Directional wavelet based features for colonic polyp classification☆

Georg Wimmer [a],[*], Toru Tamaki [c], J.J.W. Tischendorf [e], Michael Häfner [b], Shigeto Yoshida [d], Shinji Tanaka [d], Andreas Uhl [a]

[a] University of Salzburg, Department of Computer Sciences, Jakob Haringerstrasse 2, 5020 Salzburg, Austria
[b] St. Elisabeth Hospital, Landstraßer Hauptstraße 4a, A-1030 Vienna, Austria
[c] Hiroshima University, Department of Information Engineering, Graduate School of Engineering, 1-4-1 Kagamiyama, Higashi-hiroshima, Hiroshima 739-8527, Japan
[d] Hiroshima University Hospital, Department of Endoscopy, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8551, Japan
[e] Medical Department III (Gastroenterology, Hepatology and Metabolic Diseases), RWTH Aachen University Hospital, Paulwelsstr. 30, 52072 Aachen, Germany

ARTICLE INFO

ABSTRACT

In this work, various wavelet based methods like the discrete wavelet transform, the dual-tree complex wavelet transform, the Gabor wavelet transform, curvelets, contourlets and shearlets are applied for the automated classification of colonic polyps. The methods are tested on 8 HD-endoscopic image databases, where each database is acquired using different imaging modalities (Pentax's i-Scan technology combined with or without staining the mucosa), 2 NBI high-magnification databases and one database with chromoscopy high-magnification images.

To evaluate the suitability of the wavelet based methods with respect to the classification of colonic polyps, the classification performances of 3 wavelet transforms and the more recent curvelets, contourlets and shearlets are compared using a common framework. Wavelet transforms were already often and successfully applied to the classification of colonic polyps, whereas curvelets, contourlets and shearlets have not been used for this purpose so far.

We apply different feature extraction techniques to extract the information of the subbands of the wavelet based methods. Most of the in total 25 approaches were already published in different texture classification contexts. Thus, the aim is also to assess and compare their classification performance using a common framework. Three of the 25 approaches are novel. These three approaches extract Weibull features from the subbands of curvelets, contourlets and shearlets. Additionally, 5 state-of-the-art non wavelet based methods are applied to our databases so that we can compare their results with those of the wavelet based methods.

It turned out that extracting Weibull distribution parameters from the subband coefficients generally leads to high classification results, especially for the dual-tree complex wavelet transform, the Gabor wavelet transform and the Shearlet transform. These three wavelet based transforms in combination with Weibull features even outperform the state-of-the-art methods on most of the databases. We will also show that the Weibull distribution is better suited to model the subband coefficient distribution than other commonly used probability distributions like the Gaussian distribution and the generalized Gaussian distribution.

So this work gives a reasonable summary of wavelet based methods for colonic polyp classification and the huge amount of endoscopic polyp databases used for our experiments assures a high significance of the achieved results.

1. Introduction

In this paper, wavelet based methods are applied for the automated classification of colonic polyps in endoscopic images. Wavelet transforms like the discrete wavelet transform (DWT), the dual-tree complex wavelet transform (DT-CWT) and the Gabor wavelet transformation have been widely used for the purpose

of medical image analysis. In case of colonic polyp classification, especially the DT-CWT proved to be quite suitable for the distinction of different types of polyps as can be seen in numerous previous papers like e.g. Häfner et al. (2015a); 2009); 2010) Also Gabor wavelets have proved to be quite suitable for colonic polyp classification (Yuan and Meng, 2014; Häfner et al., 2009) and detection (Hwang and Celebi, 2010). The DT-CWT and the Gabor wavelets are both directional selective wavelet transforms, contrary to the classical DWT. It has been shown in Häfner et al. (2009), that these two directional selective wavelet transforms provide better results than the DWT. So enhanced directional selectivity may be an advantage classifying polyps.

Based on the wavelet theory, new multiresolution analysis tools like the curvelet, contourlet and shearlet transform have been developed. These transforms (further denoted as Lets) are even more directional selective than the DT-CWT and Gabor transformation. To the best of our knowledge, until now solely the curvelet transform was applied for the automated detection or classification of polyps, however solely for small bowel tumors using capsule endoscopy (Barbosa et al., 2009; Martins et al., 2010).

In this paper we use a common framework to compare the results of the wavelet transforms and Lets for the classification of colonic polyp in endoscopic images. To the best of our knowledge, there has not been a comparison of wavelet transforms and Lets with respect to the classification of images so far (the same applies for related issues like image retrieval or pattern and object recognition). So in spite of the similarity of wavelets and Lets, this is the first publication which systematically compares these transforms with respect to their suitability to classify texture images. In order to ensure a fair comparison of the wavelet based methods, we extract the same features (Gaussian, generalized Gaussian and Weibull distribution parameters) and use the same number of scale levels for each method. To ensure a high significance of the results, the wavelet based methods are applied to a total of 11 different endoscopic polyp databases. Feature extraction approaches using wavelet transforms already proved to be an appropriate choice in various publications. By means of our test we will see if the same applies to curvelets, contourlets and shearlets. Additionally we reimplemented some Let-based texture recognition approaches and applied them to the classification of our polyp databases to have a higher variability of extracted features and to find out which features extracted from Lets are most appropriate for our task. The results of the wavelet based approaches are compared with those of 5 non-wavelet based state-of-the-art approaches in colonic polyp classification.

But first let us introduce and motivate the employed wavelet based transforms:

Wavelet transforms use filterbanks to form a time-frequency representation for continuous-time signals. The main difference between the wavelet transform and the Fourier transform (FT) is that wavelets are localized in time and frequency whereas the standard Fourier transform is only localized in frequency. Because of the uncertainty principle, originally found and formulated by Heisenberg, the frequency and time information of a signal at some certain point in the time-frequency plane cannot be known. In other words: we cannot know what spectral component exists at any given time instant. The best we can do is to investigate what spectral components exist at any given interval. The wavelet transform deals with that problem by decomposing a signal in frequency bands (called subbands), where the higher frequency bands are better resolved in time (with less relative error) and the lower frequency bands are better resolved in frequency.

Wavelets are widely used for data compression, signal analysis, signal reconstruction, denoising, etc. One of the most useful features of wavelets is their ability to efficiently approximate signals,
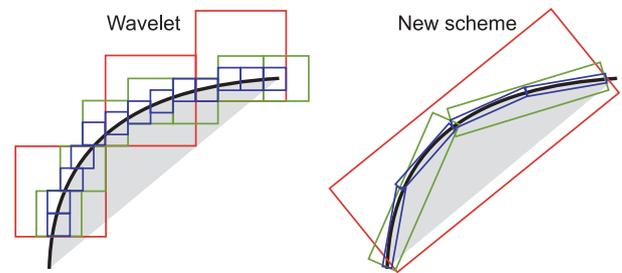

Fig. 1. Wavelet vs new scheme: an illustration of the successive refinement by the two systems near a smooth contour, which is shown as the black curve separating two smooth regions.

that means to represent a signal as accurately as possible by means of a minimum of subband coefficients. Especially for signals with pointwise singularities, the DWT is much more efficient than the Fourier transform. This motivates why wavelet transforms are now being adopted for a vast number of applications, often replacing the conventional Fourier transform.

However, the DWT does not perform as well with multidimensional data. Indeed, the DWT is very efficient in dealing with pointwise singularities only. In higher dimensions, other types of singularities (e.g. edges in images) are usually present or even dominant, and the DWT and other traditional wavelet methods are unable to handle them efficiently. In order to overcome this limitation of traditional wavelets, one has to increase their directional sensitivity. Two well known directional selective wavelet transforms are the Dual-tree complex wavelet transform (DT-CWT) (Kingsbury, 1998) and the Gabor wavelet transform (Lee, 1996).

Based on the wavelet theory, new multiresolution analysis tools have been developed that are especially designed to efficiently representate edges and curves in 2-dimensional data. The idea behind this new schemes can be described by the following scenario (Easley et al., 2008). Imagine that there are two painters, one with a "wavelet"-style and the other using the new scheme, where both wish to paint a natural scene. Both painters apply a refinement technique to increase resolution from coarse to fine. Efficiency is measured by the number of brush strokes needed to faithfully recover the scene. We consider the situation that a smooth contour has to be painted like shown in Fig. 1.

2-D wavelets are constructed from tensor products of 1-D wavelets, so the "wavelet"-style painter is limited to use square-shaped brush strokes along the contour, using sizes corresponding to the multiresolution structure of wavelets. As the resolution becomes finer, we clearly see the limitations of the painter, who needs to use many "dots" to capture the contour. The new style painter, on the other hand, is much more effective by making brush strokes with differently elongated shapes, where the directions of the shapes follows the contour. That means many wavelet coefficients are needed to account for edges or curves and it would be far more effective to have strongly anisotropic filters to represent edges of curves. This idea was implemented by a number of new wavelet-based approaches. The most established approaches using this new scheme are the curvelet transform (Candes and Donoho, 2002), the contourlet transform (Do and Vetterli, 2005) and the shearlet transform (Easley et al., 2008). We further denote these transforms as "Lets".

These Lets use non-separable filters which have elongated supports at various scales, directions and aspect ratios (the finer the scale, the higher is the aspect ration or in other words the more elongated are the supports). This allows an efficient approximation of smooth contours at multiple resolutions in much the same way as the new scheme shown in Fig. 1. Moreover, these Lets are able to use different numbers of directions at each scale (generally, the finer the scale, the more directions).

The contributions of this manuscript are as follows:

- We apply a total of 25 wavelet based methods for the automated classification of colonic polyps. 5 methods are based on the curvelet transform, 5 on the contourlet transform, 6 on the shearlet transform, 3 on the DWT, 3 on the DT-CWT and 3 on the Gabor transformation. By means of these experiments we are able to compare Lets and wavelet methods with respect to their classification performance. Most of the methods were already proposed in different texture classification contexts , but three of these methods are novel to the best of our knowledge. In these three methods, the subband coefficients of the curvelet, contourlet and shearlet transform are modeled by the 2 parameter Weibull distribution. We will show that modeling the subband coefficients by means of the Weibull distribution generally leads to the best results for classifying colonic polyps using wavelet based methods.
- We apply the Kolmogorov–Smirnov test as Goodness-of-Fit test and show that the Weibull distribution is well suited to model the subband coefficient distribution of the wavelet based transforms, which explains the superior results using Weibull features. It will turn out that the subbands are not actually Weibull distributed, but at least almost Weibull distributed.
- For our experiments we use a total of 11 different endoscopic databases. 8 databases are gathered using a HD-endoscope with 8 different imaging modalities (Pentax's i-Scan in combination with staining the mucosa), 1 databases is gathered using high magnification endoscopy (or also called zoom-endoscopy) in combination with staining the mucosa and two databases are gathered using a zoom-endoscopy in combination with narrow band imaging (NBI). So we use a quite comprehensive collection of databases for the classification of colonic polyps. The results of the methods are compared and the differences between the methods as well as their impacts to the results are analyzed.
- 5 (non wavelet based) state-of-the-art approaches for colonic polyp classification are applied to the classification of our databases to compare their results with the results of the wavelet based methods. In this way we are able to find out if there are wavelet-based methods that can compete with state-of-the-art approaches. We will see that some of the wavelet-based methods even outperform the state-of-the-art approaches, while others perform equally or inferior compared to the state-of-the-art approaches.

This paper is organized as follows. In Section 2 we briefly introduce the concept of the computer-assisted diagnosis of polyps using mucosa texture patches and review the corresponding state-of-the-art. In Section 3, we describe and compare the wavelet based approaches. The experimental setup, the used databases and the results are presented in Section 4. Section 5 presents the discussion and Section 6 concludes our work.

2. Colonic polyp classification

Colonic polyps are a rather frequent finding and are known to either develop into cancer or to be precursors of colon cancer. Hence, an early assessment of the malignant potential of such polyps is important as this can lower the mortality rate drastically. As a consequence, a regular colon examination is recommended, especially for people at an age of 50 years and older. The current gold standard for the examination of the colon is colonoscopy using a colonoscope. Modern endoscopy devices are able to take pictures or videos from inside the colon, allowing to obtain images (or videos) for a computer-assisted analysis with the goal of detecting and diagnosing abnormalities.

Colonic polyps are usually divided into hyperplastic, adenomatous and malignant polyps. In order to determine a diagnosis based
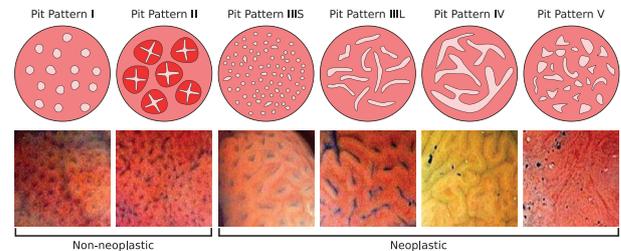

Fig. 2. The 6 pit pattern types along with exemplar images and their assigned classes in case of a two class (non-neoplastic vs neoplastic) differentiation.

on the visual appearance of colonic polyps, the pit pattern classification scheme was proposed by Kudo et al. (1994). A pit pattern refers to the shape of a pit, the opening of a colorectal crypt. The various pit pattern types and exemplar (zoom-endoscopic) images of the classes are presented in Fig. 2. The pit pattern classification scheme differentiates between six types. Type I (normal mucosa) and II (hyperplastic polyps) are characteristics of non-neoplastic lesions, type III-S, III-L and IV are typical for adenomatous polyps and type V is strongly suggestive to malignant cancer.

So this classification scheme allows to differentiate between normal mucosa and hyperplastic lesions, adenomas (a premalignant condition), and malignant cancer based on the visual pattern of the mucosal surface. The removal of hyperplastic polyps is unnecessary and the removal of malignant polyps maybe hazardous. In this work we use the 2-class classification scheme differentiating between non-neoplastic and neoplastic lesions. This classification scheme is quite relevant in clinical practice as indicated in a study by Kato et al. (2006).

For an easier detection and diagnosis of the extent of mucosal lesions, two common mucosal enhancement technologies were developed:

1. Conventional chromoendoscopy (CC) came into clinical use 40 years ago. By staining the mucosa using (indigocarmine) dye spray, it is easier to detect and differentiate colonic polyps. CC is often used in conjunction with high-resolution or magnification endoscopy.
2. Digital chromoendoscopy is a technique to facilitate "chromoendoscopy without dyes" (Kiesslich, 2009). The strategies followed by major manufacturers differ in this area:
   - In Narrow band imaging (NBI, Olympus), narrow bandpass filters are placed in front of a conventional white-light source to enhance the detail of certain aspects of the surface of the mucosa.
   - The i-Scan (Pentax) image processing technology (Kodashima and Fujishiro, 2010) is a digital contrast method which consists of combinations of surface enhancement, contrast enhancement and tone enhancement.
     The FICE system (Fujinon) decomposes images by wavelength and then directly reconstructs images with enhanced mucosal surface contrast.
     Both systems (i-Scan and FICE) apply post-processing to the reflected light and thus are called "computed virtual chromoendoscopy (CVC)".

Previous works for the computer assisted classification of colonic polyps using highly detailed images gathered from endoscopes in combination with different imaging modalities, can be divided in three categories:

- High definition (HD) endoscope combined with or without staining the mucosa and the i-Scan technology:
  In Häfner et al. (2014a), shape and contrast features were extracted from blobs and in Häfner et al. (2015b); 2014c) fractal analysis based features were extracted.

- High-magnification chromoendoscopy:
  In Häfner et al. (2012c), the pit density was estimated using Delaunay triangulation, local binary patterns based features were used in Häfner et al. (2009) and Häfner et al. (2012a) and features from wavelet transforms were extracted in Häfner et al. (2008); 2009); 2010); 2015a).
- High-magnification endoscopy combined with NBI:
  Tamaki et al. (2013) extracted dense SIFT features and Gross et al. (2012) extracted features describing the vessel structure.

In this work we use endoscopic image databases of all three categories.

One of the aims of this work is to compare the classification results of the databases of all three categories.

