

© Springer Verlag. The copyright for this contribution is held by Springer Verlag. The original publication is available at [www.springerlink.com](http://www.springerlink.com).

# Complex Wavelet Transform Variants in a Scale Invariant Classification of Celiac Disease

Andreas Uhl<sup>1</sup>, Andreas Vécsei<sup>2</sup>, Georg Wimmer<sup>1</sup>

<sup>1</sup>Department of Computer Sciences, University of Salzburg, Salzburg, Austria

<sup>2</sup>St. Anna Children's Hospital, Vienna, Austria

**Abstract** In this paper, we present variants of the Dual-Tree Complex Wavelet Transform (DT-CWT) in order to automatically classify endoscopic images with respect to the Marsh classification. The feature vectors either consist of the means and standard deviations of the subbands from a DT-CWT variant or of the Weibull parameter of these subbands. To reduce the effects of different distances and perspectives toward the mucosa, we enhanced the scale invariance by applying the discrete Fourier transform or the discrete cosine transform across the scale dimension of the feature vector.

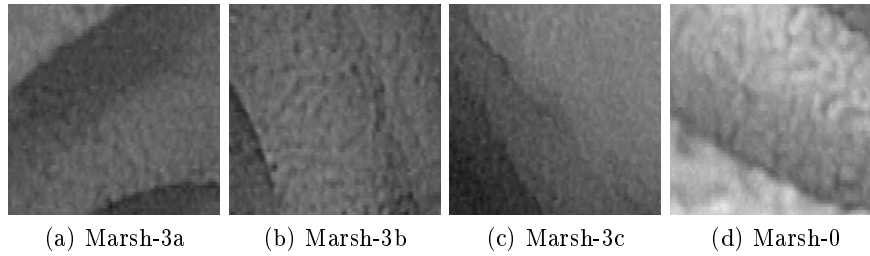
Key words: Endoscopic Imagery, Celiac Disease, Automated Classification, Scale Invariance, Dual-Tree Complex Wavelet Transform, Discrete Fourier Transform, Discrete Cosine Transform

## 1 Introduction

The celiac state of the duodenum is usually determined by visual inspection during the endoscopy session followed by a biopsy of suspicious areas. The severity of the mucosal state of the extracted tissue is defined according to a modified Marsh scheme, which divides the images in four different classes, Marsh-0 Marsh-3a, Marsh-3b and Marsh-3c (see Figure 1). Marsh-0 represents a healthy duodenum with normal crypts and villi, Marsh-3a, Marsh-3b and Marsh-3c have increased crypts and mild atrophy (3a), marked atrophy (3b) or the villi are entirely absent (3c), respectively. Types Marsh-3a to Marsh-3c span the range of characteristic changes caused by celiac disease, whereat Marsh-3a is the mildest and Marsh-3c the most severe form. We distinguish between two regions of the duodenum, the *bulbus duodeni* and the *pars descendens*.

In gastroscopic (and other types of endoscopic) imagery, mucosa texture is usually found with different perspective, zoom (see Figure 1) and distortions (barrel-type distortions of the endoscope [1]). That means that the mucosal textures shows different spatial scales, depending on the camera perspective and distance to the mucosal wall.

Consequently, in order to design reliable computer-aided mucosa texture classification schemes, the scale invariance of the employed feature sets could be



**Figure 1.** Example images for the respective classes taken from the pars descendes database

essential.

We consider feature vectors extracted from subbands of the Dual-Tree Complex Wavelet Transform (DT-CWT), the Double Dyadic Dual-Tree Complex Wavelet Transform ( $D^3T$ -CWT) and the Quatro Dyadic Dual-Tree Complex Wavelet Transform ( $D^4T$ -CWT) [2], since wavelet transforms in general excel by their respective multiscale properties. A classical way of computing scale invariant features from multi-scale methods like e.g. the DT-CWT [4,5] is to apply the discrete Fourier transform (DFT) to statistical parameters of the subband coefficients' distributions (e.g. mean and standard deviation) and compute the magnitudes of these complex values. In this work, we also use the real part of the DFT or apply the real-valued discrete cosine transform (DCT) to coefficient parameters, which enhanced the results and the scale invariance in magnification-endoscopy image classification [2]. In addition to classical coefficient distribution parameters, we employ shape and scale parameters of the Weibull distribution [3] to model the absolute values of each subband.

