Local Fractal Dimension based approaches for Colonic Polyp Classification

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

[a] *University of Salzburg, Department of Computer Sciences, Jakob Haringerstrasse 2, 5020 Salzburg, Austria*
[b] *St. Elisabeth Hospital, Landstraßer Hauptstraße 4a, A-1030 Vienna, Austria*
[c] *Hiroshima University, Department of Information Engineering, Graduate School of Engineering, 1-4-1 Kagamiyama, Higashi-hiroshima, Hiroshima 739-8527, Japan*
[d] *Hiroshima University Hospital, Department of Endoscopy, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8551, Japan*

Abstract

This work introduces texture analysis methods that are based on computing the local fractal dimension (or also called the local density function) and applies them for colonic polyp classification. The methods are tested on 8 HD-endoscopic image databases, where each database is acquired using different imaging modalities (Pentax's i-Scan technology combined with or without staining the mucosa) and on a zoom-endoscopic image database using narrow band imaging (NBI). In this paper, we present three novel extensions to a local fractal dimension based approach. These extensions additionally extract shape and/or gradient information of the image to enhance the discriminativity of the original approach. To compare the results of the local fractal dimension based approaches with the results of other approaches, 5 state of the art approaches for colonic polyp classification are applied to the employed databases. Experiments show that local fractal dimension based approaches are well suited for colonic polyp classification, especially the three proposed extensions. The three proposed extensions are the best performing methods or at least among the best performing methods for each of the employed databases.

The methods are additionally tested by means of a public texture image database, the UIUCtex database. With this database, the viewpoint invariance of the methods is assessed, an important features for the employed endoscopic image databases. Results imply that most of the local fractal dimension based methods are more viewpoint invariant than the other methods. However, the shape, size and orientation adapted local fractal dimension approaches (which are especially designed to enhance the viewpoint invariance) are in general not more viewpoint invariant than the other local fractal dimension based approaches.

Keywords: polyp classification, local fractal dimension, texture recognition, viewpoint invariance

1. Introduction

In this paper, texture analysis methods are applied for the automated classification of colonic polyps in endoscopic images under unknown viewpoint and illumination conditions. Endoscopic images occur with different scales, orientations or perspectives, depending on the distance and perspective of the camera to the object. Figure 1 shows some examples for the field of view depending on the endoscopic viewpoint to the mucosal wall.

The varying viewpoint condition combined with the large intra-class and small inter-class variations of polyps make it very difficult to distinguish between different types of polyps. The viewpoint invariance of the employed methods is an important feature to at least reduce the problem with the varying viewpoint conditions.

(Uhl et al., 2011) and (Häfner et al., 2014c) showed that methods based on fractal analysis are able to combine viewpoint invariance with high discriminativity and are quite suitable for endoscopic image classification.

The term "fractal" was first used by the mathematician Benoit Mandelbrot as an indication of objects whose complex geometry cannot be characterized by an integral dimension. Fractal geometry is able to describe the irregular or fragmented shape of natural features as well as other complex objects that traditional Euclidean geometry fails to analyze. The fractal dimension is the key quantity to describe the fractal geometry and the heterogeneity of irregular shapes. Roughly spoken, the fractal dimension is a ratio that compares how the detail of a shape changes with the scale at which it is measured.

However, the fractal dimension is only one number, which is not enough to describe an object.

As an extension to the classical fractal analysis, multifractal analysis provides more powerful descriptions. Applied to image processing, first define a point characterization on an image according to some criteria (e.g. the intensity values of the pixels), then the fractal dimensions are computed for every point set from this categorization (e.g. categorize the image pixels by their intensity and ob-

*Corresponding author
Email addresses: uhl@cosy.sbg.ac.at (Andreas Uhl), gwimmer@cosy.sbg.ac.at (Georg Wimmer)

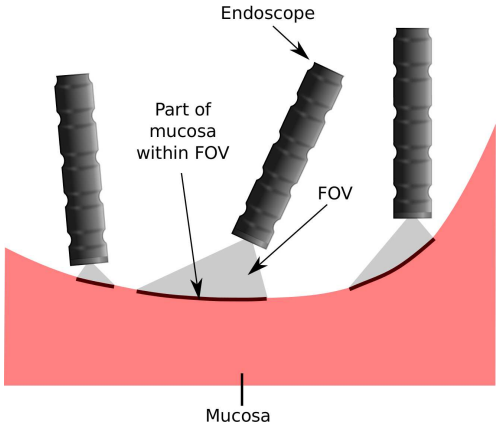


Figure 1: The field of view (FOV) depending on the endoscopic viewpoint to the mucosal wall

tain binary images by setting a pixel to 0 if its intensity value is in the considered set and to 1 otherwise). The collection of the fractal dimensions of the binary images is called a multi fractal spectrum (MFS) vector.

Another extension to the classical analysis to provide a more powerful description is to compute local fractal features. These features are already the norm in fractal based image segmentation (Chaudhuri and Sarkar, 1995; Xia et al., 2006).

In Xu et al. (2009), local fractal based features (we denote them as local fractal dimensions) are computed densely followed by applying multifractal analysis to these features (categorize the local fractal dimensions by their values, thereby obtain binary images followed by computing the fractal dimension of the binary images). Another approach (Varma and Garg, 2007) using the local fractal dimension (LFD) is pre-filtering the image with the MR8 filter bank obtaining 8 filtered images on which the local fractal dimensions are computed. Subsequently, the bag of visual words approach is used to build histograms of the LFD's. It has been shown that the LFD is invariant to bi-Lipschitz transformations, such as local affine or perspective transformations and certain smooth, non-linear transformations (Xu et al., 2009). The LFD is also invariant to local affine illumination changes as showed in Xu et al. (2009).

Roughly speaking, the LFD at an arbitrary location of an image is computed by summing up intensity values in disk shaped areas with fixed radii surrounding the considered (pixel) location followed by analyzing the increase of the sums for increasing radii. Actually, the scale and perspective of the object or texture in the image at the considered location is not taken into account, the radii are always the same and the areas are always disk shaped. In Häfner et al. (2014c), a more viewpoint adaptive approach is presented. This LFD based approach uses ellipsoidal areas instead of disk shaped areas. The sizes, shapes and orientations of the ellipsoidal areas are adapted to the local texture structure by analyzing the shape, size and orientation of connected components (blobs). Instead of a dense computation of the LFD's like in Xu et al. (2009) and Varma and Garg (2007), the size, shape and orientation adapted LFD's in Häfner et al. (2014c) are computed only for interest points, more precisely only for those points that are the centers of the area of a blob.

A review about methods using fractal and multifractal analysis is presented in Lopes and Betrouni (2009).

In this work we compare methods based on the LFD, compare their classification results on different image databases, analyze the reasons for those results and examine the affine invariance of the methods. We will test the LFD approaches on 9 different endoscopic image databases, which consist of highly detailed endoscopic images with 9 different imaging modalities. Additionally we apply the LFD based approaches on a public texture database with huge viewpoint variations, the UIUCtex database (S. Lazebnik and Ponce, 2005).

The contributions of this manuscript are as follows:

- We apply 7 LFD based methods for the automated classification of colonic polyps using 9 different endoscopic image databases. 8 databases are gathered using a HD-endoscope with 8 different imaging modalities (Pentax's i-Scan in combination with staining the mucosa) and one database is gathered using a zoom-endoscope with NBI as imaging modality. To the best of our knowledge, this is the highest number of endoscopic polyp databases that has been used in publications so far. The results of the LFD based methods are compared and the differences between the methods and their impacts to the results are analyzed.
- 5 (non LFD based) state-of-the-art approaches for colonic polyp classification are applied to the classification of our databases to compare their results with the results of the LFD based methods.
- We present three novel extensions of an LFD approach. For each database, the results of these extensions are among the best results of all the employed methods.
- We assess the viewpoint invariance of the methods by means of the a public texture database, the UIUCtex database (S. Lazebnik and Ponce, 2005). Results imply, that most of the LFD based methods are more viewpoint invariant than the other methods. The size, shape and orientation adapted LFD methods are generally not more viewpoint invariant than the other LFD based methods.

Already in Häfner et al. (2014c), a LFD-based method was proposed for the classification of colonic polyps. However, this publication used only one endoscopic image database (one of our 8 HD-endoscopic image databases) and compared the result of the proposed method with only one other LFD based approach and three non LFD based approaches. Furthermore, neither the differences between

the two LFD based approaches were analyzed nor the viewpoint invariance of the approaches was tested.

This paper is organized as follows. In Section 2 we briefly introduce the concept of the computer-assisted diagnosis of polyps by the automated classification of mucosa texture patches and review the corresponding state-of-the-art. In Section 3, we describe the feature extraction approaches and compare the approaches that are based on computing the LFD. The experimental setup, the used databases and the results are presented in Section 4. Section 5 presents the discussion and Section 6 concludes our work. The acronyms used in this work are listed in the Appendix.

2. Colonic Polyp Classification

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

Colonic polyps are a frequent finding and are usually divided into hyperplastic, adenomatous and malignant. In order to determine a diagnosis based on the visual appearance of colonic polyps, the pit pattern classification scheme was proposed by (Kudo et al., 1994). A pit pattern refers to the shape of a pit, the opening of a colorectal crypt. This classification scheme allows to differentiate between normal mucosa and hyperplastic lesions, adenomas (a pre-malignant condition), and malignant cancer based on the visual pattern of the mucosal surface. The removal of hyperplastic polyps is unnecessary and the removal of malignant polyps maybe hazardous. Thus, this classification scheme is useful to decide which lesions need not, which should, and which most likely cannot be removed endoscopically. For these reasons, assessing the malignant potential of lesions at the time of colonoscopy is important, as this would allow to perform targeted biopsy.

The various pit pattern types are presented in Figure 3 e–f. The pit pattern classification scheme differentiates between six types. Type I (normal mucosa) and II (hyperplastic polyps) are characteristics of non-neoplastic lesions, type III-S, III-L and IV are typical for adenomatous polyps and type V is strongly suggestive to malignant cancer.

To enable an easier detection and diagnosis of the extent of a lesion, there are two common image enhancement technologies:

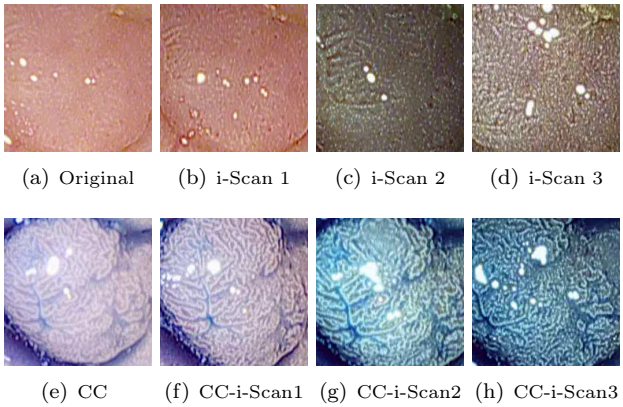
1. Conventional chromoendoscopy (CC) came into clinical use 40 years ago. By staining the mucosa using (indigocarmine) dye spray, it is easier to find and classify polyps.
2. Digital chromoendoscopy is a technique to facilitate "chromoendoscopy without dyes" (Kiesslich, 2009). The strategies followed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

Previous works for the computer assisted staging of colon polyps, which are using endoscopes producing highly detailed images in combination with different imaging modalities, can be divided in three categories: High definition (HD) endoscope combined with or without staining the mucosa and the i-Scan technology (Häfner et al., 2014c), high-magnification chromoendoscopy (Häfner et al., 2009) and high-magnification endoscopy combined with NBI (Tamaki et al., 2013; Gross et al., 2012). In this work we use highly detailed images acquired by a high definition (HD) endoscope without magnification in combination with CC and CVC (the i-Scan technology) and images acquired by a high-magnification endoscope combined with NBI.