In addition to classical endoscopy, endomicroscopy, computed tomography (CT) and wireless capsule endoscopy can be used for the examination of the gastro-intestinal tract. Endomicroscopy (Jabbour et al., 2012) is a technique to obtain histology-like images and is also known as 'optical biopsy'. For example André et al. (2011); 2012) showed approaches based on semantics and visual concepts for the automated diagnosis of colonic polyps using endomicroscopy. CT colonography, also known as virtual colonoscopy, is a minimally invasive technique for the investigation of the colon. An example showing a detection and classification system based on Curvature Analysis using CT colonography can be seen in Chowdhury et al. (2008). Wireless capsule endoscopy (Iakovidis and Koulaouzidis (2015); Yuce and Dissanayake (2012)) is mainly used to examine parts of the gastrointestinal tract that cannot or only hardly be seen with other types of endoscopes (the small bowel). The capsule has the size and shape of a pill and contains a tiny camera. After a patient swallows the capsule, it takes images of the inside of the gastro-intestinal tract. An example for the automated detection and classification of colonic polyps using capsule endoscopy can be seen in Romain et al. (2013).

2.1. HD endoscopy combined with the i-Scan technology and chromoendoscopy

In this work we use a total of 8 image databases gathered by HD endoscopy. HD-endoscopy has the advantage of an higher resolution compared to standard definition endoscopes. Each database is gathered by a different combination of the i-Scan technology and CC, respectively no CC.

The three i-Scan modes are as follows:

1. i-Scan 1 includes surface enhancement and contrast enhancement. Surface enhancement mode augments pit pattern and surface details, providing assistance to the detection of dysplastic areas. This mode enhances light-to-dark contrast by obtaining luminance intensity data for each pixel and adjusting it to accentuate mucosal surfaces.
2. i-Scan 2 includes surface enhancement, contrast enhancement and tone enhancement. Expands on i-Scan 1 by adjusting the surface and contrast enhancement settings and adding tone enhancement attributes to the image. It assists by intensifying boundaries, margins, surface architecture and difficult-to-discern polyps.
3. i-Scan 3 also includes surface enhancement, contrast enhancement and tone enhancement. Similar to i-Scan 2, with increased illumination and emphasis on the visualization of vascular features. This mode accentuates pattern and vascular architecture.

In Fig. 3 we see an image showing an adenomatous polyp without image enhancement technology (a), example images using CVC (b,c,d), an image using CC (e) and images combining CC and CVC by using the i-Scan technology to visually enhance the already stained mucosa (f,g,h).
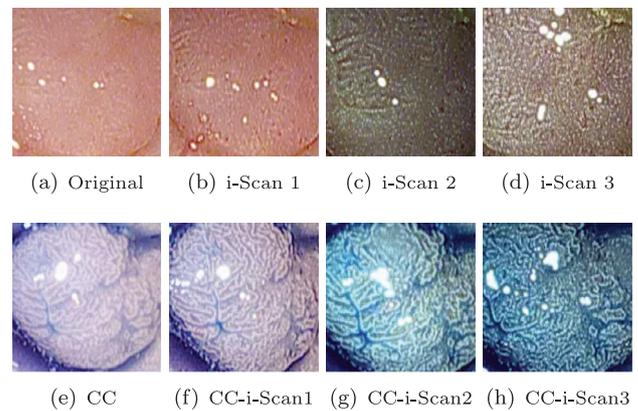

(a) Original (b) i-Scan 1 (c) i-Scan 2 (d) i-Scan 3

(e) CC (f) CC-i-Scan1 (g) CC-i-Scan2 (h) CC-i-Scan3

Fig. 3. Images of a polyp using digital (i-Scan) and/or conventional chromoendoscopy (CC).

(a) Non-neoplastic (b) Non-neoplastic (c) Neoplastic (d) Neoplastic
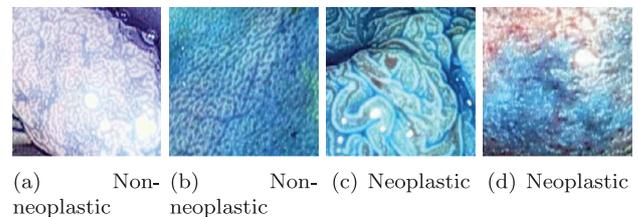
Fig. 4. Example images of the two classes obtained by a HD endoscope using a combination of CC and i-Scan mode 2.

In Fig. 4 we see exemplar images of the two classes (denoted as class "Non-neoplastic" and class "Neoplastic") obtained by a HD endoscope using a combination of CC and i-Scan mode 2.

In this work we will examine the effects of combinations of CVC and CC on the classification results.

2.2. High magnification endoscopy in combination with chromoendoscopy

High magnification endoscopes are defined by the ability to perform optical zoom by using a moveable lens in the tip of the endoscope. In that way magnified images are obtained without losing display quality. High magnification endoscopy enables the visualization of mucosal details that cannot be seen with standard endoscopy. The CC-high-magnification database is gathered using zoom-endoscopy in combination with chromoendoscopy. Example images of the classes can be seen in Fig. 2.

2.3. High-magnification endoscopy in combination with NBI

NBI (Gono et al., 2003) is a videoendoscopic system using RGB rotary filters placed in front of a white light source to narrow the bandwidth of the spectral transmittance. NBI enhances the visibility of microvessels and their fine structure on the colorectal surface. Also the pits are indirectly observable, since the microvessels between the pits are enhanced in black, while the pits are left in white.

In this work we use two NBI-high-magnification databases.

For one database, further denoted as the NBI-high-magnification database Aachen, image labels were provided according to their histological diagnosis (like for the previously presented databases). Exemplar images of the two classes of this database can be seen in Fig. 5.

For the second database, further denoted as the NBI-high-magnification database Hiroshima, image labels were provided according to the optical appearance of the polyps. The images were

Non-neoplastic   Neoplastic

Fig. 5. Examples images of the two classes from the NBI-high-magnification database Aachen.

Non-neoplastic   Neoplastic

Fig. 6. Examples images of the two classes from the NBI-high-magnification database Hiroshima.

labeled by at least two medical doctors and endoscopists who are experienced in colorectal cancer and NBI classification. Exemplar images of the two classes of this database can be seen in Fig. 6.

3. Wavelet and wavelet based feature extraction approaches

In this section we will describe the wavelet based transforms and the employed feature extraction approaches.

3.1. The 2-D discrete wavelet transform

The discrete wavelet transform (DWT) (Mallat, 1989) generates frequency bands by applying low-pass (h) and high-pass (g) filters to the input signal followed by a subsampling of the filter outputs with factor 2. To increase the frequency resolution, the decomposition is repeated by decomposing the outputs of the low pass filtering. (see Fig. 7).

This results in a binary tree with nodes representing a sub space with different time-frequency localization (see Fig. 7). This tree is known as filter bank. Starting with a mother wavelet ψ, the filters $\psi_{j,\,k}$ are shifted and scaled versions of the mother wavelet:

$$\psi_{j,k} = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t-k2^j}{2^j}\right), \tag{1}$$

where j is the scale (or decomposition level) and k is the shift parameter and both are integers.
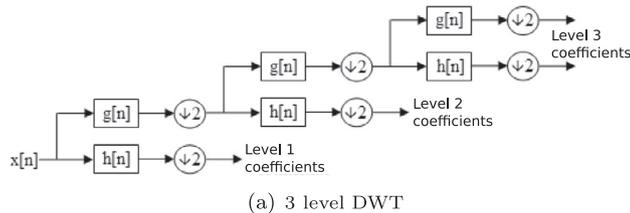
(a) 3 level DWT
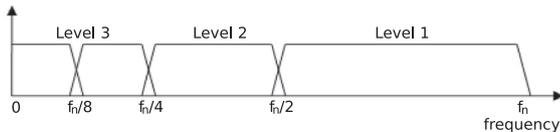
(b) Frequency domain representation of the DWT

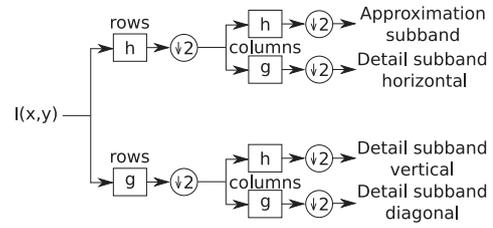**Fig. 7.** The DWT filter bank and its frequency domain representation.

Fig. 8. 1-level 2-D DWT and the resulting subbands.
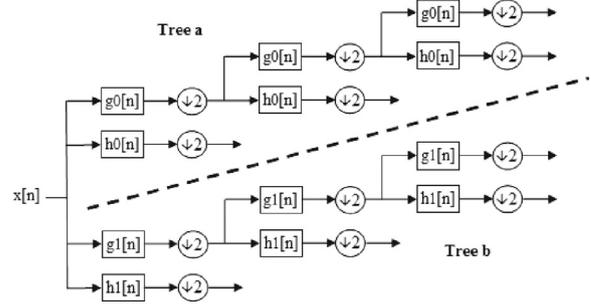

Fig. 9. 3 level DT-CWT filter bank.

Then the wavelet coefficient $\gamma_{j,\,k}$ of a signal $x(t)$ is computed as follows:

$$\gamma_{j,k} = \int_{-\infty}^{\infty} x(t)\psi_{j,k}\,dt. \tag{2}$$

Given an image, the 1-D filter bank is first applied to the rows of the image and then applied to the columns as can be seen in Fig. 8.

Like in Kwitt and Uhl (2010), we use the CDF 9/7 filters (Daubechies, 1992) for the DWT, which are biorthogonal wavelet filters. If not stated otherwise, the DWT and also the other employed wavelet based transforms are applied to RGB color images using 4 decomposition levels.

3.2. The DT-CWT

Kingsbury's dual-tree complex wavelet transform (Kingsbury, 1998) is designed to overcome two commonly known shortcomings of the 2-D DWT, the lack of shift-invariance and the poor directional selectivity. The key concept of the DT-CWT in 1-D is to use two separate DWT decompositions (see Fig. 9), where the low-pass filter of one tree is a half-sample delayed version of the low-pass filter of the other tree and the filters of one tree are the reverse of the filters of the other tree.

The outputs of one tree can be interpreted as the real parts and the outputs of the other tree can be interpreted as the imaginary parts of complex wavelet coefficients. The redundancy of 2^d (where d is the dimension of the signal being transformed) compared to the DWT provides extra information for analysis. The DT-CWT leads to a fixed number of 6 detail subbands per decomposition level in 2-D, capturing image details oriented at $\approx \pm 15°$, $\approx \pm 45°$ and $\approx \pm 75°$. In Fig. 10 we see the frequency tiling of a DT-CWT with two scales.

3.3. The Gabor wavelet transform

Gabor wavelets use complex functions constructed to serve as a basis for the Fourier transforms in information theory applications. The Gabor wavelet transform has multi-resolution as well as multi-orientation properties. Gabor wavelets minimize the product of its standard deviations in the time and frequency domain. In that way,
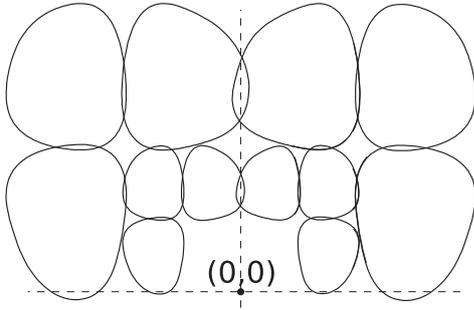
Fig. 10. Contours of 70% peak magnitude of DT-CWT filters in the frequency plane.
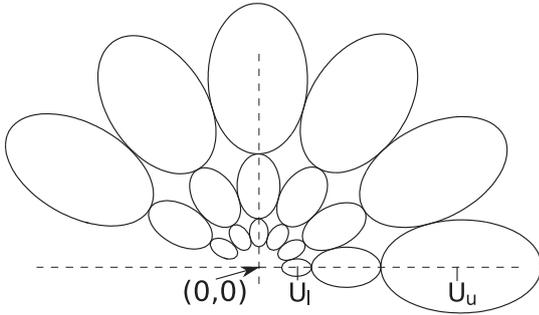

Fig. 11. Contours of half-peak magnitude of Gabor Wavelet filters in the frequency plane.

the uncertainty in information (frequency resolution vs time resolution) carried by this wavelet is minimized. It has been found that the simple cells of the visual cortex of mammalian brains are best modeled as a family of self-similar 2D Gabor wavelets (Lee, 1996).

A generic 2-D Gabor function (Manjunath and Ma, 1996) can be written as

$$g(x, y) = \left(\frac{1}{2\pi \sigma_x \sigma_y} \right) e^{\left(\frac{-1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi iWx \right)}, \qquad (3)$$

where σ_x and σ_y are the bandwidths of the filters and W is the central frequency. This function can be dilated and rotated to get a dictionary of filters.

The Gabor wavelet transform (GWT) is parametrized by the number of orientations and scales and the lower (U_l) and upper (U_u) center frequency of interest, which influences the calculation of the scaling factor for the mother wavelet. Redundancy is minimized by choosing the scaling factor and the bandwidth of the filters so that the half-peak magnitudes of the filter responses touch each other. The frequency tiling for a GWT using 6 orientations and 3 scales can be seen in Fig. 11.

Manjunath and Ma (1996) found that a choice of 4 scales and 6 orientations with center-frequency $(U_l, U_u) = (0.05, 0.4)$ (resulting in a scaling factor of $a = 2$) is optimal for their problem (texture analysis) and we chose the same parameter values.

3.4. The curvelet transform

The continuous curvelet transform (CCT) is based on tilling the 2D Fourier space in polar "wedges", with higher directional selectivity for higher frequency bands (see Fig. 12(a)).

The CCT (Candes et al., 2006) can be defined by a pair of windows $W(r)$ (the radial window) and $V(t)$ (the angular window). Both are smooth, nonnegative and real-valued and are defined as
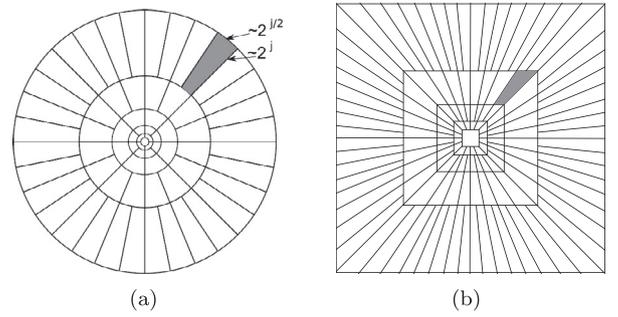

Fig. 12. The basic tiling of the frequency plane of the continuous (a) and discrete (b) curvelet transform.

scaled Meyer window functions (Daubechies, 1992):

$$W(r) = \begin{cases} \cos[\pi/2 \, \gamma \, (3r - 4)] & 4/3 \leq r \leq 5/3 \\ 1 & 5/6 \leq r \leq 4/3 \\ \cos[\pi/2 \, \gamma \, (5 - 6r)] & 2/3 \leq r \leq 5/6 \\ 0 & \text{else} \end{cases} \qquad (4)$$

$$V(t) = \begin{cases} \cos[\pi/2 \, \gamma \, (3|t| - 1)] & 1/3 \leq |t| \leq 2/3 \\ 1 & |t| \leq 1/3 \\ 0 & \text{else} \end{cases} \qquad (5)$$

where γ is a smooth function satisfying:

$$\gamma(x) = \begin{cases} 0 & 0 \leq x \\ 1 & x \geq 1 \end{cases}, \quad \gamma(x) + \gamma(1 - x) = 1 \quad x \in \mathbb{R} \qquad (6)$$

The frequency window U_j is defined in the Fourier domain by

$$U_j(r, \phi) = 2^{-3j/4} W(2^{-j}r) V\left(\frac{2^{\lfloor j/2 \rfloor} \phi}{2\pi} \right), \qquad (7)$$

where the support of U_j is a polar "wedge" defined by the support of W and V applied with scale dependent window width. The frequency window U_j corresponds to the Fourier transform of a curvelet φ_j, which can be thought of as a "mother" curvelet in the sense that the $2^{\lfloor j/2 \rfloor}$ curvelets at scale 2^{-j} are obtained by rotations and translations of φ_j.

Contrary to the DT-CWT and the GWT, which only cover part of the frequency spectrum in the frequency domain, curvelets have a complete cover of the spectrum in the frequency domain.