This paper is organized as follows. In section 2 we discuss 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 DFT or DCT. In section 3 we describe the experiments and present the results. Section 4 presents the discussion of our work.

## 2 CWT VARIANTS AND SCALE INVARIANT FEATURES

Kingbury's Dual-Tree Complex Wavelet Transform [6] divides an image into six directional ( $15^\circ$ ,  $45^\circ$ ,  $75^\circ$ ,  $105^\circ$ ,  $135^\circ$ ,  $165^\circ$ ) oriented subbands per level of decomposition. The DT-CWT analyzes an image only at dyadic scales. The  $D^3T$ -CWT [4] 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 downscaled 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  $1, 2, \dots, L$  in the DT-CWT we get the levels  $1, 1.5, 2, \dots, L+0.5$  in the

D<sup>3</sup>T-CWT, where the integer levels correspond to the levels of the DT-CWT. The D<sup>4</sup>T-CWT works similar to the D<sup>3</sup>T-CWT, with the difference that the D<sup>4</sup>T-CWT has even more additional levels between the scales. These scales are generated by applying the DT-CWT to the downscaled 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. 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 ( $\mu_{l,d}$ ) and the empirical standard deviation ( $\sigma_{l,d}$ ) of the absolute values of each subband (decomposition level  $l \in \{1, \dots, L\}$  and direction  $d \in \{1, \dots, 6\}$ ) and concatenate them to one feature vector later denoted as Classic distribution).

The second approach is to model the absolute values of each subband by a two-parameter Weibull distribution [3]. The probability density function of a Weibull distribution with shape parameter  $c$  and scale parameter  $b$  is given by

$$p(x; c, b) = \frac{c}{b} \left(\frac{x}{b}\right)^{c-1} e^{-\left(\frac{x}{b}\right)^c}. \quad (1)$$

The moment estimates ( $c, b$ ) of the Weibull parameters of each subband are then arranged into feature vectors like in the approach before. The feature extraction for the D<sup>3,4</sup>T-CWTs works the same way, but the feature vector is longer because of the non-dyadic scales.

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 [4,5] and apply the DFT to the feature vector (of the DT-CWT) as follows

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

for  $n \in \{1, \dots, L\}$  and  $d \in \{1, \dots, 6\}$ . The feature curve of a feature vector shifts if input texture is scaled. DFT magnitude makes the feature values independent of cyclic shifts of the feature curve. The DFT assumes that there is a periodic input signal; however there is no reason why the statistical features should be periodic. If the statistical features are close to zero at both ends, the approach provides good scale invariance.

For the D<sup>3</sup>T-CWT, we replace  $L$  with  $2L$  and  $n \in \{1, 1.5, 2, \dots, L+0.5\}$  and for the D<sup>4</sup>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

$$f = \{|U_{1,1}|, \dots, |U_{L,1}|, |U_{1,2}|, \dots, |U_{L,2}|, \dots, |U_{L,6}|, |S_{1,1}|, \dots, |S_{L,1}|, \dots, |S_{L,6}|\}.$$

The feature vectors for the D<sup>3</sup>T-CWT and D<sup>4</sup>T-CWT are created by analogy.

It turned out that the results of the real values of the  $U$ 's and  $S$ 's provide better results than the absolute values [2]. Because of

$$\Re \left( e^{\frac{-i 2\pi(l-1)(n-1)}{L}} \right) = \cos \left( \frac{-2\pi(l-1)(n-1)}{L} \right), \quad (3)$$

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

$$U(n, d) = w(n) \sum_{l=1}^L \mu_{l,d} \cos\left(\frac{\pi(2(l-1)(n-1))}{L}\right) \quad (4)$$

for  $n \in \{1, \dots, L\}$  and  $d \in \{1, \dots, 6\}$  (and similar for  $S(n, d)$ ), where  $w(n) = 1/\sqrt{L}$  for  $n = 1$  and  $w(n) = 2/\sqrt{L}$  for  $2 \leq n \leq L$ . 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<sup>3</sup>T-CWT works similar, but it turns out, that the transformation leads to better results if we apply the DCT or DFT on  $(\mu_{1,d}, \mu_{2,d}, \dots, \mu_{L,d})$  and  $(\mu_{1.5,d}, \mu_{2.5,d}, \dots, \mu_{L+0.5,d})$  separately, instead of  $DCT(\mu_{1,d}, \mu_{1.5,d}, \dots, \mu_{L+0.5,d})$ . The DCT or DFT for the D<sup>4</sup>T-CWT is done in a similar fashion by applying them four times separately.

