

# Delaunay Triangulation-based Pit Density Estimation for the Classification of Polyps in High-magnification Chromo-colonoscopy

M. Häfner<sup>a</sup>, M. Liedlgruber<sup>b,\*</sup>, A. Uhl<sup>b</sup>, A. Vécsei<sup>c</sup>, F. Wrba<sup>d</sup>

<sup>a</sup>*St. Elisabeth Hospital  
Department for Internal Medicine  
Vienna, Austria*

<sup>b</sup>*University of Salzburg  
Department of Computer Sciences  
5020 Salzburg, Austria*

<sup>c</sup>*St. Anna Children's Hospital  
Endoscopy Unit  
Vienna, Austria*

<sup>d</sup>*Medical University of Vienna  
Department of Clinical Pathology  
Vienna, Austria*

---

## Abstract

In this work we propose a method to extract shape-based features from endoscopic images for an automated classification of colonic polyps.

This method is based on the density of pits as used in the pit pattern classification scheme which is commonly used for the classification of colonic polyps. For the detection of pits we employ a noise-robust variant of the LBP operator. To be able to be robust against local texture variations we extend this operator by an adaptive thresholding. Based on the detected pit candidates we compute a Delaunay triangulation and use the edge lengths of the resulting triangles to construct histograms. These are then used in conjunction with the k-NN classifier to classify images.

We show that, compared to a previously developed method, we are not only able to almost always get higher classification results in our application scenario, but that the proposed method is also able to significantly outperform the previously developed method in terms of the computational demand.

*Keywords:* Colonoscopy, polyps, pit pattern, local binary patterns, delaunay triangulation, classification

---

\*Corresponding author

*Email addresses:* [mliedl@cosy.sbg.ac.at](mailto:mliedl@cosy.sbg.ac.at) (M. Liedlgruber), [uhl@cosy.sbg.ac.at](mailto:uhl@cosy.sbg.ac.at) (A. Uhl)

## 1. Introduction

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 from inside the colon, allowing to obtain images for a computer-assisted analysis with the goal of detecting abnormalities. To be able to acquire highly detailed images a magnifying endoscope can be used [1]. Such an endoscope represents a significant advance in colonoscopy as it provides images which are up to 150-fold magnified, thus uncovering the fine surface structure of the mucosa as well as small lesions. Example images of colonic polyps, acquired with such an endoscope, are given in Fig. 1.

There already exists some previous work devoted to an automated cancer staging employing different colonoscopic imaging modalities or videos. For classic white-light endoscopy several studies have demonstrated that computer-based image analysis is capable of detecting colorectal polyps [2, 3] in endoscopic video frames to a certain extent and to perform a first assessment of the malignant potential of these polyps [4, 5, 6, 7]. Narrow-band-imaging (NBI) has been shown to facilitate discrimination between neoplastic and non-neoplastic polyps relying on features of the observed microvasculature to some extent [8, 9, 10]. Confocal laser endomicroscopic images have also been used to differentiate lesions into the categories neoplastic and benign [11, 12] using a dense variant of the bag-of-visual-words features.

While diagnostic accuracies of automated staging techniques employing the imaging modalities described so far range between 70% and 94%, classification accuracies ranging from 95% up to 99% depending on employed features and classification schemes (see e.g. [13, 14, 15, 16]) have been achieved using high-

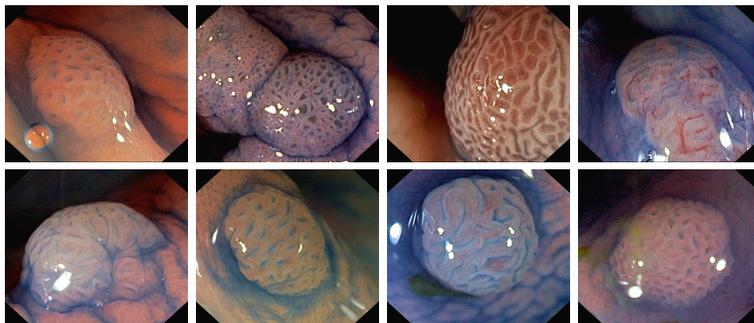


Figure 1: Example images showing colonic polyps (acquired using a high-magnification colonoscope).

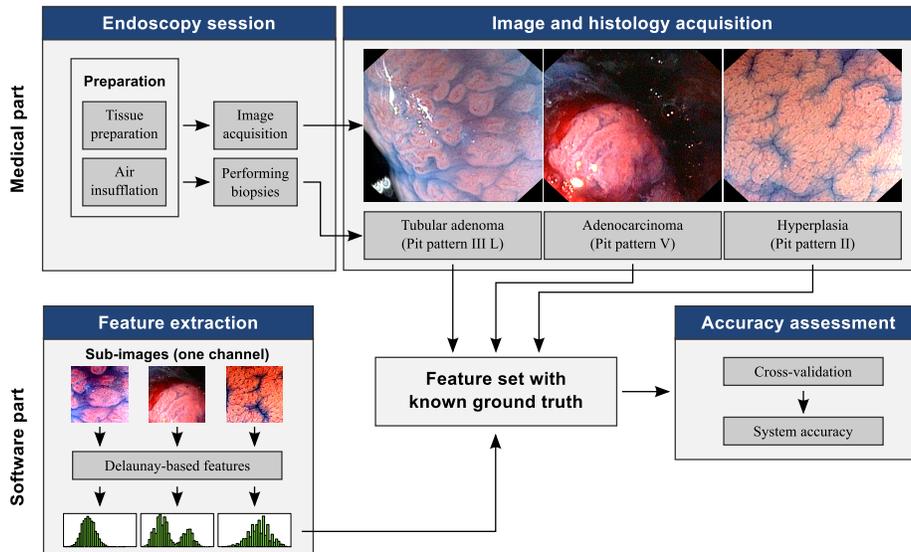


Figure 2: Overview of our system for endoscopic image classification.

magnification chromo-colonoscopy (where contrast enhancement is achieved by actual staining during colonoscopy).

In this work we specifically aim at the classification of the latter type of imagery based on the pit pattern scheme. Since in chromo-colonoscopy the mucosal crypt patterns get visually enhanced, a classification of polyps using high-level features instead of statistical texture features is a natural choice. In previous work we already showed that features based on the distribution of pits delivered promising results [17].

The method presented in this work is similar to the work in [17] in terms of the basic idea of measuring the pit densities in zoom-endoscopic images. But in contrast to the method proposed in [17] we were able to achieve a more robust pit detection by modifying the underlying Local Binary Patterns (LBP) operator variant. In addition, the proposed method has a lowered computational demand due to a greatly simplified methodology for the pit candidate detection. In contrast to [17] we also carry out a clinically more relevant 3-class classification. Besides that, this work also investigates the classification accuracies when using the more restrictive leave-one-patient-out cross-validation (LOPO-CV), while in [17] we used leave-one-image-out cross-validation (LOO-CV) only.

An overview of our system for endoscopic image classification is shown in Fig. 2. First we acquire endoscopic images and collect the respective histologic classifications. Based on the original images we then manually select sub-images for further processing. We then detect pit candidates using a noise-insensitive variant of the LBP operator and measure the density of pits within an image. This is done by computing histograms of edge lengths of a Delaunay triangulation based on the pit candidate positions. These histograms are then classified

using the k-NN classifier in order to obtain the overall accuracy of the system (based on two different cross-validation protocols).

In Section 2 we review the classification of pit patterns of the colonic mucosa. Section 3 summarizes related work on methods for polyp classification based on high-level features. The proposed method is described in more detail in Section 4, followed by the classification of the resulting features in Section 5. Experimental results and configuration details of the classification system proposed are given in Section 6. Section 7 concludes the paper.

## 2. Pit Pattern Classification

Polyps of the colon are a frequent finding and are usually divided into metaplastic, adenomatous, and malignant. As resection of all polyps is time-consuming, it is imperative that those polyps which warrant endoscopic resection can be distinguished: polypectomy of metaplastic lesions is unnecessary and removal of invasive cancer may be hazardous. For these reasons, assessing the malignant potential of lesions at the time of colonoscopy is important as this would allow to perform targeted biopsy. While such systems are still not standard-of-care, the aim of developing such automated polyp classification systems is to avoid random and, probably, unnecessary biopsies. Hence, such systems could potentially help to save time, lower the cost for a colonoscopy procedure, and reduce the risk of complications during the procedure.

The most commonly used classification system to distinguish between non-neoplastic and neoplastic lesions in the colon is the pit pattern classification, originally introduced by Kudo et al. [18]. This system allows to differentiate between normal mucosa, hyperplastic lesions (non-neoplastic), adenomas (a pre-malignant condition), and malignant cancer based on the visual pattern of the mucosal surface. Thus this classification scheme is a convenient tool to decide which lesions need not, which should, and which most likely can not be removed endoscopically. The mucosal pattern as seen after dye staining and by using magnification endoscopy shows a high agreement with the histopathologic diagnosis. Due to the visual nature of this classification it is also a convenient choice for an automated image classification.

In this classification scheme exist five main types according to the mucosal surface of the colon, as illustrated in Fig. 3. Type III is divided into types III-S and III-L, designating the size of the pit structure. It has been suggested that type I and II pattern are characteristic of non-neoplastic lesions (benign and non-tumorous), type III and IV are found on adenomatous polyps, and type V are strongly suggestive of invasive carcinoma, thus highly indicative for cancer [19].

