

# An Affine Invariant Local Ternary Patterns Method

Sebastian Hegenbart      Andreas Uhl      Andreas Vécsei<sup>a</sup>

<sup>a</sup>St. Anna Children's Hospital, Department of Pediatrics Medical University,  
Vienna

Technical Report 2013-03

July 2013

**Department of Computer Sciences**

Jakob-Haringer-Straße 2  
5020 Salzburg  
Austria  
[www.cosy.sbg.ac.at](http://www.cosy.sbg.ac.at)

**Technical Report Series**

# An Affine Invariant Local Ternary Patterns Method

Sebastian Hegenbart<sup>1</sup>, Andreas Uhl<sup>1</sup>, and Andreas Vécsei<sup>2</sup>

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

<sup>2</sup>St. Anna Children's Hospital, Department of Pediatrics  
Medical University, Vienna

**Abstract.** Local Binary Patterns and various derivatives of the method have been widely used in the field of texture recognition over the past 15 years. A restriction of these methods is their variance with respect to affine transformations of an image. This is caused by the fixed circular neighborhood and the fixed support area of sampling points. The main approach to deal with affine transformations such as rotations is based on modifying or enhancing the encoding scheme of the patterns. In this work we present an extension to Local Ternary Patterns which is based on adaptive elliptic shaped neighborhoods with adaptive support areas of sampling points. We use scale normalized Laplacian maxima in a scale-space to identify interest points within an image. Based on the scale information the multi-scale second moment matrix is computed to estimate the affine transformation at the location of a Laplacian scale-space maximum. Utilizing this information, a scale mask is computed to improve the reliability of scale estimation. Finally Local Ternary Patterns are computed along equidistant points in terms of arc length along the estimated ellipse.

## 1 Introduction

The Local Binary Pattern approach as introduced by Ojala et al. [1] as well as the Local Ternary Patterns method proposed by Tan and Triggs [2] are not invariant with regard to affine transformations.

In certain scenarios, endoscopic image classification for example, textures are captured at various perspectives and distances. These variations, caused by camera motion, lead to a visualization of textures under affine transformations. Methods that are invariant in terms of these type of transformation could therefore improve the accuracy of an automated classification.

The main approach found in literature on LBP to deal with these kind of transformation is to enhance or modify the encoding scheme of the patterns. Ojala and Mäenpää [3] for example deal with rotations by circular shifting the computed patterns until a minimum (in terms of decimal interpretation of the binary string) is found. By doing so, the patterns are implicitly aligned. A drawback however is that the number of possible patterns is heavily reduced losing discriminative power.

Ojala and Mäenpää also introduced multi-resolution Local Binary Patterns [3], using a set of different radii with appropriate sampling areas sizes. This approach however does not employ a scale selection mechanism and does not improve invariance in terms of re-scaling of a texture.

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 utilize additional information gained by the computation of multi-scale second moment matrices in combination with scale-normalized Laplacian maxima in a scale-space. The method shares the idea of using the scale-space framework with methods such as the SIFT feature detector [4] and other region detectors.

The idea of combining scale-space maxima with Local Binary Patterns has also been explored by Li et al. [5]. Their approach also uses scale-space maxima to adaptively select the radius of the Local Binary Pattern descriptor. Rotational invariance is achieved by analyzing the computed patterns, identifying subuniform patterns with the maximum statistical value. These subuniform patterns are finally circularly shifted to the first position within the patterns in order to achieve rotational invariance, a method similar to the approach as suggested by Ojala and Mäenpää [3].

In analogy with the technique proposed by Li et al. our scale selection mechanism is also based on scale-normalized Laplacian scale-space maxima. But instead of modifying or enhancing the encoding scheme of the Local Binary Patterns to achieve rotational invariance, our method is based on adaptive, elliptically shaped, neighborhoods. The shape of the neighborhoods is derived from information encoded by multi-scale second moment matrices which are computed at corresponding sampling locations. Following this approach we are able to compensate for an even larger number of affine transformations including rotation, uniform scaling as well as non-uniform scaling. Finally we improve the robustness of the scale selection mechanism, by the computation of a scale reliability mask which employs local shape and scale information using a Gaussian probability density function.

In Section 2 we will give a brief review of the traditional LBP methodology and point out the restrictions of those methods relating to affine transformations. We will cover our methodology in Section 3, including the mechanism for scale-selection, shape estimation, as well as the computation of adaptive sampling support areas, the adaptive elliptic neighborhoods and the adaptive sampling point indexing. Section 4 covers the experiments conducted to evaluate the performance of the proposed method. Section 5 concludes the paper.

## 2 Traditional LBP based Methods

The Local Binary Patterns (LBP) operator [1] is used to model a pixel neighborhood in terms of pixel intensity differences. A pixel neighborhood  $\nu$  is defined in relation to a pixel at  $(x, y)$  of the image intensity function  $f$  as a sequence of  $p$  equidistant points on a circle with radius  $r$  around  $(x, y)$ :

$$\varphi_{r,p}(x, y, k) := \begin{pmatrix} x + r \cos\left(\frac{2\pi k}{p}\right) \\ y - r \sin\left(\frac{2\pi k}{p}\right) \end{pmatrix}^T$$

$$(\nu_k) := \left( f(\varphi_{r,p}(x, y, k)) \right)_{k \in \{0 \dots p-1\}}$$

Based on a sign function a weighted sum is computed and interpreted as binary label according to the specific pixel neighborhood intensity relationship:

$$s(x) := \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

For a position  $(x, y)$  the local binary pattern of  $p$  neighbors and radius  $r$  is computed as:

$$\mathbf{LBP} := \sum_{k=0}^{p-1} 2^k s(\nu_k - f(x, y)).$$

The joint distributions of these labels are then used to characterize a texture.

