Medical Image Analysis 17 (2013) 458–474

Contents lists available at SciVerse ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Scale invariant texture descriptors for classifying celiac disease

Sebastian Hegenbart [a], Andreas Uhl [a], Andreas Vécsei [b], Georg Wimmer [a,*]

[a] University of Salzburg, Department of Computer Sciences, Salzburg, Austria
[b] St. Anna Children's Hospital, Department Pediatrics, Medical University, Vienna, Austria

ARTICLE INFO

Article history:
Received 19 March 2012
Received in revised form 29 January 2013
Accepted 1 February 2013
Available online 13 February 2013

Keywords:
Scale invariance
Texture recognition
Celiac disease

ABSTRACT

Scale invariant texture recognition methods are applied for the computer assisted diagnosis of celiac disease. In particular, emphasis is given to techniques enhancing the scale invariance of multi-scale and multi-orientation wavelet transforms and methods based on fractal analysis. After fine-tuning to specific properties of our celiac disease imagery database, which consists of endoscopic images of the duodenum, some scale invariant (and often even viewpoint invariant) methods provide classification results improving the current state of the art. However, not each of the investigated scale invariant methods is applicable successfully to our dataset. Therefore, the scale invariance of the employed approaches is explicitly assessed and it is found that many of the analyzed methods are not as scale invariant as they theoretically should be. Results imply that scale invariance is not a key-feature required for successful classification of our celiac disease dataset.

1. Introduction

Texture analysis is one of the fundamental issues in image processing. The majority of existing texture analysis methods work with the assumption that texture images are acquired from the same viewpoint (Zhang and Tan, 2002). This limitation makes these methods useless for applications, where textures occur with different scales, orientations, or translations. Therefore, scale and orientation invariant texture analysis approaches have been proposed (see Tan (1995) or Zhang and Tan (2002) for surveys on this topic). Invariance is important for many applications in medical image processing, since medical images are often acquired at different scales and viewpoints. This is especially true for endoscopic imagery since mucosal texture is seen from different perspectives and distances to the cavity wall depending on the relative position of the endoscopes tip and the mucosa surface. Fig. 1 illustrates that, depending on the angle between endoscope and the surface (middle case) and the curvature of the surface (rightmost example), different distances between camera and surface may even occur within a single image.

In gastroscopic (and other types of endoscopic) imagery, mucosal texture is usually found with different perspective and scale (see Fig. 3). That means that the mucosal texture shows different spatial scales, depending on the camera perspective and distance to the mucosal wall (see Fig. 1).

As a consequence, endoscopic imagery typically exhibits mucosal texture with different and/or mixed spatial scales, depending on the corresponding acquisition conditions (see Fig. 3 for examples from our celiac disease database).

In this work, we focus on scale invariant texture classification approaches being applied in computer-assisted diagnosis of celiac disease. While most of the used techniques in this work exhibit additional invariance to other transformations like rotation, translation, and illumination, we specifically concentrate on scale invariance for the reasons explained above. The contributions of this manuscript are as follows:

- We apply general purpose scale invariant texture descriptors for the classification of duodenal mucosa texture imagery aiming at the staging of celiac disease.
- Several approaches have been developed to achieve scale (and often orientation) invariance for multi-scale and multi-orientation wavelet transforms. These techniques are mostly applicable to any multi-scale and multi-orientation transform. We employ the Dual-Tree Complex Wavelet Transform (DT-CWT) (Selesnick et al., 2005) instead of the originally proposed transforms and are able to show that our approach works better for the target celiac disease database than other wavelet-type transforms (like e.g. Gabor filters (Fung and Lam, 2009) or steerable pyramids (Montoya-Zegarra et al., 2007) (see Table 3). An additional benefit is the improved ability to compare the different strategies to achieve scale invariance if the underlying transform is the same in all cases.
- We propose a new affine invariant method based on Local Ternary Patterns (LTPs).

* Corresponding author.
E-mail addresses: shegen@cosy.sbg.ac.at (S. Hegenbart), uhl@cosy.sbg.ac.at (A. Uhl), vecsei@stanna.at (A. Vécsei), gwimmer@cosy.sbg.ac.at (G. Wimmer).

1361-8415
http://dx.doi.org/10.1016/j.media.2013.02.001

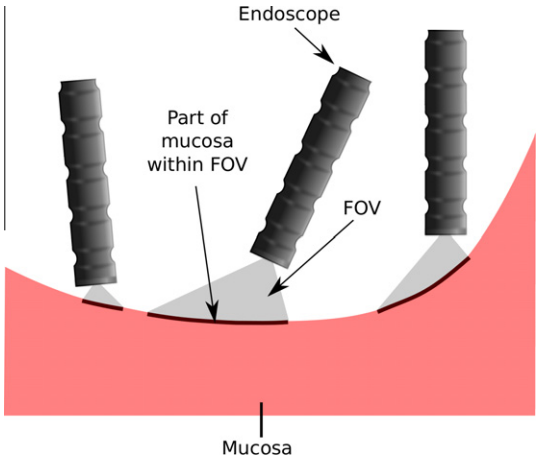

Fig. 1. The field of view (FOV) depending on the endoscopic viewpoint and distance to the mucosal wall.

- We conduct explicit experimental tests for scale invariance for all feature descriptors considered based on the Columbia–Utrecht (CUReT) (Dana et al., 1999) dataset and the Celiac Disease Scale (CDS) database (see Section 6.2.2) following ideas in Varma and Zisserman (2009), revealing that claimed scale invariance cannot be verified for many of the schemes investigated.
- Most approaches are tested for their ability of invariant texture analysis on public databases like Brodatz (Brodatz, 1966), CUReT (Dana et al., 1999), KTH-TIPS (Hayman et al., 2004), or the UIUCTex (Lazebnik et al., 2005) database. Correspondingly, most of the considered methods have been optimized for the corresponding datasets. Hence we have adjusted some of these methods (e.g. using different parameters or replacing parts of the original algorithm) to make them applicable in a sensible manner for the classification of celiac disease (e.g., use of different measures in techniques based on fractal analysis in Section 4 or application of a different clustering strategy for the dense Scale Invariant Feature Transform (SIFT) features in Section 5).
- We show, that methods extracting highly contrast sensitive information work well for the classification of celiac disease, specifically methods based on fractal analysis.

This paper is organized as follows. In Section 2 we briefly introduce the concept of computer-assisted diagnosis of celiac disease by automated classification of duodenal mucosa texture patches and review the corresponding state-of-the-art. In Section 3, we describe strategies to achieve scale invariance for wavelet transforms including the application of the discrete cosine transform (DCT) or the discrete Fourier transform (DFT) to the feature vectors of the wavelet transforms (Häfner et al., 2010; Lo et al., 2004), re-arrangement of feature vectors (cyclic shifting, dominant scale, and slide matching) (Montoya-Zegarra et al., 2007; Lo et al., 2009; Fung and Lam, 2009), or methods that preprocess the image before the wavelet transform is being applied (Pun and Lee, 2003). Section 4 describes techniques based on fractal analysis while Section 5 covers a heterogeneous set of additional approaches to generate scale invariant texture descriptors (e.g. neural nets (Ma et al., 2010; Zhan et al., 2009), SIFT features and region detectors (Fei-Fei and Perona, 2005; Zhang et al., 2006), and multiscale blob features (Xu and Chen, 2006)) as well as a new affine invariant method we propose which is based on scale-normalized Laplacian maxima combined with Local Ternary Patterns (Hegenbart and Uhl, 2013). Experimental results with respect to classification of the celiac disease dataset and with respect to effective scale invariance (by means of the CDS database and parts of the CUReT database) are presented in Section 6. Section 7 concludes our work.

2. Computer-assisted diagnosis of celiac disease

Celiac disease is a complex autoimmune disorder in genetically predisposed individuals of all age groups after introduction of gluten containing food. The gastrointestinal manifestations invariably comprise an inflammatory reaction within the mucosa of the small intestine caused by a dysregulated immune response triggered by ingested gluten proteins of certain cereals (wheat, rye, and barley), especially against gliadine. During the course of the disease, hyperplasia of the enteric crypts occurs and the mucosa eventually

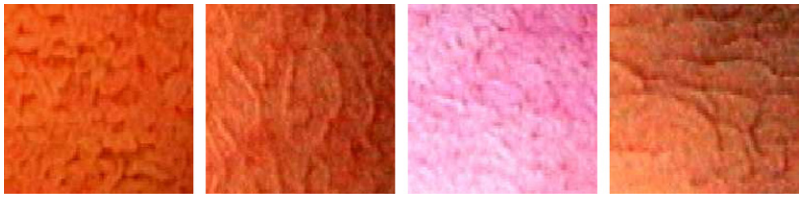

Fig. 2. Example images for the respective classes.

Fig. 3. Images with different perspective and scale.

looses its absorptive villi thus leading to a diminished ability to absorb nutrients. The real prevalence of the disease has not been fully clarified yet. This is due to the fact that most patients with celiac disease suffer from no or atypical symptoms and only a minority develops the classical form of the disease.

Since several years, prevalence data have continuously been adjusted upwards. Fasano et al. (2003) state that more than 2 million people in the United States, this is about one in 133, have the disease. People with untreated celiac disease even if asymptomatic are at risk for developing various complications like osteoporosis, infertility and other autoimmune diseases including type 1 diabetes, autoimmune thyroid disease and autoimmune liver disease. This is why early diagnosis is of highest importance. Endoscopy with biopsy is currently considered the gold standard for the diagnosis of celiac disease. Besides standard upper endoscopy, several new endoscopic approaches for diagnosing celiac disease have been applied (Chand and Mihas, 2006). The modified immersion technique described in Cammarota et al. (2006) is based on the instillation of water into the duodenal lumen for better visualization of the villi. Furthermore magnifying endoscopy (standard endoscopy with additional magnification) has been investigated (Cammarota et al. (2004)). For conducting capsule endoscopy (Petroniene et al., 2005) the patient swallows a small capsule equipped with a camera that takes images of the duodenal mucosa during its passage through the intestine. All these techniques aim for detection of total or partial villous atrophy and other specific markers that show a high specificity for celiac disease in adult patients like scalloping of the small bowel folds, reduction in the number or loss of Kerkring's folds, scalloped folds, mosaic patterns, and visualization of the underlying blood vessels (Niveloni et al., 1998).

Automated classification as a support tool is an emerging option for endoscopic treatments (e.g. (Liedlgruber and Uhl, 2011a,b)). Systems are being developed that support physicians during surgery or highlight malignant areas during an endoscopy for further inspection. Such systems could also be used for training purposes. In the context of celiac disease, an automated system identifying areas affected by celiac disease in the duodenum would offer the following benefits (among other):

- Methods that help indicating specific areas for biopsy might improve the reliability of celiac disease diagnosis. As biopsying is invasive and the number of biopsy samples should be kept small, optimal targeting is desirable. This targeting can be supported by an automated system for identification of areas affected by celiac disease.
- The whole diagnostic work-up of celiac disease, including duodenoscopy with biopsies, is time-consuming and cost-intensive. To save costs, time, and manpower and simultaneously increase the safety of the procedure it would be desirable to develop a less invasive approach avoiding biopsies. Recent studies (Cammarota et al., 2006, 2007) investigating such endoscopic techniques report reliable results. These could be further improved by analysis of the acquired visual data (digital images and video sequences) with the assistance of computers.
- The (human) interpretation of the video material captured during capsule endoscopy (Petroniene et al., 2005) is an extremely time consuming process. Automated identification of suspicious areas in the video would significantly enhance the applicability and reduce the costs of this technique for the diagnosis of celiac disease.

The celiac state of the duodenum is usually determined by visual inspection during the endoscopic session followed by a biopsy of suspicious areas. During endoscopy at least four duodenal biopsies are taken. The severity of the mucosal state of the extracted tissue can be histologically staged according to a modified Marsh scheme (Oberhuber et al., 1999) which is based on Marsh (1992). According to this staging scheme, we have divided available duodenal image material into four different classes, Marsh-0 Marsh-3a, Marsh-3b and Marsh-3c (see Fig. 2).

Marsh-0 represents a healthy duodenum with normal crypts and villi, Marsh-3a, Marsh-3b and Marsh 3c have increased crypts and mild atrophy (3a), marked atrophy (3b) or the villi are entirely absent (3c), respectively. Types Marsh-3a to Marsh-3c span the range of characteristic changes caused by celiac disease, where Marsh-3a is the mildest and Marsh-3c is the most severe form. We also consider the 2-class case, where we only differentiate between healthy (Marsh-0) and unhealthy (Marsh-3a, Marsh-3b and Marsh 3c) mucosal types, respectively.

As described in Section 1 endoscopic image material of mucosal texture is usually found at different perspective and scale (see Fig. 3 for examples from our database). Therefore, the employment of scale invariant feature description techniques is a highly intuitive idea for a computer-assisted diagnosis system.

