Modeling the Marginal Distributions of Complex Wavelet Coefficient Magnitudes for the Classification of Zoom-Endoscopy Images

Roland Kwitt
Department of Computer Science
University of Salzburg
Salzburg, Austria
rkwitt@cosy.sbg.ac.at

Andreas Uhl
Department of Computer Science
University of Salzburg
Salzburg, Austria
uhl@cosy.sbg.ac.at

Abstract

In this paper, we propose a set of new image features for the classification of zoom-endoscopy images. The feature extraction step is based on fitting a two-parameter Weibull distribution to the wavelet coefficient magnitudes of subbands obtained from a complex wavelet transform variant. We show, that the shape and scale parameter possess more discriminative power than the classic mean and standard deviation based features for complex subband coefficient magnitudes. Furthermore, we discuss why the commonly used Rayleigh distribution model is suboptimal in our case.

1. Introduction

Recent statistics of the American Cancer Society reveal that colorectal cancer is the third most common cancer in men and women and the second most common cause of US cancer deaths. Since most colorectal cancers develop from polyps, a regular inspection of the colon is recommended, in order to detect lesions with a malignant potential or early cancer. A common medical procedure to examine the inside of the colon is colonoscopy, which is usually carried out with a conventional video-endoscope. A diagnostic benefit can be achieved by employing so called zoom-endoscopes, which achieve a magnification factor of up to 150 by means of an individually adjustable lens. In combination with dye-spraying to enhance the visual appearance (chromo-endoscopy) of the colon mucosa, zoom-endoscopy can reveal characteristic surface patterns, which can be interpreted by experienced physicians. Commonly used dyes are either methylene-blue, or indigo-carmine, which both lead to a plastic effect. In the research work of Kudo *et al*. [12, 13], the macroscopic appearance of colorectal polyps is systematically described and results in the so called *pit-pattern* classification scheme.

In this work, we examine the marginal distributions of the complex wavelet coefficient magnitudes of each subband of Kingsbury's Dual-Tree Complex Wavelet Transform (DT-CWT) [9, 18, 10] for image feature extraction, to allow computer-assisted pit-pattern classification. Existing approaches in this research area include histogram- and 2-D DWT-based methods for pit-pattern classification [4, 5] or classic video-endoscopy image classification by statistical second-order measures [15, 7].

The outline of this paper is as follows. In Section 2, we briefly introduce the pit-pattern classification scheme. Section 3 discusses the principles of the DT-CWT, together with the proposed feature extraction process and provides the statistical background of our work. In Section 4, we briefly depict the classification step and Section 5 presents experimental results. Section 6 concludes the paper with a summary of our work and an outlook on further research.

2. 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[1] of metaplastic lesions is unnecessary and removal of invasive cancer may be hazardous. The classification scheme of Kudo *et al*. divides the mucosal crypt patterns into five groups (pit-patterns I-V, see Figures 1,2).

While types I and II are characteristic of benign lesions and represent normal colon mucosa or hyperplastic polyps, types III-V represent neoplastic, adenomatous and carcinomatous structures. Our classification problem can be stated as follows: the problem to differentiate pit-types I and II from III-V will be denoted as the *two-class* problem, whereas the more detailed discrimination of pit-types I-V

[1] the process of removing polyps

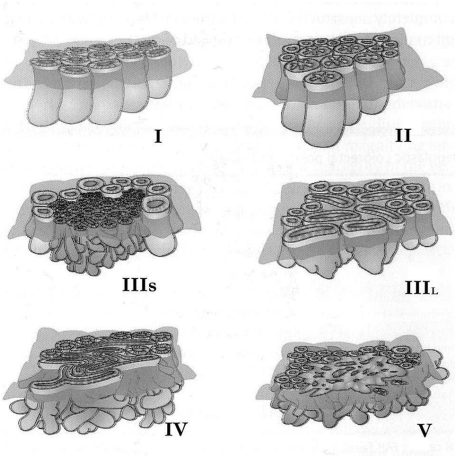


Figure 1. Pit-Patterns I-V

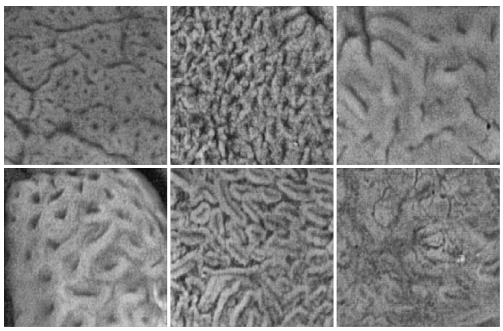


Figure 2. Endoscopy Images for pit-patterns I-V (left to right)

will be denoted as the *six-class* problem. Note, that pit-type III is subdivided into types III-S/III-L and thus accounts for two classes.