## 2.1 Local Ternary Patterns

Local Ternary Patterns [2] are derived from Local Binary Patterns. The authors modified the original method to improve it's robustness with regard to varying lighting conditions. In LTP, the sign function is changed from a binary function to a ternary function

$$s(x) = \begin{cases} 1, & \text{if } x \geq T_h \\ 0, & \text{if } |x| < T_h \\ -1, & \text{if } x \leq -T_h. \end{cases} \quad (1)$$

The ternary decision leads to two separate histograms, one representing the distribution of the patterns resulting in a  $-1$ , the other representing the distribution of the patterns resulting in a  $1$ . Two separate histograms are computed

$$H_{I,lower}(i) = \sum_{x,y} (\mathbf{LBP}_{r,p}(x, y) = -i) \quad i = 0, \dots, 2^p - 1 \quad (2)$$

$$H_{I,upper}(i) = \sum_{x,y} (\mathbf{LBP}_{r,p}(x, y) = i) \quad i = 0, \dots, 2^p - 1. \quad (3)$$

The neighbor information of pixels that lie within the threshold is encoded implicitly by this splitting. Finally, both histograms are concatenated and treated as a single histogram.

## 2.2 Restrictions of the Traditional LBP-based Methods

The main restriction of LBP based methods is their variance with respect to affine transformations of an image. This is caused by the fixed circular neighborhood and the fixed support area of sampling points. Rotations can not be compensated due to the ambiguous orientation of the circular neighborhood.

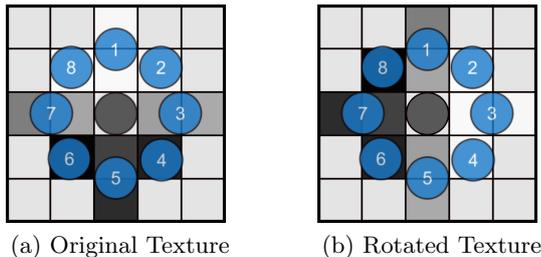


Fig. 1: Restriction Concerning Rotation

Figure 1 demonstrates how the rotation of a texture affects the calculation of the binary pattern. The sampling points along the circle are labeled by their corresponding indices in the pattern. Due to the fixed indexing of sampling points, a rotation of the texture results in a circular shift of the pattern.

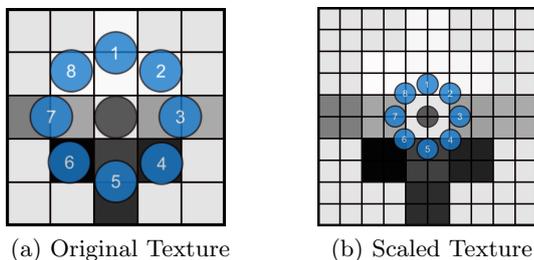


Fig. 2: Restriction Concerning Uniform Scaling

Figure 2 shows the effects of uniform scaling to the computation of the LBP patterns. Caused by the fixed sized radius of the operator, the spatial sampling locations are implicitly changed due to uniform scaling. Additionally the fixed size of the support area of the sampling points is a restriction. We see that the textures are analyzed at two different scales leading to two different, incompatible patterns.

Finally, Figure 3 shows the effects of non-uniform scaling. We can observe that the circular shape of the neighborhood is not well suited to compensate for

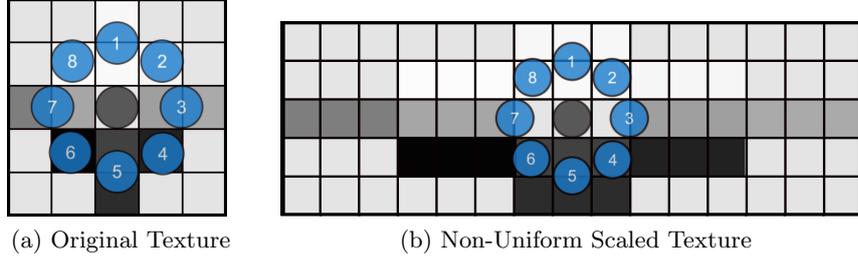


Fig. 3: Restriction Concerning Non-Uniform Scaling

this type of transformation. In parallel to uniform scaling, the textures are analyzed at different scales. Even more, different parts of the textures are analyzed.

### 3 Affine Invariant Local Ternary Patterns

Instead of modifying the encoding scheme of the computed patterns, we use an adaptive neighborhood with adaptive sampling support areas and adaptive indexing to achieve affine transformation invariance. The general methodology can be seen as an extension to the entire class of LBP based methods and will be presented based on Local Ternary Patterns in this work. The advantage of using Local Ternary Patterns is the implicit thresholding that prohibits structures detected as highly anisotropic, from contributing to the joint distributions of the patterns significantly.

Figure 4 gives a schematic overview of the separate steps of the method. The following sections cover these steps in detail.

#### 3.1 Scale Estimation based on Scale-Normalized Laplacian Maxima in Scale-Space

The scale-space theory was first extensively explored in the field of signal processing by Lindeberg [6]. It presents a framework to analyze signals at different scales. Let  $f : \mathbb{R}^2 \mapsto \mathbb{R}$  represent a continuous signal, then the linear scale-space representation  $L : \mathbb{R}^2 \times \mathbb{R}_+ \mapsto \mathbb{R}$  is defined by

$$L(\cdot; \sigma) = g(\cdot; \sigma) * f, \tag{4}$$

with initial condition  $L(\cdot, 0) = f$ . Where  $t \in \mathbb{R}_+$  is the scale parameter,  $g$  is a Gaussian function and “\*“ denotes convolution. The scale-space family  $L$  is the solution to the diffusion equation (heat equation):

$$\partial_\sigma L = \sigma \left( \frac{\partial^2 L}{\partial x^2} + \frac{\partial^2 L}{\partial y^2} \right) = \sigma \Delta L. \tag{5}$$

The popular Difference of Gaussians (DoG) approach is a finite element approximation of the Laplacian of Gaussians. This technique is used in several