Furthermore, lesions of type I and II can be grouped into non-neoplastic lesions and types III to V can be grouped into neoplastic lesions. This allows a grouping of lesions into two classes, which is more relevant in clinical practice as indicated in a study by Kato et al. [20]. In addition, Kato et al. proposed a 3-class classification which groups the six different pit pattern types into normal lesions (pit pattern types I and II), non-invasive lesions (pit pattern types

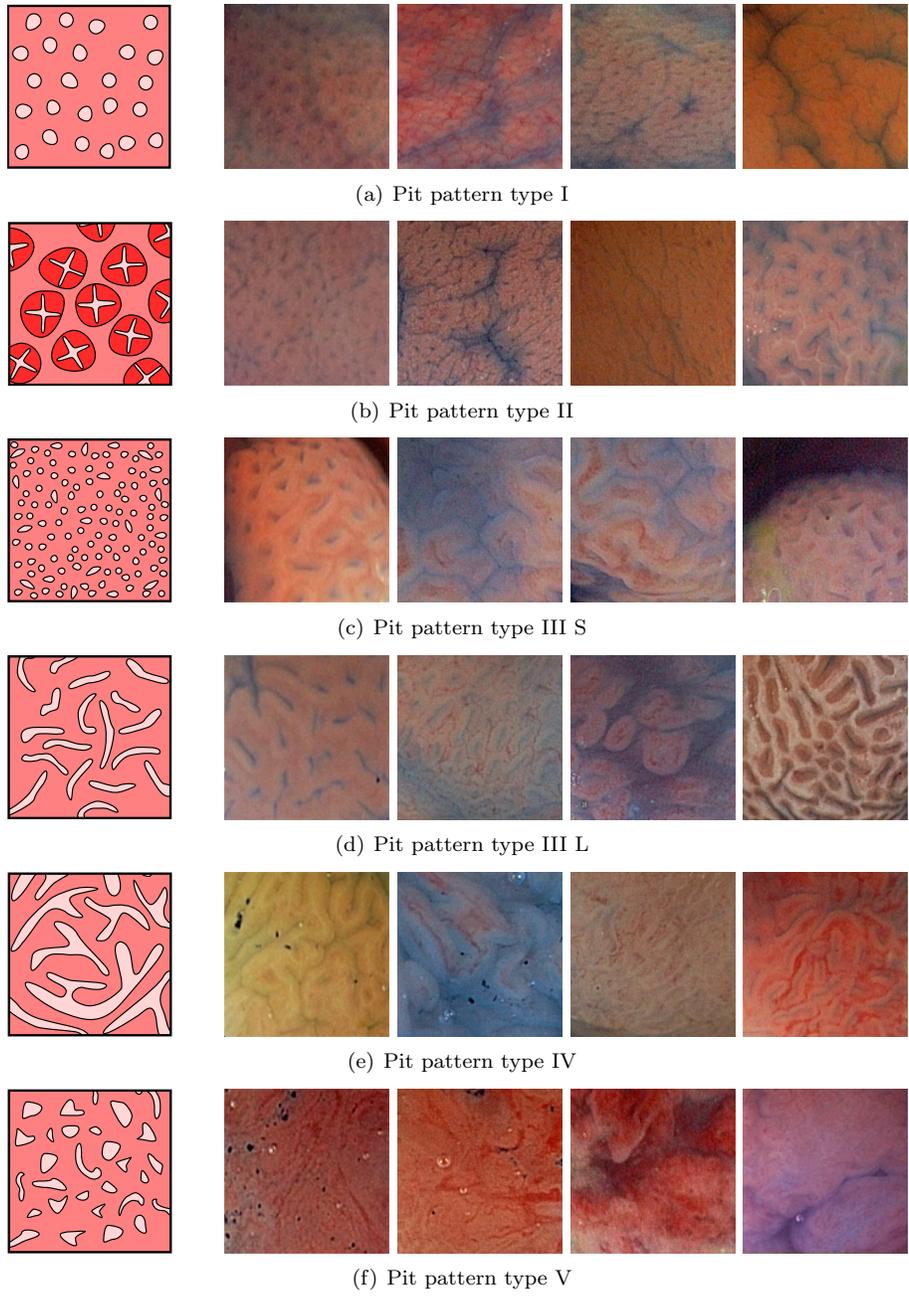


Figure 3: Schematic illustration of the pit pattern classification along with example images for each pit pattern type.

III-S, III-L, and IV), and invasive lesions (pit pattern type V). This classification scheme is of particular importance since normal mucosa needs not to be removed, non-invasive lesions must be removed endoscopically, and invasive lesions must not be removed endoscopically. Due to the clinical relevance of these two classifications (2-class and 3-class), this work will focus on these groupings only.

Using a magnifying colonoscope together with indigo carmine dye spraying, the mucosal crypt pattern on the surface of colonic lesions can be observed [19]. Several studies found a good correlation between the mucosal pit pattern and the histological findings, where especially techniques using magnifying colonoscopes led to excellent results [20].

From Fig. 3 we notice that pit pattern types I to IV can be characterized fairly well, whereas type V contains no clear pit pattern structure anymore. At a first glance this classification scheme seems to be straightforward and easy to be applied. But it needs some exercising to achieve fairly good results [21]. Apart from that, similar to the reported inter-observer variability of NBI-based colonoscopy ( $\kappa \approx 0.57$  [22],  $\kappa \approx 0.63$  [23],  $\kappa \approx 0.69$  [24]) inter-observer variability of magnification chromo-endoscopy in the interpretation of pit patterns of colonic lesions has been described ( $\kappa \approx 0.56$  [25],  $\kappa \approx 0.64$  [23]). This work aims at allowing computer-assisted pit pattern classification in order to enhance the quality of differential diagnosis.

The topical staining used in chromo-endoscopy has the effect of visually enhancing mucosal crypt patterns or vascular features. Due to this visual enhancement of mucosal structures developing a method to detect and measure the distribution of pits is a natural choice. In addition, as we notice from Fig. 3, pit pattern types I and II are regular to some extent and the pits are distributed more tightly. Types III to V, in contrast, are more irregular in terms of the pit distribution, showing a lower pit density or even a complete absence of pits. These observations are the basis for the method proposed in this work.

### 3. Related Work

Using high-level features for the detection of polyps is quite common throughout literature (e.g. [26, 27, 28, 29]). But the assessment of the malignant potential of polyps is commonly performed on the basis of texture features. Usually such approaches employ features obtained from wavelet transformed images [2, 3, 30, 31], features obtained from Fourier transformed images [16], or other texture features [32, 15, 33].

Up to our knowledge currently only a few approaches exist which make use of some sort of high-level features for the classification of polyps. In [8] the polyp regions are first segmented in a manual fashion from NBI images. Then, features describing the visible blood vessels are extracted and used for a subsequent classification (number of blood vessel pixels, average perimeter of vessels, and intensity based features). Similarly, Stehle et al. [9] also base their work on NBI images and analyze the vesselation visible within an image (in contrast to [8] the images have been acquired using a zoom-endoscope). In order

Reference	Enhancement	Feature basis
Gross et al. [8]	NBI	Blood vessel properties
Stehle et al. [9]	NBI / Zoom	Blood vessel properties
Tischendorf et al. [10]	NBI / Zoom	Blood vessel properties
Takemura et al. [34]	Color dyes / Zoom	Shape of pits
Häfner et al. [17]	Color dyes / Zoom	Density of pits

Table 1: A comparison of different approaches which use high-level features for polyp classification.

to detect vessel structures the authors compare a directional stamping algorithm in combination with a vesselness filter against measuring the phase symmetry in combination with the fast marching algorithm. For the classification the authors use vessel features expressing the length, mean perimeter, and mean intensity of vessels. The phase symmetry approach is then used in a subsequent work for the classification of polyps [10].

A different way of polyp classification is proposed in [34]. Instead of using NBI-enhanced images, the authors base their work on images obtained during a high-magnification chromo-colonoscopy. Based on the captured images, the authors extract six shape descriptors which reflect properties of pit structures found within the endoscopic images (e.g. area, perimeter, and circularity). However, a major drawback of this approach is that the method, used to detect the pits, obviously produces wrong pit regions or pit regions which are over-segmented. To overcome this problem, the authors use an image manipulation software to manually correct the pit detection. This however is problematic since the real robustness of the proposed method can not be deduced from the results presented.

Another work, which is also based on the pits found on the colonic mucosa, has recently been proposed in [17]. In this work a rotation-invariant LBP operator is used to generate a binary image, showing pit candidates. This image is then subject to morphological post-processing, followed by applying the Canny edge detector, and a subsequent post-processing of the detected pit regions by again employing different morphological operations. Based on the center points of the final pit candidates a Delaunay triangulation is computed. The lengths of the edges connecting the pit centers are used to generate histograms, based on which the classification is carried out. While being similar to the proposed method, the method proposed in [17] suffers from a rather high computational demand due to the high number of processing steps involved. Apart from that, this method achieved relatively low classification accuracies.

#### 4. The Proposed Method

In the following we describe the proposed method in more detail. An overview of the method is shown in Fig. 4.

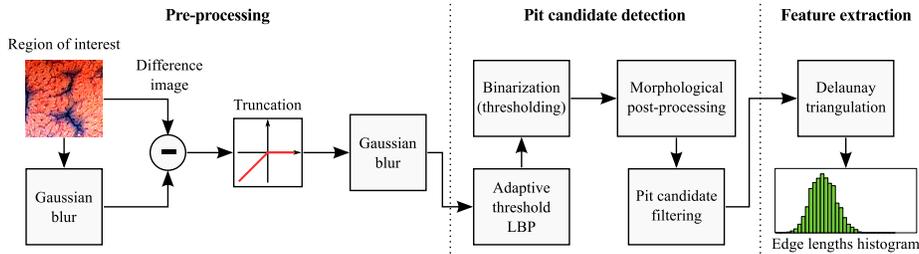


Figure 4: Overview of the proposed feature extraction method.