3. Feature Extraction

The process of feature extraction for pattern classification aims at characterizing an object by a set of measurements whose values are very similar to objects in the same class, and different for objects in different classes [2]. Without any feature extraction, we could directly consider the values of all $N \times N$ pixels as features for the classification process. However, this would immediately lead to the so called *curse of dimensionality* [1], which denotes the problem, that the number of needed training samples grows exponentially with the number of feature vector dimensions. Since we have a rather limited training sample size (see Section 5) compared to the dimensionality of the input space $\mathcal{X} \subseteq \mathbb{R}^{N \times N}$, this would lead to very poor classification results.

In this work, the feature extraction step is based on the assumption that the zoom-endoscopy images can be distinguished by textural measures. These measures are calculated from the subbands of a wavelet transform variant, known as the Dual-Tree Complex Wavelet Transform.

3.1. The Dual-Tree Complex Wavelet Transform

To obtain a multi-directional multi-resolution image representation for each image of our database, we employ Kingsbury's Dual-Tree Complex Wavelet Transform, which is an efficient realization of a complex wavelet transform and allows perfect reconstruction.

We have chosen the DT-CWT, since it is designed to overcome two commonly known shortcomings of the real, separable 2-D DWT, which are lack of shift-invariance and poor directional selectivity. In addition to that, the DT-CWT is nearly rotationally invariant in 2-D as well. These properties, which come at the expense of a limited redundancy of 2^m in m dimensions, are important due to the following reasons. First, our images exhibit structures, orientated along orientations other than horizontal, vertical or diagonal (see Figure 2), which are emphasized by the classic 2-D DWT. Second, the image acquisition process during colonoscopy is subject to several physical influences, which often cause image rotations and shifts. This fact highlights the aforementioned approximate rotation- and shift-invariance property of the DT-CWT.

Now that we justified our decision for the used image decomposition, we briefly motivate Kingsbury's dual-tree approach. First, we note that shift-invariance in 1-D can approximately be achieved with a real 1-D DWT by doubling the sampling rate at each level of the decomposition. However, this is equivalent to using two parallel, fully-decimated real DWTs (two trees), where the delay of the lowpass filters in the first stage of one tree is one sample offset from the delay of the filters in the other tree. For subsequent stages, the filters in the two trees must satisfy the requirement that the delays are half a sample different, which is also known as the *half-sample delay condition* [18]. If this condition is satisfied, the corresponding wavelets in each tree form a so called Hilbert-transform pair.

In our work, we use filters obtained from Kingsbury's Q-shift filter design procedure [8], which results in filters having a group delay of one-quarter. To fulfill the half-sample delay condition, the time-reverse of the filters in one tree are used in the opposite tree. The rather practicable property of the DT-CWT now is, that the outputs of the two trees can be interpreted as the real and imaginary part of a complex process. The resulting complex wavelet is an approximately analytic function, supported on only the positive half of the frequency axis. The indices h and g denote the filter sets $\{h_0, h_1\}, \{g_0, g_1\}$, which are used to implement the transform at levels ≥ 2. At the first level, any perfect reconstruction filters can be used, if they are offset from each other by one sample. In both 1-D and 2-D, the magnitudes of the complex wavelet coefficients are approximately shift-

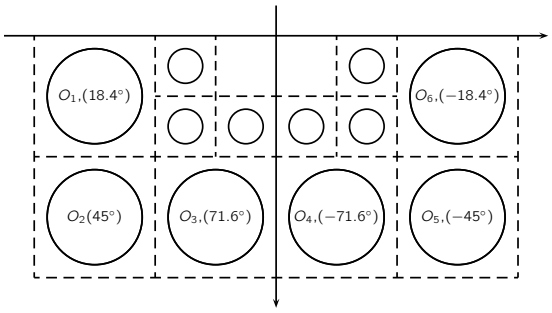


Figure 3. Orientations and subband numbering of the DT-CWT subbands

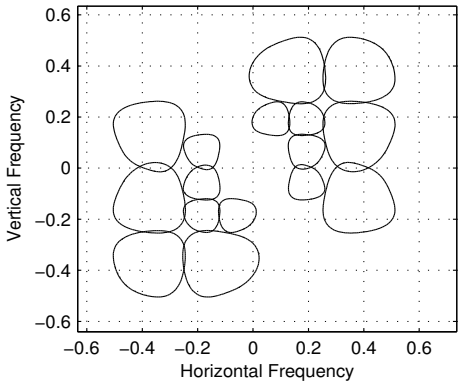


Figure 4. Frequency partitioning of the 2-D DT-CWT [16]

invariant, although the phase varies rapidly.