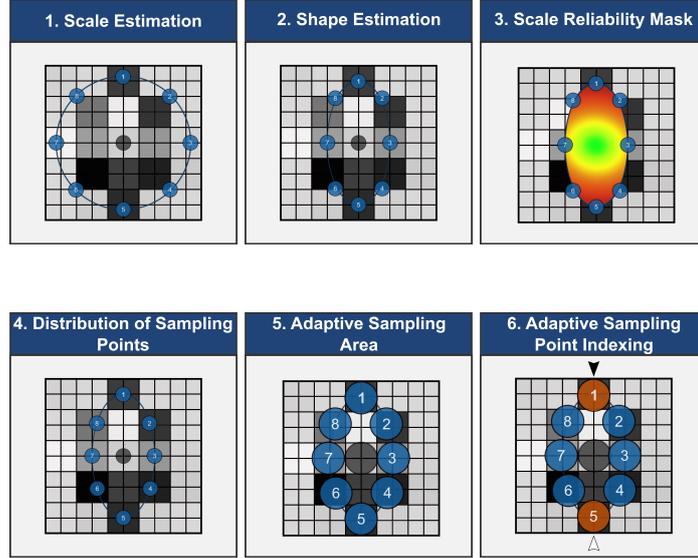


Fig. 4: Schematic Overview of the Method

methods, SIFT [4] being the most prominent. Instead of using the DoG approximation to the Laplacian of Gaussians, we construct the scale-space and compute the scale-normalized Laplacians ( $\sigma^2 |\Delta L|$ ) of each image  $I$  at each location  $x \in \mathbb{N}^2$  at different scales with  $\sigma = \frac{1.5^k}{\sqrt{2}}$ ,  $k \in \{1, \dots, 20\}$  denoted as  $(\bar{\Delta}I(x; \sigma))$ . The initial scale is chosen such that it corresponds to the LBP radius of 1.5.

### 3.2 Shape Estimation based on the Multi-Scale Second Moment Matrix

The second moment matrix (also known as structure tensor) summarizes the predominant directions of the gradient in a specific point neighborhood of an image. In contrast to the second moment matrix, the multi-scale second moment matrix [7] is defined over two scale parameters. It allows to estimate the shape of visual structures at their dominant scale, as detected by employing the maximum responses of the Laplacian of the scale-space representation.

The local scale, denoted by  $t$  determines the amount of smoothing applied prior to computing the gradient of the image. The integration scale  $s$  is used as parameter to a Gaussian function  $g$  defining the shape and weights of a specific neighborhood area in the image over which the gradients are accumulated. We compute the multi-scale second moment matrices at each location  $x$  of an image  $I$  which is attaining a local maximum. We use the detection scale of the maximum as the local scale  $t$ , the integration scale  $s = \sqrt{2}t$  depends only on the

corresponding detection scale.

$$\mu(x; t, s) = \int_{\xi \in \mathbb{R}^2} (\nabla I)(x - \xi; t) (\nabla I)^T(x - \xi; t) g(\xi; s) d\xi. \quad (6)$$

We denote  $(\nabla I)(x; t)$  as the gradient of the scale-space orientation at scale  $t$  and position  $x$ . The eigenvalues of the multi-scale second moment matrix characterize the length of the axes of an ellipse (up to some constant multiplier) while the eigenvectors describe the orientations of the axes. We use the information encoded by the multi-scale second moment matrix to adapt the shape of the LTP neighborhoods in relation to local structural properties.

Due to the fact that the orientation of the ellipse described by the second moment matrix is summarizing the gradient, the orientation is normal to a structure in the analyzed area, we use the inverse of the second moment matrix. The inverse does not modify the ratios of the axis lengths, the shape is therefore retained.

### 3.3 Computation of the Scale Reliability Mask

The SIFT method identifies locations attaining a local maximum and a maximum across the scale dimension of the scale-space to compute feature descriptors. The property that Laplacian maxima are preserved on re-scaling of a signal theoretically leads to a one-to-one mapping between descriptors of a signal at two different scales. Hence the classification of SIFT features is based on matching pairs of keypoints descriptors.

The LBP descriptor in contrast is a global descriptor, the discriminative power does not lie within the single patterns but within the joint distribution of those patterns. The identification of locations of SIFT like keypoints to compute the features would lead to a insufficient number of patterns, leading to a diminished discriminative power for classification. Therefore the patterns are computed along a fixed grid (all pixel positions) in an image and are used in combination with the scale reliability mask. Because of the fixed sampling, we can not guarantee that a pixel position attains a scale-space maximum across the scale dimension. This is in parallel to dense SIFT features [8] where the SIFT descriptors are computed along a fixed grid. Dense SIFT uses a fixed scale parameter, losing the scale invariance property.

To achieve scale invariance in our method, we employ the scale selection mechanism as explained in Section 3.1 at each pixel location and employ the scale-normalized Laplacian responses at each scale to compute a scale reliability mask.

The mask is especially useful when textures are not strictly periodic and attain multiple scales among the texture as is the case in endoscopic imagery. We exploit the fact that image structures at close spatial proximity to a Laplacian maximum are at the same scale or relatively close to the scale of the detected maximum.

We combine the computed shape information at the spatial location of a maximum with the detection scale of the maximum to identify a corresponding

area in relation to the specific maximum. The corresponding area is defined as the area of the normalized ellipse (see Section 3.5) as represented by the specific multi-scale second moment matrix. All spatial locations within the corresponding area are assigned to the scale and response of that maximum. In our terminology this maximum is called the corresponding maximum for the spatial locations within the corresponding area of the maximum. It is possible that a single spatial location attains multiple scales and responses, this is due to the fact that the support areas of multiple maxima might overlap.

We observed, that the reliability, that a local structure inside the corresponding area of a maximum attains the same scale as the maximum decreases with spatial distance.

We therefore employ a measure for the reliability, that a local structure attains the same scale as a corresponding maximum at a certain distance  $d$  from this maximum, using a Gaussian probability density function.

This function is parametrized by  $\sigma_p$ . The parameter is chosen such that the reliability of a local structure with a distance, according to the length of the minor axis ( $b$ ) of the ellipse describing the corresponding area, to the maximum is 0.5.

$$\sigma_p = \left( \frac{-(\frac{b}{2})^2}{2 \log(0.5)} \right)^{\frac{1}{2}} \quad (7)$$

We additionally use the responses of all corresponding maxima for a spatial location to ensure that maxima with lower responses have lower reliabilities. Please remember that the responses are scale normalized and can be compared across different scale levels. Let's denote  $\bar{\Delta}M(x; \cdot)$  as the set of all corresponding maximas for a location  $x$  and  $\bar{\Delta}M_i(x; \cdot)$  the  $i$ -th corresponding maximum. We utilize the responses to give the main emphasis on the corresponding maximum with the highest response. The reliability of a location  $x$  to attain scale  $l$  is then

$$r(x; l) = \frac{\bar{\Delta}M_i(x; l)}{\max_t \bar{\Delta}M(x; t)} e^{\frac{-d^2}{2\sigma_p^2}}. \quad (8)$$

The reliability measure is finally used to assign a weight, controlling the contribution of each computed pattern to the histogram. Figure 5 illustrates the distribution of reliabilities for a maximum, not taking into account multiple overlapping maxima.