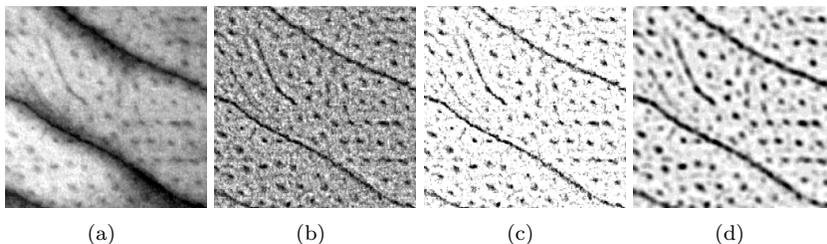


Figure 5: Illustration of the different pre-processing steps. (a) the red color channel of a sample pit pattern I image, the difference image (b) before and (c) after truncation of positive difference values, and (d) the final image after applying a gaussian blur (the contrast of the images has been manually enhanced in order to better show the effect of the different steps).

#### 4.1. Pre-processing

Endoscopic images usually suffer from inhomogeneous illumination. To be able to isolate regions which potentially contain pits, we first remove global illumination changes by computing the difference image  $I_D$  between the original color channel  $C_O$  and a Gauss-blurred version of  $C_O$  (Gaussian kernel with a width of  $K_D = 33$  with  $\sigma_D = 4$ ).

The pits we aim to detect are usually relatively dark as compared to surrounding image regions. Hence,  $I_D$  contains negative values in regions where pits are located. Thus we truncate the difference  $I_D$  by

$$I_T(i, j) = \min(0, I_D(i, j)). \quad (1)$$

As a consequence the result  $I_T$  contains negative values in potential pit regions and zeros elsewhere.

Finally, to reduce noise, which may eventually have a negative impact on the pit detection, we apply a Gaussian blur (Gaussian kernel with a width of  $K_N = 7$  with  $\sigma_N = 5$ ) to the truncated difference image  $I_T$ .

The effect of the different pre-processing steps is illustrated in Fig. 5.

#### 4.2. Pit Candidate Detection

Similar to the work in [17], we use an LBP variant for the detection of pit candidates. Theoretically it would be also possible to use a simple thresholding

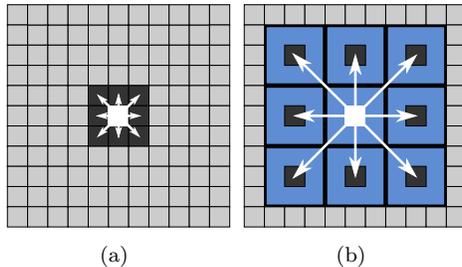


Figure 6: Illustration of (a) the pixel comparison used in the original LBP operator and (b) the block-based comparison used in [17].

for the detection. But as we notice from Fig. 5(d), our endoscopic images may contain ridges which exhibit similar brightness levels in  $I_T$  as compared to pits. Hence, a thresholding would also result in wrongly detected pits due to ridges. In addition, the pit regions within  $I_T$  are different in terms of the brightness. As a consequence, at least a global thresholding would fail.

As a consequence we decided to base the pit detection on an extended version of the LBP variant already used in [17] which allows us to detect pit regions by identifying darker regions surrounded by relatively bright regions. In case of this operator the comparison between pixel values in LBP is replaced by a comparison of pixel block intensity averages, which are computed for each pixel (hence, the blocks overlap).

The difference between the pixel comparisons used in the original LBP operator and the operator used in [17] is depicted in Fig. 6. From Fig. 6(a) we see that in case of the original LBP operator the value of the center pixel (denoted by a white square) is compared to the 8 nearest neighbors. In case of the operator used in [17] the average intensity of the center block, which has a width of  $K_{LBP} = 9$  pixels (size of  $9 \times 9$  pixels) and is centered at the pixel position denoted by the white box, is compared against the intensity averages of the neighboring blocks (the center positions of these blocks are denoted by the dark gray boxes in Fig. 6(b)). The blue boxes surrounding the dark gray boxes denote the subwindows of pixels used to compute the respective intensity averages (in this example the blocks have a width of  $K_{LBP} = 3$ ).

Applying this block-based operator to  $I_T$  results in an image containing the highest possible LBP number (i.e. 255) at the pixel positions belonging to pit regions, since all surrounding block averages contain higher values as compared to the center block.

To make this operator resistant against small fluctuations still present in  $I_T$  (especially in regions which do not contain pits), we extend the comparison of block averages by an adaptive thresholding.

In [17] the LBP number for a pixel located at  $(i, j)$  is computed as

$$LBP_{i,j} = \sum_{n=1}^N 2^{n-1} T_{n,i,j}, \quad (2)$$

with  $N = 8$  since we are using a 8-neighborhood and

$$T_{n,i,j} = \begin{cases} 1, & \text{if } A(x_n, y_n) \geq A(i, j) \\ 0, & \text{otherwise} \end{cases}, \quad (3)$$

where  $x_n$  and  $y_n$  denote the center position of the  $n$ -th neighboring block (these are ordered in a clock-wise fashion, starting with the top left neighbor) and  $A(x, y)$  denotes the average value within a  $K_{LBP} \times K_{LBP}$  block located at  $(x, y)$ .

For this work we slightly modify Equ. (3) by introducing a local threshold  $t_{i,j}$ , which results in

$$T_{n,i,j} = \begin{cases} 1, & \text{if } A(x_n, y_n) \geq A(i, j) + t_{i,j} \\ 0, & \text{otherwise} \end{cases}. \quad (4)$$

To take care of local intensity variations within a pixel block,  $t_{i,j}$  is computed for each pixel position  $(i, j)$  as the mean standard deviation within all pixel blocks belonging to the neighborhood (the blue boxes in Fig. 6(b)). If the image exhibits a higher standard deviation at some position (i.e. a higher roughness within the neighborhood) the threshold gets higher. This way the operator tends to response more to rather high intensity differences between the center block and its neighboring blocks, while being robust against small intensity variations. The threshold  $t_{i,j}$  is therefore computed as

$$t_{i,j} = \frac{1}{N+1} \left( \sum_{n=1}^N \sigma_{x_n, y_n} + \sigma_{i,j} \right), \quad (5)$$

where  $\sigma_{x_n, y_n}$  and  $\sigma_{i,j}$  denote the standard deviation within the  $n$ -th neighboring block located at  $(x_n, y_n)$  and the standard deviation within the center block, respectively. By applying the LBP operator to  $I_T$  we obtain the LBP image  $I_L$ .

Figure 7 shows the differences between the pit regions in  $I_T$ , the pit regions in  $I_T$  after applying the LBP variant as used in [17], and the same regions after transforming  $I_T$  using the LBP operator based on block averaging and adaptive thresholding. From this figure we notice that, in contrast to the different pit heights in  $I_T$ , the pits show up more clearly in  $I_L$ . In addition, we notice that the LBP variant used in [17] produces pit regions which are not as distinct as the pit regions produced by the LBP variant based on adaptive thresholding.

The flat regions at the highest points in Fig. 7(d) correspond to an LBP number equal to 255. Therefore, to obtain the binary image  $I_B$  which contains the pit regions only, we simply apply a thresholding to  $I_L$ :

$$I_B(i, j) = \begin{cases} 1, & \text{if } I_L(i, j) = 255 \\ 0, & \text{otherwise} \end{cases}. \quad (6)$$

Since in some cases  $I_B$  contains small spurs or isolated pixels (a white pixel surrounded by black pixels in the respective  $3 \times 3$  pixel neighborhood) which do not belong to pits, we post-process the binary image by applying the following morphological operations (the actual implementations of these operations are provided by the MATLAB function *bwmorph*):

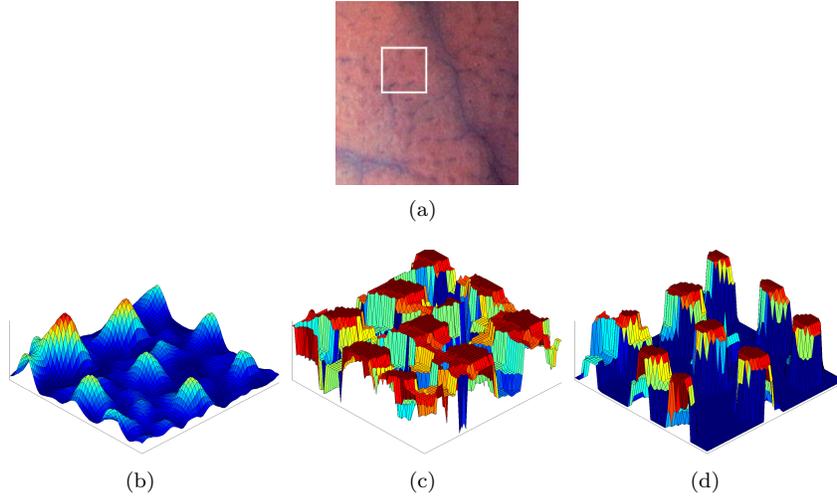


Figure 7: A sample image with a subregion (white rectangle) in (a) and a surface plot of the red color channel in this subregion (b) in the difference image  $I_T$ , (c) in  $I_T$  after applying the LBP variant used in [17], and (d) in the LBP transformed image  $I_L$  based on adaptive thresholding.

- **Spur removal**

This step removes spur pixels.

- **Majority**

This operator sets a pixel to 1 if at least half of the pixels in the  $3 \times 3$  neighborhood of the pixel are set to 1. Otherwise the pixel is set to 0. This removes small non-pit regions which may be present.

- **Cleaning**

The cleaning removes the remaining isolated pixels which are likely to be caused by noise and thus do not belong to pit regions.

After applying the morphological operations, we remove pits with an area equal to or above  $A_{max} = 100$  pixels as these are in most cases caused by ridges. Figure 8 shows a sample color channel, the respective truncated and blurred difference image  $I_T$ , and the binary image which contains the final pits detected.