Prior approaches dealing with the computer-aided diagnosis of celiac disease using endoscopic imagery do exist but they do not focus on scale invariance. With respect to feature descriptors in previous papers, we have investigated several variants of Local Binary Pattern (LBP) based operators (Vécsei et al., 2011; Hegenbart et al., 2011), band-pass type Fourier filters (Vecsei et al., 2009), as well as histogram and wavelet-transform based features (Uhl et al., 2011a; Vécsei et al., 2008). We have also systematically compared the classification performance of two different image capturing techniques and various pre-processing schemes using a set of different feature extraction and classification methods (Hegenbart et al., 2009). Smoothness/sharpness measures have been used as features in Ciaccio et al. (2011). Techniques involving temporal information computed from videocapsule endoscopy have been described recently (Ciaccio et al., 2010b,a).

3. Scale invariant wavelet based features

In this section we describe scale invariant texture descriptors, that are based on multi-scale and multi-orientation transforms like the discrete wavelet transform, the Gabor wavelet transform and the dual-tree complex wavelet transform (DT-CWT). Various wavelet-based feature extraction methods have been proposed for endoscopic image analysis (since approximately 2003) (e.g. Kwitt et al., 2009; Barbosa et al., 2008, 2009; Iakovidis et al., 2004). The subbands of these methods contain information about different scales and orientations of an image. The strategies to make these transforms invariant to scale change are to transform or reorder the corresponding transform coefficients or to find a different representation for the images before applying the respective transforms. The underlying principles how to achieve scale invariance are similar for the approaches in this section (except for the approach that re-arranges the image before the transformation). If an image is scaled, then the subbands of the scaled image are shifted across the scale dimension compared to the subbands of the unscaled image. In Fig. 4 we see two checkerboard patterns in the first row, where the right pattern is a scaled version of the left one with a scale factor of two. In the second row of Fig. 4, the corresponding subband means (when using DT-CWT) of the checkerboard patterns are shown. We can see that the subband means of the scaled checkerboard pattern (the right one) are shifted one scale level up compared to the subband means of the unscaled checkerboard pattern.

Most strategies used to achieve scale invariance of the methods described are applicable to any multi-scale and multi-orientation transform method. We propose to apply these different strategies

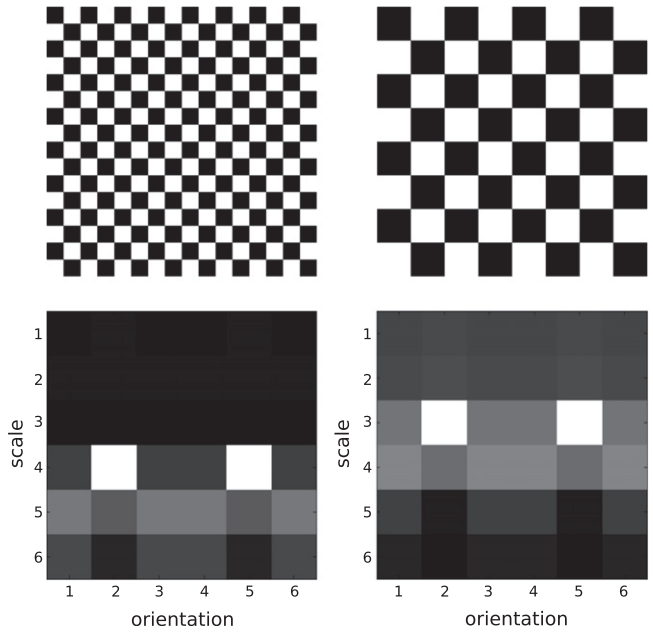

Fig. 4. Cyclic shifting of the means of the subbands across the scale dimension.

to the DT-CWT as opposed to the various (wavelet) transforms published in the original papers. The advantages of this approach are as follows: First, the DT-CWT provides better results for the classification of celiac disease than any other type of multi scale transform (as we will see in Table 3). Second, the different strategies to achieve scale invariance are easier to compare if the underlying transform is the same in all cases. In fact, the results for classifying celiac disease are better for all strategies if we use the DT-CWT instead of the originally proposed (wavelet) transforms. One possible reasons for that is the shift invariance of the DT-CWT. Shift invariance is important for the classification of celiac disease, since the representation of features of an image by wavelet coefficients should not be dependent on the position of the features in the image. Another possible reason is the high redundancy of the DT-CWT (it is using two separate Discrete Wavelet Transforms (DWT) and thus has double the redundancy of the DWT), which provides extra information for the analysis.

Kingbury's DT-CWT (Selesnick et al., 2005) divides an image into six directional (15°, 45°, 75°, 105°, 135°, 165°) oriented subbands per level of decomposition. The DT-CWT analyzes an image only at dyadic scales. For some of the strategies proposed in this section, a finer scale resolution is required. The double dyadic dual-tree complex wavelet transform (D^3T-CWT) (Lo et al., 2009) overcomes this issue by introducing additional levels between dyadic scales. These additional levels are generated by recursively applying the DT-CWT to a downscaled version of the original image (using a factor of $2^{-0.5}$ in downscaling). For each subband we calculate the statistical features' mean (μ) and standard deviation (σ) from the absolute values of the subband coefficients. We denote a statistical feature of a subband $S_{l,d}$ with scale level $l \in \{1,\dots,L\}$ and orientation $d \in \{1,\dots,D\}$ as $q_{l,d}$. The feature vector of an image is composed of these statistical features collected from the different subbands.

3.1. Applying discrete fourier transform and discrete cosine transform to DT-CWT

Features that are approximately scale invariant can be generated by applying the discrete Fourier transform (DFT) across the scale dimension of a feature vector of the D^3T-CWT (Häfner et al., 2010) (see Fig. 5):

$$Q_{n,d} = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} q_{l,d} e^{\frac{-i2\pi(l-1)(n-1)}{L}},$$

with $n \in \{1,\dots,L\}$, $d \in \{1,\dots,D\}$.

The vector $fv_{SI} = \{|Q_{1,1}|,\dots|Q_{L,1}|,|Q_{1,2}|,\dots|Q_{L,2}|,\dots|Q_{L,D}|\}$ provides a texture feature that is nearly invariant to scale. The feature curve of a feature vector shifts if input texture is scaled (see Fig. 4). If a feature curve $q^m_{l,d}$ is a cyclic shifted version of the old one $\left(q^m_{l(\mathrm{mod}L+1),d} = q_{l+m(\mathrm{mod}L+1),d}, m \in \{1,\dots,L\}\right)$, then applying DFT to the feature curves followed by taking the magnitude of it provides the same results for the old and new feature curve ($|Q_{n,d}| = \left|Q^m_{n,d}\right|$, where $Q^m_{n,d}$ is defined like $Q_{n,d}$, but with using $q^m_{l,d}$ instead of $q_{l,d}$). The reason for that follows from the Shift Theorem of the DFT: $Q^m_{n,d} = Q_{n,d} e^{\frac{2\pi i(n-1)m}{L}}$ (with $\left|e^{\frac{2\pi i(n-1)m}{L}}\right| = 1$). The problem is, that the Shift Theorem is only valid if the input signal $q_{l,d}$ is periodic, but it is questionable why these statistical features should be periodic. However if the statistical features are close to zero at both ends, the approach provides good scale invariance. In Fig. 5 we can see the means of the subband coefficients from the red color channel of an image of the celiac disease database (we separately apply the DFT to the features of the three color channels of the RGB color space). We can see that the coarse end ($l = 6$) has the highest means, which are absolutely not close to zero. This of course questions input periodicity.

Another possibility is to consider only the real part of the DFT, which is a cosine transform. This leads us to the application of the Discrete Cosine Transform (DCT) across scale dimension (see Häfner et al., 2010). The DCT is not invariant to cyclic shifts of the feature curve and so it is not theoretically clear if the DCT enhances the scale invariance of the DT-CWT in general. Even if the DCT would be invariant to cyclic shifts, this would not enhance scale invariance since the input signal $q_{l,d}$ is not periodic. Results in Häfner et al. (2010) indicate that the DCT enhances the scale invariance at least for small differences of scales (maximum scale factor ≈1.4), tests for bigger scale differences were not made.

A related approach (Lo et al., 2004) is to resize each D^3T-CWT subband to the size of the original image. In this way we get a local feature vector for each pixel consisting of the subband coefficients (absolute values) at the position of the pixel. Like in the approach before, the DFT is applied across the scale dimension of these local feature vectors (see Fig. 5), but this time to each local feature vector instead of the statistical features of the subbands:

$$Q^{local}_{n,d}(x,y) = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} S_{l,d}(x,y) e^{\frac{-i2\pi(l-1)(n-1)}{L}},$$

with $n \in \{1,\dots,L\}$, $d \in \{1,\dots,D\}$. For each "transformed subband" $Q^{local}_{n,d}$ we compute the statistical features mean and standard deviation. Since the operations for achieving scale invariance are applied to the local subband coefficients (instead to the global statistical subband features mean and standard deviation like in the approaches before) we denote this method as "D^3T-CWT with DFT (local)". In extending Lo et al. (2004) we additionally use the DCT instead of the DFT and denote this method as "D^3T-CWT with DCT (local)". A feature vector of DT-CWT or DT-CWT with DCT has a length of 216 (6 orientations × 6 scale levels × 3 color channels × 2 statistical features per subband). In case of DT-CWT with DFT, for each direction d, the following features form complex conjugates: $Q_{2,d} = Q^*_{6,d}$ and $Q_{3,d} = Q^*_{5,d}$. That means 2 of the 6 scale levels (scale levels 5 and 6) are redundant, which reduces the length of the feature vector to 144 elements. If using D^3T-CWT instead of DT-CWT, the feature vector length is doubled.

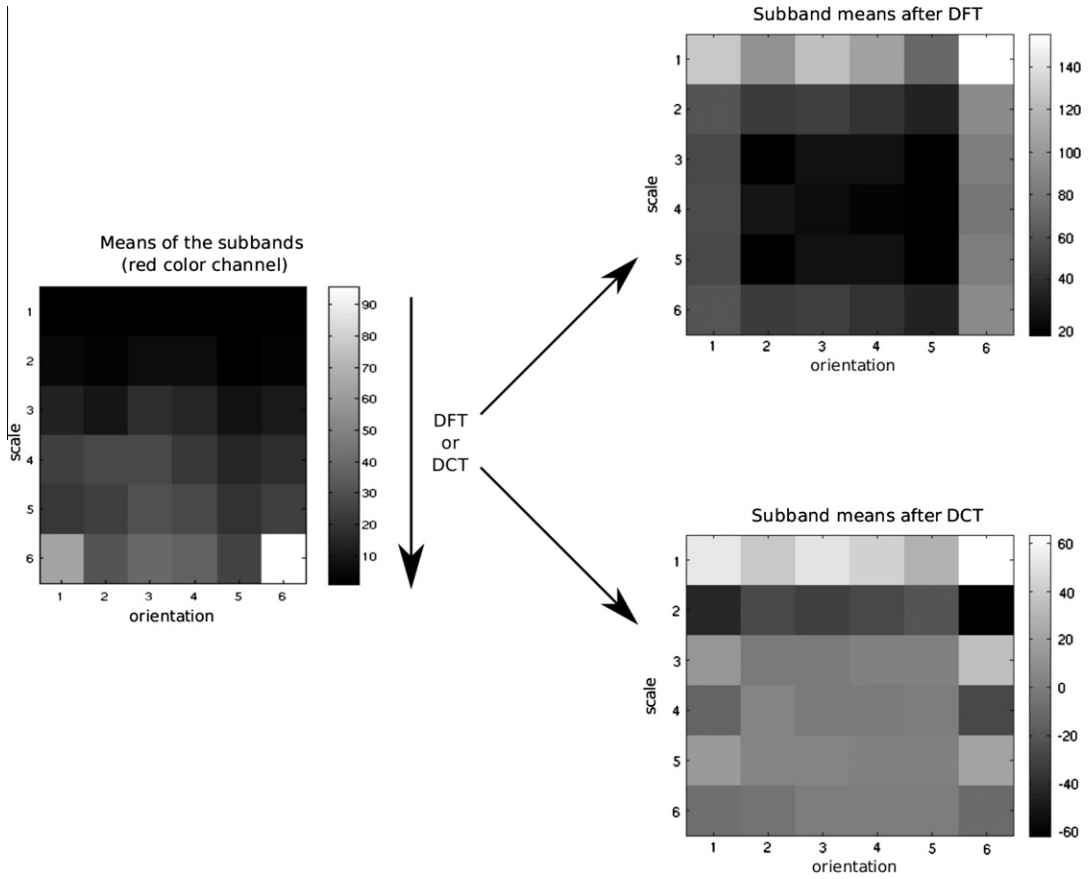

Fig. 5. Computing the discrete cosine transform (DCT) or discrete Fourier transform (DFT) across the scale dimension of the subband means.

3.2. Cyclic shifting of local features

Instead of computing the DFT across the scale dimension of the local feature vectors of the D^{3}T-CWT, these vectors are cyclically shifted across the scale dimension (in the original approach (Lo et al., 2009), the local feature vectors are additionally shifted across the orientation dimension to achieve orientation invariance, but since we are primarily interested in scale invariance we omit that). First we square each element of the local feature vectors and apply subsequently a circular-correlation (only across the scale dimension) of the squared feature vector with a specific mask M (see Fig. 6).

