Fractal Analysis for the Viewpoint Invariant Classification of Celiac Disease

Andreas Uhl
Department of Computer Sciences
University of Salzburg
Salzburg, Austria
Email: uhl@cosy.sbg.ac.at

Andreas Vécsei
St. Anna Children's Hospital
Vienna, Austria

Georg Wimmer
Department of Computer Sciences
University of Salzburg
Salzburg, Austria
Email: gwimmer@cosy.sbg.ac.at

***Abstract*—In this paper we present a viewpoint invariant method using fractal features for the automatic classification of celiac disease. First the endoscopic images of the duodenum are pre-filtered using the MR8 filter bank. Subsequently we compute the local fractal dimension of each pixel of the filter responses. After finding representative combinations of the local fractal dimensions for images with or without celiac disease, we generate frequency histograms that count how often each representative combinations occur as nearest neighbor for each image (as nearest neighbor to the local fractal dimensions of the filter responses for each pixel position of the images). We use this histograms for the classification of the images. We will see that the result of the proposed method outperforms the approach, our method is based on, as well as other methods for the classification of medical imagery.**

I. Introduction

Celiac disease is a complex autoimmune disorder in genetically predisposed individuals of all age groups after introduction of gluten containing food.
The celiac state of the duodenum is usually determined by visual inspection during the endoscopy session followed by a biopsy of suspicious areas.

The images are divided into two classes. The class denoted as No-Celiac represents a healthy duodenum with normal crypts and villi whereas the class denoted as Celiac represents a duodenum with mild or marked atrophy of the villi or the villi are even entirely absent.

In gastroscopic (and other types of endoscopic) imagery, mucosa texture is usually found with different perspective, distances to the texture (see Figure 1) and distortions (barrel-type distortion of the endoscope [1]). That means that the mucosal textures show different spatial scales, depending on the camera perspective and distance to the mucosal wall.
Consequently, in order to design reliable computer-aided mucosa texture classification schemes, the scale invariance of the employed feature sets could be essential.

To avoid overfitting, the images of the celiac disease database are divided into a training set and a evaluation set [2]. Table I lists the number of image samples and patients per class. The images of one patient are either all in the evaluation set or all in the training set. The images have a fixed size of 128×128 pixels.

(a) No-Celiac (b) No-Celiac

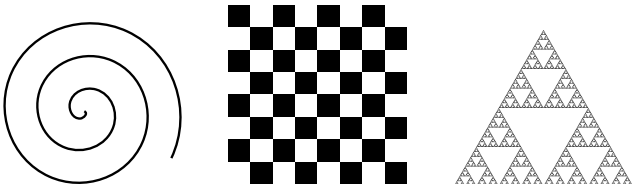
(c) Celiac (d) Celiac

Figure 1. Images of the two classes with different perspectives and zooms

Table I
NUMBER OF IMAGE SAMPLES AND PATIENTS PER CLASS (GROUND TRUTH BASED ON HISTOLOGY)

Data set	Training set		Evaluation set	
	No-Celiac	Celiac	No-Celiac	Celiac
Number of images	155	157	151	149
Number of patients	66	21	65	19

We decided to handle the problem, that images have different scales, perspectives and distortions by means of fractal analysis. Our proposed method is based on a approach of Varma et al. [**?**], which is a combination of a former approach of Varma et al. [**?**] and a approach of Xu et al. [**?**]. Fircist the images are filtered using the MR8 filterbank like in [**?**], then we compute the local densities (see [**?**]), a measure that is similar to the fractal dimension, of each pixel of the 8 filter responses of an image. The local density is invariant under the bi-Lipschitz map, which includes view-point changes and non-rigid deformations of texture surface as well as local



(a) (b) (c)

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

affine illumination changes. We now have an 8-dimensional viewpoint invariant local density vector for every pixel of an image. Typically filter responses are found by separate clustering (k-Means) the local density vectors of the 2 classes of the training set. The cluster centers of the 2 classes are denoted as texton. These textons are used to label the local density vectors of each image of the evaluation and training set. The histogram of texton frequencies is used to form models of the images with which the images are classified.
The difference between the just now explained approach of Varma et al. [**?**] and our proposed approach is, that we use a much more contrast sensitive way to compute the local densities. We will see that our proposed approach leads to clearly better classification results than the approach of Varma et al. it is based on.