#### 4.3. Feature Extraction

For the extraction of the final features we adapt the feature extraction from [17]. Hence, to measure the density of pits within a color channel we construct a mesh from the previously obtained pit centers. We then deduce the density of the pits from the edge lengths of a triangulation based on the pit centers. For this purpose we employ the Delaunay triangulation based on the Quickhull algorithm [35].

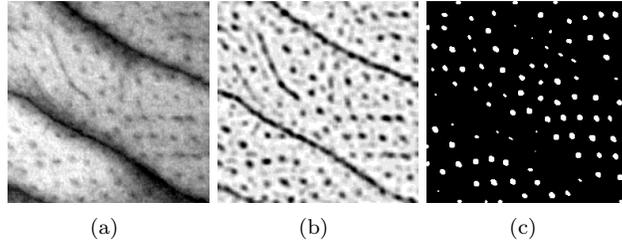


Figure 8: (a) the red color channel of a sample pit pattern I image, (b) the respective truncated and blurred difference image  $I_T$ , and (c) the final pit detection result.

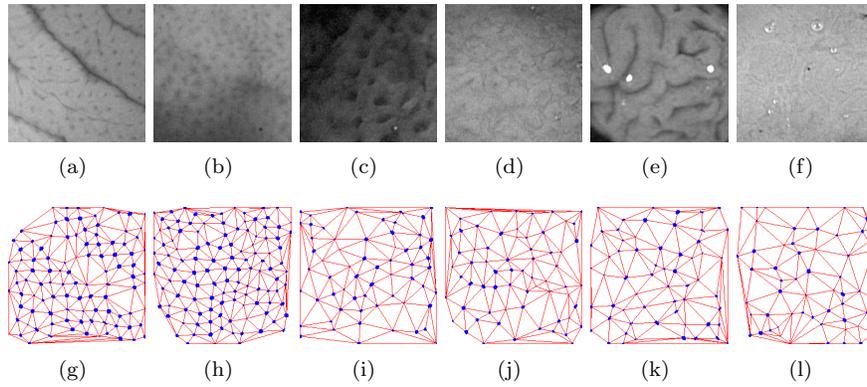


Figure 9: Results of the Delaunay triangulation along with the detected pits. (a)-(b) example images from the non-neoplastic class (red channel), (c)-(f) neoplastic images, and (g)-(l) the according Delaunay triangulations along with the detected pits (blue spots).

This algorithm transforms the 2D points to 3D (lifted to a paraboloid by computing a  $z$ -coordinate for each pit center as  $z = x^2 + y^2$ ), computes the convex hull in 3D, and projects the lower part of the hull back to 2D to obtain the triangulation [36]. This way we get a set of non-overlapping triangles with the minimum of the inner angles maximized. Since the proposed method is implemented in MATLAB, the method used for the computation of the Delaunay triangulation is the one provided by MATLAB.

Figure 9 shows the red color channel of sample images from our image database along with the respective Delaunay triangulations and the detected pits. From this figure we notice that in non-neoplastic images the number of pits found as well as the pit density is higher as compared to neoplastic images. But we also notice that in case of the non-neoplastic images ridges destroy the regularity in some parts of the images. In addition to ridges, in case of neoplastic images pits are detected although none are present (due to e.g. blood vessels).

Based on the triangulations we create 1-dimensional histograms from the edge lengths of all triangles for each color channel of an image separately. To

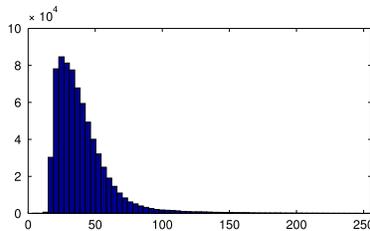


Figure 10: Distribution of the different edge lengths found in the color channels of our images.

concentrate more on triangles not belonging to the 2D convex hull of the triangulation we iterate over all triangles and use each edge of each triangle to update the histogram. This way edges shared by two triangles contribute to the histogram twice, while the remaining edges are used only once.

For a set of points located on a regular grid such a histogram has only a few peaks while the remaining bins of the histogram are empty. The more irregular the distribution of the points is the higher the spread within the resulting histogram gets. Apart from that, the density of the pit distribution is also reflected by such histograms. A dense distribution of pits results in rather short edges which results in a histogram containing the most values within the first bins. If the distribution is more sparse, the edge lengths get longer and the values within the respective histogram are shifted to the right.

Since the number of edges between images most likely will vary we normalize each histogram such that the histogram bins sum up to 1. This makes the histograms comparable during the classification process. Moreover, since all our images have a dimension of  $256 \times 256$  pixels the upper limit for an edge length is  $\sqrt{256^2 + 256^2} \approx 362$  (corresponding diagonal). But it is very unlikely that pits are only detected in the image corners. This implies that it is also unlikely that the maximum possible edge length occurs. Apart from that, the more pits we detect the more likely it is that the distances between neighboring pits get smaller.

The color channels used throughout our experiments show a maximum edge length of approximately 255, but most edge lengths lie between about 15 and 160, as can be seen from Fig. 10. Based on this observation we consider the range for the edge lengths between 1 and  $L_{max} = 256$  to be a reasonable choice and therefore use this range throughout our experiments.

The computation of the edge length histograms is sketched in Algorithm 1.

## 5. Classification

In order to predict the pathology of an image we use the k-NN classifier for the classification [37]. To measure the distance between two histograms we

---

**Algorithm 1** Computation of the edge length histogram based on the triangulation for an image

---

```

1:  $V_x \leftarrow$  matrix containing the x-coordinates of the vertices of each triangle within
   the triangulation (one row per triangle)
2:  $V_y \leftarrow$  matrix containing the y-coordinates of the vertices of each triangle within
   the triangulation (one row per triangle)
3:  $N \leftarrow$  total number of triangles in the triangulation
4:  $L \leftarrow ()$  // vector of lengths for all triangle edges (initially empty)
5:  $i \leftarrow 1$  // edge counter
6:  $M \leftarrow \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$  // indices into the vertex matrices for the triangle edges
7: for  $t = 1$  to  $N$  do
8:   for  $e = 1$  to  $3$  do
9:      $\Delta_x \leftarrow V_x(t, M(1, e)) - V_x(t, M(2, e))$ 
10:     $\Delta_y \leftarrow V_y(t, M(1, e)) - V_y(t, M(2, e))$ 
11:     $L(i) \leftarrow \sqrt{\Delta_x^2 + \Delta_y^2}$ 
12:     $i \leftarrow i + 1$ 
13:   end for
14: end for
15:  $H \leftarrow$  normalized histogram based on the values in  $L$  containing  $B$  bins and covering
   the range between 0 and  $L_{max}$ 

```

---

employ the histogram intersection distance metric, defined as

$$d(H_i, H_j) = 1 - \sum_{k=1}^B \min(H_{i,k}, H_{j,k}), \quad (7)$$

where  $H_i$  and  $H_j$  are two normalized histograms,  $B$  denotes the number of bins used in our histograms, and  $H_{i,k}$  and  $H_{j,k}$  represent the value of the  $k$ -th bin of histogram  $H_i$  and  $H_j$ , respectively. We also carried out experiments using the Euclidean distance metric and the Bhattacharyya distance metric but there was no significant difference in terms of the classification accuracies achieved. Hence, since the histogram intersection can be computed more efficiently as compared to the two other alternatives, we decided to use this distance metric for our experiments.

Besides performing a classification based on single color channels, we also aim at carrying out experiments with a combination of different color channels. This is mainly motivated by the fact that there might exist images which are misclassified when using one channel but classified correctly when using another channel. Hence, by combining color channels, hope is raised that we are able to increase the classification accuracies by overruling a wrong decision within one channel with a correct decision from one or two other color channels.

To be able to carry out a classification based on multiple color channels we had to choose a way to combine the respective histogram distances. For this purpose we compute the distances for all color channels used separately and

multiply them to obtain the final distance  $D$ . This can be formulated as

$$D(I_a, I_b) = \prod_{n=1}^C d(H_n^{(a)}, H_n^{(b)}), \quad (8)$$

where  $I_a$  and  $I_b$  denote two images,  $C$  is the number of color channels considered for combination, and  $H_n^{(a)}$  and  $H_n^{(b)}$  represent the histograms for the  $n$ -th color channel considered of images  $I_a$  and  $I_b$ , respectively. There are also other possibilities for a combination, for example by summing up the distances instead of multiplying them by replacing the product in Equ. (8) by a sum. But since the product is more tolerant against outliers – one similar color channel in terms of the histogram distance leads to a very small total distance between two images already – we favor the product instead of a sum.

## 6. Experiments

### 6.1. Experimental Setup

The image database used throughout our experiments is based on 327 endoscopic color images (either of size  $624 \times 533$  pixels or  $586 \times 502$  pixels) acquired between the years 2005 and 2009 at the Department of Gastroenterology and Hepatology (Medical University of Vienna) using a zoom-colonoscope (Olympus Evis Exera CF-Q160ZI/L) with a magnification factor of 150. In order to acquire the images 40 patients underwent colonoscopy. To obtain a larger set of images we manually extracted subimages (regions of interest) with a size of  $256 \times 256$  pixels from the original images. This resulted in an extended image set containing 716 images in total.

Lesions found during colonoscopy have been examined after application of dye-spraying with indigocarmine, as routinely performed in colonoscopy. Biopsies or mucosal resection have been performed in order to get a histopathological diagnosis. Biopsies have been taken from type I, II, and type V lesions, as those lesions need not to be removed or cannot be removed endoscopically. Type III and IV lesions have been removed endoscopically.

Table 2 shows the detailed ground truth information used for our experiments where  $\mathbf{N}_O$ ,  $\mathbf{N}_E$ ,  $\mathbf{N}_P$  denote the number of original images, the number of images in the extended image set, and the number of patients in each class, respectively. Since different types of lesions may develop inside the colon of a single patient some patients may appear in more than one class. Hence, the number of patients shown in this table is slightly higher as compared to the total number of patients who underwent colonoscopy.