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. If we use RGB-images, than we simply concatenate the feature vectors of each color channel.

### 3 EXPERIMENTAL STUDY

We employ two methods to evaluate and compare the feature sets described in the section before: The area under the ROC curve (AUC) [7] and the overall classification accuracy.

To generate the ROC curve for k-NN classifier we used the method described in [7] (for k=20). We consider the 20 nearest neighbors of each image (the 20 feature vectors with the lowest euclidean distance to the feature vector of the considered image). We employ leave-one-out cross-validation (LOOCV) to find these 20 nearest neighbors for each image. We achieve the first point on the ROC curve by classifying the images as positive, if one or more than one nearest neighbor of a considered image is positive. Because for nearly every image, there is at least one of the 20 neighbors positive (positive means that the image belongs to class Marsh-3a, Marsh-3b or Marsh-3c), the true positive rate (TPR) (= sensitivity) and the false positive rate (FPR) (= 1 - specificity) will be 100 % or near to 100 %. The second point on the ROC curve is achieved by classifying an image as positive, if two or more of the 20 nearest neighbors are positive, the third point if three or more of the 20 nearest neighbors are positive and so on till 20. The more positive nearest neighbors of the 20 nearest neighbors are needed to classify an image as positive, the lower are the TPR and FPR. The last point is achieved by classifying an image as positive, if all the 20 nearest neighbors are positive. Because there is hardly ever one of the 20 nearest neighbors of an image negative, the TPR and the FPR are in this case 0 % or near to 0 %. That is the way to generate the points on the ROC curve. To be sure that the curve reaches from the the point (0,0) (TPR=0, FPR=0) to the point (1,1) (TPR=1,FPR=1),

we add these points to the curve. The point (0,0) can be interpreted as 21 of the the 20 nearest neighbors of an image have to be positive to classify the image as positive. This is not possible and so the TPR and FPR are 0 %. The point (1,1) can be interpreted as 0 or more of the 20 nearest neighbors of an image have to be positive to classify the image as positive. This will always happens and so the TPR and FPR will be 100 %. In Figure 2 we see two examples of ROC-curves. The AUC is computed by trapezoidal integration,

$$AUC = \sum_{i=1}^{21} ((TPR_i \cdot \Delta FPR_i) + 1/2(\Delta TPR_i \cdot \Delta FPR_i)) \quad (5)$$

where

$$\Delta TPR_i = TPR_{i-1} - TPR_i, \quad (6)$$

$$\Delta FPR_i = FPR_{i-1} - FPR_i, \quad (7)$$

and  $TPR_i$  or  $FPR_i$  are those TPR or FPR, where at least  $i$  positive nearest neighbors (out of the 20 nearest neighbors) are necessary to classify an image as positive.

The second method is the 20-Nearest Neighbor (denoted by 20-NN) classifier. We already know the 20 nearest neighbors from the AUC. An image is classified as positive, if more than the half (=10) of the nearest neighbors are positive, or as negative if more than the half of the nearest neighbors are negative. If there are 10 positive and 10 negative nearest neighbors, then the image is classified as its nearest neighbor (1-NN classifier). Classification accuracy is defined as the number of correctly classified samples divided by the total number of samples.

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

The image database consists of a total of 273 bulbus duodeni and 296 pars descendes images and was taken at the St. Anna's Children Hospital using a standard duodenoscope without magnification. In order to condense information of the original endoscopic images, we cut out regions of interest of size  $128 \times 128$  [8]. Table 1 lists the number of image samples per class. Tests were carried out with 6 levels of decomposition and RGB-images. We only consider the 2-class case.