### 3.4 Adaptive Sampling Support Area

Ojala and Mäenpää [3] first used Gaussian filtering to adapt the sampling support area to various LBP scales. By employing low-pass filtering, a pixel at a single spatial location contains information of its spatial neighborhood. By employing the scale information of the corresponding Laplacian maxima at each image location we apply an adaptive Gaussian filter prior to computing the LTP pattern at a location. The width of the Gaussian filter is chosen such that the

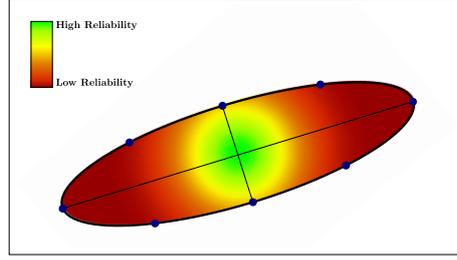


Fig. 5: Scale Reliability Mask with Fixed Responses

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 is computed as

$$f_w = \frac{\sqrt{2}\sigma 2\pi}{N}, \quad (9)$$

for  $N$  being the number of neighbors, and  $\sigma$  being the current scale. The Gaussian Filter coefficients are then computed adaptively for a given radius  $r = \frac{f_w}{2}$  such that  $P$  percent of the area below the Gaussian is covered:

$$\begin{aligned} \int_{-r}^r e^{-\frac{x^2}{2\sigma^2}} dx &= P \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx \\ 2 \int_0^r e^{-\frac{x^2}{2\sigma^2}} dx &= P\sigma\sqrt{2\pi} \\ \sigma &= \frac{r}{\sqrt{2}\text{erf}^{-1}(P)} \end{aligned} \quad (10)$$

We chose  $P$  to be 0.9 which corresponds to 90% of the area. Figure 6 illustrates the computation of the adaptive sampling area.

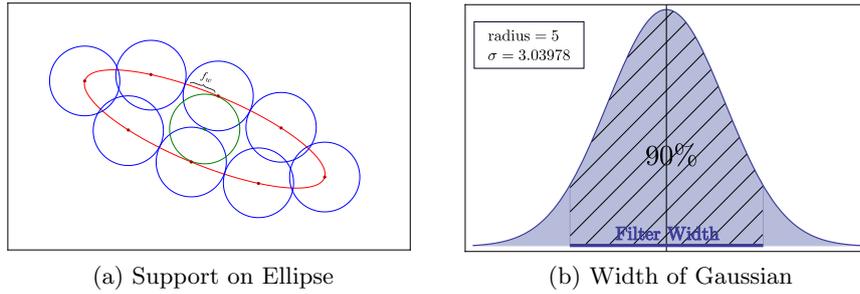


Fig. 6: Adaptive Sampling Area by Gaussian filtering

### 3.5 Adaptive Elliptic Neighborhood

Non-uniform scaling transforms a circle to an ellipse. Using a fixed circular neighborhood leads to incorrect sampling positions in case of this type of transformation. Therefore the neighborhood of sampling points is required to be adaptive to be able to compensate for non-uniform scaling. Utilizing the information encoded by the multi-scale second moment matrix at a given location we compute an adaptive LTP neighborhood.

The absolute lengths of the ellipse's axes are not known from the information encoded by the multi-scale second moment matrix. We therefore resort to utilizing the scale information gained by the scale-space representation. A Laplacian maximum found at scale  $\sigma$  corresponds to a circular shaped structure with radius  $\sqrt{2}\sigma$  (often denoted as a *Blob* in literature). By using the scale information we normalize the ellipse's axes such that the circumference is equal to the circumference of the detected Blob. Denoting  $a$  as the major semi-axis of the ellipse and  $b$  as the minor semi-axis respectively, we utilize Ramanujan's formula for approximating the circumference of the ellipse and solve the quadratic equation for a constant scaling factor  $c$

$$\sqrt{2}\sigma 2\pi = \pi \left( 3(ac + bc) - \sqrt{(3ac + bc)(ac + 3bc)} \right). \quad (11)$$

The axes of the ellipse are then re-scaled by the appropriate solution of  $c$ . Figure 7 illustrates the normalized ellipses at a fixed scale.

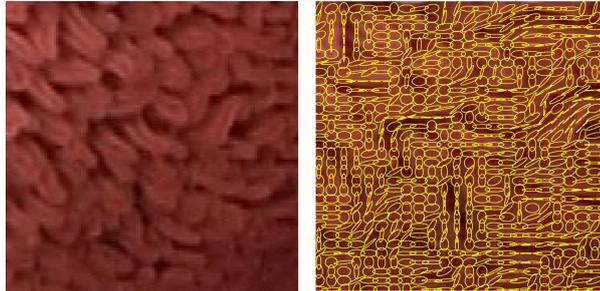


Fig. 7: Normalized Ellipses

As described in Section 3.4 we use adaptive Gaussian filtering to compensate for scaling. Please note however that non-uniform scaling of an image also changes the shape of the sampling areas from circles to ellipses. Unfortunately the pre-computation of non-uniform Gaussians for computing elliptically shaped sampling areas is currently unfeasible.

This is due to the extra degrees of freedom as compared to uniform Gaussians. Based on the assumption that the major number of this type of transformation is not highly anisotropic we use uniform Gaussian filtering as an approximation. Figure 8 illustrates the change of sampling locations after non-uniform scaling of

a circle with axis aligned equidistant sampling points. By using the approximation however the intersecting areas of adjacent sampling points would be large at certain positions along the ellipse, therefore losing discriminant information encoded by the patterns.

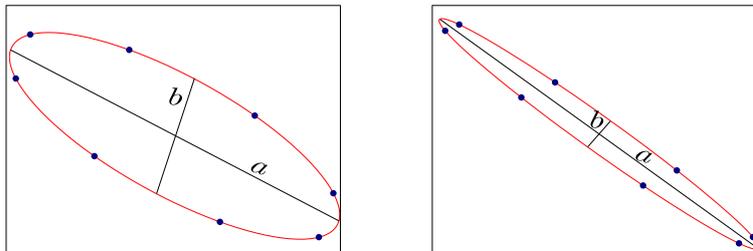


Fig. 8: Sampling Point Positions after Non-Uniform Scaling

We therefore distribute sampling points such 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 remaining  $3N$  points. To find the offset on the x-axis of the  $n$ -th point ( $\Delta x_n$ ) from the center of the ellipse the equation

$$\frac{n}{N+1} \int_0^a \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx = \int_0^{\Delta x_n} \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx \quad (12)$$

is solved for  $\Delta x_n$ . 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. Figure 9 illustrates the distribution of sampling points on the ellipse.

