s,\alpha)\}) \times (F_m(b,s,\alpha) - E\{F_m(b,s,\alpha)\})) \quad (20)$$

where a and b represent the different color channels in the covariance calculation, F_m is the statistic textural descriptor, α is the considered angle of the DWT/DCT coefficients, s the considered detail scale and $E\{Fm(a,s,\alpha)\}$ the average of the statistical textural descriptor F_m over the different angles α, in the color channel a. As it is clear from Fig.9, the Color Wavelet Covariance features will arise from a covariance of only three different angles, while the Color Curvelet Covariance features will take more angles into account in this computation. One should note that if $a = b$, the $C*C$ features will enconde the angular variance of the textural descriptor F_m. On the other hand, for $a \neq b$, the $C*C$ features will give a measure of the similarity of the angular covariation of F_m between two color channels.

The proposed $C*C$ features allow to capture additional information regarding the texture patterns and can be regarded as second order statistical modelling of the texture descriptors F_m. However, the texture descriptors F_m already possess information regarding the existing textures in the image. Thus, it is proposed to add the average value $E\{F_m\}$ to the feature set. It should be referred that $E\{F_m\}$ corresponds to 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/DWT coefficients at different angles. Nonetheless, the inclusion of the entire F_m feature set helds dimensionality problems, which can compromise the training and consequently the performance 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.

4. Implementation issues

4.1 General considerations

A 2.8 GHz Intel i7 dual core processor, with 4 GB of RAM, was used with MATLAB to run the proposed algorithm. The average processing time is between 0.2s (DWT) and 0.75s (DCT) for each CE frame. Note however that the implementation of the proposed algorithm was not optimized for speed, so the processing time can still be improved.

The DWT calculation was done using the MATLAB Toolbox Wavelets (made available by G. Peyré at http://www.ceremade.dauphine.fr/~peyre/matlab/wavelets/content.html). The basis used was the standard Daubechies wavelet with 4 vanishing moments.

The DCT calculation was done with the routines implemented in the toolbox CurveLab (available for research purposes at www.curvelet.org). The method chosen for the discretization of the Curvelet Transform was the wrapping algorithm, for the reasons referred in (Candès et al., 2006).

For both DWT and DCT, the reader is highly advised to explore the tutorial and demos offered in the used toolboxes in order to better understand the underlying principles of these mathematical tools.

The selected color space was the HSV, since it is more similar to the physiological perception of human eye (Li & Meng, 2009), and therefore more adequate than the standard RGB color space.

4.2 Feature sets

In order to assess the impact of the choice of texture descriptors, several feature sets were generated for the the analyzed data. The feature set A corresponds to a feature set comprising the mean and standard deviation of each DWT/DCT sub-band. A more advanced feature set is also formulated by using the covariance of the textural descriptors, as well as their average value for all the analyzed sub-bands. The covariance+average approach is applied to the texture descriptors mean and standard deviation and results in the feature set B. These different feature sets were computed for both the medium and finest detail scales of the DWT/DCT transforms, since it is known from before hand that low detail coefficients do not possess relevant texture information.

4.3 Classification scheme

The features were imported into the open source machine learning package WEKA (available at http://www.cs.waikato.ac.nz/ml/weka/). A stratified 10-fold cross-validation procedure was chosen to train a standard multilayer perceptron neural network. The default parameters were kept in the classifier options. The choice of a simple classification scheme, with default parameters, was done in order to make the results more representative of the choice of the features. The 10-fold cross-validation method is a standard procedure to validate machine learning classification outputs and has been found to provide an adequate and accurate estimate of the true error rate (Alpaydin, 2004). The 10-fold cross-validation algorithm splits the data into 10 partitions, where the proportion of both normal and abnormal frames in each partitions is similar to the entire dataset. The training and classification process is then repeated 10 times, where 9 partitions are used to train and 1 partition is used to assess the classification process. This way, each frame will be used exactly once as test data, allowing to efficiently use the available dataset. In order to have an accurate error estimate, the cross-validation process was repeated 10 times, being this a standard procedure (Witten & Frank, 2005). Note that the splitting of the training and testing datasets in the 10 repetitions of the 10-fold stratified cross-validation were similar for the different feature sets, allowing thus a paired comparison.

