EDGE DETECTION IN BIOMEDICAL IMAGES

HOŠŤÁLKOVÁ EVA, PROCHÁZKA ALEŠ
ICT Prague, Technická 5, 166 28 Prague 6, Czech Republic
Email: Eva.Hostalkova@vscht.cz, Ales.Prochazka@ieee.org

Abstract: Image edges convey undoubtedly the most important part of the percepted information about the objects depicted in an image. Edge detection and sharpening contribute extensively to visual enhancement and objects boundaries detection. The term *edge* denotes an abrupt change in the values of image intensity. Sharp edges of a step-function profile may by easily detected by short gradient masks. In blurred and noisy images, it is more convenient to detect edges of different spatial sizes by multi-scale methods, such as the Canny detector. In this paper, all the edge detection techniques are applied to biomedical images. Prior to edge extraction, we carry out noise reduction by wavelet coefficients shrinkage. For this purpose, we use either the Discrete Wavelet Transform (DWT) or the Dual-Tree Complex Wavelet Transform (DTCWT) designed by N. Kingsbury. The latter outperforms the DWT by its approximate shift invariance and better directional selectivity. Furthermore, owing to their sparsity and persistence, the DTCWT coefficients may be modeled by Hidden Markov models (HMM) and utilized for edge detection.

Keywords: Edge detection, wavelet transform, wavelet shrinkage, gradient masks, Sobel filter, Canny filter, dual-tree complex wavelet transform, hidden Markov models

1 INTRODUCTION

Image edges are most important for image perception and object boundaries detection [Rangayyan, 2005]. The term *edge* denotes an abrupt change in the values of image intensity represented by high frequencies in the Fourier domain. Sharp edges of a steep step-function profile may by easily detected by short gradient masks [Petrou and Bosdogianni, 2000; Hlaváč and Sedláček, 2007]. However, these short-tap filters proof insufficient for blurred or noisy edges. Extension of the filter tap length leads to blurring the originally sharp edges. It is more convenient to detect edges of different spatial sizes by multi-scale methods, such as the Canny detector [Petrou and Bosdogianni, 2000; Hlaváč and Sedláček, 2007] having the form of the first derivative of Gaussian with variable standard deviation.

In this paper, we detect edges in biomedical computed tomography (CT) and magnetic resonance (MR) images of the brain. As a preprocessing step, we attempt to reduce the amount of noise by soft global thresholding of wavelet coefficients. For this purpose, we use either the Discrete Wavelet Transform (DWT) or the Dual-Tree Complex Wavelet Transform (DTCWT) designed by N. Kingsbury [Kingsbury, 2000].

Thanks to the sparsity and persistence properties of the DTCWT coefficients, this complex wavelet representation may be modeled as a Hidden Markov model (HMM) and thus utilized for edge detection [Romberg et al., 2001]. Let us begin with describing the DTCWT algorithm.

2 THE DUAL-TREE COMPLEX WAVELET TRANSFORM

Despite its usefulness in many signal processing applications, the critically sampled discrete wavelet transform (DWT) encounters the following limitations [Selesnick et al., 2005]:

- *Oscillations* of coefficients at a singularity (coefficients at a singularity are expected to be large at all scales, however, some values can be very small due to zero-crossings)
- *Shift variance* (coefficients do not simply shift after a shift in the signal, they change chaotically)
- *Aliasing* due to downsampling and non-ideal filtering during the analysis (canceled out by the synthesis filters only provided that the coefficients are not altered

- *Lack of directional selectivity* in higher dimensions (only 3 directional subbands in 2D; unable to distinguish between $+45°$ and $-45°$ edge orientations)

