

Classification of Confocal Laser Endomicroscopic Images of the Oral Cavity to Distinguish Pathological from Healthy Tissue

Christian Jaremenko¹, Andreas Maier^{1,2}, Stefan Steidl¹, Joachim Hornegger¹, Nicolai Oetter³, Christian Knipfer³, Florian Stelzle³, Helmut Neumann⁴

¹Pattern Recognition Lab, Department of Computer Science, FAU
Erlangen-Nürnberg, Martenstr. 3, 91058 Erlangen, Germany

²SAOT Graduate School in Advanced Optical Technologies

³Department of Oral and Maxillofacial Surgery, University Hospital Erlangen

⁴Department of Medicine I, University Hospital Erlangen

`christian.jaremenko@fau.de`

Abstract. Confocal laser endomicroscopy is a recently introduced advanced imaging technique which enables microscopic imaging of the mucosa in-vivo. This technique has already been applied successfully during diagnosis of gastrointestinal diseases. Whereas for this purpose several computer aided diagnosis approaches exist, we present a classification system that is able to differentiate between healthy and pathological images of the oral cavity. Varying textural features of small rectangular regions are evaluated using random forests and support vector machines. Preliminary results reach up to 99.2% classification rate. This indicates that an automatic classification system to differentiate between healthy and pathological mucosa of the oral cavity is feasible.

1 Introduction

Cancer of the oral cavity and pharynx can only be reliably diagnosed by performing biopsies of lesions. These biopsies may lead to unfavorable sequelae such as permanent damage in sensitive areas. Another difficulty is finding an adequate resection margin of the primary tumor site as intraoperative surveillance of the resection margin is complex since it includes histological analysis. Both treatments are influenced by the experience of the physician and may introduce operator dependent errors.

Confocal laser endomicroscopy (CLE) allows intraoperative real time visualization of the mucosa as an optical biopsy with en face view, whereas the histological analysis of biopsies investigates a transverse section through the tissue. To obtain a high contrast visualization of the mucosa, a contrast agent such as fluorescein has to be administered. The resulting images allow a microscopic comparison between inconspicuous epithelium and neoplastic lesions by visualizing the different structural and architectural compositions of cells. The image quality can be enhanced by band-pass filtering [1] to reduce noise and emphasize cell boundaries.

Currently CLE has been successfully used in different diseases such as Ulcerative colitis or Crohn's disease [2]. CLE seems to be a promising technique when it comes to intraoperative analysis of the mucosa and has been rarely used for diagnosis of cancer in the oral cavity e.g. [3]. When it comes to computer aided diagnosis only a few approaches exist, all handling the gastric intestinal tract based on analyzing the arrangement of crypts e.g. [4] by using the Scale-Invariant-Feature-Transform (SIFT). Another approach using SIFT is accomplished by Mualla et al. [5] for automatic cell detection in bright-field microscope images.

This paper investigates whether it is possible to classify the acquired images into two stages *pathological* and *healthy* using textural features and pattern recognition methods.

2 Materials and methods

Three subjects, two patients and one healthy control, were examined using a probe based CLE (pCLE) system from Cellvizio (GastroFlex/UHD, Mauna Kea Technologies, Paris, France) that can be applied via the accessory channel of an endoscope. IRB approval was provided and procedures were performed after written informed consent was obtained from the patients. The 16 bit grayscale images with a resolution of 576×578 are streamed with a frame rate of 12 frames per seconds. The resulting CLE video sequences are separated into n images per second and annotated individually by experts of the Oral and Maxillofacial Surgery. Sequences of the alveolar ridge were acquired and examined leading to a database consisting of 206 images with no pathological characteristics and 45 images being cancerous (Fig. 1). The image on the left hand side visualizes the healthy oral mucosa with flat, uniform, polygonal cells. These have a regular architectural arrangement with alternating dark and light bands. The image on the right hand side shows a cancerous region highlighted by the rectangle. The architectural arrangement of cells is completely lost and cells have varying sizes. Besides this, cell piles with no architectural structures and strong

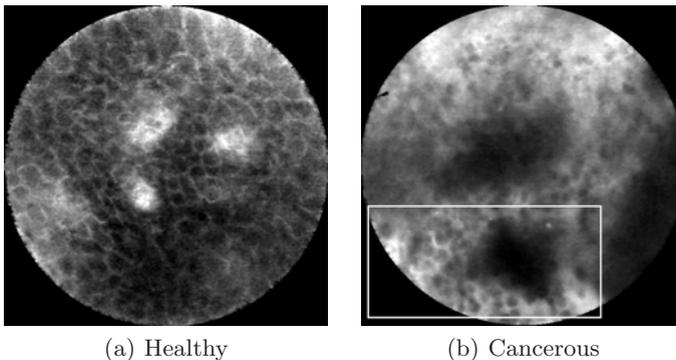


Fig. 1. Exemplary pCLE images with different image characteristics.