Since the optimal choices for the  $k$ -value for the  $k$ -NN classifier and the number of bins  $B$  used in the edge length histograms are not known, we decided to carry out an exhaustive search for both parameters which in combination lead to the highest overall classification rates ( $k \in 1, \dots, 50$  and  $B \in 16, \dots, 256$ ). Apart from that, the combination of different color channels from the RGB color space is expected to influence the classification results too. Hence, we carried

Histology	Pit Pattern	3 classes			2 classes		
		$N_O$	$N_E$	$N_P$	$N_O$	$N_E$	$N_P$
Normal	I	72	198	14	72	198	14
Hyperplasia	II						
Tubular adenoma	III-L	212	420	27	255	518	32
Tubulovillous adenoma							
Serrated adenoma	III-S						
Tubular adenoma							
Adenoma	IV	43	98	6			
Tubular adenoma							
Tubulovillous adenoma							
Adenocarcinoma	V	43	98	6			
Carcinoma							
Lymphoma							
<b>Total</b>		<b>327</b>	<b>716</b>	<b>47</b>	<b>327</b>	<b>716</b>	<b>46</b>

Table 2: The detailed ground truth information for the image database used throughout our experiments.

out experiments with all six possible combinations (i.e. R, G, B, R+G, R+B, G+B, and R+G+B).

While we could have carried out the experiments in other color spaces as well, an investigation of our images in the CIELAB color space and the HSI color space revealed that the pit structures are noticeable within the intensity channels only in these color spaces. Since, however, the pit structures are the backbone of our method the only useful channel within these spaces would be the intensity component. Hence, we conducted experiments on grayscale images and the results achieved were rather low. We therefore consider the combination of different RGB color channels to be the most promising option.

We also aim at a comparison between the proposed method and the method proposed in [17]. Hence, we carry out experiments with both methods (using the same setup for the experiments among both methods).

The parameter choices used throughout the experiments for the proposed method are summarized in Table 3. It must be noted that the values listed in this table have been determined experimentally and are thus specifically tailored to the image database used throughout this work. When applying the method to other images (e.g. different image size) some of these parameters may be subject to readjustment.

Since our image database is quite limited in terms of the number of images available, we employ two distinct cross-validation protocols to estimate the classification accuracies: LOO-CV and LOPO-CV. In case of LOO-CV one image is removed from the image set (serving as the validation sample) while the remaining images are used to train the underlying classifier. This process is repeated for each image available. LOPO-CV is more strict, since instead of removing a single image from the training set, all images of a certain patient are removed and used as validation samples (i.e. each image of the patient is classified). Then, similar to LOO-CV, the remaining images are used to train

Description	Variable	Value used
Difference image Gaussian kernel width	$K_D$	33
Difference image Gaussian kernel sigma	$\sigma_D$	4
Noise removal Gaussian kernel width	$K_N$	7
Noise removal Gaussian kernel sigma	$\sigma_N$	5
Block width for our adaptive LBP operator	$K_{LBP}$	9
Upper limit for the area of pit regions	$A_{max}$	100
Maximum edge length for histograms	$L_{max}$	256

Table 3: Summary of the parameters used for the proposed method during the experiments.

the classifier. This process is then repeated for all patients in the database.

To make a comparison of the proposed method against the method in [17] in terms of runtime performance feasible, both methods have been implemented and compared using MATLAB R2009a. The runtime measurements have been carried out on a machine equipped with an Intel Core2Quad CPU at 2.83 GHz, running Linux.

In order to be able to assess whether two different methods produce statistically significant differences in the results obtained we employ McNemar’s test [38]. For two methods  $M_1$  and  $M_2$  this test statistic keeps track of the number of images which are misclassified by method  $M_1$  but classified correctly by method  $M_2$  (denoted by  $n_{01}$ ) and vice versa (denoted by  $n_{10}$ ). The test statistic, which is approximately Chi Square distributed (with one degree of freedom), is then computed as

$$T = \frac{(|n_{01} - n_{10}| - 0.5)^2}{n_{01} + n_{10}}. \quad (9)$$

From  $T$  the  $p$ -value can be computed as

$$p = 1 - F_{\chi_1^2}(T), \quad (10)$$

where  $F_{\chi_1^2}$  denotes the cumulative distribution function of the Chi Square distribution with one degree of freedom. The null-hypothesis  $H_0$  for McNemar’s test is that the outcomes of  $M_1$  and  $M_2$  lead to equal error rates. Given a fixed significance level  $\alpha$ , there is evidence that the methods  $M_1$  and  $M_2$  produce significantly different results if  $p < \alpha$ . As a consequence we can reject the null-hypothesis  $H_0$ . Throughout this work we chose a significance level of  $\alpha = 0.05$ . This implies that, if  $M_1$  and  $M_2$  are significantly different, there is a confidence level of 95% that the differences between the outcomes of the methods are not caused by random variation.

## 6.2. Pit Detection Results

Figure 11 shows the number of detected pits plotted for all images in the different image classes (for all different color channels). From this figure we see that, compared to the method proposed in [17], the differences between the number of detected pits is much more distinct between the non-neoplastic

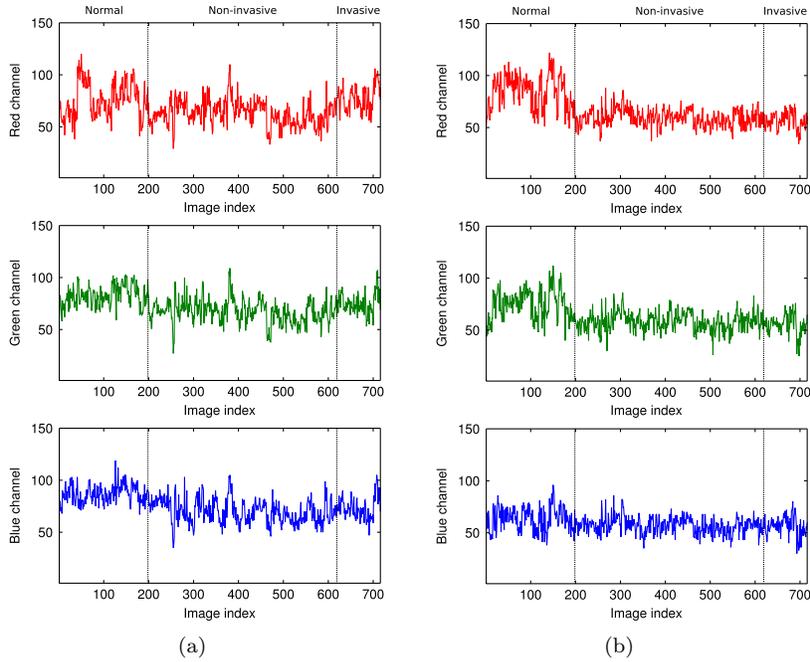


Figure 11: Number of pits found in the images across the different image classes and color channels, (a) according to [17] and (b) with the proposed method.

(normal) and neoplastic images (non-invasive and invasive). This is especially noticeable in the case of the red color channel.

Nevertheless, we also notice that even in case of images which do not contain any pits (or only a low number of pits) a rather high number of pits is detected (wrong pits in case of neoplastic images). Apart from that the number of detected pits has a rather high variance (i.e. the plots look very noisy). In addition, at least in case of the proposed method (see Fig. 11(b)), the number of pits in the invasive class is similar to the pit counts in one of the other classes, which makes the classification in the 3-classes case a non-trivial task. As a consequence the number of detected pits is not a well-suited feature to be used directly for a classification of the images.

### 6.3. Classification Results

Tables 4 and 5 show the classification results for our experiments. The results given in brackets denote the classification rates obtained with LOPO-CV, while the other results have been obtained using LOO-CV. The last column of these tables ( $\mathbf{S}_{CV}$ ) indicates whether there is a statistically significant difference between the two cross-validation protocols according to McNemar’s test. In addition, the sign given in brackets indicates whether the results obtained with LOPO-CV are significantly lower (–) or significantly higher (+) as compared to the LOO-CV results.

The accuracies, specificities, and sensitivities are computed as:

$$\begin{aligned}
 \text{Accuracy} &= \frac{\text{number of correctly classified images}}{\text{total number of images}} \\
 \text{Specificity} &= \frac{\text{number of correctly classified non-neoplastic images}}{\text{total number of non-neoplastic images}} \\
 \text{Sensitivity} &= \frac{\text{number of correctly classified neoplastic images}}{\text{total number of neoplastic images}}
 \end{aligned}$$

From tables 4 and 5 we immediately notice that, as expected, the LOPO-CV rates are always lower as compared to the LOO-CV counterparts. However, while in case of the method from [17] the LOPO-CV results are always significantly worse, this is not the case for the proposed method. Especially in case of the best performing LOPO-CV channel combinations the differences between LOO-CV and LOPO-CV vanish (they are not statistically significant). This means that the proposed method delivers more stable classification results when using the more strict LOPO-CV.

Table 4 also shows that in terms of the different channel combinations the proposed method almost always delivers higher classification accuracies as compared to the method from [17] in the 2-classes case. This holds true for LOO-CV as well as for LOPO-CV. As a consequence, also the highest result obtained with the proposed method is higher as compared to the method from [17] (93.7% versus 90.2% and 92.6% versus 84.6% in case of LOO-CV and LOPO-CV, respectively). We also notice, that the specificities of the proposed method are in general higher as compared to the method proposed in [17]. But we also notice that, even in case of the proposed method, the sensitivities are generally higher than the specificities (especially in case of the LOPO-CV results).