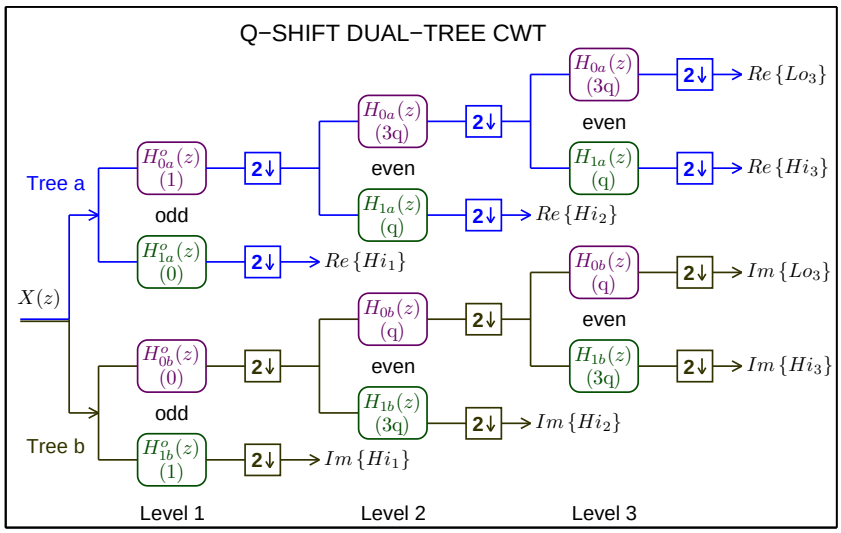

Figure 1 – Outline of 1-dimensional dual-tree complex wavelet transform analysis including convolution (multiplication in the Z-domain) with the corresponding filter and downsampling by 2 at each level

The dual-tree complex wavelet transform (DTCWT) is designed [Kingsbury, 2000; Selesnick et al., 2005] to overcome these problems taking inspiration from the Fourier transform which is shift invariant. A large magnitude in this representation implies the presence of a singularity while the phase signifies its position within the support of the basis function. Magnitude is shift invariant and phase linearly encodes the shift. To provide such magnitude-phase representation, the DTCWT needs employ *analytic* or *quadrature* wavelets.

An analytic wavelet $\psi_c(t) = \psi_r(t) + j \cdot \psi_i(t)$ is composed of two real wavelets $\psi_r(t)$ and $\psi_i(t)$ forming a Hilbert transform (HT) pair given as

$$\psi_i(t) = HT\{\psi_r(t)\} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\psi_r(t)}{t-\tau}d\tau = \psi_r(t)\frac{1}{\pi\, t} \tag{1}$$

and for their respective Fourier transform pairs $\Psi_r(\omega)$ and $\Psi_i(\omega)$

$$\Psi_i(\omega) = FT\{\psi_i(t)\} = FT\{HT\{\psi_r(t)\}\} = -j \cdot sgn(\omega)\Psi_r(\omega) \tag{2}$$

which means that they are orthogonal, i.e. shifted by $\pi/2$ in the complex plain [Shukla, 2003]. The overall Fourier spectrum of such quadrature filters is single-sided (zero for negative frequencies $\omega < 0$). Thus half the bandwidth is spared and aliasing is almost suppressed which is substantial for desired shift invariance.

Nevertheless, it is impossible to design analytic wavelets with finite (compact) support as the HT is infinitely extended in both time and frequency domain. As an implication, the HT pair of a wavelet function of finite support has infinite support Selesnick et al. [2005]. Due to this fact, wavelets may only be designed approximately analytic, shift invariant, and aliasing-free.

The dual-tree approach to complex wavelet transform construction is $2^d : 1$ redundant in comparison with the DWT in d-dimensional space which is however still far less than that for the undecimated DWT. As implied in its name, the DTCWT is realized in two filter bank trees *a* and *b* composed of real-valued filters producing real and imaginary parts of the complex coefficients, resp., as outlined in Fig. 1.

There are several solutions of filter design for the DTCWT. All these solutions endeavour for the filters to fulfil the *half-sample delay condition* [Selesnick et al., 2005] which guarantees their analicity. It states that the scaling filters h_{0a} and h_{0b} are in quadrature when offset from one another by a half sample

$$h_{0b}(n) = h_{0a}(n - 0.5) \tag{3}$$

As a result, the lowpass filters of one tree interpolate midway between the lowpass filters of the other tree and the DTCWT is shift invariant. Unfortunately, the half-sample delay condition cannot be implemented by filters with compact support which always have a certain non-zero gain in their stopband and aliasing cannot be eliminated. That is why the lowpass filters h_{0a}, h_{0b}, same as the highpass filters h_{1a}, h_{1b} of both trees, are only in *approximate* quadrature.

For this paper, we chose the *q-shift* filter design [N. G. Kingsbury, 2003] for the DTCWT. These filters got their name from their symmetry around the point $n = 0.5(N-1) - 0.25$ which is a quarter of a sampling period away from the center. As a consequence, the phase response of these filters is approximately linear while the overall phase response of the complex filters is exactly linear.