Further examples of approaches for colonic polyp classification classification are Iakovidis et al. (2005); Karkanis et al. (2003); Maroulis et al. (2003); Iakovidis et al. (2006).

In addition to classical endoscopy, endomicroscopy and wireless capsule endoscopy are used for the examination of the gastro-intestinal tract. Endomicroscopy (Jabbour et al., 2012) is a technique to obtain histology-like images and is also known as 'optical biopsy'. For example Andrė et al. (2011) and Andrė et al. (2012) show state of the art approaches based on semantics and visual concepts for the automated diagnosis of colonic polyps using endomicroscopy.

Wireless capsule endoscopy (Iakovidis and Koulaouzidis (2015); Yuce and Dissanayake (2012)) is mainly used to examine parts of the gastrointestinal tract that cannot be seen with other types of endoscopes. The capsule has the size and shape of a pill and contains a tiny camera. After a patient swallows the capsule, it takes images of the inside



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

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

Figure 2: Images of a polyp using digital (i-Scan) and/or conventional chromoendoscopy (CC)

of the gastro-intestinal tract. An example for the automated detection and classification of colonic polyps using capsule endoscopy can be seen in Romain et al. (2013).

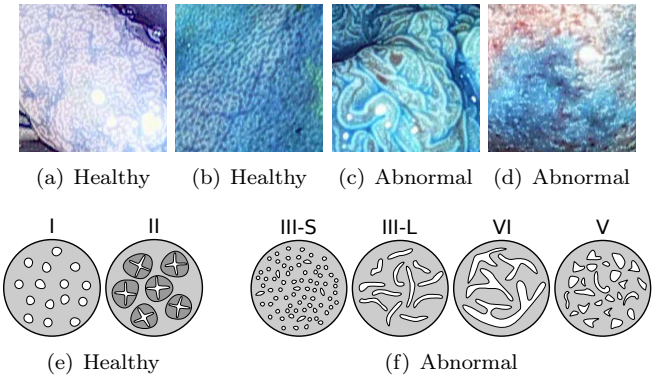
2.1. HD endoscopy in combination with the i-Scan image processing technology

In this work, the HD endoscopic images are gathered using three different i-Scan modes:

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

In Figure 2 we see an image showing an adenomatous polyp without image enhancement technology (a), example images using CVC (b,c,d), an image using CC (e) and images combining CC and CVC by using the i-Scan technology to visually enhance the already stained mucosa (f,g,h).

In our work we use a 2-class classification scheme for our 8 image databases gathered by HD endoscopy in combination with CC and the i-Scan technology. Lesions of pit pattern type I and II can be grouped into non-neoplastic lesions (healthy mucosa) and types III to V can be grouped into neoplastic lesions (abnormal mucosa). This allows a grouping of lesions into two classes, which is quite relevant in clinical practice as indicated in a study by (Kato et al., 2006). In Figure 3 we see the various pit pattern types divided into two classes (denoted as class "Healthy" and class "Abnormal") along with exemplar images of these two classes obtained by a HD endoscope using CC and i-Scan mode 2.



(a) Healthy (b) Healthy (c) Abnormal (d) Abnormal

(e) Healthy (f) Abnormal

Figure 3: Example images of the two classes (a–d) and the pit pattern types of these two classes (e–f)

One of the aims of this work is to compare classification results with respect to using CVC (i-Scan) or CC (staining). We will also examine the effects of combinations of CVC and CC on the classification results.

2.2. High-magnification endoscopy in combination with NBI

NBI (Gono et al., 2003) is a videoendoscopic system using RGB rotary filters placed in front of a white light source to narrow the bandwidth of the spectral transmittance. NBI enhances the visibility of microvessels and their fine structure on the colorectal surface. Also the pits are indirectly observable, since the microvessels between the pits are enhanced in black, while the pits are left in white. In this paper we use the classification scheme of the medical research group of the Hiroshima University Hospital (Kanao et al., 2008). This classification scheme divides the microvessel structure in an NBI image into types A, B and C. In type A microvessels are either not or only slightly observed (opaque with very low contrast). In type B, fine microvessels are visible around clearly observed pits. Type C is divided into three subtypes C1, C2, and C3. In type C3, which exhibits the most irregular texture, pits are almost invisible because of the irregularity of tumors, and microvessels are irregular and thick, or heterogeneously distorted. In Figure 4 we see examples from the classes A, B and C3 (without CC).

It has been shown that this classification scheme has a strong correlation with histological diagnosis (Kanao et al., 2008). 80% of type A corresponds to hyperplasias and 20% to tubular adenomas. 79.7% of type B

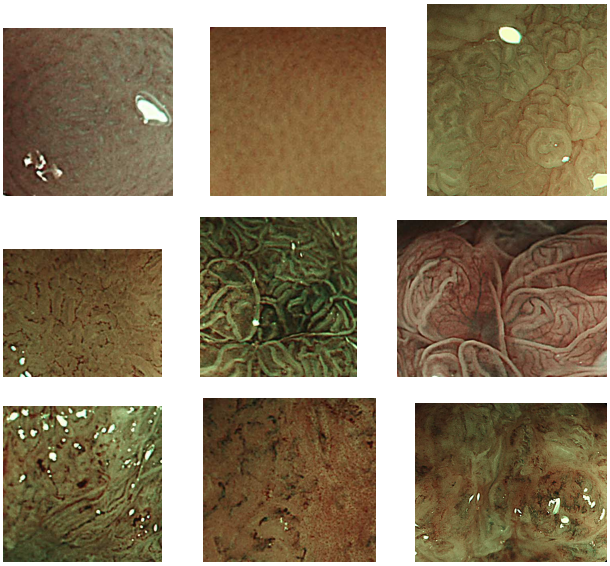


Figure 4: Example of NBI images of types A (top row), B (middle row) and C3 (bottom row)

corresponds to tubular adenomas and 20.3% to carcinomas with intramucosal invasion to scanty submucosal invasion. 100% of type C3 correspond to carcinomas with massive submucosal invasion. Intramucosal invasion to scanty submucosal invasion (Pit Pattern type V_I) demands further examinations and carcinomas with massive submucosal invasion (Pit Pattern type V_N) requires surgery. Therefore it is important to detect type C3 among other types, instead of differentiating just between the two classes of neoplastic and non-neoplastic lesions. Like in Kanao et al. (2008) and Tamaki et al. (2013), types C1 and C2 are excluded from the experiments of this paper.

3. Local Fractal Dimension based Feature Extraction Approaches

3.1. The Fractal Dimension

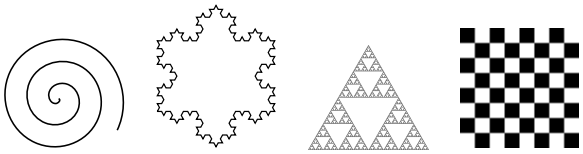
As already mentioned in the introduction, the fractal dimension is the key quantity to describe the fractal geometry and the heterogeneity of irregular shapes. Fundamental to the fractal dimension is the concept of "measurements at scale σ". For each σ, we measure an object in a way that ignores irregularity of size less than σ, and we analyze how these measurements behave as σ goes to 0. A well-known example to illustrate this concept is the length of a coastline measured with differently long measuring sticks (see Figure 5).

For most natural phenomena, the estimated quantity (e.g. the length of a coast) is proportional to $(1/\sigma)^D$ for some D. For most natural objects, D is almost the same for small scales σ. Its limit D for $\sigma \to 0$ is defined as the local fractal dimension. In case of an irregular point set E defined on $\mathbb{R}^2$, the fractal dimension of E is defined as

$$dim(E) = \lim_{\sigma \to 0} \frac{\log(N(\sigma, E))}{-\log \sigma}, \qquad (1)$$

(a) $11.5 \times 200 = 2300km$ (b) $28 \times 100 = 2800km$ (c) $70 \times 50 = 3500km$

Figure 5: As the length of the measuring stick is decreasing, the total length of the coastline measured is increasing.



(a) (b) (c) (d)

Figure 6: Fractal dimension D in 2D space. (a) Smooth spiral curve with $D = 1$, (b) the Koch snowflake with $D \approx 1.26$ (c) the Sierpinski-Triangle with $D \approx 1.58$ and (d) the checkerboard with $D = 2$.

where $N(\sigma, E)$ is the smallest number of sets of diameter less than sigma that cover E. The set consists of closed disks of radius σ or squares of side σ. In practice, the fractal dimension is usually computed using the box counting method (dividing the space with a mesh of quadratic boxes of size $\sigma \times \sigma$, and counting the boxes occupied by the point set).

The fractal dimension D of any object in $2D$ space is between 0 and 2. The fractal dimensions of a point, a smooth curve or a completely filled rectangle is the same as their topological dimension (0, 1 and 2). Irregular sets have a fractal dimension between 0 and 2 (see Figure 6). For example a curve with fractal dimension very near to 1 behaves similar to an ordinary line, but a curve with fractal dimension close to 2 winds convolutedly through space very nearly like a surface.

3.2. The Local Fractal Dimension

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. As an example, $\mu(B(x, r))$ could be the sum of all pixel intensities that lie within a closed disk of radius r centered at an image point x, i.e. $\mu(B(x, r)) = \sum_{||y-x|| \leq r} I(y)$.

The local fractal dimension (Xu et al., 2009) (or also

called the local density function) of x is defined as

$$LFD(x) = \lim_{r \to 0} \frac{\log \mu(B(x,r))}{\log r}. \quad (2)$$

The LFD measures the "non-uniformness" of the intensity distribution in the region neighboring the considered point. In Figure 7 we show examples of values of the LFD for different intensity distributions. If the intensities decrease from the center outwards, then the center point has a LFD < 2. For uniform intensities, the LFD $= 2$. Finally, if the surrounding intensities increase from the center outwards, the LFD of the center point is > 2.

(a) (b) (c)

Figure 7: LFD's at the center point using $\mu(B(x,r)) = \sum_{||y-x|| \leq r} I(y)$: (a) LFD=1.64, (b) LFD=2 (c) LFD=2.37.

In that way, the pit pattern structure of the mucosa provide high responses in terms of the LFD. Pits produce high LFD values and the peaks of the pit pattern structure produce low LFD values. So the LFD response is highlighting the pit pattern structure of the mucosa. In Figure 8 (a) and (b) we see an image of class abnormal and its LFD's and in (c) and (d) we see an image of healthy mucosa and its LFD's (both images are gathered using a HD endoscope combined with i-Scan mode 2).