The result of this process is a correlation vector, which is as long as the number of scale levels is. Now the original feature vector is cyclically shifted in the scale dimension, so that the first scale level of the new local feature vector is the scale level of the original local feature vector, in which the correlation vector had its maximum (see Fig. 6). Then the subbands (consisting of the corresponding feature values) are modeled by a Rayleigh distribution (Lo et al., 2009) and the parameters of this distribution are used to form the final feature vector of an image. Since we use only one statistical feature per subband, the number of features per image is half the number of features using the D^{3}T-CWT (216).

3.3. The dominant scale approach

The accumulated energies of the scales $l \in \{1 \dots, L\}$ are computed across the orientations $d \in \{1, \dots, D\}$ (Montoya-Zegarra et al., 2007):

$$E(l) = \sum_d \sum_x \sum_y |S_{l,d}(x,y)|.$$

A scale invariant representation is achieved by computing the dominant scale (DS) of the images followed by feature alignment. The dominant scale is defined as the scale with the highest accumulated energy $E(l)$. Now the feature vector, consisting of mean and standard deviation of the subbands, is circularly shifted, such that the features of the dominant scale are the first ones in the feature vector.

We face a problem when applying this method to our database (which is identical for the original approach (Montoya-Zegarra et al., 2007) using steerable pyramid decomposition and for our version using the DT-CWT). Due to subsampling, subbands at increasing scale have a lower number of coefficients. On the other hand, the absolute values of coefficients in higher scales are distinctively higher than those in lower levels (see the subband means in Fig. 7). Nevertheless, the dominant scale will almost ever occur at scale level $l = 1$ due to the high number of coefficients.

In fact, when using the DT-CWT on our data set, the DS is always at scale level $l = 1$, and for the steerable pyramid decomposition the DS is not at scale level $l = 1$ for 17 images only (out of 612 images). That is why we adapted the dominant scale approach by using subband means instead of the subband energies. Using this approach, for 38 images the DS is not at scale level $l = 6$ which improves the situation only slightly. Feature vector values are almost always monotonically increasing with the scale level (subband means) or monotonically decreasing with the scale level (energy of the subbands) and therefore shifting the features across the scale dimension according to the DS does not make a big difference. The original approach of Montoya-Zegarra et al. (2007) also determines the dominant orientation, but as we are more interested in scale invariance we omit this process (results have been deteriorated when using this approach). The length of the feature vector is equal to that of the DT-CWT (216).

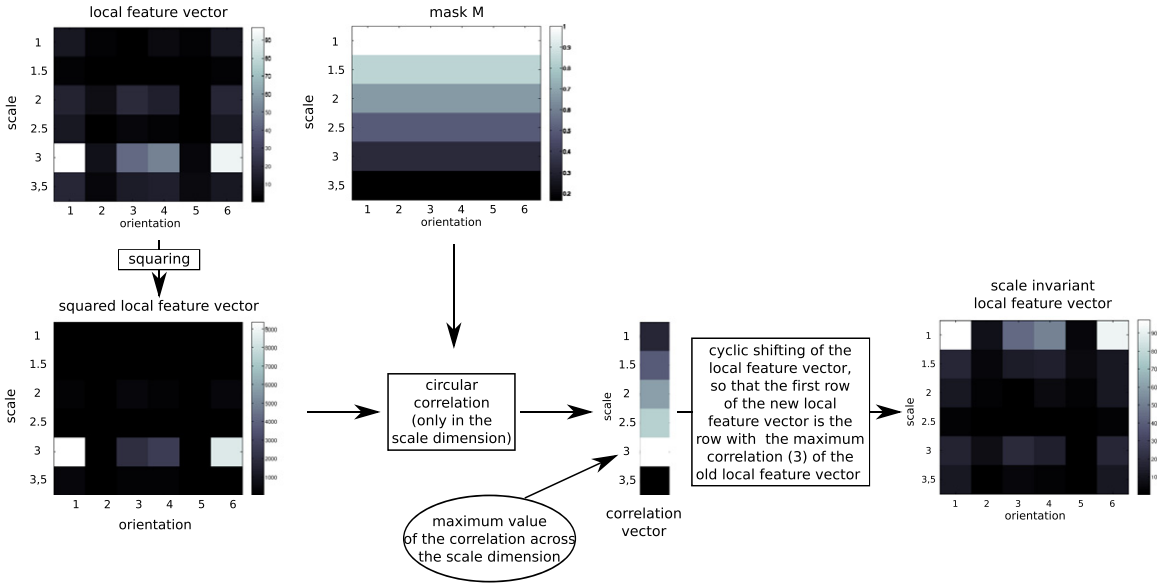

Fig. 6. Cyclic shifting of local feature vectors.

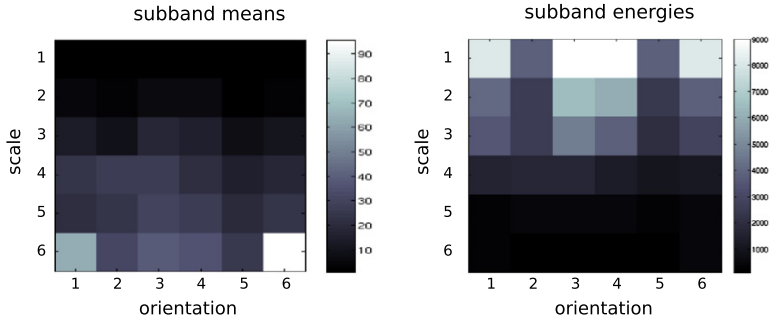

Fig. 7. Comparison of the subband means and energies.

3.4. The slide matching approach

Slide matching (Fung and Lam, 2009) was originally proposed for the Gabor transform but is used with the D^{3}T-CWT in our context. The original approach is first made orientation invariant by summing up the means and standard deviations of the subbands' coefficients with same scale level. In adapting the proposed approach to our scenario (Fung and Lam, 2009), we compute the scale levels 1, 1.5, 2, ..., 6 of the training set and the scale levels 2, 3, 4, 5 of the evaluation set. The distance between an image of the training set and an image of the evaluation set is the distance that is minimized by sliding the feature vectors along the scale dimension against each other (see Fig. 8).

Since we are primarily interested in scale invariance we also use a modified version of slide matching without summing up the subbands of the same scale level. Consequently, we have for each scale level 12 (6 orientations, 2 parameters per subband) instead of 2 features for the slide matching process. Therefore in case of the original version the length of the feature vector is equal to that of the DT-CWT (216) and in case of the modified version, the length of the feature vector is only a sixth of that of the DT-CWT (36).

3.5. The log-polar approach

The log-polar transformation maps points from the Cartesian plane (x, y) to points in the log-polar plane (ξ, η) (see Fig. 9). In this coordinate system, scaling and rotation is converted to translations.

Scale invariance (and orientation invariance) can be achieved by analyzing the transformed image with a shift invariant transform like the adaptive row shift invariant wavelet packet transform (as originally proposed in Pun and Lee (2003)) or the DT-CWT used in this work. Unlike in Pun and Lee (2003), we compute subband means and standard deviations as features (instead of energies) and do not use the best basis algorithm. Obviously, the length of the feature vector is equal to that of the DT-CWT (216).

4. Scale invariant methods based on fractal analysis

For a point set E defined on $\mathbb{R}^2$, the fractal dimension of E is defined as

$$dim(E) = \lim_{\delta \to 0} \frac{\log N(\delta, E)}{-\log \delta},$$

where $N(\delta, E)$ is the smallest number of sets with diameter less than δ that cover E. The set is made up of closed disks of radius δ or squares of side length δ. In Fig. 10 we present some examples for the fractal dimensions of different objects.

Intuitively, the fractal dimension is a statistical quantity that gives a global description of how complex, how irregular or how rough a geometric object is. However, the fractal dimension alone, as defined before, does not provide a rich description. It is just a single value.

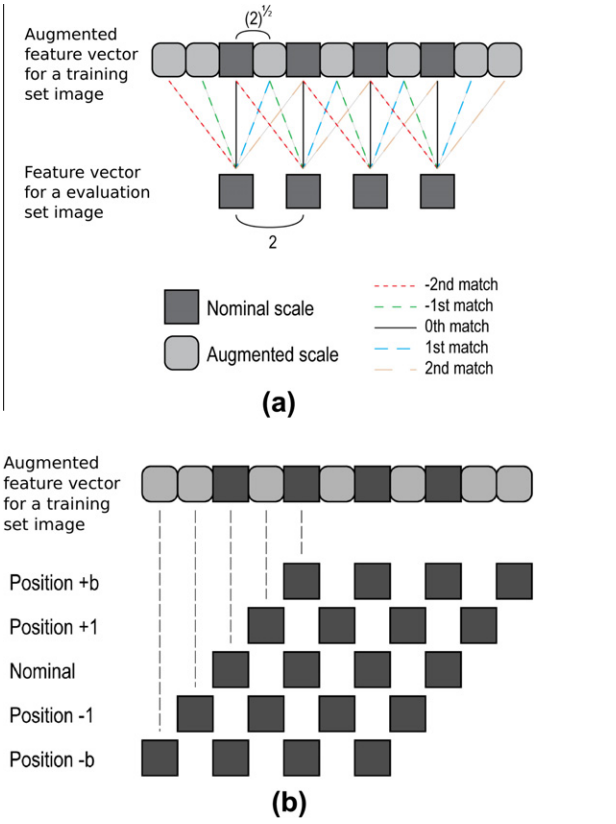

Fig. 8. (a) Different scale factors are used for the training set images and for the evaluation set images. Each node denotes two elements, a sum of means and a sum of standard deviations. (b) The sliding of evaluation set image feature vector along augmented training set image vector.

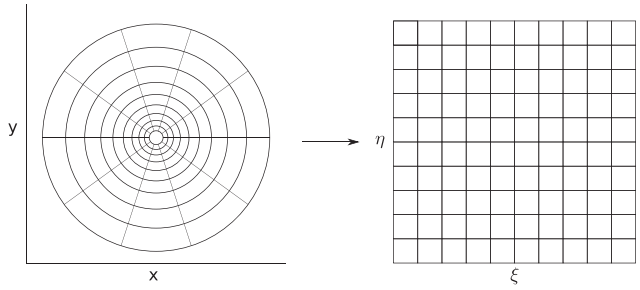

Fig. 9. The log-polar transformation.

The local fractal dimension or also called the local density function, used in two of the three methods presented in this section, provides a more powerful and adaptive description. Let μ be a finite Borel regular measure on $\mathbb{R}^2$. For $x \in \mathbb{R}^2$, denote $B(x,r)$ as the closed disk with center x and radius $r > 0$. $\mu(B(x,r))$ is considered as an exponential function of r, i.e. $\mu(B(x,r)) = c\ r^{D(x)}$, where $D(x)$ is the density function and c is some constant. The local density function (or also called local fractal dimension) of x is defined as

$$D(x) = \lim_{r \to 0} \frac{\log \mu(B(x,r))}{\log r}.$$

The density function measures the "non-uniformness" of the intensity distribution in the region neighboring the considered point.

The local density D is invariant under the bi-Lipschitz map, which includes view-point changes and non-rigid deformations of texture surface as well as local affine illumination changes. A bi-Lipschitz function f must be invertible and satisfy the constraint

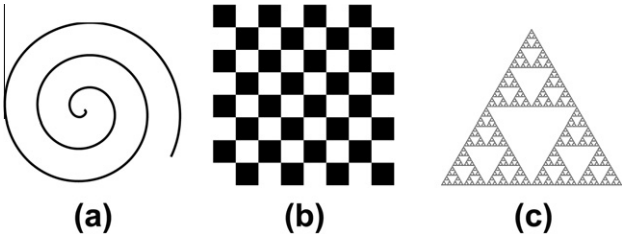

Fig. 10. Fractal dimension D in 2D space. (a) Smooth spiral curve with $D = 1$, (b) the checkerboard with $D = 2$ and (c) the Sierpinski-Triangle with $D \approx 1.6$.

$c_1\|x-y\| \leqslant \|f(x)-f(y)\| \leqslant c_2\|x-y\|$ for $c_2 \geqslant c_1 > 0$. Consequently, local fractal dimensional based approaches are especially interesting for developing scale-invariant feature descriptors, and so also for the classification of celiac disease.

By choosing different measures μ, the local density function can be adapted to different image processing tasks. As we will see, measures based on derivative information work best for our dataset, since their contrast sensitiveness is a good feature to differentiate between images with or without celiac disease. The reason for that is that images of patients with celiac disease have less or entirely no villi and therefore a lower amount of contrast compared to images of patients without celiac disease.

4.1. The multi-fractal spectrum

First, the local fractal dimension is computed for each pixel of an image (Xu et al., 2009b). Let E_α be the set of all image points x with local density in the interval α:

$$E_\alpha = \{x \in \mathbb{R}^2 : D(x) \in \alpha\}.$$

Usually this set is irregular and has a fractal dimension $f(\alpha) = dim(E_\alpha)$.