So as the Q-shift filters provide perfect reconstruction of the input signal x, the synthesis filters of each tree g_{0a}, g_{1a} and g_{0b}, g_{1b} form orthogonal pairs with the corresponding analysis filters h_{0a}, h_{1a} and h_{0b}, h_{1b} such as for the filter length N

$$h_{0b}(n) = h_{0a}(N - 1 - n) \tag{4}$$

As shown in Fig. 1, filters h^o_{0a}, h^o_{1a}, h^o_{0b}, and h^o_{1b} of the first level are different from the others and we may use any biorthogonal filters of our choice. To pick up opposite samples of the input in both trees, there is a one-sample offset between trees *a* and *b*. Past level 1, we use q-shift filters of a chosen length and the delays in one tree are 1/2 sample different form the opposite tree in order to get uniform interval between the samples of both trees and to satisfy (3).

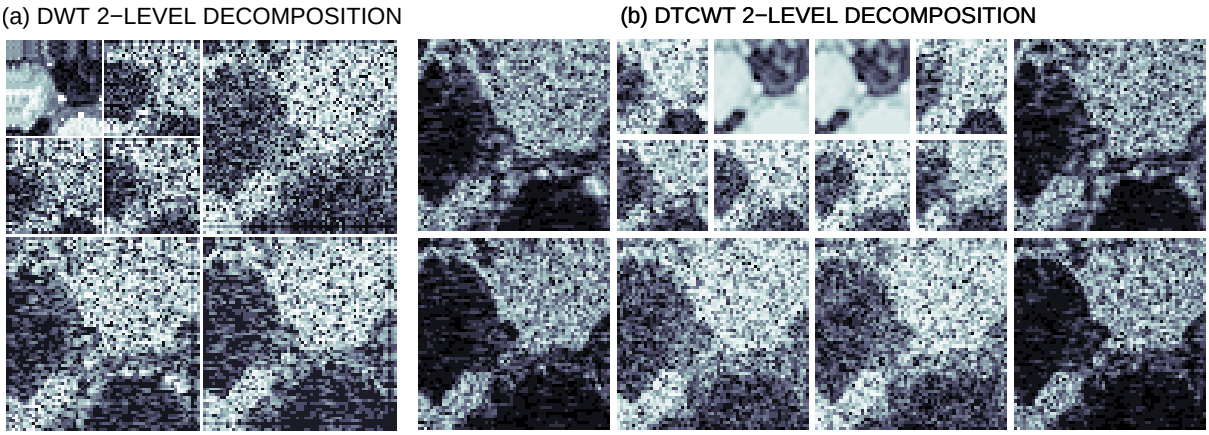

Figure 2 – Wavelet analysis into the second level of a MR brain image exploiting **(a)** the discrete wavelet transform (DWT) using the Daubechies 8-tap filter and **(b)** the dual-tree complex wavelet transform (DTCWT) with 16-tap q-shift filters [N. G. Kingsbury, 2003]

For image analysis, improved directional selectivity is another advantage of the 2-dimensional DTCWT over the DWT. While the critically sampled 2D DWT generates three directional subbands per level conveying image features oriented at the angles of $90°$, $\pm 45°$, and $0°$, the 2D DTCWT produces six directional subbands to reveal the details of an image in $\pm 15°$, $\pm 45°$ and $\pm 75°$ directions.

Fig. 2 displays a 2-level decomposition of a MR image by the both transforms. Decomposition in both cases runs first on columns and then on the rows of the resulting matrix. The output becomes the input for the next level. The 2D DTCWT is 4-times more computation expensive than the 2D DWT which is a little price for the improved directional selectivity and approximate shift invariance.

3 DENOISING THROUGH WAVELET SHRINKAGE

In this paper, we use both the DWT and the DTCWT for noise reduction by wavelet coefficients shrinkage assuming the noise additive and dominant in higher frequencies while the signal is assumed dominant in lower frequencies. We decide on soft global threshold supposing that large coefficients contain a similar amount of noise as the smaller ones.

