Automated Marsh-like Classification of Celiac Disease in Endoscopic Images using Local Texture Operators based on Local Binary Patterns

Diplomarbeit

zur Erlangung des Diplomgrades
an der Naturwissenschaftlichen Fakultät der
Paris-Lodron Universität Salzburg

eingereicht von
Sebastian Hegenbart

Gutachter:
Ao.Univ.-Prof. Dr. Andreas Uhl

Fachbereich:
Computerwissenschaften

Salzburg, Juni 2010

Abstract

Celiac disease is a complex autoimmune disorder after introduction of gluten containing food. In recent years the prevalence numbers of celiac disease has been corrected upwards continuously. Celiac disease is therefore considered as one of the most common genetic disorders in Europe and the United States. Current estimates state that the prevalence of celiac disease is approximately one in 300 in the general population. Currently a histological confirmation of the characteristic changes within the tissue of the small bowel caused by the disease is considered the gold standard for diagnosing celiac disease. Following standards, at least four biopsy specimen are extracted during endoscopy to minimize the risk of false diagnosis caused by a patchy distribution of the villous atrophy. Once detected, the only treatment of celiac disease is a life long gluten-free diet. An early diagnosis followed by a strict diet can help the tissue to heal and reduce the risks of developing related complications including type 1 diabetes, autoimmune thyroid disease and autoimmune liver disease. To improve the diagnosis an automated system for assistance in targeting of biopsies is desirable. The development towards such a system faces manifold problems. In this work solutions to several of those problems are discussed. Fundamental questions concerning image acquisition and preparation are addressed first. Among those are the evaluation of the optimal endoscopic image capturing technique and the identification of robust and characteristic celiac specific markers that are promising for automated feature extraction. Based on the acquired image data several feature extraction methods based on local texture operators (Local Binary Patterns) are evaluated. The operators describe pixel neighborhoods in terms of pixel intensity differences as binary patterns. The distribution of these patterns are then used as features. A wavelet based approach for feature extraction combining suitable operators based on local binary patterns with the wavelet transform is presented. The final step towards diagnosis is the classification of the extracted features. An optimal margin classifier (SVM) as well as the k-Nearest Neighbor classifier are used. The classification process is based on a modified Marsh scheme representing a four class model as well as a more conservative approach using two classes.

Preface

Writing this thesis was a very demanding task. However, the amount of things I learned during the work on this project outweigh all the faced difficulties by a large margin. Nevertheless this thesis would not exist if it weren't for a few people I owe thanks for their support and backup.

I want to thank my parents Reinhard and Barbara as well as my entire family for always backing up my decisions and supporting me during my academic studies.

Furthermore I want to thank Andreas Uhl for supervising my thesis and always leading me in the right directions.

Thanks to Andreas Vécsei who helped me learning the characteristics of celiac disease and is supporting the work with endoscopic image data.

Finally I want to thank Michael Liedlgruber who helped me taking my first steps in the fields of medical image classification. Thanks you as well for many inspiring discussions.

Vielen Dank!

Sebastian Hegenbart
Salzburg 2010

Contents

1 Introduction

Automated classification as a support tool is an emerging option for endoscopic treatments. Systems are being developed that support physicians during surgery or highlight malignant areas during an endoscopy for further inspection. Such a system could also be used for training purposes of physicians that are new in the specific area. In the context of diagnosing celiac disease an automated system identifying areas suspected to be affected by the disease offers several benefits. The number of extracted biopsy specimen should be kept small. Therefore optimal targeting is desirable. Methods that highlight characteristic areas could improve the targeting of biopsies. The assistance of computers in diagnosis of celiac disease could save costs, time and manpower. Once celiac disease is diagnosed in a patient, the only treatment is a life long strict gluten free diet. This allows the tissue to heal leading to a resolution of all symptoms in the most cases. Depending on the time the disease is diagnosed, associated complications can be avoided. Hence it is of high interest that the diagnosis of celiac disease is made as early as possible. More information on the adoption of automated systems for supporting endoscopic treatments can be found (e.g. [28, 2, 1, 32]).

The development of such a system faces manifold problems. Solutions to several of those problems are discussed throughout this work. The cornerstones of an automated classification system are

- **Image Acquisition and Preparation**
  The first step is to acquire images to establish a training set of data. The training set is used to evaluate several methods for feature extraction and classification. All experiments conducted within this thesis are based on this set of training data. The construction of this set is important for the establishment of reliable methods. The acquiration of images in the context of this work, was done during endoscopies of the duodenum. The captured images were then prepared further for feature extraction. This was done by extracting image regions from the originally captured images exhibiting specific celiac markers, or characteristic details for the absence of the disease. The next step was to improve the visibility of the features by applying preprocessing. The process of image acquisition and preparation is thoroughly discussed in

Section 2.

- **Feature Extraction**
  The step following the image preparation is the feature extraction. The feature extraction is used to reduce and consolidate the information present within the input data. This can be formulated as a mapping

$$\varphi : I \subset \mathbb{R}^m \rightarrow F \subset \mathbb{R}^n, n < m \tag{1}$$

  for the set of all possible images I. Depending on the method (φ) that is used for feature extraction the dimensionality of the feature vectors (F) can vary considerably. The information essential for classification can be further consolidated by applying methods for feature selection. This results in a feature vector with reduced dimensionality. The methods that are used for feature extraction in this work are discussed in Section 3.

- **Classification**
  The final step involves assigning a label to each input sample and is known as classification. In general the classification can be seen as a mapping

$$\gamma : F \subset \mathbb{R}^n \rightarrow C \subset \mathbb{N} \tag{2}$$

  assigning a class-label to each input sample of a test set. Two classification schemes will be used throughout the experiments. The reduced Marsh classification scheme as covered in Section 2.3 and the more conservative two class case which reduces the number of classes by combining all three celiac type classes (Marsh-3a, Marsh-3b and Marsh-3c) to a single class. In this work classification methods based on supervised learning are deployed. An optimal margin classifier (SVM) as well as a k-Nearest Neighbor classifier are used. Section 4 covers these methods in more detail.

This work focuses on using local texture operators based on Local Binary Patterns for feature extraction. Several such operators exist that seem promising for use within an automated system. In order to gain a comprehensive view of the reliability and efficiency of the applied methods and their corresponding parameters we evaluate the efficiency of the system by conducting numerous experiments. Using a systematic approach the optimal combination of feature extraction techniques and classification methods are identified. The experiments are described in Section 5. The final conclusion (Section 7) compares the efficiency of the applied methods in terms of classification rate using the classical two-class model and the modified Marsh classification scheme.

1.1 Celiac Disease and Endoscopic Methodology

Celiac disease is one of the most common genetically based diseases. It is a complex autoimmune disorder in individuals of all age groups after introduction of gluten-containing food. Besides the term celiac disease the disorder has several other names found in literature including cœliac disease, co(e)liac sprue, non-tropical sprue, endemic sprue, gluten enteropathy and gluten intolerance. Gluten is a composite of the proteins glutenin and gliadin. These proteins can be found in grass-related grains notably wheat, barley and rye. The ingestion of gliadin by individuals affected by celiac disease causes an inflammatory reaction of the duodenal mucosa. This is caused by a cross reaction of the immune system and the small bowel tissue following the modification of the ingested protein by the tissue transglutaminase enzyme also known as tTG or TG2. During the course of the disease, hyperplasia of the enteric crypts occurs and the mucosa eventually looses its absorptive villi. This leads to a diminished ability to absorb nutrients. The prevalence of the disease has not been fully clarified yet. This is caused by the fact that depending on the severity of the malabsorbtion the symptoms vary among individuals. A large multi center study by Fasano et al. [17] reports that more than two million people in the United States, that is one in 133, are affected by the disease. People with untreated celiac disease are at risk for developing various complications like osteoporosis, infertility and other autoimmune diseases including type 1 diabetes, autoimmune thyroid disease and autoimmune liver disease.

An endoscope used for esophagogastroduodenoscopy (short EGD) is a small flexible tube equipped with a camera, a light source at the tip and one or more working channels. These channels are used for the removal of air inside the duodenum by suction and for the instillation of water. Endoscopes are usually also equipped with snares, injecting needles and a biopsy forceps to extract mucosal tissue. Besides standard upper endoscopy, several other methods for diagnosing celiac disease have been applied.

The modified immersion technique is based on the instillation of water into the duodenal lumen for better visualization of the villi. This technique has been described by Cammarota et al. [9]. Another technique is to use an endoscope with additional optical magnification to improve the accuracy of diagnosis [10]. Capsule endoscopy has been used as a non invasive approach. For accomplishing this, a patient swallows a small capsule equipped with a camera that captures images of the duodenal mucosa during its travel. The drawback of this technique is that no histological examination following the endoscopic procedure can be performed as no biopsy specimen are extracted. This approach has been described by Petroniene al. [45].

Currently a histological confirmation of the characteristic small bowel changes caused by celiac disease is considered the gold standard for diagnosing celiac disease in subjects with positive antibody testing. Due to the fact that histological changes can be patchy, at least four biopsy samples from the duodenal mucosa are recommended to improve the diagnostic accuracy.

2 Image Acquisition and Preparation

The first step towards an automated system for classification is to establish reliable methods for feature extraction and classification. To be able to conduct the experiments a consistent set of training data has to be created. Based on the image data that was gathered during the endoscopic sessions a test set of images was assembled. Endoscopic images vary greatly in visible characteristics of mucosal changes, perspective towards the tissue and quality in terms of distortions and noise. Even the quality within an image varies considerably among image regions. Because of this, and some more practically based issues like memory consumption and processing speed, it is not optimal to use the entire image for feature extraction and classification. Regions that are optimal in terms of classification are extracted and used for computing the features, that are used for the classification in the process.

This Section covers all aspects of image acquisition and preparation that are relevant for automated classification. Including the identification of robust and frequent markers specific for celiac disease that can be used within an automated system, the impact on classification through the application of two different image capturing techniques, geometrical differences caused by the shape of the duodenal regions and considerations regarding the region extraction. Finally the last step of preparation before extracting the features is an enhancement of the images in terms of feature visibility.

2.1 Image Acquisition

The used images were taken during duodenoscopies at the St.Anna Children's Hospital using pediatric gastroscopes without magnification (GIF-Q165 and GIF-N180, Olympus, Hamburg). The mean age of the children undergoing endoscopy was 11.1 years (ranging from 0.8-20.9 years). The female to male ratio was 1.43:1. The images were recorded during a single endoscopic session applying both, the conventional and the modified immersion technique. The ground truth used for classification was established by a histological examination of the biopsy specimen extracted during the endoscopic treatment. The images were recorded within two duodenal regions, the Pars Descendens and the duodenal Bulb. The two regions are part of the human duodenum.

The duodenal Bulb is the most upper part of the duodenum while the Pars Descendens is adjacent to the duodenal Bulb. Figure 8 presents the shape of the duodenum with the corresponding regions using a schematic illustration. As discussed in Section 2.5 these regions differ significantly in terms of geometric properties and hence images are of varying quality and exhibit changing perspectives. As a result of this differences we split the images into two distinct sets. One set contains images from the duodenal Bulb whereas the other set is comprised of images from the Pars Descendens. I will refer to these image sets as the 'Bulbus'-set and the 'Pars'-set throughout this work.

2.2 Celiac Specific Markers

The visual inspection of duodenal tissue in the context of diagnosing celiac disease involves the detection and identification of several celiac specific markers that are characteristic for the pathologic changes of the mucosa caused by the disease. The visibility and appearance of features is dependent on several factors, including the camera position and the duodenal region. Not all visible changes show great promise for automated classification. It is important to identify frequent and robust markers. A subset of celiac specific markers relevant for image classification include (among others in [14]):

- **Mosaic Mucosal Textures**
  A characteristic change of mucosal tissue that is affected by celiac disease is the visibility of textures that are similar to textures resembling a mosaic. These mosaic-like textures are not visible in case of healthy tissue as these textures are hidden by the villous structures.

- **Scalloping of the duodenal Folds** The duodenal folds increase the surface of the duodenum. In healthy tissue the folds are densely populated by villi. In case of villous atrophy however the visible population of villi on the folds decreases. As a consequence the so called 'scalloping' becomes visible. The scalloping is characterized by small notches along the folds. These notches are not visible if the villi on top of the fold are present.

- **Visualization of underlying Blood Vessels** In case of villous atrophy the underlying mucosa becomes visible. Therefore underlying blood vessels become visible. This is not visible in healthy tissue as the underlying blood vessels are all hidden by the villi.

- **Villous Atrophy (reduced size and modified shape of the villi)** As a consequence of villous atrophy, the scalloping if duodenal folds and mosaic like textures can be observed. Nevertheless these specific characteristics are not always visible (e.g. there is no duodenal fold within the image region). The severeness of villous atrophy can be estimated by the shape and size of the villi structures. In healthy tissue the villi are rather long. If the tissue is affected by the disease, the size of the villi decreases. Also a thicker and rounder shape can be observed in case of milder stages of the disease. The most severe form of villous atrophy shows no villi structures at all. The shape and size of the villi is the most important characteristic for the classification between the three celiac classes.

Figure 2.2 demonstrates the celiac specific markers. The left image shows the mosaic-like textures that are visible in celiac affected tissue. The image in the middle demonstrates the scalloping of the duodenal folds and the right image illustrates complete villous atrophy combined with the visualization of the underlying blood vessels.

(a) Mosaic Mucosa (b) Scalloping (c) Villous Atrophy

Figure 1: Images demonstrating Features that are specific for Celiac Disease.

Opposed to the markers specific for tissue affected by celiac disease, healthy tissue also exhibits certain features that can be used for classification. Nevertheless the absence of features specific for non-celiac tissue does not necessarily indicate celiac disease. These markers can be masked within image regions of low quality (exhibiting distortions or containing a high degree of blur). Patchy distributions of villous atrophy have been reported. Hence healthy regions within an image do not suffice to diagnose the absence of celiac disease. Anyway it is important to identify features in the non-celiac case. Markers specific for non-celiac tissue include:

- **Shape and Size of the Villi** The shape of healthy villi is an important characteristic. Long and straight villi indicate that the mucosa is not affected by villous atrophy. In affected mucosal tissue the size of the villi is reduced often combined with a round shape.
- **Appearance of the duodenal Folds** The scalloping of the duodenal folds is only visible if the villi density is low. This indicates celiac disease. In case of healthy tissue the villi on top of the folds are dense and long. Therefore no scalloping is visible.

Figure 2.2 demonstrates the markers that are specific to healthy tissue. The left image shows healthy structures with long and straight villi. The image on the right shows a duodenal fold with a comparable perspective to the middle image in 2.2. Opposed to the celiac image, the duodenal fold is densely populated by villi and no scalloping is visible.

(a) Villi (b) Appearance of the Fold

Figure 2: Images demonstrating Features specific for healthy Tissue.

Within this thesis the markers that are used for classification are the shape and size of the villous structures. This implicitly includes also the mosaic mucosa (it is visible if there is no villous structure), the scalloping of the duodenal folds as well as the visualization of the underlying blood vessels. In case of the classification in four classes using the modified Marsh scheme, the only discriminative characteristic between the three celiac classes is the shape and size of the villi.

2.3 Marsh Classification of Celiac Disease

The severity of celiac disease in a patient can be quantified according to Marsh [39]. Oberhuber et al. [41] subsequently proposed a standardized report scheme based on Marsh where celiac stage four was split into three sub stages. This scheme is known as the modified Marsh classification and is applied throughout this work. The quantification of the severity is based on three characteristics of celiac disease. The first is based on the number of intra epithelial lymphocytes as determined by a biopsy. The other two characteristics are based on visual changes of the duodenal mucosa caused by celiac disease. These are the shape, size and density of villi as well as the magnitude of gross enlargement of the crypts that is caused by hyperplasia.

Marsh Type	IEL per 100 Enterocytes	Crypts	Villi
Marsh-0	< 40	Normal	Normal
Marsh-1	> 40	Normal	Normal
Marsh-2	> 40	Increased	Normal
Marsh-3a	> 40	Increased	Mild atrophy
Marsh-3b	> 40	Increased	Marked atrophy
Marsh-3c	> 40	Increased	Absent

Table 1: Modified Marsh Classification of Celiac Disease

Table 1, as found in [3], lists the expansion of characteristics used to decide the severity of celiac disease according to the modified Marsh scheme. The Marsh types indicate the severeness of the disease where Marsh-0 stands for a normal mucosa and a very small chance of being affected by celiac disease. Type Marsh-1 is usually encountered in patients on a gluten free diet or family members of patients with celiac disease. Marsh-2 is very rare and usually seen in patients with dermatitis herpetiformis, a lesion that is associated with untreated celiac disease. Types Marsh-3a to Marsh-3c span the range of characteristic changes caused by celiac disease where Marsh-3a is the mildest form and Marsh-3c is the most severe form. The types Marsh-3a to Marsh-3c are the relevant types for automatic image classification. This is caused by the fact that celiac types Marsh-1 and Marsh-2 show no visual characteristics caused by the disease (except the hyper plastic changes of the crypts in the Marsh-2 case, which is not used for texture classification in this work).

(a) Marsh-0 (b) Marsh-3a (c) Marsh-3b (d) Marsh-3c

Figure 3: Images from the Duodenal Bulb showing the relevant Marsh Classes for Image Classification.

Figure 2.3 demonstrates the changes of the mucosal tissue due to celiac disease in the duodenal Bulb. On the left, long and straight villi can be observed. The three images on the right side show the decreasing length and changing shape of the villi due to the disease. The image on the right shows no villous structures at all. Additionally the underlying blood vessel are visualized very clearly.

(a) Marsh-0 (b) Marsh-3a (c) Marsh-3b (d) Marsh-3c

Figure 4: Images from the Pars Descendens showing the relevant Marsh Classes for Image Classification.

Figure 2.3 illustrates the characteristic mucosal changes caused by celiac disease. The image on the right shows healthy long villi while the other images show the decrease of visible structures. The image representing Marsh-3b already shows signs of visible scalloping on top of the fold. The appearance of the folds within the right image exhibits signs of heavy scalloping.

2.4 Endoscopic Image Capturing Techniques and their Impact on Automated Classification

In this thesis two image capturing techniques for acquiring endoscopic images were evaluated. Both techniques have their benefits but also introduce problems for automated classification systems. The key elements when evaluating the optimal image capturing technique for automated classification are

- **Visualization of Villi**
  Beside the endoscopic appearance of the duodenal mucosa the shape and size of the duodenal villi are the main factors in diagnosing celiac disease. For an image based classification system the visualization of specific features is the most important parameter towards a good performance of an automated system.

- **Distortions**
  The human bowel is a challenging environment for capturing images. The instillation of fluids and suction of air leads to small bubbles possibly covering characteristic changes of the mucosa. Furthermore specular reflections can be observed on the moist tissue frequently. Depending on the used image capturing technique the distortions have a varying impact for classification. Figure 2.4 illustrates distortions that are frequently visible within endoscopic images.

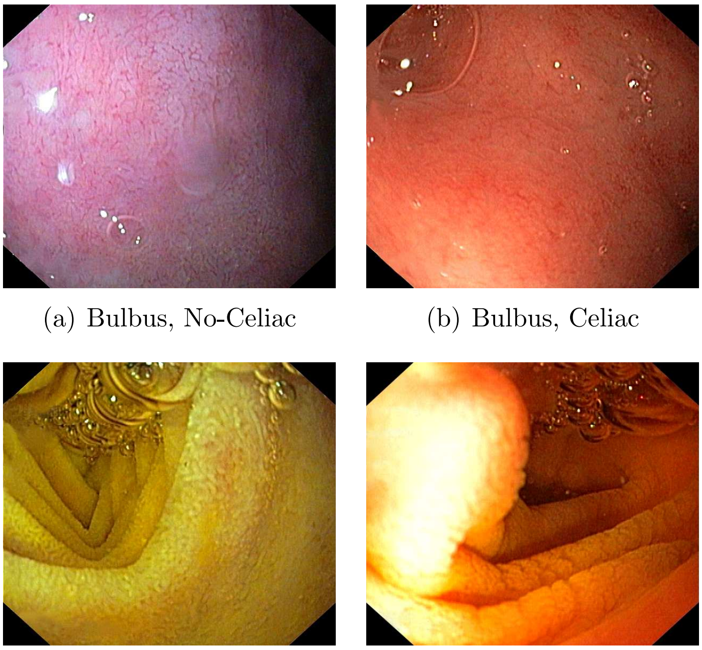
- **Blur and Exposure**
  The camera that is attached to the endoscope has a fixed focus and uses a point light source for illumination. This can lead to problems due to bad exposure. Underexposed regions contain a high degree of noise, this noise must not be interpreted as existing villi. Additionally, movements of the patient can introduce blur to the images. Blurred regions resemble areas with villous atrophy (absence of villous structures) and must not be interpreted as celiac specific markers.

(a) Reflections (b) Mirrored-Surface (c) Bubbles

Figure 5: Frequent Distortions within endoscopic Images.

2.4.1 The Modified Immersion Image Capturing Technique

Studies [18, 10] have shown that the modified immersion image capturing techniques can be used to directly visualize the duodenal villi during an endoscopy. The observation that villous structures can be better visualized in a duodenum filled with water lead to the development of this technique. The modified immersion technique uses rapid instillation of water into the duodenal lumen after removal of air by suction. The camera is then put into the water and the tissue is inspected for characteristic features. The removal of air is not always perfect and often results in an induction of bubbles. As a property of the modified immersion technique the bubbles swim on the surface of water and therefore do not cover the inspected areas. The instillation of water poses another interesting phenomenon. Some pictures (depending on the perspective) show a mirrored version of the captured area at the surface of the water. This has not been a problem so far, furthermore the additional redundancy could possibly be used for extracting features. Specular reflections can be observed caused by the instilled water. In contrast to the conventional image capturing technique, only small regions are concealed by specular reflections. Additionally the scatter of the reflections is lower when using the modified immersion technique. Figure 6 shows a selection of celiac and non-celiac images captured within the duodenal Bulb and the Pars Descendens using the modified immersion technique.



(a) Bulbus, No-Celiac

(b) Bulbus, Celiac

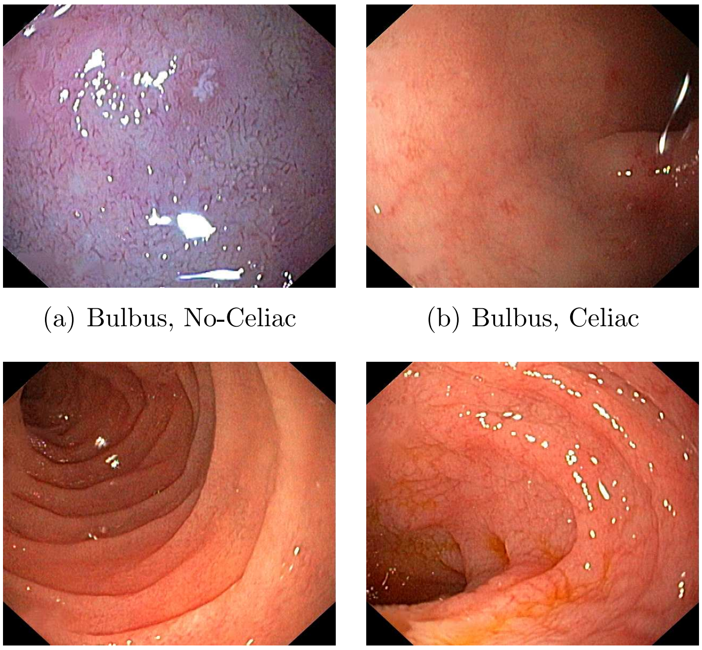
(c) Pars Descendens, No-Celiac

(d) Pars Descendens, Celiac

Figure 6: Images captured with the Modified Immersion Technique.

2.4.2 The Conventional Image Capturing Technique

The conventional image capturing technique involves no special treatment of the small bowel except the insufflation of air to expand the duodenum. In contrast to the modified immersion technique the lack of visualization of villi leads to a lower positive predictive value for the diagnosis of celiac disease [18]. In parallel to the other technique bubbles can be observed caused by the insufflation of air into the moist tissue. If both techniques are used in a single endoscopic session, remaining water in combination with air insufflation may lead to a higher density of bubbles. The bubbles in images captured by using the conventional technique frequently conceal characteristic mucosal textures. Opposed to the modified immersion technique the bubbles can not be avoided by putting the camera into the water. A type of distortion that is related to the conventional image capturing technique is the visibility of reflection on the tissue. Reflections could mask possible features used for the diagnosis. On the other hand, the way the tissue reflects light might be an indicator for celiac disease. However there has not been any research on this topic yet. Figure 7 presents images captured by using the conventional technique.



(a) Bulbus, No-Celiac

(b) Bulbus, Celiac

(c) Pars Descendens, No-Celiac

(d) Pars Descendens, Celiac

Figure 7: Images captured with the Conventional Technique.

2.4.3 Conclusion of Image Capturing Techniques

In a comparative study that was part of this project [23] we applied several feature extraction methods as well as classification techniques to images captured by using the modified immersion technique and images acquired through the conventional technique. The methods were comprised of Fourier-based, Wavelet-based and Dual Tree Complex Wavelet-based feature extraction techniques. The k-Nearest Neighbor classifier as well as a classifier based on Support Vector Machines and a statistical Bayes classifier were used for classification. Additionally an ensemble classifier as described in [21] was applied to maximize the classification rates.

For a thorough description of the methods and discussion of the results please refer to [23].

- **FFT-SVM**,
  Evolutionary Feature Selection, Classic features. [53] The FFT is used to transform the image into it's frequency spectrum. Multiple ring shaped filters are then applied to the Fourier spectrum of each color channel to select relevant subsets of the most discriminative coefficients. The feature vectors are then classified by a non-linear Support Vector Machines (SVM) classifier with radial basis function as kernel.

- **FFT-Bayes**,
  Evolutionary Feature Selection, Classic features. [53] FFT-Bayes uses the same feature extraction as the FFT-SVM method but applies a statistical Bayes classifier instead of a Support Vector Machines classifier.

- **DT-CWT, Classic**,
  Dual-Tree Complex Wavelet Transform, 6 scales, Classic features [30, 38]. The DT-CWT is used to decompose the images. Features are computed from the mean and standard deviation of the absolute values of the complex detail subband coefficients. This is the same setup as it is used in [38], except that we use the DT-CWT instead of the Gabor Wavelet Transform. The classification is performed by a 1-NN classification using the Euclidean metric.

- **LDB-WT**,
  Wavelet Decomposition Depth 3, Subband Energy (over all coefficients) as features [31]. The Local Discriminant Basis algorithm is employed to find an optimal Wavelet decomposition basis with respect to discrimination between the image classes. After transforming all images into this basis, for each of the resulting subbands the energy over all coefficients is computed for each color channel separately. The energy values of all channels are then concatenated to form the feature vectors, which are used in conjunction with the 1-NN classifier for the classification.

- **GMRF-WT**,
  Gaussian Markov Random Fields, Pyramidal Wavelet transform, Wavelet Decomposition Depth 2, Geman neighborhood of order 5 [20].

The Pyramidal Discrete Wavelet Transform is used to decompose the images. For the resulting subbands the Markov parameters are estimated for each color channel separately, using a Geman neighborhood of order 5. The concatenated parameters are subsequently used as feature vectors for the classification by using a Bayes classifier.

Method	**Classification Rates**					
	No-Celiac	Celiac	Total	No-Celiac	Celiac	Total
	Bulbus Immersion			**Bulbus Conventional**		
FFT-SVM	96.67	85.00	**93.24**	93.97	70.59	**88.66**
DT-CWT, Classic	92.80	87.50	**91.15**	88.79	55.88	**81.33**
GMRF-WT	92.05	85.00	**89.84**	91.38	58.82	**84.00**
	Pars Immersion			**Pars Conventional**		
FFT-Bayes	84.70	82.23	**83.42**	87.50	76.47	**82.93**
DT-CWT, Classic	82.51	82.74	**82.63**	64.58	52.94	**59.76**
LDB-WT	62.84	82.74	**73.15**	68.75	55.88	**63.41**

Table 2: A Selection of Results comparing the Conventional and the Modified Immersion Techniques.

Table 2 lists a subset of selected results. The experiments were all based on the two-class scheme for classification. We can see that in the case of the duodenal Bulb the overall classification rates of the immersion images are considerably above the rates achieved by the experiments based on images captured using the conventional technique. The same holds for images from the Pars Descendens.

Based on the results, the modified immersion technique is superior to the conventional technique in terms of overall classification rate. Beside the better performance of classification, also the overall image quality and utility for classification is superior to the conventional technique. As a consequence all experiments described in this work were performed on images acquired by using the modified immersion image capturing technique.

2.5 Duodenal Geometry and the Impact on Image Perspective and Classification

The duodenum is the first part of the small intestine. In adult humans the duodenum's length is approximately 30 centimeters and it's shape resembles a 'C'. The duodenum is divided into four parts. In this work we focus on the first two sections, the Pars Superior and the Pars Descendens. In humans the Pars Superior is extended to a Bulbus Duodeni (or duodenal Bulb). In the context of image acquisition and automated classification the duodenal Bulb and the Pars Descendens have some important differences.