The results for the AUC are given in Table 2. If we watch the results for the bulbus dataset, we can see that the results with DCT or without any further manipulation of the feature vector are similar. The results with the absolute values of the DFT are much worse and the results with the real values of the DFT are a little bit worse than the results with DCT or without any further

**Table 1.** Number of image samples per Marsh type (ground truth based on histology)

Data set	Bulbus				Pars			
Marsh type	0	a	b	c	0	a	b	c
Number of images (4-class case)	153	45	54	21	132	42	53	69
Number of images (2-class case)	153	120			132	164		

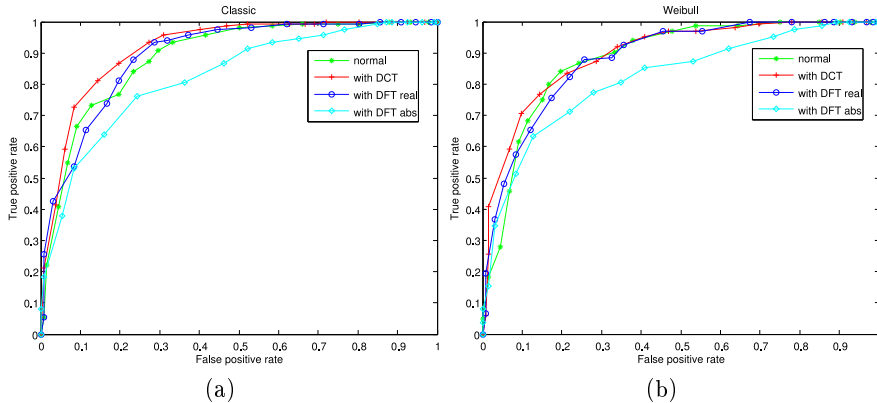
**Table 2.** Area under the ROC curve for the two data sets (bulbus and pars) with features extracted from DT-CWT variants by computing the Classic or the Weibull distribution and none or a further manipulation of the feature vectors by DFT variants or the DCT.

Feature		Classic				Weibull			
Manipulation		non	DFT abs	DFT real	DCT	non	DFT abs	DFT real	DCT
Bulbus	DT-CWT	0.97	0.82	0.96	0.98	0.97	0.78	0.95	0.97
	D <sup>3</sup> T-CWT	0.98	0.81	0.96	0.98	0.98	0.78	0.97	0.98
	D <sup>4</sup> T-CWT	0.98	0.83	0.96	0.98	0.98	0.80	0.96	<b>0.99</b>
Pars	DT-CWT	0.82	0.76	<b>0.88</b>	0.86	0.81	0.78	0.86	0.84
	D <sup>3</sup> T-CWT	0.82	0.76	0.87	0.87	0.82	0.78	0.85	0.84
	D <sup>4</sup> T-CWT	0.84	0.77	0.86	<b>0.88</b>	0.83	0.77	0.85	0.86

manipulation of the feature vector. In case of the pars descendens dataset, the results with DCT or the real part of the DFT are distinctly better than those without feature vector manipulation and the results with the absolute part of the DCT are much more worse compared to the other methods. The differences between the CWT-variants or the feature extraction methods (classic way or Weibull distribution) are very small. The best results for the both datasets in Table 2 are given in bold face numbers.

**Table 3.** Classification accuracy in % for the 20-NN classifier with the two data sets (bulbus and pars) and features extracted from DT-CWT variants by computing the Classic or the Weibull distribution and none or a further manipulation of the feature vectors by DFT variants or the DCT.

Feature		Classic				Weibull			
Manipulation		non	DFT abs	DFT real	DCT	non	DFT abs	DFT real	DCT
Bulbus	DT-CWT	94.9	74.7	91.6	94.5	93.8	72.9	91.2	92.3
	D <sup>3</sup> T-CWT	94.9	74.0	92.3	94.9	94.1	70.3	93.4	<b>95.2</b>
	D <sup>4</sup> T-CWT	94.5	79.1	92.3	94.9	94.5	74.7	91.9	<b>95.2</b>
Pars	DT-CWT	82.4	70.3	<b>84.1</b>	77.4	82.1	73.3	70.3	76.0
	D <sup>3</sup> T-CWT	82.1	68.2	82.1	78.0	81.4	70.3	80.7	74.3
	D <sup>4</sup> T-CWT	82.4	69.3	82.1	78.4	81.8	71.6	80.1	79.4



**Figure 2.** ROC curves of the different feature vector manipulation methods (normal, DCT, DFT abs, DFT real) for the pars descendens dataset, with features extracted from the  $D^4T$ -CWT by the classic way (a) or the Weibull distribution (b).