We denote the convolution $*$ between an image $I = I(x,y)$ and a Gaussian kernel $G_\sigma = G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{x^2+y^2}{2\sigma^2}}$ (i.e. Gaussian blur) as follows:

$$I(x,y,\sigma) = I(\sigma) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x+u, y+v) G(u,v,\sigma)\, du\, dv$$

and $I_x(\sigma)$ is the first derivative of $I(\sigma)$ in the direction of x.

In the original approach (Xu et al., 2009b) three different types of measures $\mu(B(x,r))$ are defined for the computation of the local density:

$$\begin{aligned} \mu(B(x,r)) &= \int_{B(x,r)} I(\sigma)\, dx \\ \mu(B(x,r)) &= \int_{B(x,r)} \sum_{k=1}^{4} (f_k * I(\sigma)^2)^{\frac{1}{2}} dx \\ \mu(B(x,r)) &= \int_{B(x,r)} |(I_{xx}(\sigma) + I_{yy}(\sigma))|\, dx, \end{aligned} \tag{1}$$

where $\{f_k, k = 1,2,3,4\}$ are four directional operators (derivatives) along the vertical, horizontal, diagonal, and anti-diagonal directions. The feature vector of an image I consists of the concatenation of the fractal dimensions $f(\alpha_i)$ for the three different measures $\mu(B(x,r))$.

In case of our dataset, it turned out that the Laplacian measure (Eq. (1)) is the only one of the three measures which leads to sensible results. The reason for that is very probably the comparatively highest contrast sensitiveness of the Laplacian measure. So contrasting to the original proposal (Xu et al., 2009b), the employed feature vector only has entries of the third type. We use 14 non-overlapping intervals α and so the length of the feature vector per image is 14.

4.2. Fractal analysis using filter banks

Instead of partitioning the local densities of images in sets $E(\alpha)$'s and computing their fractal dimensions, we first convolve the images with the MR8 filter bank (Varma and Zissermann, 2005; Geusebroek et al., 2003), a rotationally invariant, nonlinear filter-bank with 38 filters but only 8 filter responses, and compute local fractal dimension afterwards (Varma and Garg, 2007; Uhl et al., 2011b). Filters can smooth over image noise and lead to more robust features. However, they also have the drawback of lowering the level of bi-Lipschitz invariance.

Let us introduce the measures

$$\mu(B(x,r)_i) = \iint_{B(x,r)} |f(i)|\,dx \tag{2}$$

$$\mu(B(x,r)_i) = \iint_{B(x,r)} |(S_x + S_y) * (G_\sigma * f(i))|\,dx, \tag{3}$$

where $f(i)$ is the i-th MR8 filter response image with $1 \leqslant i \leqslant 8$. $S_x = [-1, 0, 1; -2, 0, 2; -1, 0, -1]/4$ and $S_y = -S(x)^T$ are Sobel filters. The first measure (Eq. (2)) is the measure originally proposed in Varma and Garg (2007), while the second measure (Eq. (3)) is proposed in Uhl et al. (2011b), where the original fractal method of Varma and Garg (2007) has been optimized for the celiac disease database. We follow the optimized approach using the second measure and computing the local density for each of the 8 filter responses (in Varma and Garg (2007), only 5 of the 8 filter responses are used, in our experiments the results are better using all 8 responses). For each pixel of an image, we result in an 8-dimensional local density vector. For each class of the training set we aggregate the local density vectors of the images of this class and learn cluster centers (called textons) by k-means clustering.

The next step is to learn models for each image of the training and evaluation sets. Given an image, its corresponding model is generated by first convolving it with the filter bank, computing the local density of each filter response and then labeling each local density vector with the texton that lies closest to it. Distances between two frequency histograms (models) are measured using the χ^2 statistic. The length of the feature vector per image is the number of classes of the according image database multiplied by 10 (10 clusters per class).

4.3. Fractal dimensions for orientation histograms

Similar to SIFT features (see next section), this method (Xu et al., 2009a) is based on computing local orientation histograms. First the gradient magnitude and the orientation of a given pixels neighborhood are computed. The orientation histogram from the neighborhood of the given pixel is formed by discretization of orientations by weighing the gradient magnitude (see Fig. 11). The histogram is then assigned to one of 29 orientation histogram templates, which are constructed based on the spatial structure of the orientation histogram (the number of significant image gradient orientations and their relative positions). We now have for each pixel (for a given neighborhood size) a value between 1 and 29, depending on the template it is assigned to. By setting a pixel to one if it is assigned to template i ($i \in \{1, \ldots, 29\}$) and to zero otherwise, 29 binary images are generated, from which we compute the fractal dimensions (by means of the box-counting method[1]). This process is applied for eight different neighborhood sizes (scale levels). In order to get better robustness to scale changes, finally a wavelet transform (a redundant tight wavelet frame system) is applied across the scale dimension (the different neighborhood sizes) of the fractal dimensions. The final feature vector of an image consists of the detail and approximation coefficients of the wavelet transform. This feature vector can be viewed as the information about the changes with respect to scale (the neighborhood sizes represent the scale levels). According to (Xu et al., 2009a), this enhances the scale invariance, since the scale changes are often consistent across multiple scales for natural textures. The length of the feature vector per image is 1160 (29 orientation histogram templates $\times$ 8 neighborhood sizes $\times$ 5 (2 different high-pass filters each with 2 decomposition levels and a low-pass filter using only the second decomposition level).

[1] rsbweb.nih.gov/ij/plugins/fraclac/FLHelp/BoxCounting.htm.

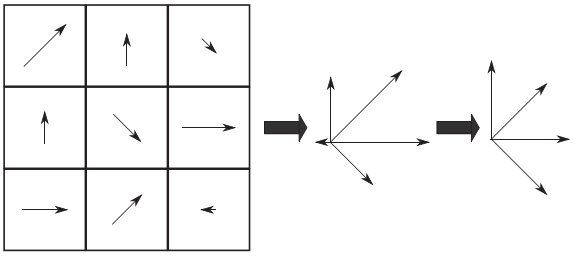

Fig. 11. The process of construction and discretization of the orientation histogram when using a neighborhood of size 3×3.

5. Further approaches

In this section we will present approaches that are neither based on wavelet transforms nor on fractal analysis. The first two approaches are based on the widely used SIFT features (Lowe, 1999) and affine invariant region detectors (Zhang et al., 2006), two approaches work with neural networks (Ma et al., 2010; Zhan et al., 2009) and one approach analyzes characteristics of connected regions (blobs) (Xu and Chen, 2006). Finally we cover the affine invariant Local Ternary Patterns which are based on the analysis of multi-scale second moment matrices in a Laplacian scale space (Hegenbart and Uhl, 2013).

5.1. SIFT features and region detectors

The Scale Invariant Feature Transform (SIFT) (Lowe, 1999) is probably the most popular feature used in computer vision (Vedaldi and Fulkerson, 2008). SIFT detects salient image regions (keypoints) and extracts discriminative yet compact descriptors of their appearance. SIFT keypoints are invariant to viewpoint changes like translation, rotation, and rescaling of an image.

First an image is convolved with Gaussian filters at different scales σ. By means of detecting the maxima/minima of the Difference of Gaussians (DoG), local scale space extrema are found. The DoG is given by

$$DoG(x, y, \sigma) = I(k\sigma) - I(\sigma).$$

Local scale space extrema which have low contrast or are poorly localized are eliminated, the rest are used as keypoints. Using the Gaussian filtered image (with keypoint scale σ), gradient magnitudes and orientations are computed from the neighboring region of the keypoint to form an orientation histogram. Now the gradient information is rotated according to the dominant orientation of the orientation histogram and weighted by a Gaussian function. A local descriptor uses 16 histograms, aligned in a 4×4 grid, each with 8 orientation bins. This results in a feature vector containing 128 ($16 * 8$) elements for each keypoint of an image.

The original approach Lowe, 1999 is suited for object recognition, in this work however we are interested in texture classifications, SIFT keypoints do not make sense in our context. We apply two different ways to deal with that problem:

1. We use dense SIFT features (Fei-Fei and Perona, 2005), that means that we compute SIFT descriptors for each pixel of an image.
2. We use a region detector that is suited for texture images and then apply the SIFT descriptor to the detected regions (Zhang et al., 2006).

A region detector suited for texture recognition is the Harris detector (Lazebnik et al., 2005; Mikolajczyk and Cordelia, 2004). The Harris detector is based on the second moment matrix M_I. This matrix must be adapted to scale changes to make it independent of the image resolution:

$$M_I(\sigma,\gamma) = G_\sigma * \begin{pmatrix} I_x^2(\gamma) & I_x(\gamma)I_y(\gamma) \\ I_x(\gamma)I_y(\gamma) & I_y^2(\gamma) \end{pmatrix},$$

The Harris corner measure μ is defined as follows:

$$\mu(\sigma,\gamma) = \det(M_I(\sigma,\gamma)) - \alpha\ \mathrm{trace}^2(M_I(\sigma,\gamma)).$$

Note that μ simultaneously lives in two scale spaces (caused by the Gaussian kernel G) with parameters γ and σ. The inner scale γ, which is less critical than the outer scale σ, is set to a constant value. Local maxima of this measure determine the location of Harris interest points, and then the Laplacian scale selection procedure is applied at these locations to find their characteristic (outer) scale σ. The Laplacian scale selection finds the characteristic scale at a given point (x, y) by maximizing the Laplacian-of-Gaussian:

$$L(x,y;\sigma) = \sigma^2 |I_{xx}(\sigma) + I_{yy}(\sigma)|. \quad (4)$$

The elliptic region around a found location is described by its principal axes corresponding to the eigenvectors of M_I and axis length depending on the eigenvalues. For affine invariance, a region is normalized by mapping it onto a unit circle and using a rotational invariant descriptor, the SIFT descriptor.

So for both ways, using dense SIFT features or using the Harris detector (combined with the affine invariant mapping and the SIFT descriptor), we get features from the SIFT descriptor as output. For both approaches we follow the strategy applied in Section 4.2, as opposed to the classical dense SIFT approach (Fei-Fei and Perona, 2005). For each class of the training set we aggregate the SIFT descriptors of the images of this class and learn cluster centers (textons) by k-means clustering. Given an image, its corresponding model is generated by labeling its SIFT descriptors with the texton that lies closest to it. Distances between two frequency histograms (models) are measured using the χ^2 statistic. For both approaches, the length of the feature vector per image is the number of classes of the according image database multiplied by 10 (10 clusters per class).

It should be noted that instead of using the Harris detector it would be possible to use other region detectors (e.g. Laplacian (Zhang et al., 2006) and Hessian region detectors (Mikolajczyk and Schmid, 2002)) and descriptors (e.g. SPIN and RIFT features (Zhang et al., 2006)), the principle of the approach however remains the same. Following the terminology in the original papers, we denote the approach using the dense SIFT features as "Dense SIFT Features" and the approach using the Harris detector as "Local Affine Regions".

5.2. Pulse-coupled neural networks based methods

Pulse-coupled neural networks (PCNN's) (Ranganath et al., 1995) are neural models proposed by modeling a cat's visual cortex. PCNN is a neural network algorithm that produces a series of binary pulse images when stimulated with an image. The intersecting cortical model (ICM) (Ma et al., 2010) and the spiking cortical model (SCM) (Zhan et al., 2009) are two methods derived from the PCNN, which are faster and provide better or similar results as compared to the PCNN (Ma et al., 2010; Zhan et al., 2009).

The ICM model consists of two coupled oscillators, a small number of connections and a non-linear function. F is the state oscillator and Θ the threshold oscillator. Together they constitute the neurons pulse sequence Y. The mathematical model of ICM is described as follows:

$$F_{ij}(n) = fF_{ij}(n-1) + I(i,j) + \sum_{kl} M_{ijkl} Y_{kl}(n-1),$$

$$\Theta_{ij}(n) = g\Theta_{ij}(n-1) + hY_{ij}(n-1),$$

$$Y_{ij}(n) = \begin{cases} 1 & \text{for } F_{ij}(n) > \Theta_{ij}(n), \\ 0 & \text{otherwise.} \end{cases}$$

where f, g and h are scalars, $M = [0.5, 1, 0.5; 1, 0, 1; 0.5, 1, 0.5]$ is the connection function through which the neurons communicate, I is the input image and $n \in \{1, \ldots, N\}$. The pair (i, j) stands for the position of the neuron in the map and (k, l) is that of its neighboring neurons. The outputs of ICM are N binary images, which represent features like texture, edges, and segments.

The mathematical model of the SCM is described as follows:

$$F_{ij}(n) = fF_{ij}(n-1) + I(i,j) + I(i,j)\sum_{kl} M_{ijkl} Y_{kl}(n-1),$$

$$\Theta_{ij}(n) = g\Theta_{ij}(n-1) + hY_{ij}(n-1),$$

$$Y_{ij}(n) = \begin{cases} 1 & \text{for } F_{ij}(n) > \Theta_{ij}(n), \\ 0 & \text{otherwise.} \end{cases}$$

As for ICM, the outputs of SCM are N binary images. The final feature vectors of the SCM and ICM, respectively, consist of the entropies of the $N = 37$ binary output images.