4.4 Evaluation of classification performance

The ultimate goal of the proposed method is to provide a binary classification of whether a given CE frame presents patterns of abnormal tissue. Thus, it is vital to understand how such classification performance is assessed. Confusion matrices offer a simple yet effective way of visualizing the classification errors:

		Outcome of classification test	
		Normal	Tumor
Ground Truth	Normal	TN	FP
	Tumor	FN	TP

A false negative, FN corresponds to a CE frame presenting tumor pathology which is misclassified as normal, while a false positive, FP, corresponds to a normal frame wrongly

considered as abnormal. While the overall accuracy can be defined as:

$$Accuracy = \frac{TP + TF}{TP + TF + FP + FN}, \tag{21}$$

it is more frequent to report the performance of a binary classification test using sensitivity and specificity:

$$Sensitivity = \frac{TP}{TP + FN}, \tag{22}$$

$$Specificity = \frac{TN}{TN + FP}. \tag{23}$$

There is an intrinsic trade-off between sensitivity and specificity of a classifier, which can be adapted to each situation in order to penalyze a specific type of error. This can be nicely visulized through Receiver Operating Characteristic (ROC) curves. Although no single number is able to capture this trade-off, being thus such 2D representation needed, the area under the ROC curve is usually used as a measure of the classifier performance (Witten & Frank, 2005).

Lastly, it was decided to include the Cohen's kappa coefficient, k, in the set of classification evaluation measures, since it is a statistical measure of agreement between the classifier output and theground truth. Furthermore, k takes into account the agreement that may occur simply by chance, being thus a better performance metric for the classification performance. It can be simply calculated as:

$$K = \frac{P(a) - P(e)}{1 - P(e)}, \tag{24}$$

where $P(a)$ is the relative observed agreement and $P(e)$ the probability of agreement due to random chance. Note that relative agreement is not more than simply the classification accuracy.

In order to assess the statistical significance of differences in the classification, a standard paired t-test was employed.

5. Experiments and results

5.1 Dataset

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 and 196 abnormal frames. Examples of the dataset frames can be observed in the Fig.10. These frames have been manually classified by the expert physician.

5.2 Influence of the scale on the classification Performance

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. Table 1 shows that the most relevant information for classification purposes is encoded as high frequency content in both DWT and DCT finest detail coefficients. Note that M and F stands for medium and finest detail scales. For comparison purposes, the feature sets A was chosen. The values in Table 1 correspond to the mean and standard deviation for a ten-fold stratified cross-validation scheme repeated 10 times.

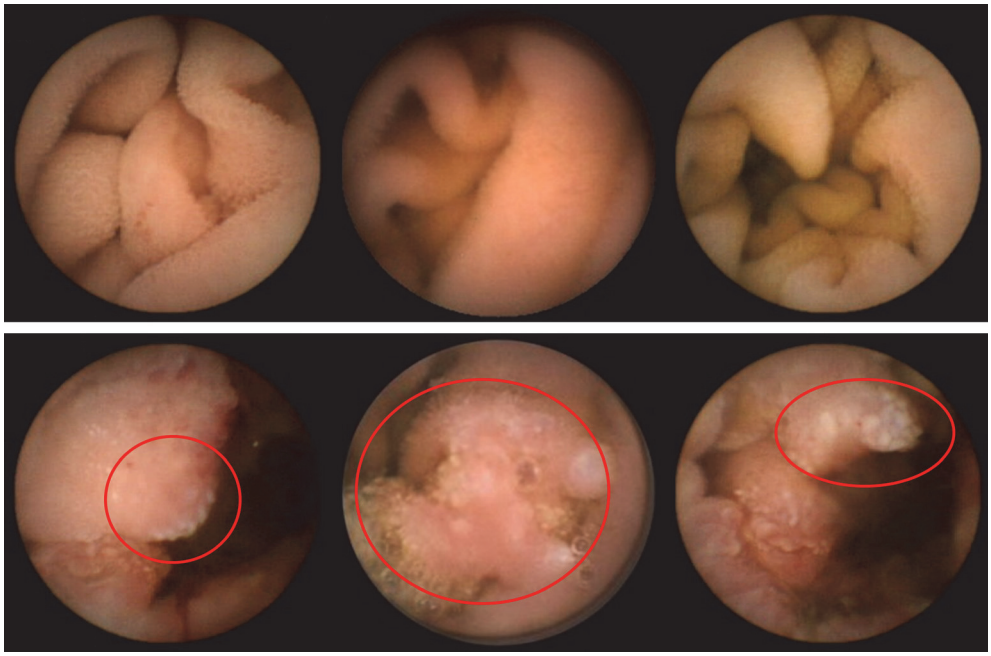

Fig. 10. Examples of Wireless Capsule Endoscopy video frames (top: normal frames, bottom: frames containing tumoral tissue).