(white) fluorescein leakage prevail. The circular shape of the pCLE images is a challenge, as the shape itself or its large dark areas may introduce artefacts during the processing of images. Instead of cropping the image to a quadratic region inside the circle and losing large parts of image information, multiple small rectangular regions of the image are extracted and examined separately. To find the coordinates of these patches, the intensity values of their corners are compared with a threshold which is delineated by the dark background. The step length in x and y direction consists of half the edge length leading to an overlap of 50%. Two different patch sizes with an edge length of 80 and 105 pixels are examined, so that a certain amount of structure is comprised within each patch and a symmetrical distribution of patches is accomplished as emphasized by Fig. 2b. This leads to a diverse amount of 110 and 52 processed patches, respectively. Fig. 2 shows a single patch with edge length of 105 pixels as well as the resulting gradually processed image.

To accomplish a classification into images with healthy and pathological characteristics, four different features are computed for every patch:

- *Histogram features*: These features describe the local distribution of gray level values. First order statistics, like mean, standard deviation, coefficient of variance, skewness, kurtosis, and entropy are computed from the histogram to describe a local patch.
- *Homogeneity features*: As images with no pathological characteristics appear more structured than pathological ones, the amount of included edges within a patch may imply the presence or absence of cancerous regions. For this reason the image is filtered with a Sobel operator. As homogeneity features, the mean, standard deviation and variance are computed from the edge image, as well as of the gray level image.
- *Gray level co-occurrence matrices (GLCM)*: Haralick et al. [6] propose different indicators to describe an image texture by numerical characteristics called Gray-Level Co-Occurrence Matrices Features. These features are based on a statistical analysis of the frequency of certain gray levels in dependence of their geometrical arrangement to each other. In this study, a

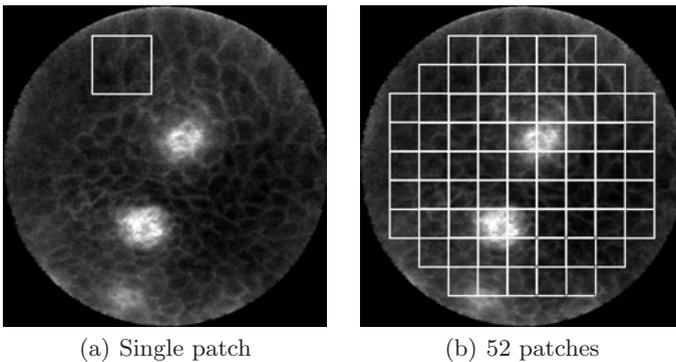


Fig. 2. Division of images into patches.

distance of one and two between the pixels as well as the orientations of 0° , 45° , 90° , 135° are investigated. This leads to a feature vector including the angular second moment, contrast, correlation, variance, inverse difference moment, entropy, energy, homogeneity and inertia. In addition the recursivity and inverse recursivity introduced by Beraldi et al. [7] are included.

- *Local binary patterns (LBP)*: LBP is another approach to describe images by using the distribution of binary patterns in the image [8]. For every pixel a binary pattern is constructed by thresholding the center pixel with its circular neighborhood. If the neighboring pixel has a larger value it is set to one and when connected with the results of all neighboring pixels this leads to a local binary pattern. This approach has been extended to uniform rotation independent local binary patterns. In this version, LBP are bitwise rotated circularly to their minimum code. A uniform pattern is defined by having ≤ 2 1 to 0 or 0 to 1 transitions and represents a certain structure. For instance, the LBP 00001111_2 denotes an edge, and 00000111_2 denotes a corner. These binary patterns are sampled into a histogram describing the image. If a pattern is non-uniform, the reject bin comprising of all non-uniform patterns is assigned. In this study the classical LBP, as well as the rotation invariant uniform LBP are evaluated. Instead of using the histogram itself, the already mentioned first order statistics are computed and used as features.

