Complex Wavelet Transform Variants and Scale Invariance in Magnification-Endoscopy Image Classification

Michael Häfner, Andreas Uhl, Andreas Vécsei, Georg Wimmer, and Friedrich Wrba

Abstract— **In this paper, scale invariant features are extracted from complex wavelet transform variants in order to classify high-magnification colon endoscopy imagery with respect to the pit pattern scheme. Superior results as compared to techniques described previously in literature are reported.**

I. Introduction

In colonoscopic (and other types of endoscopic) imagery, mucosa texture usually is found at different scales. This is due to varying distance and perspective towards the colon wall and eventually different zoom factors applied in the endoscope. Consequently, in order to design reliable computer-aided mucosa texture classification schemes, the scale invariance of the employed feature sets is essential.

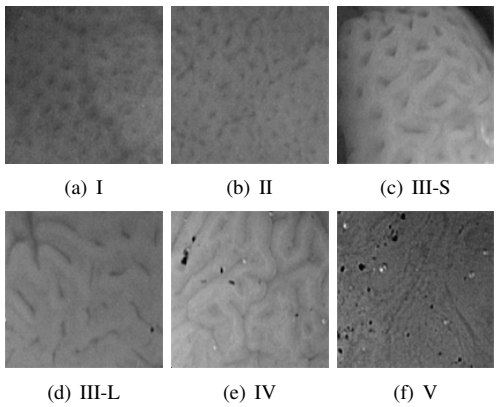
We consider feature vectors extracted from subbands of the Dual-Tree Complex Wavelet Transform (DT-CWT) and the Double Dyadic Dual-Tree Complex Wavelet Transform (D^3T-CWT), since wavelet transforms in general excel by their respective multiscale properties. The classical way of computing scale invariant features from these transform domains [4], [7] is to apply the Discrete Fourier Transform (DFT) to statistical parameters of the subband coefficients' distributions (e.g. mean and standard deviation).

In this work, we introduce the Quatro Dyadic Dual-Tree Complex Wavelet Transform (D^4T-CWT) and propose to apply the real-valued Discrete Cosine Transform (DCT) to coefficient parameters instead of the complex-valued DFT. In addition to classical coefficient distribution parameters, we employ shape and scale parameters of the Weibull distribution to construct scale invariant features by applying DCT and DFT, respectively.

This paper is organized as follows. In section II, we briefly introduce the pit-pattern classification scheme. Section III discusses the basics of the DT-CWT, the D^3T-CWT and the D^4T-CWT. Subsequently we describe the feature extraction with focus on achieving scale invariance with the DCT or DFT. In section IV we describe the experiments and present the results. Section V presents the conclusion of our work and an outlook on further research.

II. Pit-Pattern Classification

Polyps of the colon are a frequent finding and are usually divided into metaplastic, adenomatous and malignant. Since the resection of all polyps is rather time-consuming, it is imperative that those polyps which warrant resection can be distinguished. Furthermore, polypectomy of metaplastic lesions is unnecessary and removal of invasive cancer may be hazardous. To obtain images of the polyp's surface which are as detailed as possible a magnifying colonoscope can be used. This type of colonoscope provides images which are up to 150-fold magnified and thus are very detailed as they uncover the fine surface structure of the mucosa as well as small lesions. In the research work of Kudo [3], the macroscopic appearence of colorectal polyps is systematically described and results in the so called pit-pattern classification scheme, which divides the mucosal crypt patterns into five types (pit-patterns I-V, see Figure 1). It has been shown that pit patterns can be observed well using a high-magnification endoscope and thus this combination can be used to create a computer-assisted diagnosis system [2]. While types I and II are characteristic of benign lesions and represent normal colon mucosa or hyperplastic polyps (non-neoplastic lesions), types III to V represent neoplastic, adenomatous and carcinomatous structures. Our classification problem can be stated as follows: the problem of differentiating pit-patterns I and II from III-L, III-S, IV and V will be denoted as the two-class problem (non-neoplastic vs. neoplastic), whereas the more complex and detailed discrimination of all pit-patterns I to V will be denoted as the six-class problem. Note, that pit-type III is subdevided into types III-S/III-L and thus accounts for two classes.