In Candes et al. (2006), two second generation discrete curvelet transforms (DCT's) are proposed, the DCT via unequispaced FFTs (fast fourier transforms) and the DCT via wrapping. We chose the wrapping based algorithm because it is the more often used algorithm for feature extraction purposes. This algorithm is implemented in the tool CurveLab (available at http://www.curvelet.org/). The DCT via wrapping uses a spatial grid to translate curvelets at each scale and angle using 2-D FFT, with the assumption that "Cartesian" curvelets are defined in a regular rectangular grid (see Fig. 12(b)). Then for each scale $s = 2^{-j}$ and orientation n, the product of U_j (the curvelet in FT domain) and the image in FT domain is obtained. Finally the product is wrapped around the origin and the 2-D inverse FFT is applied to the wrapped product, resulting in the curvelet coefficients at scale s and orientation n. The frequency tiling of the DCT can be seen in Fig. 12(b).

It should be noted that in case of the DCT there is a different denotation of the scale levels compared to the wavelet transforms. Scale level 1 of the DCT denotes the coarsest scale level and consists of only one undirectional subband that can be considered as the approximation subband or as the lowpass subband. Scale levels 2 till L include the directional subbands and can be considered as detail subbands. The higher the scale level, the finer the scale of
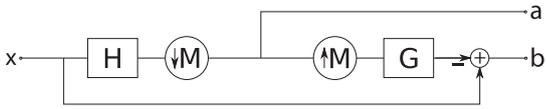
Fig. 13. Laplacian pyramid decomposition. The outputs are a coarse approximation *a* and a difference *b* between the original image and the prediction.
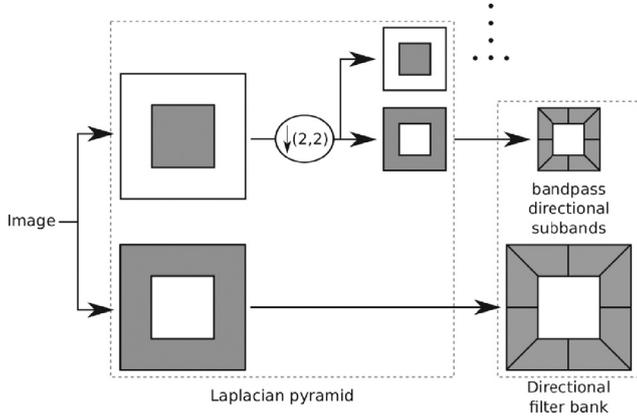

Fig. 14. The contourlet filter bank: first the multiscale decomposition into octave bands by means of the Laplacian pyramid decompositions followed by the application of a directional filter bank to the bandpass channels.

the subbands (the higher the frequency content in the subbands), which is the exact opposite of the wavelet transforms. That means comparing a 5 level DCT and a 4-level wavelet transform (WT), the level 1 subband of the DCT can be considered as the approximation subband of the wavelet transform and the DCT subbands of the levels {2,3,4,5} can be considered as the level {4,3,2,1} subbands of the WT (with respect to the frequency partition).

If not stated otherwise, we employ curvelets using the four DCT subband levels {2,3,4,5} with 16 level 2 subbands, 32 level 3 subbands, 32 level 4 subbands and 64 level 5 subbands.

3.5. The contourlet transform

In an attempt to provide a better discrete implementation of the curvelets, the contourlet representation has been proposed by Do and Vetterli (2005). The contourlet transform is designed to achieve essentially the same frequency tiling as the curvelet transform, however contourlets allow a different (selectable) number of directions at each scale and are not a discretization of the curvelets.

The multiscale decomposition of the contourlet transform (CT) is obtained using the Laplacian pyramid (LP) decomposition (Burt and Adelson, 1983). The LP decomposition at each level generates a downsampled low-pass version of the image and the difference between the image and the prediction, resulting in a bandpass image. In Fig. 13 we depict the LP decomposition, where *H* is the lowpass filter, *G* the synthesis filter and M the sampling matrix.

At each level, a directional filter bank (DFB) is applied to the bandpass image (b) that leads to a decomposition of 2^l ($l \in \mathbb{N}$) subbands with wedge-shaped frequency partitioning as shown in Fig. 14.

The DFB (Do, 2001) is constructed from two building blocks. The first one is a two-channel quincunx filter bank (Vetterli, 1984) that divides a 2-D spectrum into 2 directions: horizontal and vertical. The second building block of the DFB is a shearing operator, which amounts to just reordering of image samples.

We use the CT implementation described in Do and Vetterli (2005), which is public available (http://www.mathworks.com/matlabcentral/ fileexchange/8837-contourlet-toolbox). In all of our
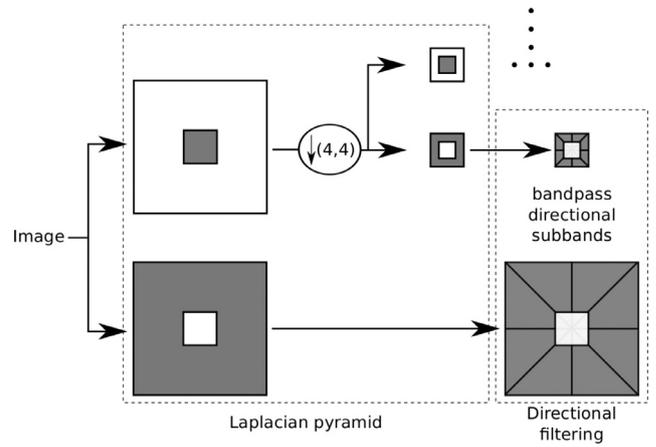
employed contourlet based approaches, the CDF 9/7 filters are used for the CT decomposition. If not stated otherwise, we employ the CT using four decomposition levels with 8 orientations per level.

3.6. The shearlet transform

The continuous shearlet transform (Easley et al., 2008) is based on parabolic scaling matrices A_a to change the resolution and on shear matrices B_s to change the orientation:

$$A_j = \begin{pmatrix} a & 0 \\ 0 & a^{1/2} \end{pmatrix}, \quad B_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, \tag{8}$$

with $a > 0$ and $s \in \mathbb{R}$.

The shearlets are given by

$$\psi_{a,s,k}(x) = a^{3/4} \psi(B_s A_a x - k), \tag{9}$$

where $k \in \mathbb{R}^2$ is the translation. The continuous shearlet transform is defined as the mapping for $f \in \mathbb{R}$:

$$SH_\Psi f(a, s, k) = \langle f, \psi_{a,s,k} \rangle. \tag{10}$$

The discrete shearlet transform (Easley et al., 2008) can be viewed as a simplifying theoretical justification of the contourlet transform. The shearlet transform offers more flexibility than the contourlet and curvelet transform (the directions per scale and the local support of the shearing filters are selectable). The first step of the discrete shearlet transform (ST) is to accomplish a multiscale partition using the Laplacian pyramid decomposition similar to the contourlet transform. Then the 2-D FFT is applied to the resulting highpass images. The samples in the frequency domain are taken not on a Cartesian grid, but along lines across the origin at various slopes, known as pseudo-polar grid. In order to obtain the directional localization, a band-pass filtering is applied using a frequency window function *W*, which is localized on a pair of trapezoids and constructed from the shearing filters using Meyer wavelets and which is also transformed to the frequency domain and taken on the pseudo-polar grid (for a more detailed description see Easley et al. (2008)). The final step is to re-assemble the Cartesian sampled values and apply the inverse 2-D FFT. The ST-scheme is showed in Fig. 15.

The ST offers large flexibility in the choice of the frequency window and allows to choose an arbitrary number of directional subbands per decomposition level to adapt the transform to specific applications. In case of the DCT, only the number of the directional subbands of the second coarsest level (the coarsest level is the low-pass subband) can be chosen, and even this number has to be a multiple of 4. All other decomposition levels of the DCT have a fixed number of directional subbands depending on the number

Fig. 15. ST decomposition using the Laplacian pyramid and directional filtering.

of the directional subbands in the second coarsest level. In case of the CT, the number of subbands can be chosen free, but the numbers have to be dyadic.

We use the 2D Shearlet Toolbox software (http://www.math.uh.edu/~dlabate/software.html), which is described in Easley et al. (2008). It should be noted that contrary to the ST scheme shown in Fig. 15, we use the nonsubsampled shearlet transform using the nonsubsampled Laplacian pyramid transform. Since already the normal ST is highly redundant because of the missing anisotropic subsampling, the nonsubsampled shearlet transform is even more redundant. The ST subbands are all of the same size as the input image (like for the GWT).

If not stated otherwise, we employ the ST using four decomposition levels with 8 orientations per level and shearing filters with a support size of 32×32.

3.7. Preprocessing

In this work we apply three preprocessing steps before each of the wavelet based methods.

The first preprocessing step removes specular reflections, which often occur in endoscopic images and have a major impact to the resulting subband coefficients in affected areas. Reflections are detected by thresholding the Saturation and grayscale values of an image. Similar to Stehle et al. (2009), a pixel is identified as part of a specular reflection if its gray value is greater as 235 and its Saturation is smaller than 0.09. As this kind of segmentation usually tends to a under-segmentation, a morphological dilation using a disc of radius $r = 5$ as the structuring element is applied to enlarge the segmented area. The segmented area is set to the average RGB color values of the adjacent pixels of the segmented area. To avoid sharp transitions between the segmented area and the surrounding area, the pixels surrounding the segmented area with less than 4 pixels distance to the segmented area are Gaussian blurred using $\sigma = 2$ and a 10×10 mask.

The most important preprocessing step is contrast-limited adaptive histogram equalization (CLAHE) Zuiderveld (1994). CLAHE is used to enhance the contrast and remove noise and intensity inhomogeneities. CLAHE is applied using 8×8 tiles and an uniform distribution for constructing the contrast transfer function. CLAHE operates on small regions in the image, called tiles, rather than the entire image. Each tile's contrast is enhanced, so that the histogram of the output image is approximately flat. CLAHE distinctly enhances the results of the wavelet based methods.

In the final preprocessing step, the images are (slightly) Gaussian blurred using $\sigma = 0.5$ and a 3×3 mask. This step is done to slightly smooth the images and to remove noise.

If not stated otherwise, the employed wavelet based methods are preprocessed using the previously described three steps, even if the methods are originally proposed using different or no preprocessing methods. For each method, the achieved classification rates increase using our preprocessing approach compared to the originally proposed preprocessing approaches or no preprocessing at all.

3.8. Feature extraction of the wavelet based methods

For each of the wavelet based transforms (DWT, DT-CWT, Gabor wavelets, curvelet, contourlet and shearlet transform), the distribution of the subband coefficients is once modeled by means of the Gaussian distribution, the GGD and the Weibull distribution (Evans and Peacock, 2000; Kwitt and Uhl, 2007). The Gaussian distribution and the Weibull distribution are used to model the subband coefficient magnitudes, whereas the GGD is used to model the original subband coefficients in case of the employed wavelet transforms producing real valued coefficients (DWT, contourlets and shearlets) and the subband coefficient magnitudes in case of the employed wavelet transforms producing complex valued subband coefficients (DT-CWT, Gabor wavelets and curvelets).

We chose the Gaussian distribution because extracting mean and standard deviation (the two parameters of the Gaussian distribution) of subband coefficients is probably the most known and most used approach to extract features of wavelet based transforms. The GGD is also a widely used feature to extract information from subbands of wavelet based transforms and it is able to model the subband coefficient distributions more accurate than the Gaussian distribution. In fact, the subbands of various types of wavelet transforms (with real valued subband coefficients) are well modeled using the GGD (Do and Vetterli, 2002). The Weibull distribution has been chosen because it already has been successfully used for the classification of polyps in combination with the DT-CWT and because it is able to accurately model the subband coefficient distribution of all employed wavelet based transforms, contrary to the Gaussian distribution and the GGD (see Section 5.1).

The probability distribution of the GGD (Do and Vetterli, 2002) is defined as

$$p(x; \mu, \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^\beta}, \tag{11}$$

where $\Gamma(.)$ denotes the gamma function, μ is the mean, α the scale parameter and β the shape parameter. Only the parameters α and β are extracted as features from the subbands for further classification. Distances between GGD feature vectors are measured using the Kullback–Leibler distance (Do and Vetterli, 2002), which is in case of the GGD defined as

$$D(p(.; \alpha_1, \beta_1), ||p(.; \alpha_2, \beta_2))$$
$$= log\left(\frac{\beta_1\alpha_2\Gamma(1/\beta_2)}{\beta_2\alpha_1\Gamma(1/\beta_1)}\right) - \frac{1}{\beta_1} + \frac{\alpha_1}{\alpha_2}\frac{\beta_2\Gamma((\beta_2+1)/\beta_1)}{\gamma(1/\beta_1)}. \tag{12}$$

The probability density function of a Weibull distribution with shape parameter c and scale parameter b is given by

$$p(x; c, b) = \begin{cases} \frac{c}{b}\left(\frac{x}{b}\right)^{c-1} e^{-\left(\frac{x}{b}\right)^c} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0, \end{cases} \tag{13}$$

where $b > 0$ and $c > 0$. The two parameters of the Weibull distribution are estimated using the method of moments (Niola et al., 2006).

In case of the Gaussian and Weibull distribution, the resulting feature vectors are L^2-normalized and distances between the feature vectors are measured using the Euclidean distance. The L^2-normalization is important to balance the different ranges of coefficient values per decomposition level. All our employed wavelet based transforms have in common that the coefficients in the subbands representing the coarser image details are much higher than the coefficients in the subbands representing the finer image details. Given our d-dimensional samples $v_1, \ldots, v_n$, the normalization formula for the mth feature of the jth feature vector is defined by

$$\tilde{v}_j(m) = \frac{v_j(m) - \bar{v}(m)}{\bar{s}(m)}, \tag{14}$$

where $\bar{v}(m)$ and $\bar{s}(m)$ denote the sample mean and the sample variance of the mth features of the n feature vectors. 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 distance metric.

The wavelet based transforms are applied using four decomposition levels (scales) for the extraction of Weibull, GGD and Gaussian features from the subbands. If not stated otherwise, all employed wavelet and wavelet based approaches are applied to RGB

color images and the final feature vector consists of the concatenation of the features of the three color channels. So the length of a feature vector $l(fv)$ resulting from extracting Gaussian, GGD or Weibull parameters from the subbands of a wavelet based transform is given by

$$l(fv) = 3 * 2 * NrSB, \tag{15}$$

where 3 is the number of color channels, 2 is the number of parameters extracted by the probability distributions and $NrSB$ is the number of subbands of the used wavelet based transform.

Since we primarily focus on Lets in this work, we additionally reimplemented 2 texture analysis approaches based on curvelets (Gomez and Romero, 2011; Barbosa et al., 2009), 2 based on contourlets (Long and Younan, 2006; Dong and Ma, 2013) and 3 based on shearlets (Schwartz et al., 2011; He et al., 2013; Dong et al., 2015), which extract features that are different to the three previously described statistical features (Gaussian distribution, GGD and Weibull distribution). These approaches were published in well known journals or conferences.

3.8.1. Rotation invariant DCT using the dominant orientation

In this approach (Gomez and Romero, 2011), the same features are extracted as in the approach extracting GGD features using the DCT (DCT-GGD). Also the same distance metric is used. To achieve rotation invariance, the features of each scale level are circular shifted, using the dominant orientation as reference. The dominant orientation is defined as the orientation whose associated directional (second level) subband has the highest sum of absolute valued coefficients (the highest energy). Contrary to the original approach, we apply the same DCT decomposition as for the DCT-GGD approach. This has the advantage of a better comparability to the DCT-GGD approach, which is basically the same approach as the considered one but without cyclic shifted features to achieve rotation invariance. We further denote this approach as DCT-DO.

3.8.2. DCT color covariance features

This approach (Barbosa et al., 2009) firstly extracts the means and standard deviations (std) of the DCT subbands. Then the color covariance of these features can be calculated as follows:

$$CC(a, b, s, m) = \sum_{\alpha} (F_m(a, s, \alpha) - E\{F_m(a, s, \alpha)\})$$
$$\times (F_m(b, s, \alpha) - E\{F_m(b, s, \alpha)\}), \tag{16}$$

where a and b represent two different color channels, F_m the statistical texture descriptor ($m = 1$: mean, $m = 2$: std), α is the considered angle of the DCT subband, s the considered DCT level and $E\{F_m(a, s, \alpha)\}$ the average of the statistical texture descriptor F_m over the different directions in the color channel a.