Transform	**DWT**		**DCT**	
Feature Set	**A**		**A**	
Detail Level	**M**	**F**	**M**	**F**
Features (#)	18	18	48	96
Accuracy (%)	77.15 ± 1.56	96.34 ± 0.42	87.44 ± 0.96	95.54 ± 0.31
Specificity (%)	84.78 ± 3.23	97.34 ± 0.29	90.61 ± 2.29	96.76 ± 0.37
Sensitivity (%)	61.64 ± 4.60	94.29 ± 1.10	81.01 ± 3.39	93.13 ± 0.62
ROC Area	0.834 ± 0.012	0.990 ± 0.003	0.936 ± 0.008	0.989 ± 0.002
Cohen's K	0.473 ± 0.032	0.917 ± 0.010	0.716 ± 0.020	0.899 ± 0.007

Table 1. Influence of the Detail Level in the Classification Performance

From the results in the previous table, there is no clear advantage arising from the use of DCT, although better results are observed for the medium detail coefficients. However, the large difference between the number of features extracted does not allow a strong conclusion in this comparison, since it is known that large feature vectors pose difficulties to the training algorithm. Thus, the key observation from the results presented in Table 1 is that for both DWT and DCT domains, the features extracted from the sub-band corresponding to the higher detail yield indeed a better classification performance.

Transform (Detail Level)	DWT (F)		DCT (F)	
Feature Set	A	B	A	B
Features (#)	18	18	96	18
Accuracy (%)	96.34 ± 0.42	96.30 ± 0.45	95.54 ± 0.31	97.65 ± 0.55
Specificity (%)	97.34 ± 0.29	97.45 ± 0.65	96.76 ± 0.37	98.47 ± 0.58
Sensitivity (%)	94.29 ± 1.10	94.04 ± 1.40	93.13 ± 0.62	96.0 ± 0.61
ROC Area	0.990 ± 0.003	0.990 ± 0.003	0.989 ± 0.002	0.994 ± 0.003
Cohen's K	0.917 ± 0.010	0.916 ± 0.010	0.899 ± 0.007	0.947 ± 0.012

Table 2. Color Covariance Texture Descriptors Classification Performance

5.3 Color covariance texture descriptors

In the present section, the added value for the color covariance texture descriptors is shown. Given the results presented in the previous section, this approach was applied only to the finest detail coefficients of both DWT and DCT transforms, as these possess the most relevant texture information for classification purposes. The results are shown in the Table 2. The values correspond to a ten-fold stratified cross-validation scheme repeated 10 times.

Using the Color Covariance approach, a visible increase in the classification performance of the features extracted from the DCT domain was observed. Nonetheless, in the performance of DWT features there were no significant changes. These results support the hypothesis that $C*C$ features can extract complex texture patterns between color channels, leading to good texture classificantion performance. The Color Curvelet Covariance approach is able to significantly reduce the feature set dimensionality, while improving its classification performance. These results also point that features extracted from the DCT domain possess more relevant information for classification purposes when compared to the DWT. This can be intrinsically linked with the higher directional sensitivity of the DCT, which helps to better handle complex two-dimensional patterns such as texture.

5.4 Statistical analysis of the classification performance

Given the results shown in subsection 5.2, the statistical analysis was limited to the features extracted from the highes detail coefficients of both DWT and DCT domains. The significance of the statistical differences found between the different feature sets extracted from both domains can be observed in the Table 3, for the metrics used to evaluate the classification performance. For sake of clarity, NS stands for not statistically significantly different, † for $p < 0.05$ and ‡ for $p < 0.001$. Note also that $DWT(F, A)$ stands for the feature set A extracted from the finest detail level of the DWT transform, while $DCT(F, B)$ corresponds to the feature set B extracted from the DCT coefficients corresponding to the finest detail level.