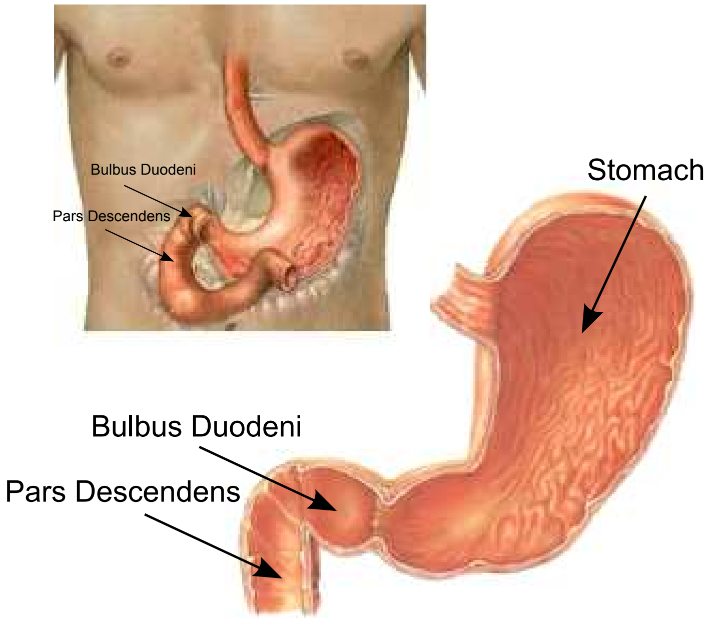


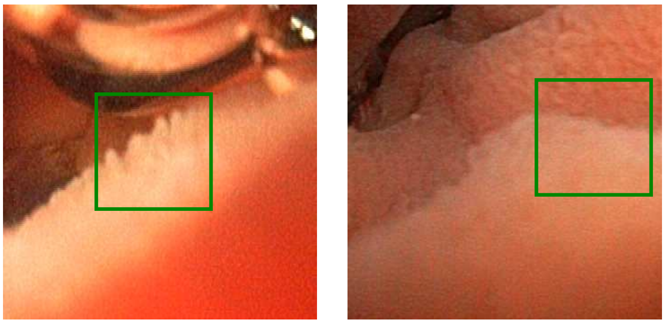
Figure 8: A Schematic of the Human Duodenum.

2.5.1 Bulbus Duodeni

The shape of the duodenal Bulb is rather flat and frequent changes in perspective are unlikely to happen. In general this region shows a high degree in homogeneity that is reflected within the images. The uniform orientation of textures is a very beneficial property for automated classification. Nevertheless the distance to the mucosa can vary heavily, depending on the position of the camera. Hence scale invariance within feature extraction methods, at least to some degree, is desirable. The duodenal Bulb does not contain duodenal folds. Therefore the scalloping of folds can not be used as a characteristic marker for images from this region.

2.5.2 Pars Descendens

The Pars Descendens is adjacent to the duodenal Bulb and in contrary has a tube-like shape. Besides the villi structures, the duodenal folds present within the Pars Descendens can indicate whether the tissue is affected by the disease or not. The tube-like shape introduces several problems relating to camera perspective and zoom. The mucosal shape on top of the folds, viewed from a lateral perspective, seems to be a good indicator. Nevertheless this kind of images usually suffers from a lack of visible texture. Figure 9 shows an example of such a problem caused by perspective. The green squares correspond to the image regions that would be used for feature extraction. Even though the distinction between celiac and non celiac markers is clearly visible at the edge in the images (corresponding to a duodenal fold), there are hardly any texture features that can be used due to the perspective towards the lumen center.



(a) No-Celiac (b) Celiac

Figure 9: Images from the Pars Descendens showing Problems caused by Perspective.

2.6 Extraction of Optimal Image Regions

The final step in preparing the image set is the extraction of optimal image regions (called patches) from within the original images captured during the endoscopic sessions. The extracted patches are then used to construct the set of training data. All this considerations are made for images with a size of 768×576 pixels (captured using the GIF-Q165 endoscope), and images with a size of 528×522 pixels (captured using the GIF-N180 endoscope).

2.6.1 Region Shape and Dimensions

It is not clear a priori which shape of the extracted region is optimal. Nevertheless most feature extraction methods require at least a rectangular shaped region. Furthermore some methods have even stricter requirements, so that a square region with the length of 2^n pixels per side was chosen as the basic shape for extraction. The next question targets the actual size of the region. Large regions are able to capture a lot of specific information while small regions might miss important features. On the other hand it is important to find an optimal positioning within the images. Distortions, blur and bad exposure are frequently observed within endoscopic images, hence a large size increases the possibility of containing such low quality image areas. This might influence the classification performance negatively. Also geometric properties have to be considered. As discussed in Section 2.5 the Pars Descendens has a tube-like shape. As a matter of fact, a too large region would contain tissue that is close to the camera as well as tissue that is more distant and therefore lacks resolution.

In the case of feature extraction methods based on Local Binary Patterns (see Section 3.2) the following considerations have to be taken into account:

- **Histogram Density**
  Small region sizes result in a lower number of pixels, and hence in a lower number of histogram entries. This problem is even increased when applying a threshold based operator. As the histograms are filled only sparsely the distance metrics become unreliable and the classification performance degrades.

- **Scale Invariance**
  As mentioned in Section 2.5 at least some degree of scale invariance within the texture operator is desirable for a promising classification performance. The local binary pattern based operators achieve this by considering pixel relations at a variable distance from the corresponding center. As the region size gets too small only a limited number of

pixels can be considered at a specific scale to avoid crossing the patch boundaries. This again results in sparse histograms as mentioned in the point above.

- **Histogram Noise**
  Large regions usually contain distortions visible within the extracted area. Without using some sort of distortion masking the operator can not distinguish between distorted pixels and informative pixels. These kind of distortions introduce noise to the histograms and decrease the classification performance.

The experiments show that when using regions of sizes larger than 128×128 pixels, distortions within the images can not reliably be avoided. Regions with a side length of 64 pixels or smaller pose the problem of creating sparse histograms. Using a region size of 128×128 pixels seems to be the best trade off between avoiding distortions and losing specific information. Therefore regions of this size were used for feature extraction throughout the experiments. Figure 10 demonstrates the different sizes. The blue square is of size 32×32 pixels and shows that not enough information can be captured. The green square demonstrates the actually used size of 128×128 pixels. The red region has a side length of 256 pixels and demonstrates that distortions and complications caused by geometric properties can not be avoided.

2.6.2 Region Extraction

A fully automated system would apply segmentation techniques to decide which parts of the image are suitable for feature extraction. As a first step towards automated diagnosis we need to establish reliable texture classification of the training data. As the main focus of this work lies on classification and feature extraction, this step was performed manually. We have created a set of textured image patches with optimal quality to assess if the classification is feasible under 'idealistic' conditions. Therefore the captured data was inspected and filtered by several qualitative factors (sharpness, distortions, visibility of features) as discussed in more detail in Section 2.4. The final step was extracting the regions from within the images. As illustrated in the previous section, texture patches with a fixed size of 128×128 pixels were extracted.

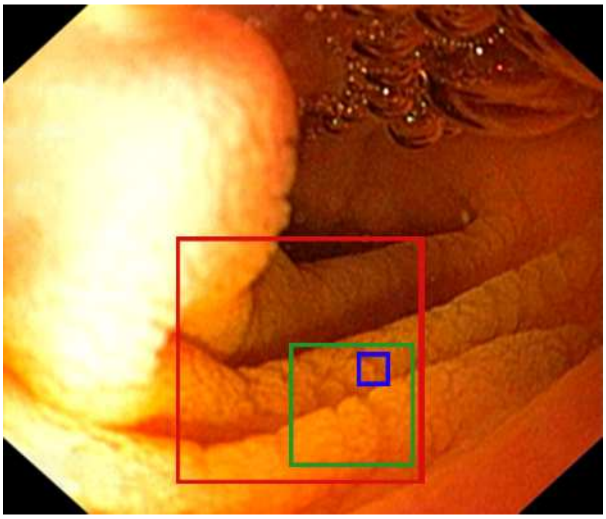


Figure 10: Image from the Pars Descendens demonstrating different Patch sizes.

As mentioned before, the extraction of the image regions was performed manually. To ensure that image regions contain characteristic markers, the extraction process was supported and supervised by an expert in the field of gastroenterology from the St.Anna Children's Hospital in Vienna (Dr. Andreas Vécsei). The image set that was used throughout this thesis was later built based on the extracted image regions from this two day lasting session.

2.7 Image Preprocessing

After acquiring the image data and extracting the corresponding regions, features can be extracted. Endoscopic images are subject to bad image quality (low contrast) and distortions, therefore an additional step of image processing is added. The image preprocessing is used to enhance the image contrast. In this project the contrast limited adaptive histogram equalization (CLAHE) [47] is used to improve the image quality. CLAHE was originally developed for medical imaging and has been applied successfully for enhancing low-contrast images. The histogram equalization improves the contrast of an image by spreading out the most frequent intensity values according to some predefined intensity distribution. Through this, the intensity values are better distributed and the visibility of features is improved.

A problem of the general histogram equalization is that uneven illumination can have a negative effect on the equalization process. This is very likely to happen within medical images. Therefore an adaptive histogram equalization (AHE) is used. The adaptive histogram equalization is based on the observation that our eyes adapt to the local context of images to evaluate their contents. Therefore the local image contrast is optimized instead of the global image contrast. To accomplish this, the image is divided in a grid of rectangular contextual regions. The number of contextual regions was determined by experimentation within this thesis and is set to 8×8 which corresponds to 64 contextual regions of size 16×16 pixels.

For each contextual region, the histogram of the contained pixels is calculated. The contrast is optimized for each contextual region by calculating the corresponding gray level assignment table for each of the corresponding cumulative histograms. The histogram equalization is therefore based on local image data. To avoid the visibility of region boundaries, a bilinear interpolation scheme along the boundaries is used (see Section 3.2.2 for more details).

Noise within images are a major drawback of the AHE method. Homogeneous areas are characterized by a high peak within the histogram since many pixels fall inside the same gray range. This leads to a higher visibility of noise. The problem is reduced by limiting the contrast enhancement specifically in homogeneous areas. With CLAHE, the slope associated with the gray level assignment scheme is limited. This is accomplished by allowing only a maximum number of pixels in each of the bins associated with local histograms. After clipping the histogram, the pixels that were clipped are equally redistributed over the whole histogram to keep the total histogram count identical. The chosen clip limit was found by experimentation and is set to 0.05 throughout the experiments. Please note that this number is implementation specific and does not directly relate to the discussion of CLAHE in [57]. The used implementation of CLAHE was provided by the Matlab Image Processing Toobox version 7.0 (please refer to Section 5.1 for an overview of the used software packages in this thesis).

The preprocessing of color images is slightly different from gray scale images.

The pixel intensity values in gray scale images represent the luminosity as the weighted sum of each color channel. Therefore no color information is lost through the histogram equalization. The naive approach for handling color images is to apply the preprocessing method to each separate color channel. Nevertheless in this case substantial inter-channel information is destroyed. To preserve this information the image is transformed into a color model with a luminance component.

In the experiments described in this thesis, sRGB is transformed to the CIE 1976 (L^*, a^*, b^*) (CIELAB) color space. The XYZ color system that was defined in 1931 by the Commision Internationale de L'Éclairage (CIE), builds the foundation of the CIELAB color space. The XYZ color space has a linear relationship with non-gamma-corrected RGB

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.240479 & -1.53715 & -0.498535 \\ -0.969256 & 1.875992 & 0.041556 \\ 0.055648 & -0.204043 & 1.057311 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{3}$$

The three coordinates of CIELAB represent the lightness of color ($L^* = 0$ yields black), its position between red, magenta and green (a^*, positive values indicate magenta while negative values indicate green) and its position between yellow and blue (b^*, negative values indicate blue while positive values indicate yellow). The lightness component is dependent on a reference white-point. Let X_n,Y_n,Z_n be the XYZ values of a reference white-point. The white point that is used in the implementation is the standard D65 white point, $X_n = 0.950456$, $Y_n = 1$, $Z_n = 1.088754$. The transformation from XYZ to CIELAB is defined as

$$g(t) = \begin{cases} t^{\frac{1}{3}}, & \text{if } t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29}, & \text{otherwise,} \end{cases} \tag{4}$$

$$
\begin{aligned}
L^* &= 116\left[g\left(\frac{Y}{Y_n}\right)\right] - 16 \\
a^* &= 500\left[g\left(\frac{X}{X_n}\right) - g\left(\frac{Y}{Y_n}\right)\right] \\
b^* &= 200\left[g\left(\frac{Y}{Y_n}\right) - g\left(\frac{Z}{Z_n}\right)\right]
\end{aligned} \tag{5}
$$

Figure 11 demonstrates the representation of color in the CIELAB color space.

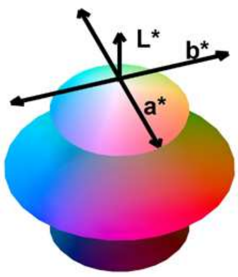


Figure 11: Schematic of the CIELAB Color Space.

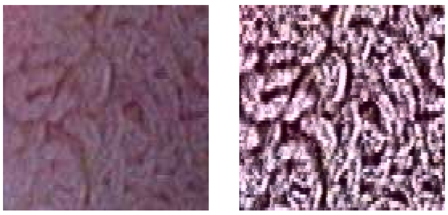
The contrast equalization is performed on the luminance component (L^*) without destroying the color information. The color space is then transformed back to sRGB and each color channel is used separately for feature extraction. The color conversion is realized by using XYZ as an intermediate color space for each transformation. Let us assume that $\delta = \frac{6}{29}$. The transformation from CIELAB to XYZ is defined as

$$
\begin{aligned}
f_y &= \left(\frac{L^* + 16}{116}\right) \\
f_x &= f_y + \left(\frac{a^*}{500}\right) \\
f_z &= f_y - \left(\frac{b^*}{200}\right).
\end{aligned} \tag{6}
$$

The specific coordinates in the XYZ color space are then calculated as

$$
\begin{aligned}
X &= \begin{cases} X_n\, f_x^3, & \text{if } f_x > \delta \\ \left(f_x - \frac{16}{116}\right) 3\delta^2\, X_n, & \text{otherwise} \end{cases} \\
Y &= \begin{cases} Y_n\, f_y^3, & \text{if } f_y > \delta \\ \left(f_y - \frac{16}{116}\right) 3\delta^2\, Y_n, & \text{otherwise} \end{cases} \\
Z &= \begin{cases} Z_n\, f_z^3, & \text{if } f_z > \delta \\ \left(f_z - \frac{16}{116}\right) 3\delta^2\, Z_n, & \text{otherwise} \end{cases} .
\end{aligned}
\tag{7}
$$

Figure 12 demonstrates the effect of applying the contrast limited adaptive histogram equalization (CLAHE) to a color image of class Marsh-0. We can see that the visibility of the villi structures is greatly enhanced by applying the preprocessing.



(a) Original (b) Preprocessed

Figure 12: Marsh-3c Image showing the Effect of Preprocessing.

3 Feature Extraction

The feature extraction is a vital step within the processing chain of an automated system. It can be seen as a consolidation of information that is present within an input signal. For a given input signal $I \in \mathbb{R}^m$ relevant information (according to the feature extraction method) is mapped to a feature vector $F \in \mathbb{R}^n$. This Section covers the relevant theoretical details behind the methods that were used for feature extraction. It can roughly be subdivided into two parts. The first part covers the mathematical details of the wavelet transform that is used in combination with the Local Binary Patterns to extract features. The latter part covers the theoretical details of the feature extraction methods that are based on LBP. Most implementation specific details as well as details that are directly related to the experiments are omitted until Section 5.

3.1 Introduction to Wavelets

The wavelet analysis has a wide range of applications in the fields of science. Ranging from pure mathematics, quantum physics, electrical engineering and seismic geology it has also made it's way into the field of digital signal processing. Nowadays wavelets are used for pattern recognition, signal compression, speech recognition and computer graphics, to mention just a few.

The wavelet analysis shows many different origins in the history of mathematics. The main branch in mathematics that led to wavelet analysis was the work of Joseph Fourier published in 1807 and is known as Fourier synthesis. In his work, Fourier showed that any continuous 2π-periodic function $f : \mathbb{R} \to \mathbb{R}$ is the sum of it's Fourier series

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos(nx) + b_n \sin(nx), \tag{8}$$

with the Fourier coefficients

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx, \quad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\cos(nx)\,dx, \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\sin(nx)\,dx. \tag{9}$$

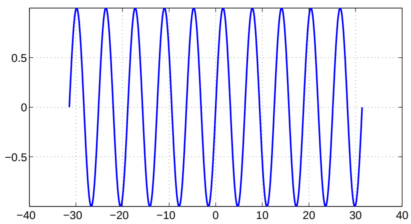
In subsequent works, mathematicians were gradually led from the notion of frequency analysis to the notion of scale analysis. Scale analysis is accomplished by analyzing functions using mathematical structures that vary in scale. In this case a function is constructed, shifted and changed in scale. This function is then used in approximating a signal. The procedure is repeated by shifting and rescaling the basic structure again and using it to approximate the same signal.

The term wavelet was first mentioned by Haar in 1909. He was interested in finding a basis in a function space similar to Fourier's basis in frequency space. During the 1930s researches from different fields picked up the idea of wavelets. In physics, wavelets were used in the characterization of Brownian motion, wavelets were also used for the analysis of coherent states of a particular quantum system. Between 1960 and 1980 research towards finding the simplest elements in a function space (called atoms), with the goal to be able to reconstruct all elements within this function space was done by Coifman and Weiss. In 1985, Stephane Mallat discovered the benefits of wavelets within the field of signal processing. In subsequent works Meyer developed the first non-trivial wavelets. Later Ingrid Daubechies constructed a set of wavelet orthonormal basis functions that are among the cornerstones of wavelet analysis today.

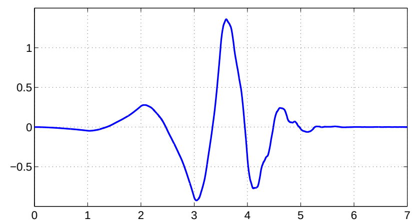
This Section covers only the basics of wavelet analysis. For a more detailed description please consult [12],[54],[37],[48],[19],[8].

3.1.1 Wavelets

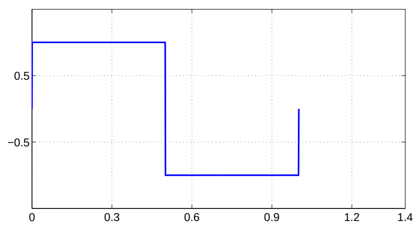
Wavelets are mathematical functions that are used to study signals at different frequency components with a resolution matched to it's scale. A frequently used phrase in the context of wavelets that describes the idea well is "seeing both, the wood and the trees". Wavelets have advantages over the sines and cosines used by Fourier when analyzing signals with sharp spikes or discontinuities. The problems lie within the periodicity of the sines and cosines (they are not localized in space). The wavelet analysis uses functions with compact support as basis. This means that wavelet functions vanish outside a finite interval and are in contrast to the Fourier basis functions localized in space. Figure 13 demonstrates two wavelet functions with compact support compared to the 2π-periodic sine function.



(a) Sine



(b) Daubechies 4 Wavelet



(c) Reverse Biorthogonal Wavelet 1.1

Figure 13: Wavelet Functions compared to a Sine Wave.

3.1.2 The Continuous Wavelet Transform

For a signal with infinite energy, it is impossible to cover it's frequency spectrum and it's time duration with wavelets. Hence wavelets can only be used for square-integrable functions

$$\int_{-\infty}^{\infty} |f(t)^2| dt < \infty. \tag{10}$$

Two square-integrable real-valued functions have an inner product

$$\langle f, g \rangle = \int f(t) g(t)\, dt. \tag{11}$$

The norm defined by the inner product is

$$\|f\|^2 = \langle f, f \rangle. \tag{12}$$

A wavelet system is generated from a wavelet by simple translation and scaling. The two-dimensional parametrization is achieved from the function (called the mother wavelet or generating wavelet) $\psi(t)$ by

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\, \psi\left(\frac{t-b}{a}\right) \qquad a \in \mathbb{R} \setminus 0,\ b \in \mathbb{R}. \tag{13}$$

This is in contrast to the Fourier transform, the basis functions are not defined at this point. The continuous wavelet transform of a square-integrable signal $f(t)$ is

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\, \psi\left(\frac{t-b}{a}\right)\, dt, \tag{14}$$

for the scale parameter a and the translation parameter b. Please note that ψ is implicitly assumed to be real. If it is complex, complex conjugates have to be introduced to the inner products. The term of the integral wavelet transform can be simplified further by using Equation (13)

$$T(a,b) = \int_{-\infty}^{\infty} f(t)\, \psi_{a,b}(t)\, dt. \tag{15}$$

The signal f is represented using the coefficients in a linear expansion of the wavelet functions (the inverse wavelet transform), C_ψ is a constant value depending on the mother wavelet function

$$f = C_\psi^- 1 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{da\, db}{a^2}\, T(a,b)\, \psi_{a,b}. \tag{16}$$

Assuming that $\|\psi\| = 1$, and all ψ are orthogonal, $T(a,b)$ can be calculated using the inner product, as

$$T(a,b) = \langle f, \psi_{a,b} \rangle. \tag{17}$$

If the set of basis functions is not orthogonal, a dual basis set $\widetilde{\psi_{i,j}}$ exists such that using Equation (17) with the dual basis gives the desired coefficients.

While theoretically elegant the continuous wavelet transform poses several practical problems. The wavelet transform is calculated by continuous shifting of a continuously scalable function over a signal and calculating the correlation of these two. It is unlikely that the scaled wavelets build an orthogonal basis in this case. Hence the obtained wavelet coefficients will be highly redundant. Another problem is the infinite number of used wavelets in the continuous wavelet transform.

3.1.3 The Discrete Wavelet Transform

Discrete wavelets are not scalable and translatable continuously. They can only be scaled and translated in discrete steps. Hence the scale parameter a and the translation parameter b are integers in the discrete wavelet transform. The wavelet parametrization is therefore changed to

$$\begin{aligned} \psi_{a,b}(t) &= \frac{1}{\sqrt{\sigma_0^a}}\, \psi\left(\frac{t - b\tau_0\, \sigma_0^a}{\sigma_0^a}\right) \\ &= \frac{1}{\sqrt{\sigma_0^a}}\, \psi\left(\left(t - b\tau_0\, \sigma_0^a\right) \sigma_0^{-a}\right) \\ &= \frac{1}{\sqrt{\sigma_0^a}}\, \psi\left(t\sigma_0^{-a} - b\tau_0\right) \qquad a, b \in \mathbb{Z}. \end{aligned} \tag{18}$$

We restrict $\tau_0 > 0\,, \sigma_0 > 1$. In this case τ_0 and σ_0 represent stepping factors for the translation and scaling parameters. In the case of $\sigma_0 = 2$ and $\tau_0 = 1$ the sampling frequency of the time and frequency axis corresponds to the dyadic sampling. In analogy to the continuous wavelet transform, the discrete wavelet transform of a continuous square-integrable signal $f(t)$ can be expressed as

$$\begin{aligned} T(a,b) &= \frac{1}{\sqrt{\sigma_0^a}} \int_{-\infty}^{\infty} f(t) \, \left(t\sigma_0^{-a} - b\tau_0\right) \, dt \\ &= \int_{-\infty}^{\infty} f(t) \, \psi_{a,b}(t) \, dt. \end{aligned} \tag{19}$$

A numerically stable reconstruction of the signal can be achieved if the wavelet coefficients of the discrete wavelet transform lie between positive bounds

$$A \, \|f\|^2 \leq \sum_{a,b} |\langle f, \psi_{a,b} \rangle|^2 \leq B \, \|f\|^2 \qquad A > 0, B < \infty \in \mathbb{R}. \tag{20}$$

If this equation is satisfied, the set of wavelets $\psi_{a,b}$ build a so called frame with bounds A and B. If $A = B$ the frame is called tight and the discrete wavelets behave like an orthonormal basis. In this case the inverse discrete wavelet transform is

$$f = \sum_a \sum_b T(a,b) \, \psi_{a,b}. \tag{21}$$

In the case of $A \neq B$ a dual frame is used for reconstruction of the signal. In a dual frame discrete wavelet transform, the decomposition wavelet is different from the reconstruction wavelet. Such wavelets are called biorthogonal. In analogy to the continuous wavelet transform $T(a,b)$ can again be calculated using the inner product

$$T(a,b) = \langle f, \psi_{a,b} \rangle. \tag{22}$$

3.1.4 The Scaling Function

Although the redundancy of the wavelet coefficients can be resolved by using the discrete wavelet transform (in the case of orthogonal wavelets) the practical problems mentioned in Section 3.1.2 still hold for the discrete wavelet transform. In parallel to the continuous wavelet transform still an infinite number of scalings and translations is needed. For practical applications a finite number of wavelets have to suffice. An upper bound on the number of translations is given by the length of the signal. A wavelet can be seen as a band-pass filter and scaling it creates the problem of reducing the covered bandwidth. Hence an infinite number of wavelets would be needed to cover the entire spectrum. This is where the scaling function comes into use.

A set of scaling functions is described in terms of integer translations of the basic scaling function by (in the case of the dyadic sampling)

$$\phi_{a,b}(t) = \frac{1}{\sqrt{2^a}}\,\phi(2^{-a}t - b). \tag{23}$$

These functions span a subspace V_0 of $L^2(\mathbb{R})$. By fulfilling the following requirements, the subspaces have a nesting structure

$$V_j \subset V_{j+1}, \qquad \forall j \in \mathbb{Z} \tag{24}$$

with

$$V_{-\infty} = \{0\}, \qquad V_{\infty} = \mathbb{L}^2. \tag{25}$$

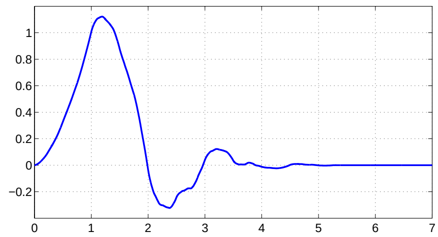
A space containing signals of higher resolution, will also contain those of lower resolution. The subspaces satisfy a natural scaling condition

$$f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1}. \tag{26}$$

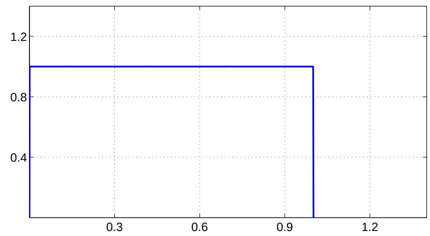
This means that $\phi(t)$ can be represented as a linear expansion of $\phi(2t)$ as

$$\phi(t) = \sum_{n} h(n)\, \sqrt{2}\, \phi(2t - n), \qquad n \in \mathbb{Z}. \tag{27}$$

The coefficients $h(n)$ are called the scaling function coefficients (or the scaling filter). The $\sqrt{2}$ maintains the norm of the scaling function with the scale of two. Figure 14 illustrates the scaling functions of two popular wavelet functions.



(a) Daubechies 4



(b) Reverse Biorthogonal Wavelet 1.1

Figure 14: Scaling Functions.

3.1.5 The Wavelet Function

The important features of a signal can better be described, not by using $\phi_{a,b}(t)$ and increasing a to increase the size of the subspace spanned by the scaling function, but by defining a slightly different set of functions $\psi_{a,b}(t)$ that span the differences between the spaces spanned by the various scales of the scaling function.

There are several advantages to requiring that the scaling functions and wavelets be orthogonal. Orthogonal basis functions allow simple calculation of expansion coefficients and have a Parseval's theorem that allows a partitioning of the signal energy in the wavelet transform domain. The orthogonal complement of V_j in V_{j+1} is defined as W_j. This means that all members of V_j are orthogonal to all members of W_j. We require

$$\langle \phi_{a,b}(t), \psi_{i,j}(t) \rangle = 0 \tag{28}$$

for all appropriate $a, b, i, j \in \mathbb{Z}$. We now define the wavelet spanned subspace W_0 such that

$$V_1 = V_0 \oplus W_0. \tag{29}$$

In general this gives

$$\mathbb{L}^2 = V_0 \oplus W_0 \oplus W_1 \oplus \ldots \tag{30}$$

V_0 is the initial space spanned by the scaling function. The symbol $\oplus$ has the meaning of the disjoint differences except for the zero element ($U = V \oplus W \Leftrightarrow (U = V \cup W) \wedge (V \cap W = \{0\})$). The scale of the initial space is arbitrary and is usually chosen to represent the coarsest detail of interest in a signal. Since these wavelets reside in a space represented by the next narrower scaling function, they can be represented by a weighted sum of shifted scaling functions

$$\psi(t) = \sum_n g(n)\, \sqrt{2}\, \phi(2t - n), \qquad n \in \mathbb{Z}. \tag{31}$$

The wavelet coefficients are related to the scaling function coefficients by

$$g(n) = (-1)^n\, h(1 - n). \tag{32}$$

The function generated in Equation (31) gives the following prototype mother wavelet (in the case of dyadic sampling)

$$\psi_{a,b}(t) = \frac{1}{\sqrt{2^a}}\,\psi(2^{-a}t - b). \tag{33}$$

3.1.6 Filter Banks

In practice only the coefficients of $g(n)$ and $h(n)$ of Equations (31) and (27) have to be considered. The scaling coefficients for a scale of j can be calculated as

$$c_j(k) = \sum_m h(m - 2k)\, c_{j+1}(m). \tag{34}$$

while the wavelet coefficients can be calculated as

$$d_j(k) = \sum_m g(m - 2k)\, c_{j+1}(m). \tag{35}$$

Equation (34) in combination with Equation (35) form one stage of an iterated filter bank. The coefficients $h(k)$ are known as the scaling filter and the coefficients $g(k)$ are known as the wavelet filter.