Please note that using uniform Gaussians leads to a decreased invariance with regard to non-uniform scaling. In case of highly anisotropic scaling the approximation does not suffice. Therefore using non-uniform Gaussians to compute adaptive sampling areas is a topic for future work.

### 3.6 Adaptive Sampling Point Indexing

The shape information employed in the LTP neighborhood allows to define an adaptive indexing of points in the neighborhood. The first sampling point defines

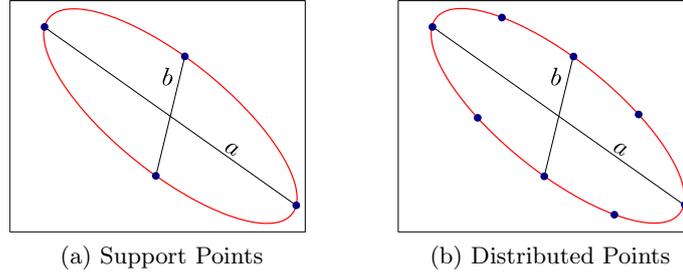


Fig. 9: Sampling Point Positions used with Uniform Gaussian Filtering

the orientation of the computed pattern. In standard LBP based methods the sampling start indexing is fixed. This is caused by the missing shape information.

To cope with the ambiguous orientation of an ellipse we use two starting points instead of one (computing two patterns per position). The computation of two patterns per location leads to an implicit alignment of orientation between two patterns when computing the histogram distances. Experiments have shown that the discriminative power of the computed histograms does not decrease significantly by the computation of extra patterns. Alternatively the dominant orientation could be computed in the respective area to determine the orientation unambiguously.

In our methodology the first sampling point is defined as the point of the intersection of the major axis with the ellipse. In case of ellipses that are close to a circle this definition becomes unreliable. By choosing the larger axis the first sampling point would "jump" in case of small scaling differences (due to the swaps of the major and minor axis). We therefore compute an isotropy measure as the ratio of eigenvalues of the corresponding second moment matrix to identify this case. If the ratio of eigenvalues  $\frac{\lambda_{\min}}{\lambda_{\max}}$  is greater or equal to 0.95 the vertical axis is always considered as the major axis.

Due to the fact that rotation of a texture will be reflected by the multi-scale second moment matrix we are able to compensate for this kind of transformation by employing this approach. Figure 10 illustrates the computation of patterns with adaptive sampling start points. The arrows indicate that direction of computation. The numbers give the position of the sampling point in the pattern where the numbers in circles indicate the start of indexing (and computation of the patterns).

### 3.7 Proposed Solutions

In this Section we will briefly summarize the proposed solutions to the restrictions of the traditional LBP methods regarding affine transformations as discussed in Section 2.2. Figure 11 illustrates how the adaptive indexing of sampling points leads to invariance with regard to rotation. Please remember that two patterns are computed due to the ambiguous orientation of the ellipse. The

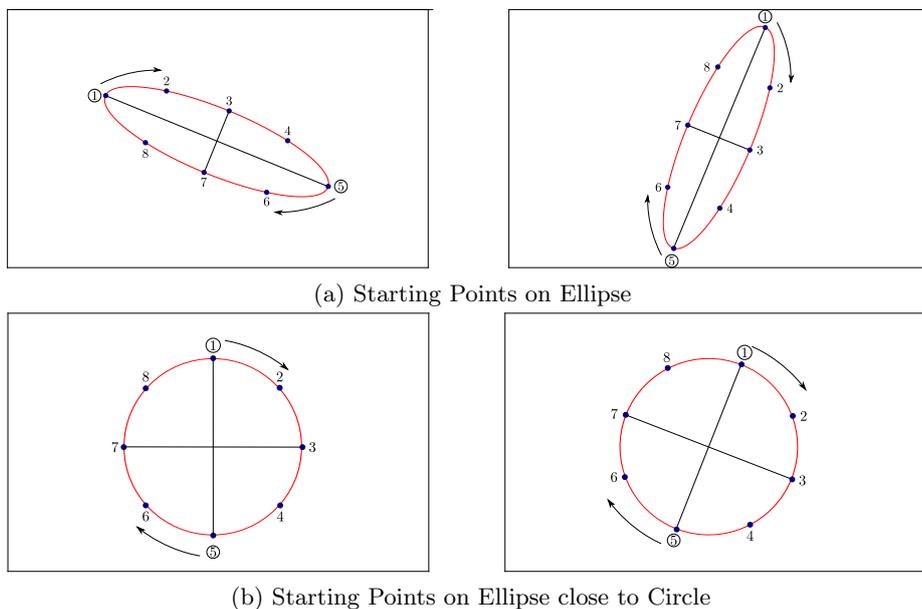


Fig. 10: Definition of Sampling Start Points

orange sampling points, highlighted using an arrow, indicate the respective starting points for the computation of each pattern.

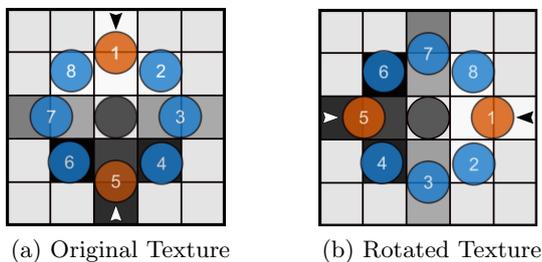


Fig. 11: Solution Concerning Rotation

Figure 12 illustrates how the adaptive sampling support area and adaptive neighborhood size improves the scale invariance with regard to uniform scaling. Note, that in contrast to the traditional LBP methods the same textural areas are analyzed.

Finally, Figure 13 presents the theoretical approach to invariance with respect to non-uniform scaling. Please note that, as discussed in Section 3.5 we resort to using circular shaped sampling areas due to the computational complexity.