As already mentioned before, the LFD 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 (Xu et al., 2009). A bi-Lipschitz function g must be invertible and satisfy the constraint $c_1||x-y|| \leq ||g(x)-g(y)|| \leq c_2||x-y||$ for $c_2 \geq c_1 > 0$. The core of the proof in Xu et al. (2009) shows that for an bi-Lipschitz transform g applied to an image $I(x)$ with $I'(x) = I(g(x))$, the LFD of $I(x)$ and $I(g(x))$ are identical:

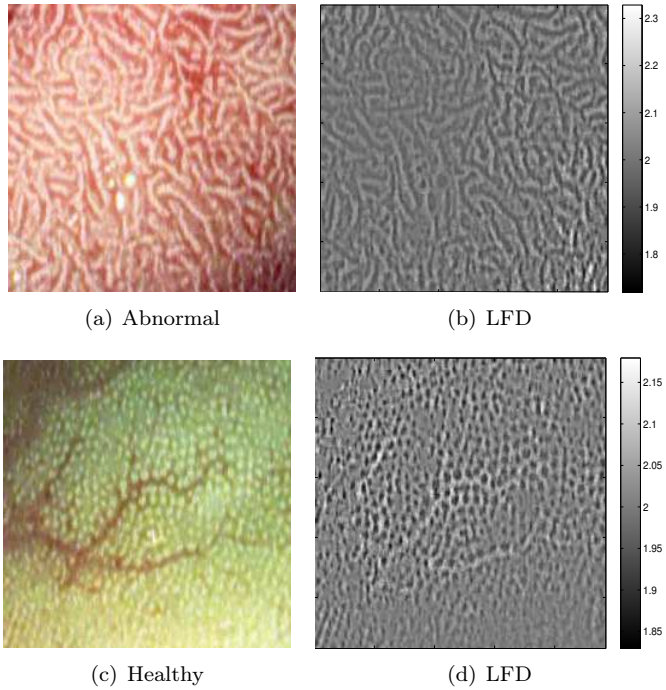
$$\frac{\log(c_1^2 \mu(B(x,r)))}{\log r} \leq \frac{\log(\mu(B(g(x),r)))}{\log r} \leq \frac{\log(c_2^2 \mu(B(x,r)))}{\log r}.$$

Since

$$\lim_{r \to 0} \frac{\log(c_i^2 \mu(B(x,r)))}{\log r} = \lim_{r \to 0} \frac{2 \log c_i}{\log r} + \lim_{r \to 0} \frac{\log \mu(B(x,r))}{\log r}$$

for $i \in \{1,2\}$ and since $\frac{\log 2c_i}{\log r}$ is zero for $r \to 0$ ($\log r \to -\infty$), the fractal dimensions $D(x)$ and $D(g(x))$ are identical.

However, the proof shows that the LFD is invariant in a continuous scenario, but not in case of a discrete scenario (e.g. an image), since $r \to 0$ is not possible for an image with limited resolution. So the LFD is not proven to be viewpoint invariant in case of any image processing tasks. Of course, total viewpoint invariance in image processing tasks is impossible since images appear totally different for huge differences in scale. Despite their missing actually viewpoint invariance, the viewpoint invariance of the two approaches using the LFD (Xu et al., 2009; Varma and Garg, 2007) seems to be sufficient to achieve high classification rates on the UIUCtex database (S. Lazebnik and Ponce, 2005), a texture database consisting of texture images which are acquired under quite different viewpoint conditions.



(a) Abnormal (b) LFD

(c) Healthy (d) LFD

Figure 8: Example images of class abnormal and healthy and their LFD's using $\mu(B(x,r)) = \sum_{||y-x|| \leq r} I(y)$.

In practical computation, the LFD at each pixel location x of an image is computed by linear fitting the slope of the line in a scaling plot of $\log \mu(B(x,r))$ against $\log r$ for $r = \{1, \ldots, 8\}$. In Figure 9, we visually show the computation of the LFD for the pixel location x of an image I using $\mu(B(x,r)) = \int_{B(x,r)} I(x)\, dx = \sum_{||y-x|| \leq r} I(y)$.

3.3. Feature extraction methods based on the LFD

3.3.1. The MFS-LFD approach

In the approach of Xu et al. (2009), three different definitions of $\mu(B(x,r))$ are used, which capture different aspects of the structure of textures:

$$\mu_1(B(x,r)) = \int_{B(x,r)} I(\sigma)\, dx \quad (3)$$

$$\mu_2(B(x,r)) = \int_{B(x,r)} \sum_{k=1}^{4} (f_k * (I(\sigma)^2)^{\frac{1}{2}}\, dx \quad (4)$$

$$\mu_3(B(x,r)) = \int_{B(x,r)} |I_{xx}(\sigma) + I_{yy}(\sigma)|\, dx, \quad (5)$$

where $I(\sigma)$ is the Gaussian blurred image I using variance σ^2, $I_{xx}(\sigma)$ is the second derivative in x-direction, " $*$ " is

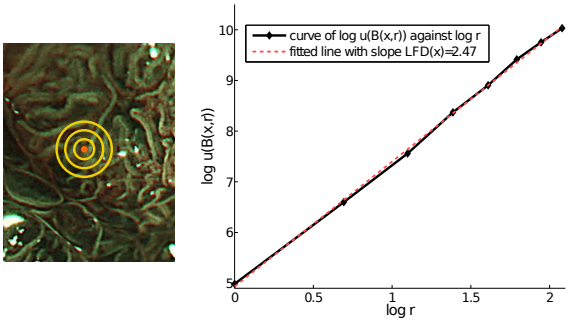


Figure 9: In the image to the left we see the schematic representation of a pixel location x (orange dot) and the corresponding disks $B(x,r)$ (yellow). The plot to the right visually shows the computation of the LFD by linear fitting the slope of the line of $\log \mu(B(x,r))$ against $\log r$.

the 2D convolution operator and $\{f_k,\ k = 1,2,3,4\}$ are four directional operators (derivatives) along the vertical, horizontal, diagonal, and anti-diagonal directions.

Let E_α be the set of all image points x with LFD's in the interval α:

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

Usually this set is irregular and has a fractional dimension $f(\alpha) = dim(E_\alpha)$. The feature vector of an image I consists of the concatenation of the fractal dimensions $f(\alpha)$ for the three different measures $\mu_k(B(x,r)),\ k \in \{1,2,3\}$.

That means the range of values of the LFD's is splitted into N equally sized intervals $\alpha_i,\ i \in \{1,\ldots,N\}$ ($N = 26$ in Xu et al. (2009)). So for each of the three measures $\mu_k(B(x,r))$, we generate 26 binary images $I_b^{\alpha_i}$, where $I_b^{\alpha_i}(x,y) = 1$ if $LFD(x,y) \in \alpha_i$ and $I_b^{\alpha_i}(x,y) = 0$ otherwise. The final feature vector consists of the fractal dimensions of the 26 binary images per measure $\mu_k(B(x,r))$. So the feature vector of an image consists of $3 * 26 = 78$ features per image. We furtherly denote this approach as the multi fractal spectrum LFD (MFS-LFD) approach.

3.3.2. The MR8-LFD approach

In the approach presented in Varma and Garg (2007), the images are convoluted with the MR8 filter bank (Varma and Zissermann, 2005; Geusebroek et al., 2003), a rotationally invariant, nonlinear filterbank with 38 filters but only 8 filter responses. It contains edge and bar filters, each at 6 orientations and 3 scales, as well as a rotationally symmetric Laplacian and Gaussian filter (see Figure 10). Rotation invariance is achieved by taking only the maximum response over all orientations for each scale of the edge and bar filters.

The LFD's are computed for each of the 8 filter responses $f_i(I),\ i \in \{1,\ldots,8\}$ using the measure

$$\mu(B(x,r)) = \int_{B(x,r)} |f_i(I)|\, dx.$$

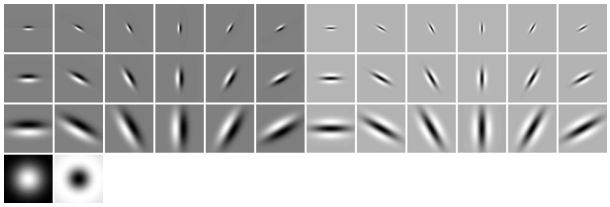


Figure 10: The filters of the MR8 filter bank

So for each pixel of an image there is an 8-dimensional LFD vector. Finally, the bag of visual words approach is applied to the LFD vectors. The visual words are learned by k-means clustering the LFD vectors using 100 cluster centers per image class. The feature vector of an image consists of the resulting histograms of the bag of visual words approach. We furtherly denote this approach as the MR8-LFD approach.

For both, the MFS-LFD and the MR8-LFD approach, disks $B(x,r)$ with $r = \{1,\ldots,8\}$ are used to sum the intensity values $I(y)$ (where $I(y)$ is the Gaussian blurred image $I(\sigma)$, the gradient image or the Laplacian of the image in case of the MFS-LFD approach and one of the 8 MR8 filter responses in case of the MR8-LFD approach) surrounding the considered pixel x with $||x-y|| \leq r$. We can interpret these disks as circle shaped binary filters, with which the image (respectively its filter responses or its derivatives) is filtered.

3.3.3. The Blob-Adapted LFD approach

In Häfner et al. (2014c), we proposed a feature extraction method that is derived from the local fractal dimension. However, instead of disk shaped filters with preassigned radii ($B(x,r)$), we used ellipsoidal binary filters and anisotropic, ellipsoidal Gaussian filters fitted to the shape, size and orientation of the local texture structure. The shapes, orientations and sizes of the filters are adapted to the shapes, orientations and sizes of connected components (blobs).

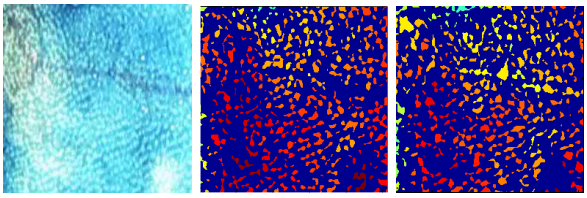
These blobs are generated by a segmentation algorithm (Häfner et al., 2014c), that applies local region growing to the maxima and minima of the image in a similar way as the watershed segmentation by immersion (Vincent and Soille, 1991; Roerdink and Meijster, 2000).

The blobs represent the local texture structures of an image. We differentiate between blobs evolved from local minima (pit blobs) and blobs evolved from local maxima (peak blobs) of an image (see Figure 11). Roughly said, beginning with a local minima (maxima), the algorithm adds those neighboring pixels to the considered minima (maxima), which have the smallest (highest) intensity value of all neighboring pixels. In this way we generate a blob and this blob is growing as long as the darkest (brightest) neighboring pixel of the blob is brighter (darker) or equally bright (dark) as the brightest (darkest) pixel of the blob. If the darkest (brightest) neighboring pixel is darker (brighter) as the brightest (darkest) pixel of the blob, the

region growing algorithm stops resulting in a pit (peak) blob b evolved from the local minima (maxima).

The idea behind this segmentation approach is that different classes of polyps have different typical pit pattern types (see Figure 3). By filling up the pits and peaks of a mucosal image, the resultant blobs represent the shapes of local structures of the image including the different types of pit pattern. In that way the shape of the blobs contain information that enables an distinction between healthy and abnormal mucosa (see Häfner et al. (2014a)).

For further feature extraction (computing the local fractal dimension derived feature), only the blobs with $N \geq 8$ pixels are used. In this way it is ensured that only these blobs are used which represent a distinct pit or peak and exclude those blobs which evolve of minima or maxima that are caused by noise. For each resulting blob, the inertia matrix is computed and from these matrices we determine the eigenvectors and eigenvalues.