II. Fractal Features

Before we describe the proposed fractal features and discuss their properties we want to review fractal theory and how it is applicable for our scenario. Simplified, a perfect fractal is a shape which appears similar at all scales of magnification. Because of that property, a perfect fractal can be decomposed into N copies of itself, all downscaled by a factor of δ, which tile the original shape exactly. N and δ are often related by a power law, i.e. $N \propto \delta^{-D}$, where D is the fractal dimension. Many natural phenomena satisfy this power law (for example the length of a coast).
For the case of a point set E defined on $\mathbb{R}^2$, the fractal dimension of E is defined as

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

where $N(\delta, E)$ is the smallest number of sets with diameter less than δ that cover E. The set is made up of closed disks of radius δ or squares of side length δ. In Figure 2 we see some examples for the fractal dimensions of different objects.

Let μ be a finite Borel regular measure on $\mathbb{R}^2$. For $x \in \mathbb{R}^2$, denote $B(x, r)$ as the closed disk with center x and radius $r > 0$. $\mu(B(x, r))$ is considered as an exponential function of r, i.e. $\mu(B(x, r)) = c\, r^{D(x)}$, where $D(x)$ is the density function and c is some constant. We can see that if $\mu(B(x, r))$ converges to 0 then r converges to 0 as well. The local density function (or also called local fractal dimension) of x is defined as

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

The local density function describes how the measurement μ satisfies the power law behavior at the measured point. The density function measures the "non-uniformness" of the intensity distribution in the region neighboring the measured point. Figure 3 shows three images and their corresponding local densities at the center point, where the intensities are defined on the range [1,256] with black=1 and white=256. The faster $\mu(B(x, r))$ is increasing for an increasing r, the higher the local density is.

In the caption of Figure 3 we see the local densities of the center points for two different measures. One measure $\mu(B(x, r))$ only sums up the intensity values inside the disc $B(x, r)$ (the values in brackets), the other measure extracts the edge information of an image (the values before brackets).

Unlike in equation (4) and (3), the local densities in the examples of Figure 3 are computed on the image itself instead of computing them on the filter responses of the image.

(a) (b) (c)

Figure 3. Local densities at the center points, where $\mu(B(x, r))$ is defined like in equation (4) (respectively (3)) on page 3: (a) 3.8 (2.2), (b) 0 (2.4) (c) 3.8 (5.0)

In practical computation, we compute the local density $D(x)$ for each pixel of an image by linear fitting the slope of the line of $\log \mu(B(x, r)$ against $\log r$ with $r \in \{1, \ldots, 8\}$ (see Figure 4).
As we can see in Figure 4, the eight points lie along a straight line, indicating that the power law is being followed faithfully.

The local density D is invariant under the bi-Lipschitz map, which includes view-point changes and non-rigid deformations of texture surface as well as local affine illumination changes. A bi-Lipschitz function f must be invertible and satisfy the constraint $c_1||x - y|| \leq ||f(x) - f(y)|| \leq c_2||x - y||$ for $c_2 \geq c_1 > 0$.
Theorem: $D(f(x)) = D(x)$, where D is the local density function and f denotes a bi-Lipschitz transform on the spatial variable x.
Proof: See [**?**].

III. Feature Extraction

Most fractal methods are based on measures calculated over filter responses. Our approach also computes fractal features (the local density function) over filter responses and is based on the method of [**?**].

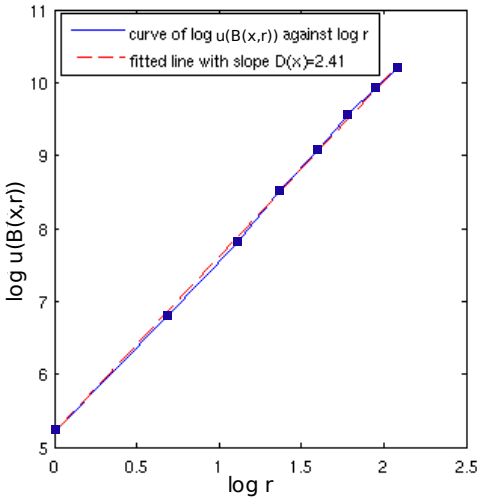


Figure 4. Linear fitting the slope of the line of $\log \mu(B(x,r)$ against $\log r$

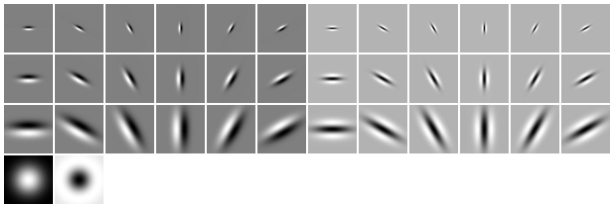


Figure 5. The MR8 filter bank

The filters we are using have the advantage that they smooth image noise and lead to more robust features. The drawback of them is that they lower the bi-Lipschitz invariance. However, in our case, we empirically verified that using filters leads to a clearly better classification performance.