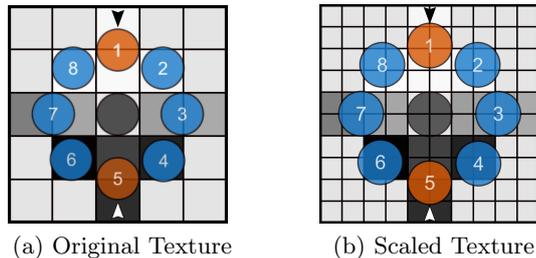


Fig. 12: Solution Concerning Uniform Scaling

In general, the amount of anisotropy in natural images in terms of scaling is not as significant as the illustrated Figure however.

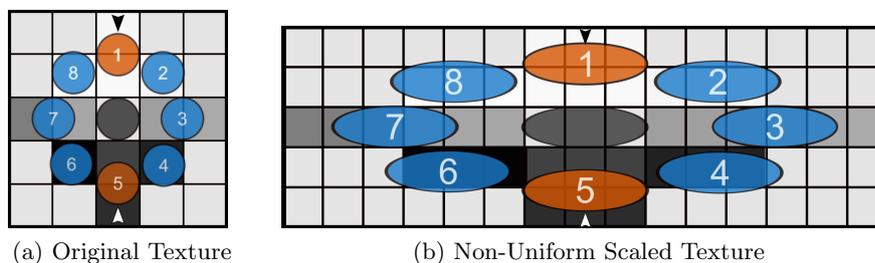


Fig. 13: Solution Concerning Non-Uniform Scaling

## 4 Experiments

### 4.1 Feature Extraction and Classification

For our experimental evaluation we use a set of scale invariant methods which have been published in the recent past and have gained attention by the community. We restrict the set of methods to scale-invariant methods to be capable of analyzing and comparing the scale-invariance property of the proposed method in an experimental setup especially designed for evaluating scale-invariance.

We present a short overview of the evaluated methods. Please refer to the original publications for more detailed information. Where available the implementation of the authors of the corresponding publications have been used. The methods without publicly accessible code have been re-implemented in Matlab. The proposed method was implemented in Java.

The various methods published in the original manuscripts employ a wide variety of different classifiers. We chose to use the k-nearest neighbor classifier (kNN) to put the main emphasis on the features as well as to allow comparisons

between methods. All data was split into distinct training and evaluation sets for experimentation. Where appropriate, feature optimization was performed on the training set using cross validation.

- **Fractal Analysis using Filter Banks** [9]:  
The fractal dimension gives a measure of how the detail of a pattern changes with the analyzing scale. We compute the local fractal dimension [9] using the Laplacian measure on images pre-filtered using the MR8 filter bank [10, 11]. The local fractal dimensions combined with the bag of visual words approach is then used to form the feature vector.
- **Multi-Fractal Spectrum** [12]:  
The method is based on the computation of the local fractal dimension for each pixel in an image. Again, only the Laplacian measure is used to compute the local fractal dimension. Based on the local fractal dimension, binary images are generated by applying a set of appropriate thresholds. Finally, the feature vector of an image consists of the concatenation of the global fractal dimensions of the binary images.
- **Multiscale Blob Features (number)** [13]:  
The method uses a series of flexible threshold planes which are computed for an image to construct a set of binary images. Geometrical attributes are then used to describe the image texture. We use the number of identified Blobs to form a feature vector.
- **DT-CWT** [14]:  
The dual-tree complex wavelet transformation is applied prior to computing statistical features from the resulting subbands. The mean and standard deviations of the absolute subband coefficients are computed and used as features.
- **DT-CWT with DCT** [15]:  
The method is based on the DT-CWT method. The features are computed by applying the discrete cosine transform across the scale dimension of a feature vector of the DT-CWT.
- **D<sup>3</sup>T-CWT** [16]:  
The D<sup>3</sup>T-CWT is applied to the image data. We compute the mean and standard deviation of the absolute subband coefficient values as feature.
- **D<sup>3</sup>T-CWT with DCT** [15]:  
Features are computed by applying the DCT across the scale dimension of a feature vector of the D<sup>3</sup>T-CWT.
- **D<sup>3</sup>T-CWT with DCT (local)** [15]:  
Each D<sup>3</sup>T-CWT is resized to match the size of the original image. A local

feature vector for each pixel is computed. Finally the DCT is applied across the scale dimension of these local feature vectors.

- **D<sup>3</sup>T-CWT with DFT (local)** [17]:  
Each D<sup>3</sup>T-CWT is resized to match the size of the original image. A local feature vector for each pixel is computed. Finally the DFT is applied across the scale dimension of these local feature.
- **D<sup>3</sup>T-CWT with DFT** [15]:  
Features are computed by applying the discrete Fourier Transform DFT across the scale dimension of a feature vector of the D<sup>3</sup>T-CWT.
- **Fractal Dim. for Orientation Histograms** [18]:  
The first step is the computation of a local orientation histograms, followed by the computation of binary images which are based on the orientation histograms. The fractal dimension is then computed using the binary images. This is done for eight different scale levels. Finally a wavelet transform is applied across the scale dimension. The feature vector consists of the detail coefficients of this wavelet transform.
- **Cyclic Shifting of Local Features** [16]:  
The D<sup>3</sup>T-CWT is applied and a local feature vector for each pixel is computed. The local feature vectors are cyclically shifted across the scale dimension. A circular-correlation is computed between a specific mask and the squared feature vectors across the scale dimension. The original feature vector is then cyclically shifted in the scale dimension, so that the first scale level of the new local feature vector is at the same scale level as the original local feature vector.
- **Log-Polar Approach** [19]:  
The log-polar method transforms the image into the log-polar domain which converts scaling and rotation to translations. The DT-CWT [14] which is shift invariant is then applied. Therefore the DT-CWT features are scale invariant in the log-polar domain. The feature vectors are the subband coefficients' means and standard deviations.
- **Dominant Scale Approach** [20]:  
The accumulated energies of the image is computed for a set of scales across several orientations. The dominant scale is defined by the scale with the highest accumulated energy. We adapted the approach to use subband means instead of energies. The feature vectors consist of the means and standard deviations of the subband coefficients.
- **Slide Matching (original)** [21]:  
We use the slide matching approach with the D<sup>3</sup>T-CWT. The features con-

sist of the means and standard deviations of subbands with same scale level.