(a) Image (b) Peak blobs (c) Pit blobs

Figure 11: The extracted peak and pit blobs of the image

The orientation and shape of an elliptic filter is derived from the eigenvectors and eigenvalues of the inertia matrix of a blob. That means for each blob b, a specific filter is generated and its shape and orientation is adapted to the considered blob. The size of the elliptic filters is adapted to the number of pixels of the corresponding blob (the higher the number of pixels, the bigger the size of the filter).

Like in the two previous approaches, 8 differently sized binary filters are used (disks $B(x,r)$ with $r = \{1,\dots,8\}$ in case of the previous approaches). The size of the 8 elliptic binary filters is controlled by 8 threshold parameters $t_i \times \sqrt{N/\pi}$, $i \in \{1,\dots 8\}$ (t_i, $i \in \{1,\dots 8\}$ is fixed and strictly monotonic increasing and N is the number of pixels of the considered blob). Additionally to the 8 binary filters $E_b^{t_i}$, 8 Gaussian filters are used, whose shape and orientation is equally determined as for the binary filters. Instead of the threshold parameters t_i, 8 standard deviations $\sigma_i \times \sqrt{N/\pi}$, $i \in \{1,\dots 8\}$ are used as size-perimeters for the Gaussian filters $G_b^{\sigma_i}$ (see (Häfner et al., 2014c)), where σ_i, $i \in \{1,\dots 8\}$ is fixed and strictly monotonic increasing.

The parameters t_i and σ_i are chosen so that the filters uniformly gain in size with increasing i.

In Figure 12 we see an image patch containing a blob b and the corresponding binary and Gaussian filters.

For a given Blob b with center position $(\overline{x},\overline{y})$ in the image I and the corresponding filters $G_b^{\sigma_i}$ ($E_b^{t_i}$ analogous)

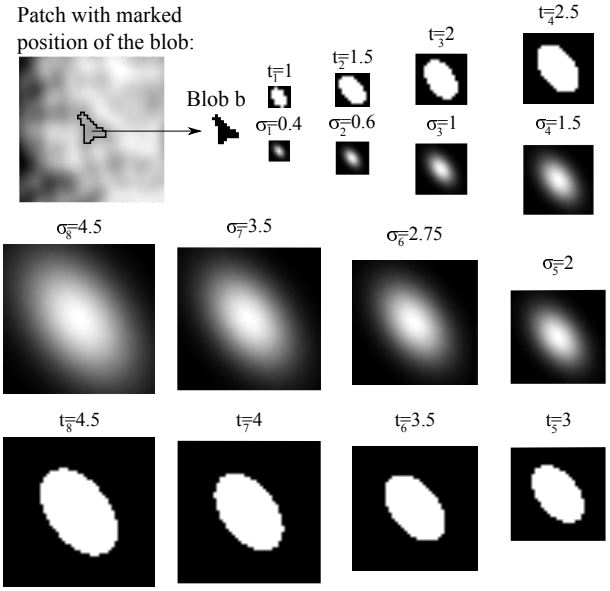


Figure 12: A patch containing a blob b in his center and the corresponding binary elliptic filter masks $E_b^{t_i}$ and elliptic Gaussian filter masks $G_b^{\sigma_i}$.

with filter size $f \times f$, μ is defined as follows:

$$\mu(G_b^{\sigma_i}) = \sum_{x=-\frac{f-1}{2}}^{\frac{f-1}{2}} \sum_{y=-\frac{f-1}{2}}^{\frac{f-1}{2}} I(\overline{x}-x,\overline{y}-y) \quad G_b^{\sigma_i}(x,y)$$

The LFD derived features are computed separately for binary and Gaussian filters and only for interest points, which are defined as the centers of the blobs. The two local fractal dimensions derived features for a Blob b are defined as:

$$LFD_E(b) = \lim_{i\to 0} \frac{\log \mu(E_b^{t_i})}{\log i}, LFD_G(b) = \lim_{i\to 0} \frac{\log \mu(G_b^{\sigma_i})}{\log i}, \tag{6}$$

where σ_i and t_i are strictly monotonic increasing. Equally to the original LFD, the practical computation of the LFD_E (LFD_G) is done by linear fitting the slope of the line in a scaling plot of $\log \mu(E_b^{t_i})$ ($\log \mu(G_b^{\sigma_i})$) against $\log i$ with $i \in \{1,\dots,8\}$.

Since the two features LFD_E and LFD_G in this approach are derived from the LFD as defined in the two previous approaches (MFS-LFD and MR8-LFD), we will further denote them as blob-adapted LFD (BA-LFD).

The BA-LFD measures the "non-uniformity" of the intensity distribution in the region and neighboring region of a blob. Starting with the center region of a pit or peak, it analyzes the changes in the intensity distribution with expanding region. In that way it analyzes the changing intensity distribution from the inside to the outside of a pit or peak in an image. Since size, shape and orientation of the filters are adapted to the blob representing the pit or peak, the BA-LFD should be even more invariant to varying viewpoint conditions as the LFD using disks with fixed radii (Xu et al., 2009).

The BA-LFD approach was especially designed to classify polyps using the CC-i-Scan databases. It finds the pits and peaks of the pit pattern structure and then filters the area in and surrounding the detected pits with filters that are shape, size (= scale) and orientation adapted to the pits and peaks.

The final feature vector of an image consists of the concatenation of the histograms of the LFD_E's separately computed for the pit and peak blobs of an image and the histograms of the LFD_G's separately computed of the pit and peak blobs of an image. Each of the 4 histograms consists of 15 bins. All parameter values (e.g. the number of bins per histogram, σ_i and t_i) are taken from the original approach (Häfner et al., 2014c).

Distances between two feature vectors are measured using the χ^2 statistic, which has been frequently used to compare probability distributions (histograms) and is defined by

$$\chi^2(x,y) = \sum_i \frac{(x_i - y_i)^2}{x_i + y_i}. \tag{7}$$

Also the 3 extensions of the BA-LFD approach (see Section 3.5) use the χ^2 statistic as distance metric. The histograms of the BA-LFD approach (and its 3 extensions) are not normalized. In case of the experiments using the NBI database, the values of the histograms of an image are divided by the number of pixels of the considered image, to balance the different sizes of the NBI images. This approach will be further denoted as the BA-LFD approach.

3.4. Closing the gap between LFD and BA-LFD

As already mentioned before, there are major differences between the LFD and the BA-LFD. Contrary to the LFD, the filters of the BA-LFD are

- scale-adapted by fitting the size of the filters to the number of pixels per blob,
- shape and orientation-adapted by fitting the shape and orientation of the filters to the shape and orientation of the blobs,
- only applied on interest points, which are defined as the centers of peak and pit blobs that are detected by an segmentation algorithm,
- partly Gaussian filters and partly binary filters (instead of only binary filters).

To analyze the weak and strong points of the BA-LFD compared to the LFD and to analyze which of the adaptions make sense and which not, we will create methods that are intermediate steps between the LFD and the BA-LFD. That means we leave out one or several of the four adaptation steps that turn the LFD into the BA-LFD. For a better comparability of the results, for each intermediate step the histograms of the LFD's (or BA-LFD's) are used as features. It should be noted that the computation of LFD's (BA-LFD's) only on interest points means that we separately compute histograms of the LFD's (BA-LFD's) of pit and peak blobs, whereas a dense computation of the LFD's means that we compute only one histogram of the LFD's.

	Adaption			
Nr.	Scale	Shape	Int. Points	Gaussian Filters
1	x	x	x	x
2	x	x	x	✓
3	x	x	✓	x
4	x	x	✓	✓
5	✓	x	✓	x
6	✓	x	✓	✓
7	✓	✓	✓	x
8	x	✓	✓	✓
9	✓	✓	✓	✓

Table 1: The adaptions of each of the 9 intermediate steps beginning with the DLFD (1) and ending with the BA-LFD (9).

Altogether, we analyze 9 methods that are intermediate steps between LFD and BA-LFD:

1. Dense computation of the LFD's without any adaption and disk radii $r = 1 - 8$. We furtherly denote this approach as dense LFD (DLFD).
2. Like in (1.), but we additionally use isotropic Gaussian filters with standard deviations σ_i, $i \in \{1, \dots 8\}$ without any scale-adaption.
3. The LFD's are computed like in (1.), but only on interest points.
4. The LFD's are computed only on interest points like in (3.). Additionally Gaussian filters are used (like in 2.)).
5. The LFD's are computed on interest points and the sizes of the circle shaped binary filters are adapted to the number of pixels of the blobs and the thresholds t_i, $i \in \{1, \dots 8\}$.
6. Like in (5.), but we additionally use non-isotropic (elliptic) Gaussian filters whose size is adapted to the number of pixels of the blobs.
7. Like the BA-LFD approach, but without Gaussian filters. The difference to (5) is the elliptic shape and the adapted orientation of the filters.
8. Like the BA-LFD approach, but without the scale adaption of the binary and Gaussian filters. The differences to (4.) are the elliptic shape and the adapted orientation of the filters and the use of the size parameters t_i, $i \in \{1, \dots 8\}$ instead of the disk radii $r = 1 - 8$.
9. The BA-LFD approach.

In Table 1, we see which of the 4 major differences between the LFD and BA-LFD (the adaptions of the BA-LFD to the LFD) are applied to each of the 9 intermediate steps between LFD and BA-LFD.

3.5. Extensions to the BA-LFD approach

In this section we propose three new variations of the BA-LFD approach.

3.5.1. The Blob-Adapted Gradient LFD Approach

This approach especially analyzes the edge information of an image. First the BA-LFD approach is applied to the image. In the second part of the approach we apply the BA-LFD approach to the gradient magnitude image I_G with

$$I_G = \sqrt{I_x^2 + I_y^2},$$

where I_x is the derivative of I in x-direction and I_y is the derivative in y-direction. The final feature vector of an image consists of the concatenation of the four histograms of the original BA-LFD approach and the four histograms of the BA-LFD's from the gradient magnitude image I_G.

It should be noted that the segmentation approach generates a higher number of blobs if it is applied to the gradient magnitude images than if it is applied to the original image (about 1.5 times as much) and thus the values of the histograms of the gradient magnitude image are about 1.5 times as high than those of the original image. Since the histograms are not normalized, the histograms of the gradient magnitude image have a slightly higher impact on the classification of the images than those of the original image.

We will furtherly denote this approach as blob-adapted gradient LFD (BA-GLFD) approach.

3.5.2. The Blob Shape adapted LFD Approach

Our second approach additionally analyzes the shape and contrast of the blobs. Already in Häfner et al. (2014a), we proposed an approach that used the shape and contrast of the blobs as features for the classification of endoscopic images. The segmentation algorithm to generate the blobs in Häfner et al. (2014a) is similar to the segmentation algorithm used in the BA-LFD approach (see Section 3.3.3).

In Häfner et al. (2014a), the following shape features are computed from a blob b:

- A convex hull feature (CH):

$$CH(R) = \frac{\#\text{ Pixels of Convex Hull}(b)}{\#\text{ Pixels of } b}.$$

- A skeletonization feature (SK):

$$SK(R) = \frac{\#\text{ Pixels of Skeletonization}(b)}{\sqrt{\#\text{ Pixels of } b}},$$

- A perimeter feature (PE):

$$PE(R) = \frac{\#\text{ Pixels of Perimeter}(b)}{\sqrt{\#\text{ Pixels of } b}}.$$

(a) Convex hull (b) Skeletonization (c) Perimeter

Figure 13: Examples of the blob features

In Figure 13 we see examples of the three shape features.