The authors (Ma et al., 2010; Zhan et al., 2009) state that their approaches (ICM and SCM) are scale invariant (and rotation and translation invariant), however their manuscripts miss a valid justification for this statement. They reference a further publication (Johnson, 1994) in which scale invariance is explained. The problem is that there a special kind of PCNN is considered and that scale invariance is only shown for objects on a uniform background, not for textures.

5.3. Multiscale blob features

In order to derive multiscale blob features (Xu and Chen, 2006), we apply a series of flexible threshold planes to a textured image and then use the topological and geometrical attributes of the generated blobs in the obtained binary images to describe image texture. The flexible threshold planes FP are determined by Gaussian blurring:

$$FP(x,y;\sigma,b) = b + I(x,y,\sigma),$$

where σ^2 is the variance to control the spread of the window and b is the bias. By applying the flexible threshold planes to the grayscale image I, we obtain binary images

$$g_b(x,y;\sigma) = \begin{cases} 1 & \text{if } I(x,y) > FP(x,y;\sigma,b), \\ 0 & \text{otherwise.} \end{cases}$$

In each binary image, all 1-valued pixels and 0-valued pixels are grouped into two sets of connected regions called blobs (see Figs. 12 and 13).

The original approach uses two features to describe an image, the number of blobs and the shapes of the blobs. The shape feature indicates how compact the blobs of an image are (the compactness of a blob is here defined as the maximum of the distances from the

pixels of a blob to the centroid of the blob, divided by the square root of the number of pixels of the blob). The multiscale blob features are invariant to rotation and gray-level scaling (the bias b of $FP(x, y; \sigma, b)$ is depending on the standard deviation of the input image). The shape features are invariant to spatial scaling within a small range (the compactness is similar for spatial scaling within a small range), but the number of blobs change to some extent. As opposed to the original approach (Xu and Chen, 2006), we separately classify the images for the two blob features, since the shape feature is scale invariant and the number of blobs is not. The length of the final feature vector is 480 (30 values for $\sigma \times 8$ values for $b \times 2$ (black and white regions of a binary image)).

5.4. Affine invariant local ternary patterns

The Local Binary Pattern approach as introduced by Ojala et al. (1996) as well as the Local Ternary Patterns method proposed by Tan and Triggs (2007) are not affine invariant. An extension to the method suggested by Mäenpää (2003) uses multiple Gaussian filters with varying sizes to improve the support area of the operator. This extension adds multi resolution to the operator but misses a scale selection mechanism. We propose an affine invariant method based on Local Ternary Patterns that employs scale-normalized derivatives of local scale space maxima for scale selection. We compute the multi-scale second moment matrices at given scales to analyze textures according to their affine shape along an ellipse. The method shares the idea of using the scale space framework with methods such as the SIFT feature detector and other region detectors as discussed in Section 5.1. The idea of combining scale space maxima with Local Binary Patterns has also been explored by Li et al. (2012).

Instead of using the DoG (difference of Gaussian) approximation to the Laplacian of Gaussians as used by SIFT we construct the scale space by computing the scale-normalized Laplacian (see Eq. (4)) of each image I at each location $x \in \mathbb{N}^2$ at different scales with $\sigma = \frac{1.5^k}{\sqrt{2}}$, $k \in \{1, \ldots, 20\}$ denoted as $(\bar{\Delta} I(x; \sigma))$. The initial scale is chosen such that it corresponds to the standard radius of LBP (1.5).

Due to the fact that not all locations in an image attain a local maximum and a maximum between scales, we compute a scale mask to improve the reliability of the scale estimation. This is especially useful when textures are not strictly periodic and attain multiple scales as is the case in celiac disease. The computation involves the detection of local maxima at each scale. We exploit the fact that pixels in close spatial proximity to a maximum are at the same or a relatively close scale to the detection scale of the corresponding maximum.

We compute the multi-scale second moment matrices at each location x of an image I which is attaining a local scale space maximum. We use the detection scale of the maximum as the local scale t, the integration scale $s = \sqrt{2}t$ is depending on the detection scale.

$$\mu(x; t, s) = \int_{\xi \in \mathbb{R}^2} (\nabla I)(x - \xi; t)(\nabla I)^T (x - \xi; t)\; g(\xi; s)\, d\xi.$$

With $(\nabla I)(x; t)$ denoting the gradient of the scale space orientation at scale t and g denoting a Gaussian function. The second moment matrix summarizes the gradient distribution of the area around a pixel location at a given scale. The eigenvalues of the matrix characterize the length of the axes of an ellipse (up to some constant multiplier) while the eigenvectors describe the orientation of the axes. Due to the fact that the orientation of the ellipse described by the second moment matrix is normal to the detected blob we compute the inverse of the second moment matrix. The inverse results in a rotation by ninety degrees without modifying the ratios of the axis lengths.

The absolute sizes of the axes given by the second moment matrices are unknown, we therefore normalize the ellipses such that the circumference is equal to the circumference of the detected maxima treated as circle (the radius at scale σ is $\sqrt{2}\sigma$). To do so, we apply Ramanujan's formula for approximating the circumference of the ellipse (with axes a and b) and solve the quadratic equation for a constant scaling factor c

$$\sqrt{2}\sigma 2\pi = \pi\left(3(ac + bc) - \sqrt{(3ac + bc)(ac + 3bc)}\right).$$

The axes of the ellipse are then re-scaled by the appropriate solution of c.

After the computation of the support area of a local maximum (the ellipse given by the second moment matrix at the position and scale of the maximum), all locations within this area are assigned to the scale and response of that maximum (we call this a corresponding maximum for a location). In our methodology it is possible that a single location contains multiple possible scales and responses, this is due to the fact that the support areas of multiple maxima might overlap.

We observed that the reliability of a location attaining the same scale as a corresponding maximum decreases with spatial distance. To compensate for this, we compute the reliability of a scale at a distance d from a corresponding maximum using a Gaussian probability density function choosing σ_p such that the reliability of a given scale at a distance of the length of the semi minor axis b of the corresponding maximum's normalized support area is 0.5. We additionally use the responses of all corresponding maxima for a pixel location to ensure that maxima with lower responses have lower reliabilities.

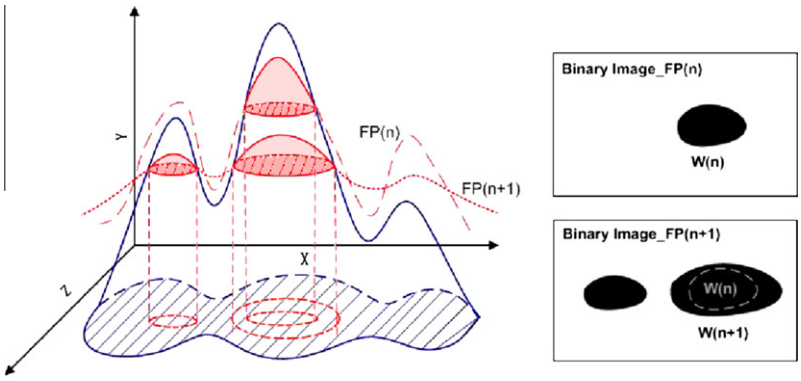

Fig. 12. Process of extracting binary blobs.

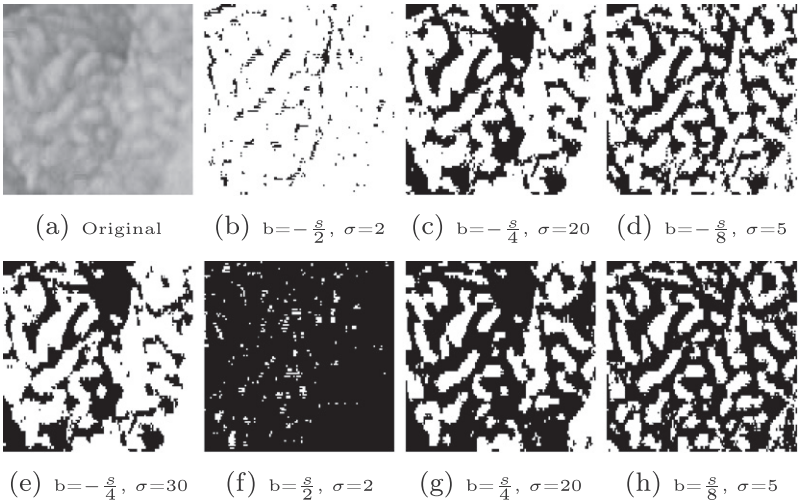

Fig. 13. Binary blob images with different sigmas and biases denoted by b, where s denotes the standard deviation of the original image. 0-valued pixels are displayed as black pixels and 1-valued pixels are displayed as white ones.

$$r(x;l) = \frac{\bar{\triangle} I(x;l)}{\max_t \bar{\triangle} I(x;t)} e^{\frac{-d^2}{2\sigma_p^2}} \quad \text{and} \quad \sigma_p = \left(\frac{-\left(\frac{b}{2}\right)^2}{2\ \log(0.5)} \right)^{\frac{1}{2}}.$$

The reliability measure is finally used to assign a weight controlling the contribution of each computed pattern to the histogram. Once the scales of each location have been determined we apply an adaptive Gaussian filter to the data, prior to computing the LTP code at a location. We select the width of the Gaussian filter such that the area covered by the operator in relation to the local scale is the same across all scales. This gives invariance to uniform scaling. The width of the Gaussian filter (f_w) is selected in a way that 90% of the area of the Gaussian function are in the area of the computed filter.

$$f_w = \frac{\sqrt{2}\sigma 2\pi}{n} \quad \text{and} \quad \sigma_{\text{Gauss}} = \frac{f_w\ 0.5}{\sqrt{2}\ \text{erf}^{-1}(0.9)}.$$

For n being the number of considered LTP neighbors, σ being the scale at the location.

To compute a pattern at a location we estimate the second moment matrices using the detection scales of all corresponding maxima at that location. We distribute the sample points of the operator such that they lie along the normalized ellipses described by the second moment matrices. By using this approach non uniform scaling of the data can be compensated, because this type of transformation would change the shape of the ellipses accordingly. We then distribute sample points so that the distance in terms of arc length between adjacent points is equal, giving n-equidistant points along the ellipse. To speed up the computation we define four support points on the ellipse which lie on the ends of the major and minor axes respectively. The definition of support points limits the method to distribute a number of $4N + 4$ equidistant points along the ellipse but reduces the computation to N points. We use the fact that all ellipses can be described as a scaled and rotated version of a canonical ellipse. To distribute the points on a canonical ellipse in parametric form, the positions of N points in the first quadrant are computed and symmetries are exploited to gain the other $3N$ points. To find the offset on the x-axis of the n-th point (Δx_n) from the center of the ellipse the equation

$$\frac{n}{N+1} \int_0^a \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = \int_0^{\Delta x_n} \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$$

is solved for Δx, where a is the length of the horizontal semi-axis, N is the number of points to distribute per quadrant and the second additive term is the derivative of the canonical implicit equation of an ellipse.

The definition of support points also provides the possibility of defining a fixed starting point for the computation of the patterns. Due to the ambiguous orientation of an ellipse we define two starting points (computing two patterns per position) to compensate. These are by definition the points on the intersection of the major axis with the ellipse.In case of ellipses that are close to a circle this definition becomes unreliable, we therefore treat second moment matrices with a ratio of eigenvalues $\frac{\lambda_{\min}}{\lambda_{\max}} \geqslant 0.95$ as a circle treating the vertical axis as the major axis. By defining a starting point we are able to compensate rotations, this is due to the fact that this kind of affine transformation is reflected by the orientation of the computed ellipses. Please see Hegenbart and Uhl (2013) for a more thorough explanation of the method. The feature vector of an image consists of a single histogram with 59 bins.

6. Experimental results

We use the software provided by the Robotic Research Group[2] for region detection (Harris detector) and description (SIFT) in Section 5.1, the VLFEAT implementation (Vedaldi and Fulkerson, 2008) for the dense SIFT features in Section 5.1 and the implementation of Geusebroek et al. (2003) for the MR8 filter in Section 4.2. For the remaining algorithms custom implementations from earlier work (Kwitt and Uhl, 2007; Uhl et al., 2011b) (DT-CWT, Fractal Analysis using Filter Banks) or specifically developed for this work have been used (all using Matlab except for the affine invariant LTP method which we developed using Java).

The original manuscripts employ a wide variety of different classifiers. Since higher developed classifiers (e.g. the SVM classifier) will mostly produce better results than more simple classifiers (e.g. the k-NN classifier) and since the focus lies on scale invariant feature extraction strategies and not on classification methods, all methods are classified using the k-NN classifier. The advantage of that approach is the better comparability of the results with respect to feature expressiveness.

Classification accuracy is computed using an evaluation set and a training set. An image from the evaluation set is classified into

[2] http://www.robots.ox.ac.uk/vgg/research/affine.

the class, where most of the k nearest neighbors from the training set belong to. The k for the k-NN classifier, used to classify the evaluation set, is optimized on the training set (the k with the highest overall classification rate (OCR) using leave-one-out cross-validation (LOOCV) on the training set).

The algorithms using k-means clustering provide different results each run. For these methods we provide average results from 10 runs per method.