Figure 15 demonstrates two steps of an iterated filter bank. The circles demonstrate the subsampling property of filter banks. As can be seen in Equations 34 and 35 the scaling and wavelet filters have a step size of 2 in the variable k hence only half of each filter output characterizes the signal. The coefficients d_i are also known as detail coefficients while the coefficients c_i are known as approximation coefficients.

3.1.7 The Pyramidal Wavelet Decomposition

So far, the discussion of the wavelet transform has covered only one dimensional signals. Nevertheless images as a frequent subject to this technique consist of two dimensional signals. In Mallat's vertical and horizontal analy-

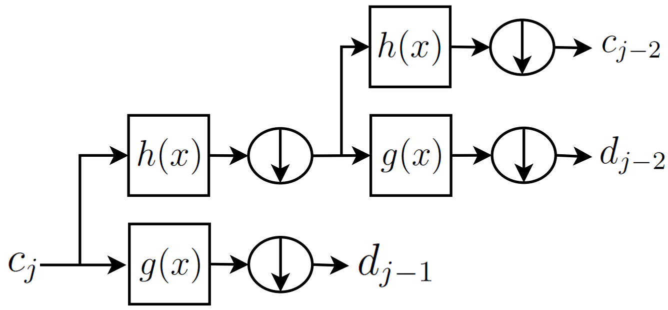


Figure 15: Two Stages of an Iterated Filter Bank.

sis [37] the decomposition algorithm is based on two variables x and y leading to a priorization of each direction. The scaling function is defined as

$$\phi(x,y) = \phi(x)\,\phi(y). \tag{36}$$

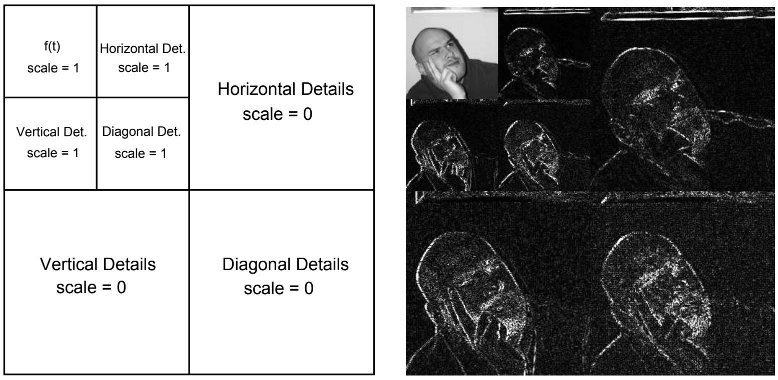
The signal is derived from three wavelets in this case

- vertical wavelet:
$$\psi^1(x,y) = \phi(x)\,\psi(y) \tag{37}$$

- horizontal wavelet:
$$\psi^2(x,y) = \psi(x)\,\phi(y) \tag{38}$$

- diagonal wavelet:
$$\psi^3(x,y) = \psi(x)\,\psi(y) \tag{39}$$

This leads to three subimages each representing the details of the corresponding spatial orientation. Figure 16 illustrates this process by using a schematic on the left side and a wavelet decomposed natural image on the right. For demonstrating purposes, an image of a person was chosen (due to the nature of endoscopic imagery the properties of the subbands are difficult to spot for the unexperienced eye).



(a) Pyramidal Decomposition (b) Example

Figure 16: The Pyramidal Wavelet Decomposition.

3.2 Introduction to Local Binary Patterns

Methods for describing textures have been divided into two categories traditionally. The first category, called the statistical or stochastic approach, models textures by their statistical properties. This is done in terms of pixel intensity values and pixel positions. Co-occurrence matrices [22] and difference histograms [55] are popular methods within this category. The other class of methods is based on describing structural properties of textures. This is done by using basic structural elements called textons. Textons [27] are considered as the basic units of human preattentive texture discrimination (orientation elements, crossings and terminators). Tuceryan et al. [52] divide the recent developments of texture classification methods in geometric, model based, statistical and signal processing approaches. In the recent past Gaussian Markov Random fields were extensively used to model the process of texture formation.

The Local Binary Patterns fill the gap between statistical and structural methods (for a schematic demonstrating the relations see [34]). The operator was first introduced by Ojala et al. [42]. Since then many modifications and improvements have been suggested. These improvements range from multiscale, multiresolution, rotation invariance and gradient based operators to wavelet-based operators. The entire family of operators is used to model a pixel neighborhood in terms of pixel intensity differences. The operators as-sign a binary number to each possible pixel neighborhood. The distributions of these neighborhoods are then used as features. In contrary to the wavelet transform, the local binary pattern operators do not modify the input signal, they are solely used to extract features from the signal that are later used for classification.

Local binary patterns have been successfully used in several applications. The applications comprise image segmentation, biometrics and texture classification in general. Additionally, the LBP operator is invariant to monotonic intensity variations which is beneficial to texture classification in environments with varying lighting conditions. This properties make the local binary patterns suitable for extracting features in endoscopic images. The following Sections will cover the theoretical details of the various modifications and extensions to the local binary pattern operator. For simplicity and to improve the readability the term 'local binary patterns' will be abbreviated as LBP when suitable in the following text.

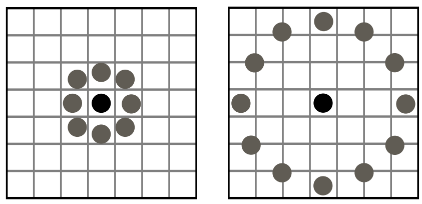
3.2.1 Local Binary Patterns (LBP-Operator)

The basic LBP operator was introduced to the community by Ojala et al. in [42]. In [36] Malik et al. extended the Texton model to gray scale textures. Their method includes Gabor filtering and hence includes calculating the weighted mean of pixel values in a small neighborhood. The LBP operator considers each pixel in a neighborhood separately. Hence the LBP could be considered as a micro-texton.

The operator is used to model a pixel neighborhood in terms of pixel intensity differences. This means that several common structures within a texture are represented by a binary label. The joint distributions of these labels are then used to characterize a texture. The LBP operator is parametrized by two variables, namely the number of neighbors and the radius. The pixel intensity differences in the LBP operator are expressed by using a sign function

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}. \tag{40}$$

Two popular approaches of defining neighbor positions exist. A fixed 8-neighborhood uses the neighboring pixels on a pixel grid while the symmetric circular neighborhood uses equidistant pixel positions from a center pixel. The circular neighborhood has the advantage over the fixed 8-neighborhood that a variable number of neighbors can be used. A drawback is that the circular neighborhood requires some sort of pixel interpolation mechanism as the centers of the pixel neighbors are not exactly on pixel boundaries in the general case. In this work the circular neighborhood is used in combination with a bilinear interpolation mechanism. Figure 17 illustrates the neighbor pixel positions considered when using a circular symmetric neighborhood. The abbreviation 'p' refers to the number of used neighbors, 'r' stands for the used radius from the center.



(a) p = 8, r = 1.0 (b) p = 12, r = 2.5

Figure 17: Pixels on a Circular Neighborhood.

The coordinates of neighbor number $k \in 1, \cdots, p$ of the center pixel at location (x, y) for an LBP operator with radius r using p neighbors is calculated by

$$\phi^x_{r,p}(x, k) = x + r \cos\left(\frac{2\pi k}{p}\right) \tag{41}$$

$$\phi^y_{r,p}(y, k) = y - r \sin\left(\frac{2\pi k}{p}\right) \tag{42}$$

For simplicity we combine Equations 41 and 42 as

$$\phi_{r,p}(x, y, k) = (\phi^x_{r,p}(x, k), \phi^y_{r,p}(y, k)). \tag{43}$$

Let $I(x, y)$ be the pixel at location (x, y) of an image I. The operator $LBP_{r,p}(x, y)$ for location (x, y) of an image I using p neighbors with a radius of r is then defined as

$$LBP_{r,p}(x, y) = \sum_{k=0}^{p-1} 2^k \, s(I(\phi_{r,p}(x, y, k) - I(x, y)). \tag{44}$$

The Local Binary Pattern operator assigns a label ranging from 0 to $2^p - 1$ to each pixel location within an image, describing the pixel neighborhood. The labels are collected into a histogram with 2^p bins describing the distribution of the pixel neighborhoods.

Figure 18 demonstrates the calculation of a local binary pattern. The intermediate windows show the pixel intensity relations which is assigned by the sign function ($s(x)$ as defined in Equation (40)) and the corresponding weights. The window on the right shows the process of assigning a weighted sum to the pixel intensity neighborhood as defined in Equation (44).

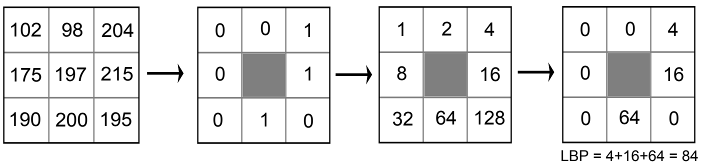


Figure 18: Schematic demonstrating the calculation of an LBP String.

Unfortunately the long notation used in the Equations above is inconvenient to read. I will therefore stick to a simplified notation as can be found throughout the literature concerning local binary patterns,

$$LBP_{r,p}(x, y) = \sum_{k=0}^{p-1} 2^k \, s(I_k - I_c). \tag{45}$$

With I_k being the value of neighbor number k and I_c being the value of the corresponding center pixel. Whenever this notation leads to ambiguity the long notation is used instead. The LBP histogram of an image I is formally defined as

$$H_I(i) = \sum_{x,y}(LBP_{r,p}(x,y) = i) \qquad i = 0, \cdots, 2^p - 1. \tag{46}$$

3.2.2 Interpolation

By using a symmetric circular neighborhood the operator extracts information from subpixels rather than pixels. In most cases the neighbors lie on the bounds between two or more pixels. Therefore a mechanism is needed to calculate the values of those subpixels. In mathematics interpolation is used if a set of discrete values is given and a corresponding approximating continuous function which fits the given values exactly is needed. This function is then said to interpolate the given values. Considering digital images using pixels, the values at each pixel position are known, however the values that do not lie on integer locations are unknown. In order to be able to use the symmetric circular neighborhood interpolation is used.

Many methods for interpolation exist. However the complexity of the applied interpolation method plays a big role in the decision which method is used. For each computed pattern at maximum all considered neighbors have to be interpolated. This accounts for a huge part of the entire computational effort. Although the linear interpolation is not as precise as other methods (e.g. spline interpolation) the tradeoff between computational complexity and preciseness favors this method. Therefore the linear interpolation method is used to interpolate the values of neighbors that lie on pixel boundaries in this thesis.

In one dimension, linear interpolation of an unknown value is based on two anchor locations for which the values are known. For this, the location of the searched (interpolated) value has to lie between these two known locations. If x_l is the location left of the interpolated value x, and x_r lies on the right respectively, the interpolated value $f(x)$ lies on a straight line between $f(x_l)$ and $f(x_r)$ (hence the name linear interpolation).

$$\frac{f(x) - f(x_l)}{x - x_l} = \frac{f(x_r) - f(x_l)}{x_r - x_l} \tag{47}$$

Solving this equation for $f(x)$ which is the desired value at location x, the interpolation formula is

$$f(x) = f(x_l) + (x - x_l)\frac{f(x_r) - f(x_l)}{x_r - x_l}. \tag{48}$$

In case of using images, the interpolated function is two-dimensional however. Therefore the interpolation mechanism has to be adapted. This is done by first interpolating the values in the x-axis. As a second step the locations of the values that were interpolated in the first step are used to interpolate the value in the y-axis. In case of interpolating subpixel values the four diagonal neighboring pixels are used as sampling points.

A chronological summary of interpolation methods ranging from the ancient to modern signal processing with their applications in science can be found in [40]. Figure 19 demonstrates the two steps involved in the bilinear interpolation of a subpixel.

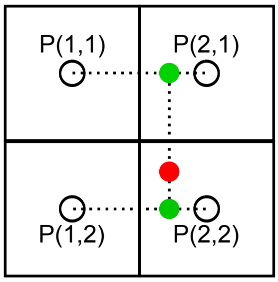


Figure 19: Bilinear Interpolation.

3.2.3 Multiscale Local Binary Patterns

Local binary patterns describes pixel intensity relations in a circular neighborhood from a given center. In general not all textural information can be captured with a small spatial support area. Also the LBP operator is not invariant in terms of texture scale. In endoscopic images the camera distance towards the mucosa varies depending on the camera position relating to the mucosa. In natural images pixels with close proximity have a higher correlation than pixels with a big distance to each other. Therefore a large radius leads to unreliable patterns as the pixels within the neighborhood are not necessarily related. Additionally the operator is not very robust against local changes in the texture caused by either distortions or a changing viewpoint. Hence an operator with a larger spatial support area is needed for describing mucosal textures effectively. To avoid unreliable patterns within the histograms and still be able to capture structures at different scales, the multiscale extension to the LBP operator was suggested in [34].

The straightforward approach towards multiscale is to combine multiple operators with different radii and a different number of neighbors. The most accurate information would be obtained using the joint distribution of these LBP patterns. In practice this approach is not applicable as the histograms would grow extremely sparse assuming an image of moderate size is used. Therefore the marginal distributions of each operator are considered although the patterns might not be entirely statistically independent. Nevertheless the sparse sampling is suboptimal when using a small number of neighbors. As a consequence low pass filters are used to gather information from an area of pixels. The size and positions of the filters are designed that the neighborhoods are covered as well as possible. After the low pass filtering each pixel represents a weighted sum of itself and it's neighboring pixels. The operator can then applied using larger radii without the problems introduced by sparse sampling.

Figure 20 demonstrates the effect of designing the filter sizes accordingly. The dashed circles represent the radius of the LBP operator. The solid circles indicate the area of pixels that are represented as weighted sum by the low pass filtering. These values are then treated as neighborhood pixels. The radii of the effective areas are calculated as

$$r_n = r_{n-1}\left(\frac{2}{1-\sin\left(\frac{\pi}{P_n}\right)} - 1\right), \qquad n \in \{2,\ldots,N\}, \tag{49}$$

where P_n is the number of neighborhood samples at scale n. For $P_n = 8$, low pass filtering only makes sense if the radius is larger than one. Therefore scale one (r_1) is used with a radius of 1.5. The radii are chosen so that the effective areas touch each other. Therefore the radius for the LBP operator

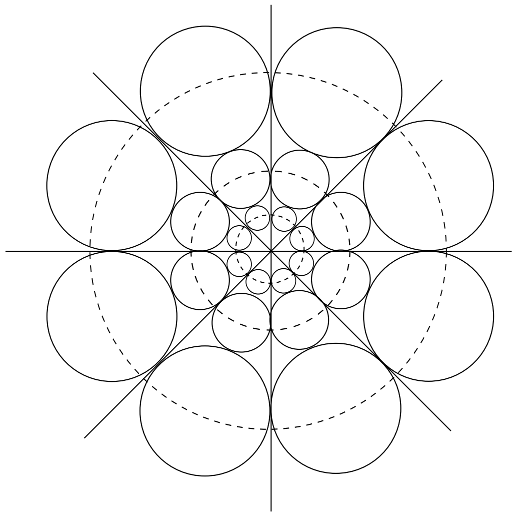


Figure 20: Pixel Neighborhoods with optimized Filter Size.

at scale n lies between r_n and r_{n-1}.

$$R_n = \frac{r_{n-1} + r_n}{2} \tag{50}$$

The size of the low pass filter is then calculated by

$$w_n = 2\left\lceil \frac{r_n - r_{n-1}}{2} \right\rceil + 1 \tag{51}$$

In the original publication regarding multiscale local binary patterns using a Gaussian low pass filter is suggested. Experiments based on the given data however showed that by using an averaging filter, superior classification rates could be reached. Therefore an averaging filter is used for the low pass filtering instead of a Gaussian filter.

3.2.4 Uniform Local Binary Patterns

The length of each histogram representing the distribution of the LBP patterns is dependent on the number of neighbors. In general 2^p histogram bins are needed to represent all possible patterns describing a p-neighborhood. Therefore the size of each histogram grows exponentially with the number of neighbors. This fact leads to problems in practice. Beside the amount of used memory, the computational complexity of calculating similarity measures and distances also grows exponentially. Another problem is, that an automated system supporting a physician during an endoscopy demands properties similar to a real time system. A step towards improving the classification rate and computational performance lies in selecting appropriate subsets of the histogram bins for classification.

Considering all 2^p histogram bins it is likely that not all patterns have the same importance in describing properties of a given texture. Even more, some patterns have preferable properties in terms of robustness against geometric distortions. Mäenpää et al. investigated this characteristics in [35]. An exhaustive search for finding the optimal subset of patterns would require N full classification runs with

$$N = \sum_{i=1}^{2^p} \frac{(2^p)!}{(2^p - i)!i!} \tag{52}$$

It is obvious that this is an impractical number even for small p. Mäenpää et al. suggest using an adopted Beam-Search algorithm for finding a good (although possibly suboptimal) subset of patterns. The algorithm incrementally builds possible pattern sets increasing the number of used sets up to a maximum number of elements. The best n patterns according to the classification error are then used to build a new pattern set. This is done up to a predefined number of iterations and results in an optimized set of patterns. In terms of feature selection algorithms this behaves like a feature subset optimization. A drawback lies within the possibly poor generalization of the classifier. Classifiers using a large number of parameters are likely to learn the training set without error. In theory a set of randomly distributed histograms with a sufficiently large number of bins and a rather small number of samples could be optimized using this algorithm to produce good classification rates. Although this is the extreme case the selection of histogram bins according to the classification error seems risky in terms of over fitting as long as the size of our image set is moderately small. This is a side effect of the huge number of possible subsets.

The subset of optimal patterns selected in accordance to the classification error might not be entirely related to certain textural properties in general. Nevertheless a subset of patterns exists that is promising for describing frequent textural properties independent of the given data. This subset is called the uniform local binary patterns. The measure of non-uniformity relates to the number of bitwise transitions of a given pattern in a circular bitwise representation. For example, the patterns 0 and 255 have a non-uniformity measure of 0 (00000000, 11111111). The patterns $1, 2, 4, 8, 16, 32, 64, 128$ have a non-uniformity measure of 2 (00000001, 00000010, ...). The idea is that patterns with a non-uniformity measure of 2 or less are more robust against geometric transformations like rotation and tilting. In total 58 (in a 256-bin histogram) different patterns with a non-uniformity measure less than 3 exist. This seems like a waste of information but this approximation is reported ([35]) to contribute most of the spatial patterns within textures.

3.2.5 Rotation Invariant Local Binary Patterns

Rotational invariance is a desired property for many texture classification applications. In the case of classifying textures captured during endoscopies the camera position in relation to the mucosa is unclear. Therefore rotational invariance might be a necessary property for the operators.

Binary patterns are used to represent the intensity differences of pixels within a certain neighborhood. If this neighborhood is rotated the resulting binary string is shifted in a circular manner by a certain displacement. Considering this fact, a modification adding rotational invariance to the LBP operators can be made. Considering that the rotation of a texture introduces a circular shift in the LBP string, this shift can be compensated by defining a standard displacement. Ojala et al. [43] suggest using the displacement that results in the smallest number if the LBP string is interpreted as a binary number. The displacement is found by continuous circular shifting until the smallest number that is represented by the binary (LBP) string is found. Let's call the circular right shift function of a binary string 'ROR'. The rotation invariant LBP operator is then defined as

$$LBP_{r,p}^{ror}(x, y) = min\{\mathrm{ROR}(LBP_{r,p}(x, y), i) : \qquad i = 0, \dots, p\} \qquad (53)$$

The modification made to the LBP can be seen as a quantization to the pattern occurrence statistics. Orientational information is lost as a cost for gaining rotational invariance. A subset of certain patterns in microstructures is represented by this operator. Nevertheless there are some problems inherent in this modification:

- **Angular Resolution**
  If an 8-neighborhood is used, only rotations of a multiple of 45° can be compensated exactly. A solution is to use a neighborhood with a finer angular resolution, a 16-neighborhood can represent rotations that are a multiple of 22.5°. Nevertheless, large neighborhoods create sparse histograms and a higher computational complexity. Therefore rotational invariance is optimally used in combination with Uniform Local Binary Patterns.

- **Discriminative Patterns**
  Not all patterns in the subset of the rotational invariant LBP operator sustain rotation in the digital domain equally well. Using all possible patterns therefore leads to suboptimal results. Again the Uniform Local Binary Patterns are suggested to be used in combination with rotational invariant LBP operators.

3.2.6 Local Ternary Patterns (LTP-Operator)

During the endoscopic procedure, the bowel is illuminated by a point source located at the tip of the used endoscope. The used camera has a fixed focus, hence some areas that are either too close or too far away from the position of the camera are blurred. The three dimensional property of the bowel also causes uneven illumination leading to noise within certain areas of the image. Therefore the pixel intensity relations of a pixel within a noisy or blurred region are rather unpredictable and contain almost no useful information. Tan et al. [51] suggest to use a threshold based LBP operator to compensate uneven illumination. We use this operator to ensure that pixel regions that are influenced by these kind of distortions do not contribute to the LBP histogram. This operator is also referenced as Peripheral Ternary Sign Correlation (PTESC) as used in [56]. The LTP operator uses a modified sign function to describe the intensity differences of the form

$$s(x) = \begin{cases} 1, & \text{if } x \geq T_h \\ 0, & \text{if } |x| < T_h \\ -1, & \text{if } x \leq -T_h \end{cases} . \tag{54}$$

The ternary decision leads to two separate histograms. One representing the distribution of the patterns resulting in a -1, the other representing the distribution of the patterns resulting in a 1.

$$H_{I,lower}(i) = \sum_{x,y} (LBP_{r,p}(x,y) = -i) \qquad i = 0, \cdots, 2^p - 1 \tag{55}$$

$$H_{I,upper}(i) = \sum_{x,y} (LBP_{r,p}(x,y) = i) \qquad i = 0, \cdots, 2^p - 1 \tag{56}$$

The neighbor information of pixels that lie within the threshold is encoded implicitly by this splitting. A problem is that not the joint distribution of lower and upper patterns is considered but the marginal distributions. An alternative is to encode the patterns as trinary numbers. Nevertheless this approach creates rather huge and therefore sparse histograms. This can result in instable results of the histogram measures. All tests show inferior results of this trinary encoding, therefore the experiments are conducted using the concatenation of both histograms.

3.2.7 Local Binary Patterns with Contrast Measure (LBP/C-Operator)

Texture can be seen as a combination of the texture patterns and the strength of this patterns. Textures can therefore be described by two orthogonal properties. The spatial structures (patterns) and the strength of these structures (contrast). As the LBP operators are invariant in terms of monotonic grayscale changes the strength of a pattern can not be represented. Therefore Ojala et al. [42] introduce the LBP/C operator to combine both properties. This operator is also used for segmentation. Due to the invariance regarding monotonic intensity variations the basic LBP operator can distinguish between background and foreground only by texture differences.

The contrast and the local binary patterns supplement each other in a very useful way. The LBP are sensitive to rotational changes but invariant to monotonic grayscale variations where the contrast measure is rotation invariant but sensitive to grayscale changes.

The rotation invariant local contrast measure for a pattern calculated at center (x, y) with a radius r considering p neighbors is calculated as

$$C_{r,p}(x, y) = \frac{1}{p} \sum_{k=1}^{p} (I_k - \mu_{r,p}(x, y))^2 \tag{57}$$

with

$$\mu_{r,p}(x, y) = \frac{1}{p} (\sum_{k=1}^{p} I_k). \tag{58}$$

The histogram is extended to two dimensions using the contrast measure as index in one dimension modeling the joint distribution of both random variables. Usually the contrast values (c) are quantized to reduce the numbers of indices into the histogram. The best number of intervals is unclear a priori. A small number leads to bad discrimination where a large number leads to sparse histograms.

$$H_I(i, c) = \sum_{x,y} (LBP_{r,p}(x, y) = i \wedge C_{r,p}(x, y) = c) \qquad i = 0, \cdots, 2^p - 1 \tag{59}$$

3.2.8 Local Binary Patterns with Intensity Measure (LBP/I-Operator)

In parallel to LBP/C, Wang et al. [46] use the LBP operator in combination with an intensity measure. Their results suggest that using the average intensity of the pixels within the kernel generates more stable results in the case of classification of endoscopic image classification compared to the plain patterns. The motivation is, that the LBP operator can not capture the intensity of a region as only intensity differences are modeled. Nevertheless the intensity values within an image might contain information that are beneficial for classification purposes (for example finding lesions within the mucosa). The used measure for intensity is

$$\mu_{r,p}(x,y) = \frac{1}{p}(\sum_{k=1}^{p} I_k) \tag{60}$$

In parallel to the LBP/C operator the intensity values are remapped to avoid sparse histograms.

$$H_I(i,j) = \sum_{x,y}(LBP_{r,p}(x,y) = i \wedge \mu_{r,p}(x,y) = j) \qquad i = 0, \cdots, 2^p - 1 \tag{61}$$

Opposed to the LBP/C operator the intensity measure is not invariant regarding different lighting conditions. This seems to be a considerable disadvantage compared to the contrast measure employed by the LBP/C operator. Therefore the LBP/I is not used in favor of the LBP/C operator throughout the experiments in this thesis.

3.2.9 Improved Local Binary Patterns (ILBP-Operator)

Jin et al. [26] propose a modified version of the LBP operator. They state that the LBP operator can miss the local structure in case of certain lighting variations. They model both, the local shape and texture of a region by extending the LBP operator. Instead of thresholding the neighboring pixels based on the value of the center pixel, all pixels within the kernel are thresholded with the mean over all pixels within the kernel.

$$\mu_{r,p} = \frac{1}{p+1}\left(\sum_{k=1}^{p} I_k\right) + I_c \tag{62}$$

$$ILBP_{r,p}(x,y) = \left(\sum_{k=0}^{p-1} 2^k\, s(I_k - \mu_{r,p})\right) + s(I_c - \mu_{r,p})2^p. \tag{63}$$

An additional effect of this extension is, that the center pixel is also used for building the binary pattern. Therefore the number of possible patterns is doubled. Because the contrast limited histogram equalization is performed within all experiments using considering local contextual regions of size 16×16 pixels, the contrast within the pixel neighborhoods is assumed to be adequate. Therefore the ILBP operator is not tested within the experiments of this thesis.

3.2.10 Extended Local Binary Patterns (ELBP-Operator)

A drawback of the LBP operator is that it can not describe the velocity of local variation. Only the first derivative information can be reflected. For this purpose a gradient based Local Binary Pattern operator is used. In fact, the operator itself is not modified but applied to a gradient intensity image instead of the original image. The name of this operator can be misleading as the naming conventions are not consistent within literature. We will stick to the naming in [24] and reference to the gradient based LBP operators as the Extended Local Binary Pattern (ELBP) operator.

The gradients image is calculated by approximating the derivatives in x and y by using a Sobel filter of the form

$$S_{horiz} = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \qquad S_{vert} = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix} \tag{64}$$

Please note that the y-gradient captures the horizontal edge information, where the x-gradient captures the vertical edge information. The original image is then convoluted with the corresponding filters. The convolution for a position x, y is defined as

$$I(x,y) \otimes S = \sum_{i=0}^{i=2} \sum_{j=0}^{j=2} S(i+1, j+1)\, I((x-1)+i, (y-1)+j) \tag{65}$$

Figure 21 shows how the y-gradient is computed by using the Sobel operator.

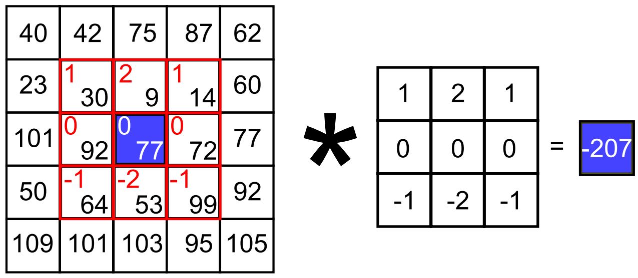


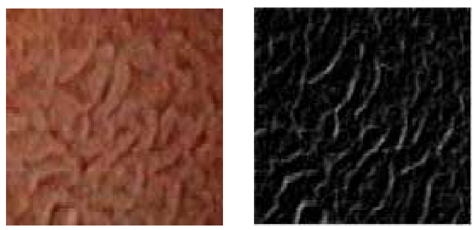
Figure 21: The Computation of a Convolution.

Figure 22 demonstrates the effects of Sobel-filtering an image. The image was created by calculating the mean of the gradient images that were created by convolution with the horizontal and vertical Sobel filter. This is referred to as the diagonal orientation throughout this thesis.