For each of the three shape features, histograms are computed separately for peak and pit blobs, resulting in 6 shape histograms per image.

Additionally, a contrast feature (CF) for each pixel of a blob is computed in Häfner et al. (2014a). For each pixel x contained in a blob b, a normalized gray value is computed as

$$CF(x) = \frac{I(x) - \text{mean}_{b(x)}(I)}{\sqrt{\text{var}_{b(x)}(I)}}, \tag{8}$$

where $b(x)$ is the blob containing x, $\text{mean}_{b(x)}$ and $\text{var}_{b(x)}$ are the mean and the variance of the gray values inside the considered blob $b(x)$, respectively.

The CF is computed separately for pixels contained in peak and pit blobs, respectively. This results in two contrast feature histograms, computed by scanning all pixels contained in peak or pit blobs.

The feature vector of an image in Häfner et al. (2014a) consists of the histograms of the shape and contrast features.

In our new approach, the feature vector of an image consists of the concatenation of the BA-LFD features and the shape and contrast features (using the segmentation algorithm of the BA-LFD approach). Combining the BA-LFD features with the shape and contrast features makes sense, since they extract very different informations which compliment each other. The BA-LFD approach extracts the information about the changes in the intensity distribution for growing regions centered at the considered point of interest (the center of a blob), and the BS approach extracts the information about the shape of the blob and the contrast inside of the blob. The feature vector of the BA-LFD approach consists of 60 feature elements (4 histograms with 15 bins per histogram) per image and the shape and contrast histograms consist of 140 feature elements (6 shape histograms with 15 bins per histogram and 2 contrast histograms with 25 bins per histogram). The shape histograms and the BA-LFD histograms have the same range of values ((the same blobs are used to extract BA-LFD and shape features), however the contrast histograms have distinctly higher values. For example, a blob generates one perimeter feature and one BA-LFD feature (one for binary filters and one for Gaussian filters), but each pixel of the blob generates one contrast feature. So the sum over a contrast histogram divided by the sum

over a shape or BA-LFD feature histogram results in the average number of pixels per blob (peak or pit blob) in an image. For example the average number of pixels of a blob over all images of the NBI database is about 49.

As already mentioned before, the distance between 2 feature vectors is measured using the χ^2 distance. When we compare the χ^2 distance between two arbitrary values with the χ^2 distance of these values multiplied by a factor f, then the distance between the 2 values is f times smaller than the distance between the multiplied values. Since we use the χ^2 distance metric and the histograms are not normalized, the contrast features would have an inflated impact to the classification of the images. To balance the inequality of the range of the feature values, we weight the BA-LFD features distinctly stronger than the contrast (and shape) features. We set the weighting to combine the BA-LFD features and the shape and contrast features to (10,1). The weighting is applied by multiplying the values of the BA-LFD histograms with 10. Experimental results showed that the weighting factor $f = 10$ is suitable for the CC-i-Scan databases as well as for the NBI database.

We furtherly denote this approach as blob shape adapted LFD (BSA-LFD) approach.

The BSA-LFD approach combines the shape and contrast information of peaks or pits with the information about the changes of the intensity distribution from the center of a pit or peak to the area surrounding the pit or peak. Since the same segmentation algorithm is used for the BA-LFD features as well as for the shape and contrast features, the BSA-LFD approach requires hardly any additional computation time compared to the BA-LFD approach.

To assess the influence of the combined shape and contrast features compared to the BA-LFD features to the results of the BSA-LFD, we additionally compute the shape and contrast features alone like in Häfner et al. (2014a) (but with our slight modification of the segmentation algorithm).

We denote this approach, using the six histograms of the shape features and the two contrast histograms, as Blob Shape (BS) approach.

3.5.3. The Blob Shape adapted Gradient LFD Approach

This approach combines the BA-GLFD approach with the BSA-LFD. That means we compute BA-LFD, shape and contrast histograms of the original image as well as of the gradient image.

The final feature vector of an image consists of the concatenation of the BA-LFD features (of the original and gradient image) and the shape and contrast features (also of the original image and the gradient image). Once again, the BA-LFD features are higher weighted by means of a multiplication of the BA-LFD features with a factor of 10. Experimental results showed that the weighting factor 10 is suitable for the CC-i-Scan databases as well as for the NBI database. We will furtherly denote this approach as blob shape adapted gradient LFD approach (BSA-GLFD).

3.6. Other methods

In this sections we describe a variety of state of the art methods for colonic polyp classification used in corresponding literature that are not based on the LFD. We furtherly want to compare the results of these approaches with the results of the LFD based approaches.

3.6.1. Dense SIFT Features

This approach (Tamaki et al., 2013) combines densely computed SIFT features with the bag-of-visual-words (BoW) approach. The SIFT descriptors are sampled at points on a regular grid. By means of the SIFT descriptors, cluster centers (visual words) are learned 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. We use the same parameters that led to the best results in Tamaki et al. (2013) (grid spacing = 5, SIFT scale 5 and 7, 6000 visual words). In Tamaki et al. (2013), this approach is used for the colonic polyp classification in NBI endoscopy, however, there is no reason why this approach should not also be suited for other imaging modalities like the i-Scan technology or chromoendoscopy. Drawbacks of this method are the huge dimensionality of its feature vectors (6000 feature elements per feature vector of an image) and the huge computational effort to learn the cluster centers.

3.6.2. Vascularization Features

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

3.6.3. Dual-Tree Complex Wavelet Transform (DT-CWT)

The DT-CWT (Häfner et al., 2009) is a multi-scale and multi-orientation wavelet transform. The final feature vector of an image consists of the statistical features mean and standard deviation of the absolute values of the subband coefficients (6 decomposition levels $\times$ 6 orientations $\times$ 3 color channels $\times$ 2 features per subband = 216 features per image). The DT-CWT showed to be well suited for the classification of polyps for different imaging modalities like high-magnification chromoendoscopy (Häfner et al., 2009) or HD-chromoendoscopy combined with the i-Scan technology (Häfner et al., 2014b).

3.6.4. LBP

Based on a grayscale image, this operator generates a binary sequence for each pixel by thresholding the neighbors of the pixel by the center pixel value. The binary sequences are then treated as numbers (i.e. the LBP numbers). Once all LBP numbers for an image are computed, a histogram based on these numbers is generated and used as feature vector. There are several variations of the LBP operator and they are used for a variety of image processing tasks including endoscopic polyp detection and classification (e.g. Häfner et al. (2012)). Two examples of such LBP variants are local ternary patterns Tan and Triggs (2010) and fuzzy local binary patterns Eystratios et al. (2012). Because of its superior results compared to the standard LBP operator $LBP_{(8,1)}$ (with block size = 3), we use a multiscale block binary patterns (MB-LBP) operator (Liao et al., 2007) with three different block sizes (3,9,15). The uniform LBP histograms of the 3 scales (block sizes) are concatenated resulting in a feature vector with $3 \times 59 = 177$ features per image.

4. Experimental Results

We use the software provided by the Center for Automation Research [1] for the MFS-LFD approach. The implementations of the BA-LFD approach and the BS approach are the ones we already used in (Häfner et al., 2014c). The algorithm of the MR8-LFD approach is custom implemented following the description in publication Xu et al. (2009) (using Matlab). We use the implementation of the phase symmetry filter (Kovesi, 2000) for the vascularization feature approach, the remaining code for this approach is custom implemented following the description in Gross et al. (2012) (using Matlab). The SIFT descriptors and the following k-means clustering is done using the Matlab software provided by the VLFeat open source library (Vedaldi and Fulkerson, 2008). The DT-CWT is implemented using the same software as in (Häfner et al., 2009). The remaining algorithms are specifically developed for this work using Matlab.

For a better comparability of the results and to put more emphasis to the feature extraction, all methods are evaluated using a k-NN classifier.

4.1. The CC-i-Scan database

The CC-i-Scan database is an endoscopic image database consisting of 8 sub-databases with 8 different imaging modalities. Our 8 image sub-databases are acquired by extracting patches of size 256 x 256 from frames of HD-endoscopic (Pentax HiLINE HD+ 90i Colonoscope) videos either using the i-Scan technology or without any CVC (¬CVC in Table 3). The mucosa is either stained or not stained. The patches are extracted only from regions having histological findings. The CC-i-Scan database is provided the St. Elisabeth Hospital in Vienna and was already used e.g. in Häfner et al. (2014b,c).

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

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

In Table 3 we see the overall classification rates (OCR) for our experiment using the CC-i-Scan database. To balance the problem of varying results depending on k, we average the 10 results of the k-NN classifier using $k = 1, \ldots, 10$. The column $\varnothing$ shows for each method the averaged accuracies across all image enhancement modalities. The highest results for each image enhancement modality across all methods are given in bold face numbers.

As we can see in Table 3, all methods perform distinctly better without staining the mucosa. But this does not necessarily mean that the classification is easier without staining. It could also be based on the fact that the proportion of the number of healthy images to the number of abnormal images is more unbalanced (in favor to the number of abnormal images) in case of the 4 image databases without staining than in case of the 4 image databases with stained mucosa (see section 5).

The i-Scan modes distinctly enhance the OCR results, especially the two modes i-Scan 1 and i-Scan 2.

When we compare the results of the original BA-LFD approach with the results of its three extensions, then we see that the three extensions perform slightly better. The two extensions using additional shape information (BSA-LFD and BSA-GLFD), the approach using only shape information (BS) and the approach using the MR8 filter bank (MR8-LFD) perform best in our experiments. However, we see that there is no method that provides constantly high results over all databases. Altogether, the differences of the averaged accuracies are quite small between the methods (except of the vascularization features, whose averaged accuracy is lower than those of the other methods), but the differences between the accuracies of the methods of single databases are partly much higher. In case of the databases with stained mucosa, the vascularization features provide very poor results because the pits of the mucosal structure, which are filled with dye, are wrongly recognized as vessels. The MFS-LFD, the DLFD, the SIFT and especially the vascularization feature approach are the methods with the lowest accuracies.

By means of the McNemar test (McNemar, 1947), we assess the statistical significance of our results. With the McNemar test we analyze if the images from a database are classified differently by the various LFD based methods, or if most of the images are classified identical by the various LFD based methods (whereat we only differentiate between classifying an image as right or wrong).

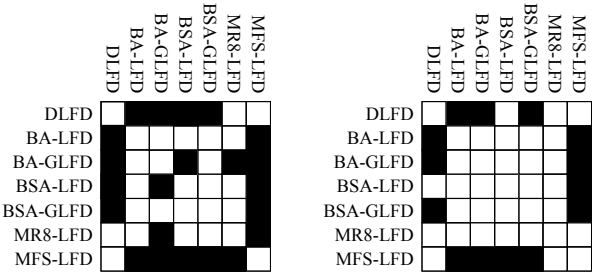
[1] http://www.cfar.umd.edu/~fer/website-texture/texture.htm

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

Table 2: Number of images and patients per class with and without CC (staining) and computed virtual chromoendoscopy (CVC)

Methods	No staining				Staining				∅
	¬CVC	i-Scan1	i-Scan2	i-Scan3	¬CVC	i-Scan1	i-Scan2	i-Scan3	
DLFD	75	78	78	80	72	67	77	61	74
BA-LFD	74	87	81	79	70	76	**85**	64	77
BA-GLFD	77	**90**	78	85	70	73	81	65	78
BSA-LFD	76	86	84	85	68	**81**	83	69	**79**
BSA-GLFD	**80**	89	82	**86**	68	75	82	68	**79**
MR8-LFD	77	84	80	81	**73**	78	82	**74**	**79**
MFS-LFD	69	75	80	72	68	77	79	62	73
BS	79	85	**87**	87	66	77	80	71	**79**
SIFT	74	82	78	72	65	76	76	65	74
Vasc. F.	64	73	76	72	58	48	63	60	64
DT-CWT	78	84	85	85	70	72	73	68	77
MB-LBP	71	83	80	76	66	74	73	73	75

Table 3: Accuracies of the CC-i-Scan databases.