(a) I (b) II (c) III-S

(d) III-L (e) IV (f) V

Fig. 1. Example images for the respective classes taken from the image database used.

III. CWT Variants and Scale Invariant Features

Kingbury's Dual-Tree Complex Wavelet Transform [1] has already been sucessfully applied in the context of pit

This work is partially supported by the Austrian National Bank "Jubiläumsfonds" Project No. 12514 .

M. Häfner is with the St. Elisabth Hospital, Vienna, Austria

A. Uhl and G. Wimmer are with the University of Salzburg, Salzburg, Austria uhl@cosy.sbg.ac.at

A. Vécsei is with the St. Anna Children's Hospital, Vienna, Austria

F. Wrba is with the Vienna Medical University, Austria

pattern classification [2]. The DT-CWT devides an image into six directional ($15°$, $45°$, $75°$, $105°$, $135°$, $165°$) oriented subbands per level of decomposition. The DT-CWT analyses an image only at dyadic scales. The D^{3}T-CWT [5] overcomes this issue, by introducing additional levels between dyadic scales. These additional levels between dyadic scales are generated by applying the DT-CWT to a scaled-down version of the original image using a factor of $2^{-0.5}$. We use the bicubic interpolation to scale down the image. Instead of the levels 1, 2,..., L in the DT-CWT we get the levels 1, 1.5, 2,..., L+0.5 in the D^{3}T-CWT, where the integer levels correspond to the levels of the DT-CWT (Fig. 2).

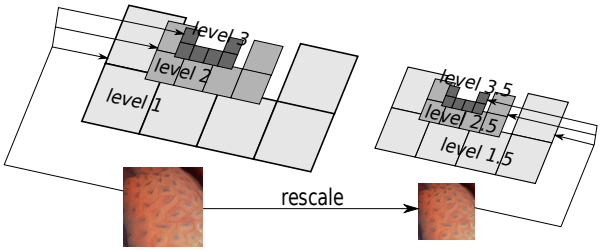


Fig. 2. The process of decomposing an image using the D^{3}T-CWT

The D^{4}T-CWT works similar to the D^{3}T-CWT, with the difference that the D^{4}T-CWT has even more additional levels between the scales. These scales are generated by applying the DT-CWT to the scaled-down versions of the original image using the factors $\sqrt{3/8}$, $\sqrt{1/2}$, and $\sqrt{3/4}$. The advantages of these three complex wavelet transforms are their approximately shift-invariance, their directional selectivity and the very efficient implementation scheme by four (eight) [sixteen] parallel 2-D DWT at the DT-CWT (D^{3}T-CWT) [D^{4}T-CWT]. All of these properties come at the very low cost of four times (six times $(4+4\times(\sqrt{1/2})^2)$) [10.5 times $(4+4\times(3/8+1/2+3/4))$] redundancy in 2-D in the DT-CWT (D^{3}T-CWT) [D^{4}T-CWT].

In this paper, we use two ways to generate the feature set from the DT-CWTs. The first and most common approach is to compute the empirical mean and the empirical standard deviation of the absolute values of each subband and concatenate them to one feature vector (classic distribution):

$$\mu_{l,d} = \frac{1}{N_{l,d}} \sum_{i=1}^{N_{l,d}} |x_{l,d}(i)| \tag{1}$$

$$\sigma_{l,d} = \frac{1}{N_{l,d}} \Big(\sum_{i=1}^{N_{l,d}} (|x_{l,d}(i)| - \mu_{l,d})^2 \Big)^{\frac{1}{2}}, \tag{2}$$

where $|x_{l,d}(i)|$ denotes the absulute value of the coefficients in subband (l,d) and $N_{l,d}$ denotes the total number of coefficients in subband (l,d) (at decomposition level l and direction $d \in \{1,\dots,6\}$).
The second approach is to model the absolute values of each subband by a two-parameter Weibull distribution [2]. The probability density function of a Weibull distribution with shape parameter c and scale parameter b is given by

$$p(x;c,b) = \frac{c}{b}\Big(\frac{x}{b}\Big)^{c-1} e^{-(\frac{x}{b})^c}. \tag{3}$$