The extension of the DT-CWT to 2-D is straightforward by using two-trees for the rows and two trees for the columns, which results in a quad-tree structure. The great advantage of the DT-CWT in 2-D is its true directional selectivity, which results from the fact, that the spectrum of an approximately analytic wavelet is supported on the positive frequency axis only. Suppose, we have a complex scaling function $\phi(x) = \phi_h(x) + j\phi_g(x)$ and a complex (approx. analytic) wavelet $\psi(x) = \psi_h(x) + j\psi_g(x)$. Then, by taking the real and imaginary parts of $\psi(x)\psi(y)$, $\psi(x)\overline{\psi(y)}$, $\psi(x)\phi(y)$, $\phi(x)\overline{\psi(y)}$ and $\psi(x)\overline{\phi(y)}$, we obtain a total of twelve real-valued wavelets or six complex wavelets, oriented along six different directions. Figure 3 shows a schematic frequency partitioning of the DT-CWT, together with the orientations and the subband numbering, we use throughout our work.

For a concrete implementation of the DT-CWT, the aforementioned wavelet construction implies that the detail coefficients of each of the four trees have to be combined (sum and difference) to form the complex subbands [18]. To illustrate the real frequency partitioning of the DT-CWT, Figure 4 shows the 70% peak magnitude of a selection of filter responses in the frequency domain at levels ≥ 2. Regarding the choice of filters, all results in this paper were obtained by using near-symmetric (13,19)-tap filters at level 1 and Q-Shift (14,14)-tap filters at levels ≥ 2 [10].

3.2. Modeling the Marginal Distributions

Before decomposing the grayscale images with the DT-CWT, we conduct two preprocessing steps, to enhance visual quality. First, we employ adaptive histogram equalization using the CLAHE [20] (contrast limited adaptive histogram equalization) algorithm with 8×8 tiles and an uniform distribution for constructing the contrast transfer function. Second, we blur the images with a Gaussian 3×3 mask and $\sigma = 0.5$.

In the remainder of the paper, we will adhere to the following notation. The complex wavelet coefficients of subband k will be denoted by $x_{kn} \in \mathbb{C}, 1 \leq n \leq N_k$, with N_k denoting the total number of coefficients at subband k. The real and imaginary part of x_{kn} will be denoted by $\Re(x_{kn}) =: x_{r,kn}$ and $\Im(x_{kn}) =: x_{i,kn}$ respectively.

Probably the most commonly used statistic for the marginal wavelet coefficient magnitude distributions in texture retrieval [16] or texture classification [6] problems is the empirical mean and the empirical standard deviation, which are given by

$$m_k = \frac{1}{N_k} \sum_{j=1}^{N_k} |x_{kj}| \tag{1}$$

and

$$s_k = \frac{1}{N_k} \left(\sum_{j=1}^{N_k} (|x_{kj}| - m_k)^2 \right)^{\frac{1}{2}}, \tag{2}$$

with $|x_{kn}| = \left(x_{r,kn}^2 + x_{i,kn}^2 \right)^{1/2}$. Given, that the coefficient magnitudes at subband k follow a normal distribution with mean μ_k and standard deviation σ_k, equations (1) and (2) give the maximum likelihood estimates (MLE) for these parameters. Other commonly measures to characterize the marginal distributions are the information entropy or the subband energy for example.

However, we propose that modeling the marginal distributions of the wavelet coefficient magnitudes provides a good basis for feature extraction and implicitly improves classification accuracy for our classification problem. In [17], the complex coefficients are considered to be two-dimensional random vectors, where the random variables are statistically uncorrelated. This is justified by the Hilbert transform pair property we mentioned in Section 3.1. If we further follow the assumptions of [19], where the real and imaginary parts of the complex wavelet coefficient magnitudes are modeled by zero-mean Gaussian random variables X, Y with equal variances, then the random variable

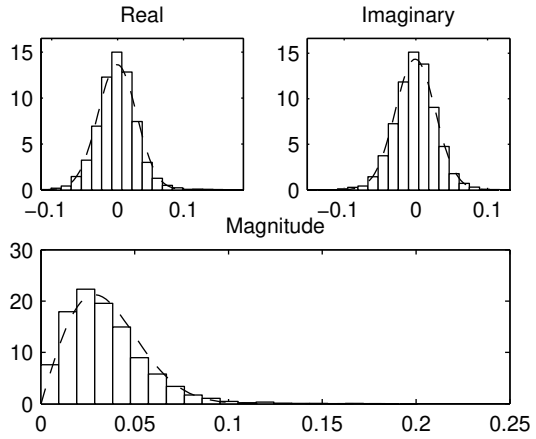


Figure 5. Histograms of the magnitude and real/imaginary part of the complex wavelet coefficients on level 2 of a DT-CWT decomposition together with fitted Gaussians (top) and a fitted Rayleigh distribution (bottom)

$Z = \sqrt{X^2 + Y^2}$ follows a Rayleigh distribution [11] with parameter b (denoted by $Z \sim \mathcal{R}(b)$). Formally, we have the well known relationship

$$X \sim \mathcal{N}(0,b), Y \sim \mathcal{N}(0,b) \Rightarrow Z \sim \mathcal{R}(b), \qquad (3)$$

which can easily be proven by first transforming the variables X, Y to polar coordinates (r, ϕ) and integrating over the angle ϕ of the joint distribution of r and ϕ. To see, if the Rayleigh distribution, given by the probability density function

$$p(x;b) = \frac{x}{b^2} \exp\left(-\frac{x^2}{2b^2}\right), x > 0, b > 0 \qquad (4)$$

is a suitable model for the subband coefficient magnitudes, we have to check whether the assumptions of normality, zero-mean and homogeneity of variances are satisfied. Note, that the following statements are based on our image database, which will be introduced in Section 5.