Huang et al. [24] compute the gradient magnitudes to generate the LBP histograms as

$$I_{mag} = \sqrt{(I \otimes S_{horiz})^2 + (I \otimes S_{vert})^2}. \tag{66}$$

In general however, mucosal images may have a dominant orientation (this might be related to the physician's style however). Hence a filter direction might be superior over the other. If one direction is dominant within the image, the calculation of the gradient magnitude might introduce an error. Therefore the gradient images are directly used for computing the LBP histograms.



(a) Original Images (b) Gradient Image

Figure 22: The Effect of Diagonal Sobel Filters.

3.2.11 Extended Local Ternary Patterns (ELTP-Operator)

The extended LTP operator is, in perfect analogy to the ELBP operator, based on gradient images which are produced by convoluting the original input image with Sobel kernels. In contrary to the ELBP the feature extraction is based on the LTP operator rather than the LBP operator. The differences between ELTP and ELBP are the thresholding that is introduced by the LTP operator as well as the computation of the histograms (see Section 3.2.6). Please note that using a threshold of zero is therefore not equivalent to the ELBP operator.

3.2.12 Wavelet based Local Binary Patterns (WTLBP-Operator)

By comparing the properties of each operator, we can identify two main tracks towards improving the discriminative power of the extracted information. The extended LBP (ELBP, ELTP) operators extract information from the high frequency components of a signal describing the speed of pixel intensity variation. Other operators use orthogonal features to improve the discriminative power of the extracted features (LBP/C and LBP/I). To be able to improve the description of textures at different scales and textures that are comprised of large structural elements, the LBP multiscale extension is used.

When considering the properties of the wavelet transform, one can see that there is a natural relation to the improvements suggested to the LBP operators:

- **Multiscale**
  The scaling function used within the wavelet transform leads to downscaling of the transformed signal. This corresponds to a decrease in resolution. When considering the LBP multiscale extension, pixel intensities are described as a weighted sum of the pixels within a neighborhood. As an averaging filter is used, this also corresponds to a decrease in resolution.
- **High Frequency Information**
  In Mallat's vertical and horizontal analysis, the decomposition algo-

rithm is based on two variables x and y leading to a priorization of each direction. The detail subbands contain high frequency information of the input signal. High frequency components in an image correspond to edge information. As the magnitude of each coefficient represent the strength of an edge we can interpret the detail subband coefficients as the speed of variation of pixel intensity differences. This is used within the operator based on using gradient filtering (ELBP and ELTP).

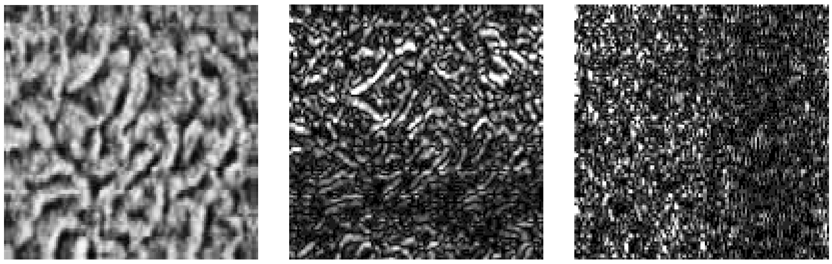
- **Orthogonal Features**
  The coefficients of the detail subbands represent the information that is lost due to the downscaling of the approximation subband. Therefore the information present in the detail subbands complements the information present within the approximation subband in a natural way.

Local Binary Patterns have been used in conjunction with the wavelet transform to describe textures in [33, 49]. Su et al. use Gabor wavelets in conjunction with the LBP operator in an Active Appearance Model. Liu et al. use non-separable wavelets to improve the discriminative power of the information that is extracted by the LBP operator. Although the combination of LBP and wavelet decomposition is used within these works, a combination of better suited operators (better suited than LBP) could improve the performance.

In this work a new wavelet based operator is constructed by combining suitable variants of the LBP operator. The properties of the specific operators and the wavelet decomposition is taken into account when constructing this new WTLBP operator. Both the approximation and detail subbands are used for feature extraction. By using all subbands different components of the textures can be optimally described.

- **Detail Subbands**
  The detail subbands contain high frequency components and are similar to the information that is represented by gradient images. In contrast to the ELBP and ELTP operators the detail subband coefficients contain the information that is lost within the approximation subband due to the downscaling process of the wavelet transform. Therefore no

information is lost overall. This is in contrast to the Sobel filtering. We are interested in the energy distribution of the coefficients, therefore the absolute values of the coefficients are used. Figure 23 shows the approximation subband as well as the absolute values of the coefficients of the horizontal and vertical detail subbands of a wavelet decomposed image.



(a) Approximation (b) Horizontal Details (c) Vertical Details

Figure 23: Coefficients of Wavelet Subbands.

As can be seen, due to using a discrete signal, the detail coefficients contain some amount of noise. To avoid introducing this noise to the computed histograms the LTP operator is used to extract features from the detail subbands. Applying the LTP operator is similar to the quantization of coefficients.

- **Approximation Subbands**
  The approximation subband represents the low frequency components of the image. By using dyadic sampling the bandwidth of the image is halved during each iteration. This is a problem, as we can not guarantee that the size of texture elements correspond to this sampling. It is possible to miss texture components by applying the basic LBP operator to the approximation subband coefficients. Therefore the LBP multiscale extension is used to extract features from the approximation subband. The maximum scale is set to three. As the LTP and LBP operator can not describe the strength of the patterns LBP/C is used to extract features from the approximation subbands.

By using a combination of operators based on the wavelet transform both the high and low frequency components of a signal can be analyzed by using the corresponding (optimal) operator. Even more, the high frequency components can be used at different scales without losing information due to the bandwidth filtering. The set of wavelet functions spans the differences between the spaces spanned by the various scales of the scaling function. Therefore by combining the high frequency and low frequency components of the signal no information is lost. This is in contrast to the gradient based operators (ELBP, ELTP), as only the high frequency information is used for feature extraction.

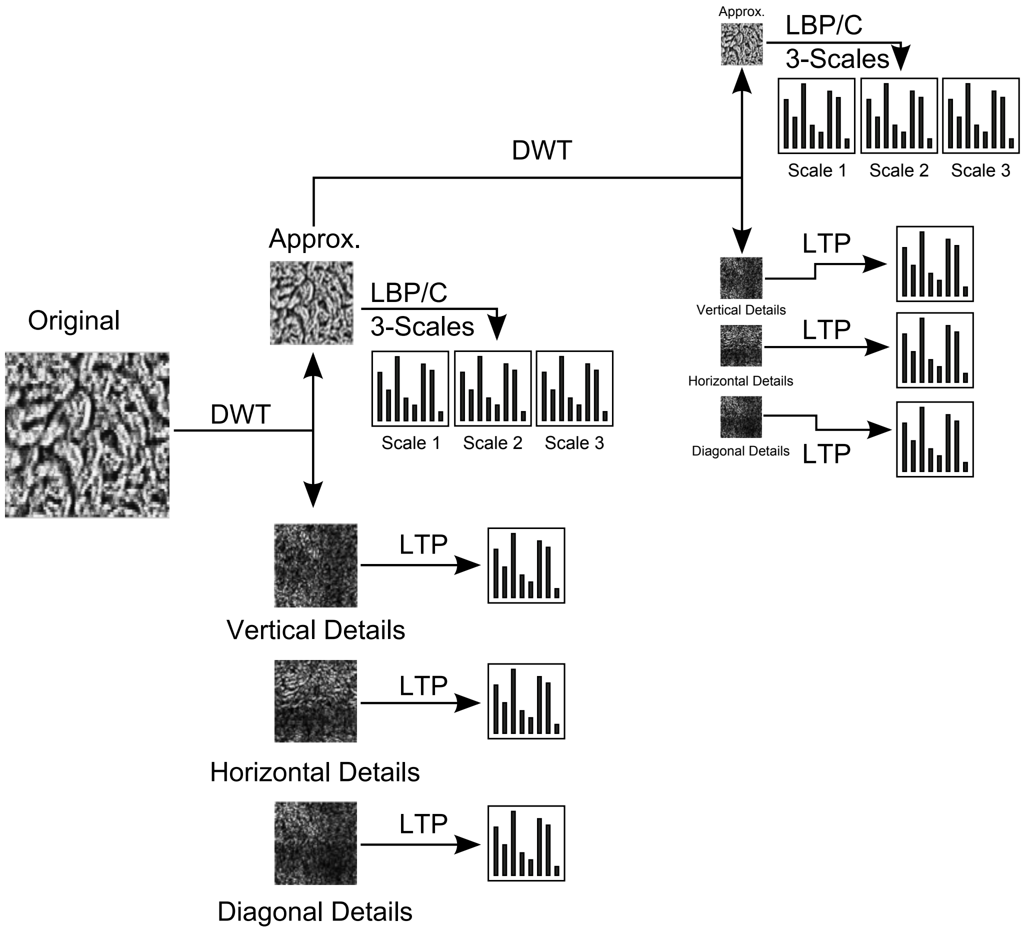


Figure 24: Two-Scale Wavelet Based LBP Operator (WTLBP-Operator).

The LBP/C operator is used to extract features from the approximation subbands. By using a maximum scale set to 3 and a minimum scale set to 1 the downscaling of the dyadic decomposition is supported. The LTP operator that is applied to the detail coefficients does not use the multiscale extension in order to avoid the low pass filtering of the high frequency information. The radius of the LTP that is used within the WTLBP is set to 1.5 pixels. The mother wavelet that is used is the biorthogonal Cohen-Daubechies-Feauveau (CDF) 9/7 analysis filter also used within Jpeg2000. Figure 24 shows a schematic of a two-scale wavelet based LBP (WTLBP) operator.

3.2.13 Combining Operators

Each presented operator involves some sort of tradeoff. For example the gradient based operators (LTP, ELTP) are not able to describe contrast. Therefore combining different operators seems promising. Two approaches towards the combination of operators exist.

- **Direct Approach**
  The direct approach involves incorporating the properties of two or more operators directly into the process of feature extraction. The LTP operator could be combined with the LBP/C operator by introducing thresholding to the LBP/C. Some combinations on the other hand are more problematic. A gradient based operator can not directly be combined with a non-gradient based operator for example.

- **Indirect Approach**
  The indirect approach combines operators by combining the computed histograms. The histograms are combined by simple concatenation. This results in a single large histogram. Although this approach does introduce redundancy all type operators can easily be combined.

Both approaches have their benefits and disadvantages. It is up to the given problem to decide which one is better suited. In case of the WTLBP operator the indirect approach was used. A direct combination of LTP and LBP/C seems to be a promising idea for further research though.

4 Classification

The classification process is the final step in the logical chain of automated classification. It uses the features that were extracted by the LBP operators as input data and assigns a class label to each sample. Classification methods can be divided into supervised and non-supervised. Supervised methods are based on a set of training data. This means that the input samples are labeled accordingly. The data is then used in combination with a training algorithm that tries to find an optimal decision function. In contrary, non-supervised classification methods do not use a training algorithm. Therefore they can be used if the data is unlabeled.

Both types of classification methods have their advantages and disadvantages. Non-supervised methods are closely related to density estimation. Therefore a large number of input samples is needed to guarantee stables results. This is not given in this project. On the other side supervised methods are based on a training algorithm. With a moderately small number of samples a lot of care has to be taken that this training process does not over fit the model towards the given input data. Additionally, special techniques for valid evaluation of the classification rates are needed.

In this work two supervised classification methods were used. A classifier based on Support Vector Machines as well as a k-Nearest Neighbor classifier. A variety of alternative options exist. The reason this methods are used are

- **Feature Vector Dimensionality**
  The LBP operators store the distribution of patterns within histograms. In most cases histograms of size 256- or 512-bins are used. It is obvious that this features can not be separated using a linear or quadratic classifier. The SVM classifier is designed to be able to handle high dimensional data and is therefore one option that is used. On the other hand the k-Nearest Neighbor classifier bases the classification decision on the proximity (in an adequate space) of similar samples. Using histogram based distance measures we can apply the k-NN method in a natural way.

- **Amount of available Data**
  Non-supervised classifiers a closely related to density estimation. So is

the popular statistical Bayes classifier. Nevertheless the rather limited number of input data limits the effectiveness of density estimation.

- **Availability of Ground Truth**
  Supervised classification methods are based on a predetermined ground truth. This ground truth is then used in the training process of the method. The images that build the training data were all extracted during a single endoscopic session. This session was followed by a histological examination of extracted tissue. This pathological diagnosis based on the extracted tissue build the ground truth for the used training data.

4.1 Support Vector Machines

One of the classification method that is used in this work is an optimal margin classifier. This method treats the extracted features not as histograms but as high dimensional vectors. The training process of the classifier tries to find a decision function that separates vectors belonging to distinct classes. In the case of p-dimensional features (in this case vectors) a $p-1$ dimensional hyperplane is constructed to linearly separate the samples. The classifier uses a training algorithm that optimizes the margin between training samples and class boundary. The resulting classification function is based on so called supporting samples or supporting vectors. The supporting vectors are those, that are closest to the decision boundary. The training algorithm for the maximum margin classifier was described by Boser et al. in [6].

The algorithm solves an optimization problem. Assuming the training data is of the form $(x_i, y_i), i = 1, \dots, l,\ y_i \in \{-1, 1\}, x_i \in \mathbb{R}^n$, with x_i the feature vectors and y_i the corresponding class label. If the data is linearly separable a hyperplane exists that separates all samples in this space. For points $x \in \mathbb{R}^n$ that lie on this hyperplane $wx + b = 0$ holds. In this case w is normal to the hyperplane and $|b|/\|w\|$ is the perpendicular distance from the origin to the hyperplane. All (x_i, y_i) then have the following properties

$$\begin{aligned} x_i w + b &\geq 1, & y_i &= +1 \\ x_i w + b &\leq -1, & y_i &= -1. \end{aligned} \tag{67}$$

This can be written more compactly as

$$y_i(x_i w + b) - 1 \quad \geq \quad 0, \qquad \forall i. \tag{68}$$

To optimize the margin we consider the points that lie either on the hyperplane $H_1 : x_i w + b = 1$ or $H_2 : x_i w + b = -1$. The euclidean distance of a vector x to a hyperplane (w, b) is

$$d((w, b), x) = \frac{xw + b}{\|w\|}. \tag{69}$$

Considering the assumptions made in Equation 67 we see that the margin is $\frac{2}{\|w\|}$ wide. Therefore the optimal margin is found by minimizing $\|w\|$. This results in a low generalization error as samples that are not included in the training set could be closer to the hyperplane than the support vectors. In this case the maximal margin optimization gives the highest probability of a correct classification.

Obviously most non-trivial classification problems are not linearly separable. Therefore Boser et al. [6] suggest a way to create non-linear classifiers. This approach maps the input data into a higher dimensional space until the input data is linearly separable. Nevertheless it is not very practical to use an infinite dimensional space. Therefore the so-called Kernel-Trick is used.

Let us assume a mapping of our data to some other Euclidean space H

$$\Phi : \mathbb{R}^n \to H. \tag{70}$$

The margin optimization would be dependent on the data through the dot product in H $(\Phi(x_i) \cdot \Phi(x_j))$. If a kernel function exists that satisfies $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ we can simply use K within the training algorithm, even though H might be an infinite dimensional space. The decision function in the dual space is then

$$D(x) = \sum_{k=1}^{p} \alpha_k K(x_k, x) + b. \tag{71}$$

The α_k are the parameters that are adjusted during the optimization and the x_k are the training samples. The function K is the predefined kernel function. Three popular kernels are

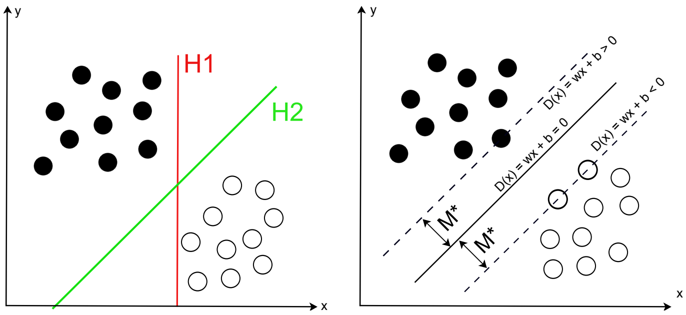
- Radial Basis Function Kernel
$$K(x_i, x_j) = e^{-\|x_i - x_j\|^2/2\sigma^2} \tag{72}$$
- Polynomial Kernel
$$K(x_i, x_j) = (x_i^T x_j)^d \tag{73}$$
- Linear Kernel
$$K(x_i, x_j) = (x_i^T x_j) \tag{74}$$

The kernel functions are parametrized in practice to adjust them according to the input data. In the case of two parameters, an exhaustive search for these parameters is not very efficient in terms of computational time. Therefore a method called grid-search can be used. This method arranges the parameters as a grid corresponding the the geometrical representation of the possible values and searches through all grid points at a certain scale. This is done by computing the classification rate using the parameters represented by the grid point. Depending on these rates, the optimal area on the grid is used to continue the search at a finer scale. This process is iterated until a certain predefined threshold is reached.

Figure 25 demonstrates separating hyperplanes in a two dimensional case. The left image demonstrates that possibly more than one (but not optimal) hyperplane exists that linearly separates the data (the red line). The figure on the right demonstrates an optimal hyperplane maximizing the margin between support vectors and decision boundary.

4.1.1 Multi Class Support Vector Machines

Due to various difficulties, using a single SVM formulation for problems with more than two classes is usually avoided. In general, several binary SVM classifiers are combined to perform the classification of a multi class problem. Popular methods are



(a) Separating Hyperplanes (b) Optimal Hyperplane

Figure 25: Linear Decision Functions

- **Winner-takes-all Strategy**
  This approach constructs M binary classifiers (for an M-class problem). The i-th classifier assigns all samples of class i to class one and all other samples to class two. The winner-takes-all strategy classifies a sample to the class which receives the highest value among the binary classifiers. This type of SVM classifier is used for the multi class problems within this work.

- **Max-wins-voting Strategy**
  The max-wins-voting method creates a binary classifier for every pair of distinct classes. Each binary classifier (C_{ij}) is trained assigning the samples of class i to class one and the samples of class j to class two. The outcome of each classifier counts as a vote. A sample x is classified using the class with the highest number of votes.

- **Error-correcting Codes**
  This method constructs M binary classifiers. Each class is assigned a codeword. A vector containing the class probabilities for each class is constructed. The probabilities are then viewed as the probabilities that its corresponding bit in the codeword is set. The decision is based on the smallest distance of the assigned codeword and the constructed vector.

The error-correcting codes approach is described in [15]. A comparative study of multi class Support Vector Machines can be found in [16].

4.2 The k-Nearest Neighbor Classifier

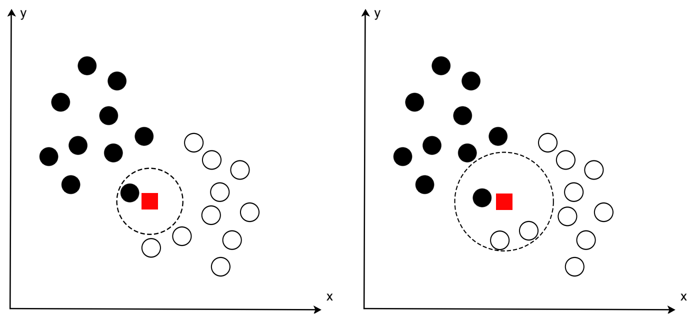
The nearest neighbor classification method (Cover and Hart [11]) is based on the observation that half of the classification information in an infinite sample set is contained in the nearest neighbor. The decision of classifying a sample x_i as class y_i is based on a collection of n-correctly classified samples $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. The assumption is that the (x_i, y_i) are independently identically distributed according to the joint distribution of samples x and class labels y. In other words, it is reasonable to assume that samples with close proximity will belong to the same class.

There are some problems with the basic assumption.

- **Outliers**
  In a moderately large natural data set, samples exist that do not resemble their class members. This can be a cause of distortions or bad image quality. As a result, the position of these samples within space can not reflect their class membership. When applying classifications with low k-values outliers can lead to misclassification.
- **Twins**
  Due to the limited size of endoscopic imagery sometimes multiple patches of a patient are extracted and used for classification. These patches tend to be similar in terms of perspective and illumination (they were taken during a short distance of time). These twin images can fool the classification process and (although correctly) lead to classification rates that are above the rates that are to be expected in a real world experiment.

The k-Nearest Neighbor Classifier is parametrized by the number of considered neighboring samples. This value has an influence on the classification rates. Figure 26 demonstrates the effect of choosing the k-value. On the left, a k-value of one was chosen. The right image shows that a k-value of three changes the result.

The voting strategy employed by this classification method can be ambiguous at certain situations. Depending on the number of used classes within the classification problem and the chosen k-value a tied vote can occur. In this



(a) 1-NN decision Radius (b) 3-NN decision Radius

Figure 26: Impact of k-value on the Classification Result

case, the prior probability of a sample belonging to a certain class is used for classification. The class frequencies are used as prior probabilities.

4.2.1 Histogram Similarity Measures

The assumption that the k-Nearest Neighbor method is based on implies that a meaningful distance measure exists for the given samples. The standard measure used in various classification scenarios is the Euclidean distance. Nevertheless a geometric interpretation of the LBP distributions might not be optimal. Therefore three different metrics for measuring the similarities between two probability distributions, which seems more appropriate for this scenario, are used.

For two histograms (H_1, H_2) with N bins and bin number i being referenced to as $H(i)$, the similarity measures are defined as

- Discrete Bhattacharyya Distance ([5])

$$B(H_1, H_2) = \sqrt{1 - \sum_{i=1}^{N} \sqrt{H_1(i)H_2(i)}} \tag{75}$$

- Chi-square Metric ([44])

$$\chi^2(H_1, H_2) = \sum_{i=1}^{N} \frac{(H_1(i) - H_2(i))^2}{H_1(i) + H_2(i)} \tag{76}$$

- Histogram Intersection Metric ([50])

$$H(H_1, H_2) = \sum_{i=1}^{N} \min(H_1(i), H_2(i)) \tag{77}$$

It should be noted that these measures are no distance measures in the natural sense. All measures map to $[0; 1]$. The discrete Bhattacharyya as well as the Chi-square metric return zero on a perfect match where the Histogram Intersection returns one at a perfect match. Taking this into account the classifier's decision is based on a similarity rather than a distance.

4.3 Cross Validation Techniques

Cross validation is used to predict the accuracy of the general prediction of the classification method. As a subset of samples is used as the training set, the classification of a sample within this very set would produce inaccurate results. In the case of a 1-NN classification the result would be always correct in the case no other duplicates exist (the closest neighbor is the sample itself). The training algorithm in the Support Vector classifier optimizes the margins for the given training set. Therefore each sample within the set can optimally be classified. To avoid this problem cross validation techniques are used.

In this work all classification rates are computed by applying the leave-one-out cross validation technique. In this approach a sample (x_i, y_i) is extracted from the set of all samples. A new set is constructed that is constituted of all samples except the extracted sample. This set is commonly called the training set and is used as input for the training algorithms of the classification methods. The estimated decision function based on this training set then classifies the left-out sample. This process is iterated for all samples within the data set.

For a data set with n samples, the leave-one-out cross validation protocol consists of n leave-one-out cross validations (each performing $n - 1$ single classifications). A generalization of the leave-one-out cross validation technique uses disjoint subsets of samples. This generalization is called k-fold cross validation. The classification uses $k - 1$ subsets as training input and classifies samples from the left out subset. This process is repeated k times. This generalization results in a lower accuracy of the estimation of the classification methods performance, but can be a good trade off for computational complexity. The leave-one-out cross validation is a special case of the k-fold cross validation (for k samples). More information about cross validation techniques can be found in [29, 7, 13].

4.4 Classification of LBP-based Histograms

The classification of the computed subset of histograms is performed in a natural way corresponding to the used classification method. The selected subset of histograms is combined to a single large histogram by using simple concatenation. This histogram is then treated corresponding to the used classification method. In case of the k-NN method, a histogram similarity metric (as listed in Section 4.2.1) is used to calculate the distances used within the classification. This is performed on the concatenated large histogram. The SVM method interprets the concatenated histogram as a high dimensional vector and uses a kernel that is based on a radial basis function to perform the classification.

5 Experiments

The various operators as well as the classification methods are defined without ambiguity. Many details can influence the performance in terms of classification rate in an experiment however. In order to be able to compare the achieved results, a common experimental environment for all experiments has to be defined. Additionally to the well defined environment, all methods that were described are influenced by numerous unknown parameters. These parameters include operator specific parameters, parameters that influence the specific classification methods as well as parameters that are used to modify the input signal. The optimal combination of these parameters (towards optimal classification rates) is unclear and often involves handling tradeoffs. The identification of an optimal parameter combination can generally not be achieved by an exhaustive search. The overwhelmingly large number of possibilities renders this approach inapplicable in practice. Therefore a systematic approach has to be applied. The experimental methodology used to identify the basic parameters together with the more specific parameters that were later used throughout the experiments are described in this Section.

5.1 Experimental Environment

The experiments that were conducted in this thesis were applied using a clearly defined environment. The compilation of test data that was used throughout the experiments is described in Section 2. Table 3 lists the distribution of images among their pathologically determined Marsh class. In the experiments based on the two-class scheme for classification, the images of type Marsh-3a, Marsh-3b and Marsh-3c are combined as the 'celiac' class and the images of type Marsh-0 are used as the 'non-celiac' class.

	Marsh-0	Marsh-3a	Marsh-3b	Marsh-3c	Total
Bulbus Duodeni	165	47	54	23	289
Pars Descendens	151	47	60	76	334

Table 3: Distribution of the Image Data

The software package that was used for the development and execution of self written code is Matlab 2009b [1] using the Image Processing Toobox version 7.0 and the Wavelet Toolbox version 4.5. The third-party Matlab Toolbox for Pattern Recognition (PRTools version 4.1.4 [2]) was used in experiments with the SVM method to perform the classification (the SVM classifier was provided by PRTools).

The applied methods were implemented in the Matlab language. Performance critical components were additionally implemented using the C language in order to optimize the computational performance. The MEX interface to Matlab was used to combine the Matlab and C code. The GCC suite version 4.3.2 [3] was used to compile the C code. The host system for computing the experimental results was Linux version 2.6.26 [4].

5.2 Experimental Setup

In the context of this thesis the term experiment is used to describe a specific test setup that is used to compute a classification result. An experiment consists of three steps:

- **Image Acquisition and Preparation**
  This step was performed as described in Section 2. The image sets that were constructed build the basis for the evaluation of the feature extraction and classification methods. The images are preprocessed using CLAHE within each experiment.
- **Feature Extraction**
  The feature extraction process uses the LBP-based operators that were described in Section 3.
- **Classification**
  The final step involves the actual classification. This is performed by using cross validation (leave-one-out). Additionally to the classification

[1] www.mathworks.com
[2] www.prtools.org
[3] gcc.gnu.org
[4] www.linux.org

a feature subset selection is performed using a sequential forward subset selection algorithm. The details can be found in Section 4.

All three steps within an experiment are based on several parameters. The combination of these specific parameters defines an experiment. These parameters comprise:

- Classification Scheme
- Image Set
- Image Type
- Feature Extraction Method
- Classification Method

Additionally to the parameters specific for a single experiment, the steps involving feature extraction and classification are also dependent on a set of parameters. Each LBP-based operator that is used for feature extraction is parametrized by a number of a priori unknown values. So are the classification methods. Experiments are conducted for all combinations of classification schemes, image sets and image types.

The identification of optimal parameters regarding the feature extraction and classification methods, are the main concern towards establishing reliable classification rates. Considering the needed number of experiments to cover all possible combinations of feature extraction methods, classification methods as well as image sets, 96 single experiments have to be conducted. As the feature extraction and classification method specific parameters are unknown it is obvious that optimizing these unknown parameters are impossible to compute in a reasonable amount of time based on the full set of possible experiments. Therefore a systematic approach is needed.

By separating the feature extraction and classification method based parameters according to several categories the number of experiments can be reduced while still maintaining a high level of confidence with respect to the found parametric values. The separation is based on the assumption that parameters within a category have a higher dependency among each other than parameters within different categories.

According to this parameter dependency property, the following categories are defined:

- **Basic Parameters**, are parameters that are directly related to all operators. This includes the used number of scales (in the context of the multiscale extension), the number of considered neighboring pixels as well as histogram based extensions such as the rotational invariance and the uniform patterns.
- **Classification Method Specific Parameters**, are related to the classification methods such as the used distance metrics, kernel parameters (SVM) and the k-value (k-NN).
- **Operator Specific Parameters**, include parameters that are specific for a single operator. These are the values used for thresholding (LTP based operators) and the used values for quantization (LBP/C based operators).

Considering the assumption that parameters within a category have a higher dependency among each other while parameters within different categories have only a small impact on each other, the parameters within a category can be optimized while neglecting parameters of other categories. This reduces the number of possible combinations. In general the deciding measure of the optimization process is the overall classification rate. For some parameters a theoretical analysis of possible options suffice while other parameters have to be evaluated by experimental results.