6.1. Celiac disease

We use a database of duodenal endoscopic images employed in earlier work (Hegenbart et al., 2011) to enable easier comparison. Table 1 lists the number of image samples and patients per class. To avoid overfitting and to test the methods in a practice-related context, the images of a patient are either all in the evaluation set or all in the training set. In this way it is impossible that the nearest neighbors of an image and the image itself come from the same patient. This is important to avoid any bias in the result, the setup of this data set resembles in a way how LOPO (Leave-one-patient-out) and LOOCV (Leave-one-out cross-validation) work.

The original endoscopic images (which are of size 620×530 or 520×510 depending on the used endoscope) often exhibit only small areas that permit a distinction between healthy mucosa and mucosa affected by celiac disease. This is due to the facts, that endoscopic images in general show a high amount of distortions such as bubbles, specular reflections and occlusions due to the geometric properties of the duodenum. Additionally, the distribution of villous atrophy caused by the disease could be restricted to certain areas within the visible area (this is known as patchy distribution of celiac disease). Therefore we extract non overlapping patches of size 128×128 (under supervision of a physician). Our celiac disease database consists of these patches.

We observed that the overall classification rate (OCR) varies significantly depending on the chosen number of nearest neighbors of the k-NN classifier. Therefore, we use a second measure for the 2-class case to evaluate the methods, the area under the ROC (receiver operating characteristic) curve (AUC) (Bradley, 1997). We generate the ROC curve by considering the class membership of the 20 nearest neighbors for each image of the evaluation set, where the area under the ROC curve is calculated by trapezoidal integration (Bradley, 1997). The AUC uses the information how many of the 20 nearest neighbors of each evaluation set image are positive (celiac disease) or negative (healthy), whereas the OCR only uses the information if more or less of the k nearest neighbors are positive than negative.

The results in Table 2 are sorted according to the OCR results of the 2-class case. In the last four rows of the table we display methods not designed to be scale invariant, DT-CWT, D^{3}T-CWT and two earlier results using the same database (Hegenbart et al., 2011). One method is the original LBP approach, the other one, denoted "WT–LBP", is the best performing approach for this dataset so far. (Except for the approach 'Fractal Analysis using Filter Banks', which is proposed in Uhl et al. (2011b).) WT–LBP is a combination of Local Binary Patterns and the discrete wavelet transform (for details see Hegenbart et al., 2011).

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

Data set: Marsh type	0	3a	3b	3c	Total
Training set					
Number of images	155	50	56	51	312
Number of patients	66	6	7	8	87
Evaluation set					
Number of images	151	45	58	46	300
Number of patients	65	5	6	8	84

Table 2
Results of the different methods in OCR (%) and AUC. In the 4-class case we only present the OCR.

Method	2-Class case		4-Class case
	OCR	AUC	OCR
Fractal analysis using filter banks	91.7	95.0	65.8
Multi-fractal spectrum	89.0	90.5	62.0
D^{3}T-CWT with DCT	88.3	92.9	63.0
Multiscale blob features (number)	86.3	89.9	57.7
DT-CWT with DCT	86.0	92.7	63.0
Affine invariant LTP	85.6	92.4	61.3
Fractal dim. for orientation histograms	84.0	90.6	62.7
Dense SIFT features	83.6	87.5	62.0
D^{3}T-CWT with DCT (local)	82.3	89.3	60.0
Cyclic shifting of local features	81.0	88.4	61.7
Log-polar approach	80.0	86.9	57.0
Dominant scale approach	78.3	87.5	56.7
D^{3}T-CWT with DFT (local)	78.3	86.4	55.7
Slide matching (original)	76.3	81.7	57.3
Slide matching (modified)	74.7	85.5	62.3
Local affine regions	70.9	88.1	56.3
Multiscale blob features (shape)	70.8	76.2	54.3
ICM	67.7	71.7	52.3
D^{3}T-CWT with DFT	66.0	70.2	50.0
SCM	64.0	66.4	51.3
DT-CWT	84.7	90.2	60.3
D^{3}T-CWT	82.3	90.1	58.0
WT-LBP	88.0	–	63.7
LBP	84.0	–	61.4

The two fractal methods using the local density function perform best for our celiac disease database, especially "Fractal Analysis using Filter Banks" works very good. Also the second fractal method ("Multi-Fractal Spectrum") performs reasonably well. However, the AUC of the latter fractal method is not high compared to other methods. This is because this method has the highest OCRs when we consider many nearest neighbors ($30 \leqslant k \leqslant 70$ in the kNN classifier), while the AUC only uses information of the 20 nearest neighbors (all other methods have their highest OCRs for ks between one and thirty). The third method using fractal features, "Fractal Dimensions for Orientation Histograms", also provides useful results. Overall, the considered fractal methods are quite well suited for classifying celiac disease.

When we consider the results of different strategies for achieving scale invariance using the DT-CWT or the D^{3}T-CWT, we see that DCT computed across the scale dimension of the statistical subband features (DT-CWT and D^{3}T-CWT with DCT) can clearly enhance the results compared to the DT-CWT or the D^{3}T-CWT without any further feature manipulation. All other modifications of the DT-CWT or D^{3}T-CWT decrease the results. The results of the methods, where operations for achieving scale invariance are applied to the local subband coefficients of the D^{3}T-CWT ("D^{3}T-CWT with DCT (local)", "D^{3}T-CWT with DFT (local)", and "Cyclic Shifting of Local Features"), are in the middle of the results range and give pretty similar OCR. The results of the methods, where operations for achieving scale invariance are applied to the global statistical subband features (e.g. mean and standard deviation) of the D^{3}T-CWT, differ a lot. Some are better than their local counterparts ("DT-CWT and D^{3}T-CWT with DCT"), some are worse ("Slide Matching" and "D^{3}T-CWT with DFT"), while the others ("Log-Polar" and "Dominant Scale Approach") give comparable OCR.

"Multiscale Blob Features" using the scale dependent number of blobs as feature works well whereas using the scale invariant shape of the blobs as feature did not provide useful rates for classifying celiac disease. This result, together with the well

performing DT-CWT and D^{3}T-CWT techniques without any technique for further scale invariance being applied, questions the importance of scale invariance in general for our dataset. "Dense SIFT Features", which do not use any keypoint selection, provides a clearly better result compared to "Local Affine Regions" using a keypoint selection strategy specifically tuned for textured data. "Affine Invariant LTP" shares the same idea of key point detection with "SIFT" and "Local Affine Regions", the computed scales mask however increases the reliability in case of non periodic textures. Overall it provides a better performance as compared to these two methods.

The results of the neural nets approaches ("ICM" and "SCM") and the two slide matching variants are not competitive at all.

If we compare results of the 4-class case to the 2-class case, then we see that the differences among the results are smaller in case of the 4-class case. The ranking among the different approaches is similar to the 2-class case. Overall, the OCRs in the 4-class case are not suited for any application scenario and do not improve over earlier work.

Having applied DT-CWT and D^{3}T-CWT instead of the originally proposed transforms for several techniques, we shed light on the reason for this decision. In Table 3 we show classification results of the wavelet based methods, which originally did not use CWTs. We compare the OCRs (2-class case) of the methods using the originally proposed wavelet transforms with the results using DT-CWT variants instead of the original transforms.

As we can see in Table 3, using CWTs works distinctly better for classifying celiac disease as compared to the originally proposed transforms.

Finally, we want to assess statistical significance of our results. The aim is to analyze if the images from the celiac disease database are classified differently by the various methods considered or if all techniques fail for the same set of images. We use the McNemar test (McNemar, 1947), to test if two methods are significantly different for a given level of significance (α) by building test statistics from incorrectly classified images. Tests were carried out for the 2-class case with three different levels of significance ($\alpha = 0.05$, $\alpha = 0.01$ and $\alpha = 0.001$). Results are displayed in Fig. 14 (the methods are sorted according to the OCR results of the 2-class case), where we can observe, that methods with similar OCRs are never found to be significantly different. Fig. 14 shows that only methods with clearly different OCR results are rated as significantly different. That indicates that for methods with similar OCR results, almost the same images are classified wrong, independent of the extracted features.

6.2. Testing scale invariance explicitly

We have employed a set of methods, explicitly introduced to provide scale invariance, motivated by the observation that our celiac disease database contains features at various scales. We want to investigate if these methods are really as scale invariant as they theoretically should be. Further, we want to assess if the techniques' scale invariance really enhances the results for detecting celiac disease, or if the obtained results depend primarily on the general feature extraction ability, independent of scale invariance properties.

The training and the evaluation sets of the celiac disease database both contain images with features at different scales (as well as various orientations, brightnesses and viewpoints). Since each class in the training or evaluation sets has at least 45 images, for almost every image in the evaluation set there might exist images of the same class in the training set with rather similar scales. That means, that a technique does not necessarily have to be scale invariant to work well on our dataset. Therefore, for assessing the scale invariance of a method it is not adequate to test if the method works well for a database containing images with various scales. We need to use two databases, where one database contains differently scaled images as compared to the other.

A further problem of testing scale invariance of the employed approaches with the celiac disease database is that we do not have the information which actual spatial scale an image belongs to (i.e. the distance and perspective of the camera to the mucosal wall) therefore it is difficult to separate the database into two disjoint sets depending on the scale of the images. Another possibility to get two data sets with different scales would be to synthetically scale the database, but this changes the characteristics of the images too much (e.g. interpolation effects, eventual contrast changes, etc.).

Table 3
Results of the wavelet based methods in % (2-class case). The column "original" shows the results using the originally proposed wavelet transforms, the column "CWT" shows the results using CWTs instead of the original transforms.

Methods	Original	CWT
Log-polar approach	58.0	80.0
Dominant scale approach	74.3	78.3
Slide matching approach	74.0	76.3

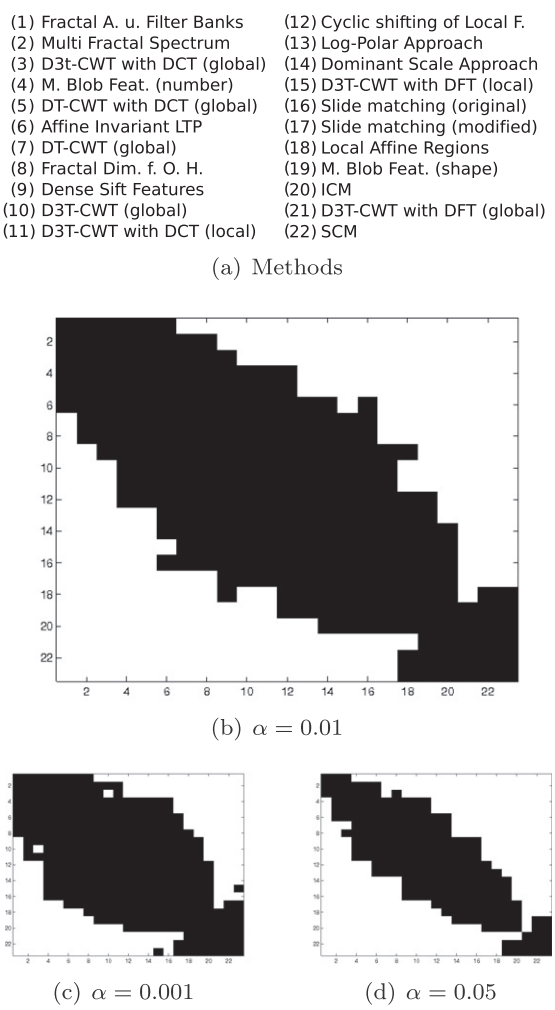

Fig. 14. Results of the McNemar test for the 2-class case. A white square in the *i*th row and *j*th column or in the *j*th row and *i*th column of a plot means that the *i*th and the *j*th method are significantly different with significance level α. If the square is black then there is no significant difference between the methods. The methods are sorted beginning with the best ones (OCR 2-class case) like in Table 2.

We solved this problem by extracting patches from frames of endoscopy videos instead of extracting them from endoscopic images (like done for the celiac disease database). Since it is possible to choose any (suitable) frame of a video from an endoscopy session, it is easier to find patches with a specific distance to the mucosal wall as compared to choosing the patches of some images taken during the endoscopic session. Additionally it is easier to estimate distances in a video than estimating them by means of single images.

As a second texture database to verify the scale invariance of the employed methods, we use parts of the CUReT database (Dana et al., 1999). The advantage of this database compared to our celiac disease scale database is that we have the exact information to which scale an images actual belongs to and that images of one texture class are gathered under exactly the same scale conditions.

That is why we decided to test scale invariance by using two different databases, the celiac disease scale database and parts of the CUReT database.

6.2.1. CUReT database

The cropped version of the CUReT database[3] contains 92 images per texture with different viewing and illumination conditions. There are four texture classes from the CUReT database (material numbers 2, 11, 12, and 14), for which additional scaled data is available (as material numbers 29, 30, 31, 32). The scale difference between these two sets is approximately 1.7. These materials are shown in Fig. 15. The material classes are evenly divided into one part for the evaluation set and one part for the training set, where the images of 46 viewpoint and illumination conditions are used for the training set and the images of the remaining 46 viewpoint and illumination conditions are used for the evaluation set.