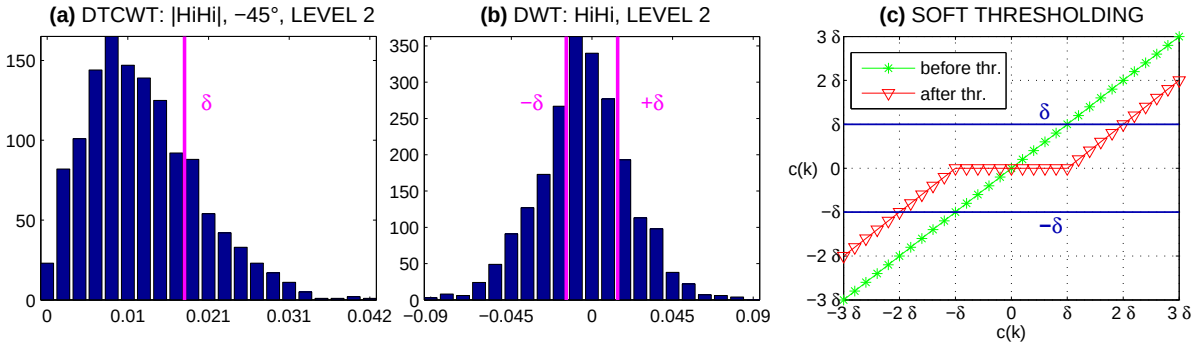

Figure 3 – Soft thresholding of a CT image for noise reduction presenting histograms of **(a)** the magnitudes of the DTCWT coefficients, and **(b)** the DWT coefficients of a selected subband before wavelet shrinkage, and **(c)** the soft thresholding scheme

For soft global thresholding, modification of the wavelet coefficients of all levels $\{c(k)\}_{k=0}^{M-1}$ where M is their total number by the threshold δ results in zeroing out coefficients in the absolute value smaller than δ and decreasing those larger than δ by δ according to Fig. 3c and the formula

$$\overline{c}_s(k) = \begin{cases} \text{sign } c(k)\ (|\,c(k)\,| - \delta) & \text{if } |\,c(k)\,| > \delta \\ 0 & \text{if } |\,c(k)\,| \leq \delta \end{cases} \tag{5}$$

To settle the thresholding level, we need estimate the noise standard deviation. For this purpose, the median absolute deviation (MAD) estimator is computed from the first level HiHi coefficients, which contain the highest frequencies and thus are supposed to be noise dominated

$$\hat{\sigma}_{mad} = \frac{median\{|c_1^{hh}(0)|, |c_1^{hh}(1)|, \ldots, |c_1^{hh}(N/4-1)|\}}{0.6745} \tag{6}$$

where $|c_1^{hh}(n)|$ is the absolute value of the n-th HiHi coefficient of level 1 and N is the image size. The constant in the denominator applies to independent identically distributed Gaussian noise. The median approach is robust against large deviations of noise variance.

The estimated value of the standard deviation is used to calculate the soft global threshold limit from the Donoho formula

$$\delta = \sqrt{2\ \hat{\sigma}_{mad}^2\ log(N)} \tag{7}$$

After thresholding, the image is reconstructed from the altered wavelet coefficients leaving the scaling coefficients unchanged. In case of the DTCWT, we threshold only coefficients magnitudes preserving their phase. Choice of wavelet filters and the number of decomposition levels is influenced by visual consideration of the results. The noise reduction is used in this work as a preliminary step to edge detection described in the next section.

4 EDGE DETECTION

4.1 Gradient Masks

By an *edge*, we understand a singularity of the pixel intensity represented by high frequencies in the Fourier domain. The simplest method of edge detection [Petrou and Bosdogianni, 2000; Rangayyan, 2005; Hlaváč and Sedláček, 2007] is application of short masks approximating the intensity gradient. The mask centered each time at a different *root pixel* slides along the image. It is convenient for the mask to have an odd size. At each shift, we compute 2D convolution between the filter and the corresponding area of the image and assign a new value to the root pixel.

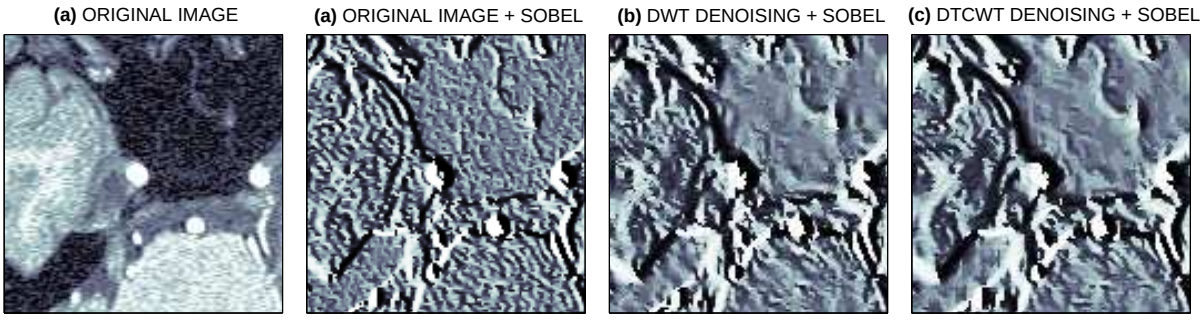

Figure 4 – Results of convolution with the Sobel filter displaying **(a)** the original brain axial MR image and application of the Sobel filter **(b)** on the original image, **(c)** the image after DWT denoising, and **(d)** on the image after DTCWT denoising