5.3 Identifying the Basic Parameters

The basic parameters influence the feature extraction process. All applied feature extraction methods are a modified form of the LBP operator, therefore it is reasonable to use a common basis set of parameters for all operators. The histogram based extensions do not alter the extraction process itself. Therefore it is safe to assume that they provide stable results among operators. The optimal basic parameters are evaluated by using a theoretical analysis accompanied by a series of tests using a subset of all operators. The found parameters are later used throughout all experiments. The following list presents the basic parameter settings that are evaluated.

- **Number of Operator Scales**
  The multiscale extension to the LBP operators adds scale invariance to some degree. A priori it is not known what is the best number of scales to use (in terms of computational efficiency and classification rate).

- **Number of Pixel Neighbors**
  The LBP operators are able to work with a variable number of neighbors. A small number of neighbors leads to coarse angular resolution introducing problems with rotational invariance. A larger number of neighbors improves the angular resolution but leads to huge, probably sparse histograms. This creates problems with both, the histogram measures and the computational efficiency (in terms of memory consumption and speed). The combination of scale and number of neighbors is important.

- **Rotational Invariance**
  Rotational invariance is a beneficial property for many texture classification applications. Nevertheless some discriminative power is lost. Due to the nature of the LBP, the rotation invariance is related to the number of pixel neighbors in terms of angular resolution.

- **Uniform Patterns**
  Using uniform patterns is a method to reduce the size of each histogram. A subset of prominent codes that describe the majority of patterns within a texture are selected and used for classification. Nevertheless discriminative information might get lost.

5.3.1 Number of Operator Scales

The multiscale extension to the LBP helps to describe pixel relations at different scales. At each single scale one histogram is computed. Multiple scales can be combined by concatenating the respective histograms. As the histograms are concatenated, histograms that were computed using the same LBP-scale are compared during the classification. In order to provide scale invariance, the histograms would have to be aligned according to the specific scale of the texture not the used scale of the operator. This is a point that is open for improvements in future works and is not considered in this thesis.

Textures can be a combination of small scale and large scale patterns. The optimal combination of scales within an experiment is highly dependent on the specific classification problem. Therefore it does not make sense to predefine a fixed subset of scales used within all experiments. The multiscale extension does have some limitations. Small scales do not allow to capture large textures. Larger scales help to capture these textures but there is an upper limit in scale that is reasonable to be used. By analyzing how the multiscale extension affects the calculation of the patterns two bounds can be determined. The discussed calculations are based on the dimensions of the images that were used throughout this thesis (128 $\times$ 128 pixels).

- **Lower Bound**
  Scale 1 defines the lower bound. The theoretical maximum percentage of considered pixels is 96.8 percent (in the case of using a 1-pixel radius. This is caused by using a margin at the border of the images to avoid crossing the image boundaries. For efficiency purposes integer margins are used. The minimum radius of the operators is defined as 1.5 (this corresponds to scale 1). The maximum percentage of considered pixels (using scale 1) is 93.8 percent.

- **Upper Bound**
  The multiscale extension represents pixels as a weighted sum of itself and it's neighbors. This is accomplished by low pass filtering and is used to represent large structures within textures. For larger scales, the required margin along the image borders grow. This is a side effect of the growing radius of the operators. Therefore the number of pixels, that can be considered, decreases. At a scale of 5 only 31.6 percent of the pixels can be used as center for calculating the LBP. Although the pixels within the margin are implicitly used as they are part of the weighted sum of the neighboring pixels the number of patterns decreases and the histograms become sparse. At scale 6 only 7 percent

of all pixels can be used as centers (this is only 1156 pixels considering our image test set). Therefore an upper bound is specified as scale 5.

Figure 27 shows the effect of growing scales in relation to the number of usable center pixels due to the needed margin. As the optimal scale combination is highly dependent on the given texture classification problem all scales within our designated range (1 to 5) are used to compute the histograms.

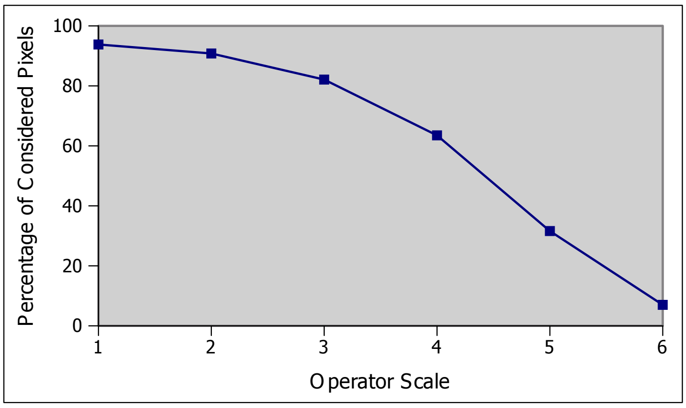


Figure 27: Impact of Operator Scales on the Number of Usable Pixels.

5.3.2 Number of Pixel Neighbors

The LBP operators describe the pixel intensity differences taking a number of neighboring samples into account. The number of used neighbors is variable. A large number produces longer binary patterns and might therefore contain more information. Also by using a large number of neighbors a finer angular resolution is provided. In the case of rotational invariance this is desirable. On the other hand using large numbers of neighbors creates numerous problems.

Using large numbers of neighboring samples only makes sense at a large scale. In general however, most of the descriptive patterns in textures are at small scales. When using a large number of neighbors in combination with a small scale, a significant number of values have to be interpolated. This is caused by the fact that the theoretical size of a neighbor (in the context of the LBP operator) becomes smaller than a pixel. Therefore no additional information can be gained for small samples.

Another problem is that large images have to be used to guarantee dense histograms. The number of histogram bins is directly related to the number of considered neighbors ($2^n - 1$ for n neighbors). Therefore a high number of neighbors introduces large histograms (which are probably sparsely filled). It is usually not the case that large image regions can be used in the context of endoscopic imaging. The reason are the frequent distortions within the images, rendering the visible features useless. This was discussed in more detail in Section 2.6.1.

Finally an efficient implementation of the classification methods (in terms of computational speed) demands that a set of training data, or a set of support vectors is in memory during classification. Assuming 16 neighboring samples a histogram takes up to 65535 bytes of memory. For a large number of used histograms during classification (taking into consideration various scales, color channels, wavelet subbands, ...) the resulting memory consumption can be critical.

Overall the advantages of using a large number of neighboring samples does not compensate for the created problems. Therefore all experiment in this work are based on operators using 8 neighbors.

5.3.3 Rotational Invariance

Rotational invariance can be achieved by a circular shift of the binary patterns by a corresponding offset. Due to the decision that all operators use a fixed number of 8 neighbors, a shift of a single pattern position corresponds to a rotation of 45 degrees. The angular resolution is highly limited. It is unlikely that endoscopic images differ in terms of rotational alignment of a multiple of 45 degrees. Also the rotational invariance leads to a reduce number of usable patterns as all information regarding the rotational alignment is lost. Therefore the rotational invariance extension to the LBP histograms is not used throughout the experiments in this thesis.

5.3.4 Uniform Patterns

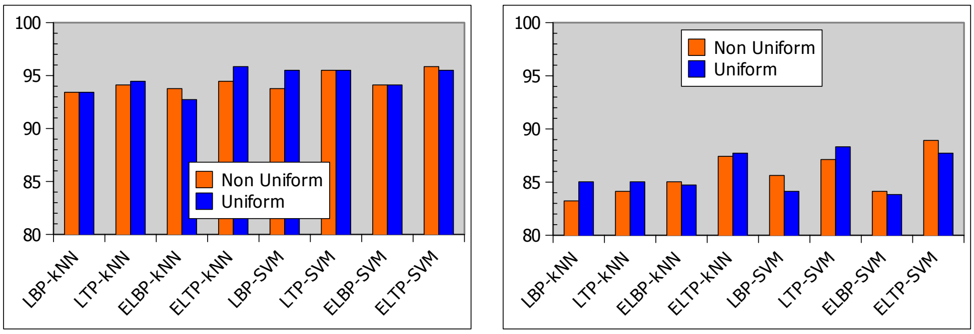
Considering only patterns with a non-uniformity measure lower than 3 reduces the size of the histograms. This subset of patterns is expected to behave well under geometric distortions and is called uniform patterns. Although it is assumed that the uniform patterns contribute to most of the patterns in textures, information is lost due to the reduction of histogram bins. Theoretical conclusions do not suffice to decide how well the uniform patterns fit the present classification problems. Therefore a number of experiments were conducted to identify if the uniform patterns are feasible for the classification of celiac disease.

In the experiments four, of the LBP operators are applied based on the two-class scheme for classification. The chosen methods for feature extraction are the LBP, ELBP, LTP and ELTP operators. This set of operators covers all operator specific modifications. Additionally the gradient filtering is considered. To be able to support a reliable decision only the two-class problem was considered. In case of the four-class scheme, the basic classification rates are to low to reliably support a decision. Figure 28 shows the overall classification rates of the experiments.

As can be seen, the classification rates of experiments based on using uniform patterns reach results comparable to the non-uniform based experiments. In fact the uniform patterns have a slight advantage in classification rate over the non-uniform patterns for the majority of operators. Keeping in mind that the image data set was built under idealistic conditions it is likely that the overall degree of geometric distortions within the images will be higher in a real world system. Based on the beneficial properties of the uniform patterns (smaller histograms, more robust in terms of geometrical changes), this extension is used throughout all the experiments in this thesis.

5.4 Identifying the Operator Specific Parameters

The operator specific parameters are highly dependent on the present input data. These parameters include the specific threshold values as well as the number of used contrast bins. No general assumption can be made about the optimal values for these parameters. Therefore they are optimized within



(a) Duodenal Bulb

(b) Pars Descendens

Figure 28: Impact of using Uniform Patterns on the Classification Rate.

each single experiment. Nevertheless we can again define certain bounds that help to reduce the number of tested values.

5.4.1 Threshold Parameters

When optimizing a threshold value the following considerations have to be taken into account.

- **Image Properties**
  First of all, the image properties have to be considered. When calculating the standard pixel deviations within a 3×3 window (which corresponds to an LBP neighborhood at scale 1), one can observe that the deviation is rather small (4.1). By applying preprocessing (CLAHE) the contrast is enhanced. Therefore the mean intensity deviation increases from 4.1 to 37.9.

- **Operator Scale**
  When applying a threshold, the multiscale extension to the LBP operator must not be neglected. The multiscale extension uses low-pass filtering. Therefore the mean pixel intensity deviation decreases. Considering this fact, it is not optimal to use a single threshold for all operator scales. Finding the optimal threshold for each scale increases the algorithms complexity from $O(N)$ to $O(N^s)$, where N is the number of considered thresholds and s is the number of used scales. A more

efficient option of optimizing the threshold is to apply a scaling function to the threshold according to the decrease in pixel intensity deviation at each scale. Nevertheless a large number of input data is needed to optimally estimate this scaling function. The experiments based on using operators with thresholds (LTP, ELTP) use fixed thresholds among scales in this thesis. This is not optimal and might be subject to further improvements.

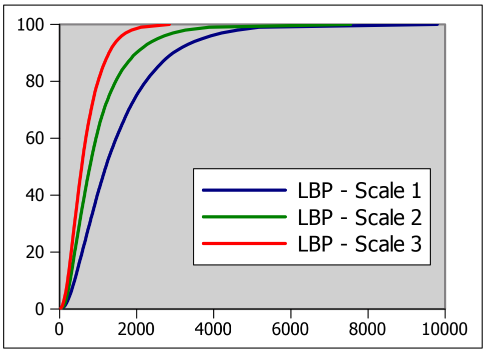
Throughout all the experiments the threshold parameter was optimized in a range from 0 to 30. Using a threshold of 0 does not exactly correspond to the LBP operator though. The histograms are still split between positive and negative pixel intensity differences as explained in Section 3.2.6.

5.4.2 Contrast Quantization Parameters

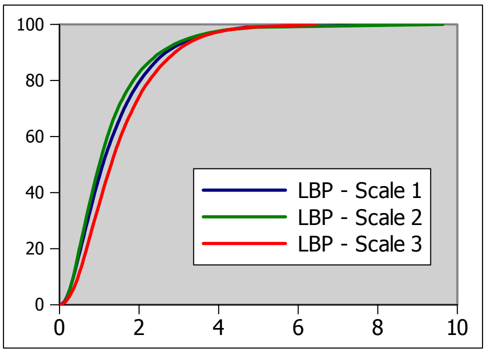
The LBP/C operator (Section 3.2.7) uses a contrast measure to combine two orthogonal properties in order to improve the discriminative power of the extracted features. The contrast measure has to be quantized to avoid sparse histograms. The set of possible contrast values ranges from 0 to 16265.25 (the highest value results from a set of the neighboring pixels with half of the pixels having the minimum value (0) and half of the pixels with the maximum value (255)). Obviously it is highly unlikely to find a pixel neighborhood with these properties in a natural image. The distribution of contrast values is far from being uniform. Instead the contrast values are assumed to be exponentially distributed. Therefore a linear mapping of the contrast value to the corresponding bin index is inadequate as it would result in a number of very sparse histograms.

The properties of the multiscale extension have to be considered. The larger the scale, the lower the variation of contrasts. This is again a result of the low-pass filtering. Figure 29 presents the cumulative distribution of the contrast values at specific operator scales. To avoid problems due to the low-pass filtering the contrast values are normalized through division by the standard deviation of the contrast values.

To find a mapping that results in equally dense histograms the image sets were used to estimate the distribution of contrast values. According to these



(a) Deviation by Scale



(b) Normalized

Figure 29: Cumulative Distributions of the Contrast Measures

values the binning was performed. The number of contrast bins is optimized within each single experiment in a range from 1 to 20.

5.5 Identifying the Classification Method Specific Parameters

The classification specific parameters influence the classification process directly. Some of the parameters are highly dependent on the input data such as the k-value and the subset of selected features. These parameters are optimized during each experiment. On the other hand, parameters such as the used distance metric and kernel are more stable. These parameters are found by running a series of tests and are later used throughout all experiments. Additionally not all histograms that are computed contain the same amount of discriminative information. Therefore a subset of features which is locally optimal is selected by using a feature subset selection algorithm.

- **Classification Method**
  The methods used for classification are not related to each other. Therefore both methods are applied in experiments for all operators. The parameters specific to the classification methods (for example the k-value in the k-NN classifier) are optimized during each experiment.

- **Histogram Distance Metric**
  The histogram distance metrics are relevant for the k-NN classification method only. The best measure was evaluated by a series of experiments (see Section 5.5.2).

5.5.1 Parameters of the SVM Classification Method

As discussed in Section 4 the kernel functions used by the SVM classification method can be parametrized. In this work a radial basis function kernel is used with the parametrization (p)

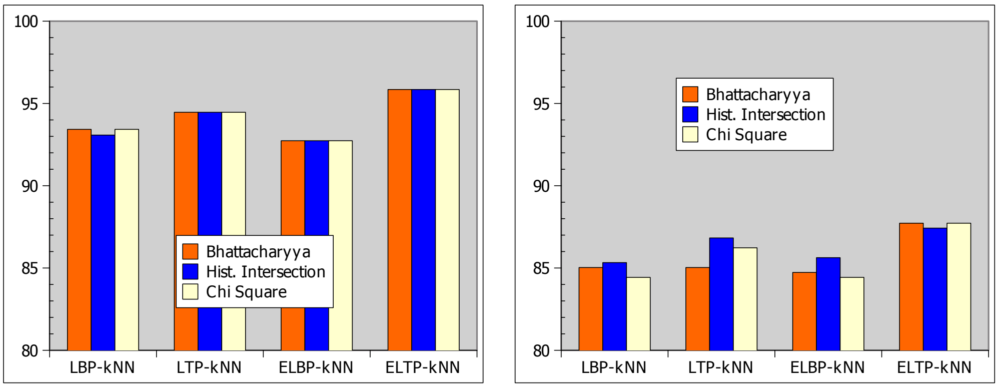
$$K(x_i, x_j, p) = e^{\frac{-\|x_i - x_j\|^2}{p^2}} \tag{78}$$

This parameter was found by a linear search in a range from 0.1 to 1 in steps of 0.1. Throughout all the experiments $p = 0.1$ is used.

5.5.2 Parameters of the k-NN Classification Method

The k-NN classification method is parametrized by two variables. The number of considered neighbors (k-value) and the used distance measure. In contrast to the SVM classifier it is cheap (in terms of computational speed) to optimized the k-value during each classification run. The reason is, that the entire test set can be kept in memory with the corresponding distances to each sample. The distances have to be calculated only once during the entire classification. Therefore the k-value is optimized during each classification. On the other hand, the used distance measure influences the classification outcomes. Opposed to the k-value this can not efficiently be optimized during the classifications. For this reason the optimal distance measure is evaluated by a series of tests.

Figure 30 demonstrates the impact of using different histogram similarity measures. As we can see the results vary by only a small margin. The histogram intersection metric provides a better classification performance for the Pars image set by a mean of 0.5 percentage points. The metrics only differ by a mean of 0.1 percentage points for images from the Bulbus



(a) Images from the Bulbus (b) Images from the Pars

Figure 30: Comparison of Histogram Similarity Measures

image set. Based on this observations the histogram intersection is used as similarity measure for the k-Nearest Neighbor classifier.

5.5.3 Feature Subset Selection

The operators compute a histogram for an image at each single scale, subband and color-channel. It is unknown which combination of histograms provides the most discriminative power for a given texture or classification problem in general. A naive way to find the optimal combination of these histograms is using an exhaustive search. This approach computes the classification rates for all possible combinations of histograms. For n histograms, this results in a total number of classification runs of $\binom{n}{m}$ for a fixed value of m. The optimal number of features used for the classification (m) is also unknown and has to be optimized as well resulting in $2^n - 1$ possible combinations (the empty set is certainly not the best). Considering the basic LBP operator three color subbands and five scales are used to compute a total of 15 histograms. This is the operator with the smallest number of histograms. A total of 32767 combinations have to be tested. Assuming a speed of a second per classification (which is optimistic), the entire experiment takes over nine hours to finish.

A way to speed up the search is to put constraints on the combinations of histograms. For example constraining that a combination of subbands uses the same set of histograms cuts the number down to $(2^{cs}-1)(2^{sb}-1)(2^{sc}-1)$ for cs color subbands with sb subbands and sc scales.

Although this constraint considerably limits the time needed to perform an experiment, all threshold based operators have to be run multiple times to find the optimal threshold value. The same holds for the LBP/C operator to find the optimal number of contrast bins. A fast classifier such as the k-Nearest Neighbor method is feasible under this conditions and can run such a test within a day (with the currently available hardware resources). Nevertheless the Support Vector Machines classifier is also considered. This method requires over two minutes for a single leave-one-out cross validation. Therefore a feature subset selection algorithm is used to find the possibly best combination of histograms while maintaining a reasonable computational speed. To be able to compare the results between both classification methods, the feature subset selection algorithm is also applied when using the k-NN classifier.

A feature subset selection algorithm has two main purposes:

- **Reduce the Feature Dimensionality**
  The computational speed of the classification method is improved by using a smaller number of features.
- **Identify an Optimal Subset of Features**
  For a given classification problem the optimal combination of discriminative features is usually unknown. Suboptimal features are likely to interfere with the classification. Therefore a subset of features is selected that is locally optimal.

Feature subset selection algorithms generally involve a search algorithm and a criterion function. The search algorithm finds and compares several subsets of features by applying the criterion function as a measure of effectiveness. In our case the criterion function is the overall classification rate of a cross validation using the selected feature subset. The algorithm that is applied is called sequential forward subset selection algorithm (SFS [25]). This algorithm is a greedy forward selection algorithm. The search function simply combines the currently (locally) optimal set of histogram with all remaining histograms. The histogram that improves the classification rate most is added to the local set of optimal histograms.

For the set of all histograms $H = \{h_i | i = 1, \ldots, n\}$ and a function C giving the overall classification rate, the algorithm applies the following steps

1. Start with the empty optimal set:
   $\mathbb{O} = \emptyset$
2. Select the best histogram:
   $\underset{i, h_i \notin \mathbb{O}}{\text{argmax}}\{C(\mathbb{O} \cup \{h_i\})\}$
3. Add the histogram to the optimal set:
   if $(C(\mathbb{O} \cup \{h_i\}) > C(\mathbb{O}))$ then: $\mathbb{O} = \mathbb{O} \cup \{h_i\}$
   else: go to step 5.
4. If the number of iterations is beyond the predefined threshold, go to step 2.
5. $\mathbb{O}$ is the selected subset of histograms.

A predefined threshold for the maximal number of histograms of 10 is used as abortion criterion for the algorithm. Using the SFS algorithm reduces the computational costs while delivering very good results. In the case of SVM a small adaption is applied. Instead of considering the classification rate of a leave-one-out cross validation, a ten-fold cross validation is applied to compute the classification rate. To ensure that the random partitioning does not influence the decision negatively the cross validation is repeated twice. The mean classification rate of both repetitions is then used as the criterion. This adaption was done to further improve the computational speed of the experiments based on the SVM classification method.

5.6 Summary of Identified Parameter Setting

The parameter setting that was used to compute all the results presented in Section 6 are as follows:

- **Basic Parameters**
  The number of used scales in the context of the multiscale LBP-extension

is 5, with a minimum of 1 and a maximum of 5. The number of considered neighbor pixels is 8. All experiments are based on uniform patterns. The rotational invariance extension is not used.

- **Operator Specific Parameters**
  All operator specific parameters are optimized during each single experiment. The range of the threshold values for the LTP based operators is 0 to 30. The set of possible numbers of bins used by the LBP/C based operators ranges from 1 to 20.

- **Classification Method Specific Parameters**
  The used distance metric of the k-NN method is the Histogram Intersection. The SVM method uses a radial basis function as kernel. An optimal k-value is searched during each experiment within a range from 1 to 25. The maximum number of selected histograms by the features subset selection algorithm is 10.

6 Experimental Results

The following Section presents the results of the experiments using six LBP-based operators. The experiments were conducted with the gathered image data from the Bulbus Duodeni as well as the Pars Descendens in both, the four and two-class scheme for classification. For simplicity the duodenal regions will be abbreviated in the discussion of specific experiments as 'Bulbus' in case of the duodenal Bulb and as 'Pars' when referring to the Pars Descendens.

The classification scheme applied in the four-class case is as discussed in Section 2.3 with type Marsh-3c being the most severe form of celiac disease. Marsh-0 indicates no celiac specific changes of the duodenal mucosa. In case of the two-class scheme the union of images corresponding to types Marsh-3a, Marsh-3b and Marsh-3c are used as the second class (this is referred to as the 'celiac' class). The result tables indicate the two-class scheme by combining the three celiac type Marsh columns. In parallel to the four-class scheme the Marsh-0 type images are referenced as the 'non-celiac' class.

The method names are encoded in a way that helps referencing the results unambiguously. This encoding uses the operator name as used in the title of the corresponding Subsections in Section 3 (e.g.' LTP'). Additionally an abbreviation for the classification method is used. The k-Nearest Neighbor classifier is abbreviated as 'k-NN' while the Support Vector Machines classifier is abbreviated as 'SVM'. Finally the image type is appended in braces indicating if the experiment was based on the grayscale versions of the images or was using the color versions of the images. The abbreviation 'GS' stands for grayscale while the color type experiments are indicated by the appended term 'Color'.

The numbers presented within the result tables are the percentages of correctly classified images together with their specificity and sensitivity. The specificity is the percentage of correctly classified images actually showing a normal mucosal state, while the sensitivity is the percentage of correctly classified images indicating mucosal changes caused by celiac disease. In the case of the four-class scheme for classification, the absolute classification rate of each separate class is listed. Therefore the term sensitivity refers to the classification rates of the Marsh classes indicating celiac disease. In case of interesting rates a more detailed discussion of each class is given. To improve the readability the results are rounded to one decimal place in the discussion.

Additionally to the classification rates the specific parameters are listed. The presented parameters are dependent on the operator used for feature extraction as well as the classification method. Experiments that optimized a threshold value present the optimal threshold in a separate column indicated by 't'. If the optimized parameter was the number of used contrast bins (in case of experiments based on LBP/C) the optimal value is listed in a column indicated by 'bins'. If the result was achieved by the k-NN classification method, the k-parameter was optimized using a range of 1 to 25. The optimal value is displayed within the column labeled as 'k'.

During each experiment a subset of locally optimal histograms was selected by using the sequential forward feature subset selection algorithm. This subset was subsequently used for classification. This subset is encoded in a way that avoids ambiguity and supports readability and is listed in the 'features' column. The subset encoding is arranged such that each column describes a single histogram that was selected by the algorithm. It is read as $\begin{bmatrix} \text{subband} \\ \text{scale} \\ \text{color channel} \end{bmatrix}$.

The term subband is not clear without ambiguity among the operators unfortunately. It is either used to indicate a filter orientation (in case of LTP and ELTP) or a wavelet subband (in case of the WTLBP operator). The specific meaning is explained within the preamble of each operator's result Section. In the case of experiments based on grayscale images as well as experiments using operators without subbands, the corresponding elements are marked as omitted by using a hyphen ('-'). The color channels are abbreviated as 'R' for the red color subband, 'G' and 'B' for the green and the blue color subband respectively. The scales range from 1 to 5 and have the same meaning as explained in Section 3.2.3.

6.1 Results of the LBP-Operator

The Local Binary Pattern (LBP) operator using the multiscale extension and uniform patterns builds the basis for all other operators that are applied in this thesis. In total 5 histograms are computed in experiments based on grayscale images and 15 histograms in case of using color images. Each histogram is 59-bins wide. Table 4 lists the results of the experiments based on the LBP operator.

Method	**Classification Rates**						
	Marsh-0	Marsh-3a	Marsh-3b	Marsh-3c	Total	k	Features
Bulbus							
LBP-kNN (GS)	97.58	87.10			**93.08**	10	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{3}}$
LBP-kNN (Color)	98.18	87.90			**93.77**	6	$\overset{-}{\underset{R}{1}}\overset{-}{\underset{R}{2}}\overset{-}{\underset{R}{3}}\overset{-}{\underset{R}{4}}\overset{-}{\underset{G}{1}}\overset{-}{\underset{G}{2}}$
LBP-SVM (GS)	96.97	93.55			**95.50**	-	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{3}}\overset{-}{\underset{-}{4}}$
LBP-SVM (Color)	95.76	93.55			**94.81**	-	$\overset{-}{\underset{R}{1}}\overset{-}{\underset{R}{3}}\overset{-}{\underset{B}{1}}$
LBP-kNN (GS)	98.79	57.45	74.07	30.43	**82.01**	7	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{3}}\overset{-}{\underset{-}{4}}$
LBP-kNN (Color)	95.76	65.96	72.22	52.17	**83.04**	3	$\overset{-}{\underset{R}{1}}\overset{-}{\underset{R}{3}}\overset{-}{\underset{R}{4}}\overset{-}{\underset{G}{1}}\overset{-}{\underset{B}{1}}$
LBP-SVM (GS)	95.76	61.70	72.22	52.17	**82.35**	-	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{3}}\overset{-}{\underset{-}{5}}$
LBP-SVM (Color)	96.97	59.57	74.07	56.52	**83.39**	-	$\overset{-}{\underset{R}{2}}\overset{-}{\underset{R}{3}}\overset{-}{\underset{G}{1}}$
Pars Descendens							
LBP-kNN (GS)	84.11	86.34			**85.33**	5	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{3}}\overset{-}{\underset{-}{4}}\overset{-}{\underset{-}{5}}$
LBP-kNN (Color)	78.81	88.52			**84.13**	11	$\overset{-}{\underset{R}{1}}\overset{-}{\underset{R}{2}}\overset{-}{\underset{R}{3}}\overset{-}{\underset{R}{4}}\overset{-}{\underset{G}{1}}\overset{-}{\underset{G}{3}}$
LBP-SVM (GS)	82.12	85.79			**84.13**	-	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{2}}\overset{-}{\underset{-}{3}}\overset{-}{\underset{-}{4}}$
LBP-SVM (Color)	82.78	85.79			**84.43**	-	$\overset{-}{\underset{R}{2}}\overset{-}{\underset{R}{3}}\overset{-}{\underset{G}{3}}\overset{-}{\underset{B}{1}}\overset{-}{\underset{B}{4}}$
LBP-kNN (GS)	86.75	23.40	43.33	69.74	**66.17**	14	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{2}}\overset{-}{\underset{-}{3}}$
LBP-kNN (Color)	86.09	19.15	56.67	68.42	**67.37**	5	$\overset{-}{\underset{R}{1}}\overset{-}{\underset{R}{2}}\overset{-}{\underset{R}{3}}$
LBP-SVM (GS)	89.40	36.17	56.67	65.79	**70.66**	-	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{3}}$
LBP-SVM (Color)	88.74	8.51	58.33	69.74	**67.66**	-	$\overset{-}{\underset{R}{1}}\overset{-}{\underset{R}{3}}\overset{-}{\underset{B}{1}}\overset{-}{\underset{B}{3}}$

Table 4: Experimental Results Based on the LBP-Operator.

The best classification rate of the Bulbus set was achieved by the SVM method using grayscale images with 95.5 percent. This is in contrast to the majority of the results based on other operators, where the classification of the color images performs better compared to the grayscale images. The better overall classification rates of the SVM methods in case of the Bulbus two-class scheme is caused by a higher sensitivity, compared to the rates of the k-NN method, by an average of 6 percentage points. On the other hand we can observe that the SVM method has a lower specificity of 0.6 percentage points in the grayscale case and 2.4 percentage points when using color images.