We begin with generating multiple measures by first pre-filtering the images using MR8 filter [**?**]. The MR8 filter bank is a rotationally invariant, nonlinear filterbank with 38 filters but only 8 filter responses. It contains edge and bar filters, each at 6 orientations and 3 scales, as well as a rotationally symmetric Gaussian and Laplacian of Gaussian filter (see Figure 5). Rotational invariance of the edge and bar filters is achieved by only taking the maximum response over all orientations at each level of scale. In the approach [**?**], the results are better if only 5 of the 8 filter responses are used (skipping the Gaussian filter, the medium scale bar filter and the smallest edge filter), but in case of the celiac disease database it turned out that the results are better if all 8 responses are used. We used anisotropic Gaussian filtering [3], a fast implementation for filtering the images with the MR8 filter bank. Before the filtering process, the images are normalized, so that the mean of the intensity values over an image is zero and the standard deviation is one. Hence our method is invariant to shifts in the illumination intensity.

Now we compute the local density function of each pixel of the eight filter responses from the MR8 filter bank. Instead of the measure

$$\mu(B(x,r)_i) = \int_{B(x,r)} |f(i)|\ dx \tag{3}$$

used in the approach of [**?**], we propose the measure

$$\mu(B(x,r)_i) = \int_{B(x,r)} | \quad S_1 \quad *(G_r * f(i)) + \quad S_2 \quad *(G_r * f(i))|\ dx. \tag{4}$$

$f(i)$ denotes the i-th filter response image with $1 \leq i \leq 8$,

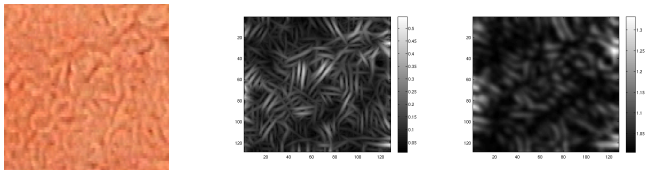
$$S_1 = \frac{1}{4}\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$

and $S_2 = -S_1^T$ are Sobel filters, $x \in \mathbb{R}^2$, '$*$' denotes the convolution and G is a smoothing kernel with variance r:

$$G(r) = \frac{1}{r\sqrt{2\pi}} \exp \frac{-||x||^2}{r^2} \tag{5}$$

In other words, the approach of [**?**] (equation 3) sums up the intensity values inside the disc B(x,r). Our proposed approach (equation 4) blurs the image depending on r (the bigger r, the stronger the blurring). The two Sobel filter S_1 and S_2 are similar to directional differential operators along the horizontal and vertical direction and primarly provides information about the edges of an image. By the convolution of the blurred image with the Sobel filters we are able to extract the contrast information of the image. So our proposed measure is much more contrast sensitive than the measure used in approach [**?**].

Figure 6 demonstrates a No-Celiac image the filter response of the image and the corresponding local densities of the filter response. As filter response we used the edge filter of the MR8 filter bank with the medium scale. The measure $\mu(B(x,r))$ used in Figure 6 is our proposed contrast sensitive measure defined in equation (4).



(a) (b) (c)

Figure 6. (a) Endoscopic image, (b) the corresponding edge filter response (medium scale) from the MR8 filter bank and (c) the local densities of the filter response, where $\mu(B(x,r))$ is defined like in equation (4)

With these measures (equation 3 and 4) we can compute the local density for each of the 8 filter responses. For each pixel of an image we now have a 8-dimensional local density vector. Now we aggregate all local density vectors of each image of a class and find cluster centres by k-means clustering. These cluster centres are the representative local density vectors of a class and are known as textons (see Figure 7).

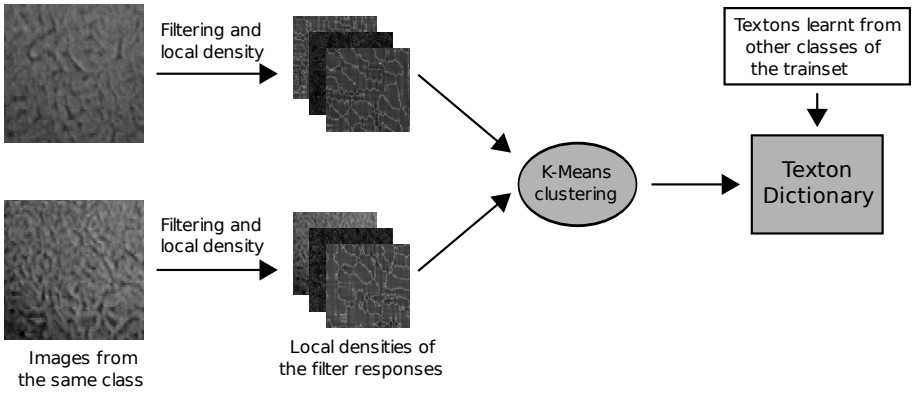


Figure 7. Learning the texton dictionary by the images of the training set.