The Sobel filter given by the following matrices represents one of the gradient approximating masks

$$\mathbf{Q^{(1)}}=\begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \mathbf{Q^{(2)}}=\begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{pmatrix} \mathbf{Q^{(3)}}=\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \mathbf{Q^{(4)}}=\begin{pmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix} \quad (8)$$

By rotation, this mask approximates the gradient in all eight possible directions to detect horizontal edges ($0°$), diagonal edges ($\pm 45°$) and vertical edges $90°$. For every root pixel, we chose the orientation with the absolute maximum value of convolution with the intensity of the neighboring pixels.

Such short-tap edge detectors are sensitive to noise since designed to operate on sharp edges of a steep step-function profile. Their performance under noisy conditions may be improved by implementing the denoising process as a preprocessing step as shown in Fig. 4.

When applied straight to blurred and noisy images, these short-tap filters either fail to detect an edge or tend to give false alarms. However, extending the tap length leads to blurring the originally sharp edges. The problem lies in attempting to detect edges of different spatial sizes by a single-scale filter. It is more convenient to analyze images by multi-scale methods, such as the Canny detector.

4.2 Canny Edge Detector

The Canny filter [Rangayyan, 2005; Hlaváč and Sedláček, 2007] approximates the first derivative of a 2D Gaussian in the direction of the gradient. This method is robust against noise owing to the use of a Gaussian filter to smooth the data prior to edge detection and a selective algorithm for weak edges pixels identification.

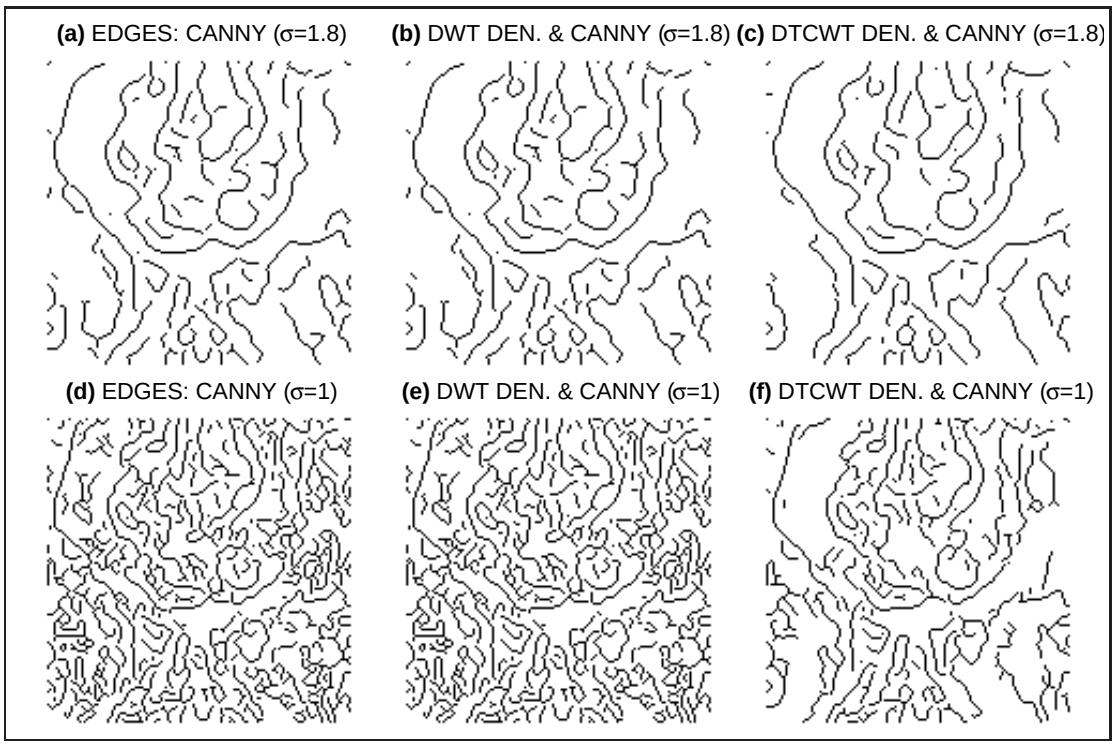

Figure 5 – Edge extraction using the Canny method depicting the edges obtained for $\sigma = 1.8$ and $\sigma = 1$ from **(a)**, **(d)** the original CT brain image, **(b)**, **(e)** the image after denoising using the DWT, and **(c)**, **(f)** the DTCWT