For explicitly testing scale invariance, two experiments are performed following ideas in Varma and Zisserman (2009). In the first experiment (E1), the training set consists of original textures (4×46 images of material numbers 2,11,12, and 14, each with 46 different viewpoint and illumination conditions), while the evaluation set consists of original textures and scaled versions of the original textures (8×46 images, images of material numbers 2, 11, 12, 14, 29, 30, 31 and 32 with the remaining 46 viewpoint and illumination conditions). For this experiment, scale invariance is obviously crucial since half of the evaluation set consists of data scaled differently than the data in the training set. In the second experiment (E2), the evaluation set is like in the first experiment, but this time the training set consists of original textures as well as scaled versions of the original textures. The lower the difference between the classification results of the first and the second experiment, the higher the scale invariance of a method is.

The classification results are shown in Table 5.

6.2.2. Celiac disease scale database

The celiac disease scale (CDS) database consists of patches extracted from endoscopy videos. To determine the scale invariance of the employed approaches, we divided the patches into the two categories "Regular" and "Far", depending of the distance to the mucosa wall. Images of the category "Regular" have optimal distances to differentiate between "healthy" and "affected" tissue. Because of the larger distances, the differentiation between the two classes is harder for images of the category "Far". The assessment of distance was performed manually based on the visibility of features (there is no ground truth about the actual distance of the endoscope to the mucosal wall).

We only used images of sequences showing the same mucosal area at a regular distance as well as at a further distance (like done in Hegenbart et al. (2012)). That means for each extracted image of regular distance, we extracted exactly one image with further distance (and vice versa). Similar to the CUReT database, the CDS database consists of a training set (named training set "Regular–Far"), consisting of images with far and regular distances, a second training set (training set "Regular") consisting of images with only regular distances (the images of training set *Regular–Far* with regular distance), and an evaluation set consisting of images with regular and far distances (evaluation set "Regular–Far") (see Fig. 16).

In parallel to the celiac disease database, the images of the training sets are gathered from different patients as of these contained in the evaluation set. Table 4 lists the number of image samples and patients per class.

For explicitly testing scale invariance we perform two experiments. In the first experiment, we use training set *Regular–Far* and evaluation set *Regular–Far*. In the second experiment, we use training set *Regular* and evaluation set *Regular–Far*. Similar to the CUReT database, scale invariance is only needed for the second experiment and not for the first, since only in the second experiment the training and evaluation set are gathered under different scale conditions. The Classification results are shown in Table 5. The lower the difference between the classification results of the first (E3) and the second experiment (E4), the higher is the scale invariance of a method.

6.2.3. Results of testing the scale invariance

The presented results in Table 5 are the mean values of the results using a k-NN classifier with $k = 1$–20. In that way we balance the problem of varying results depending on the number of nearest neighbors of the k-NN classifier. The lower the difference between the classification results of experiment 1 (E1) and experiment 2 (E2) respectively experiment 3 (E3) and experiment 4 (E4), the higher is the scale invariance of a method.

The methods showing the highest degree of scale invariance for a database are marked with a "+", the ones showing the least degree of scale invariance are marked with a "−", and the methods showing average scale invariance are marked with a "○". The ones that are hard to interpret, because they even do not work without scale changes (E1 or E3), are marked with a "?".

Results shown in Table 5 are quite unexpected, especially the ones of the CUReT database.

The absolute OCR results of the first (E1 respectively E3) and second experiments 2 (E2 respectively E4) are not relevant for us, but the differences between them, indicating the extent of scale invariance, are very interesting.

In case of the CDS database, some results are hard to interpret (the two slide matching approaches, D^3T-CWT with DFT and SCM), because even the results without scale changes (E3) are pretty near to the results of randomly classifying images (50%).

In case of the CUReT database, the three methods using fractal analysis are rated as not scale invariant (except of "Fractal Analysis using Filter Banks", which is rated as average). In case of the CDS database all three methods are rated as scale invariant. So the ratings with respect to scale invariance of the CUReT database are contrary to those of the CDS database.

When we consider the methods based on DT-CWT or D^3T-CWT, we also see a clear difference between the two databases. In case of the CUReT database the original (not scale invariant) approaches (DT-CWT and D^3T-CWT) are more scale invariant than their variations (except of the Dominant Scale Approach). In case of the CDS database, the methods applying a transformation to the local wavelet features or shifting them across the scale dimension (D^3T-CWT with DCT (local), D^3T-CWT with DFT (local) and Cyclic Shifting of Local Features) provide more scale invariance than the original approaches. The methods which apply the transformation to global wavelet features or shift them across the scale dimension

[3] www.robots.ox.ac.uk/vgg/research/texclass/data/curetcol.zip.

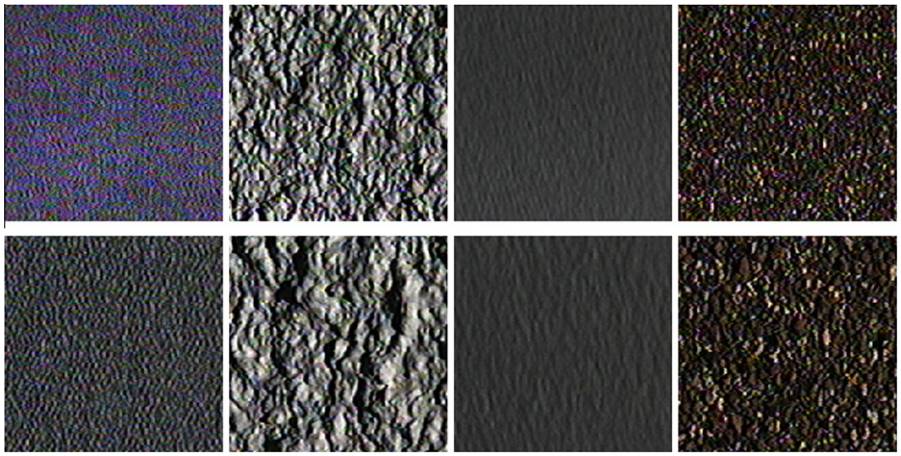

Fig. 15. The top row shows one image each from material numbers 2, 11, 12, and 14 from the CUReT database, while the bottom row shows the textures with higher zoom factor (as material numbers 29, 30, 31, and 32).

Fig. 16. Example images of the CDS database gathered from regular and further distances.

Table 4
Number of image samples and patients of the CDS database.

Data set: Class	Healthy	Celiac disease	Total
Training set Regular–Far			
Number of images	40	40	80
Number of patients	20	12	32
Training set Regular			
Number of images	20	20	40
Number of patients	20	12	32
Evaluation set Regular–Far			
Number of images	38	38	76
Number of patients	19	10	29

are either average scale invariant (DT-CWT with DCT and D^3T-CWT with DCT), not scale invariant (Dominant Scale Approach) or they are hard to interpret, because already the results without scale change (E3) are rather low (the two slide matching approaches and D^3T-CWT with DFT). The Log-Polar-Approach turned out to be scale invariant for both databases.

The methods ICM and SCM are rated as average scale invariant or unratable for both databases.

The Multiscale Blob Features are rated as quite scale invariant (the shapes of the blobs) or as average scale invariant (the number of blobs) in case of the CUReT database (corresponding to the theoretical considerations). But in case of the CDS database they are both rated as not scale invariant (especially when using the shape of the blobs).

The Dense SIFT Features are for both databases more scale invariant as compared to the Local Affine Regions. In case of the CUReT database both methods are rated as average scale invariant, in case of the CDS database the Dense SIFT are rated as quite scale invariant and the Local Affine Regions are rated as not scale invariant.

The Affine Invariant LTP is rated as quite scale invariant for both databases.

Overall, the results of the two databases with respect to the scale invariance are quite different. Of course the two databases for testing the scale invariance are very different.

The intra-class variability of the CUReT database is significantly smaller than those of the CDS database. The visual distinction between images with or without celiac disease is quite different, since there are many images that do not look like typical representatives of their class or they even look like belonging to the other class. In case of the CUReT database the visual distinction between the classes is quite easy. Another difference between the two databases is, that the images of the CUReT database are much more homogeneous than those of the CDS database (an image of the CUReT database looks similar at different positions of the image, that is usually not the case for images of the CDS database). One additional problem is, that one of the most important feature to differentiate between the two classes of the CDS database, the villi, are often less visible at bigger distances of the endoscope to the mucosal wall. This complicates the differentiation between images showing healthy mucosa from further distances to those showing celiac disease affected mucosa from closer distances Hegenbart et al. (2012).

Because of these big differences between the two databases, there are features proving to be more scale invariant for one database than for the other. The scale invariance of the extracted features of a method vary with the application of the method.

7. Conclusion

It seems that especially contrast sensitive methods work very well for the celiac disease database, specifically the fractal methods. The "Multi-Fractal Spectrum" is originally using a combination of three different measures ($\mu(B(x,r))$), but we only use the Laplacian measure, which is the most contrast sensitive of the three. The second fractal method, "Fractal Analysis using Filter Banks" behaves similarly. Other contrast sensitive methods are the third fractal method "Fractal Dimensions for Orientation Histograms" and the method "Multiscale Blob Features (number)", both methods performed well for our celiac disease database. The affine invariant LTP method performed comparably to the best methods.

Table 5
OCR results for the CDS and CUReT database.The columns "E1" and "E3"show the results of the experiments using same scale levels in training and evaluation set, and the columns "E2" and "E4" show the results of the experiments using different scale levels in training and evaluation set. The columns "Diff" show the relative differences between the results of E1 and E2 respectively E3 and E4. The column "SI" rates the scale invariance of the methods as high (+), low (−), average (∘) or unratable (?).

Method	CUReT				CDS			
	E1	E2	Diff	SI	E3	E4	Diff	SI
Fractal analysis using filter banks	91.1	85.8	5.8	∘	71.8	70.5	1.8	+
Multi-fractal spectrum	91.4	77.0	15.8	−	69.1	71.1	−2.8	+
D^3T-CWT with DCT	98.3	88.7	9.8	∘	76.1	73.8	3.0	∘
Multiscale blob features (number)	97.2	89.6	7.8	∘	65.9	59.9	9.1	−
DT-CWT with DCT	98.4	87.1	11.5	∘	75.3	70.9	5.8	∘
Affine invariant LTP	99.0	95.7	3.3	+	74.4	75.8	−1.9	+
Fractal dim. for orientation histograms	86.9	74.1	14.7	−	71.7	70.6	1.5	+
Dense SIFT features	71.5	67.4	5.7	∘	68.1	66.7	2.1	+
D^3T-CWT with DCT (local)	97.7	92.9	4.9	∘	72.2	71.1	1.5	+
Cyclic shifting of local features	98.8	95.1	3.7	+	73.3	72.8	0.7	+
Log-polar approach	90.7	88.3	2.6	+	72.6	73.4	−1.1	+
Dominant scale approach	92.7	93.9	−1.3	+	72.4	64.8	11.7	−
D^3T-CWT with DFT (local)	96.3	89.7	6.9	∘	72.9	73.0	−0.1	+
Slide matching (original)	93.5	81.4	12.9	∘	58.8	54.3	7.6	?
Slide matching (modified)	97.6	75.3	22.8	−	63.2	60.5	4.3	?
Local affine regions	96.1	89.8	6.6	∘	67.6	59.0	12.7	−
Multiscale blob features (shape)	96.9	93.9	3.1	+	72.0	62.0	16.1	−
ICM	90.2	81.4	9.8	∘	64.8	59.8	7.7	∘
D^3T-CWT with DFT	95.9	89.2	7.0	∘	63.7	69.1	−8.5	?
SCM	97.9	92.6	5.4	∘	60.3	56.5	6.3	?
DT-CWT	99.2	97.0	2.2	+	72.4	68.4	5.5	∘
D^3T-CWT	99.1	96.8	2.3	+	73.0	69.7	4.5	∘

When we consider the methods using the DT-CWTs, we see that many of the techniques designed to be scale invariant perform worse than the original CWTs without any specific tuning, except for the methods applying DCT across global subband descriptors, for which it is not even theoretically clear why they should enhance scale invariance of the DT-CWT.

Our results indicate that scale invariance is not important for the classification of celiac disease, at least when considering our dataset to be representative. There is no positive correlation between the performance of the methods (in terms of OCR) and their (determined) scale invariance.

There is a big difference between theoretical concepts for scale invariance and practical scale invariance actually achieved in experiments. It also turned out that the practical scale invariance of a method is not fixed, it depends on the application the method is used for. The determined scale invariance of the methods using the CUReT database (application texture recognition) is quite different to the determined scale invariance using the CDS database (application endoscopic image classification).

In case of endoscopic image classification, it turned out that the methods which have not been designed to be scale invariant are nearly as scale invariant than those explicitly designed to be scale invariant. The affine invariant LTP method exhibited the highest degree of scale invariance.

The behavior of methods is interesting in the case of texture recognition. Our results indicate that scale variant methods turn out to be more effective than their scale invariant counterparts. This is quite surprising, since these methods were especially designed to be scale invariant for texture recognition tasks. From this point of view we have to state that techniques claimed to be scale invariant should be actually tested for this property in properly designed experiments. It contradicts good scientific practice to state properties which do not hold in actual applications.