The moment estimates (c,b) of the Weibull parameters of each subband are than arranged into feature vectors like in the approach before. The feature extraction for the D^{3,4}T-CWTs works the same way, but the feature vector is longer because of the non-dyadic scales. For example, the feature vector of a RGB-image using the DT-CWT with five decomposition levels has length $5\times 6\times 2\times 3 = 180$ (5 levels (scales), 6 directions, 2 (mean and standard deviation), 3 color channels). In case of the D^{3}T-CWT the feature vectors length is 360.

A common approach to achieve scale-invariance for wavelet-based features is to use the absolute values of a Discrete Fourier Transformation (DFT) applied to extracted statistical moments. We use the method from [4], [7] and apply the DFT to the feature vector (of the DT-CWT) as follows

$$U_{n,d} = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} \mu_{l,d}\, e^{\frac{-i\,2\pi(l-1)(n-1)}{L}} \tag{4}$$

for $n \in \{1,\dots,L\}$ and $d \in \{1,\dots,6\}$, and

$$S_{n,d} = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} \sigma_{l,d}\, e^{\frac{-i\,2\pi(l-1)(n-1)}{L}}. \tag{5}$$

For the D^{3}T-CWT, we replace L with $2L$ and $n \in \{1, 1.5, 2, \dots, L+0.5\}$ and for the D^{4}T-CWT we replace L by $4L$ and $n \in \{1, 1.25, 1.5, 1.75, 2, \dots, L+0.75\}$. The new feature vector (for the DT-CWT) is

$$\begin{aligned} f \;=\; & \{|U_{1,1}|,\dots|U_{L,1}|,|U_{1,2}|,\dots|U_{L,2}|,\dots|U_{L,6}|, \\ & |S_{1,1}|,\dots|S_{L,1}|,|S_{1,2}|,\dots|S_{L,2}|,\dots|S_{L,6}|\}. \end{aligned} \tag{6}$$

The feature vectors for the D^{3}T-CWT and D^{4}T-CWT are created by analogy.
Contrasting to proposals in literature [4] it turned out that the absolute values of the $U's$ and $S's$ degrade the pit-pattern classsification results, whereas the real values ot the $U's$ and $S's$ enhance the results. Because of

$$\Re\big(e^{\frac{-i\,2\pi(l-1)(n-1)}{L}}\big) = \cos\big(\frac{-2\pi(l-1)(n-1)}{L}\big), \tag{7}$$

the real values of the DFT are obtained by a cosinus transform. Hence we propose to use the Discrete Cosinus Transform (DCT). The DCT of one of our feature vector is computed by

$$U(n,d) = w(n) \sum_{l=1}^{L} \mu_{l,d}\, \cos\big(\frac{\pi(2(l-1)(n-1)}{L}\big) \tag{8}$$

and

$$S(n,d) = w(n) \sum_{l=1}^{L} \sigma_{l,d}\, \cos\big(\frac{\pi(2\,l-1)(n-1)}{L}\big) \tag{9}$$

for $n \in \{1,\dots,L\}$ and $d \in \{1,\dots,6\}$, where

$$w(n) = \begin{cases} \frac{1}{\sqrt{L}} & \text{for } n = 1, \\ \frac{2}{\sqrt{L}} & \text{for } 2 \le n \le L. \end{cases}$$

Pit Type	I	II	III-L	III-S	IV	V
2 class	178		449			
6 class	114	64	18	119	232	80

TABLE I
NUMBER OF IMAGE SAMPLES PER PIT-PATTERN CLASS (GROUND TRUTH)

For the Weibull parameter case, $\mu_{l,d}$ and $\sigma_{l,d}$ are simply replaced by $c_{l,d}$ and $b_{l,d}$.