A simple visual inspection of the histograms of the real and imaginary part together with fitted normal distributions, can lead to the elusive conclusion, that the model requirements are satisfied (see Figure 5).

However, the requirements imposed on the random variables are quite rigorous and we will show that they cannot be satisfied for every subband. Deviations from the fundamental assumptions of normality, zero-mean and homogeneity of variances will lead to other distributional models fitting better to the wavelet coefficient magnitudes than the Rayleigh distribution. Since we will not rely on visual inspections alone, we conduct several statistical tests to verify the model requirements. We perform a Lilliefors [14] test on the real and imaginary part of the complex subband coefficients to check whether the null-hypothesis for normality cannot be rejected at 1% significance. If this is the case, we further perform a F-test to check for homogeneity of variances and a T-test to test against the null-hypothesis of zero-mean, both times at 1% significance.

To show, that visual inspection of the histograms alone can actually be quite elusive, we take a look at Figure 5. Here, the null-hypothesis of normality would be rejected at 1% significance level for the imaginary part. However, it is virtually impossible to capture that visually. A total listing of the percentage of all subbands, where the null-hypothesis of normality could not be rejected is given in Table 1. The column labels $O_1, \ldots, O_6$ denote the six subband orientations, whereas the row labels $S_1, \ldots, S_4$ denote the four decomposition scales we used.

	O_1	O_2	O_3	O_4	O_5	O_6
S_1	0	0	0	0	0	0.21
S_2	50.00	78.72	57.44	52.69	76.65	46.49
S_3	38.22	70.04	33.47	36.16	67.77	38.43
S_4	63.43	65.29	67.98	67.56	73.97	62.60

Table 1. Percentage of all subbands, where the null-hypotheses (normality) of the Lilliefors tests could not be rejected at the 1% significance level

The percentage of all subbands, where the null-hypotheses for all three tests could not be rejected, is listed in Table 2 for all orientations and scales.

	O_1	O_2	O_3	O_4	O_5	O_6
S_1	0	0	0	0	0	0
S_2	0	0	0	0	0	0
S_3	0.41	0.41	1.86	1.65	0	0.21
S_4	47.52	49.38	53.51	53.72	59.50	48.97

Table 2. Percentage of subbands, where the null-hypotheses for all three tests (normality, zero-mean, homogeneity of variance) cannot be rejected at the 1% significance level

We point out, that at the first scale (denoted by S_1), the null-hypothesis for normality is rejected for almost all orientation subbands. A possible reason for that is, that the frequency response of the first stage is far from being analytic, since a different set of filters is used. In combination with the numbers in Table 2, where the total rates are even lower, the test results indicate, that a Rayleigh distribution is not the best model for our subband coefficient magnitudes as far as the model assumptions are concerned. We therefore consider the Weibull distribution [11, 3] as a reasonable alternative. The probability density function of a Weibull distribution with shape parameter c and scale parameter b is given by

$$p(x;c,b) = \frac{c}{b}\left(\frac{x}{b}\right)^{c-1} \exp\left\{-\left(\frac{x}{b}\right)^{c}\right\}, b > 0, c > 0. \tag{5}$$

A comparison of the probability density functions in equations (4) and (5) reveals, that the Rayleigh distribution is just a special case of the Weibull distribution with $b = \sqrt{2}b$ and fixed shape parameter $c = 2$. In order to justify, that the Weibull distribution is a better model for our subband coefficients, we check the probability plots for both distributions (see Figure 6), which constitute a reliable visual tool to assess which distribution fits best to our data.

The probability plots for both Weibull and Rayleigh distributions, can easily be constructed from the inverse cumulative distribution functions (ICDF), which can be given explicitly [11]. Based on the ICDFs, a so called probability paper can then be constructed, where the data points should follow a straight line, given that the assumed distribution constitutes a good model. Any deviation from a straight line is an evidence that the data does not stem from the assumed distributional model.

As we can see from probability plots in Figure 6 (randomly chosen subbands), the data points depart significantly from a straight line (dashed) in case of the Rayleigh probability plots. Compared to the Weibull probability plots, where the data points actually do form a straight line (at least approximately), this is a clear indication that the Weibull model provides a better fit to our data.