The results for the 20-NN classifier are given in Table 2. We can see that the results for the bulbus dataset are similar with DCT or without any further manipulation of the feature vector. The results with the real valued DFT are a little bit worse than the results mentioned before. In case of the pars descendens dataset, the results with DCT are worse than the results without any further feature vector manipulation and the results with the real part of the DFT are in case of the classic features similar and in case of the Weibull features worse than the results without feature vector manipulation. The results with the absolute part of the DFT are always worse than the other results. Once again, the results of the different CWT-variants are similar and the results of our feature extraction methods (classic and weibull) are also similar, apart from the case with the real valued DFT and the pars descendens dataset. The best results for each of the both datasets in Table 3 are given in bold face numbers.

## 4 Discussion

It is hard to interpret these results because the AUC and the overall classification results are often contradictory. For an example let us consider the results of the results of the DCT for the pars descendens dataset. The AUC is distinctly larger with DCT than without feature vector manipulation, whereas the classification accuracy for the 20-NN classifier is distinctly higher without feature vector manipulation than with DCT. If we watch the results, then it is impossible to say if we should favor a feature vector manipulation like the DCT or the real valued DFT or prefer no further feature vector manipulation. The advantages of a better balancing of different perspectives, zooms and distortions seems to be equal than the drawbacks like losing scale information by making the feature vector more scale invariant or by destroying information by the transformation. Maybe



the results of the AUC are more significant than the overall classification results, because an overall classification result uses only the information whether there are more positive or negative nearest neighbors for an image, whereas the AUC uses the information how much nearest neighbors are positive or negative.

There are small improvements of the CWT's with additional scales in between dyadic scales ( $D^3T$ -CWT,  $D^4T$ -CWT) compared to the standard DT-CWT, but because of their higher computational complexity it is questionable if the small improvements justify their application.

The classic way and the Weibull distribution are equally suited to extract the information from the subbands of the CWT's.

## References

1. Gschwandtner, M., Liedlgruber, M., Uhl, A., Vécsei, A.: Experimental Study on the Impact of Endoscope Distortion Correction on Computer-assisted Celiac Disease Diagnosis. In: Proceedings of the 10th International Conference on Information Technology and Applications in Biomedicine (ITAB'10), Corfu (2010)
2. Häfner, M., Uhl, A., Vécsei, A., Wimmer, G., Wrba, F.: Complex Wavelet Transform Variants and Scale Invariance in Magnification-Endoscopy Image Classification. In: Proceedings of the 10th International Conference on Information Technology and Applications in Biomedicine (ITAB '10), Corfu (2010)
3. Kwitt, R., Uhl, A.: Modeling the Marginal Distributions of Complex Wavelet Coefficient Magnitudes for the Classification of Zoom-Endoscopy Images. In: Proceedings of the IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '07), pp. 1–8, Rio de Janeiro, Brasil (2007)
4. Lo, E. H. S., Pickering, M. R., Frater, M.R., Arnold, J.F.: Image segmentation from scale and rotation invariant texture features from the double dyadic dual-tree complex wavelet transform. *Image and Vision Computing*, vol. 29(1), pp. 15–28, ElsevierScience INC., 2011
5. Manthalkar, R, Biswas, P.K., Chatterji, B.N.: Rotation and scale invariant texture features using discrete wavelet packet transform. *Pattern Recognition Letters*, vol. 24(14), pp. 2455–2462, Elsevier Science Inc., 2003
6. Kingsbury, N.G.: The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters. In: Proceedings of the IEEE Digital Signal Processing Workshop, DSP '98, pp. 9–12, Bryce Canyon, Utah, USA (1998)
7. Bradley, A.P.: The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, vol. 20, pp.1145–1159, 1997
8. Hegenbart, S., Kwitt, R., Liedlgruber, M., Uhl, A., Vécsei, A.: Impact of Duodenal Image Capturing Techniques and Duodenal Regions on the Performance of Automated Diagnosis of Celiac Disease. In: Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis (ISPA '09), pp. 718–723, Salzburg, Austria (2009)