From these results, some important highlights appear, confirming statistically some previous observations. First and foremost, the proposed Color Curvelet Covariance approach presents significantly higher classification performance for all the metrics used to compare the different methods. Second, the $C*C$ approach in the DWT domain does not lead to a significant difference in the classification performance. Last, the feature vector A extracted from the DCT scale of finest details helds a much higher dimensionality than the remaining ones. Thus, no strong conclusions can be taken from the statistical comparison with the remaining feature sets, since the high dimensionality of this feature set may pose a problem to the performance of the classifier.

Accuracy

Transform(Detail Level,Feature Set)	DWT(F,A)	DWT(F,B)	DCT(F,A)	DCT(F,B)
DWT(F,A)		NS	‡	‡
DWT(F,B)			‡	‡
DCT(F,A)				‡
DCT(F,B)				

Specificity

Transform(Detail Level,Feature Set)	DWT(F,A)	DWT(F,B)	DCT(F,A)	DCT(F,B)
DWT(F,A)		NS	†	‡
DWT(F,B)			†	‡
DCT(F,A)				‡
DCT(F,B)				

Sensitivity

Transform(Detail Level,Feature Set)	DWT(F,A)	DWT(F,B)	DCT(F,A)	DCT(F,B)
DWT(F,A)		NS	†	‡
DWT(F,B)			†	‡
DCT(F,A)				‡
DCT(F,B)				

ROC Area

Transform(Detail Level,Feature Set)	DWT(F,A)	DWT(F,B)	DCT(F,A)	DCT(F,B)
DWT(F,A)		NS	†	‡
DWT(F,B)			†	‡
DCT(F,A)				‡
DCT(F,B)				

K

Transform(Detail Level,Feature Set)	DWT(F,A)	DWT(F,B)	DCT(F,A)	DCT(F,B)
DWT(F,A)		NS	†	‡
DWT(F,B)			†	‡
DCT(F,A)				‡
DCT(F,B)				

Table 3. Statistical Significance of the Differences for the Classification Accuracy

6. Conclusion and future work

The more significant information content for classification purposes is encoded as high frequency information. Indeed, texture corresponds to medium/high frequency image content. However, and contrarily to what has been found in (Karkanis et al., 2003) for colonoscopy videos, the most relevant sub-bands for texture classification correspond to the highest detail levels, for both DWT and DCT domains.

The proposed color covariance approach achieves an optimal balance between feature vector size and classification performance, presenting promising results that support the feasibility of the proposed method. The Color Curvelet Covariance method yielded the higher classification accuracy, achieving $97.65 \pm 0.55\%$ of correctly classified frames.

Future work will include the extraction of different texture descriptors from the DCT and the use of different classifiers. The enlargement of the available dataset will be equally pursued, in order to further validate the observed results over a wider range of images and pathologies.

7. References

Alpaydin, E. (2004). *Introduction to machine learning*, MIT Press.

Barbosa, D., Correia, J. H., Ramos, J. & Lima, C. S. (2009). A multi-scale comparison of texture descriptors extracted from the wavelet and curvelet domains for small bowel tumor detection in capsule endoscopy exams, *Proc. World Congress on Medical Physics and Biomedical Engineering 2009*, pp. 1546–1549.

Barbosa, D., Ramos, J., Correia, J. H. & Lima, C. (2009). Automatic detection of small bowel tumors in capsule endoscopy based in color curvelet covariance statistical texture descriptors, *Proc. 31th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC2009)*, pp. 6683 – 6686.

Barbosa, D., Ramos, J. & Lima, C. S. (2008). Detection of small bowel tumors in capsule endoscopy frames using texture analysis based on the discrete wavelet transform, *Proc. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC2008)*, pp. 3012 –3015.

Candès, E., Demanet, L., Donoho, D. & Ying, L. (2006). Fast discrete curvelet transforms, *Multiscale Modeling and Simulation* 5(3): 861–899.

Candès, E. & Donoho, D. (2000). Curvelets, multi-resolution representation, and scaling laws, *Wavelet Applications in Signal and Image Processing VIII, SPIE vol. 4119* pp. 1–12.

Chen, X., Ran, Z. & Tong, J. (2007). A meta-analysis of the yield of capsule endoscopy compared to double-balloon enteroscopy in patients with small bowel diseases, *World Journal of Gastroenterology* 13(32): 4372–4378.

Daubechies, I. (1988). Orthogonal bases of compactly supported wavelets, *Communications on Pure and Applied Mathematics* 41: 909 – 996.