These features are combined during the processing of the patches leading to two different characteristic vectors. The first one is specified by a concatenation of the features of every patch representing the whole image while preserving local image information. The second feature vector consists of the average (mean, std. deviation and variance) of all features and patches describing the mean or global representation of the patches.

On the previously introduced features, a support vector machine (SVM) [9] classifier as well as a random forest (RF) [10] classifier are trained. Within the SVM approach, a separation between the classes is realized by transforming the feature space to a higher dimension. The optimal hyperplane is then constructed using the critical points next to it as support vectors by solving a maximization problem to increase the margin using a polynomial kernel function. The other approach being applied is the RF. This classifier constructs many unpruned decision trees and includes bagging as well as random selection of features. These random combinations of input data lead to a reduction of correlation between trees. Furthermore, the influence of strong features as chosen predictors is reduced as every feature vector can contain duplicate values or miss features. The final class of the examined feature vector is then assigned by the majority votes of the particular decision trees.

3 Results

The SVM and RF classifier are implemented using WEKA, with a 10 fold cross-validation approach as a result of the small image database. In addition, different

Table 1. Accuracy (acc) and average recall (rec) of the particular features of the concatenated feature vector (Patchsize 105).

Features	Property	SVM		RF	
		Acc	Rec	Acc	Rec
Histogram	256 bins	88.5%	70.4%	90.4%	78.6%
Histogram	512 bins	89.3%	72.7%	92.4%	81.5%
Histogram	768 bins	89.2%	72.6%	89.6%	77.2%
Homogeneity	—	90.8%	77.1%	96.4%	91.7%
GLCM	8 Imglvl	99.2%	97.8%	92.8%	79.8%
GLCM	16 Imglvl	99.2%	97.8%	90.8%	74.5%
GLCM	32 Imglvl	98.8%	97.6%	90.0%	73.1%
LBPc	R1 N8	86.1%	71.6%	89.6%	73.5%
LBP _r	R1 N8	85.7%	73.9%	87.3%	65.3%
LBP _r	R2 N16	83.7%	65.8%	92.0%	83.0%

settings of the feature types are investigated. For the histogram features varying amounts of bins (256, 512, 768) are used. The GLCM matrices are evaluated with 8, 16, 32 gray value quantization levels. The classical LBP is evaluated with a radius of one and a neighborhood of 8. In case of the rotation invariant uniform LBP, a radius of one and two as well as a neighborhood of 8 and 16 are evaluated.

The recognition rate (acc) as well as the average recall (rec) for the two different investigated patch sizes are illustrated in Tabs. 1 and 2 in case of the non-averaged feature vector. The corresponding results of the averaged feature vector are shown in Tabs. 3 and 4. In both cases the classification rate is illustrated in dependance of the underlying feature types and its adjusted properties. Overall the GLCM based classification reaches the best results with a recognition rate of 99.2% and average recall of 97.8% in case of the SVM with 80px patchsize in the non-averaged classification problem. For the average feature vector a classification rate of 98.0% and an average recall of 95.3% is reached in case of the 80px patchsize using the RF classifier.

4 Discussion and conclusion

The GLCM features achieve the best classification result (99.2%) without strong dependance on the patchsize as they deliver nearly equal classification results. Though the larger patchsize strongly reduces computation time. Even in case of the averaged feature vector, a strong accuracy of 98.0% is reached. This feature vector may be of advantage when a classification between different stages of the disease into hyperplasia, dysplasia or carcinoma is conducted, especially when multiple disease characteristics are apparent on a single image. Besides this, the simple homogeneity features reach a strong accuracy of up to 96.8%. This indicates that the amount of included edges gives insight about the state of

Table 2. Accuracy (acc) and average recall (rec) of the particular features of the concatenated feature vector (Patchsize 80).