A 2D Gaussian low-pass filter is separable. Hence we may subsequently convolve the image with two 1D Gaussian masks in the row and the column direction x and y, resp. For the standard deviation σ, a zero-mean Gaussian is given as

$$G_{\sigma,0}(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{9}$$

The outcome of smoothing is then convolved with the derivatives of the 2D Gaussian in the x-direction

$$\frac{\partial G_{\sigma,0}(x,y)}{\partial x} = -\frac{x}{\sqrt{2\pi}\sigma^3} \cdot exp\left(-\frac{(x^2+y^2)}{2\sigma^2}\right) \tag{10}$$

and in the y-direction. These two results are then combined together into a matrix whose elements are thresholded to identify pixels corresponding to strong edges. The rest of the pixels may be assigned to weak edges only if their values are greater than the lower threshold limit and if their gradient corresponds to the direction of the stronger edges in the neighborhood. The edges after erosion extracted from the CT image at two different scales σ are shown in Fig. 5.

By adjusting the value of σ, the Canny detector may operate at various scales. For the MR image in Fig. 6 we selected $\sigma = 2.5$ and for the CT image in Fig. 7 $\sigma = 1.8$. Both figures display the extracted edge images combined with the de-noised images.

4.3 Hidden Markov Models

The DTCWT coefficients may also be utilized for edge detection as they depend on the height and the location of an edge. Since independent on the edge height, the phase of the coefficients provides

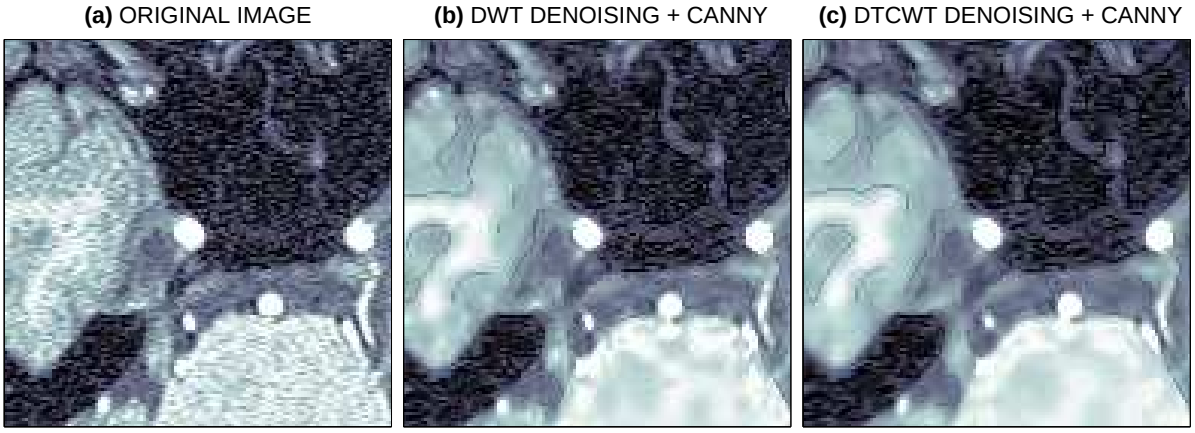

Figure 6 – Axial MR brain image processing presenting **(a)** the original image, **(b)** and **(c)** images processed by the Canny edge detector after de-noising by wavelet shrinkage exploiting the DWT (14-tap symlet filters, 3 levels) and the DTCWT (14-tap q-shift filters, 3 levels), resp.

us with information about the location. Owing to their sparsity and persistence, the DTCWT coefficients may be modeled by Hidden Markov models (HMM) [Shaffrey, 2003; Choi and Baraniuk, 2001; Reeves and Kingsbury, 2000] and thus utilized for edge detection.