It should be noted that in this approach the HSV color space is used instead of the RGB color space like for all other approaches and that we did not apply CLAHE as preprocessing step (CLAHE cannot be applied to all HSV color channels). The DCT decomposition results in a lowpass subband and two levels of directional subbands with 8 and 16 orientations (we only use the directional subbands) resulting in a feature vector of an image of length 24 (6 combinations of color channels × 2 scale levels × 2 parameters). We further denote this approach as DCT color covariance feature (DCT-CCF).

3.8.3. CT coefficients modeled by histograms

In this approach (Long and Younan, 2006), the CT subband coefficients are modeled using histograms with 10 bins. Distances between two feature vectors are measured using the χ^2 distance metric:

$$\chi^2(x, y) = \sum_i \frac{(x_i - y_i)^2}{x_i + y_i}. \tag{17}$$

The CT is applied using 3 decomposition levels with 8 directional subbands per level resulting in 24 directional subbands and the low-pass subband. The final feature vector of an image has length 750 (3 color channels × 25 subbands × 10 bins per histogram). We further denote this approach as CT-Histogram.

3.8.4. CT subband clustering

In this approach (Dong and Ma, 2013) the k-means clustering algorithm is used to find 3 cluster centers of the CT subband coefficients, which are used as features for further classification. First, the CT decomposes an image into $L = 4$ levels with 8 directional subbands per level and the low-pass subband. With increasing decomposition level i (from fine ($i = 1$) to coarse ($i = 4$)), the average amplitude of the CT coefficients increases almost exponentially. The by far highest coefficient values are in the low-pass subband. To balance the different ranges of coefficient values per decomposition level, the low-pass subband coefficients are multiplied by the factor of $1/4^L$ and the detail subband coefficients of decomposition level i ($i \in \{1, 2, \ldots, L\}$) are multiplied by the factor of $1/4^{i-1}$ (in the publication describing the approach (Dong and Ma, 2013) the authors wrote $1/4^i$, but we think this is a typo since this factor would not consider the far higher coefficient values in the low-pass subband). Additionally, the variance and norm-2 energy of each subband is extracted, resulting in a feature vector's length of 990 (3 color levels × 33 subbands × 5 parameters (3 clusters, variance and energy)). We further denote this approach as CT-Cluster.

3.8.5. Energy of the ST coefficients

In this approach (Schwartz et al., 2011), the energy of the subband coefficients is used as feature:

$$E(s) = \sum |s(x)|, \tag{18}$$

where s denotes a subband. The energy feature is computed from the subbands of a 4 level ST decomposition with 8 directional subbands per level. The resulting feature vectors are L2-normalized and have length 81 (3 color channels × 4 levels × 8 orientations per level × 1 parameter). This approach will be further denoted as the ST-Energy approach.

3.8.6. ST combined with a LBP based feature extraction

In this approach (He et al., 2013), a feature based on local binary patterns (LBP) (Ojala et al., 2002) is extracted from the subband coefficients of the ST decomposition. First, two local features are computed as follows:

$$e_{i,j}^{l,d} = \frac{1}{9} \sum_{p=-1}^{1} \sum_{q=-1}^{1} |s_{i+p,j+q}^{l,d}| \tag{19}$$

$$g_{i,j}^{l,d} = \frac{-1}{\log(9)} \sum_{p=-1}^{1} \sum_{q=-1}^{1} \frac{|s_{i+p,j+q}^{l,d}|}{norm_{i,j}^{l,d}} \log \left(\frac{|s_{i+p,j+q}^{l,d}|}{norm_{i,j}^{l,d}} \right) \tag{20}$$

where $s_{i,j}$ is the shearlet coefficient at (i, j) in the d'th directional subband within the l'th decomposition level and

$$norm_{i,j}^{l,d} = \sum_{p=-1}^{1} \sum_{q=-1}^{1} |s_{i+p,j+q}^{l,d}|. \tag{21}$$

Then these features are normalized and by means of thresholds t_n^l ($n \in \{0, 1, 2\}$) with $t_1^l < t_2^l < t_2^l$, an integer value m between 0 and 3 is assigned to each local feature $e_{i,j}^{l,d}$ ($g_{i,j}^{l,d}$ analogous with different

threshold values) in each decomposition level l:

$$m_{i,j}^{l,d} == \begin{cases} 0 \text{ for } e_{i,j}^{l,d} < t_0^l \\ 1 \text{ for } t_0^l < e_{i,j}^{l,d} < t_1^l \\ 2 \text{ for } t_1^l < e_{i,j}^{l,d} < t_2^l \\ 3 \text{ for } e_{i,j}^{l,d} > t_2^l \end{cases} \tag{22}$$

The local shearlet-based energy pattern (LSEP) is defined as

$$LSEP_{i,j}^l \sum_{d=1}^{D} m_{i,j}^{l,d} 3^{d-1}. \tag{23}$$

To achieve orientation invariance $m_{i,j}^{l,d}$ is sorted before the LSEP computation, so that the values of a given level l and position (i, j) are ascending in the orientation dimension ($D = 4$ directions are used and so e.g. $m_{i,j}^l = 2, 1, 2, 0$ becomes $m_{i,j}^l = 0, 1, 2, 2$).

The support size of the shearlet filters is 16×16. The final step is to build histograms of the LSEP's and to concatenate these histograms into a feature vector. The final feature vector of an image consists of 270 elements (3 color channels $\times$ 3 scales $\times$ 2 local features $\times$ 15 bins per histogram). Distances between feature vectors are measured using the χ^2 distance metric. We further denote this approach as ST-LSEP.

3.8.7. Linear regression of ST subbands

In this approach (Dong et al., 2015), regression is used as a tool to investigate the dependences between shearlet subbands at different scale levels.

By applying the $L = 3$ level shearlet transform using shearlet filters with support size 30×30, we obtain one low-pass subband and $D = 10$ directional subbands at each scale. From each subband the norm-1 (mean) and norm-2 energy is computed. Such a subband feature at scale i and direction d from an image n ($n \in \{1, \ldots, N\}$) of class c is further denoted as $q_{c,n}^{i,d}$. Then the samples $\{(q_{c,n}^{i-1,d}, q_{c,n}^{i,d})\}_{n=1}^N$ (further denoted as $\{(x_n, y_n)\}_{n=1}^N$) can be seen as the N observations of (X, Y). The following linear regression models the dependences between the shearlet subband features at neighboring scale levels:

$$E(Y|X = x) = \beta_0 + \beta_1 x \tag{24}$$

Using the training images, the estimates $\hat{\beta}_0^{c,i,d}$ and $\hat{\beta}_1^{c,i,d}$ are computed for each class c, where d denotes the direction and i the scale level.

Given a test image and a pair of extracted features ($q^{i-1,d}, q^{i,d}$), the residual $d_c i, d$ is computed as follows:

$$d_c^{i,d} = |q^{i,d} - \hat{q}^{i,d}| \tag{25}$$

where

$$\hat{q}^{i,d} = E(Y|X = q^{i-1,d}) = \hat{\beta}_0^{c,i,d} + \hat{\beta}_1^{c,i,d} q^{i-1,d} \tag{26}$$

The distance from the test image I to the cth class T_c is defined as the weighted summation of residuals (WSR):

$$D_{WSR}(I, T_c) = \sum_{d=1}^{D} \sum_{i=1}^{L} 2^i d_{c,norm1}^{i,d} + \sum_{d=1}^{D} \sum_{i=1}^{L} 2^i d_{c,norm2}^{i,d}, \tag{27}$$

where $d_{c, norm1}$ ($d_{c, norm2}$) is the residual using norm-1 (norm-2) energy as subband feature. The test image I is assigned to the class corresponding to the minimum of $\{D_{WSR}(I, T_c)\}_{c=1}^C$.

So contrary to the other methods, there is no feature vector as output of a image. The output of an evaluation set image is the predicted class. This approach will be further denoted as the ST-Reg approach.

3.9. Other methods

In this sections we will describe a variety of state of the art methods for colonic polyp classification which are not based on wavelets or Lets. By means of these methods we are able to compare the results of the wavelet based approaches with the results of state-of-the-art methods.

3.9.1. Blob-adapted local fractal dimension

This feature extraction method (Häfner et al., 2014c) is derived from the local fractal dimension (LFD) (Varma and Garg, 2007; Xu et al., 2009). For a given pixel location $x = (x_1, x_2)$, the local fractal dimension LFD(x) analyzes the changes of the intensity distribution of differently sized circle shaped regions of the image centered at the point x. This is usually done by filtering the image I with circle shaped binary filters with $r = 1, 2, 3, \ldots, 8$ and the LFD is computed for each pixel location by estimating the slope of the filter responses with increasing radii.

Contrary to the original LFD approach, the considered approach (Häfner et al., 2014c) enhances the viewpoint invariance using elliptic shaped binary and Gaussian filters, whose shape, size and orientation is adapted to the local texture structure. The final feature vector consists of the histograms of the LFD's.

3.9.2. Blob shape and contrast features

This approach (Häfner et al., 2014a) consists of two steps. The first step is a segmentation algorithm, that applies local region growing to the maxima and minima of the image in a similar way as the watershed segmentation by immersion (Roerdink and Meijster, 2000). The resulting blobs represent the local texture structures of an image.

In the second step, 3 shape features and a contrast feature are extracted from the blobs. The final feature vector consists of the histograms of these 4 features.

3.9.3. Dense SIFT features

This approach (Tamaki et al., 2013) combines densely computed SIFT features with the bag-of-visual-words (BoW) approach. The SIFT descriptors are sampled at points on a regular grid. From these SIFT descriptors, cluster centers (visual words) are learned by means of k-means clustering. Given an image, its corresponding model is generated by labeling its SIFT descriptors with the texton that lies closest to it. We use the same parameters that led to the best results in Tamaki et al. (2013) (grid spacing = 5, SIFT scale 5 and 7), but with a lower number of visual words (only 600 instead of up to over 10000 visual words in (Tamaki et al., 2013)). In our experiments, the lower number of visual words led to better results and less (but still huge) computational cost. In Tamaki et al. (2013), this approach is used for the colonic polyp classification in NBI endoscopy, however, there is no reason why this approach should not also be suited for other imaging modalities like the i-Scan technology or chromoendoscopy. The computation of the SIFT descriptors and the following k-means clustering is done using the Matlab software provided by the VLFeat open source library (Vedaldi and Fulkerson, 2008).

3.9.4. Vascularization features

This approach (Gross et al., 2012) segments the blood vessel structure on polyps by means of the phase symmetry (Kovesi, 1999). Vessel segmentation starts with the phase symmetry filter, whose output represents the vessel structure of polyps. By thresholding the output, a binary image is generated, and from this image 8 features are computed that represent the shape, size, contrast and the underlying color of the connected components (the segmented vessels). This method is especially designed to analyze the vessel structures of polyps in NBI images and is probably not

Table 1
Summary and characterization of the used feature extraction methods except for the Gaussian, GGD and Weibull features.

Method	Description
DCT-DO	Rotation invariant version of DCT-GGD
DCT-CCF	Color covariance feature applied on means and std's of the subbands
CT-Histogram	Histograms are built of subband coefficients
CT Cluster	Cluster centers of subband coefficients are used as features
ST-Energy	Extracts the energy of subbands
ST-LSEP	LBP based feature is extracted from the subbands
ST-Reg	Regression is used to investigate dependencies across different subband levels
BA-LFD	A viewpoint invariant feature analyzing changes in the intensity distribution
Blob-SC	Shape and contrast description of segmented blobs
SIFT	The BoW approach is applied to densely computed SIFT features
Vasc. F.	Blood vessel structure is segmented and described using 8 features
MB-LBP	Multiscale LBP variant

suited for imaging modalities that are not designed to highlighting the blood vessel structure. Hence, this method is most probably not suited for any other image processing task than endoscopic polyp classification using NBI.

We use the implementation of the phase symmetry filter (Kovesi, 2000) for the vascularization feature approach.

3.9.5. MB-LBP

Based on a grayscale image, the LBP operator generates a binary sequence for each pixel by thresholding the neighbors of the pixel by the center pixel value. The binary sequences are then treated as numbers (i.e. the LBP numbers). Once all LBP numbers for an image are computed, a histogram based on these numbers is generated and used as feature vector. There are several variations of the LBP operator and they are used for a variety of image processing tasks including endoscopic polyp detection and classification (e.g. Häfner et al. (2012b)). Because of its superior results compared to the standard LBP operator $LBP_{(8, 1)}$ (with block size = 3), we use a multiscale block binary patterns (MB-LBP) operator (Liao et al., 2007) with three different block sizes (3,9,15). The uniform LBP histograms of the 3 scales (block sizes) are concatenated resulting in a feature vector with $3 \times 59 = 177$ features per image.

3.10. Summary of the employed methods

For each type of the employed wavelet-based transforms we employ three approaches extracting three different types of statistical features (Gaussian GGD and Weibull features), which describe the subband coefficient distributions. The remaining methods are listed and characterized in Table 1.

4. Experimental results

In this paper we use a total of 11 different endoscopic databases to classify colonic polyps.

For a better comparability of the results and to put more emphasis to the feature extraction, all methods are evaluated using a k-NN classifier. To balance the problem of varying results depending on k, we average the 10 results of the k-NN classifier with $k = 1, \ldots, 10$.

Since we employ a high number of feature extraction approaches on many different databases, we decided to use the accuracy as the only performance measure and resigned to use other classification measures like e.g. sensitivity and specificity or precision and recall. The advantages of the accuracy is the easy comparability of the results (the accuracy is only one performance measure compared to the two performance measures for sensitivity and specificity or precision and recall).

In Fig. 16 we show a flowchart summarizing our experimental setup.

4.1. The CC-i-Scan database

The CC-i-Scan database is an endoscopic image database consisting of 8 sub-databases with 8 different imaging modalities. The 8 image sub-databases are acquired by extracting patches of size 256×256 from frames of HD-endoscopic (Pentax HiLINE HD+ 90i Colonoscope) videos either using the i-Scan technology or without any computed virtual chromoendoscopy (¬CVC in Table 2). The mucosa is either stained or not stained. The patches are extracted only from regions having histological findings. The CC-i-Scan database is provided by the St. Elisabeth Hospital in Vienna and was already used for colonic polyp classification e.g. in Häfner et al. (2014b); 2014c).

Table 2 lists the number of images and patients per class and database.

Classification accuracy is computed using Leave-one-patient-out (LOPO) cross validation. The advantage of LOPO compared to leave-one-out cross validation is the impossibility that the nearest neighbor of an image and the image itself come from the same patient. In this way we avoid over-fitting.

In Table 3 we see the overall classification rates (OCR) for our experiment using the CC-i-Scan database. The column ∅ shows for each method the averaged accuracies across all image enhancement modalities. The row ∅ shows the averaged accuracies across all wavelet-based methods. The highest results for each image enhancement modality across all methods are given in bold face numbers. In Fig. 17 we once again show the averaged accuracies across all image enhancement modalities (column ∅) for an easier comparison of the methods results.

As we can see in Table 3 and Fig. 17, extracting the Weibull parameters as features leads to the best results for each wavelet-based method. The two directional wavelet transforms DT-CWT and GWT extracting Weibull features are the best performing
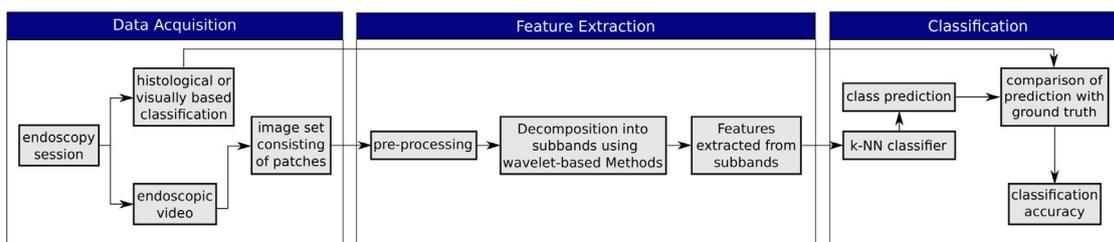

Fig. 16. Flowchart of the experimental setup.