The main reason for this phenomenon is an unstable pit detection in case of the non-neoplastic images. As we notice from Fig. 11(b), the number of detected pits is in general higher for non-neoplastic images. But quite often there are downward spikes present with a rather low pit count in this image class. These can be explained by the fact that our image database contains many images in this class which are rather tough in terms of the pit detection (in case of quite a few images it is even problematic for a human observer to clearly identify the pits).

In order to verify that the spikes lower the specificities we investigated the outcome of the 2-class LOPO-CV experiment which is based on the red color channel only. Figure 12 shows the number of detected pits within the red color channel of the images along with indicators for misclassified images. From this figure we clearly notice that non-neoplastic images with a rather low number of detected pits indeed tend to get misclassified. We also notice that images with an upward spike in case of neoplastic images tend to get misclassified too. But since the relative frequency of downward spikes in the non-neoplastic class is higher as compared to upward spikes in the neoplastic class the specificity tends to be lower as compared to the sensitivity.

In the case of the 3-class experiments (see Table 5) the overall picture is

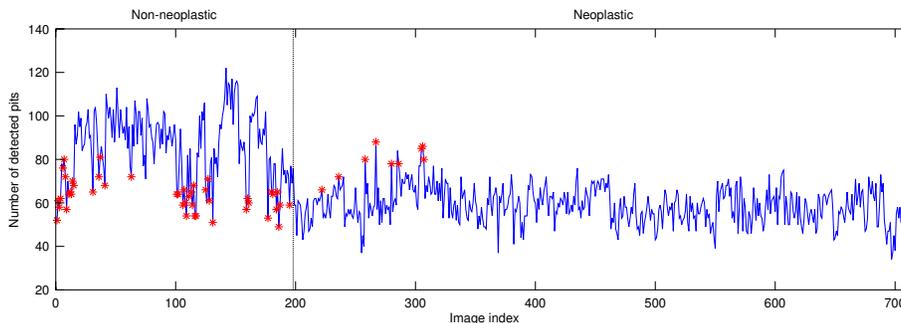


Figure 12: Number of pits found in the red color channel along with indicators for misclassified images (indicated by red stars).

Channels	Specificity	Sensitivity	Accuracy	Scv
<b>Method from [17]</b>				
R	56.1 (48.0)	98.5 (98.6)	86.7 ( <b>84.6</b> )	✓ (-)
G	47.5 (42.4)	97.5 (95.0)	83.7 (80.4)	✓ (-)
B	63.6 (62.1)	91.1 (87.1)	83.5 (80.2)	✓ (-)
R+G	62.6 (49.0)	98.5 (97.5)	88.5 (84.1)	✓ (-)
R+B	67.2 (55.1)	98.1 (95.4)	89.5 (84.2)	✓ (-)
G+B	64.6 (55.6)	94.0 (91.3)	85.9 (81.4)	✓ (-)
R+G+B	70.7 (52.5)	97.7 (96.1)	<b>90.2</b> (84.1)	✓ (-)
<b>Proposed method</b>				
R	79.3 (77.3)	98.6 (98.3)	93.3 (92.5)	
G	55.6 (50.0)	98.3 (97.5)	86.5 (84.4)	✓ (-)
B	43.4 (35.9)	97.3 (97.7)	82.4 (80.6)	✓ (-)
R+G	80.8 (74.7)	98.6 (98.3)	93.7 (91.8)	✓ (-)
R+B	81.3 (77.3)	98.5 (98.5)	<b>93.7 (92.6)</b>	
G+B	61.6 (48.0)	97.5 (98.6)	87.6 (84.6)	✓ (-)
R+G+B	78.3 (72.7)	99.4 (98.8)	93.6 (91.6)	✓ (-)

Table 4: Detailed classification results for the 2-classes case (highest results are highlighted in bold).

rather similar. For most channel combinations the proposed method outperforms the method from [17] in terms of the overall classification rate. This also holds true for LOO-CV as well as for LOPO-CV. Similar to the 2-classes case, the highest result obtained with the proposed method is higher as compared to the method from [17] (81.8% versus 80.7% and 78.8% versus 71.8% in case of LOO-CV and LOPO-CV, respectively). We also notice, that in the 3-classes case the class containing invasive lesions constantly delivers poor results as compared to the other classes. This may be attributed to the fact that, as already mentioned in Section 6.2 and noticeable from Fig. 11(b), the number of detected pits is rather similar between the classes non-invasive and invasive (which may also have a negative influence on the respective edge length histograms). In addition, the number of images in this class is rather low as compared to the

Channels	Normal	Non-invasive	Invasive	Accuracy	S <sub>CV</sub>
<b>Method from [17]</b>					
R	56.1 (51.5)	95.5 (92.6)	25.5 (13.3)	75.0 (70.4)	✓ (-)
G	65.2 (52.5)	91.2 (93.3)	11.2 (0.0)	73.0 (69.3)	✓ (-)
B	74.2 (69.2)	85.5 (84.5)	6.1 (0.0)	71.5 (68.7)	✓ (-)
R+G	65.7 (57.6)	92.6 (91.2)	38.8 (10.2)	77.8 (70.8)	✓ (-)
R+B	72.7 (68.7)	90.7 (90.0)	41.8 (2.0)	79.1 (72.1)	✓ (-)
G+B	72.7 (65.7)	91.0 (90.0)	12.2 (0.0)	75.1 (70.9)	✓ (-)
R+G+B	75.3 (68.7)	92.4 (89.5)	41.8 (2.0)	<b>80.7 (71.8)</b>	✓ (-)
<b>Proposed method</b>					
R	77.3 (78.8)	97.4 (97.1)	13.3 (0.0)	80.3 ( <b>78.8</b> )	✓ (-)
G	57.1 (51.0)	97.4 (96.4)	0.0 (0.0)	72.9 (70.7)	✓ (-)
B	46.0 (37.4)	96.7 (96.0)	3.1 (0.0)	69.8 (66.6)	✓ (-)
R+G	78.8 (76.8)	93.1 (96.7)	33.7 (0.0)	81.0 (77.9)	✓ (-)
R+B	77.8 (75.8)	97.6 (98.3)	21.4 (0.0)	81.7 (78.6)	✓ (-)
G+B	62.6 (45.5)	96.7 (98.3)	10.2 (0.0)	75.4 (70.3)	✓ (-)
R+G+B	78.3 (74.2)	94.5 (97.4)	34.7 (0.0)	<b>81.8 (77.7)</b>	✓ (-)

Table 5: Detailed classification results for the 3-classes case (highest results are highlighted in bold).

other two classes.

Concerning the different color channel combinations in case of the proposed method, we notice from tables 4 and 5 that in most cases a combination of two or three channels only slightly outperforms the best single channel result, which is always obtained by using the red color channel only. The superiority of the red color channel can be explained by the plots shown in Fig. 11(b). From these plots we notice that the class border between non-neoplastic and neoplastic images shows up most clearly in case of the red color channel. While the plot for the green color channel also shows a distinct class border, this plot is highly correlated with the plot for the red color channel (which can be explained by the usually high correlation between these color channels within the images). Nevertheless, the plot for the green color channel also shows a pit count which is too low in case of quite a few non-neoplastic images. As a consequence the classification accuracies for this channel are lower as compared to the red channel results. The classification accuracies for the blue channel are even lower as compared to the green channel results, but the pit count plot for this channel shows no noticeable correlation with the plot for the red channel. As a consequence it seems that the information contained within the blue channel complements the red channel best, which is a reasonable explanation why in many cases the combination of the red and the blue channel yields the highest classification accuracy.

While until now we only investigated the statistical significance of differences between LOO-CV and LOPO-CV results, Table 6 shows a comparison between the previously proposed method [17] and the method proposed in this work. From this table we notice that the proposed method almost always delivers statistically significantly higher results in the 2-classes case (in case of the LOO-

Channels	2 classes		3 classes	
	$S_{\text{LOO-CV}}$	$S_{\text{LOPO-CV}}$	$S_{\text{LOO-CV}}$	$S_{\text{LOPO-CV}}$
R	✓ (+)	✓ (+)	✓ (+)	✓ (+)
G	✓ (+)	✓ (+)		
B				
R+G	✓ (+)	✓ (+)		✓ (+)
R+B	✓ (+)	✓ (+)		✓ (+)
G+B		✓ (+)		
R+G+B	✓ (+)	✓ (+)		✓ (+)
Best	✓ (+)	✓ (+)		✓ (+)

Table 6: Outcome of the tests for statistical significance (proposed method compared against the method from [17]).

CV results as well as in case of the LOPO-CV results; denoted in Table 6 by  $S_{\text{LOO-CV}}$  and  $S_{\text{LOPO-CV}}$ . Even when comparing the best performing channel combinations from both methods we see that the proposed method is able to significantly outperform the previously developed one in terms of the overall classification accuracy.

In the 3-classes case the picture is somewhat different. While the proposed method never performs significantly worse as compared to the previously developed method, in case of LOO-CV the observed result differences are in most cases not statistically significant. Only in case of LOPO-CV the proposed method is mostly able to significantly outperform the method proposed in [17].

In Tables 7 and 8 we present the highest results achieved by the proposed method along with accuracies yielded by the most promising previously developed methods (using wavelet-based features [13, 15, 31]). From these tables we notice that in the 2-classes case as well as in the 3-classes case the proposed method is outperformed by almost any previously developed method in case of LOO-CV. However, when using the more restrictive LOPO-CV the proposed method always delivers the highest classification accuracies. From the tables we also notice that – compared to the other methods – the proposed method is much more resistant against overfitting since most previously developed methods suffer from a rather high drop in terms of the overall accuracy when using LOPO-CV instead of LOO-CV, which is not the case for the proposed method. While in case of the proposed method the accuracies drop slightly too, the observed differences between LOO-CV and LOPO-CV are not statistically significant. For the previously developed methods, however, the observed accuracy drops are always statistically significant.