Applying the DCT or DFT for the D^{3}T-CWT works similar, but it turns out, that the transformation leads to better results if we apply the DCT or DFT on $(\mu_{1,d}, \mu_{2,d}, \ldots, \mu_{L,d})$ and $(\mu_{1.5,d}, \mu_{2.5,d}, \ldots, \mu_{L+0.5,d})$ seperately, instead of $DCT(\mu_{1,d}, \mu_{1.5,d} \ldots, , \mu_{L+0.5,d})$. The DCT or DFT for the D^{4}T-CWT is done in a similar fashion by applying them four times seperatly.
Further we have to note, that in case of the DFT, parts of the feature vector will be deleted after the DFT, because the complex conjugates are redundant in the feature vector.

IV. EXPERIMENTAL STUDY

A. Experimental Settings

We employ a simple 1-Nearest Neighor (dennoted by 1-NN) classifier, which uses the Euclidean distance to measure the distance between two d-dimensional feature vectors u and $v \in \mathbb{R}^d$ in order to give more emphasis to the quality of the extracted features than to the classifier. Classification accuracy is defined as the number of correctly classified samples devided by the total number of samples. We employ leave-one-out crossvalidation (LOOCV) to reliably estimate classification accuracy.
Euclidian distance is very sensitive to large differences in the numerical range of single features. This is especially important in case of the Weibull parameters, since the range of scale and shape differs significantly. The solution of this problem is to normalize the features of each feature vector. Given our d-dimensional training samples $v_1, \ldots, v_n$, the normalization formula for the m-th element of the j-th feature vector is defined by

$$\tilde{v}_{jm} = \frac{v_{jm} - \bar{v}_m}{s_m}, \tag{10}$$

where $\bar{v}_m, s_m$ denote the sample mean and the sample variance of the m-th feature. In this way we obtain re-scaled features with zero-mean and unit standard deviation. Now each feature contributes equally to the calculation of the metric. If we apply the DCT or DFT, the range of the feature-values increases and we re-apply the normalization a second time.
Our image database consists of a total of 627 images, acquired between the years 2005 and 2008 at the Department of Gastroenterology and Hepatology (Medical University of Vienna) using a zoom-endoscope (Olympus Evis Excera CF-Q160ZI/L) with a magnification factor of 150. In order to condense information of the original endoscopic images, we cut out regions of interest of size 312 x 312 and scaled them to a size of 256 x 256. Table I lists the number of image samples per class.

Before decomposing the images with the CWTs, we first employ adaptive histogram equalization using the CLAHE (contrast-limited adaptive histogram equalization) [6] algorithm with 8×8 tiles and a uniform distribution for constructing the contrast transfer function. Second, we blur the image by a Gaussian 3×3 mask with $\sigma = 0.5$.

B. Experimental Results

Tests were made with 4, 5 and 6 levels of decomposition and with grayscale- and RGB-images. In most cases the best results were obtained with 6 levels of decomposition of RGB-images and so only these results are presented. The best results in the tables are given in bold face numbers. Table II shows a comparison between the DT-CWT variants, results with DCT being applied are given in parantheses().

Features	2-class		6-class	
	Classic	Weibull	Classic	Weibull
DT-CWT	98.3 (98.4)	98.4 (98.6)	92.3 (**95.9**)	91.2 (90.3)
D^{3}T-CWT	98.3 (**99.2**)	98.4 (98.4)	92.2 (95.7)	92.3 (93)
D^{4}T-CWT	98.4 (**99.2**)	**98.6** (98.7)	92.3 (94.4)	**92.5** (93.5)

TABLE II
CLASSIFICATION ACCURACY IN % WITHOUT (AND WITH) DCT.

There are no big differences among DT-CWT, the D^{3}T-CWT and the D^{4}T-CWT, if we do not employ the DCT. We can also see, that the DCT distinctly enhances the results of the classic distribution (especially in the 6-class case), whereas the results of the Weibull parameters nearly remain the same. The D^{3}T-CWT and the D^{4}T-CWT are slightly superior to the DT-CWT, especially in the Weibull case.
In Table III we display the results when using the DFT instead of the DCT. The results for applying the absolute values of the DFT are put in parentheses. We see that the results of the real DFT values are sometimes even better than those of the DCT ("Classic" case, see Table II) in contrast to the absolute values, which are really bad.