At the beginning of this section, we stated, that commonly used features computed from the wavelet coefficient magnitudes, are the empirical mean and the empirical standard deviation (see equations (1) and (2)). We further noted, that by using these statistical measures as features for classification purposes, we implicitly use the maximum likelihood estimates of a normal distribution. We propose, that an improvement in classification accuracy can be achieved by using the shape and scale parameter of fitted Weibull distributions as features for the classification process. However, before we can go further to explain the classification procedure, we have to take a closer look at parameter estimation, since the parameters constitute our feature vectors.

Regarding the estimation of shape and scale for the two-parameter Weibull distribution, the MLEs have to be determined numerically, since they cannot be given explicitly. Let $y_{ki} := |x_{ki}|, i = 1, \ldots, N_k$ be our sample observations drawn from a Weibull distribution and let $z_{ki} = \log(y_{ki}), i = 1, \ldots, N_k$. Then, the maximum likelihood estimate $\hat{c}_k$ of the shape parameter at subband k is given by the solution to equation [11]

$$\hat{c}_k = \left[\sum_{i=1}^{N_k} y_{ki}^{\hat{c}_k} z_{ki} / \sum_{i=1}^{N_k} y_{ki}^{\hat{c}_k} - \overline{y_k}\right]^{-1}, \tag{6}$$

with

$$\overline{y}_k = \frac{1}{N_k}\sum_{i=1}^{N_k} y_{ki}. \tag{7}$$

The scale parameter $\hat{b}_k$ at subband k then follows from

$$\hat{b}_k = \left(\frac{1}{N_k}\sum_{i=1}^{N_k} y_{ki}^{\hat{c}_k}\right)^{1/\hat{c}_k}. \tag{8}$$

In our implementation, we have used MATLAB's optimization toolbox for the numerical calculations. Since we need a starting value for the calculation of $\hat{c}_k$, we exploit the fact, that the cumulative distribution function (CDF) of the Weibull distribution, given by

$$F(x) := F(x;b,c) = 1 - \exp\left\{-\left(\frac{x}{b}\right)^{c}\right\}, b > 0, c > 0 \tag{9}$$

can easily be transformed into a Weibull *identification plot* by taking advantage of the following property:

$$\log(-\log(1 - F(x))) = c(log(x) - log(b)) \tag{10}$$

Now, let $\hat{F}_k(x)$ be the empirical CDF of the wavelet coefficient magnitudes of subband k, then, by plotting $\log(-\log(1 - \hat{F}_k(x)))$ against $log(x)$ we can fit a straight line $\hat{\alpha}_k x + \hat{\beta}_k$ to the data points and obtain estimates $\hat{b}_k, \hat{c}_k$ through

$$\hat{c}_k = \hat{\alpha}_k, \hat{b}_k = \exp\left(\frac{\hat{\alpha}_k}{\hat{c}_k}\right). \tag{11}$$

The numerical value of $\hat{c}_k$ in (11), obtained by a least-squares fit, can now be used as a starting point to solve (6). Since we will also evaluate the discriminative power of the Rayleigh distribution parameter b, we need its maximum likelihood estimate $\hat{b}$ as well. Assuming that our sample $y_{ki}, i = 1, \ldots, N_k$ is drawn from a Rayleigh distribution with parameter b, the explicit solution for the MLE $\hat{b}_k$ is given by [11]

$$\hat{b}_k = \left(\frac{1}{2N_k}\sum_{i=1}^{N_k} y_{ki}^2\right)^{1/2}. \tag{12}$$

By now, we can calculate all necessary estimates for the parameters of our distributional models and are ready to discuss the classification process. In a formal context, we have three different mappings $\omega_i : \mathcal{X} \to \mathcal{F}_i, i = 1, 2, 3$, from our input space $\mathcal{X} \subseteq \mathbb{R}^{N \times N}$ into three lower dimensional feature spaces $\mathcal{F}_i \subseteq \mathbb{R}^d, i = 1, 2, 3$, with either $d = 48$ in case of the Weibull or classic features and $d = 24$ in case of the Rayleigh-based features. An exemplary feature vector for image j, composed by the MLEs of the Weibull distribution parameters has the form