#### 6.4. Performance Analysis

It is out of question that the primary goal in medical image classification systems must be a high diagnostic accuracy. However, considering the fact that an endoscopic procedure should allow a real-time diagnosis (to enable the examiner to set an appropriate reaction like taking a biopsy or similar) it is also

Method	Specificity	Sensitivity	Accuracy	$S_{CV}$
Proposed	81.3 (77.3)	98.5 (98.5)	93.7 (92.6)	
[17]	56.1 (48.0)	98.5 (98.6)	86.7 (84.6)	✓ (-)
[13]	93.9 (58.0)	98.6 (98.0)	97.3 (87.0)	✓ (-)
[31]	96.5 (46.9)	97.9 (94.0)	97.5 (81.0)	✓ (-)
[15]	98.0 (75.3)	99.2 (96.3)	98.9 (90.4)	✓ (-)

Table 7: Comparison of the proposed method with previously published work with respect to the 2-classes classification results.

Method	Normal	Non-invasive	Invasive	Accuracy	$S_{CV}$
Proposed	77.3 (78.8)	97.4 (97.1)	13.3 (0.0)	80.3 (78.8)	
[17]	75.3 (68.7)	92.4 (89.5)	41.8 (2.0)	80.7 (71.8)	✓ (-)
[13]	93.9 (60.1)	97.1 (96.4)	84.7 (6.1)	94.6 (74.0)	✓ (-)
[31]	96.5 (50.0)	95.9 (83.8)	86.7 (5.1)	94.8 (63.7)	✓ (-)
[15]	98.0 (67.8)	98.8 (88.8)	95.9 (23.5)	98.2 (76.3)	✓ (-)

Table 8: Comparison of the proposed method with previously published work with respect to the 3-classes classification results.

important to create algorithms steering into that direction, and thus to develop fast and efficient feature extraction and classification methods.

The method proposed in this work is greatly simplified as compared to the method proposed in [17]. As a consequence, we also carried out a performance analysis to be able to assess whether the simpler method can also be computed more efficiently. The respective timing measurements can be found in Table 9. The columns  $\mathbf{T}_F$ ,  $\mathbf{T}_C$ , and  $\mathbf{T}_T$  denote the time needed to extract the features from one color channel, the time needed for the classification of a single color channel, and the total time needed by the respective method (i.e.  $\mathbf{T}_F + \mathbf{T}_C$ ), respectively. The column  $\mathbf{SF}$  indicates how many times higher the value of  $\mathbf{T}_T$  is, compared to the fastest method.

Since with our methods it is also possible to carry out the classification using a single color channel only, we provide the timing results for a single channel only. However, if combining two or three channels the time  $\mathbf{T}_T$  will be approximately two or three times higher, respectively.

From Table 9 we see, that the time needed for the classification based on a single color channel of an image is equal for both methods and does not contribute much to the total time  $\mathbf{T}_T$ . This is not surprising, since both methods use the same type of features (1D-histograms), classifier, and distance metric for the classifier. When looking at the time needed for the extraction of features from one color channel we see that the proposed method is roughly four times faster as compared to the previously developed one.

In order to allow a comparison to previously developed methods, Table 9 contains timing measurements for previously published methods too [13, 15, 31]. When comparing the timing measurements for these methods with the measurements for the proposed method we immediately notice that the proposed method is always significantly faster (up to 7.39 times).

Method	$T_F$	$T_C$	$T_T$	SF
<b>High-level features</b>				
Proposed	82	1	83	1.0
Häfner et al. [17]	325	1	326	3.93
<b>Wavelet-based features</b>				
Häfner et al. [13]	613	1	614	7.39
Häfner et al. [31]	143	1	144	1.73
Häfner et al. [15]	22	336	358	4.31

Table 9: Result of the performance analysis of the different methods (time measurements are given in milliseconds).

If we assume an endoscopy video to be captured at a frame rate of 25 frames per second a real-time application demands processing times of at most 40 milliseconds for a single frame. From the timing results in Table 9 we notice that this requirement is currently not met by any of the methods compared. Nevertheless, the significant speedup between the method from [17] and the proposed method represents a significant step into the direction of a real-time application.

## 7. Conclusion

In this work we proposed a method for the extraction of shape-based features. Since these features rely on the frequency and distribution of pits on the colonic mucosa, the method is specifically tailored to the pit pattern classification scheme for endoscopic polyp images.

In the past we already developed a similar method, which however suffered from rather low classification results and a poor performance in terms of the time needed to extract features from an image.

Since the method proposed in this work is greatly simplified as compared to the previously developed method, it is no surprise that the new method performs about four times faster.

But also in terms of the classification accuracy the newly developed method mostly outperforms the method proposed in [17]. We have shown that, no matter which cross-validation protocol we use (LOO-CV or LOPO-CV), the proposed method almost always delivers significantly higher classification accuracies at least in the 2-classes case. In the 3-classes case this holds true at least for LOPO-CV.

But, despite the superiority of the proposed method in terms of the classification accuracy and low computational demand, we have also seen that this method shares shortcomings with the method proposed in [17]. The specificities are usually rather low as compared to the sensitivities in the 2-classes case. We identified the rather unstable pit detection in case of non-neoplastic images as the main cause for this problem. In the 3-classes case the classification accuracy for the invasive class is also rather low as compared to the other image classes. This problem can be attributed to the fact that the differences between the

non-invasive class and the invasive class in terms of the number of detected pits are not that distinct as compared to the 2-classes case.

We have also shown that the proposed method is also able to outperform previously developed methods – in terms of the achieved classification accuracies as well as in case of the timing measurements. While the accuracies in case of LOO-CV are usually higher for the previously developed methods, in case of LOPO-CV the proposed method delivers superior results.

Nevertheless, since the method in [17] suffers from the same problems, the proposed method may be well regarded as a prospective alternative due to the rather low computational demand and the superiority in terms of the classification results obtained.

## 8. Acknowledgments

This work is partially funded by the Austrian Science Fund (FWF) under Project No. L366-N15 and Project No. TRP-206 and by the Austrian National Bank “Jubiläumsfonds” Project No. 12514.

## References

- [1] M. J. Bruno, Magnification endoscopy, high resolution endoscopy, and chromoscopy; towards a better optical diagnosis, *Gut* 52 (4) (2003) 7–11.
- [2] S. A. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. A. Karras, M. Tzivras, Computer-aided tumor detection in endoscopic video using color wavelet features, *IEEE Trans Inform Tech Biomed* 7 (3) (2003) 141–152.
- [3] D. E. Maroulis, D. K. Iakovidis, S. A. Karkanis, D. A. Karras, CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames, *Comput Meth Programs Biomed* 70 (2) (2003) 151–166.
- [4] S. M. Krishnan, X. Yang, K. L. Chan, S. Kumar, P. M. Y. Goh, Intestinal abnormality detection from endoscopic images, in: *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS'98)*, Hong Kong, China, 1998, pp. 895–898.
- [5] S. A. Karkanis, D. Iakovidis, D. Karras, D. Maroulis, Detection of lesions in endoscopic video using textural descriptors on wavelet domain supported by artificial neural network architectures, in: *Proceedings of the IEEE International Conference in Image Processing (ICIP'01)*, Thessaloniki, Greece, 2001, pp. 833–836.
- [6] D. K. Iakovidis, D. E. Maroulis, S. A. Karkanis, A. Brokos, 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 (CBMS'05)*, Dublin, Ireland, 2005, pp. 575–580.

- [7] D. K. Iakovidis, D. E. Maroulis, S. A. Karkanis, An intelligent system for automatic detection of gastrointestinal adenomas in video endoscopy, *Comput Biol Med* 36 (10) (2006) 1084–1103.
- [8] S. Gross, T. Stehle, A. Behrens, R. Auer, T. Aach, R. Winograd, C. Trautwein, J. Tischendorf, A comparison of blood vessel features and local binary patterns for colorectal polyp classification, in: *Proceedings of Medical Imaging 2008: Computer-Aided Diagnosis*, Vol. 6918, Orlando, USA, 2009.
- [9] T. Stehle, R. Auer, S. Gross, A. Behrens, J. Wulff, T. Aach, R. Winograd, C. Trautwein, J. J. W. Tischendorf, Classification of colon polyps in NBI endoscopy using vascularization features, in: *Proceedings of Medical Imaging 2009: Computer-Aided Diagnosis*, Vol. 7260, Orlando, USA, 2009.
- [10] J. J. W. Tischendorf, S. Gross, R. Winograd, H. Hecker, R. Auer, A. Behrens, C. Trautwein, T. Aach, T. Stehle, Computer-aided classification of colorectal polyps based on vascular patterns: a pilot study, *Endoscopy* 42 (3) (2010) 203–207.
- [11] B. André, T. Vercauteren, A. Perchant, A. M. Buchner, M. B. Wallace, N. Ayache, Endomicroscopic image retrieval and classification using invariant visual features, in: *Proceedings of the 6th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'09)*, Boston, Massachusetts, USA, 2009, pp. 346–349.
- [12] B. André, T. Vercauteren, A. Perchant, M. B. Wallace, A. M. Buchner, N. Ayache, Introducing space and time in local feature-based endomicroscopic image retrieval, in: *Proceedings of the MICCAI 2009 Workshop - Medical Content-based Retrieval for Clinical Decision (MCBR-CDS'09)*, London, UK, 2009, pp. 18–30.
- [13] R. Kwitt, A. Uhl, Modeling the marginal distributions of complex wavelet coefficient magnitudes for the classification of zoom-endoscopy images, in: *Proceedings of the IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'07)*, Rio de Janeiro, Brasil, 2007, pp. 1–8.
- [14] M. Häfner, A. Gangl, R. Kwitt, A. Uhl, A. Vécsei, F. Wrba, Improving pit-pattern classification of endoscopy images by a combination of experts, in: *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'09)*, London, UK, 2009, pp. 247–254.
- [15] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, F. Wrba, Pit pattern classification using extended local binary patterns, in: *Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine (ITAB'09)*, Larnaca, Cyprus, 2009, pp. 1–4.