Table 2
Number of images and patients per class of the CC-i-Scan databases gathered with and without CC (staining) and computed virtual chromoendoscopy (CVC).

i-Scan mode	No staining				Staining			
	¬CVC	i-Scan 1	i-Scan 2	i-Scan 3	¬CVC	i-Scan 1	i-Scan 2	i-Scan 3
Non − neoplastic								
Number of images	39	25	20	31	42	53	32	31
Number of patients	21	18	15	15	26	31	23	19
Neoplastic								
Number of images	73	75	69	71	68	73	62	54
Number of patients	55	56	55	55	52	55	52	47
Total nr. of images	112	100	89	102	110	126	94	85

Table 3
Accuracies of the methods for the CC-i-Scan databases in %. The highest results for each image enhancement modality are given in bold face numbers.

Methods	No staining				Staining				∅
	¬CVC	i-Scan1	i-Scan2	i-Scan3	¬CVC	i-Scan1	i-Scan2	i-Scan3	
DWT-Gaussian	74.0	82.4	84.2	81.2	63.2	65.7	69.3	67.2	73.4
DWT-GGD	75.0	82.8	84.0	81.3	67.2	70.0	80.1	68.1	76.1
DWT-Weibull	74.2	80.0	81.1	86.5	68.4	73.3	82.6	64.6	76.3
DTCWT-Gaussian	73.2	83.5	85.4	82.8	66.4	69.5	72.7	68.7	75.3
DTCWT-GGD	75.0	86.0	85.6	83.7	74.2	67.6	68.6	71.5	76.5
DTCWT-Weibull	79.6	86.4	84.8	**89.5**	**72.3**	77.0	82.6	67.1	79.9
GWT-Gaussian	75.8	82.1	85.4	80.6	67.5	73.2	74.2	66.0	75.6
GWT-GGD	79.3	82.9	83.4	82.2	75.0	69.5	74.4	72.5	77.4
GWT-Weibull	**83.5**	**88.0**	85.1	85.2	71.3	78.8	82.8	68.0	**80.3**
DCT-Gaussian	75.6	79.3	82.7	76.5	63.5	67.8	70.4	65.8	72.7
DCT-GGD	77.8	82.1	81.6	77.7	69.6	69.4	69.6	68.7	74.5
DCT-Weibull	80.0	80.2	82.6	81.6	65.1	71.2	77.1	66.0	75.5
DCT-DO	67.3	76.0	78.9	73.3	66.9	60.6	66.8	63.2	69.1
DCT-CCF	74.8	71.8	77.9	70.0	64.5	65.7	74.7	65.5	70.6
CT-Gaussian	73.8	81.9	83.7	81.1	68.6	68.5	70.4	68.0	74.5
CT-GGD	77.6	85.1	85.8	82.9	74.1	72.0	75.3	69.7	77.8
CT-Weibull	79.8	83.3	87.2	86.0	71.5	71.0	81.7	69.9	78.8
CT-Histogram	68.0	78.4	82.1	78.2	62.8	68.3	75.5	67.2	72.6
CT-Cluster	75.6	80.1	84.3	79.9	70.4	66.0	67.9	65.2	73.7
ST-Gaussian	72.8	83.2	82.6	80.1	63.1	69.7	72.5	66.4	73.8
ST-GGD	75.8	85.6	84.6	82.8	70.8	72.9	75.5	68.6	77.1
ST-Weibull	7.59	85.7	86.9	83.8	69.2	73.3	79.7	68.0	78.2
ST-Energy	72.4	82.4	83.6	79.8	63.5	70.3	72.8	67.3	74.0
ST-LSEP	71.6	77.7	84.2	77.8	65.4	69.4	81.8	66.8	74.3
ST-Reg	76.6	79.0	83.2	84.3	63.6	73.8	67.0	**75.3**	75.4
∅	75.5	81.8	83.6	81.1	67.9	70.2	74.6	67.8	75.3
BA-LFD	74.4	86.7	80.9	79.0	70.6	76.1	**84.6**	63.5	77.0
Blob SC	78.6	84.7	**87.4**	86.6	66.3	77.0	79.5	70.8	78.9
SIFT	75.4	86.9	84.0	78.8	70.5	**79.4**	77.0	65.3	77.2
Vasc. F.	63.7	72.6	76.0	72.5	58.2	48.5	62.9	59.5	64.2
MB-LBP	70.5	82.9	79.6	76.4	65.7	74.3	73.3	73.3	74.5

methods. Their results are even higher than those of the state-of-the-art approaches. Also ST and CT combined with Weibull features are among the best performing approaches. Only the Blob-SC approach achieves similarly high results as the 4 wavelet-based transforms combined with Weibull features. DWT and DCT combined with Weibull features only achieve average results.

The results of the wavelet-based methods extracting Gaussian parameters as features are quite similar to each other and average compared to the results of other methods. The GGD parameters provide constantly higher results than the Gaussian parameters and constantly lower results than the Weibull parameters.

As already mentioned before, DCT-GGD and DCT-DO are basically the same approaches with the exception that DCT-DO aims to achieve rotation invariance by cyclic shifting features. When we compare their results we see that the cyclic shifting process distinctly decreases the results.

Like expected, the Vascular Features are not suited for the classification of polyps for this database. Especially the results of the subdatabases with stained mucosa are particularly bad because

the pits of the mucosal structure, which are filled with dye, are wrongly recognized as vessels.

When we compare the results of the wavelet-based methods across the different imaging modalities (see row ∅ in Table 3), it becomes clear that staining the mucosa leads to a degradation of the results whereas the i-Scan modes improve the results. The same applies to the state-of-the-art methods.

By means of the McNemar test (McNemar, 1947), we assess the statistical significance of our results. With the McNemar test we analyze if the images from a database are classified differently by the various wavelet-based methods, or if most of the images are classified identical by the various wavelet-based methods (whereat we only differentiate between classifying an image as correct or incorrect). The McNemar test tests if the classification results of two methods are significantly different for a given level of significance (α) by building test statistics from incorrectly classified images. The test is carried out on the i-Scan 2 database without staining using significance level $\alpha = 0.05$. We chose this subdatabase because it provides the best results over all 8 subdatabases
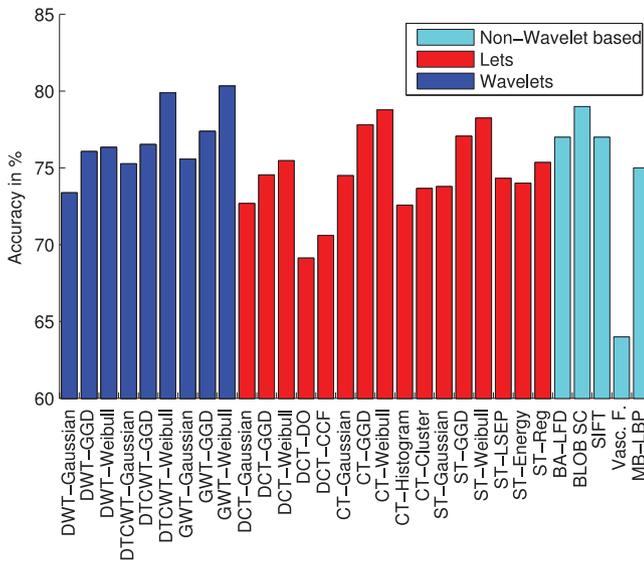
Fig. 17. Averaged accuracies across all image enhancement modalities of the CC-i-Scan database.
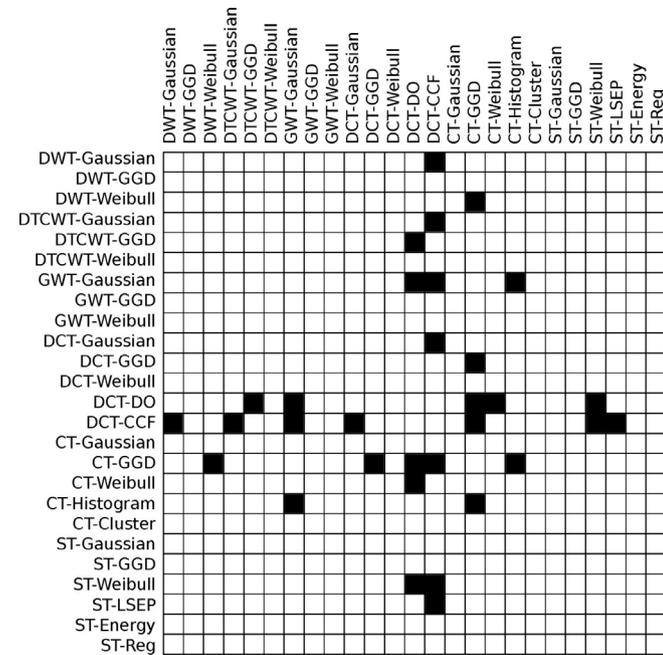

Fig. 18. Result of the McNemar test for the i-Scan 2 database without staining. A black square in the plot means that the two considered wavelet-based method are significantly different with significance level $\alpha = 0.05$. A white square means that there is no significant difference between the methods.

Table 4
Ground truth information based on histology for the CC-high-magnification image database.

	Non-neoplastic	Neoplastic	Total
Images	198	518	716
Patients	14	32	40
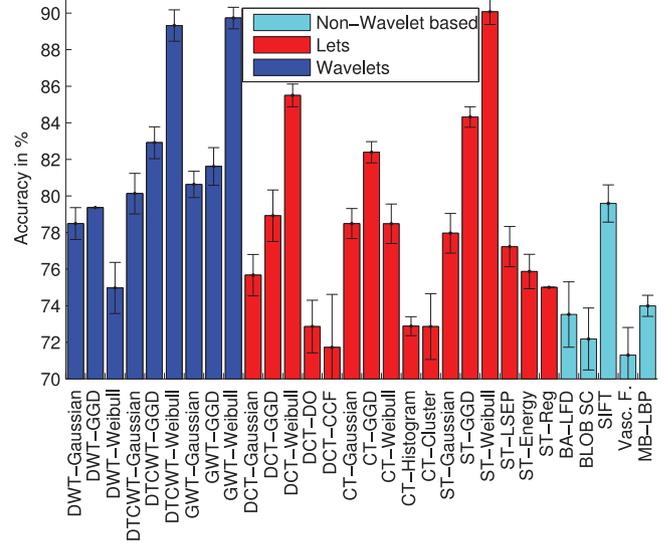
Fig. 19. Accuracies and the standard deviations of the methods for the CC-high-magnification database.

4.2. The CC-high-magnification database

The high-magnification images are acquired at the Department of Gastroenterology and Hepatology of the Medical University of Vienna using a zoom-colonoscope (Olympus Evis Exera CF-Q160ZI/L) with a magnification factor of 150 and indigocarmine dye-spraying. The database is acquired by extracting patches of size 256×256 from 327 endoscopic color images (either of size 624×533 pixels or 586×502 pixels) of 40 patients. Table 4 lists the number of images and patients per class.

Classification accuracy is computed using LOPO cross validation. In Fig. 19 we see the averaged accuracies (for $k = 1, \ldots, N$) of our employed methods using the CC-high-magnification image database.

The error bars in the figure indicate the standard deviations across the 10 k-values of the kNN-classifier. As we can see in Fig. 19, the wavelet-based methods extracting Weibull parameters as features achieve the best results. Especially DT-CWT, GWT and ST combined with Weibull features outperform the other methods, but also DCT combined with Weibull features performs very well. The state-of-the-art-approaches provide poor results compared to the wavelet-based methods.

Comparing the results using Gaussian, GGD and Weibull features across the different wavelet-based approaches, the DT-CWT, GWT and ST provide the best results. The results of the DCT and CT are already distinctly lower and the DWT provides the worst results.

Like in the previous database, extracting GGD features provides constantly better results than extracting Gaussian features and mostly worse results than extracting Weibull features. Similar to the previous database, the cyclic shifting of the features of the DCT-DO approach decreases the results compared to the DCT-GGD approach without cyclic shifting.
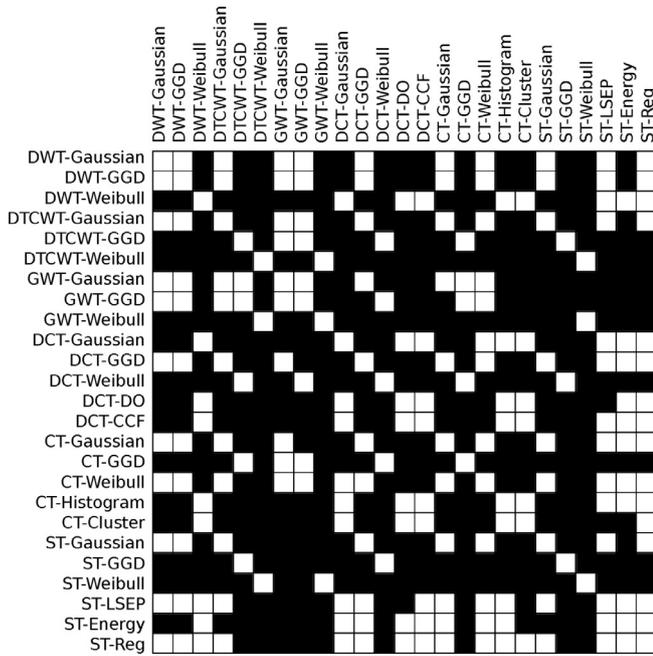
with different imaging modalities. Results are displayed in Fig. 18.

As we can see in Fig. 18, the DCT-CCF, the DCT-DO and the CT-GGD approach are classifying images significantly different to some other approaches. For $\alpha = 0.01$ there remains only one significant difference between the DCT-DO and the CT-GGD approach. However, the outcomes of the McNemar test (and any other significance test) are highly dependent on the sample size (the higher the sample size, the more likely significant differences will occur using the McNemar test). So the low number of significant differences is probably mainly caused by the low number of images of the i-Scan 2 database (89), since there are distinct differences in the results in Table 3.

Fig. 20. Result of the McNemar test for the CC-high-magnification database. A black square in the plot means that the two considered wavelet-based method are significantly different with significance level $\alpha = 0.01$. A white square means that there is no significant difference between the methods.

Table 5
Ground truth information based on histology for the NBI image database Aachen.

	Non-neoplastic	Neoplastic	Total
Images	173	214	387
Patients	98	135	211

In Fig. 20 we see the result of the McNemar test for the CC-high-magnification database using significance level $\alpha = 0.01$. As we can see in Fig. 20, there are much more significant differences between the employed approaches than at the i-Scan 2 database. This of course is partly mainly caused by the far higher number of images of the CC-high-magnification database (716) than of the i-Scan 2 database (89).

The DT-CWT , GWT and ST approaches using Weibull features are significant better than the remaining approaches.

4.3. NBI-high-magnification database Aachen

The NBI high-magnification images are acquired at the University Hospital Aachen using a NBI zoom endoscope, which can magnify the image to a maximum of 150-fold (CF-Q160ZI, Olympus Medical Systems Europe). The database is acquired by extracting patches of size 256×256 from 387 endoscopic color images of size 502×586 from 211 patients, where only one patch is extracted per image. Table 5 lists the number of images and patients per class.