Acknowledgments

This work is partially supported by the Austrian Science Fund, TRP Project 206, and the Austrian National Bank Jubiläumsfonds Project 12991.

References

Barbosa, D.J.C., Ramos, J., Lima, C.S., 2008. Detection of small bowel tumors in capsule endoscopy frames using texture analysis based on the discrete wavelet transform. In: Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008 (EMBS'08), Vancouver, British Columbia, Canada. pp. 3012–3015.

Barbosa, D.J.C., Ramos, J.,, Correia, J.H., Lima, C.S., 2009. Automatic detection of small bowel tumors in capsule endoscopy based on color curvelet covariance statistical texture descriptors. In: Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009 (EMBC'09), Minneapolis, Minnesota, USA. pp. 6683–6686.

Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition 30, 1145–1159.

Brodatz, P., 1966. Textures: A Photographic Album for Artists and Designers. Dover Publications, New York.

Cammarota, G., Martino, A., Pirozzi, G., 2004. Direct visualization of intestinal villi by high-resolution magnifying upper endoscopy: a validation study. Gastrointestinal Endoscopy 60, 732–738.

Cammarota, G., Cesaro, P., Martino, A., et al., 2006. High accuracy and cost-effectiveness of a biopsy-avoiding endoscopic approach in diagnosing coeliac disease. Alimentary Pharmacology and Therapeutics 23, 61–69.

Cammarota, G., Cuoco, L., Cesaro, P., et al., 2007. A highly accurate method for monitoring histological recovery in patients with celiac disease on a gluten-free diet using an endoscopic approach that avoids the need for biopsy: a double-center study. Endoscopy 39, 46–51.

Chand, N., Mihas, A.A., 2006. Celiac disease: current concepts in diagnosis and treatment. Journal of Clinical Gastroenterology 40, 3–14.

Ciaccio, E.J., Tennyson, C.A., Lewis, S.K., Bhagat, G., Green, P.H., 2010a. Classification of videocapsule endoscopy image patterns: comparative analysis between patients with celiac disease and normal individuals. BioMedical Engineering Online 9.

Ciaccio, E.J., Tennyson, C.A., Lewis, S.K., Krishnareddy, S., Bhagat, G., Green, P.H., 2010b. Distinguishing patients with celiac disease by quantitative analysis of videocapsule endoscopy images. Computer Methods and Programs in Biomedicine 100, 39–48.

Ciaccio, E.J., Bhagat, G., Tennyson, C.A., Lewis, S.K., Hernandez, L., Green, P.H., 2011. Quantitative assessment of endoscopic images for degree of villous atrophy in celiac disease. Digestive Disease and Science 56, 805–811.

Dana, K., Van-Ginneken, B., Nayar, S., Koenderink, J., 1999. Reflectance and texture of real world surfaces. ACM Transactions on Graphics (TOG) 18, 1–34.

Fasano, A., Berti, I., Gerarduzzi, T., Not, T., Colletti, R.B., Drago, S., Elitsur, Y., Green, P.H.R., Guandalini, S., Hill, I.D., Pietzak, M., Ventura, A., Thorpe, M., Kryszak, D., Fornaroli, F., Wasserman, S.S., Murray, J.A., Horvath, K., 2003. Prevalence of celiac disease in at-risk and not-at-risk groups in the united states: a large multicenter study. Archives of Internal Medicine 163, 286–292.

Fei-Fei, L., Perona, P., 2005. A bayesian hierarchical model for learning natural scene categories. In: Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society, pp. 524–531.

Fung, K.K., Lam, K.M., 2009. Rotation- and scale-invariant texture classification using slide matching of the gabor feature. In: Proceedings of Intelligent Signal Processing and Communication Systems, pp. 521–524.

Geusebroek, J.M., Smeulders, A.W.M., van de Weijer, J., 2003. Fast anisotropic gauss filtering. IEEE Transactions on Image Processing 12, 938–943.

Häfner, A., Uhl, A., Vécsei, A., Wimmer, G., Wrba, F., 2010. Complex wavelet transform variants and scale invariance in magnification-endoscopy image classification. In: Proceedings of the 10th International Conference on Information Technology and Applications in Biomedicine (ITAB'10), Corfu, Greece.

Hayman, E., Caputo, B., Fritz, M., Eklundh, J.O., 2004. On the significance of real-world conditions for material classification. In: Proceedings of the European Conference on Computer Vision, pp. 253–266.

Hegenbart, S., Uhl, A., 2013. An Affine Invariant Texture Descriptor based on Local Ternary Patterns. Technical Report 2013-01. Department of Computer Sciences, University of Salzburg, Austria. <http://www.cosy.sbg.ac.at/research/tr.html>.

Hegenbart, S., Kwitt, R., Liedlgruber, M., Uhl, A., Vecsei, A., 2009. Impact of duodenal image capturing techniques and duodenal regions on the performance of automated diagnosis of celiac disease. In: Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis (ISPA '09), Salzburg, Austria. pp. 718–723.

Hegenbart, S., Uhl, A., Vécsei, A., 2011. Systematic assessment of performance prediction techniques in medical image classification – a case study on celiac disease. In: Proceedings of the 22nd International Conference on Information Processing in Medical Imaging (IPMI'11), Monastery Irsee, Germany, pp. 498–508.

Hegenbart, S., Uhl, A., Vécsei, A., 2012. On the implicit handling of varying distances and gastrointestinal regions in endoscopic video sequences with indication for celiac disease. In: Proceedings of the IEEE International Symposium on Computer-Based Medical Systems (CBMS'12).

Iakovidis, D., Maroulis, D., Karkanis, S., Papageorgas, P., Tzivras, M., 2004. Texture multichannel measurements for cancer precursors identification using support vector machines. Measurement 36, 297–313.

Johnson, J., 1994. Pulse-coupled neural nets: translation, rotation, scale, distortion, and intensity signal invariance for images. Applied Optics 33, 6239–6253.

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., Häfner, M., Gangl, A., Wrba, F., Vécsei, A., 2009. Feature extraction from multi-directional multi-resolution image transformations for the classification of zoom-endoscopy images. Pattern Analysis and Applications 12, 407–413.

Lazebnik, S., Schmid, C., Ponce, J., 2005. A sparse texture representation using local affine region. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 1265–1278.

Liedlgruber, M., Uhl, A., 2011a. Computer-aided decision support systems for endoscopy in the gastrointestinal tract: a review. IEEE Reviews in Biomedical Engineering.

Liedlgruber, M., Uhl, A., 2011b. Predicting pathology in medical decision support systems in endoscopy of the gastrointestinal tract. In: Jao, C. (Ed.), Efficient Decision Support Systems – Practice and Challenges in Biomedical Related Domain. InTech, Rijeka, Croatia, pp. 195–214.

Li, Z., Liu, G., Yang, Y., You, J., 2012. Scale- and rotation-invariant local binary pattern using scale-adaptive texton and subuniform-based circular shift. IEEE Transactions on Image Processing 21, 2130–2140.

Lo, E.H.S., Pickering, M.R., Frater, M.R., Arnold, J.F., 2004. Scale and rotation invariant texture features from the dual-tree complex wavelet transform. In: Proceedings of the International Conference on Image Processing, ICIP '04, IEEE, Singapore. pp. 227–230.

Lo, E.H.S., Pickering, M.R., Frater, M.R., Arnold, J.F., 2009. Query by example using invariant features from the double dyadic dual-tree complex wavelet transform. In: CIVR '09: Proceeding of the ACM International Conference on Image and Video Retrieval, ACM, Santorini, Fira, Greece, pp. 1–8.

Lowe, D.G., 1999. Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision. IEEE, pp. 1150–1157.

Mäenpää, T., 2003. The Local Binary Pattern Approach to Texture Analysis – Extensions and Applications. Ph.D. thesis. University of Oulu.

Ma, Y., Liu, L., Zhan, K., Wu, Y., 2010. Pulse coupled neural networks and one-class support vector machines for geometry invariant texture retrieval. Image and Vision Computing 28, 1524–1529.

Marsh, M., 1992. Gluten, major histocompatibility complex, and the small intestine. a molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). Gastroenterology 102, 330–354.

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

Mikolajczyk, K., Cordelia, S., 2004. Scale & affine invariant interest point detectors. International Journal of Computer Vision 60, 63–86.

Mikolajczyk, K., Schmid, C., 2002. An affine invariant interest point detector. In: Proceedings of the European Conference on Computer Vision. Springer Verlag, pp. 128–142.

Montoya-Zegarra, J.A., Leite, N.J., Torres, R., 2007. Rotation-invariant and scale-invariant steerable pyramid decomposition for texture image retrieval. In: Proceedings of the XX Brazilian Symposium on Computer Graphics and Image Processing, pp. 121–128.

Niveloni, S., Florini, A., Dezi, R., et al., 1998. Usefulness of videoduodenoscopy and vital dye staining as indicators of mucosal atrophy of celiac disease: assessment of interobserver agreement. Gastrointestinal Endoscopy 47, 223–229.

Oberhuber, G., Granditsch, G., Vogelsang, H., 1999. The histopathology of coeliac disease: time for a standardized report scheme for pathologists. European Journal of Gastroenterology and Hepatology 11, 1185–1194.

Ojala, T., Pietikäinen, M., Harwood, D., 1996. A comparative study of texture measures with classification based on feature distributions. Pattern Recognition 29, 51–59.

Petroniene, R., Dubcenco, E., Baker, J., 2005. Given capsule endoscopy in celiac disease: evaluation of diagnostic accuracy and interobserver agreement. The American Journal of Gastroenterology 100, 685–694.

Pun, C.M., Lee, M.C., 2003. Log-polar wavelet energy signatures for rotation and scale invariant texture classification. IEEE Pattern Analysis and Machine Intelligence 25, 590–603.

Ranganath, H., Kuntimad, G., Johnson, J., 1995. Pulse coupled neural networks for image processing. In: Proceedings of the IEEE Southeastcon '95, 'Visualize the Future', pp. 37 –43.

Selesnick, I., Baraniuk, R., Kingsbury, N., 2005. The dual-tree complex wavelet transform. IEEE Signal Processing Magazine 22, 123–151.

Tan, T.N., 1995. Geometric transform invariant texture analysis. In: Proceedings of SPIE 2488, pp. 475–485.

Tan, X., Triggs, B., 2007. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: Analysis and Modelling of Faces and Gestures, pp. 168–182.

Uhl, A., Vécsei, A., Wimmer, G., 2011a. Complex wavelet transform variants in a scale invariant classification of celiac disease. In: Proceedings of the 5th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2011), Las Palmas de Gran Canaria, Spain, pp. 742–749.

Uhl, A., Vécsei, A., Wimmer, G., 2011b. Fractal analysis for the viewpoint invariant classification of celiac disease. In: Proceedings of the 7th International Symposium on Image and Signal Processing (ISPA 2011), Dubrovnik, Croatia, pp. 727 –732.

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.

Varma, M., Zisserman, A., 2009. A statistical approach to material classification using image patch exemplars. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 2032–2047.

Varma, M., Zissermann, A., 2005. A statistical approach to texture classification from single images. International Journal of Computer Vision (IJCV) 62, 61–81.

Vécsei, A., Fuhrmann, T., Uhl, A., 2008. Towards automated diagnosis of celiac disease by computer-assisted classification of duodenal imagery. In: Proceedings of the 4th International Conference on Advances in Medical, Signal and Information Processing (MEDSIP '08), Santa Margherita Ligure, Italy, pp. 1–4.

Vecsei, A., Fuhrmann, T., Liedlgruber, M., Brunauer, L., Payer, H., Uhl, A., 2009. Automated classification of duodenal imagery in celiac disease using evolved fourier feature vectors. Computer Methods and Programs in Biomedicine 95, S68–S78.

Vécsei, A., Amann, G., Hegenbart, S., Liedlgruber, M., Uhl, A., 2011. Automated Marsh-like classification of celiac disease in children using an optimized local texture operator. Computers in Biology and Medicine 41, 313–325.

Vedaldi, A., Fulkerson, B., 2008. VLFeat: An open and portable library of computer vision algorithms. <http://www.vlfeat.org/>.

Xu, Q., Chen, Y.Q., 2006. Multiscale blob features for gray scale, rotation and spatial scale invariant texture classification. In: Proceedings of 18th International Conference on Pattern Recognition (ICPR), pp. 29–32.

Xu, Y., Huang, S.B., H. Ji, C.F., 2009a. Combining powerful local and global statistics for texture description. In: Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE, pp. 573–580.

Xu, Y., Ji, H., Fermüller, C., 2009b. Viewpoint invariant texture description using fractal analysis. International Journal of Computer Vision 83, 85–100.

Zhan, K., Zhang, H., Ma, Y., 2009. New spiking cortical model for invariant texture retrieval and image processing. IEEE Transactions on Neural Networks 20, 1980–1986.

Zhang, J., Tan, T., 2002. Brief review of invariant texture analysis methods. Pattern Recognition 35, 735–747.

Zhang, J., Marszalek, M., Lazebnik, S., Schmid, C., 2006. Local features and kernels for classification of texture and object categories: a comprehensive study. In: Conference on Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06, p. 13.