- [16] M. Häfner, L. Brunauer, H. Payer, R. Resch, A. Gangl, A. Uhl, A. Vécsei, F. Wrba, Computer-aided classification of zoom-endoscopic images using fourier filters, *IEEE Trans Inform Tech Biomed* 14 (4) (2010) 958–970.
- [17] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, F. Wrba, Classification of endoscopic images using Delaunay triangulation-based edge features, in: *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR'10)*, Vol. 6112 of Springer LNCS, Povoá de Varzim, Portugal, 2010, pp. 131–140.
- [18] S. Kudo, S. Hirota, T. Nakajima, S. Hosobe, H. Kusaka, T. Kobayashi, M. Himori, A. Yagyuu, Colorectal tumours and pit pattern, *J Clin Pathol* 47 (1994) 880–885.
- [19] S. Kudo, S. Tamura, T. Nakajima, H. Yamano, H. Kusaka, H. Watanabe, Diagnosis of colorectal tumorous lesions by magnifying endoscopy, *Gastrointest Endosc* 44 (1) (1996) 8–14.
- [20] S. Kato, K.-I. Fu, Y. Sano, T. Fujii, Y. Saito, T. Matsuda, I. Koba, S. Yoshida, T. Fujimori, Magnifying colonoscopy as a non-biopsy technique for differential diagnosis of non-neoplastic and neoplastic lesions, *World J Gastroenterol* 12 (9) (2006) 1416–1420.
- [21] D. P. Hurlstone, High-resolution magnification chromoendoscopy: Common problems encountered in “pit pattern” interpretation and correct classification of flat colorectal lesions, *Am J Gastroenterol* 97 (2002) 1069–1070.
- [22] A. Rastogi, K. Pondugula, A. Bansal, S. Wani, J. Keighley, J. Sugar, P. Callahan, P. Sharma, Recognition of surface mucosal and vascular patterns of colon polyps by using narrow-band imaging: interobserver and intraobserver agreement and prediction of polyp histology, *Gastrointest Endosc* 69 (3 Suppl) (2009) 716–722.
- [23] C.-C. Chang, C.-R. Hsieh, H.-Y. Lou, C.-L. Fang, C. Tiong, J.-J. Wang, I.-V. Wei, S.-C. Wu, J.-N. Chen, Y.-H. Wang, Comparative study of conventional colonoscopy, magnifying chromoendoscopy, and magnifying narrow-band imaging systems in the differential diagnosis of small colonic polyps between trainee and experienced endoscopist, *Int J Colorectal Dis* 24 (2009) 1413–1419.
- [24] M. Raghavendra, D. G. Hewett, D. K. Rex, Differentiating adenomas from hyperplastic colorectal polyps: narrow-band imaging can be learned in 20 minutes, *Gastrointest Endosc* 72 (3) (2010) 572–576.
- [25] E. C. A. Zaroni, R. Cutait, M. Averbach, L. A. de Oliveira, C. R. Teixeira, P. A. Corrêa, J. L. Paccos, G. F. Rossini, L. H. C. Lopes, Magnifying colonoscopy: interobserver agreement in the assessment of colonic pit patterns and its correlation with histopathological findings, *Int J Colorectal Dis* 22 (2007) 1383–1388.

- [26] J. Kang, R. Doraiswami, Real-time image processing system for endoscopic applications, in: Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE'03), Vol. 3, Niagara Falls, Ontario, Canada, 2003, pp. 1469–1472.
- [27] A. Karargyris, N. Bourbakis, Identification of polyps in wireless capsule endoscopy videos using log gabor filters, in: Proceedings of the Life Science Systems and Applications Workshop (LiSSA'09), Bethesda, MD, USA, 2009, pp. 143–147.
- [28] S. Gross, M. Kennel, T. Stehle, J. Wul, J. Tischendorf, C. Trautwein, T. Aach, Polyp segmentation in NBI colonoscopy, in: W. Brauer, H.-P. Meinzer, T. M. Deserno, H. Handels, T. Tolxdorff (Eds.), *Bildverarbeitung für die Medizin 2009, Informatik aktuell*, Springer Berlin, 2009, pp. 252–256.
- [29] S. Hwang, M. E. Celebi, Polyp detection in wireless capsule endoscopy videos based on image segmentation and geometric feature, in: Proceedings of the 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'10), Dallas, Texas, USA, 2010, pp. 678–681.
- [30] M. Häfner, R. Kwitt, A. Uhl, A. Gangl, F. Wrba, A. Vécsei, Computer-assisted pit-pattern classification in different wavelet domains for supporting dignity assessment of colonic polyps, *Pattern Recogn* 42 (6) (2009) 1180–1191.
- [31] M. Häfner, R. Kwitt, A. Uhl, A. Gangl, F. Wrba, A. Vécsei, Feature-extraction from multi-directional multi-resolution image transformations for the classification of zoom-endoscopy images, *Pattern Analysis and Applications* 12 (4) (2009) 407–413.
- [32] M. Häfner, C. Kendlbacher, W. Mann, W. Taferl, F. Wrba, A. Gangl, A. Vécsei, A. Uhl, Pit pattern classification of zoom-endoscopic colon images using histogram techniques, in: Proceedings of the 7th Nordic Signal Processing Symposium (NORSIG'06), Reykjavik, Iceland, 2006, pp. 58–61.
- [33] T. Tamaki, J. Yoshimuta, T. Takeda, B. Raytchev, K. Kaneda, S. Yoshida, Y. Takemura, S. Tanaka, A system for colorectal tumor classification in magnifying endoscopic nbi images, in: R. Kimmel, R. Klette, A. Sugimoto (Eds.), *Computer Vision – ACCV 2010*, Vol. 6493 of LNCS, Springer Berlin, 2011, pp. 452–463.
- [34] Y. Takemura, S. Yoshida, S. Tanaka, K. Onji, S. Oka, T. Tamaki, K. Kaneda, M. Yoshihara, K. Chayama, Quantitative analysis and development of a computer-aided system for identification of regular pit patterns of colorectal lesions, *Gastrointest Endosc* 72 (5) (2010) 1047–1051.
- [35] C. B. Barber, D. P. Dobkin, H. Huhdanpaa, The Quickhull algorithm for convex hulls, *ACM Trans Math Software* 2 (4) (1996) 469–483.

- [36] K. Q. Brown, Voronoi diagrams from convex hulls, *Information Processing Letters* 9 (5) (1979) 223–228.
- [37] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, 2nd Edition, Wiley & Sons, 2000.
- [38] B. Everitt, *The Analysis of Contingency Tables*, Chapman and Hall, 1977.

**Michael Häfner** was born on December 9, 1968 in Vienna, Austria. He studied Medicine in Viennimagea and Perugia, Italy and graduated on January 23, 1995. He started working at the Department of Gastroenterology and Hepatology of the Medical University of Vienna in 1994, where he also was trained in Internal Medicine. He is specialist in Gastroenterology and Hepatology with his main focus on interventional endoscopy. Michael Häfner is currently secretary of the Austrian Society of Gastroenterology and Hepatology and organizer of several courses on interventional endoscopy, including an annual postgraduate course on endoscopy in Vienna. He acts as tutor at various national and international hands-on training courses. Recently he spent half a year in Malaysia to start up an endoscopy unit in Kuala Lumpur. In September 2009 he became chief physician of the Department of Internal Medicine at the St. Elisabeth Hospital, Vienna, Austria.

**Michael Liedlgruber** received his Master degree in Computer Sciences from the University of Salzburg (Austria) in 2006. Since 2007 he is a member of the Multimedia Signal Processing and Security Lab (WaveLab). In 2011 he received his PhD degree from the University of Salzburg (Austria) for his thesis focused on the computer-aided classification of endoscopic images from the gastrointestinal tract. He has published on automated polyp classification in colonoscopy, automated detection of celiac disease, and decision support systems in medical endoscopy of the gastrointestinal tract. Currently he is working as a Postdoc researcher at the Department of Computer Sciences at the University of Salzburg (Austria) in the same field of research. His main research interests include medical image processing and pattern recognition.

**Andreas Uhl** is an associate professor at the Department of Computer Sciences (University of Salzburg) where he heads the Multimedia Processing and Security Lab. His research interests include image and video processing and compression, wavelets, media security, medical imaging, biometrics, and number-theoretical numerics.

**Andreas Vécsei** graduated (Dr. med. univ.) at the University of Vienna. His post-graduate training in Austria and Kerala (India) made him receive his license to practice general pediatrics. After having been instructed in

pediatric gastroenterology according to the syllabus of GPGE (an association of German-speaking pediatric gastroenterologists) at the Medical University of Vienna and Munich, he received his certificate in 2006. Since 2006 he is in charge of the Endoscopy Unit of the St. Anna Children's Hospital, Vienna. His current research interests include *Helicobacter pylori* infection, celiac disease and computer-assisted evaluation of endoscopic images including the assessment of villous atrophy on endoscopic examination of the small intestine.

**Friedrich Wrba** is professor of Pathology at the Department of Clinical Pathology of the University of Vienna. After training as a general practitioner in various Austrian hospitals he started his special training in clinical Pathology, which he finished successfully at the Department of Clinical Pathology at the University of Vienna in the year 1987. In 1990–1992 he spent 2.5 years as a Visiting Research Fellow at the Imperial Cancer Research Fund Laboratories in London, England, (Walter Bodmer's laboratory). Since 1991 he is lecturer and Professor of Pathology, and staff member of the department. His areas of special interest are gastroenterology, hepatology, and molecular pathology.