Like in the approach of [**?**], we take 10 textons per class (that are 20 textons for our 2-classes of the celiac disease database). All the textons together build the dictionary.
The next step is to learn models for each class. Given a training image, its corresponding model (frequency histogram) is generated by first convolving it with the filter bank, computing the local density of each filter response and then labeling each local density vector with the texton that lies closest to it (see Figure 8).

In the classification stage, we repeat the procedure of the last step to build the histogram corresponding to a novel image of the evaluation set. Then we compare the new histogram with the models learnt during the training and classify it on the basis of the comparison. Distances between two frequency histograms are measured using the χ^2 statistic (see Figure 9), which is defined by

$$\chi^2(x,y) = \sum_i \frac{(x_i - y_i)^2}{x_i + y_i} \tag{6}$$

IV. Results

We employ two methods to evaluate and compare the feature sets described in the section before: The area under the ROC curve (AUC) [4] and the overall classification accuracy (OCR).
To generate the ROC curve by means of a k-NN classifier, we used the method described in [4, 5] (for k=20). We consider the 20 nearest neighbors of each image of the evaluation set (20 images from the training set that have the smallest χ^2-distance to the image from the evaluation set). We get the i'th point on the ROC curve with $i \in \{0, \ldots, 21\}$ by classifiying an images as positive, if i or more than i nearest neighbors (out of the 20 nearest neighbors) of the considered image are positive (that means that they belong to the class Celiac). In this way the range of the true positive rate (TPR) (= sensitivity) and the false positive rate (FPR) (= 1 - specificity) is exactly between zero and one (zero for $i = 21$, one for $i = 0$). In Figure 10 we see the ROC curves of the methods listed in Table II.

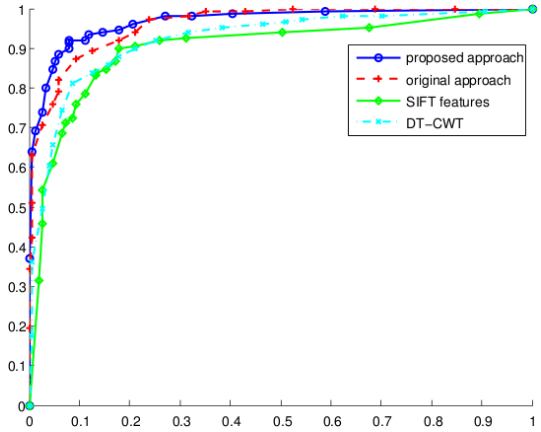


Figure 10. ROC curves of the methods listed in Table II

The AUC is computed by trapezoidal integration,

$$AUC = \sum_{i=1}^{21} \left(TPR_i \cdot \Delta FPR_i + \frac{\Delta TPR_i \cdot \Delta FPR_i}{2} \right) \tag{7}$$

where

$$\Delta TPR_i = TPR_{i-1} - TPR_i, \tag{8}$$
$$\Delta FPR_i = FPR_{i-1} - FPR_i, \tag{9}$$

and TPR_i or FPR_i are those TPR or FPR, where at least i positive nearest neighbors are necessary to classify an image as positive.

For the evaluation of the overall classification accuracy we used the k-NN classifier . The k for the k-NN classifier, used to classify the evaluation set, is optimized based on the training set.

In Table II we see the results of our method and of the approach our method is based on (original approach). To compare these results with some other methods that are often applied for medical image classification, we additionally implemented two more methods:

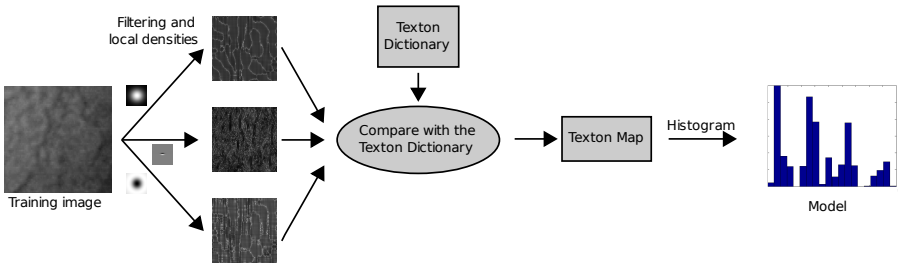


Figure 8. Learning the model of a given image from the training set.