In Fig. 21 we see the averaged accuracies and the standard deviations of our employed methods for the NBI-high-magnification database Aachen. We can see that the results of most of the methods are quite similar and that only the ST-LSEP approach achieves a result over 80%. Contrary to the previous databases, the wavelet-based transforms combined with Weibull features only achieve average results. The DCT-CCF approach achieves the distinctly worst result. The results for the GGD features are constantly better than these of the Gaussian and Weibull features.
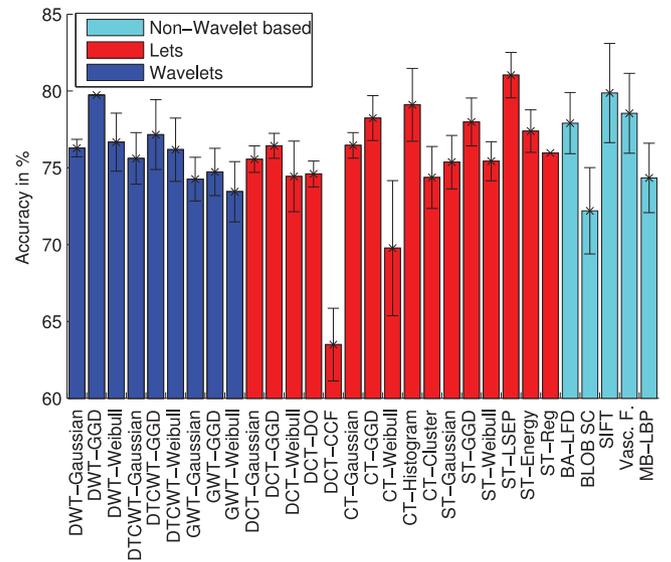

Fig. 21. Accuracies and the standard deviations of the methods for the NBI-high-magnification database Aachen.
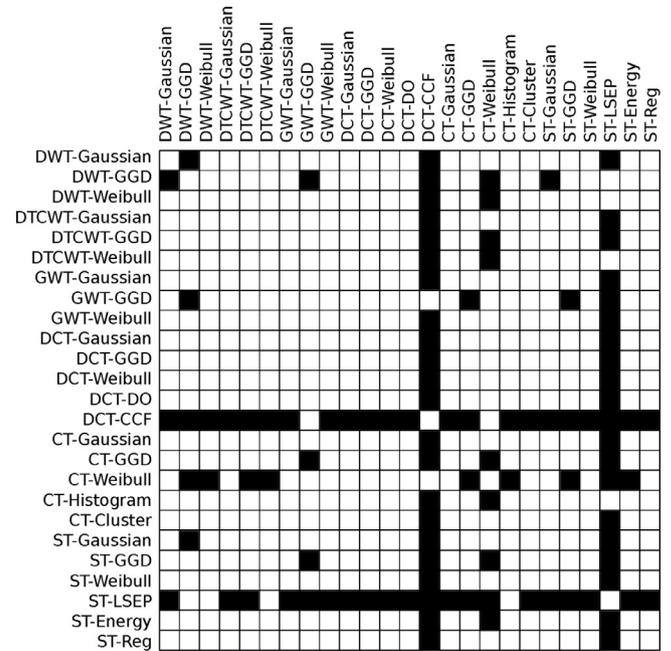

Fig. 22. Result of the McNemar test of the NBI-high-magnification database Aachen. A black square in the plot means that the two considered wavelet-based method are significantly different with significance level $\alpha = 0.01$. A white square means that there is no significant difference between the methods.

This results are also reflected in the results of the MC-Nemar test shown in Fig. 22. The DCT-CCF approach is significantly worse than most of the other approaches and the ST-LSEP approach is significantly better than most of the other approaches.

4.4. NBI-high-magnification database Hiroshima

The NBI high-magnification images are acquired at the Department of Endoscopy of the Hiroshima University Hospital using a NBI zoom endoscope, which can magnify the image to a maximum of 75-fold (CF-Q160ZI, Olympus Medical Systems Europe). Then the images were digitalized into 1440×1080 pixels. Care was taken that the lighting conditions, zooming and optical magnification were kept as similar as possible across different images.

Table 6
Ground truth information based on optical appearance of the polyps for the NBI-high-magnification database Hiroshima.
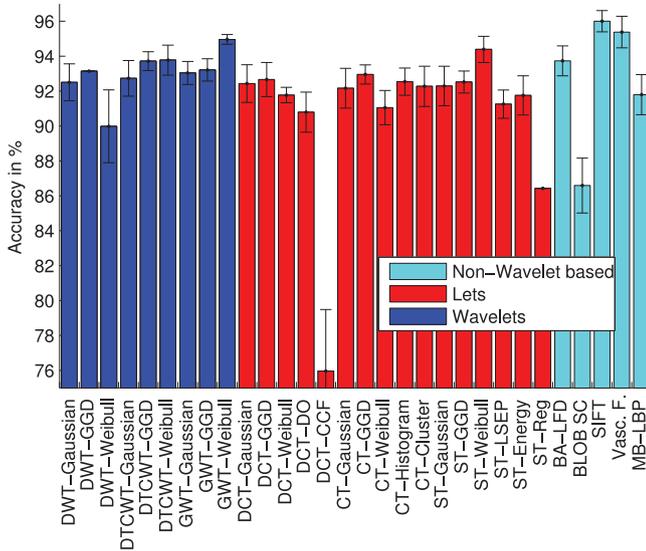
	Non-neoplastic	Neoplastic	Total
Images	168	392	560

Fig. 23. Accuracies and the standard deviations of the methods for the NBI-high-magnification database Hiroshima.

Fig. 24. Result of the McNemar test of the NBI-high-magnification database Hiroshima (2-class case). A black square in the plot means that the two considered wavelet-based method are significantly different with significance level $\alpha = 0.05$. A white square means that there is no significant difference between the methods.

The captured images were trimmed by medical doctors and endoscopists to a rectangle so that this rectangle contains an area in which typical microvessel structures appear. The size of this (908) rectangles varies between about 100×100 and 800×900 pixels. The database is acquired by extracting patches of size 256×256 from the rectangular trimmed images (the wavelet based approaches need images of size $2^n \times 2^n$). Unfortunately, we do not have any information from which patient or which polyp the rectangular trimmed images were taken, we only know from which rectangular images the 256×256 patches were extracted. So we cannot use LOPO cross classification like for the other databases or at least leave-one-polyp out cross validation to avoid biased results. We only can (and did) ensure that a 256×256 patch from one rectangular trimmed image is not permitted to be a neighbor of another 256×256 patch from the same rectangular trimmed image using the kNN-classifier. So we use a sort of Leave-one-parent image-out cross validation. Table 6 lists the number of images per class.

In Fig. 23 we see the averaged accuracies and the standard deviations of our employed methods for the NBI-high-magnification database Hiroshima. As we can see, GWT and ST combined with Weibull features, the vascularization feature approach and especially the SIFT approach achieve the highest results. So this is the only database where a non-wavelet based state-of-the-art approach achieves the best results.

Comparing the results using Gaussian, GGD and Weibull features of the different wavelet-based approaches, the DT-CWT, GWT and partly also the ST provide the best results (like for the CC-high-magnification database and similar to the CC-i-Scan database). The results of the DCT, CT and DWT are slightly lower. Once again, extracting GGD features provides constantly better results than extracting Gaussian features. The distinctly worst results are provided by the DCT-CCF approach. Also the ST-Reg approach performed comparatively weak.
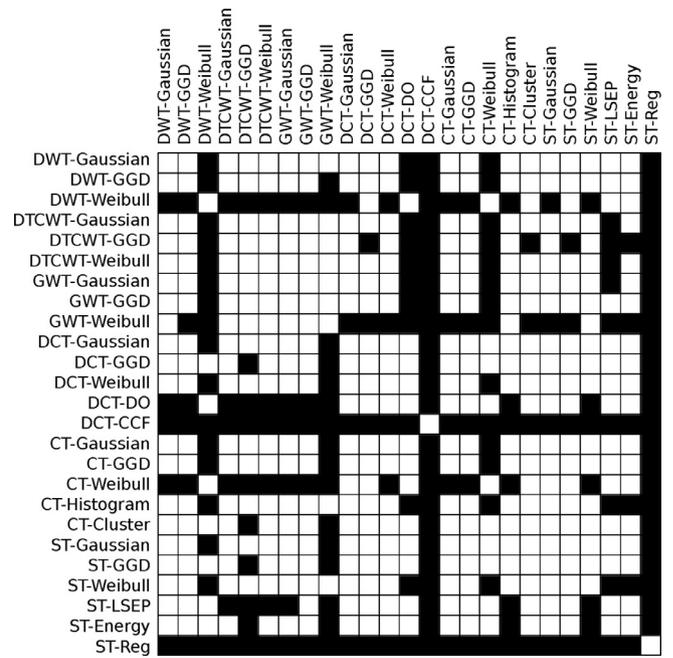
This results are also reflected in the results of the MC-Nemar test shown in Fig. 24, where it is shown that DWT-Weibull, DCT-CCF and ST-Reg are significantly worse than most of the other approaches.

5. Discussion

In this section we will analyze the problems of some of the approaches and the reasons why some approaches work better or worse than others. First we will analyze which probability distribution is best suited to model the subband coefficients, then we will analyze why the DCT-DO approach performs worse than the DCT-GGD approach, and finally we will discuss the problems of other employed approaches.

5.1. Subband coefficient distribution

We have seen in Section 4, that for most of the databases the Weibull features achieved the highest results of all features extracted from the wavelet-based transforms. The Gaussian features achieved average results. The results of the GGD features were constantly better than these of the Gaussian features, but mostly worse than these of the Weibull features. So altogether, we modeled the subband coefficient distribution using three different probability distributions. We further denote these three approaches as density-model approaches.

In this section we will show that the Weibull distributions is most suited to model the subband coefficient distribution.

But first let us take a closer look to the density functions of the three probability distributions. The simplest distribution is the Gaussian distributions. The density function has always the shape of the Gaussian bell curve. μ is the location parameter and σ is the scale parameter of the curve. Examples of three different Gaussian density functions can be seen in Fig. 25(a).

The density function of the generalized Gaussian distribution (GGD) with shape parameter β and scale parameter α can already form slightly different shapes as can be seen in Fig. 25(b).
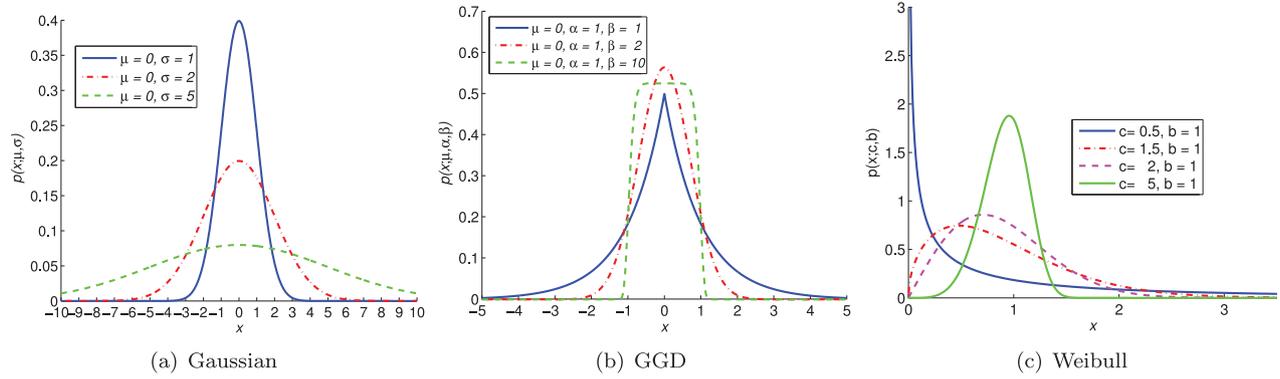
Fig. 25. Gaussian probability density functions with different scale parameters σ (a), GGD probability density function with different shape parameters β (b), and Weibull probability density function with different shape parameters c (c).
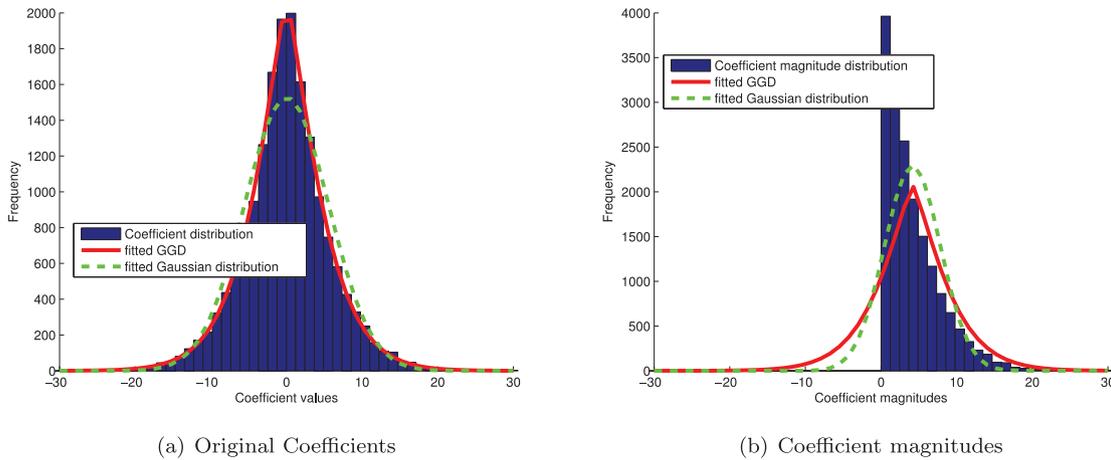

Fig. 26. DWT level 1 subband coefficients modeled by the Gaussian distribution and the GGD, where either the original coefficients (a) or the coefficient magnitudes (b) are modeled.

The Weibull density function can form totally different shapes as can be seen in Fig. 25(c). By means of the shape parameter c the Weibull distribution can take the shape of other probability distributions like the exponential distribution (for $c = 1$), the Rayleigh distribution (for $c = 2$) and the Gaussian distribution (for $k \approx 3.602$).

Now let us take a closer look to the subband coefficient distributions of our employed wavelet based methods. Here we have to differentiate between the wavelet based methods producing complex valued subband coefficients (DT-CWT, GWT, DCT) and those producing real valued subband coefficients (DWT, CT, ST). In case of the complex valued coefficients, it is most common to model the subband coefficient magnitudes. In case of the Wavelet based methods producing real valued coefficients, it is possible to model the original subband coefficients (as done with the GGD) or the subband coefficient magnitudes (as done with the Gaussian and Weibull distribution).

Using the Weibull distribution, we should always model the subband coefficient magnitudes, since the Weibull probability density function is zero for $x < 0$ as we can see in Eq. 13. In case of the Gaussian distribution and the GGD, it is theoretically possible to model the subband coefficients of the original subband coefficients (in case of real valued coefficients) as well as the coefficient magnitudes. However, both distributions are modeling the original (real valued) subband coefficients far more accurately than the subband coefficient magnitudes as can be seen in an example in Fig. 26.

In Fig. 27 we compare the results of the wavelet based methods producing real valued coefficients, where GGD and Gaussian features are extracted either from the original coefficients or from the coefficient magnitudes.

As we can see in Fig. 27(a), the classification rates for the Gaussian distribution are higher modeling the coefficient magnitudes than those modeling the original coefficients. In case of the GGD, the results are similar for modeling the original coefficients and the coefficient magnitudes.

So the results for extracting Gaussian and GGD features from wavelet based transforms producing real valued coefficients are higher or at least similar high when modeling the subband coefficient magnitudes instead of modeling the original coefficients, although the original coefficients can be modeled far more accurate than the coefficient magnitudes. This is a strong indication that the coefficient magnitudes provide a higher discriminativity for the differentiation of polyps than the original coefficients.