(a) $\alpha = 0.05$ (b) $\alpha = 0.01$

Figure 14: Results of the McNemar test. A black square in the plot means that the two considered LFD based method are significantly different with significance level α. A white square means that there is no significant difference between the methods.

The McNemar test tests if the classification results of two methods are significantly different for a given level of significance (α) by building test statistics from incorrectly classified images. Tests were carried out for two different levels of significance ($\alpha = 0.05$ and $\alpha = 0.01$) using the i-Scan1 sub-database without staining the mucosa. Results are displayed in Figure 14. Roughly summarized, the results of the two methods DLFD and MFS-LFD are significantly worse than the results of most of the other LFD based methods.

In Table 4 we show the results of the different stages between the LFD and BA-LFD approach for the CC-i-Scan databases. That means we show the results of the DLFD approach and the BA-LFD approach and the 7 intermediate steps between the two approaches like specified in Section 3.4. In this way, we are able to analyze the effects on the results of each of the 4 adaption steps that distinguish the LFD from the BA-LFD approach. Once again, the column ∅ shows the averaged accuracies over all databases. The highest result of each image enhancement modality is given in bold face numbers.

We can see in Table 4 that the scale adaption is the most effective adaptation step in case of the CC-i-Scan databases. When we compare step 8 and 9, then the scale adaption improves the averaged results for about 3%. Using Gaussian filters (+1%) (Nr. 7 → Nr. 9) and filtering only on interest points (+2%) (Nr. 7 → Nr. 9) also slightly increase the results. The shape and orientation adaptions neither increases nor decreases the results. However, these effects don't appear for each combination of adaption steps. For example the combinations of the scale adaption and filtering only on interest points (Nr. 5) does not improve the results compared to the DLFD approach (Nr. 1). Furthermore, the improvements of the results after each adaptation step are rather low.

4.2. The NBI database

The NBI database is an endoscopic image database consisting of 908 patches extracted from frames of zoom-endoscopic (CF-H260AZ/I, Olympus Optical Co) videos using the NBI technology. The patches are rectangular and have sizes between about 100*100 and 800*900 pixels. The database consists of 359 images of type A, 462 images of type B and 87 images of type C3. Image labels were provided by at least two medical doctors and

Nr.	Adaption Scale	Shape	Int.P.	Gauss.	No staining ¬CVC	i-Scan1	i-Scan2	i-Scan3	Staining ¬CVC	i-Scan1	i-Scan2	i-Scan3	∅
1	x	x	x	x	75	78	78	80	72	67	77	61	74
2	x	x	x	✓	**76**	83	81	76	**73**	74	75	59	75
3	x	x	✓	x	72	85	79	79	**73**	73	83	63	76
4	x	x	✓	✓	74	85	80	78	73	73	**85**	60	76
5	✓	x	✓	x	71	80	**82**	76	63	78	81	64	74
6	✓	x	✓	✓	**76**	86	**82**	**82**	69	75	81	**66**	**77**
7	✓	✓	✓	x	74	85	80	76	68	**80**	83	63	76
8	x	✓	✓	✓	72	84	80	79	65	68	82	64	74
9	✓	✓	✓	✓	74	**87**	81	79	70	76	**85**	64	**77**

Table 4: Accuracies for the of the DLFD approach (Nr. 1) and the BA-LFD approach (Nr. 9) and the 7 intermediate steps between the two approaches using the CC-i-Scan databases. The columns 2 – 5 show which adaptions are used for each of the 9 methods.

endoscopists who are experienced in colorectal cancer diagnosis and familiar with pit pattern analysis and NBI classifications. The NBI database is provided by the Hiroshima University and the Hiroshima University Hospital and was already used in Tamaki et al. (2013).

In Tamaki et al. (2013), 10-fold cross validation was used to classify the NBI database. We decided to use a similar test setup with a higher reliability. Classification accuracy is computed using a training set and an evaluation set. 90% of the images of each class are randomly chosen for the training set, the remaining 10% of the images per class build up the evaluation set. The classification results are defined as the averaged result of 100 runs with randomly chosen training and evaluation sets. So the main difference between our test setup and 10-fold cross validation is that we use the averaged results of 100 runs instead of the averaged results of 10 runs.

To balance the problem of varying results depending on k, we average the 10 results of the k-NN classifier using $k = 1, \ldots, 10$. The results given in Table 5 are the averaged results from 100 runs with k-values $k = 1, \ldots, 10$. The standard deviations of the results are given in brackets.

Only in case of the SIFT features, we use a 10-fold cross validation because of the huge computational effort to learn the cluster centers in each validation run.

As we can see in Table 5, the BA-GLFD, the BSA-GLFD, the MR8-LFD and the vascularization features approach provide the highest results. The BS, the MFS-LFD and the MB-LBP approach are the least adequate approaches to classify NBI images. Combining LFD based features with shape and contrast features (BSA-LFD) enhances the results, but not as much as additionally applying the BA-LFD approach to the gradient magnitudes of the images (BA-GLFD). The combination of both extensions (BSA-GLFD) is the best performing approach for the NBI database.

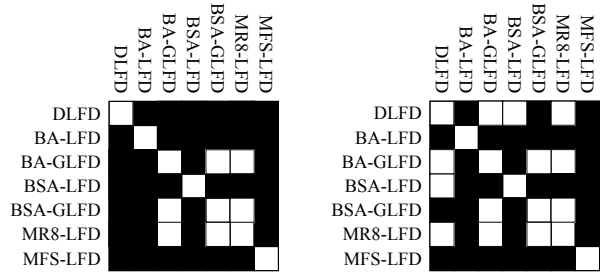
In Tamaki et al. (2013), the Dense SIFT approach achieved results of 96% for the same NBI database, whereas we achieved only 83.5% with the same feature extraction approach. Both results are achieved using 10-fold cross validation. The huge difference in the results is caused by different classification strategies. We simply average the k-NN classifier results for $k = 1, \ldots, 10$ and

Methods	Accuracy
DLFD	86.9 (2.8)
BA-LFD	83.7 (3.8)
BA-GLFD	88.0 (3.2)
BSA-LFD	85.8 (3.6)
BSA-GLFD	**88.2** (3.2)
MR8-LFD	87.5 (2.8)
MFS-LFD	80.0 (3.5)
BS	77.0 (4.8)
SIFT	83.5 (2.8)
Vasc. F.	88.1 (3.0)
DT-CWT	82.8 (3.0)
MB-LBP	81.2 (3.9)

Table 5: Accuracies and standard deviations of the NBI database in %.

use those parameters for the dense SIFT approach that achieved the best results in Tamaki et al. (2013), whereas in Tamaki et al. (2013) a variety of different support vector machine kernels and a variety of different parameters for the dense SIFT approach were tested and the classification rate of 96% was the highest classification rate of all these combinations.

Since the classification of the NBI database is done using 100 runs with different training and evaluation sets, the McNemar test is not adequate to assess the statistical significance of the results. Instead of the McNemar test, we use the Wilcoxon rank-sum test (Fay and Proschan, 2010). As input parameter for the Wilcoxon rank-sum test, we use the averaged accuracies of the 10 k's of the kNN classifier of two methods (and of course α). The input parameter of one method is of length 100 (one accuracy for each of the 100 runs). Tests were carried out for two different levels of significance ($\alpha = 0.05$ and $\alpha = 0.001$). Results for the LFD based methods are displayed in Figure 15. Only those LFD based methods with quite similar accuracies (BA-GLFD, BSA-GLFD and MR8-LFD) in Table 5 are not assessed as significant different.



(a) $\alpha = 0.05$ (b) $\alpha = 0.001$

Figure 15: Results of the Wilcoxon rank-sum test for the NBI database. A black square in the plot means that the results of the two considered method are significantly different with significance level α. A white square means that there is no significant difference between the results of the methods.

5. Discussion

5.1. Balancing the number of images per class in the CC-i-Scan databases

As already mentioned in Section 4.1 and as can be seen in Table 2, the proportion of the number of healthy images to the number of abnormal images is in favor to the number of abnormal images in case of the CC-i-Scan databases, especially for those databases without staining the mucosa. This affects the classification results, since it causes the kNN-classifier to classify more healthy and abnormal images as abnormal (because of the higher number of training images of class abnormal), as it would classify with an equal number of healthy and abnormal images. This effect is additionally increased by the relative small number of images of the CC-i-Scan databases.

To avoid this unwanted effect, we recomputed the classification accuracies using an adaption of the LOPO cross validation. In case of the "normal" LOPO cross validation, for a given image, all images from other patients than the patient of the considered image are permitted as possible nearest neighbors of the considered image. This of course leads to a higher number of abnormal images as possible nearest neighbor than healthy images (because there are more abnormal images than healthy images in case of the CC-i-Scan databases).

Our adaption of the LOPO cross validation works as follows: For a given image of patient A, we count the number of images per class that are not from patient A. Then one class will have a lower number of images that are not from patient A (class healthy) than the other. This number of images is the number of permitted images per class as nearest neighbor for the considered image of patient A. Lets say we have n permitted images per class as nearest neighbor, then the images that are permitted as nearest neighbors for the kNN classifier (the training images for the considered image) are the n images of class healthy and n randomly chosen images from the images of the class abnormal that are not from patient A.

Our adaption leads to fairer classification results than in case of the normal LOPO cross validation. However, it has the drawback of a lower number of available training images. This will probably decrease the results of the adapted LOPO cross validation compared to the normal LOPO cross validation. However, the results of the adapted LOPO cross validation should be more meaningful than those of the normal LOPO cross validation.

In Table 6 we can see the results using the adapted LOPO cross validation. The gray numbers in brackets are the accuracies using normal LOPO cross validation. Like expected, the results are lower using the adapted LOPO cross validation compared to the normal one (except the SIFT approach). However, the degradations are only quite small for most of the methods except of those methods that didn't even worked so well using the normal LOPO cross validation (DLFD, MFS-LFD and the vascularization features). In case of the i-Scan 3 mode, the results are even increasing for most of the methods by using the adapted LOPO cross validation. The best performing methods are the BA-LFD extensions (especially BSA-LFD), the MR8-LFD approach and the SIFT approach.

From the results in Table 6 we can conclude that most of the methods are in fact performing better without staining the mucosa and by using the i-Scan technology. The most possible reason why staining the mucosa has a negative impact to the results is that the colorant flows into the pits and thus the pits of the mucosa are filled with colorant whereas the peaks of the mucosa are relatively unstained. This has the effect that the pit pattern structure is easier to recognize for the physicians. But it also changes the intensity distribution between pits and peaks. Since most of the employed methods analyze this intensity distribution it is quite possible that these changes in the intensity distribution make it harder for the methods to differentiate between healthy and abnormal mucosa.