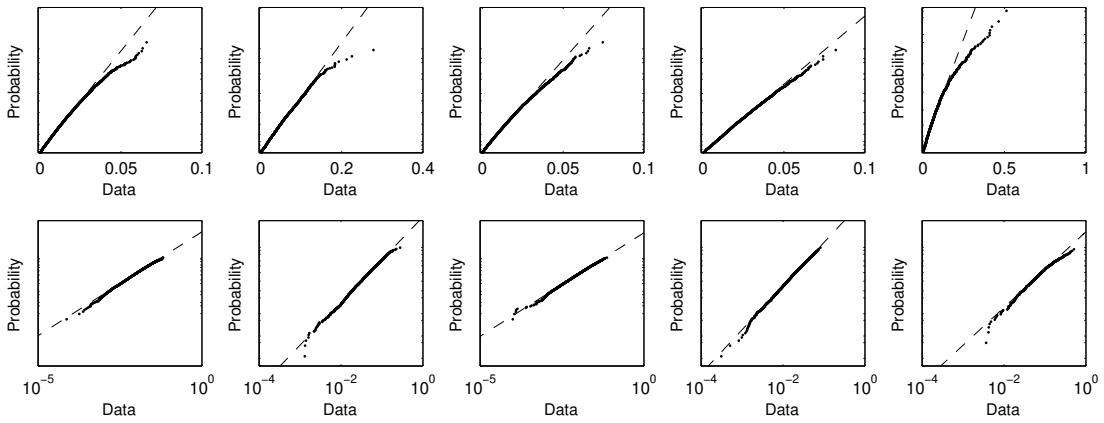


Figure 6. Rayleigh (top) and Weibull (bottom) probability-probability plots for a selection of subband coefficient magnitudes

$$\mathbf{v}_j = (v_{j1}, \ldots, v_{jd}) \tag{13}$$

with

$$(v_{j1}, \ldots, v_{jd}) := (\hat{b}_{j1}, \hat{c}_{j1}, \ldots, \hat{b}_{jd/2}, \hat{c}_{jd/2}). \tag{14}$$

4. Classification

In this work, we employ a simple k-Nearest Neighbor (denoted by k-NN) classifier, which uses the euclidian formula in d dimensions as a distance metric $D : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$. Given two sample feature vectors $\mathbf{v}_m, \mathbf{v}_n \in \mathbb{R}^d$, we thus have

$$D(\mathbf{v}_m, \mathbf{v}_n) = \left(\sum_{i=1}^{d} (v_{mi} - v_{ni})^2 \right)^{1/2} \tag{15}$$

Now let $\mathcal{D} = (\mathbf{v}_i, y_i) \in \mathcal{X} \times \mathbf{Y} := \{1, \ldots, L\}$ be a collection of N labeled feature vectors (our training set), where $\mathbf{Y}$ denotes the set of possible class labels. Further, let $\mathbf{v}$ be a new (unclassified) feature vector. According to the k-NN classification rule, the new sample is classified by assigning it the label most frequently represented among the k nearest samples (calculated according to our distance metric) [2].

To estimate the classification accuracy in our classification problem, defined as the number of correctly classified samples divided by the total number of samples, we employ the method of leave-one-out crossvalidation (LOOCV). This method is defined as follows: training the k-NN classifier is done N times, each time using $\mathcal{D}$ from which a different single sample has been deleted [2]. The classifier is then tested on the single left-out sample and the overall classification accuracy is determined by averaging the results of all N iterations.

However, we still have to remedy one more problem, which is related to our metric. It is well known, that the 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 differ significantly. In Figure 5 for example, the MLEs would be $(\hat{b}, \hat{c}) = (0.074, 1.88)$, which differ by several orders of magnitude. Since we do not want the shape parameter $\hat{c}$ to have a greater influence on the distance metric, we apply a linear transformation on the features of each feature vector. Given our d-dimensional training samples $\mathbf{v}_1, \ldots, \mathbf{v}_N$, the normalization formula (see [1]) for the n-th element of the j-th feature vector is defined by

$$\tilde{v}_{jn} = \frac{v_{jn} - \overline{v}_n}{s_n}, \tag{16}$$

where $\overline{v}_n, s_n$ denote the sample mean and the sample variance of the n-th feature. Thus, we obtain re-scaled features with zero-mean and unit standard deviation. Now, each feature contributes equally to the calculation of the metric in (15). Of course, we must take care, that this linear transform is repeated in every single iteration of the LOOCV process to avoid taking too much information into account. Practically, this means that at iteration j, $\overline{v}_n$ and s_n are calculated on the basis of the N-1 training samples and the elements of the left-out test sample are then normalized by (16) using these values.

5. Experimental Results

Our image database consists of a total of 484 images, acquired in 2005/2006 at the Department of Gastroenterology and Hepatology (Medical University of Vienna) using a zoom-endoscope (Olympus Evis Exera CF-Q160ZI/L) with a magnification factor of 150. To enhance visual appearance, dye-spraying with indigo-carmine was applied

I	II	III-L	III-S	IV	V
126	72	62	18	146	60

Table 3. Number of images per pit-pattern class (ground truth)

Features	2-class	6-class
Classic	93.60	80.17
Weibull	**95.87**	**81.61**
Rayleigh	87.60	75.41

Table 4. LOOCV accuracy results for all three discussed feature sets

and biopsies or mucosal resections were taken to obtain a histopathological diagnosis. For pit-pattern types I,II and V, biopsies were taken, since these types need not be removed. Lesions of pit-pattern types III-S/III-L and IV have been removed endoscopically. Table 3 lists the number of image samples per class.