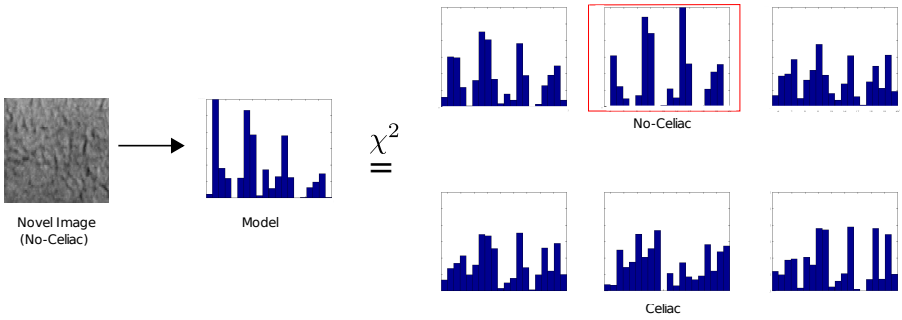


Figure 9. Classification of an image from the evaluation set

Table II
OVERALL CLASSIFICATION ACCURACY (OCR) AND AREA UNDER CURVE (AUC) (BOTH IN PERCENT) OF OUR METHOD COMPARED TO OTHER METHODS

Method	Evaluation	
	OCR	AUC
Proposed method	91.7 ± 2.0	95.0 ±0.5
Original approach	84.7 ± 2.0	93.5±0.5
Dense SIFT features	83.6 ±2.3	87.5±1.9
DT-CWT	84.7	90.2

The first method is a applying of the widely used scale-invariant feature transform (SIFT) [**?**]. We used the VLFEAT implementation (see VLFeat.org) of dense SIFT features, which extracts SIFT features for each pixel. In this way we get a 128 dimensional local SIFT feature vector for each pixel. We act with these local SIFT feature vectors in the same way than with the local density vectors (building a dictionary and then building models for each image).
The second method is the Dual-tree Complex Wavelet Transform [6], which performs well for the classification of celiac disease [7, 5].

Due to the k-Means clustering, the proposed method, the original approach [**?**] and the dense SIFT features provides different results each run. The given results in the table are the means of ten results per method.

As we can see in Table II our proposed method works very good for the classification of celiac disease. The proposed method is distinctly better than the approach it is based on (the original approach). Our method is also clearly better than the other two methods.

V. DISCUSSION

Our proposed method achieves better results as compared to the original approach. This is because our method uses a measure $\mu((B(x,r))$, that is much more contrast sensitive as compared to the measure of the original approach. Since celiac disease causes a decrease of visible mucosal structures, celiac images usually show less contrast as compared to images showing a healthy mucosa.
Our proposed method and the original approach are scale and rotation invariant, whereas the SIFT features are only scale invariant and the DT-CWT features are neither scale nor rotation invariant.
The images of the celiac disease database have different orientation and scale and so our method and the original approach have a big advantage and that is one of the reasons why they perform better as compared to the other two methods.

REFERENCES

[1] M. Gschwandtner, M. Liedlgruber, A. Uhl, and A. Vécsei. Experimental study on the impact of endoscope distortion correction on computer-assisted celiac disease diagnosis. In *Proceedings of the 10th International Conference on Information Technology and Applications in Biomedicine (ITAB'10)*, pages 1–6, Corfu, Greece, November 2010.

[2] Sebastian Hegenbart, Andreas Uhl, and Andreas Véscei. Systematic assessment of performance prediction techniques in medical image classification - a case study on celiac disease. In *Proceedings of the 22nd International Conference on Information Processing in Medical Imaging (IPMI'11)*, Monastery Irsee, Germany, July 2011. Accepted.

[3] J. M. Geusebroek, A. W. M. Smeulders, and J. van de Weijer. Fast anisotropic gauss filtering. *IEEE Transactions on Image Processing*, 12(8):938–943, 2003.

[4] A. P. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145 – 1159, 1997.

[5] A. Vécsei A. Uhl and G. Wimmer. Complex wavelet transform variants in a scale invariant classification of celiac disease. In *Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis*, Las Palmas de Gran Canaria. Spain, June 2011. accepted.

[6] I.W. Selesnick, R.G. Baraniuk, and N.C. Kingsbury. The dual-tree complex wavelet transform. *Signal Processing Magazine, IEEE*, 22(6):123 – 151, nov. 2005.

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