The best result when classifying images from the Pars was also achieved by using grayscale features. In contrast to the Bulbus two-class results the k-NN method performed better with a maximum overall classification rate of 85.3 percent. Opposed to the results of the Bulbus set, the method's general sensitivity is higher than the method's specificity. The differences are approximately 3 percentage points except in the case of the k-NN method using color features with 9.7 percentage points.

In the four-class case a considerable decrease of the overall classification rates can be observed. The highest rate achieved considering the Bulbus set is 83.4 percent using the SVM method with color features. The classification rates of the Marsh-0 image class is excellent among all four presented results averaging at 96.8 percent. This can be observed throughout all experiments and can be explained by the imbalance of class members within the image sets. The class with the smallest number of members performs worst in this experiments with an average classification rate of only 47.8 percent. The two other celiac classes perform at classification rates around 60 percent (Marsh-3a) and slightly above 72 percent (Marsh-3b).

The best classification result of the Pars four-class experiments was achieved by SVM (grayscale) with an overall rate of 70.7 percent. Compared to the other overall results this method's performance was over 3 percentage points higher. The specificity in the four-class experiments is comparable to the two-class results. Again the image set with the smallest number of members (in this case Marsh-3a) performs worse with a horrible 8.5 percent in the case of SVM (color) and an average of only 21.8 percent among all results. The classification rates of Marsh-3b average at 53.8 percent with an outlying result of k-NN (grayscale) with only 43.3 percent. The classification rates of Marsh-3c average at 68.4 percent.

The values of considered neighbors (k-values) in case of the k-NN classifier range from 3 to 14 with a mean of 7.6 and show no significant pattern. The relative number of selected histograms (relative to the maximum number of usable histograms) by the feature selection algorithm is 60 percentage points (grayscale) and 43 percentage points (color).

Considering the basic LBP operator we can see that the classification performance in the two class case is well over 93 percent and 84 percent which is a good basis towards further improvements. On the other hand, the overall classification rates of the four class experiments are at 82 percent and close to 70 percent. It is interesting that the color features did not significantly improve the classification results.

6.2 Results of the ELBP-Operator

The extended LBP (ELBP) operator uses Sobel filtering to generate gradient images. In this case the operators capture the velocity of pixel intensity variations. When applying a Sobel filter, two possible options in terms of orientation exist. One orientation approximates the gradient in the horizontal, the other approximates the gradient in the vertical direction. Observations show that mucosal images usually have a dominant orientation. Hence it is not known if a certain direction is beneficial to the classification process for a given image set. Therefore three different filters are applied. The first filter is the horizontal Sobel filter and is labeled 'H' as subband element in the feature encoding. The second filter is the vertical Sobel filter and is abbreviated as 'V'. The third filter takes into account that diagonal information could be beneficial. Therefore the mean of the horizontal and vertical filter responses is used. This is called the diagonal direction throughout this Section. The diagonal direction is indicated by a 'D'. The ELBP operator computes 15 histograms in case of grayscale and 45 histograms when using color images. Each histogram is 59-bins wide.

Table 5 shows the classification rates achieved by the experiments using the ELBP operator. Overall, the color features yield better results for all experiments. Considering the experiments based on the two-class scheme the best results in the case of the Bulbus images is 95.2 percent, achieved by using the k-NN method with color features. Compared to the grayscale features this is an improvement of 2.3 percentage points and can be explained by the improved sensitivity of 4.8 percentage points. The SVM method could improve the overall rate only insignificantly by 0.3 percentage points using color images. The best result of the experiments based on the Pars images was achieved by SVM (Color) with 88.0 percent. In this case the improvement compared to grayscale features was 4.2 percentage points. The results of the k-NN method improved only by a small margin by using color features (0.9 percentage points) and reached an overall classification rate of 86.5 percent.

Method	**Classification Rates**						
	Marsh-0	Marsh-3a	Marsh-3b	Marsh-3c	Total	k	Features
Bulbus							
ELBP-kNN (GS)	94.55	90.32			**92.73**	10	H H V V D D 2 4 1 3 2 3 - - - - - -
ELBP-kNN (Color)	95.15	95.16			**95.16**	3	H H D D D 1 4 1 5 1 G B R R B
ELBP-SVM (GS)	93.94	94.35			**94.12**	-	H H V V V D D 2 3 1 2 4 1 2 - - - - - - -
ELBP-SVM (Color)	95.76	92.74			**94.46**	-	H D D D 1 1 2 2 R R G B
ELBP-kNN (GS)	96.36	63.83	62.96	30.43	**79.58**	19	H V V D D 2 1 4 3 5 - - - - -
ELBP-kNN (Color)	94.55	70.21	70.37	52.17	**82.70**	3	H H H V D D D 1 4 4 1 1 2 1 R R B R R G B
ELBP-SVM (GS)	95.15	55.32	48.15	26.09	**74.39**	-	H V 2 3 - -
ELBP-SVM (Color)	96.36	61.70	53.70	30.43	**77.51**	-	H V D 2 3 1 G B R
Pars Descendens							
ELBP-kNN (GS)	78.81	91.26			**85.63**	9	H H H H D D D 1 3 4 5 1 2 5 - - - - - - -
ELBP-kNN (Color)	82.78	89.62			**86.53**	11	H H H H D D 2 4 3 1 3 1 R R G B R B
ELBP-SVM (GS)	80.79	86.34			**83.83**	-	V D D 1 1 3 - - -
ELBP-SVM (Color)	83.44	91.80			**88.02**	-	H H H V V V D D D 4 1 3 5 1 2 2 3 5 R G B G B B G B B
ELBP-kNN (GS)	86.75	23.40	55.00	71.05	**68.56**	9	H H H V D 1 3 4 3 1 - - - - -
ELBP-kNN (Color)	89.40	21.28	60.00	73.68	**70.96**	11	H H D D 1 1 1 2 G B G G
ELBP-SVM (GS)	91.39	12.77	58.33	64.47	**68.26**	-	H H V D 3 4 1 1 - - - -
ELBP-SVM (Color)	90.07	19.15	63.33	76.32	**72.16**	-	D D D D D 1 4 2 1 3 R R G B B

Table 5: Experimental Results based on the ELBP-Operator.

Considering the experiments based on the four-class scheme, the best result was provided by the k-NN method using color features with 82.7 percent. Compared to the corresponding grayscale experiment the improvement was 3.1 percentage points. The other experiments did not reach the 80 percent mark which is the worst performance among all operators. The SVM method reached a rate of only 77.5 percent. The best result of the Pars-experiments is 72.2 percent (SVM color). This is an improvement of 3.9 percentage points compared to the corresponding grayscale experiment. The improvement is caused by a superior classification rate of Marsh-3b (63.3%) and Marsh-3c (76.3%). The kNN method reached a result of 71.0 percent (an improvement of 2.4 percentage points compared to the grayscale experiment).

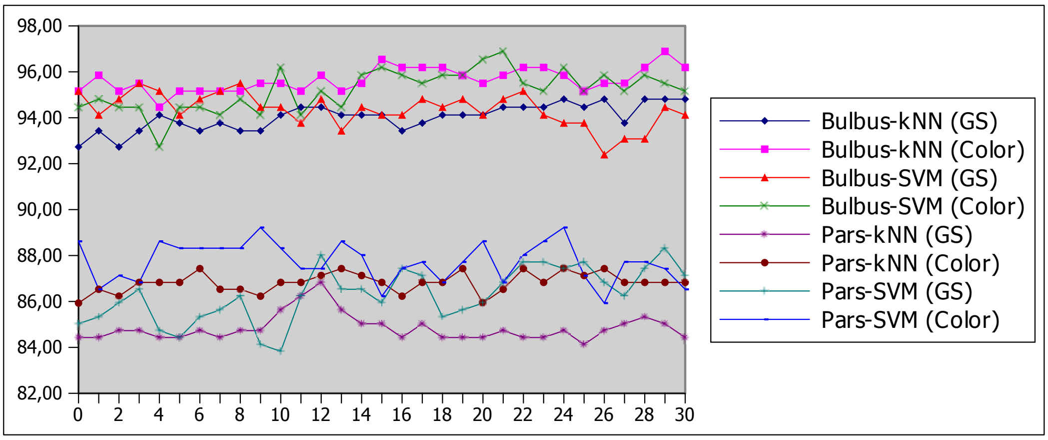
It is interesting that the results based on the Bulbus set in the four-class case are significantly lower than the results of the other operators. The results of the ELTP operator indicate that the discriminative information is present in gradient images. Therefore this significant decrease in classification rate might be an effect of noise amplification that is introduced by the Sobel filtering.

The spread of k-values is 3 to 19. The relative number of selected histograms (relative to the maximum number of usable histograms) by the feature selection algorithm is 48.7 percent for the grayscale features and 53.7 percent for the color features.

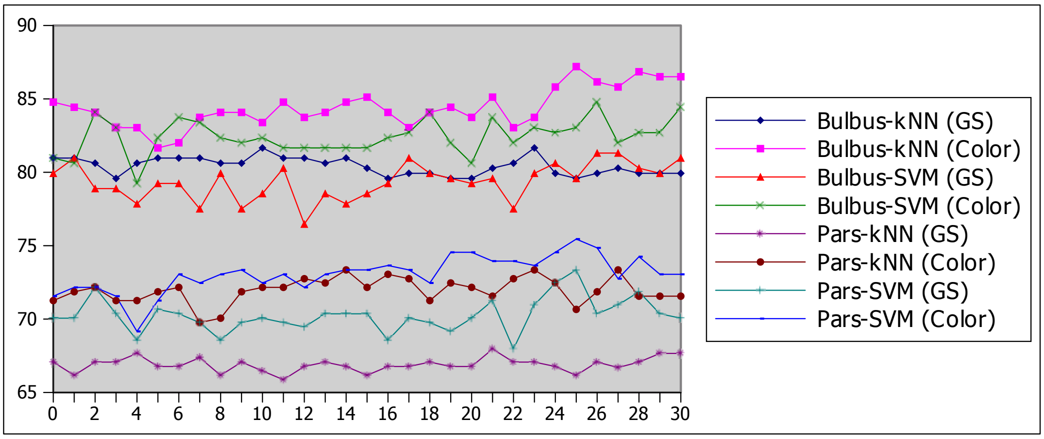
6.3 Results of the LTP-Operator

The Local Ternary Pattern operator uses a threshold to minimize the noise within histograms that is caused by bad image quality. In parallel to the LBP operator no subbands exist. In total 5 histograms are computed when using grayscale features and 15 when using color features. Each histogram contains 116-bins.

Figure 31 shows the results according to the specific threshold values. For each threshold the set of used histograms was optimized for classification. So was the k-value when using the k-NN method. In the case of classifying the images from the Bulbus (two-class case) we can observe that the degree of variation of the results caused by changing threshold values was rather small. In numbers, the margin between the worst and the best results was 2.1 percentage points (grayscale) as well as 2.4 percentage points (color) for the k-NN method and 3.1 percentage points (grayscale) as well as 4.2 percentage points (color) for results achieved by the SVM method. It is interesting to note that the k-NN method behaves more robust in terms of classification rate compared to the SVM-method for different threshold values. The same property holds for the results of the Pars image set. The results of the k-NN method vary by 2.7 percentage points (grayscale) and 1.5 percentage points (color). Again the variation of the results is higher in case of the SVM classifier with 4.5 percentage points (grayscale) and 3.3 percentage points for the color features. In case of the four-class scheme used for classification we can see that the different threshold values have a higher impact on the overall classification results compared to the two-class case. This is caused by the generally lower discrimination of features between different classes. We can see that in parallel to the two-class case the SVM method shows a higher degree of variation. In the results based on experiments with the Bulbus set, the variation is 2.1 percentage points (grayscale) and 5.5 percentage points (color) in case of the k-NN method, compared to 4.8 percentage points (grayscale) as well as 5.5 percentage points (color) for the SVM method. In case of classifying images from the Pars set, the difference is even clearer with 2.1 percentage points (grayscale) and 3.6 percentage points (color) for results of the k-NN method and 5.4 percentage points (grayscale) as well as 6.3 percentage points (color) in case of the SVM classifier.



(a) LTP-Operator (two-class scheme).



(b) LTP-Operator (four-class scheme).

Figure 31: Effect of Threshold Values on the Overall Classification Rate.

The results with the highest overall classification rate combined with the specific parameters and selected feature subsets are presented in Table 6. In the case of two different thresholds with the same overall classification rate, the result with the larger threshold is presented.

Method	**Classification Rates**							
	Marsh-0	Marsh-3a	Marsh-3b	Marsh-3c	Total	k	t	Features
				Bulbus				
LTP-kNN (GS)	96.97	91.94			**94.81**	3	24	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{3}}\overset{-}{\underset{-}{4}}$
LTP-kNN (Color)	97.58	95.97			**96.89**	1	29	$\overset{-}{\underset{R}{1}}\overset{-}{\underset{R}{4}}\overset{-}{\underset{G}{3}}\overset{-}{\underset{G}{5}}\overset{-}{\underset{B}{1}}\overset{-}{\underset{B}{4}}\overset{-}{\underset{B}{5}}$
LTP-SVM (GS)	96.97	93.55			**95.50**	-	27	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{2}}\overset{-}{\underset{-}{4}}\overset{-}{\underset{-}{5}}$
LTP-SVM (Color)	97.58	95.97			**96.89**	-	23	$\overset{-}{\underset{R}{1}}\overset{-}{\underset{R}{2}}\overset{-}{\underset{G}{4}}\overset{-}{\underset{B}{1}}\overset{-}{\underset{B}{2}}\overset{-}{\underset{B}{4}}$
LTP-kNN (GS)	98.18	61.70	66.67	39.13	**81.66**	5	23	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{3}}$
LTP-kNN (Color)	96.97	72.34	83.33	56.52	**87.20**	1	25	$\overset{-}{\underset{R}{1}}\overset{-}{\underset{R}{4}}\overset{-}{\underset{R}{5}}\overset{-}{\underset{G}{1}}\overset{-}{\underset{G}{3}}\overset{-}{\underset{B}{5}}$
LTP-SVM (GS)	96.97	72.34	59.26	39.13	**81.31**	-	26	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{2}}$
LTP-SVM (Color)	98.18	61.70	70.37	69.57	**84.78**	-	25	$\overset{-}{\underset{R}{1}}\overset{-}{\underset{R}{4}}\overset{-}{\underset{G}{2}}$
				Pars Descendens				
LTP-kNN (GS)	84.11	89.07			**86.83**	11	12	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{3}}\overset{-}{\underset{-}{4}}\overset{-}{\underset{-}{5}}$
LTP-kNN (Color)	86.09	88.52			**87.43**	5	26	$\overset{-}{\underset{R}{1}}\overset{-}{\underset{R}{3}}\overset{-}{\underset{R}{4}}\overset{-}{\underset{G}{4}}\overset{-}{\underset{G}{5}}\overset{-}{\underset{B}{1}}\overset{-}{\underset{B}{3}}$
LTP-SVM (GS)	84.77	91.26			**88.32**	-	29	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{2}}\overset{-}{\underset{-}{3}}\overset{-}{\underset{-}{5}}$
LTP-SVM (Color)	86.09	91.80			**89.22**	-	23	$\overset{-}{\underset{R}{1}}\overset{-}{\underset{R}{3}}\overset{-}{\underset{R}{4}}\overset{-}{\underset{B}{1}}\overset{-}{\underset{B}{1}}\overset{-}{\underset{B}{2}}$
LTP-kNN (GS)	88.08	29.79	45.00	69.74	**67.96**	11	21	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{3}}\overset{-}{\underset{-}{4}}$
LTP-kNN (Color)	81.46	38.30	73.33	78.95	**73.35**	1	27	$\overset{-}{\underset{R}{1}}\overset{-}{\underset{R}{3}}\overset{-}{\underset{R}{5}}\overset{-}{\underset{G}{1}}\overset{-}{\underset{G}{3}}\overset{-}{\underset{B}{1}}$
LTP-SVM (GS)	92.72	38.30	58.33	68.42	**73.35**	-	24	$\overset{-}{\underset{-}{1}}\overset{-}{\underset{-}{2}}\overset{-}{\underset{-}{3}}\overset{-}{\underset{-}{5}}$
LTP-SVM (Color)	91.39	34.04	71.67	72.37	**75.45**	-	25	$\overset{-}{\underset{R}{1}}\overset{-}{\underset{R}{3}}\overset{-}{\underset{R}{5}}\overset{-}{\underset{G}{2}}\overset{-}{\underset{B}{1}}\overset{-}{\underset{B}{2}}$

Table 6: Experimental Results based on the LTP-Operator.

In contrast to the results of the LBP operator, the color experiments reach a higher classification rate when using the LTP operator compared to the results based on grayscale images. The best result in the two-class experiments using Bulbus images is 96.9 percent. It is interesting that both, the k-NN and SVM method, have exactly the same classification rates but use different thresholds (29 in the case of k-NN and 23 in the case of SVM). Also the selected feature subsets were different. The better classification rates of the experiments based on color features compared to the grayscale experiments (1.7 percentage points by average) can be explained by a considerably higher sensitivity of 4.0 percentage points in case of the k-NN method and 2.4 percentage points for the SVM classifier. The specificity was improved slightly by 0.6 percentage points in case of both classification methods.

In the case of the Pars-experiments, the SVM method based on color features achieved the highest overall classification rate of 89.2 percent. In parallel to the experiments based on the Bulbus images, the classification rates of the experiments using color features are higher compared to the results achieved by using grayscale images. Opposed to the results of the Bulbus-experiments the difference is smaller with an average of 0.8 percentage points. It is interesting that the sensitivity of the k-NN (color) method decreased compared to the grayscale method (0.6%). The better overall classification rate can be explained by an increase in specificity by 2.0 percentage points. In case of the SVM method both, the specificity (1.3%) and the sensitivity (0.5%) could be improved by using color features.

The best results within the experiments based on the four-class scheme is reached by the k-NN method using color features with 87.2 percent. This result is 5.5 percentage points higher compared to the grayscale features, which is a considerable improvement. The improvement can be explained by an overall increase in classification rate of all three celiac classes (Marsh-3a 10.6%, Marsh-3b 16.7%, Marsh-3c 17.4%). Although the method's specificity decreased slightly (1.2%) the overall classification rate could be considerably improved. This might be an indication that color information is beneficial for intra-celiac classification.

Using color features also improved the overall classification rates of the SVM method. The best result reached 84.8 percent, this is 3.5 percentage points above the best grayscale result of this classification method. Using the color features led to an improvement of the classification rates of images from the Marsh-3b (11.1%) and Marsh-3c (30.4%) classes. The lower result compared to the k-NN method is caused by a decrease in classification rate of Marsh-3a (10.6%).

Considering the results of the experiments based on the four-class scheme using images from the Pars we observe that using color features could also improve the classification rates. The k-NN method reached an overall rate of 73.4 percent which is an improvement of 5.4 percentage points compared to the grayscale experiment. The improvement was caused by higher classification rates of the celiac classes (8.5% Marsh-3a, 28.3% Marsh-3b, 9.2% Marsh-3c). This is in parallel to the Bulbus color experiments. Interestingly the specificity also dropped, this time by 6.6 percentage points. The improvements gained by using color features combined with the SVM method were only 2.1 percentage points compared to the grayscale experiment. This led to the best overall rate of 75.5 percent and was caused, in parallel to the k-NN method, by an improvement of the classification rates of Marsh-3b (13.3%) and Marsh-3c (4.0%).

The mean threshold value of the presented results is 24.3. The only outlying value was 12, used by the k-NN (grayscale) method in the Pars two-class experiments. The number of considered neighbors of the k-NN classification method (k-value) varied between 1 and 11.

The relative number of selected histograms (relative to the maximum number of usable histograms) by the feature subset selection algorithm is 81.3 percent in case of using the grayscale features and 58.8 percent in case of the color features.

6.4 Results of the ELTP-Operator

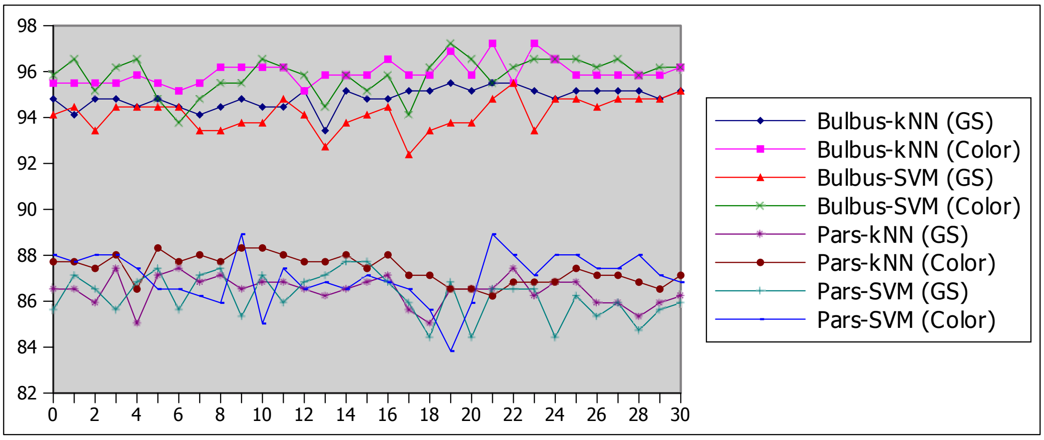
The extended LTP (ELTP) operator uses Sobel filtering in combination with thresholding to capture information present within the input data. In the context of ELTP, subbands have the meaning of Sobel filter directions. This is in parallel to the ELBP operator. The abbreviation 'H' stands for a horizontally filtered input signal, 'V' stands for a vertically filtered input signal. Additionally to the horizontal and vertical filter also a combination of both is used. This is accomplished by calculating the mean of the horizontal and vertical filter responses and is denoted as 'D' throughout the results. The ELTP operator extracts 5 histograms for each subband this corresponds to 15 histograms in case of grayscale experiments and 45 histograms in case of using color images. Each histogram is 118-bins wide. The threshold values were optimized in a range between 0 and 30.

Figure 32 shows the effects of specific threshold values to the overall classification rates. Reviewing the results we observe that the properties of the classification methods in combination with the LTP operator also hold for the ELTP operator. In case of the two-class scheme we can see that the k-NN method is more robust against changing threshold values. Considering experiments based on the Bulbus images the variation is 2.1 percentage points for color and grayscale features. The SVM method is more sensitive. The results vary between 3.1 percentage points (grayscale) and 3.5 percentage points (color). In case of the experiments based on images from the Pars set, the same relation can be observed. The results of the k-NN method vary between 2.4 percentage points (grayscale) and 2.1 percentage points (color) while the results of the SVM method vary by 3.3 percentage points (grayscale) and 5.1 percentage points (color).

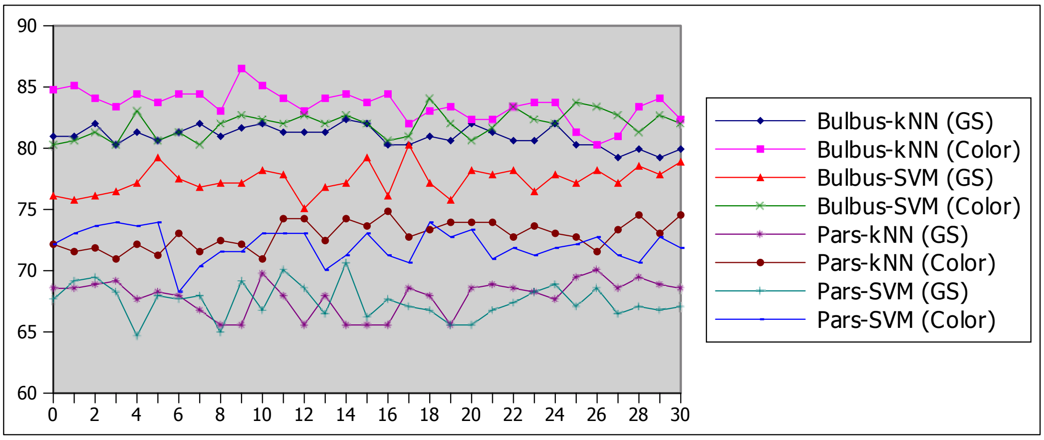
Considering the four-class case we can observe that the variation increases. This is in parallel to the LTP operator. Opposed to the results of the LTP operator no clear distinction between experiments based on k-NN and SVM can be observed. Considering the Bulbus images, the results of the k-NN method vary by 3.1 percentage points (color) and 6.2 percentage points (grayscale) while the results of the SVM method vary by 5.2 percentage points (grayscale) and 3.8 percentage points (color). The classification of features that were extracted by the ELTP operator on images from the Pars set show again that the k-NN method is more robust compared to the SVM method. The k-NN had a variation in results of 4.5 percentage points (grayscale) and 3.9 percentage points (color) while the SVM method's results varied by 6.0 percentage points (grayscale) as well as 5.7 percentage points (color).

The results with the highest overall classification rate combined with the specific parameters and selected feature subsets are presented in Table 7. In the case of two different thresholds with the same overall classification rate, the result with the larger threshold is presented in the table.

Considering the experiments focused on the two-class problem we can see that in case of the Bulbus images both methods reach an overall classification rate of 97.2 percent. This is an improvement of 1.7 percentage points compared to the grayscale results, which happened to be equal for both methods as well (95.5%). The improvement was caused by an increased sensitivity of 4.3 percentage points in case of the k-NN method and 2.4 percentage points in case of the SVM method. In case of SVM the method's specificity could also be improved by 1.2 percentage points. The k-NN method's specificity could not be improved. The results of the experiments based on the Pars images show that using color features could improve the results only slightly. The best result was achieved by the SVM method with an overall classification rate of 88.9 percent. This was an improvement of 1.2 percentage points



(a) ELTP-Operator (two-class scheme).



(b) ELTP-Operator (four-class scheme).

Figure 32: Effect of Threshold Values on the Overall Classification Rate.

compared to the grayscale experiment and can be explained by an improved specificity of 2.0 percentage points as well as an improved sensitivity of 0.5 percentage points. In parallel the results of the k-NN classifier could be improved by using color features by 0.9 percentage points to an overall rate of 88.3 percent.

The results of the experiments based on the four-class scheme show that using color features could considerably improve all results. In case of the Bulbus images, the best result was achieved by the k-NN method with 86.5

Method	Classification Rates							
	Marsh-0	Marsh-3a	Marsh-3b	Marsh-3c	Total	k	t	Features
				Bulbus				
ELTP-kNN (GS)	97.58	92.74			**95.50**	4	22	H H H D D 1 2 4 1 2 - - - - -
ELTP-kNN (Color)	97.58	96.77			**97.23**	3	20	H H H D D D D 1 2 5 5 1 3 4 R R R R B B B
ELTP-SVM (GS)	96.97	93.55			**95.50**	-	22	H H H H V D D 1 3 4 5 5 1 4 - - - - - - -
ELTP-SVM (Color)	98.18	95.97			**97.23**	-	19	H H V D D 1 5 4 5 1 R G G R B
ELTP-kNN (GS)	98.18	57.45	68.52	52.17	**82.35**	9	14	H H V V D 2 4 1 4 3 - - - - -
ELTP-kNN (Color)	97.58	72.34	79.63	52.17	**86.51**	3	9	H H H V D D D 1 4 4 4 3 1 2 R R G G R B B
ELTP-SVM (GS)	98.79	74.47	48.15	34.78	**80.28**	-	17	H H H H V D 1 3 4 5 3 1 - - - - - -
ELTP-SVM (Color)	99.39	76.60	62.96	39.13	**84.08**	-	18	H H H V D 1 1 4 4 1 R G G G B
				Pars Descendens				
ELTP-kNN (GS)	84.11	90.16			**87.43**	9	22	H H D D D 3 4 1 2 3 - - - - -
ELTP-kNN (Color)	86.75	89.62			**88.32**	5	10	H H V V D D D 3 4 4 5 4 1 1 R G G G R G B
ELTP-SVM (GS)	85.43	89.62			**87.72**	-	15	H H H 1 3 5 - - -
ELTP-SVM (Color)	87.42	90.16			**88.92**	-	19	H V D D D 1 4 3 1 4 R B R G B
ELTP-kNN (GS)	87.42	25.53	56.67	73.68	**70.06**	4	26	H H H V V V D D D 2 3 5 1 4 5 1 3 4 - - - - - - - - -
ELTP-kNN (Color)	90.07	42.55	65.00	72.37	**74.85**	5	16	H D D D 2 1 4 1 B G G B
ELTP-SVM (GS)	92.72	21.28	63.33	63.16	**70.66**	-	13	H H H H V V V D D D 1 2 3 5 1 3 5 1 3 5 - - - - - - - - - -
ELTP-SVM (Color)	92.72	8.51	71.67	78.95	**73.95**	-	19	H H H H V D D D D 1 2 3 5 2 1 3 1 1 R R G B G R R G B

Table 7: Experimental Results based on the ELTP-Operator.

percent. This is an improvement of 4.2 percentage points compared to the best grayscale result. The improvement was caused by an improved classification rate of the Marsh-3a (14.9%) and Marsh-3b (11.1%) classes while being consistent for class Marsh-3c. The specificity slightly decreased (0.6%). Considering the Pars-experiments, the same trend can be observed. The k-NN method provided the best overall classification rate of 74.9 percent, an improvement of 4.8 percentage points compared to the best grayscale result. The improvement was caused by a superior specificity of 2.7 percentage points and a higher classification rate of classes Marsh-3a (improved by 17.0%) and Marsh-3b (improved by 8.3%). The result of Marsh-3c slightly decreased by 1.3 percentage points. The SVM method also reached a higher classification rate by employing color features. The improvement was 3.3 percentage points to an overall classification rate of 74.0 percent. It is interesting to note that the smallest class (Marsh-3a) had a classification rate of only 8.5 percentage points in this experiment.