As we have noted in Section 3, a four level DT-CWT leads to a total of 24 directional subbands. However, it is questionable that all subbands contribute substantial information to the classification process. It might as well be, that a special combination of decomposition scales and subbands leads to better classification rates than using all subbands at each scale. For that reason, we vary the number of decomposition scales from 1-4 and choose the very scale-subband combination which leads to the highest LOOCV accuracy. In other words, exhaustive search in the space of all possible 2^{24} scale-subband combinations would have to be performed. However, this is hardly acceptable. To reduce computation time, we decided that the same subband combination should be used at each scale. This intuitive decision, which will need further investigation, is based on the idea, that if a specific orientation contributes a lot of information to the classification process, it should be used throughout all scales. The number of possible combinations for a four-scale decomposition thus reduces to

$$\sum_{k=1}^{4}\binom{4}{k}\sum_{k=1}^{6}\binom{6}{k} = (2^4-1)(2^6-1) \qquad (17)$$

which is feasible to calculate. Two possible scale-subband combinations are illustrated in Figure 7. As we can see, the same subbands are selected at each scale. For the following results, we will use the same subband numbering as in Figure 3.

Table 5 presents the results for the two-class and six-class classification problem for all three discussed feature sets. The results were obtained on the basis of a 1-NN classifier, since our experiments have shown, that $k = 1$ produced the best results when varying k between 1 and 20. The best LOOCV results are marked bold.

Features	**2-class**.
Classic	$\{S_1, S_3, S_4\} - \{O_2, O_3, O_6\}$
Weibull	$\{S_2, S_4\} - \{O_2, O_3, O_4, O_6\}$
Rayleigh	$\{S_1, S_2, S_3, S_4\} - \{O_1, O_3, O_4, O_5, O_6\}$
Features	**6-class**.
Classic	$\{S_1, S_2, S_3, S_4\} - \{O_1, O_2, O_3, O_4, O_6\}$
Weibull	$\{S_1, S_2, S_3, S_4\} - \{O_1, O_3, O_4, O_5, O_6\}$
Rayleigh	$\{S_1, S_2, S_3, S_4\} - \{O_2, O_3, O_4, O_5, O_6\}$

Table 5. Optimal scale-orientation combinations for all three feature sets

Problem	Proposed	[4]	[5]
2-class	**95.87**	85.6	67.3
6-class	**81.61**	75	57

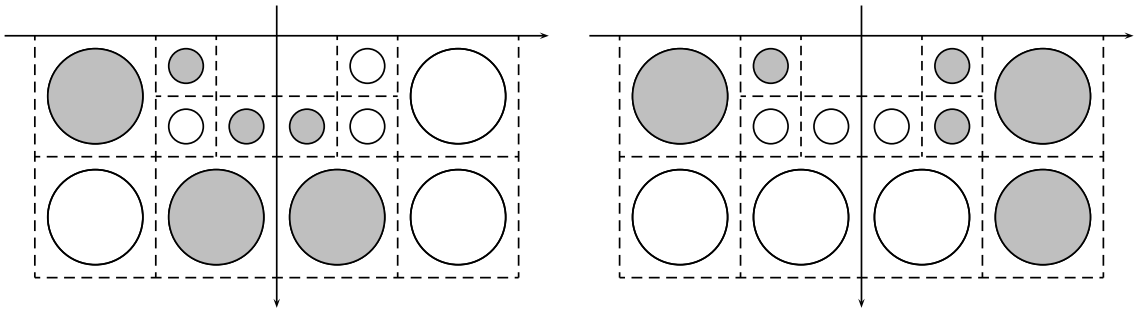
Table 6. Comparison of our proposed approach to two other approaches in terms of LOOCV accuracies

The optimal scale-subband combinations for all three feature sets with regards to our scale-subband selection procedure are given in Table 5. We use the notation $\{S_1, \ldots, S_4\} - \{O_1, \ldots, O_6\}$, where the variables in the first curly brackets denote the scales 1-4, and the variables in the second curly brackets denote the used subbands. For example, the specification $\{S_1, S_2\} - \{O_3, O_4\}$ signifies that subbands 3 and 4 are used on scales 1 and 2.

The results clearly indicate, that the Weibull-based features outperform classic mean/standard deviation and Rayleigh-based features in terms of LOOCV accuracy. Especially in the two-class problem, the rate is more than two percent higher than the rate obtained from the classic features. We therefore conclude, that for our classification problem, the shape and scale of the fitted Weibull distributions provide more discriminative power than the features commonly used for texture discrimination or texture image retrieval. Compared to the results in [4] and [5], which were obtained on the same image database, the proposed Weibull features lead to superior results by approximately 10% in the two-class problem and 6% in the six-class problem, both times referring to the highest k-NN LOOCV accuracies of [4, 5]. The numbers are listed in Table 6.