- **Slide Matching (modified)** [22]:  
In contrast to the original approach the means and standard deviations at the same scale level are not summed up. For each scale level 12 features (6 orientations, 2 parameters per subband) instead of 2 features for the slide matching process are used.
- **Local Affine Regions** [23, 24]:  
The Harris detector is used to identify interest points by using a multi-scale second moment matrix. The region described by the multi-scale second moment matrix is then mapped to a unit circle, finally SIFT [4] descriptors are computed. For each class cluster, centers of the aggregated SIFT descriptors are learned by using k-means clustering. The identified cluster centers are then used to label the SIFT descriptors with the closest distance to the respective center. The resulting histograms are used as features.
- **Multiscale Blob Features (shape)** [13]:  
The method uses a series of flexible threshold planes which are computed for an image to construct a set of binary images. Geometrical attributes are then used to describe the image texture. We use the shape of the blobs to form a feature vector.
- **Multiscale Blob Features (number)** [13]:  
In parallel to the Multiscale Blob Features (shape) multiple binary images are computed using flexible threshold planes. The number of identified blobs is then used as feature.
- **ICM** [25]:  
The intersecting cortical model (ICM) is a method derived from Pulse Coupled Neural Networks. Image data from the spatial domain is used as input to the ICM with a series of binary images as output. The entropies of the binary outputs are then used to form a feature vector.
- **SCM** [26]:  
The Spiking Cortical Model (SCM) is a derivation of the Pulse Coupled Neural Network. In analogy to ICM the output of the network is a series of binary images. The features consist of the entropies of the binary output.
- **Dense SIFT Features** [8]:  
Dense SIFT features are computed for each pixel within an image. For each class cluster, centers of the aggregated SIFT descriptors are learned, employing k-means clustering. The identified cluster centers are then utilized to label the SIFT descriptors with the closest distance to the respective center. The resulting histograms are used as features.

## 4.2 Data

The experimentation is based on two datasets with different properties. The first dataset is the CURET texture data set while the second dataset consists of a set of endoscopic images with indication for celiac disease.

### 1. CURET Data Set

We use the cropped version of the CURET database which consists of 92 images per texture with different viewing and illumination conditions. We use the CURET textures with available scale information only. The CURET textures were evenly divided into an evaluation set and two training sets. We follow the experimental setup used in [27] to test scale invariance. The first training set consists of textures at a single scale, while the second training set is based on textures at both scales. The evaluation set contains textures at both scales. Experiment 1 (E1) is based on the first training set while Experiment 2 (E2) is based on the second training set. The differences between classification accuracy will give an indication for the degree of scale invariance.

### 2. Celiac Disease Scale Data Set (CDS)

Endoscopic imagery exhibits a high amount of variance in perspective. Due to the tubic shape of the bowel, similar image textures are often captured in a rotated and scaled form. We created a database based on endoscopic video data with visually determined distances to the mucosal tissue. We took care that approximately the same mucosal area of the same patient is existent in both scales. We denote this database as the celiac disease scale (CDS) database. Table 1 shows the distribution of data. The captured endoscopic image data was inspected and filtered by several qualitative factors (sharpness, lack of distortions like specular reflections, visibility of features, etc.). In the next step, texture patches with a fixed size of  $128 \times 128$  pixels were extracted from the full sized frames, a size which turned out to be optimally suited in previous experiments [28]. The condition of the mucosal areas covered by the images was determined by histological examination of biopsies from the corresponding regions. Severity of villous atrophy was classified according to the modified Marsh classification proposed in [29].

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**. Scale invariance is only needed in the second experiment. The difference in classification rate of the first and second experiment (which are denoted as E3 and E4 respectively in Table 2) gives an indication for the degree of scale invariance of the methods.

<b>Training set Regular-Far</b>			
	<b>healthy</b>	<b>celiac</b>	<b>total</b>
<b>Images</b>	40	40	80
<b>Patients</b>	20	12	32

<b>Training set Regular</b>			
	<b>healthy</b>	<b>celiac</b>	<b>total</b>
<b>Images</b>	20	20	40
<b>Patients</b>	20	12	32

<b>Evaluation set Regular-Far</b>			
	<b>healthy</b>	<b>celiac</b>	<b>total</b>
<b>Images</b>	38	38	76
<b>Patients</b>	19	10	29

Table 1: Image and Patient Distribution of the CDS Database

### 4.3 Experimental Results

The presented results in Table 2 are the average classification accuracies using a k-NN classifier in a range of  $k = 1 - 20$ . The column labeled as Diff lists the relative differences in classification accuracy in percent. The methods were ordered by the average classification accuracy of all four experiments.

We see that the proposed method shows a high degree of scale invariance in both experimental setups. The non-uniform and rotational invariance is implicitly tested by using the CDS dataset where the proposed method is among the best. This indicates that the affine invariance is beneficial to the classification accuracy of endoscopic images.

## 5 Conclusion

We have utilized the information gained by computing local scale-normalized Laplacian maxima in a scale-space to employ an adaptive neighborhood. The scale information utilizing the scale-space representation was used to improve the scale invariance of the LBP method. We used elliptically formed neighborhoods to improve invariance to anisotropic scaling and used the orientation of eigenvectors of the multi-scale second moment matrices to add invariance to rotations.

We note that the use of uniform Gaussians for computing adaptive sampling areas leads to a diminished invariance to non-uniform scaling in case of highly anisotropic scaling. We see a lot of potential for improvement by using anisotropic Gaussian filters and sampling areas which should be employed in future work. Additionally the stated properties of invariance with respect to non-uniform scaling as well as rotation were only implicitly tested using the CDS database and have to be validated more thoroughly in future work.