Features	2-class		6-class	
	Classic	Weibull	Classic	Weibull
DT-CWT	98.9 (95.5)	98.3 (89.2)	95.7 (86)	89.8 (71.1)
D^{3}T-CWT	**99.4** (95.1)	98.3 (90.3)	**96.2** (87.4)	91.9 (76.6)
DT-CWT	99 (94.9)	98,4 (91.4)	95.5 (85.5)	92 (79.4)

TABLE III
CLASSIFICATION ACCURACY IN % FOR REAL (AND ABSOLUTE) DFT VALUES.

It seems that the intended improvements with respect to scale invariance do not enhance the classification results significantly, which may be partly due to the nature of the magnification-endoscopy image set (fixed zoom factor, small distance to colon wall and corresponding small amount of perspective changes). In order to give more emphasis to the scale invariance aspect, we crop quadratic images from the middle of the original images (sizes 180, 208, and 232) and resize them to the size of the original image (256) per bicubic interpolation. Now we fuse the original images and

the resized ones of a single size into a new artificial dataset. We classify these 1254 images by the 1-NN classifier, but as nearest neighbors for the original images only resized images are accepted and for the resized images only the original images are accepted as nearest neighbors. Additionally, for any of these 1254 images, the according resized or original image is not accepted as nearest neighbor, although we count how often this would happen (the values given in brackets in Tables V and VI).

Table IV shows the results of the two class case (results with DCT are again given in parentheses, the size, e.g. 208, represents the size of the cropped images before resizing). In case of mean and standard deviation ("classic"), applying the DCT improves the results slightly in all cases. For the Weibull parameters, this is also the case except for the smallest version of the resized images (180). With respect to the DT-CWT variants, at least all the best results (given in bold face) are improving moderately the more in-between scales are considered.

size		180	208	232
DT-CWT	Classic	76.9 (78.9)	86.7 (94.8)	96.7 (**98.2**)
	Weibull	**89.9** (88.2)	96.5 (**97**)	97.2 (98)
D^3T-CWT	Classic	78.5 (80.3)	93.1 (97.7)	97.5 (97.7)
	Weibull	**92.7** (91.5)	97.5 (**97.9**)	97.5 (**98.2**)
D^4T-CWT	Classic	79.7 (79.8)	94.4 (98)	97.4 (97.8)
	Weibull	**93.1** (92.5)	**98.4** (**98.4**)	98.1 (**98.8**)

TABLE IV
OVERALL CLASSIFICATION ACCURACY IN % FOR THE 2-CLASS CASE FOR DIFFERENTLY SCALED DATASETS.

In Table V we display the results of the 6-class case. We note that the DCT again enhances scale invarince in case of the classic parameters more distinctively as in case of the Weibull parameters, where in the case of the smallest images the results without DCT are superior. When considering the different DT-CWT variants we also see, that the D^3T-CWT and D^4T-CWT provide more scale invariance since especially for the two smaller images sizes, the best results improve the more in-between scales are used. It should also be noted that the Weibull parameters provide clearly better scale invariance compared to the classic parameters for these image sizes.

size		180	208	232
DT-CWT	Classic	41.8 (0)	64 (18)	88.1 (260)
	Classic (DCT)	50.6 (2)	83.9 (57)	**92.5** (482)
	Weibull	**65.2** (0)	84.1 (148)	88.1(686)
	Weibull (DCT)	63.4 (0)	**87.4** (154)	89.4 (749)
D^3T-CWT	Classic	46.2 (1)	77 (63)	90.1 (499)
	Classic (DCT)	52.4 (2)	88.4 (110)	**92.4** (718)
	Weibull	**74.5** (6)	88.4 (217)	90 (904)
	Weibull (DCT)	70.8(0)	**90** (207)	91.8 (911)
D^4T-CWT	Classic	48.7 (0)	80.2 (99)	90.3 (632)
	Classic (DCT)	52.3 (2)	88.6 (136)	92.2 (832)
	Weibull	**76.2** (11)	89.6 (293)	91.5 (1000)
	Weibull (DCT)	71.4 (14)	**91.4** (254)	**92.2** (1009)