6. Conclusion

In this paper, we have shown that the marginal distributions of complex wavelet coefficient magnitudes from a DT-CWT can be well modeled by a Weibull distribution. By using the maximum likelihood estimates for shape and scale as features for each subband, we could significantly improve the LOOCV accuracy for two classification problems in medical imaging, compared to classic mean and standard deviation based features. Future research on this



(a) $\{S_1, S_2\} - \{O_1, O_3, O_4\}$ (b) $\{S_1, S_2\} - \{O_1, O_5, O_6\}$

Figure 7. Two possible scale-subband combinations (filled) for a schematic frequency partitioning of the DT-CWT

topic will include the incorporation of color information and scale dependencies of wavelet coefficients across subbands into the feature extraction process. Furthermore, the performance of other classifiers, such as support vector machines for example, will have to be evaluated.

Acknowledgments

This work is funded by the Austrian Science Fund (FWF) under Project No. L366-N15. Further, the authors wish to thank Dr. Karl Entacher for his support and stimulating discussions.

References

[1] C. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[2] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley & Sons, 2nd edition, Nov. 2000.

[3] M. Evans and N. H. B. Peacock. *Statistical Distributions*. Wiley Series in Probability and Statistics. Wiley, 3rd edition, 2000.

[4] M. Häfner, C. Kendlbacher, W. Mann, W. Taferl, F. Wrba, A. Gangl, A. Vécsei, and A. Uhl. Pit pattern classification of zoom-endoscopic colon images using histogram techniques. In J. R. Sveinsson, editor, *Proceedings of the 7th Nordic Signal Processing Symposium (NORSIG 2006)*, pages 58–61, Reykavik, Iceland, June 2006. IEEE.

[5] M. Häfner, M. Liedlgruber, F. Wrba, A. Gangl, A. Vécsei, and A. Uhl. Pit pattern classification of zoom-endoscopic colon images using wavelet texture features. In W. Sandham, D. Hamilton, and C. James, editors, *Proceedings of the International Conference on Advances in Medical Signal and Image Processing (MEDSIP 2006)*, Glasgow, Scotland, UK, July 2006. paper no. 0038.

[6] S. Hatipoglu, S. K. Mitra, and N. G. Kingsbury. Texture classification using dual-tree complex wavelet transform. In *Seventh International Conference on Image Processing and Its Applications*, volume 1, pages 344–347, Manchester, UK, July 1999.

[7] S. A. Karkanis. Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE Transactions on Information Technology in Biomedicine*, 7(3):141–152, Sept. 2003.

[8] N. Kingsbury. A dual-tree complex wavelet transform with improved orthogonality and symmetry properties. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'00)*, volume 2, pages 375–378, Vancouver, Canada, 2000.

[9] N. G. Kingsbury. 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*, pages 9–12, Bryce Canyon, USA, Aug. 1998.

[10] N. G. Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Applied and Computational Harmonic Analysis*, 10(3):234–253, May 2001.

[11] K. Krishnamoorthy. *Handbook of Statistical Distributions with Applications*. Chapman & Hall, 2006.

[12] S. Kudo, S. Hirota, T. Nakajima, S. Hosobe, H. Kusaka, T. Kobayashi, M. Himori, and A. Yagyuu. Colorectal tumorous and pit pattern. *Journal of Clinical Pathology*, 47:880–885, 1994.

[13] S. Kudo, S. Tamura, T. Nakajima, H. Yamano, H. Kusaka, and H. Watanabe. Diagnosis of colorectal tumorous lesions by magnifying endoscopy. *Gastrointestinal Endoscopy*, 44(1):8–14, July 1996.

[14] H. Lilliefors. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62:399–402, June 1967.

[15] D. E. Maroulis, D. K. Iakovidis, S. A. Karkanis, and D. A. Karras. CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames. *Computer Methods and Programs in Biomedicine*, 70(2):151–66, February 2003.

[16] P. Rivaz and N. Kingsbury. Complex wavelet features for fast texture image retrieval. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'99)*, pages 109–113, Kobe, Japan, 1999.

[17] J. Romberg, H. Choi, R. Baraniuk, and N. Kingsbury. Multiscale classification using complex wavelets. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'00)*, volume 2, pages 371–374, Vancouver, Canada, 2000.

[18] I. W. Selesnick, R. G. Baraniuk, and N. Kingsbury. The dual-tree complex wavelet transform - a coherent framework for multiscale signal and image processing. *IEEE Signal Processing Magazine*, 22(6):123–151, November 2005.

[19] C. Shaffrey, N. Kingsbury, and I. Jermyn. Unsupervised image segmentation via markov trees and complex wavelets. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'02)*, volume 3, pages 801–804, Rochester, New York, United States, 2002.

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