Method	CURET			CDS		
	E1	E2	Diff	E3	E4	Diff
<b>Affine Invariant LTP</b>	<b>99.0</b>	<b>95.7</b>	<b>3.3</b>	<b>74.4</b>	<b>75.8</b>	<b>-1.9</b>
Cyclic Shifting of Local Features	98.8	95.1	3.7	73.3	72.8	0.7
D <sup>3</sup> T-CWT with DCT	98.3	88.7	9.8	76.1	73.8	3.0
D <sup>3</sup> T-CWT with DCT (local)	97.7	92.9	4.9	72.2	71.1	1.5
D <sup>3</sup> T-CWT with DFT (local)	96.3	89.7	6.9	72.9	73.0	-0.1
DT-CWT with DCT	98.4	87.1	11.5	75.3	70.9	5.8
Log-Polar Approach	90.7	88.3	2.6	72.6	73.4	-1.1
Multiscale Blob Features (shape)	96.9	93.9	3.1	72.0	62.0	16.1
Dominant Scale Approach	92.7	93.9	-1.3	72.4	64.8	11.7
Fractal Analysis using Filter Banks	91.1	85.8	5.8	71.8	70.5	1.8
D <sup>3</sup> T-CWT with DFT	95.9	89.2	7.0	63.7	69.1	-8.5
Multiscale Blob Features (number)	97.2	89.6	7.8	65.9	59.9	9.1
Local Affine Regions	96.1	89.8	6.6	67.6	59.0	12.7
Multi-Fractal Spectrum	91.4	77.0	15.8	69.1	71.1	-2.8
SCM	97.9	92.6	5.4	60.3	56.5	6.3
Fractal Dim. for Orientation Histograms	86.9	74.1	14.7	71.7	70.6	1.5
Slide Matching (modified)	97.6	75.3	22.8	63.2	60.5	4.3
ICM	90.2	81.4	9.8	64.8	59.8	7.7
Slide Matching (original)	93.5	81.4	12.9	58.8	54.3	7.6
Dense SIFT Features	71.5	67.4	5.7	68.1	66.7	2.1

Table 2: OCR results for the CDS and CURET database.

## References

1. T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, January 1996.
2. Xiaoyang Tan and Bill Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *Analysis and Modelling of Faces and Gestures*, volume 4778, pages 168–182, October 2007.
3. T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution Gray-Scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, July 2002.
4. D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer on Computer Vision*, volume 2, pages 1150 – 1157. IEEE, 1999.
5. Zhi Li, Guizhong Liu, Yang Yang, and Junyong You. Scale- and rotation-invariant local binary pattern using scale-adaptive texton and subuniform-based circular shift. *Image Processing, IEEE Transactions on*, 21(4):2130 –2140, april 2012.
6. Tony Lindeberg. *Discrete Scale-Space Theory and the Scale-Space Primal Sketch*. PhD thesis, Royal Institute of Technology, 1991.
7. Tony Lindeberg. *Scale-space theory in computer vision*, 1994.
8. L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society*, volume 2, pages 524 – 531, June 2005.

9. A. Uhl, A. Vécsei, and G. Wimmer. Fractal analysis for the viewpoint invariant classification of celiac disease. In *Proceedings of the 7th International Symposium on Image and Signal Processing (ISPA 2011)*, pages 727–732, Dubrovnik, Croatia, September 2011.
10. M. Varma and A. Zissermann. A statistical approach to texture classification from single images. *International Journal of Computer Vision (IJCV)*, 62(1–2):61–81, October 2005.
11. J. M. Geusebroek, A. W. M. Smeulders, and J. van de Weijer. Fast anisotropic gauss filtering. *IEEE Transactions on Image Processing*, 12(8):938–943, 2003.
12. Y. Xu, H. Ji, and C. Fermüller. Viewpoint invariant texture description using fractal analysis. *International Journal of Computer Vision*, 83(1):85–100, 2009.
13. Q. Xu and Y. Q. Chen. Multiscale blob features for gray scale, rotation and spatial scale invariant texture classification. In *Proceedings of 18th International Conference on Pattern Recognition (ICPR)*, volume 4, pages 29–32, September 2006.
14. I.W. Selesnick, R.G. Baraniuk, and N.C. Kingsbury. The dual-tree complex wavelet transform. *Signal Processing Magazine, IEEE*, 22(6):123–151, November 2005.
15. A. Häfner, A. Uhl, A. Vécsei, G. Wimmer, and F. Wrba. 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, November 2010.
16. E. H. S. Lo, M. R. Pickering, M. R. Frater, and J. F. Arnold. Query by example using invariant features from the double dyadic dual-tree complex wavelet transform. In *CIVR '09: Proceeding of the ACM International Conference on Image and Video Retrieval*, pages 1–8, Santorini, Fira, Greece, 2009. ACM.
17. E. H. S. Lo, M. R. Pickering, M. R. Frater, and J. F. Arnold. Scale and rotation invariant texture features from the dual-tree complex wavelet transform. In *Proceedings of the International Conference on Image Processing, ICIP '04*, volume 1, pages 227–230, Singapore, October 2004. IEEE.
18. Y. Xu, S.-B. Huang, and C. Fermüller H. Ji. Combining powerful local and global statistics for texture description. In *Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE*, pages 573–580, June 2009.
19. C.-M. Pun M.-C. Lee. Log-polar wavelet energy signatures for rotation and scale invariant texture classification. 25(5):590–603, May 2003.
20. J. A. Montoya-Zegarra, N. J. Leite, and R.S. Torres. 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*, pages 121–128, October 2007.
21. K.-K. Fung and K.-M. Lam. Rotation- and scale-invariant texture classification using slide matching of the gabor feature. In *Proceedings of Intelligent Signal Processing and Communication Systems*, pages 521–524, January 2009.
22. Sebastian Hegenbart, Andreas Uhl, Andreas Vécsei, and Georg Wimmer. Scale invariant texture descriptors for classifying celiac disease. *Medical Image Analysis*, 2013. accepted.
23. C. Schmid S. Lazebnik and J. Ponce. A sparse texture representation using local affine region. *Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1265–1278, 2005.
24. K. Mikolajczyk and S. Cordelia. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, October 2004.

25. Y. Ma, L. Liu, K. Zhan, and Y. Wu. Pulse coupled neural networks and one-class support vector machines for geometry invariant texture retrieval. *Image and Vision Computing*, 28(11):1524–1529, 2010.
26. K. Zhan, H. Zhang, and Y. Ma. New spiking cortical model for invariant texture retrieval and image processing. 20(12):1980–1986, December 2009.
27. M. Varma and A. Zisserman. A statistical approach to material classification using image patch exemplars. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(11):2032–2047, November 2009.
28. S. Hegenbart, R. Kwitt, M. Liedlgruber, A. Uhl, and A. Vecsei. 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)*, pages 718–723, Salzburg, Austria, September 2009.
29. G. Oberhuber, G. Granditsch, and H. Vogelsang. The histopathology of coeliac disease: time for a standardized report scheme for pathologists. *European Journal of Gastroenterology and Hepatology*, 11:1185–1194, nov 1999.