Compared to the LTP operator the overall optimal threshold values decreased to 17.5 with a minimum value of 9. The k-values are within the range of 3 to 9. The relative number of used histograms was 62.5 percent (grayscale) and 62.3 percent (color).

6.5 Results of the LBP/C-Operator

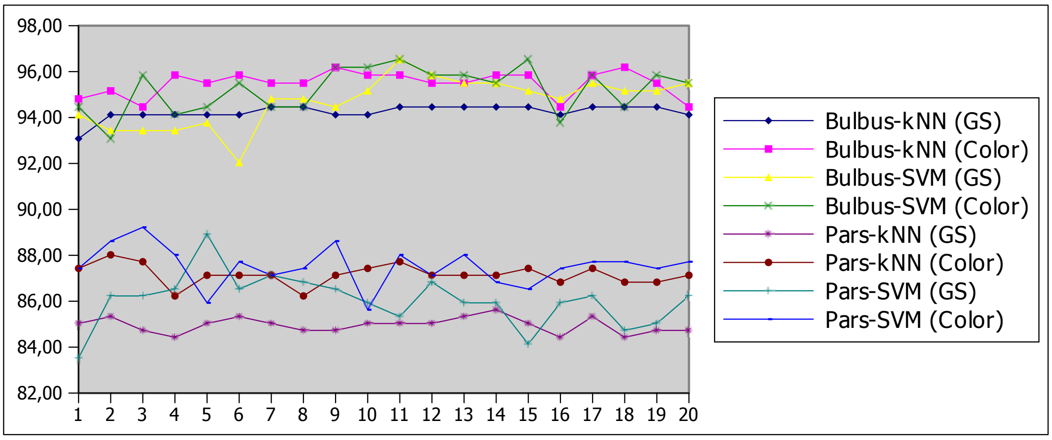
The LBP/C operator uses a complementary contrast measure to improve the discriminative information gained by the LBP. The number of bins was optimized within each experiment in a range from 1 to 20. The specific number of bins does not change the number of computed histograms. Instead of using a two-dimensional approach the histograms are concatenated (this is logically equivalent). Therefore the number of used contrast bins changes the length of the computed histograms. If the number of used contrast bins is n each histogram contains $59n$-bins. The total number of computed histograms by the LBP/C operator is 5 in case of grayscale features and 15 in case of using color features.

Figure 33 demonstrates the effect of using different values for the binning of the contrast values. In the two-class case the mean variation between the best and worst results is higher when using the SVM method (4% - Bulbus and 4.5% - Pars). The k-NN method behaves considerably smoother with a mean variation of only 1.5 percentage points in case of both images sets.

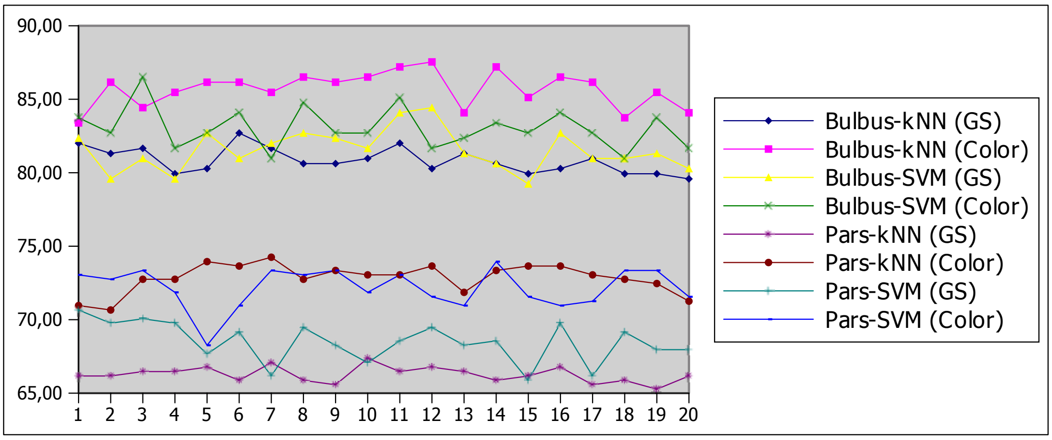
In parallel to the two-class model, the SVM method behaves more sensitive to changing number of bins in case of the four-class problem. The mean variation in case of the Bulbus images is 5.4 percentage points compared to only 3.6 percentage points by the k-NN method. This property also holds for the results of the experiments focused on the Pars images with a mean of 5.2 percentage points (SVM) compared to 2.8 percentage points (k-NN).

The results with the highest overall classification rate combined with the specific parameters and selected feature subsets are presented in Table 8. In the case of two results with an equal overall classification rate, the result with the lower number of bins is presented in the table.

Reviewing the experimental results of the LBP/C operator we observe that



(a) LBP/C-Operator (two-class scheme).



(b) LBP/C-Operator (four-class scheme).

Figure 33: Effects of the Number of Contrast Bins on the Overall Classification Rates.

the color features provide the best overall classification rates again. In case of the two-class experiments the best classification rate based on the images from the Bulbus is 96.5 percent which is achieved by the SVM method. It is interesting that the grayscale experiment reached exactly the same result in this case. Although the number of used bins is equal, we can see that the feature subset is very similar as well. Both results of the k-NN method are below this mark and reach a maximum of 96.2 percent in case of the color features. Another interesting point is that the k-NN method based on

Method	Classification Rates							
	Marsh-0	Marsh-3a	Marsh-3b	Marsh-3c	Total	k	bins	Features
Bulbus								
LBP/C-kNN (GS)	100.00	87.10			**94.46**	22	7	$\bar{1}_{-}\bar{5}_{-}$
LBP/C-kNN (Color)	96.36	95.97			**96.19**	1	9	$\bar{1}_{G}\bar{2}_{G}\bar{3}_{G}\bar{1}_{B}\bar{3}_{B}$
LBP/C-SVM (GS)	98.18	94.35			**96.54**	-	11	$\bar{1}_{-}\bar{5}_{-}$
LBP/C-SVM (Color)	98.18	94.35			**96.54**	-	11	$\bar{1}_{R}\bar{5}_{R}\bar{1}_{G}\bar{5}_{G}$
LBP/C-kNN (GS)	96.36	63.83	74.07	43.48	**82.70**	3	6	$\bar{1}_{-}\bar{2}_{-}$
LBP/C-kNN (Color)	96.97	70.21	81.48	69.57	**87.54**	1	12	$\bar{1}_{R}\bar{1}_{G}\bar{2}_{G}\bar{3}_{G}\bar{4}_{G}\bar{5}_{G}\bar{1}_{B}\bar{2}_{B}\bar{3}_{B}$
LBP/C-SVM (GS)	95.15	80.85	77.78	30.43	**84.43**	-	12	$\bar{1}_{-}\bar{2}_{-}\bar{3}_{-}\bar{4}_{-}$
LBP/C-SVM (Color)	96.36	78.72	70.37	69.57	**86.51**	-	3	$\bar{1}_{R}\bar{2}_{R}\bar{5}_{G}\bar{3}_{B}$
Pars Descendens								
LBP/C-kNN (GS)	83.44	87.43			**85.63**	15	14	$\bar{1}_{-}\bar{3}_{-}\bar{4}_{-}$
LBP/C-kNN (Color)	88.08	87.98			**88.02**	7	2	$\bar{1}_{R}\bar{3}_{R}\bar{5}_{R}\bar{1}_{G}\bar{4}_{G}\bar{1}_{B}\bar{2}_{B}\bar{4}_{B}$
LBP/C-SVM (GS)	88.08	89.62			**88.92**	-	5	$\bar{1}_{-}\bar{2}_{-}\bar{3}_{-}\bar{4}_{-}$
LBP/C-SVM (Color)	87.42	90.71			**89.22**	-	3	$\bar{1}_{R}\bar{3}_{R}\bar{1}_{G}\bar{2}_{G}\bar{3}_{G}\bar{1}_{B}\bar{2}_{B}\bar{3}_{B}$
LBP/C-kNN (GS)	88.74	19.15	46.67	71.05	**67.37**	21	10	$\bar{1}_{-}\bar{2}_{-}\bar{3}_{-}$
LBP/C-kNN (Color)	78.81	51.06	73.33	80.26	**74.25**	1	7	$\bar{1}_{R}\bar{3}_{R}\bar{1}_{G}\bar{1}_{B}$
LBP/C-SVM (GS)	89.40	36.17	56.67	65.79	**70.66**	-	1	$\bar{1}_{-}\bar{3}_{-}$
LBP/C-SVM (Color)	91.39	21.28	68.33	76.32	**73.95**	-	14	$\bar{1}_{R}\bar{3}_{R}\bar{4}_{R}\bar{5}_{R}\bar{1}_{G}\bar{2}_{G}\bar{1}_{B}\bar{2}_{B}$

Table 8: Experimental Results based on the LBP/C-Operator.

grayscale features reached a perfect (100%) specificity but a low sensitivity of only 87.1 percent. Considering the Pars-experiments the best result is 89.2 percent by the SVM method with color features. This is only a small improvement of 0.3 percentage points compared to the grayscale method. The kNN method reached a result of 88.0 percent which improves the rate by 2.4 percentage points compared to the grayscale experiment.

The experiments based on the four-class scheme show that using color information improves the discriminative power between celiac classes. The best results in case of the Bulbus are 87.5 percent by the k-NN method and 86.5 percent by the SVM method. Both classification rates could be improved by using color information (k-NN 4.8% and SVM 2.1%). In case of the SVM classifier only the classification rate of Marsh-3c could be improved (39.1%), the classes Marsh-3a and Marsh-3b suffered from a small decrease in classification rate. The best result in case of the Pars images was achieved by k-NN with 74.3 percent, an improvement of 6.9 percentage points compared to the grayscale experiment. As observed for most other operators, the improvement was caused by much higher intra-celiac class discrimination (Marsh-3a 31.9%, Marsh-3b 26.7%, Marsh-3c 9.2%), but a decrease in specificity (9.9%). The SVM method provided a lower result with 74.0 percent but could also improve compared to using grayscale features.

In case of the LBP/C operator we can observe that the optimal k-values vary considerably from 1 to 22. Please note that a high value used within the four-class scheme leads to a lower classification rate of the smaller classes. This was the case in the experiment based on the Pars image set using the k-NN classifier (k = 21). By booting out the smallest class, the rates of the classes with a higher number of images increase which significantly improves the overall classification rate. On the other hand, the corresponding color experiment uses a k-value of 1 and reaches a considerably higher classification rate of the smallest class (Marsh-3a). The used bin configurations vary from 1 to 15. The SVM method based on the four-class Pars-experiment used only a single bin. This corresponds to LBP and can also be observed in the experiments with the WTLBP operator. In that case the corresponding color experiment used only a single bin.

The relative number of selected histograms is 55.0 percent in case of the grayscale features and 61.3 percent in case of using the color features.

6.6 Results of the WTLBP-Operator

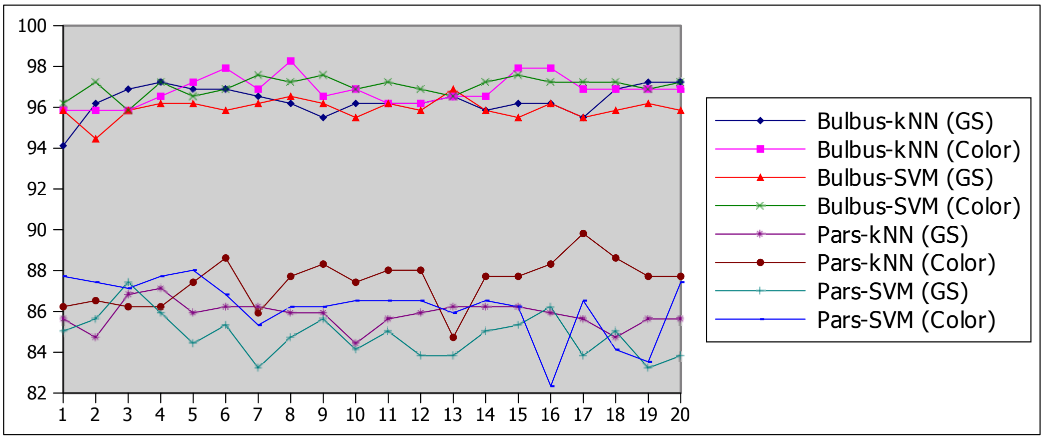
The WTLBP operator uses the LBP/C and the LTP operator in combination with the wavelet transform. The LBP/C operator is applied to the approximation subbands while the LTP operator is used to extract features from the detail subbands. In contrast to the LTP and ELTP experiments the thresholds were fixed during this experiments. The thresholds that are used to extract features from the horizontal and vertical subbands are set to 5. Due to the fact that the diagonal subband coefficients energy is lower overall, the threshold used for this subband was set to 3. The number of contrast bins was optimized within the possible range of 1 to 20. In the context of WTLBP the subbands refer to wavelet subbands. As the wavelet decomposition is performed until scale 3 and the multiscale extension is used for extracting features from the approximation subband the encoding of the used histogram subset is slightly changed. 'A1' stands for the approximation subband combined with LBP/C-scale 1. 'A2' stands for the approximation subband combined with LBP/C-scale 2 and 'A3' stands for the approximation subband combined with LBP/C-scale 3. The scale elements within the feature description vectors refer to the wavelet decomposition scale. The number of computed histograms is 18 when using grayscale features and 54 when using color features. The histograms are indirectly combined (see Section 3.2.13). The histograms extracted from the detail subbands contain 118-bins. The histograms extracted from the approximation subbands are $59n$-bins wide (assuming n is the number of used contrast bins).

Figure 34 presents the results based on different numbers of contrast bins. In case of the two-class scheme, the maximum difference between the best and the worst result of the Bulbus set was 3.2 percentage points by the k-NN method using grayscale features. This is in contrast to other results where the SVM method exhibited a higher degree of variation (SVM varied only by a maximum of 2.4 when using grayscale images). In case of the Pars image set the overall variation is considerably higher, with a maximum of 5.7 percentage points by the SVM method using color features compared to 5.0 percentage points by the k-NN method (color). In case of the WTLBP operator it can be observed that the SVM (color) and k-NN (color) methods do not behave as smoothly as the grayscale methods.

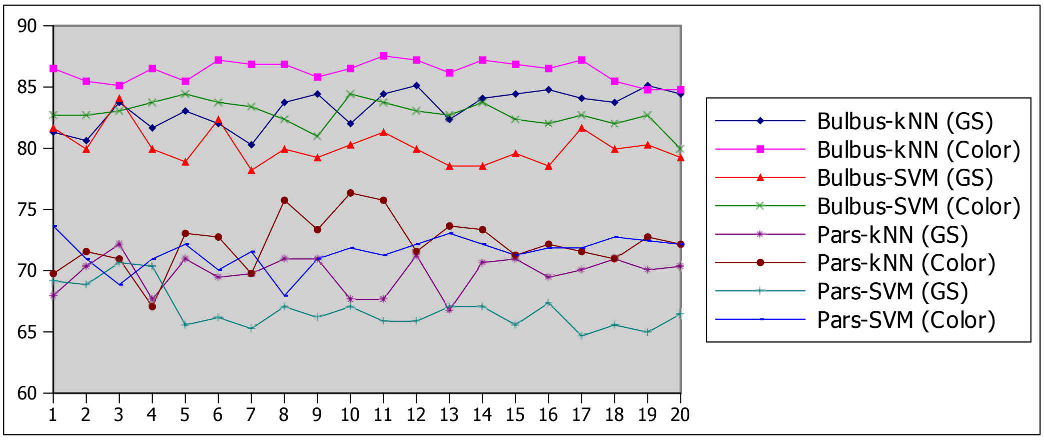
The results of the four-class experiments show a higher variation by an average of 4.5 percentage points in the case of the Bulbus images. The maximum is 5.9 percentage points achieved by SVM-grayscale. Opposed to the two-class model, the k-NN method behaves more smoothly. In case of the Pars images the experiment using the k-NN (color) method shows a huge variation of the results by 9.3 percentage points. The other methods have a variation comparable to other operators with a mean of 5.7 percentage points.

The results with the highest overall classification rate combined with the specific parameters and selected feature subsets are presented in Table 9. In the case of two results with an equal overall classification rate, the result with the lower number of bins is presented in the table.

When considering the experimental results of the WTLBP operator we can observe, that the characteristic, that color features provide better results compared to grayscale features holds. In the two-class Bulbus-experiments



(a) WTLBP-Operator (two-class scheme).



(b) WTLBP-Operator (four-class scheme).

Figure 34: Effects of the Number of Contrast Bins on the Overall Classification Rates.

the best result was achieved by the k-NN method with 98.3 percent. Both the specificity (0.6%) and the sensitivity (1.6%) could be improved. The result of the SVM method could as well be improved by 0.7 percentage points to an overall rate of 97.6 percent. In case of the Pars-experiments the best rate was achieved by k-NN (89.8%). This is an improvement of 2.7 percentage points and is caused by an improved specificity of 6.0 percentage points (the sensitivity did not change). The improvement of the classification rate caused by using color features was smaller when considering the SVM method (0.6%)

Method	Classification Rates							
	Marsh-0	Marsh-3a	Marsh-3b	Marsh-3c	Total	k	bins	features
Bulbus								
WTLBP-kNN (GS)	98.18	95.97			**97.23**	3	4	A1 A2 H D D 1 2 1 2 3 - - - - -
WTLBP-kNN (Color)	98.79	97.58			**98.27**	1	8	A1 A1 A1 A2 A2 A2 A3 A3 H H 3 1 2 3 2 1 2 1 1 1 R G G R G B R B R B
WTLBP-SVM (GS)	97.58	95.97			**96.89**	-	13	A1 A3 H V D 1 2 1 3 1 - - - - -
WTLBP-SVM (Color)	97.58	97.58			**97.58**	-	7	A1 A3 H H V 1 3 1 1 3 B G R B B
WTLBP-kNN (GS)	96.36	70.21	83.33	39.13	**85.12**	1	12	A1 A1 A2 A2 A2 A3 A3 H D D 1 3 1 2 3 1 2 1 1 3 - - - - - - - - - -
WTLBP-kNN (Color)	95.76	74.47	81.48	69.57	**87.54**	1	11	A2 A2 A2 A3 A3 H H V V V 2 1 3 2 3 2 1 3 2 1 G B B R R R B R G B
WTLBP-SVM (GS)	97.58	65.96	77.78	39.13	**84.08**	-	3	A1 A2 A3 D D 1 2 3 1 3 - - - - -
WTLBP-SVM (Color)	97.58	61.70	66.67	78.26	**84.43**	-	5	A1 A1 A2 A2 V 1 3 1 3 1 G G R B B
Pars Descendens								
WTLBP-kNN (GS)	82.78	90.71			**87.13**	7	4	A2 A3 H 1 2 1 - - -
WTLBP-kNN (Color)	88.74	90.71			**89.82**	3	17	A1 A1 A2 A3 A3 H H H V D 3 1 2 3 2 2 1 3 3 2 R B G G B R B B G G
WTLBP-SVM (GS)	85.43	89.07			**87.43**	-	3	A2 A2 H V 1 3 1 3 - - - -
WTLBP-SVM (Color)	83.44	91.80			**88.02**	-	5	A1 A2 A2 A2 H V V 2 3 1 3 2 1 2 R R B B G B B
WTLBP-kNN (GS)	88.08	40.43	60.00	69.74	**72.16**	4	3	A2 A2 A3 A3 H H V V 1 3 2 3 1 3 1 3 - - - - - - - -
WTLBP-kNN (Color)	87.42	53.19	78.33	67.11	**76.35**	1	10	A1 A1 A1 A3 A3 A3 H H D D 1 3 1 1 3 2 1 3 3 1 R R B R R B G G R B
WTLBP-SVM (GS)	88.74	38.30	55.00	67.11	**70.66**	-	3	A1 A1 A1 A2 A3 H V 1 2 3 2 2 3 1 - - - - - - -
WTLBP-SVM (Color)	92.05	14.89	66.67	78.95	**73.65**	-	1	A1 A2 H 1 2 1 R R B

Table 9: Experimental Results based on the WTLBP-Operator.

and led to an overall rate of 84.3 percent. In contrast to the k-NN method, the specificiy decreased by 2.0 percentage points while the sensitivity increased by 2.7 percentage points.

Reviewing the results of the four-class scheme experiments we see that the best results in case of the Bulbus image set was again achieved by k-NN with 87.5 percent. The improvement was caused by a higher classification rate of Marsh-3c (which is the smallest class). An improvement of 30.4 percentage points. In case of the SVM method, using the color features could not significantly improve the result. This is mainly caused by a decrease in classification rate of Marsh-3a and Marsh-3b. Considering the experiments focused on the Pars image sets, we observe the same property. The color features are again superior to grayscale images. The best result was gained by using the k-NN method (76.4%). This corresponds to an improvement of 4.2 percentage points compared to the grayscale experiment. This improvement can again be explained by an improved classification rate of the marsh classes. In this case Marsh-3a (12.8%) and Marsh-3b (18.3%). The rates of Marsh-0 and Marsh-3c dropped by (0.7 percentage points as well as 2.6

percentage points). The best result of the SVM method was 73.7 percent an improvement of 3.0 percentage points compared to the grayscale results. It is interesting to note that the rate of Marsh-3a is very low with a rate of only 14.9 percent.

Considering the k-values we see that the range is between 1 and 7. The average number of bins is 6.8. The relative number of used histograms is 58.8 percent (grayscale) and 75.0 percent when using color features.

6.7 General Conclusion

The results of numerous experiments indicate that the automated classification of celiac disease is feasible. In case of the two-class experiments based on images from the Bulbus, the overall classification rates were excellent (98.3 percent - WTLBP-kNN (Color)). As expected, the overall classification rates of the experiments based on the Pars images were lower. This was caused by the geometrical properties of the duodenal region as discussed in Section 2.5. The best classification rate was achieved by WTLBP-kNN (Color) with an overall rate of 89.8 percent. This is a very promising basis for further improvements. A hybrid approach that uses a cascading classification scheme seems possible. Such an approach would identify the camera position relative to the mucosa in the first step (among the two dominant perspectives). In a second step the actual classification would take place. The classification would be based on the corresponding training set of image data (which was determined by step one).

The results gained by the experiments based on the modified Marsh scheme (four-class case) back the assumption that the four-class problem is a significantly more challenging problem compared to the two-class case. The best classification rate of the Bulbus image set was 87.5 percent achieved in parallel by two operators (LBP/C (Color) and WTLBP (Color)) using the k-NN classification method. Although clearly below the classification rates of the two-class problems, this result is remarkable and also forms an excellent basis. These two results show that the power of discriminative information in case of the celiac classes is strong enough for classification based on visual features in case of the Bulbus. The classification rates of the celiac classes is balanced which indicates that no fitting towards the larger classes happened during the subset selection.

In the case of the four-class scheme, another hybrid approach towards solving the problem of favoring classes with a high number of images is imaginable. The problem could be alleviated by using a cascading classifier that applies the two-class scheme in the first step. According to the result, a three-class based classifier could then be used to perform the intra-celiac classification. Another possible option is to use a one-class based SVM classifier to decide the probability of the imagery showing an affected mucosa in the first step.

Considering the most difficult problem (Pars, four-class) the best result is 76.4 percent, achieved by WTLBP-kNN (color). Although the classification rate of the smallest class (Marsh-3a in this case) is only at 53.2 percent, the overall classification rates are stable. This is a result that also seems promising for further improvements. For example a combination of the two hybrid approaches that were mentioned seems possible.

Besides the evaluation of the competing operators we could observe that using color features could improve the results. Except for the experiments using the LBP operator. In this case the results remained stable. In case of the other experiments the improvement was significant. The best results of the experiments based on the two-class scheme improved by 1.3 percentage points (Bulbus) as well as 1.5 percentage points (Pars) by using color features. Considering the experiments based on the four-class scheme the improvement of using color features was 3.3 percentage points using images from the Bulbus and 3.9 percentage points for images from the Pars.

6.8 Operator Comparison

When working towards a robust classification method for a real world system for diagnosing celiac disease, care has to be taken when evaluating an optimal operator or operator combination.

- **Overall Classification Rate**
  The overall classification rate is the classic measure for evaluating the classification method's and implicitly the feature extraction's effectiveness. However the overall classification rate can be misleading in cer-

tain cases. Considering the experiments based on the modified Marsh classification scheme (four-classes), the used set of images is rather imbalanced in terms of the number of images associated with each specific class. This can lead to an interesting phenomenon. The class with the highest number of images contributes most to the overall classification rate. This means that intentionally decreasing the rate of the smallest class in favor of the largest class leads to an improvement of the overall classification rate. In general, decreasing the rate of a smaller class in favor of a larger class leads to an improvement. Therefore the specific rates of each class should also be taken into consideration when evaluating an optimal operator.

- **Statistical Significance of Results**
  Several operators use variable values for thresholding or binning which are optimized for the specific problem. As presented in Section 6 even small variations of these values can have a serious impact to the results. A robust operator with low sensitivity to varying values (at least within a certain interval) is desired.

- **Computational Performance**
  The computational performance of the used methods play a crucial role within a real world system as the physician can not wait for more than a few seconds (at maximum) for a results. The main part of the computational complexity of each classification method is the training phase and the feature subset selection. This can be precomputed in a real world system and leads to fast classification. The other part is the feature extraction. The LBP can be implemented very efficiently. As the algorithmic differences in the applied operators is rather small, the computational complexity is close among each operator. Not all operators compute the same number of histograms however. The wavelet based operator (WTLBP) as well as the gradient based operators (ELBP and ELTP) are additionally based on an intermediate transform stage of the input data. These transformations also influence the overall performance.

Table 10 illustrates the computation time for the entire feature extraction from a single input sample (128×128 pixel-region). Please note that the LBP, LTP and LBP/C operators are entirely implemented in C. The wavelet transform as well as the Sobel-filtering are implemented in Matlab however. Therefore this performance estimation is slightly biased towards the operators without an intermediate transformation stage. The measurement was performed on an AMD-3600 CPU with 2 Gigabyte of RAM running on Windows XP. The measurement was conducted twice extracting features from 289 image samples. The calculation of the estimated time for the extraction of features from a single sample was based on the arithmetic mean of both measurements.

Operator	**Computation Time**
LBP	486 ms
LTP	526 ms
LBP/C	689 ms
ELBP	1134 ms
ELTP	1237 ms
WTLBP	2309 ms

Table 10: Computational Performance of the Operators.

Table 11 lists the best results in case of the two-class scheme ranked by their overall classification rate. If two operators achieved an equal overall classification rate, the operator with the higher sensitivity value is listed first. If an operator reached exactly the same result in both the grayscale and color experiments (this is the case for LBP/C), the experiment with the smaller feature vector (corresponding to the selected subset of histograms) is presented.

When analyzing the results, we have to consider that we are dealing with overall classification rates well above 95 percent in case of the Bulbus and rates above 85 percent in case of the Pars images. The two-class problem is not as challenging in terms of classification rates as dealing with four-classes. Therefore large improvements are rather unlikely to be encountered.

Considering Table 11 we can see that the basic LBP operator is an excellent basis for the classification of celiac disease in the two-class case. The basic operator reaches results of 95.5 percent (Bulbus, two-class) as well as 85.33 percent (Pars, two-class). An interesting property of the LBP operator is, that this is the only operator that did not yield a considerable improvement by using color features. Although the classification results are respectable, there is still room for improvement. The LBP operator ranks fifth and sixth out of the six operators in terms of overall classification rate.

Method	Classification Rates				Method	Classification Rates		
	No-Celiac	Celiac	Total			No-Celiac	Celiac	Total
Bulbus					**Pars Descendens**			
WTLBP-kNN (Color)	98.79	97.58	**98.27**		WTLBP-kNN (Color)	88.74	90.71	**89.82**
ELTP-kNN (Color)	97.58	96.77	**97.23**		LTP-SVM (Color)	86.09	91.80	**89.22**
LTP-SVM (Color)	97.58	95.97	**96.89**		LBP/C-SVM (Color)	87.42	90.71	**89.22**
LBP/C-SVM (GS)	98.18	94.35	**96.54**		ELTP-SVM (Color)	87.42	90.16	**88.92**
LBP-SVM (GS)	96.97	93.55	**95.50**		ELBP-SVM (Color)	83.44	91.80	**88.02**
ELBP-kNN (Color)	95.15	95.16	**95.16**		LBP-kNN (GS)	84.11	86.34	**85.33**

Table 11: Best Results ranked by Overall Classification Rate.