In Fig. 28, we see histograms of the subband coefficient magnitudes from an endoscopic image. For each of the employed wavelet based transforms, there is one histogram of a subband from the finest scale level. As we can see in Fig. 28, the appearance and shape of the subband coefficient distributions (the blue bars) of the wavelet based methods producing complex valued coefficients (top row) nearly remind of the exponential distribution, whereas those wavelet based methods producing real valued coefficients (bottom row) remind of the Rayleigh distribution. So the Weibull distribution could be able to accurately represent subband coefficient magnitude distributions, whereas the Gaussian and generalized Gaussian distribution are obviously unable to represent the subband
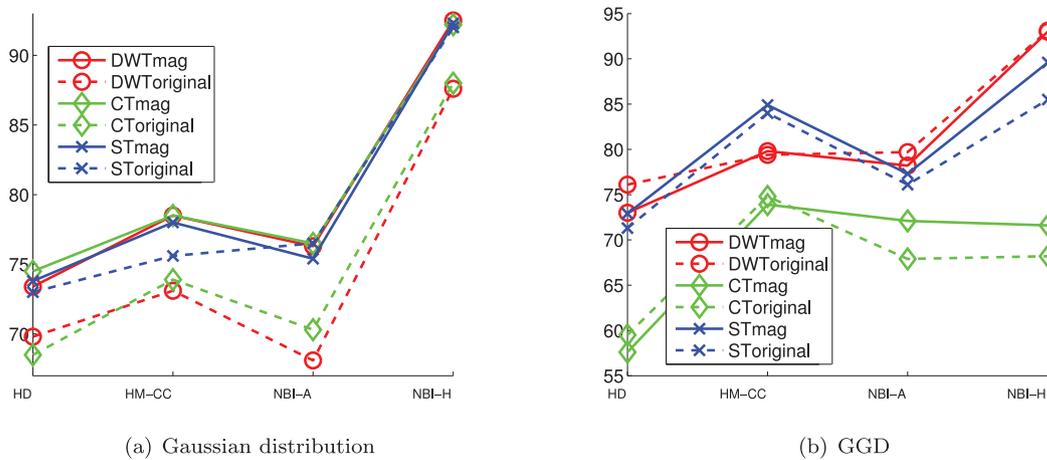
(a) Gaussian distribution

(b) GGD

Fig. 27. Classification rates using DWT, CT and ST for extracting Gaussian (a) and GGD (b) features either from the subband coefficient magnitudes or from the original subband coefficients.
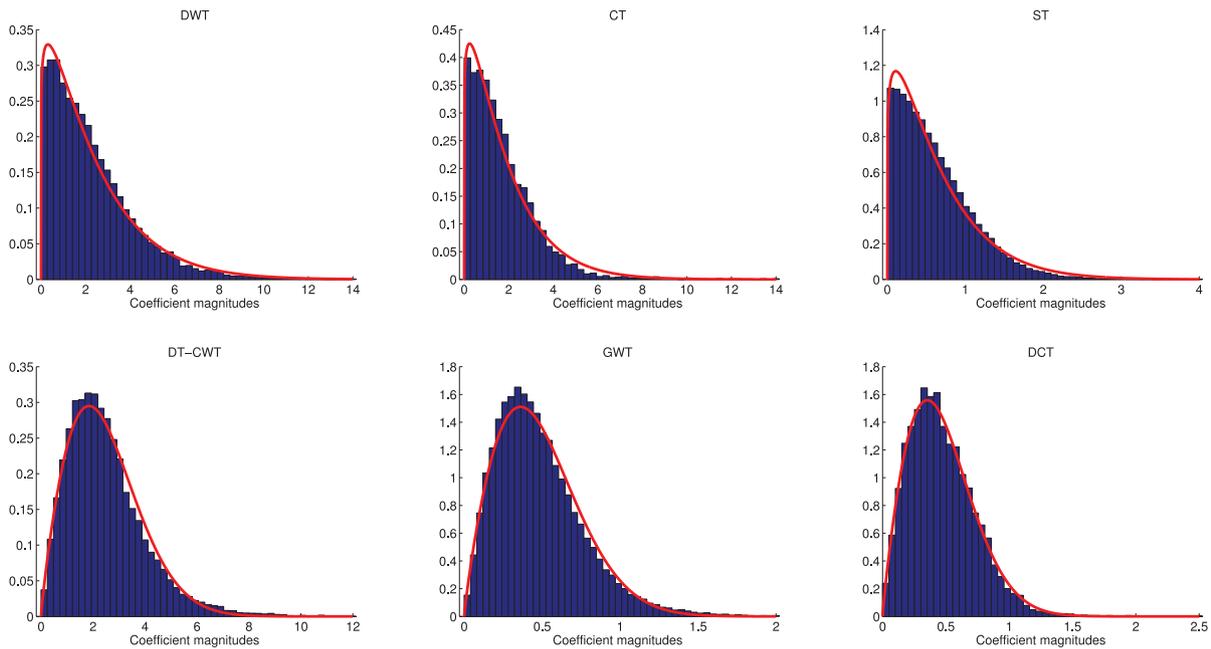

Fig. 28. Histograms of the subband coefficient magnitudes of the employed wavelet based transforms (blue bars) and the fitted Weibull density function (red line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

coefficient magnitudes. The red lines in Fig. 28 show the density functions of the Weibull distributions fitted to the corresponding subband coefficients. So for this example, the Weibull distribution is in fact fitting the subband coefficients of all wavelet based transforms quite well but not perfect.

To see if the Weibull distribution actually fits the subband coefficient distribution of the subbands of our endoscopic images, we apply the Kolmogorov–Smirnov test (Chakravarti et al., 1967) as Goodness-of-Fit (GoF) test to all images (respectively the resulting subbands using our wavelet based transforms) of our employed databases. In case of the CC-i-Scan database, we merge the images of all 8 sub-databases to one image database. The Kolmogorov–Smirnov statistic is a nonparametric test that quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. The null distribution of this statistic is calculated under the null hypothesis that the sample is drawn from the reference distribution. We apply the Kolmogorov–Smirnov test at the 5% significance level. A problem of the Kolmogorov–Smirnov test (and any other

GoF test) is that the outcome is strongly effected by the number of coefficients per subband. This is caused by the fact that a higher number of coefficients reduces the uncertainty (the higher the number of coefficients, the smaller the chance that the subband distribution is only caused by coincidence) and vice versa. So for a high number of coefficients (e.g. 256^2), the actual subband distribution has to be nearly perfectly fitted by the Weibull distribution so that the Kolmogorov–Smirnov does not reject the null-hypothesis, whereas for a low number of coefficients the Weibull distribution only needs to adequately fit the coefficients distribution.

Now we have the problem that the size of the subbands (the number of coefficients per subband) varies across the different wavelet based transforms and the different scale levels. E.g. the subbands of the GWT and ST all have the size of the original image (256×256), whereas the subband size of the DCT and CT goes down to a minimum of 356 respectively 126 coefficients per subband. That means we need a fixed number of coefficients per subband to achieve a reliably information about how well the
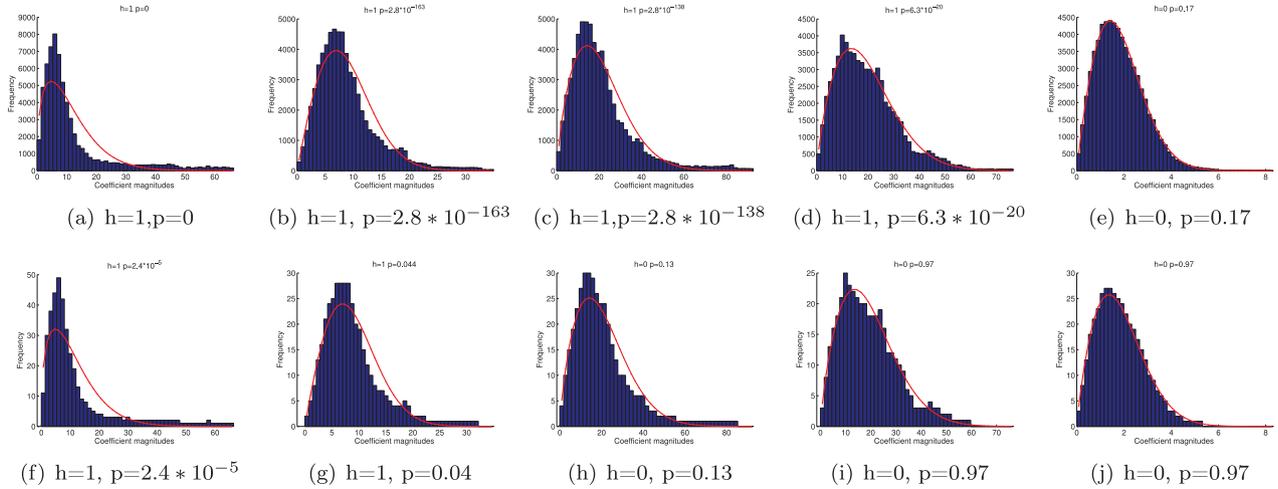
Fig. 29. GWT subband coefficient histograms (blue bars) with 50 bins, the fitted Weibull distributions (red line) and the null-hypothesis decisions and p-values of the Kolmogorov–Smirnov GoF test. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

subband distribution is fitted by the Weibull distribution using the Kolmogorov–Smirnov test. That is the reason why we reduced each subband with more than 400 coefficients to about 400 coefficients. Empirical evaluation showed that $n = 400$ is a well chosen number of coefficients per subband so that the Kolmogorov–Smirnov test does not reject the null-hypothesis for a well fitted subband distribution and reject the null-hypothesis for a misfitted subband distribution. To get a representative sample of an original subband, we divide the range of coefficient values of the subband in 50 equidistant, non-overlapping intervals I_n, $n \in \{1, \ldots, 50\}$ and randomly sample $\frac{|I_n| * 400}{N}$ (rounded to integer values) coefficients with values inside of the interval I_n, where $|I_n|$ is the number of coefficients inside of I_n and N is the number of coefficients in the whole subband. We further denote these subbands as reduced subbands.

In Fig. 29 we see some examples of GWT subbands, their reduced subbands, the corresponding p-values and the decisions if the null-hypothesis (subband is Weibull distributed) applies (h=0, which means $p \geq \alpha = 0.05$)) or not (h=1 which means $p < \alpha$). In the first row we see the original subbands and in the second the corresponding reduced subbands. The higher the p-value, the better the subband coefficients are modeled by the Weibull distribution according to the Kolmogorov–Smirnov test.

As we can see in the first row of Fig. 29, a 256×256 sized subband has to be fitted nearly perfect by the Weibull distribution, so that the Kolmogorov–Smirnov test does not reject the null-hypothesis (Fig. 29(e)). Even if the subband is quite well fitted by the Weibull distribution (Fig. 29(d)), the null-hypothesis is clearly rejected.

For the reduced subbands on the other hand, a reasonable fitting is enough so that the subbands are Weibull distributed regarding to the Kolmogorov–Smirnov test. Fig. 29(h) shows an example of a very narrowly rejected null-hypothesis and Fig. 29(i) shows an example where the null-hypothesis narrowly applies. If a reduced subband is clearly not Weibull distributed (Fig. 29(f)), then the Kolmogorov–Smirnov test still results in a clearly rejection of the null-hypothesis. On the other hand, if a reduced subband is quite well fitted by the Weibull distribution (Fig. 29(i,j)), then the null-hypothesis clearly applies.

So we can see from the examples in the bottom row of Fig. 29, that if a reduced subband is Weibull distributed according to the Kolmogorov–Smirnov test, then the extracted Weibull parameters are representing the actual subband distribution relatively accurate. That means that the Weibull parameters provide sufficient in-
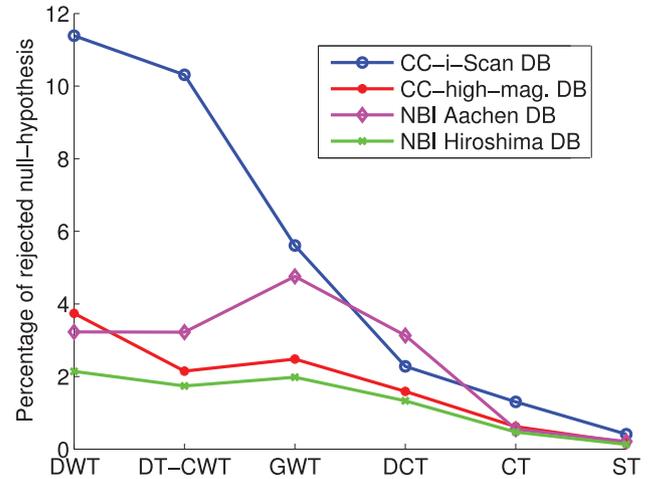

Fig. 30. Kolmogorov–Smirnov GoF test results for the 4 databases and all wavelet based transforms using the reduced subbands.

formation about the subband distribution for the further classification of the images.

In Fig. 30 we see the percentage of rejected null-hypothesis of each database and wavelet based transform for applying the Kolmogorov–Smirnov GoF test at the 5% significance level to the reduced subbands of each image of the databases.

As we can see in Fig. 30, the vast majority of the subbands of each of the wavelet based transforms are sufficiently well fitted by the Weibull distribution. Especially in case of the CT and ST, nearly all subbands are well represented using the Weibull distribution. Only in case of the CC-i-Scan database and DWT or DT-CWT, more than 10% of the subbands are not sufficiently well fitted by the Weibull distribution.

5.2. DCT-DO and DCT-GGD

As already mentioned before, the DCT-GGD and the DCT-DO approach extract the same features. The only difference is that the DCT-DO approach realigns the features by circular shifting them to achieve rotation invariance. However, the results of the DCT-DO approach are worse than the results of the DCT-GGD for each of the employed databases. In this section we will find out the reason for the worse DCT-GGD results. For this, we analyze distribution
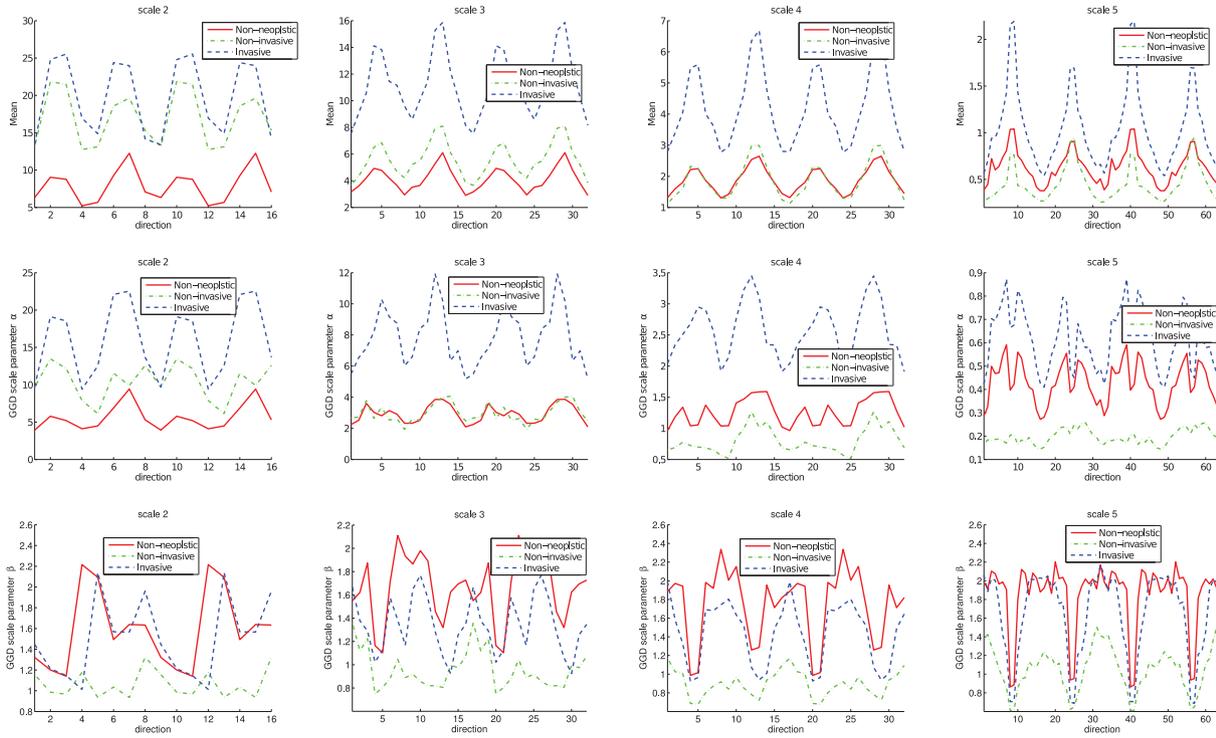
Fig. 31. Gaussian and GGD distribution parameters of DCT subband coefficients across different orientation.

parameters (mean (the location parameter of the Gaussian distribution), GGD scale parameter α and GGD shape parameter β) of the DCT subband coefficients for the different directions per scale. In Fig. 31 we see the values of these parameters for three different images of the CC-high-magnification database. One image is from a non-neoplastic polyp, one from a adenomatous polyp and one from a cancerous polyp. As we can see in Fig. 31 (apart from the fact that the parameters of the first half of directions are identical to the one of the second half of directions), the curves have a periodic pattern, which is similar for all three images at all 4 scale levels. The curves of all three images have mostly identical peaks and pits. These is highly visible for the curves of the means in the first row of the figure, but also visible for the GGD parameters.