5.2. Assessing the viewpoint invariance of the methods

As already mentioned in the introduction, viewpoint invariance is an important feature for methods classifying endoscopic image databases. In colonoscopic (and other types of endoscopic) imagery, mucosa texture is usually found at different viewpoint conditions. This is due to varying distance and perspective towards the colon wall during an endoscopy session. The differences in scale are for example much higher using HD-endoscopes (especially because of the highly variable distance) than for using high-magnification endoscopes, where the distance of the endoscope to the mucosa is relatively constant. Consequently, in order to design reliable computer-aided mucosa texture classification schemes, the viewpoint invariance of the employed feature sets could be essential, especially for the CC-i-Scan database. In Figure 16 we see examples of endoscopic images of two different polyps under different viewpoint conditions. The images showed in Figure 16 are frames of two of the HD-endoscopic videos (of two patients), which were used to extract patches for the CC-i-Scan database.

Methods	No staining				Staining				∅
	¬CVC	i-Scan1	i-Scan2	i-Scan3	¬CVC	i-Scan1	i-Scan2	i-Scan3	
DLFD	72(75)	69(78)	56(78)	71(80)	68(72)	66(67)	70(77)	55(61)	66(74)
BA-LFD	76(74)	**83**(87)	81(81)	82(79)	69(70)	78(76)	79(85)	64(64)	76(77)
BA-GLFD	**81**(77)	**83**(90)	81(78)	85(85)	68(70)	72(73)	80(81)	66(65)	77(78)
BSA-LFD	80(76)	82(86)	**85**(84)	**88**(85)	64(68)	**82**(81)	77(83)	71(69)	**79**(79)
BSA-GLFD	80(80)	81(89)	83(82)	87(86)	66(68)	76(75)	80(82)	70(68)	78(79)
MR8-LFD	73(77)	80(84)	82(80)	82(81)	**75**(73)	80(78)	**82**(82)	76(74)	**79**(79)
MFS-LFD	69(69)	66(75)	72(80)	61(72)	67(68)	76(77)	75(79)	58(62)	68(73)
BS	79(79)	77(85)	84(87)	**88**(87)	58(66)	73(77)	77(80)	75(71)	76(79)
SIFT	74(74)	**83**(82)	79(78)	85(72)	74(65)	81(76)	80(76)	**77**(65)	**79**(74)
Vasc. F.	66(64)	64(73)	63(76)	60(72)	55(58)	43(48)	55(63)	51(60)	57(64)
DT-CWT	76(78)	76(84)	81(85)	82(85)	74(70)	73(72)	75(73)	73(68)	76(77)
MB-LBP	71(71)	77(83)	78(80)	78(76)	68(66)	78(74)	74(73)	**77**(73)	75(75)

Table 6: Accuracies of the CC-i-Scan databases using adapted LOPO cross validation. The gray numbers in brackets are the accuracies using normal LOPO cross validation.

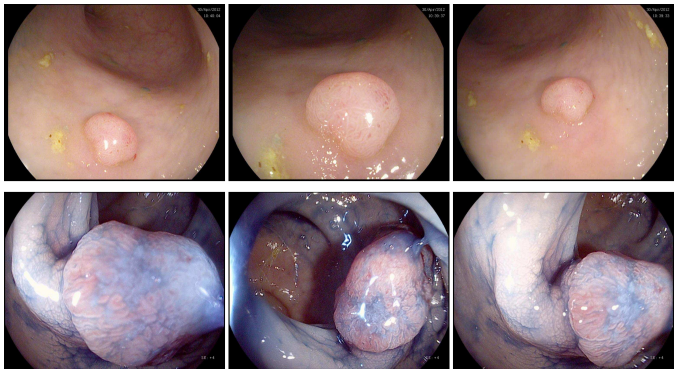
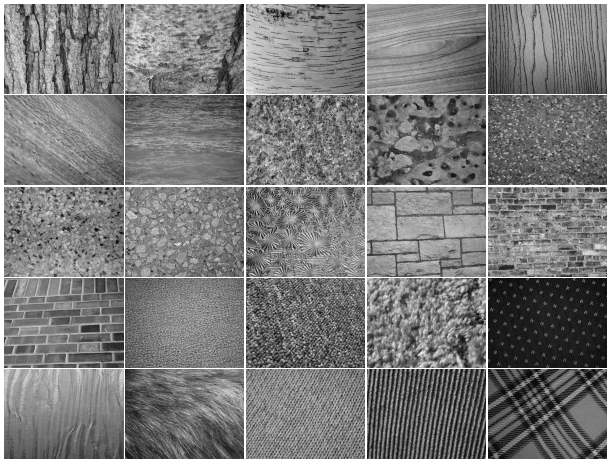


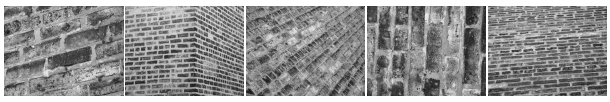
Figure 16: Two polyps shown under varying viewpoint conditions.

In this section we assess the viewpoint invariance of the employed methods by means of a public texture database, the UIUCtex database. Contrary to the endoscopic images, where images of same classes often look very different and often have quite different texture structures, the images of the classes of the UIUCtex database are half-way homogeneous (apart from the viewpoint and illumination conditions). The higher homogeneity, the huge differences of the viewpoint conditions and the high number of image classes (25) are the reasons why we choose the UIUCtex database to estimate the viewpoint invariance instead of an endoscopic image database. We estimate the viewpoint invariance of the methods by comparing the classification accuracies and by image retrieval.

The UIUCtex database (S. Lazebnik and Ponce, 2005) is a public texture database consisting of 25 different texture classes with 40 images per texture class. The resolution of the images is 640*480. Significant viewpoint changes are present within each class, and illumination conditions are uncontrolled. Additional sources of variability can be the non-planarity of textured surfaces, significant non-rigid deformations, inhomogeneities of the texture patterns and viewpoint dependent appearance variations. In Figure 17 we see an example image of each of the 25 texture classes and an example of the differences of the viewpoint conditions.



(a) Examples of the 25 texture classes of the UIUCtex database.



(b) Examples of the different viewing conditions of the UIUCtex database (by means of the texture class brick).

Figure 17: The UIUCtex database

5.2.1. Classifying the UIUCtex database

Classification accuracy is computed using a training set and an evaluation set. A fixed number of images per class (1–20) is randomly chosen to build up the training set, the remaining images build up the evaluation set. Like in Xu et al. (2009) (MFS-LFD) and Varma and Garg (2007) (MR8-LFD), a k-NN classifier is used with $k = 1$.

The results given in Figure 18 are the averaged results of 100 runs with randomly chosen training and evaluation sets. Only in case of the SIFT features, we use the result of only one run with randomly chosen training and evaluation set because of the huge computational effort to learn the cluster centers for each validation run (the computation for one run takes more than a week using a Quad-Core PC).

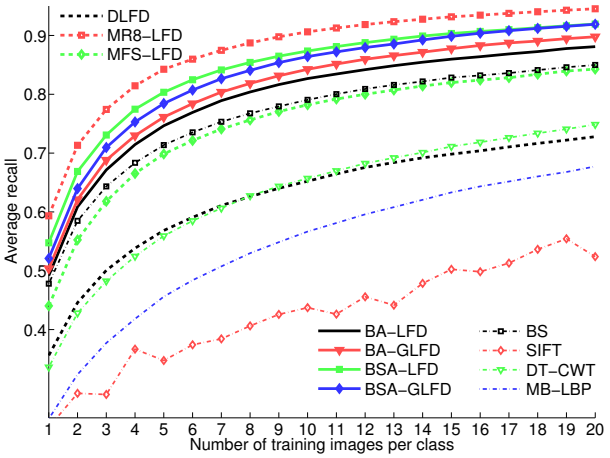


Figure 18: Classification results of the UIUCtex database (best viewed in color).

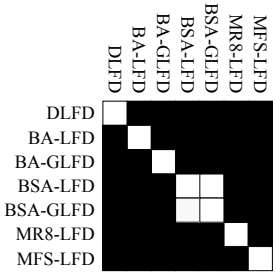


Figure 19: Result of the Wilcoxon rank-sum test for the UIUCtex database. A black square in the plot means that the results of the two considered method are significantly different with significance level $\alpha = 0.01$. A white square means that there is no significant difference between the results of the methods.

As we can see in Figure 18, the extension of the BA-LFD using additionally blob-shape features (BSA-LFD) as well as the extension additionally applying the BA-LFD approach to the gradient image (BA-GLFD) both improve the results for the UIUCtex texture database, compared to the original BA-LFD approach. The combination of both extensions (BSA-GLFD) provides slightly worse results than those of the BSA-LFD approach, which is the best performing BA-LFD based approach. The shape and contrast features of the gradient image decrease the results of the BSA-LFD approach. Without these features, the BSA-GLFD approach would outperform the BSA-LFD approach. The MR8-LFD approach provides the best results, DLFD and all not LFD based approaches provides worse results than the BA-LFD based approaches.

The BA-LFD approach in Häfner et al. (2014c) was especially developed for classifying polyps using the CC-i-Scan databases and not for general texture recognition. It finds the pits and peaks of the pit pattern structure and then filters the area in and surrounding the detected pits with filters that are shape, size and orientation adapted to the pits and peaks. Maybe the BA-LFD approach and its extensions need to be adapted for classifying texture databases, however the results for the UIUCtex database are quite respectable. By adapting the BA-LFD based approaches to general texture recognition, most likely the results of these approaches would even be higher.

We did not test the vascularization features on the UIUCtex database, since this approach is not suited for texture classification and so it would be pointless to compare its results with the other methods.

The results presented in the original publication of the MR8-LFD approach are slightly higher than the results of our reimplementation of the original MR8-LFD approach (the accuracies are about one percent higher in the original publication). This is probably caused by minor implementation differences and by the fact that we use all 8 filter responses instead of a feature subset selection using only 5 of the 8 filter responses like proposed in the original publication.

Like for the NBI database, the statistical significance of the tests for the LFD based methods is assessed using the Wilcoxon rank-sum test. Contrary to the NBI database we use the results for $k = 1$ of the kNN classifier per run instead of the averaged results over $k = 1, \ldots, 10$ per run. Results are displayed in Figure 19 for significance level $\alpha = 0.01$ and 10 training images per class.

As we can see in Table 19, the results of the LFD based methods are all significantly different except of the BSA-LFD and the BSA-GLFD approach.

5.2.2. Assessing the viewpoint invariance

As already mentioned before, significant viewpoint changes are present within each class of the UIUCtex database. It is very hard to develop a texture descriptor that is able to identify two images from one class as images from the same class, if the images are acquired under quite different viewpoint conditions. Of course it is much easier if the images are acquired under similar viewpoint conditions. For a given image of the evaluation set, the nearest neighbor classifier only needs to find the image of the training set that has the closest distance to the considered image and then the evaluation set image is classified to the class the training image belongs to. If there are several images per class in the training set, then there will probably be a training set image of the same class than the considered evaluation set image with similar viewpoint conditions. That means, the higher the number of training images per class, the lower the required viewpoint invariance of a method. That means especially for a higher number of training images per class, the feature expressiveness probably dominates the issue of viewpoint invariance. So if there is a high difference between the classification results using 1 and 20 training images per class for classifying the UIUCtex database, then this is an indicator that the considered method is not viewpoint invariant.