The first possible improvement is applying a Sobel filter to the input data. We can see that using the extended LBP (ELBP) operator for feature extraction could not improve the results of the Bulbus images however. The best result was 95.2 percent which is the lowest result in case of the Bulbus overall. Considering the classification results based on the Pars images, the classification features extracted by the ELBP operator could improve the results compared to the LBP operator. The best result is 88.0 percent. The improvement compared to the LBP operator was caused by an increased sensitivity of 6.1 percentage points. Although the result is close to the best results, it is only ranked fifth. Overall we can see that using the ELBP operator could improve the results slightly in case of the Pars compared to the LBP operator.

The LTP operator uses thresholding to cope with noise present within the images. The best result was reached by using color features with an overall classification rate of 96.9 percent. It is interesting to note, that both SVM and k-NN reached exactly the same result. Compared to the LBP operators this is an improvement of 1.4 percentage points. Compared to the other operator this result is ranked third. In case of the experiments based on the Pars images the best result reached 89.2 percent. This is a considerable improvement compared to the LBP operator by 3.9 percentage points. This result is ranked second among all operators. We can see that features from the LTP operator were able to improve the results compared to the LBP and ELBP operators. This is an indication that the thresholding is beneficial to the quality of the extracted features.

The extended LTP (ELTP) operator uses thresholding in combination with applying a Sobel filter to the input data. The best results using the ELTP features was 97.2 percent. The same overall classification rate was reached by the k-NN and SVM method. Compared to the LBP operator this is an improvement of 1.7 percentage points. Compared to the LTP operator, the result could only be slightly improved by 0.3 percentage points. Among all the other operators the ELTP reached the 2nd rank for the classification rate of the Bulbus images. Considering the best classification rate of the Pars images, the result of the ELTP operator is 88.9 percent. This result is superior to the results of the LBP and ELBP operators but is lower than the corresponding classification rate of the LTP operator (by 0.3 percentage points). We can see that the improvements compared to the LBP and ELBP operators is clear, but there is no clear advantage over the LTP operator. The ELTP operator ranks only fourth in case of the Pars. Considering the results of the ELBP and ELTP operators in relation to the LBP and LTP operators, this indicates that using gradient filtering does not necessarily improve the discriminative information in case of the two-class problem.

The LBP/C operator uses a complementary contrast measure to improve the discriminative power of the extracted features. In case of the Bulbus the best result is 96.5 percent (achieved by k-NN and SVM). This is an improvement compared to the LBP and ELBP operator of over 1 percentage point but is slightly lower than LTP and ELTP. The LBP/C ranks in the mid field as fourth. Considering the Pars-experiments the best result is 89.2 percent which is equal to the LTP and above the LBP, ELBP and ELTP operators. The LBP/C operator ranks third among the operators due to a lower sensitivity compared to the LTP operator. We can see that the LBP/C operator is a solid operator. The variation of results is lower compared to the LTP operator. The results achieved by the LBP and ELBP operators could be improved. Overall the performance of the LBP/C operator is close to the LTP and ELTP operators.

The WTLBP operators uses the LTP and LBP/C operator combined with the wavelet transform for feature extraction. In case of the Bulbus images the best result is 98.3 percent. This is a considerable improvement of over 1 percentage point compared to ELTP (the second best). Compared to the LBP operator the improvement is 2.8 percentage points. The WTLBP operator was able to improve both the sensitivity and the specificity and is therefore ranked first with a considerable margin in case of the Bulbus results. Considering the best result of the Pars-experiments the WTLBP operator reaches a classification rate of 89.8 percent. This is an improvement of 4.5 percentage points compared to the LBP operator and an improvement of 0.6 percentage points compared to the second ranked LTP operator.

Table 12 lists the best results ranked by the overall classification rate for the experiments based on the modified Marsh scheme with four-classes. In opposition to the two-class problem, the four-class scheme is a much more challenging problem. The base classification rates are considerably lower (approximately 83 percent in case of the Bulbus and slightly below 70 percent in case of the Pars) images. Besides the overall classification rate, the rates of the celiac classes is important. This is caused by an effect that the feature optimization can favor a class with a higher number of images at the expense of a class with a low image count (see Section 6.8). Therefore both the overall classification rates and the rates of the specific celiac classes have to be considered.

Method	**Classification Rates**				
	Marsh-0	Marsh-3a	Marsh-3b	Marsh-3c	Total
Bulbus					
WTLBP-kNN (Color)	95.76	74.47	81.48	69.57	**87.54**
LBP/C-kNN (Color)	96.97	70.21	81.48	69.57	**87.54**
LTP-kNN (Color)	96.97	72.34	83.33	56.52	**87.20**
ELTP-kNN (Color)	97.58	72.34	79.63	52.17	**86.51**
LBP-SVM (Color)	96.97	59.57	74.07	56.52	**83.39**
ELBP-kNN (Color)	94.55	70.21	70.37	52.17	**82.70**
Pars Descendens					
WTLBP-kNN (Color)	87.42	53.19	78.33	67.11	**76.35**
LTP-SVM (Color)	91.39	34.04	71.67	72.37	**75.45**
ELTP-kNN (Color)	90.07	42.55	65.00	72.37	**74.85**
LBP/C-kNN (Color)	78.81	51.06	73.33	80.26	**74.25**
ELBP-SVM (Color)	90.07	19.15	63.33	76.32	**72.16**
LBP-SVM (GS)	89.40	36.17	56.67	65.79	**70.66**

Table 12: Best Results ranked by Overall Classification Rate.

Reviewing the results of the LBP operator we see that the four-class scheme is indeed much more challenging than the two-class problem. All classification rates are considerably lower. This might be caused by a lower degree of discriminative power in terms of textural properties. We can see that the LBP operator achieves a classification rate of 83.4 percent in case of the Bulbus image set. The classification rates of Marsh-3a and Marsh-3c are below 60 percent. This result is ranked fifth among all operators. In case of the Pars-experiments the best classification rate is 70.7 percent. It is interesting that this result is approximately 3 percentage points higher than the other LBP based results within the Pars-experiments. The classification rate of the smallest class (Marsh-3a) is only 36.2 percent. The LBP operator ranks last in case of classifying the Pars image set. This is comparable to the ranking in the two-class case.

The extended LBP (ELBP) operator again could not improve the results of the Bulbus-experiments compared to the results achieved by the LBP operator. It is interesting that all results except that of k-NN (color) are below the 80 percent mark. The best result with 82.7 percent is also below the LBP operator. The decrease in overall performance can be explained by a decrease in classification rate of the Marsh-3c class. Also the Marsh-3b class shows a decreased classification rate. This might indicate that the gradient filtering without thresholding is not able to capture the discriminative information of these two classes. The Pars-experiments of the ELBP operator are better compared to the results of the LBP operator, this is again in parallel to the two-class problem. The best result was achieved by the SVM method using color features and reached a classification rate of 72.2 percent. The classification rate of Marsh-3a reached a very low 19.2 percent. The ELBP operator is ranked fifth among the operators in case of the Pars.

Considering the results of the LTP operator shows that using thresholding considerably improved the classification rate of the Bulbus images. The best result is 87.2 percent which is an improvement of 3.8 percentage points compared to the LBP operator. The classification rates of the celiac classes is rather stable with a good 56.5 percent of the smallest class (Marsh-3c in this case). The LTP operator is ranked third among all operators in case of the Bulbus. Considering the results of the experiments based on the Pars images, another improvement can be seen. The improvement compared to the best result of the LBP operator is 4.8 percentage points with an overall rate of 75.5 percent. The rates of Marsh-3c (72.4%) and Marsh-3b (71.7%) are very good. Nevertheless class Marsh-3a is classified by a rate of only 34.0 percent. The LTP operator ranks second in case of the Pars experiments.

The results achieved by using the extended LTP (ELTP) operator shows as well as the ELBP operator that using gradient information does not improve the classification rate in case of the four-class scheme. The best result of the Bulbus-experiments is 86.5 percent which is 0.7 percentage points below the LTP operator. This is caused by a decrease of classification rate of classes Marsh-3b and Marsh-3c. This happened in case of the ELBP and is another indication that the gradient filtering has a negative influence to the discriminative information regarding these two classes. The better overall performance compared to the ELBP operator is probably caused by the noise removal through the applied thresholding. This manifests in a higher specificity of 3.0 percentage points. The overall classification rate of the Bulbus-experiments is ranked fourth compared to the other operators. The best result in case of the Pars images is 74.9 percent. This is also lower than the best results of the LTP operator. The rates of the celiac classes is comparable between the two operators though. Compared to the ELBP and LBP operators this is an improvement that is mainly caused by a much better classification rate of Marsh-3a with 42.6 percent (which is the class with the lowest number of images in this case). Overall the ELTP operator is ranked third considering the best result of the Pars-experiments.

The LBP/C operator achieved an overall classification rate of 87.5 percent in case of the experiments based on the Bulbus. Compared to the LBP and ELBP operators this is a serious improvement of more than 4 percentage points. The improvement can be explained by a considerable improvement of the classification rate of all celiac classes. Generally all classes show very stable classification rates. The LBP/C operator is ranked second in case of the Bulbus. Please note that this is caused only by a lower classification rate of Marsh-3a compared to the WTLBP operator. The overall classification rates of those two operators are equal. Considering the experiments using images from the Pars set the LBP/C operator could also improve the classification rate compared to the LBP (3.6%) and ELBP (2.1%) operators. The result is inferior compared to the threshold based operators. This is caused by a considerably lower specificity. This is interesting and could come into use when applying a cascading classifier based explicitly on celiac classes. In the overall ranking of the Pars-experiments, the LBP/C is placed only fourth.

The WTLBP operator reaches the same overall classification rate as the LBP/C operator within the Bulbus-experiments. Only a higher classification rate of Marsh-3a of 74.5 percent (this is 4.3 percentage points above the rate of the LBP/C operator) led to the better ranking. Compared to the LBP operator the overall improvement is 4.2 percentage points. The

WTLBP ranks first considering the Bulbus-experiments. In case of classifying images from the Pars sets the best result is 76.4 percent. Compared to the LBP operator this improves the best result by 5.7 percentage points. It is important to note that this result is not influenced by a high specificity. The classification rates of the celiac classes are very promising (Marsh-3a - 53.2%, Marsh-3b - 78.3%, Marsh-3c - 67.1%), only LBP/C has a higher rate in Marsh-3c. Compared to all operators the WTLBP ranks first with a margin of 0.9 percentage points compared to the second best result (LTP).

6.8.1 Classification Method Comparison

By comparing the best overall classification rates the specific classification methods were not taken into account. This is dangerous as the observed operator-properties might be influenced by the classification method. Therefore Table 13 and Table 14 present the best results grouped by the used classification method.

Method	**Classification Rates**				**Method**	**Classification Rates**		
	No-Celiac	Celiac	Total			No-Celiac	Celiac	Total
Bulbus (k-NN)					**Bulbus (SVM)**			
WTLBP-kNN (Color)	98.79	97.58	**98.27**		WTLBP-SVM (Color)	97.58	97.58	**97.58**
ELTP-kNN (Color)	97.58	96.77	**97.23**		ELTP-SVM (Color)	98.18	95.97	**97.23**
LTP-kNN (Color)	97.58	95.97	**96.89**		LTP-SVM (Color)	97.58	95.97	**96.89**
LBP/C-kNN (Color)	96.36	95.97	**96.19**		LBP/C-SVM (GS)	98.18	94.35	**96.54**
ELBP-kNN (Color)	95.15	95.16	**95.16**		LBP-SVM (GS)	96.97	93.55	**95.50**
LBP-kNN (Color)	98.18	87.90	**93.77**		ELBP-SVM (Color)	95.76	92.74	**94.46**
Pars Descendens (k-NN)					**Pars Descendens (SVM)**			
WTLBP-kNN (Color)	88.74	90.71	**89.82**		LBP/C-SVM (Color)	87.42	90.71	**89.22**
ELTP-kNN (Color)	86.75	89.62	**88.32**		LTP-SVM (Color)	86.09	91.80	**89.22**
LBP/C-kNN (Color)	88.08	87.98	**88.02**		ELTP-SVM (Color)	87.42	90.16	**88.92**
LTP-kNN (Color)	86.09	88.52	**87.43**		WTLBP-SVM (Color)	83.44	91.80	**88.02**
ELBP-kNN (Color)	82.78	89.62	**86.53**		ELBP-SVM (Color)	83.44	91.80	**88.02**
LBP-kNN (Color)	78.81	88.52	**84.13**		LBP-SVM (Color)	82.78	85.79	**84.43**

Table 13: Best Results grouped by Classification Method and ranked by Overall Classification Rate.

In both classification schemes (two-class as well as four-class), we can see that the general trends are consistent among the classification methods. The LBP/C operator behaves well and is among the top feature extraction methods. The threshold based operators (LTP and ELTP) perform better than the LBP and ELBP operators. The Sobel-filtering in general can not improve the classification rates in case of the k-NN method as well as the SVM classifier. The rates of the wavelet based LBP (WTLBP) operator are superior in case of the k-NN classifier compared to the SVM classification method. Overall we can observe that the rates drop slightly when using the SVM classification method.

Method	**Classification Rates**				
	Marsh-0	Marsh-3a	Marsh-3b	Marsh-3c	Total
Bulbus (k-NN)					
WTLBP-kNN (Color)	95.76	74.47	81.48	69.57	**87.54**
LBP/C-kNN (Color)	96.97	70.21	81.48	69.57	**87.54**
LTP-kNN (Color)	96.97	72.34	83.33	56.52	**87.20**
ELTP-kNN (Color)	97.58	72.34	79.63	52.17	**86.51**
LBP-kNN (Color)	95.76	65.96	72.22	52.17	**83.04**
ELBP-kNN (Color)	94.55	70.21	70.37	52.17	**82.70**
Bulbus (SVM)					
LBP/C-SVM (Color)	96.36	78.72	70.37	69.57	**86.51**
LTP-SVM (Color)	98.18	61.70	70.37	69.57	**84.78**
WTLBP-SVM (Color)	97.58	61.70	66.67	78.26	**84.43**
ELTP-SVM (Color)	99.39	76.60	62.96	39.13	**84.08**
LBP-SVM (Color)	96.97	59.57	74.07	56.52	**83.39**
ELBP-SVM (Color)	96.36	61.70	53.70	30.43	**77.51**
Pars Descendens (k-NN)					
WTLBP-kNN (Color)	87.42	53.19	78.33	67.11	**76.35**
LBP/C-kNN (Color)	78.81	51.06	73.33	80.26	**74.25**
ELTP-kNN (Color)	90.07	42.55	65.00	72.37	**74.85**
LTP-kNN (Color)	81.46	38.30	73.33	78.95	**73.35**
ELBP-kNN (Color)	89.40	21.28	60.00	73.68	**70.96**
LBP-kNN (Color)	86.09	19.15	56.67	68.42	**67.37**
Pars Descendens (SVM)					
LTP-SVM (Color)	91.39	34.04	71.67	72.37	**75.45**
LBP/C-SVM (Color)	91.39	21.28	68.33	76.32	**73.95**
ELTP-SVM (Color)	92.72	8.51	71.67	78.95	**73.95**
WTLBP-SVM (Color)	92.05	14.89	66.67	78.95	**73.65**
ELBP-SVM (Color)	90.07	19.15	63.33	76.32	**72.16**
LBP-SVM (GS)	89.40	36.17	56.67	65.79	**70.66**

Table 14: Best Results grouped by Classification Method and ranked by Overall Classification Rate.

6.8.2 Conclusion of the Operator Comparison

Considering the presented results we can see that the automated classification of celiac disease can be improved by using appropriate extensions to the feature extraction methods.

- Using threshold based operators is beneficial to the classification performance in terms of classification rates. The threshold based operators LTP and ELTP were superior to the non-threshold based operators (LBP and ELBP) in all experiments.
- Using Sobel filtering could not significantly improve the classification rates compared to the LBP and LTP operators. Even more in the case of using four classes the rates within the celiac classes dropped.
- The LBP/C operator provided promising results. In case of the four-class problems, high rates among the celiac classes could be achieved. This operator could be a candidate for a cascading classification considering only the three celiac classes.
- The best classification rates could be achieved by using the WTLBP operator. In case of the Bulbus two-class problem, 98.3 percent could be reached. This is an excellent rate towards establishing a reliable classification. Considering the Pars 89.8 percent were achieved, again by using the WTLBP operator. A promising rate that could be improved further by a cascading classification.
- The overall classification rates of the experiments based on the modified Marsh scheme using four classes could considerably be improved by deploying color features. The WTLBP operator reached the highest rates in both cases. Nevertheless there is still need for further improvements in case of the four-class problem.
- The computational complexity of the WTLBP operator is almost twice compared to the gradient based operators and approximately four times as high compared to LBP, LTP and LBP/C. The computational performance itself was not a main focus during the work on this thesis however. The bigger part of the computational complexity of all operators is caused by the bilinear interpolation. By an optimized implementation the performance of all operators could be further improved.
- The combination of LBP based operators is feasible. Even more is the combination of the wavelet decomposition with LBP operators. The results could all be improved by using the wavelet based approach.

7 Conclusion

Within this thesis numerous parameters that are optimal in terms of feature extraction and classification in the context of diagnosing celiac disease were identified. Section 2 discussed the acquisition and preparation of the used image training set that was used towards establishing a reliable classification. Conducted experiments showed that using the modified immersion technique is beneficial to the feature extraction process and consequently also to the classification. Section 5 presented the systematic approach towards identifying an optimal set of parameters for six LBP based operators. Finally in Section 6 the results of 48 experiments were presented. Additionally to that, all results that were computed during the optimization of thresholds and number of contrast bins are shown. This accommodates to a total of 148 single results (ignoring the results of experiments used to find the optimal parameters). Besides employing the well known LBP modification, a new combined operator based on the wavelet decomposition was introduced. This operator combined the beneficial properties of the wavelet transform with a set of suited LBP operators to improve the feature extraction process.

The presented experiments showed that automated classification of celiac disease is possible. Even in case of the more challenging modified Marsh scheme (four-classes) very promising results could be achieved. The wavelet based operator (WTLBP) was found to be the best among all classification problems. This shows that combining LBP based operators with the wavelet transform is reasonable. Within this thesis a solid basis for the further improvement of the more challenging problems in diagnosing of celiac disease was built in terms of features extraction methods and parameters as well as classification methods.

Some possible improvements were not addressed in this thesis. The following points could improve the performance of the specific methods in future work even further:

7.1 Possible Future Improvements

- The computational complexity of the gradient based operators (ELBP and ELTP) as well as the wavelet based operator (WTLBP) is still rather high. Towards the deployment of a real world system the computational performance should be optimized. This could be done, for example, by optimizing or changing the interpolation mechanism.

- The distance measures used for classification are meant to measure the overlap of two probability distributions. In all experiments however, the normalization process implicitly assumes uniformly distributed histograms. It is most certainly the case that the histograms are uniformly distributed. Maybe a better fitting probability distribution could be found.

- The quantization into contrast bins (LBP/C) is based on estimating the distribution of the contrast values. This estimation was performed using the grayscale images. When using color features however, this estimation could be based on the separate color channels to improve the discriminative power of the contrast measure.

- Signal noise is affected by low pass filtering. This should be taken into account when using threshold based operators.

- The LBP/C operator does not have a noise reduction mechanism such as the threshold based operators. An LTP/C operator is imaginable and could perform well in the context of classifying celiac disease.

- The SVM classifier uses a radial basis function as kernel. This corresponds to a geometric interpretation of the feature vectors which might not be optimal. Barla et al. ([4]) proved that the Histogram Intersection has the mathematical properties to be used as a kernel function for Support Vector Machines. Therefore the Histogram Intersection could also be applied when using the SVM classification method.

Future work towards the deployment of a real world system for classification of celiac disease has to face various additional problems. Some of them include automated segmentation for an optimal automated region extraction and improvements of the classification rates in case of using the modified Marsh scheme.

References

[1] L. Alexandre, N. Nobre, and J. Casteleiro. Color and position versus texture features for endoscopic polyp detection. In *Proceedings of the International Conference on BioMedical Engineering and Informatics, 2008 (BMEI'08)*, volume 2, pages 38–42, Sanya, Hainan, China, May 2008.

[2] S. Ameling, S. Wirth, D.Paulus, G. Lacey, and F. Vilarino. Texture-based polyp detection in colonoscopy. In *Bildverarbeitung für die Medizin 2009*, number 15 in Informatik aktuell, pages 346–350. Springer Berlin, June 2009.

[3] D.A. Antonioli. Celiac disease: a progress report. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc*, 16(4):342–6, 2003.

[4] A. Barla, F. Odone, and A. Verri. Histogram intersection kernel for image classification. *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, 3:III–513–16 vol.2, 2003.

[5] A. Bhattacharyya. Bulletin of the calcutta mathematical society. *Journal of Artificial Intelligence Research*, 35:99–109, 1943.

[6] B. E. Boser. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152. ACM Press, 1992.

[7] E. Bradley and R. Tibshirani. Improvements on cross-validation: The .632+ bootstrap method. *Journal of the American Statistical Association*, 92(438):548–560, 1997.

[8] C. S. Burrus, R. A. Gopinath, and H. Guo. *Introduction to Wavelets and Wavelet Transforms: A Primer*. Prentice-Hall, 1998.

[9] G. Cammarota, P. Cesaro, and A. Martino. High accuracy and cost-effectiveness of a biopsy-avoiding endoscopic approach in diagnosing coeliac disease. *Alimentary Pharmacology and Therapeutics*, 23(1):61–69, January 2006.

[10] G. Cammarota, A. Martino, and G.A. Pirozzi. Direct visualization of intestinal villi by high-resolution magnifying upper endoscopy: a validation study. *Gastrointestinal Endoscopy*, 60(5):732–738, 2004.

[11] T. Cover and P. Hart. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967.

[12] I. Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF Reg. Conf. Series in Applied Math. SIAM, Philadelphia, PA, 1992.

[13] P.A. Devijver and J.V. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice Hall, 1982.

[14] W. Dickey and D. Hughes. Prevalence of celiac disease and its endoscopic markers among patients having routine upper gastrointestinal endoscopy. *American Journal of Gastroenterology*, 94:2182–2186, August 1999.

[15] T. G. Dietterich and G. Bakiri. Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2:263–286, 1995.

[16] K. Duan and S.S. Keerthi. Which is the best multiclass svm method? an empirical study. In *Proceedings of the Sixth International Workshop on Multiple Classifier Systems*, pages 278–285, 2005.

[17] A. Fasano, I. Berti, T. Gerarduzzi, T. Not, R. B. Colletti, S. Drago, Y. Elitsur, P.H. R Green, S. Guandalini, I.D. Hill, M. Pietzak, A. Ventura, M. Thorpe, D. Kryszak, F. Fornaroli, S. S. Wasserman, J.A. Murray, and K. Horvath. Prevalence of celiac disease in at-risk and not-at-risk groups in the united states: a large multicenter study. *Archives of internal medicine*, 163:286–92, February 2003.

[18] A. Gasbarrini, V. Ojetti, L. Cuoco, G. Cammarota, A. Migneco, A. Armuzzi, P. Pola, and G. Gasbarrini. Lack of endoscopic visualization of intestinal villi with the immersion technique in overt atrophic celiac disease. *Gastrointestinal endoscopy*, 57:348–351, March 2003.

[19] A. Graps. An introduction to wavelets. *IEEE Computational Science and Engineering*, 2:50–61, 1995.

[20] M. Haefner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Combining gaussian markov random fields with the discrete wavelet transform for endoscopic image classification. In *Proceedings of the 17th International Conference on Digital Signal Processing (DSP'09)*, Santorini, Greece, 2009. to appear.

[21] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Pit pattern classification using multichannel features and multiclassification. In D.I. Fotiadis T.P. Exarchos, A. Papadopoulos, editor, *Handbook of Research on Advanced Techniques in Diagnostic Imaging and Biomedical Applications*, pages 335–350. IGI Global, Hershey, PA, USA, 2009.

[22] Haralick, Dinstein, and Shanmugan. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3:610–621, 1973.

[23] S. Hegenbart, R. Kwitt, M. Liedlgruber, A. Uhl, and A. Vécsei. Impact of duodenal image capturing techniques and duodenal regions on the performance of automated diagnosis of celiac disease. In *Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis, ISPA '09*, Salzburg, Austria, September 2009.

[24] X. Huang, S.Z. Li, and Y. Wang. Shape localization based on statistical method using extended local binary pattern. *Image and Graphics, International Conference on*, 0:184–187, 2004.

[25] A. Jain and D. Zongker. Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:153–158, 1997.

[26] H. Jin, Q. Liu, H. Lu, and X. Tong. Face detection using improved lbp under bayesian framework. *International Conference on Image and Graphics*, 0:306–309, 2004.

[27] B. Julesz. Textons, the elements of texture perception, and their interactions. *Nature*, 290(5802):91–97, March 1981.

[28] S.A. Karkanis. Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE Transactions on Information Technology in Biomedicine*, 7(3):141–152, September 2003.

[29] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI*, pages 1137–1145, 1995.

[30] R. Kwitt and 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 (MM-BIA'07)*, Rio de Janeiro, Brasil, 2007.

[31] M. Liedlgruber and A. Uhl. Statistical and structural wavelet packet features for pit pattern classification in zoom-endoscopic colon images. In P. Dondon, V. Mladenov, S. Impedovo, and S. Cepisca, editors, *Proceedings of the 7th WSEAS International Conference on Wavelet Analysis & Multirate Systems (WAMUS'07)*, pages 147–152, Arcachon, France, October 2007.

[32] M. Liedlgruber and A. Uhl. Endoscopic image processing - an overview. In *Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis, ISPA '09*, Salzburg, Austria, September 2009.

[33] X. Liu, X. You, and Y. Cheung. Texture image retrieval using non-separable wavelets and local binary patterns. In *CIS (1)*, pages 287–291. IEEE Computer Society, 2009.

[34] T. Mäenpää. *The Local Binary Pattern Approach to Texture Analysis - Extensions and Applications*. PhD thesis, University of Oulu, 2003.

[35] T. Mäenpää;, T. Ojala, M. Pietikäinen, and M. Soriano. Robust texture classification by subsets of local binary patterns. *Pattern Recognition, International Conference on*, 3:3947, 2000.

[36] J. Malik, S. Belongie, J. Shi, and T. Leung. Textons, contours and regions: Cue integration in image segmentation. In *ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2*, page 918, Washington, DC, USA, 1999. IEEE Computer Society.

[37] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

[38] B.S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(8):837–842, 1996.

[39] M. N. Marsh. Gluten, major histocompatibility complex, and the small intestine. a molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology*, 102(1):330–54, 1992.

[40] E. Meijering. A chronology of interpolation: From ancient astronomy to modern signal and image processing. In *Proceedings of the IEEE*, pages 319–342, 2002.

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

[42] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, January 1996.

[43] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[44] K. Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to hove arisen from random sampling. *Philosophical Magazine*, 50:157–172, 1900.

[45] R. Petroniene, E. Dubcenco, and JP. Baker. Given capsule endoscopy in celiac disease: evaluation of diagnostic accuracy and interobserver agreement. *The American Journal of Gastroenterology*, 100(3):685–694, March 2005.

[46] P.Wang, S.M. Krishnan, C. Kugean, and M.P. Tjoa. Classification of endoscopic images based on texture and neural network. In *Proceedings of the 23rd IEEE Engineering in Medicine and Biology*, volume 4, pages 3691–3695, 2001.

[47] J. Rosenman, C.A. Roe, R. Cromartie, K.E. Muller, and S.M. Pizer. Portal film enhancement: Technique and clinical utility. *Int. J. Radiat. Oncol. Biol. Physics*, pages 333–338, 1993.

[48] J.L. Starck, F. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge University Press, 1998.

[49] Y. Su, D. Tao, X. Li, and X. Gao. Texture representation in aam using gabor wavelet and local binary patterns. In *SMC'09: Proceedings of the 2009 IEEE international conference on Systems, Man and Cybernetics*, pages 3274–3279, Piscataway, NJ, USA, 2009. IEEE Press.

[50] M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7:11–32, 1991.

[51] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *Analysis and Modelling of Faces and Gestures*, volume 4778 of *LNCS*, pages 168–182. Springer, October 2007.

[52] M. Tuceryan and A. Jain. *Handbook of Pattern Recognition and Computer Vision*. World Scientific, 1999.

[53] A. Vécsei, T. Fuhrmann, M. Liedlgruber, L. Brunauer, H. Payer, and A. Uhl. Automated classification of duodenal imagery in celiac disease using evolved fourier feature vectors. *Comput Methods Programs Biomed*, Apr 2009.

[54] M. Vetterli. Wavelets and filter banks for discrete-time signal processing. In *Wavelets and their applications*, pages 17–52. Jones and Bartlett, Boston, MA, 1992.

[55] J.S. Weszka, C.R. Dyer, and A. Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6:269–286, April 1976.

[56] K. Yokoi. Illumination-robust change detection using texture based features. In *MVA*, pages 487–491, 2007.

[57] K. Zuiderveld. Contrast limited adaptive histogram equalization. In Paul S. Heckbert, editor, *Graphics Gems IV*, pages 474–485. Morgan Kaufmann, 1994.