TABLE V
CLASSIFICATION ACCURACY IN % FOR THE 6-CLASS CASE FOR DIFFERENTLY SCALED DATASETS

In Table VI we present the results when applying the two DFT variants to the fused datasets. We can see that the DFT degrades scale invariance as compared to the DCT case significantly for both feature set variants. Especially when using the absosolute value, results are very poor.
For these results, the D^4T-CWT is more scale invariant than the D^3T-CWT and both are clearly more scale invariant than the DT-CWT.

size		180	208	232
DT-CWT	abs	30.8 (0)	35.6 (20)	55.3 (197)
	real	29.2 (3)	44.9 (8)	70.5 (113)
D^3T-CWT	abs	32.5 (2)	48.4 (45)	75 (399)
	real	39.9 (0)	60.6 (21)	88.6 (261)
D^4T-CWT	abs	35 (3)	53 (76)	77.4 (563)
	real	42 (0)	66.4 (28)	89.3 (398)

TABLE VI
CLASSIFICATION ACCURACY IN % FOR THE 6-CLASS CASE USING DFT ON THE CLASSIC PARAMETERS.

V. CONCLUSIONS

We have shown that applying the DCT instead of the DFT enhances scale invariance properties of the resulting feature sets in many cases. Using Weibull distribution parameters have turned out to be highly beneficial when scale invariance properties are important, however, the potential of improvement by applying a DCT to feature vectores of this type is much smaller compared to the classical parameters mean and standard deviation. We have found improvements with respect to scale invariance for the D^4T-CWT, but given the significantly higher computational complexity and feature set size, it is questionable if the small improvements justify its employment.

In future work, we will apply the proposed feature sets to endoscopic imagery where scale invariance properties are more important per se.

REFERENCES

[1] N. Kingsbury, "The Dual-Tree Complex Wavelet Transform: a new technique for shift-invariance and directional filters," in *Proceedings of the 8th IEEE DSP Workshop*, Bryce Canyon, Utah, USA, Aug. 1998, pp. 9-12.

[2] R. Kwitt and A. Uhl, "Modelling the marginal Distribution of complex Wavelet coefficient magnitudes for the classification of Zoom-Endoscopy images," in *Proceedings of the IEEE Computer Society Workshop an Mathematical Methods in Biomedical Image Analysis (MMBIA'07)*, Rio de Janeiro, Brazil, 2007, pp. 1-8.

[3] S.Kudo, S. Hirota, T. Nakajima, S. Hosobe H. Kusaka, T. Kobayashi, M. Himori, and A. YagYuu. Diagnosis of colorectal tumorous lesions by magnifying endoscopy. *Gastrointestinal Endoscopy*, 44(1):8-14, July, 1996, 47:880-885, 1994.

[4] E. H. S. Lo, M. R. Pickering, M. R. Frater and J. F. Arnold, Scale and Rotation Invariant Texture Features from the Dual-Tree Complex Wavelet Transform. In *Proc. Int'l Conf. Image Processing,* vol. 1, pp. 227-230 Singapore, Oct. 2004, IEEE.

[5] E. H. S. Lo, M. R. Pickering, M. R. Frater and J. F. Arnold, Query by Example using Invariant Features from the Double Dyadic Dual-Tree Complex Wavelet Transform. In *Proc. Int'l Conf. Image & Video Retrieval,* Santorini, Greece, Jul. 2009, ACM,

[6] K.Zuiderveld. Contrast limited adaptive histogram equalization. In P. S. Heckbert, editor, *Graphics Gems IV,* pages 474-485. Morgan Kaufmann, 1994.

[7] R. Manthalkar, P. K. Biswas and B. N. Chatterji, Rotation and scale invariant texture features using discrete wavelet packet transform, Pattern Recognition Letters, 24 (14):2455-2462, Oct. 2003.