That means that the coefficient magnitudes of the DCT are in general higher for the subbands of certain directions than those of certain other directions and hence not independent distributed across the different directional subbands. There are typical shapes and magnitudes for the subband coefficients at different directions (relative to each other).

So if the features are circular shifted on each scale level, using the dominant orientation as reference, then the periodic pattern along the directions is shifted. That means that two relative similar feature vectors can become completely different after circular shifting them if the two feature vectors have different dominant orientations. The circular shifting does only make sense if the coefficient magnitudes of the DCT are independent distributed across the different directional subbands. This is probably the reason for the worse results of the DCT-DO approach compared to the results of the DCT-GGD approach.

5.3. Problems of various wavelet based approaches

5.3.1. DCT-CCF

The DCT-CCF approach achieved the worst results of all our employed methods, although it was originally proposed for endoscopic image processing. To see if the covariance between features

of subbands from different color channels is generally suited for the distinction of colonic polyps, we changed the original approach and computed the covariance using the Weibull features instead of the originally proposed Gaussian features (mean and standard deviation). However, it turned out that this even decreases the results compared to the original approach. We also made an experiment using the RGB color space instead of the originally proposed HSV color space, but that only led to a small improvement of some of the results. So the covariance of features extracted from subbands of different color channels seems to be inadequate for the classification of colonic polyps.

5.3.2. CT-Histogram and CT-Cluster and ST-Energy

The problem with these three approaches is that their extracted features are not able to accurately give the information about how the subband coefficients are distributed in a subband. The parameters of a distribution which is reasonably fitting the subband coefficient distribution are able to give more information about the distribution of the subband coefficients as compared to a histogram with 10 bins (where the location of the 10 bins has to be fixed and hence not optimal to model the coefficient distribution of each subband) or 3 cluster centers of a subband (CT-Cluster) or only the energy of a subband (ST-Energy).

5.3.3. ST-Reg

The probably biggest problem with the ST-Reg approach for the classification of endoscopic polyp images is the quite big intraclass variability inside of our polyp databases. The images of polyps of one class can have quite different appearances, but the ST-Reg approach generates only one regression model per class. The ST-Reg approach computes a regression model which estimates the regression parameters β_0^c and β_1^c of a whole class c from the training images and then assigns a evaluation images to that class, whose regression model fits best to the considered evaluation set image. The problem is that only one regression model per class is only appropriate if the images of one class are quite similar. This is the

case for the Brodatz and VisTex texture database, the two texture databases for which this approach is originally proposed, but certainly not for our endoscopic image databases.

6. Conclusion

In this work we showed that wavelet based approaches are well suited for the automated classification of colonic polyps. The best results were achieved by modeling the subband coefficient distribution by means of the Weibull distribution. For the DT-CWT, GWT and ST, these results even mostly outperformed the state-of-the-art approaches for colonic polyp classification.

We found a strong indication that modeling the subband coefficient magnitude distribution provides a higher discriminativity for the differentiation of polyps than modeling the original coefficients for those wavelet based transforms producing real valued subband coefficients. By means of the Kolmogorov–Smirnov GoF test we found out that the subband coefficient magnitudes of our employed wavelet based transforms are not actually Weibull distributed, but at least almost Weibull distributed. All other employed probability distributions (Gaussian distribution and generalized Gaussian distribution) are totally unable to model the subband coefficient magnitudes, which explains their worse results compared to the results using the Weibull distribution.

Comparing the results using Gaussian and Weibull features across the different wavelet-based approaches, the DT-CWT, GWT and ST are generally providing clearly better results than the DWT, DCT, and CT.

The already published approaches based on the contourlets, curvelets and shearlets were not able to compete with approaches extracting Weibull features except for the NBI-high-magnification database Aachen, were the ST-LSEP approach achieved the best result. Some of these already published approaches performed quite poor like e.g. the DCT-CCF and the DCT-DO approach.

Acknowledgments

References

André, B., Vercauteren, T., Buchner, A., Wallace, M., Ayache, N., 2012. Learning semantic and visual similarity for endomicroscopy video retrieval. IEEE Trans. Med. Imaging 31, 1276–1288.

André, B., Vercauteren, T., Buchner, A.M., Wallace, M.B., Ayache, N., 2011. A smart atlas for endomicroscopy using automated video retrieval. Med. Image Anal. 15, 460–476.

Barbosa, D.J., Ramos, J., Correia, J., Lima, C., 2009. Automatic detection of small bowel tumors in capsule endoscopy based on color curvelet covariance statistical texture descriptors. In: Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2009), pp. 6683–6686.

Burt, P., Adelson, E., 1983. The Laplacian pyramid as a compact image code. IEEE Trans. Commun. 31, 532–540.

Candes, E., Demanet, L., Donoho, D., Ying, L., 2006. Fast discrete curvelet transforms. Multiscale Model. Simul. 5, 861–899.

Candes, E.J., Donoho, D.L., 2002. New tight frames of curvelets and optimal representations of objects with piecewise c2 singularities. Comm. Pure Appl. Math 219–266.

Chakravarti, I.M., Laha, R.G., Roy, J., 1967. Handbook of Methods of Applied Statistics, vol. 1. John Wiley and Sons.

Chowdhury, T., Whelan, P., Ghita, O., 2008. A fully automatic cad-ctc system based on curvature analysis for standard and low-dose ct data. IEEE Trans. Biomed. Eng. 55, 888–901.

Daubechies, I., 1992. Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics.

Do, M., Vetterli, M., 2002. Wavelet-based texture retrieval using generalized gaussian density and Kullback–Leibler distance. IEEE Trans. Image Process. 11, 146–158.

Do, M., Vetterli, M., 2005. The contourlet transform: an efficient directional multiresolution image representation. IEEE Trans. Image Process. 14, 2091–2106.

Do, M.N., 2001. Directional Multiresolution Image Representations. University of Canberra.Technical report.

Dong, Y., Ma, J., 2013. Feature extraction through contourlet subband clustering for texture classification. Neurocomputing 116, 157–164.

Dong, Y., Tao, D., Li, X., Ma, J., Pu, J., 2015. Texture classification and retrieval using shearlets and linear regression. IEEE Trans. Cybern. 45, 358–369.

Easley, G., Labate, D., Lim, W.Q., 2008. Sparse directional image representations using the discrete shearlet transform. Appl. Comput. Harmon. Anal. 25, 25–46.

Evans, M., Peacock, N.H.B., 2000. Statistical Distributions. Wiley Series in Probability and Statistics, 3rd ed. Wiley.

Gomez, F., Romero, E., 2011. Rotation invariant texture characterization using a curvelet based descriptor. Pattern Recognit. Lett. 32, 2178–2186.Advances in Theory and Applications of Pattern Recognition, Image Processing and Computer Vision.

Gono, K., Yamazaki, K., Doguchi, N., Nonami, T., Obi, T., Yamaguchi, M., Ohyama, N., Machida, H., Sano, Y., Yoshida, S., Hamamoto, Y., Endo, T., 2003. Endoscopic observation of tissue by narrowband illumination. Opt. Rev. 10, 211–215.

Gross, S., Palm, S., Tischendorf, J.J.W., Behrens, A., Trautwein, C., Aach, T., 2012. Automated classification of colon polyps in endoscopic image data 8315.83150W–83150W–8.

Häfner, A., Uhl, A., Wimmer, G., 2014a. A novel shape feature descriptor for the classification of polyps in HD colonoscopy. In: Medical Computer Vision. Large Data in Medical Imaging (Proceedings of the 3rd Intern. MICCAI - MCV Workshop 2013), pp. 205–213.

Häfner, M., Gangl, A., Liedlgruber, M., Uhl, A., Vécsei, A., Wrba, F., 2009. Pit pattern classification using extended local binary patterns. In: Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine (ITAB'09). Larnaca, Cyprus.

Häfner, M., Kwitt, R., Uhl, A., Gangl, A., Wrba, F., Vécsei, A., 2008. Computer-assisted pit-pattern classification in different wavelet domains for supporting dignity assessment of colonic polyps. Pattern Recognit. 42, 1180–1191.

Häfner, M., Kwitt, R., Uhl, A., Gangl, A., Wrba, F., Vecsei, A., 2009. Feature extraction from multi-directional multi-resolution image transformations for the classification of zoom-endoscopy images. Pattern Anal. Appl. 12, 407–413.

Häfner, M., Liedlgruber, M., Uhl, A., 2015a. Colonic polyp classification in high- definition video using complex wavelet-packets. In: Proceedings of Bildverarbeitung für die Medizin 2015 (BVM'15), pp. 365–370.

Häfner, M., Liedlgruber, M., Uhl, A., Vécsei, A., Wrba, F., 2012a. Color treatment in endoscopic image classification using multi-scale local color vector patterns. Med. Image Anal. 16, 75–86.

Häfner, M., Liedlgruber, M., Uhl, A., Vécsei, A., Wrba, F., 2012b. Color treatment in endoscopic image classification using multi-scale local color vector patterns. Med. Image Anal. 16, 75–86.

Häfner, M., Liedlgruber, M., Uhl, A., Vécsei, A., Wrba, F., 2012c. Delaunay triangulation-based pit density estimation for the classification of polyps in high-magnification chromo-colonoscopy. Comput. Methods Progr. Biomed. 107, 565–581.

Häfner, M., Liedlgruber, M., Uhl, A., Wimmer, G., 2014b. Bridging the resolution gap between endoscope types for a colonic polyp classification. In: Proceedings of the 22nd International Conference on Pattern Recognition (ICPR'14), pp. 2734–2744.

Häfner, M., Tamaki, T., Tanaka, S., Uhl, A., Wimmer, G., Yoshida, S., 2015b. Local fractal dimension based approaches for colonic polyp classification. Med. Image Anal. 26, 92–107.

Häfner, M., Uhl, A., Vecsei, A., Wimmer, G., Wrba, F., 2010. Complex wavelet transform variants and discrete cosine transform for scale invariance in magnification-endoscopy image classification. In: 10th IEEE International Conference on Information Technology and Applications in Biomedicine (ITAB), pp. 1–5.

Häfner, M., Uhl, A., Wimmer, G., 2014c. Shape and size adapted local fractal dimension for the classification of polyps in hd colonoscopy. In: Proceedings of the IEEE International Conference on Image Processing 2014 (ICIP'14), pp. 2299–2303.

He, J., Ji, H., Yang, X., 2013. Rotation invariant texture descriptor using local shearlet-based energy histograms. Signal Process. Lett. IEEE 20, 905–908.

Hwang, S., Celebi, M., 2010. Polyp detection in wireless capsule endoscopy videos based on image segmentation and geometric feature. In: Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 2010, pp. 678–681.

Iakovidis, D.K., Koulaouzidis, A., 2015. Software for enhanced video capsule endoscopy: challenges for essential progress. Nat. Rev. Gastroenterol. Hepatol 12, 172–186.

Jabbour, J.M., Saldua, M.A., Bixler, J.N., Maitland, K.C., 2012. Confocal endomicroscopy: instrumentation and medical applications. Ann. Biomed. Eng. 40, 378–397.

Kato, S., Fu, K.I., Sano, Y., Fujii, T., Saito, Y., Matsuda, T., Koba, I., Yoshida, S., Fujimori, T., 2006. Magnifying colonoscopy as a non-biopsy technique for differential diagnosis of non-neoplastic and neoplastic lesions. World J. Gastroenterol. 12, 1416–1420.

Kiesslich, R., 2009. Advanced imaging in endoscopy. Eur. Gastroenterol. Hepatol Rev. 5, 22–25.

Kingsbury, N.G., 1998. 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. Bryce Canyon, USA, pp. 9–12.

Kodashima, S., Fujishiro, M., 2010. Novel image-enhanced endoscopy with i-scan technology. World J. Gastroenterol. 16, 1043–1049.

Kovesi, P.D., 1999. Image features from phase congruency. Videre: J. Comput. Vis. Res. 1, 2–26.

Kovesi, P. D., 2000. MATLAB and Octave functions for computer vision and image processing. Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia. Available from: <http://www.peterkovesi.com/matlabfns/> (accessed 23.02.16).

Kudo, S.E., Hirota, S., Nakajima, T., Hosobe, S., Kusaka, H., Kobayashi, T., Himori, M., Yagyuu, A., 1994. Colorectal tumours and pit pattern. J. Clin. Pathol. 47, 880–885.

Kwitt, R., Uhl, A., 2007. 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). Rio de Janeiro, Brasil, pp. 1–8.

Kwitt, R., Uhl, A., 2010. Lightweight probabilistic texture retrieval. IEEE Trans. Image Process. 19, 241–253.

Lee, T.S., 1996. Image representation using 2d Gabor wavelets. IEEE Trans. Pattern Anal. Mach. Intell. 18, 959–971.

Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.Z., 2007. Learning multi-scale block local binary patterns for face recognition. In: Proceedings of the 2007 International Conference on Advances in Biometrics (ICB'07). Springer-Verlag, Seoul, Korea, pp. 828–837.

Long, Z., Younan, N., 2006. Contourlet spectral histogram for texture classification. In: Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation, pp. 31–35.

Mallat, S., 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. 11, 674–693.

Manjunath, B.S., Ma, W.Y., 1996. Texture features for browsing and retrieval of image data. IEEE Trans. Pattern Anal. Mach. Intell. 18, 837–842.

Martins, M., Barbosa, D., Ramos, J., Lima, C., 2010. Small bowel tumors detection in capsule endoscopy by gaussian modeling of color curvelet covariance coefficients. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2010), pp. 5557–5560.

McNemar, Q., 1947. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12, 153–157.

Niola, V., Oliviero, R., Quaremba, G., 2006. A Method of Moments for the Estimation of Weibull pdf Parameters. World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA, pp. 382–386.

Ojala, T., Pietikäinen, M., Mäenpää, T., 2002. Multiresolution Gray-Scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24, 971–987.

Roerdink, J.B., Meijster, A., 2000. The watershed transform: definitions, algorithms and parallelization strategies. Fundamenta Informaticae 41, 187–228.

Romain, O., Histace, A., Silva, J., Ayoub, J., Granado, B., Pinna, A., Dray, X., Marteau, P., 2013. Towards a multimodal wireless video capsule for detection of colonic polyps as prevention of colorectal cancer. In: IEEE 13th International Conference on Bioinformatics and Bioengineering (BIBE), pp. 1–6.

Schwartz, W., da Silva, R., Davis, L., Pedrini, H., 2011. A novel feature descriptor based on the shearlet transform. In: IEEE International Conference on Image Processing (ICIP), pp. 1033–1036.

Stehle, T., Auer, R., Gross, S., Behrens, A., Wulff, J., Aach, T., Winograd, R., Trautwein, C., Tischendorf, J., 2009. Classification of colon polyps in NBI endoscopy using vascularization features. Medical Imaging 2009: Computer-Aided Diagnosis. SPIE, Orlando, Florida, USA.

Tamaki, T., Yoshimuta, J., Kawakami, M., Raytchev, B., Kaneda, K., Yoshida, S., Takemura, Y., Onji, K., Miyaki, R., Tanaka, S., 2013. Computer-aided colorectal tumor classification in NBI endoscopy using local features. Med. Image Anal. 17, 78–100.

Varma, M., Garg, R., 2007. Locally invariant fractal features for statistical texture classification. In: Proceedings of the IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil, pp. 1–8.

Vedaldi, A., Fulkerson, B., 2008. VLFeat: an open and portable library of computer vision algorithms. http://www.vlfeat.org/ (accessed 23.02.16).

Vetterli, M., 1984. Multi-dimensional sub-band coding: Some theory and algorithms. Signal Process. 6, 97–112.

Xu, Y., Ji, H., Fermüller, C., 2009. Viewpoint invariant texture description using fractal analysis. Int. J. Comput. Vis. 83, 85–100.

Yuan, Y., Meng, M.H., 2014. A novel feature for polyp detection in wireless capsule endoscopy images. In: International Conference on Intelligent Robots and Systems (IROS 2014), pp. 5010–5015.

Yuce, M., Dissanayake, T., 2012. Easy-to-swallow wireless telemetry. Microw. Mag. IEEE 13, 90–101.

Zuiderveld, K., 1994. Contrast limited adaptive histogram equalization. Graphics Gems IV. Morgan Kaufmann, pp. 474–485.