Additionally to the classification results, the viewpoint invariance of the methods is assessed by image retrieval. Image retrieval is done as in Xu et al. (2009). Given a

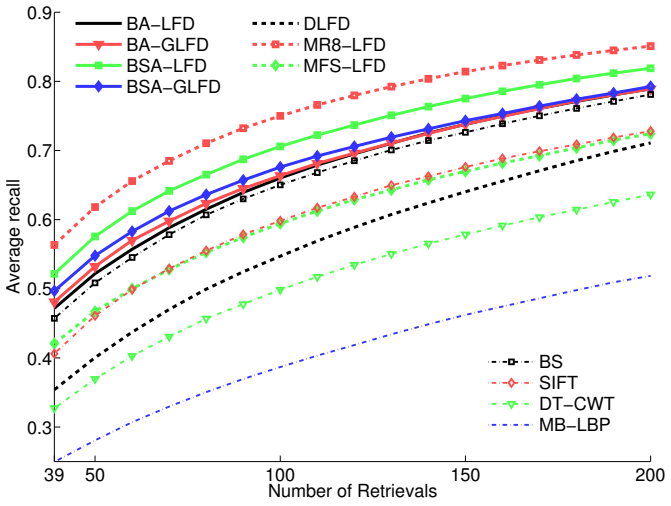


Figure 20: Image retrieval results of the UIUCtex database (best viewed in color).

query image, the other images of the database are sorted in an increasing order of distance to the query image, i.e. from the most similar to the least similar. Each image of the UIUCtex database is used once as a query image, and the performance is summarized as a plot of average recall vs. the number of retrieved images. Average recall is defined as the number of images retrieved from the same class as the query image divided by the number of images in the class minus one (40 -1 (the query image) =39) averaged over all queries. For example, perfect performance for a given class would correspond to an average recall of 100 % after 39 retrieved images. Scale and viewpoint invariance is essential for good retrieval results in case of the UIUCtex database, since the distances from the query image to the other images from the same class as the query image should be smaller than the distances to images of other classes, no matter how big the viewpoint differences are between the query image and the remaining images from the same class.

In Figure 20 we see the retrieval results of the UIUCtex database and in Figure 18 we see the classification results for all numbers of training images per class between 1 and 20.

As we can see in Figure 20 and Figure 18, the results of the average recall are visually similar to the classification results. The methods performance compared to each other is nearly similar for image retrieval and classification (except of the SIFT approach).

The clearly lowest recall rates and the clearly lowest classification rates (especially for low numbers of training images per class) of the MB-LBP approach imply that the MB-LBP approach is less viewpoint invariant than the other approaches.

When we compare the results of the DLFD and the DT-CWT approach, we can observe two facts which together imply that the DLFD approach is more viewpoint invariant than the DT-CWT approach. First, we see that the accuracies of the DLFD approach are higher than those of the DT-CWT for lower numbers of training images per class and lower for higher numbers of training images per class. Second, the recall rates of the DLFD approach are higher than those of the DT-CWT approach.

The recall and classification curves of the BA-LFD-based approaches and the BS approach are similar which indicates that these approaches are similarly viewpoint invariant. Based on the recall and classification curves, the only approach that is more viewpoint invariant than those approaches is the MR8-LFD approach.

The results of the two plots (Figure 20 and Figure 18) imply, that the BA-LFD based methods are not generally more viewpoint invariant than the approaches based on the original LFD. So the adaption of the shape, size and orientation of the filters of the BA-LFD does not increase the viewpoint invariance of the BA-LFD based approaches compared to the approaches based on the original LFD. However, since the BA-LFD based methods are amongst the best methods for each of the tested databases, the adaptions of the BA-LFD increase the feature expressiveness. When we compare the results of the BA-LFD and the DLFD approach, we see that the 4 adaptions of the BA-LFD approach (viewpoint adaption, computation only on interest points and Gaussian filters additional to the binary filters) distinctly improve the results for all databases except of the NBI database.

Results imply that at least most of the LFD based approaches are more viewpoint invariant than the other approaches.

Generally, since the ranking of the methods with respect to their accuracy for lower numbers of training images (viewpoint invariance should be an advantage) and higher number of training images (viewpoint invariance is not essential) as well as the ranking of the methods with respect to their recall rate (viewpoint invariance should definitely be an advantage) is nearly identical (except of the SIFT feature), it seems that even for the UIUCtex database, a database with huge viewpoint variations, the feature expressiveness is more important than the viewpoint invariance.

The accuracy curve of the SIFT feature is the lowest of all methods whereas the recall curve is amidst the other curves. This is caused by the fact that in case of the classification, the dictionary is build using only the images of the training set, whereas in case of image retrieval, the dictionary is build using all images of the image database (including the query image). The accuracy curve of the SIFT feature is not smooth since we used the result of only one run instead of the average result of 100 runs in case of the other methods (more noise).

6. Conclusion

In this work we showed that methods based on computing the LFD and BA-LFD are well suited for colonic

BA-GLFD	...	Blob-adapted gradient local fractal dimension
BA-LFD	...	Blob-adapted local fractal dimension
BSA-GLFD	...	Blob shape adapted gradient LFD
BSA-LFD	...	Blob shape adapted local fractal dimension
CC	...	Conventional chromoendoscopy
DLFD	...	Dense local fractal dimension
CVC	...	Computed virtual chromoendoscopy
DT-CWT	...	Dual-tree complex wavelet transform
HD	...	High definition
kNN	...	k nearest neighbor
LBP	...	Local binary patterns
LFD	...	Local fractal dimension
LOPO	...	Leave-one-patient-out
MB-LBP	...	multiscale block binary patterns
MFS-LFD	...	Multi fractal spectrum local fractal dimension
MR8-LFD	...	Maximum response 8 local fractal dimension
NBI	...	Narrow band imaging
OCR	...	Overall classification rate

Table 7: Acronyms and their meaning in alphabetical order.

polyp classification. When we compare the results of the employed methods for the 8 CC-i-Scan databases and the NBI database, we see that the proposed extensions of the BA-LFD approach are the best performing methods or at least among the best performing methods. The extension using additionally shape and contrast information (BSA-LFD) as well as the extension using additional gradient information (BA-GLFD) enhance the results, but the combination of both extensions (BSA-GLFD) is the best suited method to classify polyps on our databases. Also in case of the UIUCtex texture database, the BA-LFD extensions are amongst the best performing methods.

In case of the HD-endoscopic databases, it has been shown that most of the employed methods are performing better without staining the mucosa and by using the i-Scan technology.

Most of the LFD based approaches are more viewpoint invariant than the other approaches. The scale, shape and orientation adaptions of the BA-LFD approach and its extensions do not improve the viewpoint invariance compared to the approaches based on the original LFD. However, the 4 adaptions of the BA-LFD approach (scale, shape and orientation adaption, computation only on interest points and Gaussian filters additional to the binary filters) distinctly improve the results for all databases except of the NBI database.

Acknowledgments

This work is supported by the Austrian Science Fund, TRP Project 206.

Appendix

The used acronyms in this work and their meaning are listed in alphabetical order in Table 7.

References

Andrė, B., Vercauteren, T., Buchner, A., Wallace, M., Ayache, N., 2012. Learning semantic and visual similarity for endomicroscopy video retrieval. IEEE Transactions on Medical Imaging 31, 1276–1288.

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

Chaudhuri, B., Sarkar, N., 1995. Texture segmentation using fractal dimension. IEEE Transactions on Pattern Analysis and Machine Intelligence 17, 72–77.

Eystratios, K., Dimitris, I., Dimitris, M., 2012. Fuzzy binary patterns for uncertainty-aware texture representation. Electronic Letters on Computer Vision and Image Analysis 10.

Fay, M.P., Proschan, M.A., 2010. Wilcoxon-mann-whitney or t-test? on assumptions for hypothesis tests and multiple interpretations of decision rules. Statistics Surveys 4, 1–39.

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

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

Gross, S., Palm, S., Tischendorf, J.J.W., Behrens, A., Trautwein, C., Aach, T., 2012. Automated classification of colon polyps in endoscopic image data. SPIE Proceedings 8315, 83150W–83150W–8.

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

Häfner, M., Kwitt, R., Uhl, A., 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.

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

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

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

Iakovidis, D.K., Koulaouzidis, A., 2015. Software for enhanced video capsule endoscopy: challenges for essential progress. Nature Reviews Gastroenterology & Hepatology .

Iakovidis, D.K., Maroulis, D.E., Karkanis, S.A., 2006. An intelligent system for automatic detection of gastrointestinal adenomas in video endoscopy. Computers in Biology and Medicine 36, 1084–1103.

Iakovidis, D.K., Maroulis, D.E., Karkanis, S.A., Brokos, A., 2005. A comparative study of texture features for the discrimination of gastric polyps in endoscopic video, in: Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems, 2005 (CBMS'05), Dublin, Ireland. pp. 575–580.

Jabbour, J.M., Saldua, M.A., Bixler, J.N., Maitland, K.C., 2012. Confocal endomicroscopy: Instrumentation and medical applications. Annals of Biomedical Engineering 40, 378–397.

Kanao, H., Tanaka, S., Oka, S., Kaneko, I., Yoshida, S., Arihiro, K., Yoshihara, M., Chayama, K., 2008. Clinical significance of type vi pit pattern subclassification in determining the depth of invasion of colorectal neoplasms. World Journal of Gastroenterology 14, 211–217.

Karkanis, S.A., Iakovidis, D.K., Maroulis, D.E., Karras, D.A., Tzivras, M., 2003. Computer-aided tumor detection in endoscopic

video using color wavelet features. IEEE Transactions on Information Technology in Biomedicine 7, 141–152.

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

Kiesslich, R., 2009. Advanced imaging in endoscopy. European Gastroenterology & Hepatology Review 5, 22–25.

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

Kovesi, P.D., 1999. Image features from phase congruency. Videre: Journal of Computer Vision Research 1, 2–26.

Kovesi, P.D., 2000. MATLAB and Octave functions for computer vision and image processing. Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia. Available from: <http://www.csse.uwa.edu.au/∼pk/research/matlabfns/>.

Kudo, S.E., Hirota, S., Nakajima, T., Hosobe, S., Kusaka, H., Kobayashi, T., Himori, M., Yagyuu, A., 1994. Colorectal tumours and pit pattern. Journal of Clinical Pathology 47, 880–885.

Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S., 2007. Learning multi-scale block local binary patterns for face recognition , 828–837.

Lopes, R., Betrouni, N., 2009. Fractal and multifractal analysis: A review. Medical Image Analysis 13, 634–649.

Maroulis, D.E., Iakovidis, D.K., Karkanis, S.A., Karras, D.A., 2003. Cold: A versatile detection system for colorectal lesions in endoscopy video-frames. Computer Methods and Programs in Biomedicine 70, 151–66.

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

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

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

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

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

Tan, X., Triggs, B., 2010. Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Transactions on Image Processing 19, 1635–1650.

Uhl, A., Vécsei, A., Wimmer, G., 2011. 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., Zissermann, A., 2005. A statistical approach to texture classification from single images. International Journal of Computer Vision (IJCV) 62, 61–81.

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

Vincent, L., Soille, P., 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence 13, 583–598.

Xia, Y., Feng, D., Zhao, R., 2006. Morphology-based multifractal estimation for texture segmentation. IEEE Transactions on Image Processing 15, 614–623.

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

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