Daubechies, I. & Sweldens, W. (1998). Factoring wavelet transforms into lifting steps, *Journal of Fourier Analysis and Applications* 4(3): 247–269.

Delvaux, M. & Gay, G. (2006). Capsule endoscopy in 2005: Facts and perspectives, *Best Practice & Research Clinical Gastroenterology* 20(1): 23 – 39.

Dettori, L. & Semler, L. (2007). A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography, *Computers in Biology and Medicine,* 37(2): 486–498.

Do, M. N. & Vetterli, M. (2005). The contourlet transform: an efficient directional multiresolution image representation, *IEEE Transactions on Image Processing* 14(12): 2091–2106.

Gay, G., Pennazio, M., Delmotte, J. & Rossini, F. (1998). *Atlas of Enteroscopy*, Springer, chapter Intraoperative enteroscopy, p. 51 – 54.

Herrerías, J. & Mascarenhas-Saraiva, M. (2007). *Atlas of Capsule Endoscopy*, Sulime Diseño de Soluciones, Sevilla.

Iddan, G., Meron, G., Glukhovsky, A. & Swain, P. (2000). Wireless capsule endoscopy, *Nature* 405(6785): 417–417.

Kaffes, A. (2009). Achieving total enteroscopy with capsule endoscopy in all patients: are we stretching the limits of technology?, *Gastrointestinal Endoscopy* 69(1): 81 – 83.

Karargyris, A. & Bourbakis, N. (2010). Wireless capsule endoscopy and endoscopic imaging: A survey on various methodologies presented, *IEEE Engineering in Medicine and Biology Magazine* 29(1): 72 –83.

Karkanis, S., Iakovidis, D., Maroulis, D., Karras, D. & Tzivras, M. (2003). Computer-aided tumor detection in endoscopic video using color wavelet features, *IEEE Transactions on Information Technology in Biomedicine* 7(3): 141–152.

Kodogiannis, V., Boulougoura, M., Wadge, E. & Lygouras, J. (2007). The usage of soft-computing methodologies in interpreting capsule endoscopy, *Engineering Applications of Artificial Intelligence* 20(4): 539–553.

Lee, N. & Eisen, G. (2010). 10 years of capsule endoscopy: An update, *Expert Reviews Gastroenteroly and Hepatoly* 4(4): 503–512.

Li, B. & Meng, M. (2009). Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments, *Computers in Biology and Medicine* 39(2): 141–147.

Mackiewicz, M., Berens, J. & Fisher, M. (2008). Wireless capsule endoscopy color video segmentation, *IEEE Transactions on Medical Imaging* 27(12): 1769 –1781.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representatio, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(7): 674–693.

Mergener, K., Ponchon, T., Gralnek, I., Pennazio, M., Gay, G., Selby, W., Seidman, E. G., Cellier, C., Murray, J., de Franchis, R., Rösch, T. & Lewis, B. S. (2007). Literature review and recommendations for clinical application of small-bowel capsule endoscopy, based on a panel discussion by international experts, *Endoscopy* 39(10): 895–909.

Munoz-Navas, M. (2009). Capsule endoscopy, *World Journal of Gastroenterology* 15: 1584–1586.

Pennazio, M. (2006). Capsule endoscopy: Where are we after 6 years of clinical use?, *Digestive and Liver Disease* 38(12): 867 – 878.

Pennazio, M., Arrigoni, A., Risio, M., Spandre, M. & Rossini, F. (1995). Clinical evaluation of push-type enteroscopy, *Endoscopy* 27(2): 164–70.

Swain, P. & Fritscher-Ravens, A. (2004). Role of video endoscopy in managing small bowel disease, *Gut* 53(12): 1866–1875.

Unser, M., Aldroubi, A. & Laine, A. (2003). Guest editorial: Wavelets in medical imaging, *IEEE Transactions On Medical Imaging* 22(3): 285–288.

Westerhof, J., Koornstra, J. & Weersma, R. (2009). Can we reduce capsule endoscopy reading times?, *Gastrointestinal Endoscopy* 69(3, Part 1): 497 – 502.

Witten, I. H. & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann.

Yamamoto, H. & Kita, H. (2006). Double-balloon endoscopy: From concept to reality, *Gastrointestinal Endoscopy Clinics of North America* 16(2): 347 – 361. Small Bowel Enteroscopy.