The *sparsity* property is entailed by a large number of small coefficients from smooth regions and fewer large coefficients corresponding to singularities. Hence the marginal distribution of the coefficients magnitudes of scale n may be modeled as a 2-component mixture of Rayleigh distributions $R(\sigma_n)$ of a small variance $\sigma_{n,S}$ and a large variance $\sigma_{n,L}$. Using the Rayleigh distribution on magnitudes is implied by the assumption of independent identically distributed zero-mean Gaussian distribution of the real and the imaginary parts which form approximate Hilbert pairs [Shaffrey et al., 2002]. Markovian dependencies do not tie together the coefficients magnitudes but the hidden states s_n taking on values $q = S, L$ with the probability mass function (pmf) $p(s_n = q)$. The overall marginal density function [Romberg et al., 2001] is given as

$$f(|c_n|) = p(s_n = q)\, f(|c_n| \,|\, s_n = q) \tag{11}$$

where the conditional probability of the coefficients magnitude $|c_n|$ given the state s_n corresponds to the Rayleigh distribution

$$f(|c_n| \,|\, s_n = q) = \frac{|c_n|}{\sigma_{n,q}^2}\, exp\left(-\frac{|c_n|^2}{2\,\sigma_{n,q}^2}\right) \tag{12}$$

where the variance $\sigma_{n,q}^2$ depends on the state q and the scale n when *tying* [Crouse et al., 1998] over the whole scale.

The *persistence* property denotes strong parent-child relations in the decomposition hierarchy. In one dimension, the DTCWT runs in binary trees so as each parent coefficient has two children. In images, quad trees are employed and thus each parent has four children. The relative size of the coefficients propagates through their children across scale. To describe these dependencies, our 2-state model uses state transition probabilities between the hidden states of the parent and its children

$$f(s_n = m \,|\, s_{p(n)} = n)) = \begin{pmatrix} f(s_n = S \,|\, s_{p(n)} = S) & f(s_n = L \,|\, s_{p(n)} = S) \\ f(s_n = S \,|\, s_{p(n)} = L) & f(s_n = L \,|\, s_{p(n)} = L) \end{pmatrix} \tag{13}$$

where according to the persistence assumption $f(s_n = S \,|\, s_{p(n)} = S) \gg f(s_n = L \,|\, s_{p(n)} = S)$ and $f(s_n = L \,|\, s_{p(n)} = L) \gg f(s_n = S \,|\, s_{p(n)} = L)$ and $s_{p(n)}$ is the state of a parent node.

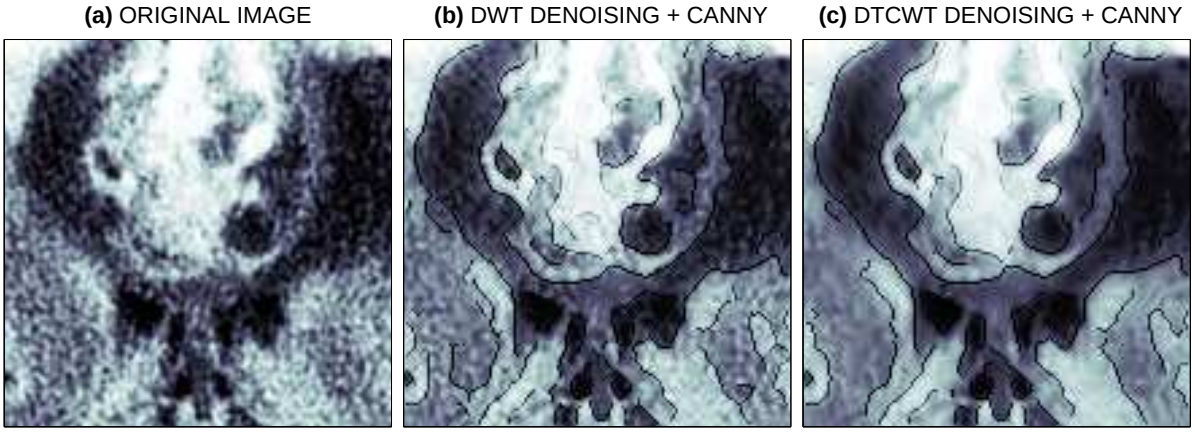

Figure 7 – Canny edge detection applied to a CT brain image presenting **(a)** the original CT image with a large malignant tumor gliom IV (tumor - white, necrosis - dark spots inside the tumor, edema around the tumor - dark), **(b)** and **(c)** images processed using the Canny detector (scale $\sigma = 1.8$) after de-noising by wavelet shrinkage exploiting the DWT (16-tap symlet filters) and the DTCWT (16-tap q-shift filters), resp.

For the modeling purposes, the DTCWT is more suitable than the DWT for its approximate shift invariance of magnitude which does not oscillate across scale at the location of a singularity and for its near linear phase encoding.

The HMM may be trained by various algorithms such as the EM (also forward-backward) algorithm based on an iterative expectation maximization and very well described in Crouse et al. [1998].

As suggested in Romberg et al. [2001], it is possible to modify our HMM to a 3-state model of the state values $q = S, T, E$ to distinguish between texture (T) and edge (E) singularities. Texture and edges differ in coherency of phase. While isolated edges have coherent phase across scale as the edge location does not change across scale, in case of texture, the phase effects interfere with one another resulting in incoherence across scale. Phase can be modeled as an independent uniform (between 0 and 2π) distribution.

Another improvement of the HMM performance can lie in clustering of the coefficients within each scale [Crouse et al., 1998]. Then we speak of *hidden Markov fields* which demand more complicated training algorithms than the EM or Viterbi algorithm.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we describe the dual-tree complex wavelet transform (DTCWT) and its advantages over the critically decimated discrete wavelet transform (DWT) which are approximate shift invariance, steady values of the magnitude across scale, phase representation of edges orientation and improved directional selectivity in higher dimensions.

We exploited these both transforms for noise reduction in biomedical images by soft wavelet shrinkage. The resulting images are then used for edge detection by gradient approximating masks and the canny method. We also suggest the possible use od the DTCWT in edge detection via hidden Markov modeling as proposed in literature.

In near future, we would like to proceed with studying and improving various edge and line detection methods particularly those using the DTCWT-base hidden Markov models. They may also be utilized in denoising.

6 ACKNOWLEDGEMENTS

Many thanks to Oldrich Vysata, MD from Neurocenter Caregroup in Rychnov nad Kneznou, Czech Republic for providing us with the biomedical images, and to Prof. N. Kingsbury from the Engineering Department of the University of Cambridge, U.K. for his dual-tree complex wavelet transform toolbox. The paper has been supported by the Research grant No. MSM 6046137306.

References

CHOI, H.; BARANIUK, R. G. 2001. Multiscale image segmentation using wavelet-domain hidden markov models. *IEEE Transactions on Image Processing*, 10, 9, 1309 – 1321.

CROUSE, M. S.; NOWAK, R. D.; BARANIUK, R. G. 1998. Wavelet-based statistical signal processing using hidden markov models. *IEEE Transactions on Signal Processing*, 46, 4, 886 – 902.

HLAVÁČ, V.; SEDLÁČEK, M. 2007. *Zpracování signálů a obrazů*. Praha: Nakladateltví ČVUT.

KINGSBURY, N. G. 2000. A dual-tree complex wavelet transform with improved orthogonality and symmetry properties. In *Proceedings of the IEEE International Conf. on Image Processing, Vancouver*, 375–378. IEEE.

N. G. KINGSBURY 2003. Design of Q-shift Complex Wavelets for Image Processing Using Frequency Domain Energy Minimisation. In *Proceedings of the IEEE Conf. on Image Processing, Barcelona*, NK1–4. IEEE.

PETROU, M.; BOSDOGIANNI, P. 2000. *Image Processing*. John Wiley & Sons.

RANGAYYAN, R. M. 2005. *Biomedical Image Analysis*. Biomedical Engineering. U.S.A.: CRC Press.

REEVES, T. H.; KINGSBURY, N. G. 2000. Prediction of coefficients from coarse to fine scales in the complex wavelet transform. In *Proceedings of the IEEE International Conf. on Acoustics, Speech, and Signal Processing, Istanbul*, 508 – 511. IEEE.

ROMBERG, J. K.; CHOI, H.; BARANIUK, R. G. 2001. Multiscale edge grammars for complex wavelet transforms. In *Proceedings of the IEEE International Conf. on Image Processing*, 614 – 617. IEEE.

SELESNICK, I. W.; BARANIUK, R. G.; KINGSBURY, N. G. 2005. The dual tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22, 6, 123–151.

SHAFFREY, C. W. 2003. *Multiscale Techniques for Image Segmentation, Classification and Retrieval*. Ph.D. thesis, University of Cambridge.

SHAFFREY, C. W.; KINGSBURY, N. G.; JERMYN, I. H. 2002. Unsupervised image segmentation via markov trees and complex wavelets. In *Proceedings of the IEEE International Conf. on Image Processing, Rochester, USA*, 801 – 804. IEEE.

SHUKLA, P. D. 2003. *Complex Wavelet Transforms and Their Applications*. Ph.D. thesis, The University of Strathclyde in Glasgow.