Features	Property	SVM		RF	
		Acc	Rec	Acc	Rec
Histogram	256 bins	88.4%	74.7%	92.4%	82.4%
Histogram	512 bins	88.8%	75.0%	91.6%	79.3%
Histogram	768 bins	88.5%	74.7%	92.4%	81.5%
Homogeneity	—	94.8%	87.3%	95.2%	86.7%
GLCM	8 Imglvl	99.2%	98.7%	90.8%	74.5%
GLCM	16 Imglvl	99.2%	98.7%	90.0%	72.0%
GLCM	32 Imglvl	99.2%	98.7%	90.4%	74.2%
LBPc	R1 N8	89.6%	79.8%	86.1%	63.7%
LBPr	R1 N8	88.8%	77.6%	88.4%	74.7%
LBPr	R2 N16	83.3%	65.5%	91.2%	88.5%

Table 3. Accuracy (acc) and average recall (rec) of the particular features of the averaged feature vector (Patchsize 105).

Features	Property	SVM		RF	
		Acc	Rec	Acc	Rec
Histogram	256 bins	82.1%	50.0%	91.2%	85.2%
Histogram	512 bins	82.1%	50.0%	92.0%	83.0%
Histogram	768 bins	82.1%	50.0%	91.2%	83.3%
Homogeneity	—	82.1%	50.0%	94.0%	87.7%
GLCM	8 Imglvl	96.0%	91.5%	96.0%	91.5%
GLCM	16 Imglvl	96.0%	91.5%	95.2%	90.2%
GLCM	32 Imglvl	96.0%	90.6%	96.4%	91.7%
LBPc	R1 N8	84.1%	60.8%	86.9%	74.7%
LBPr	R1 N8	82.1%	50.0%	86.9%	70.3%
LBPr	R2 N16	84.5%	60.2%	90.8%	78.8%

disease. The main drawback of this study is the small patient database and the consequent small image database. Besides this, only preselected images without a high level of noise or image artefacts are investigated, which would decrease the classification process significantly. In addition, only the basic problem, the differentiation between pathological and healthy images is conducted, which leads to the overall very optimistic classification rate. Nevertheless, these results show that a classification in pathological and healthy images can be accomplished with a high accuracy of 99.2%. As the clinical trial is ongoing, the image database is currently increased and in future work a discrimination between different disease stages will be investigated. Therefore, our proposed algorithm may help to improve real time diagnosis of suspicions lesions in the oral cavity to guide subsequent therapy.

Table 4. Accuracy (acc) and average recall (rec) of the particular features of the averaged feature vector (Patchsize 80).

Features	Property	SVM		RF	
		Acc	Rec	Acc	Rec
Histogram	256 bins	82.1%	50.0%	92.0%	83.9%
Histogram	512 bins	82.1%	50.0%	90.8%	84.5%
Histogram	768 bins	82.1%	50.0%	91.2%	83.4%
Homogeneity	—	82.1%	50.0%	93.2%	84.6%
GLCM	8 Imglvl	97.2%	94.8%	95.6%	90.4%
GLCM	16 Imglvl	97.2%	95.7%	94.4%	97.3%
GLCM	32 Imglvl	96.0%	92.4%	98.0%	95.3%
LBPc	R1 N8	82.1%	50.0%	86.1%	70.7%
LBP _r	R1 N8	84.9%	64.7%	87.7%	76.9%
LBP _r	R2 N16	86.1%	64.6%	90.0%	78.3%

References

1. Mualla F, Schöll S, Bohr C, et al. Epithelial cell detection in endomicroscopy images of the vocal folds. *Springer Proc Phys.* 2014;154:201–5.
2. Neumann H, Vieth M, Langner C, et al. Cancer risk in IBD: how to diagnose and how to manage DALM and ALM. *World J Gastroenterol.* 2011;17(27):3184–91.
3. Volgger V, Conderman C, Betz CS. Confocal laser endomicroscopy in head and neck cancer: steps forward? *Curr Opin Otolaryngol Head Neck Surg.* 2013;21(2):164–70.
4. Couceiro S, Barreto P, Freire P, et al. Machine learning in medical imaging: description and classification of confocal endomicroscopic images for the automatic diagnosis of inflammatory bowel disease. *Lect Notes Comput Sci.* 2012;7588:144–51.
5. Mualla F, Schöll S, Sommerfeldt B, et al. Automatic cell detection in bright-field microscope images using SIFT, random forests, and hierarchical clustering. *IEEE Trans Med Imaging.* 2013;32(12):2274–86.
6. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern.* 1973;3(6):610–21.
7. Baraldi A, Parmiggiani F. An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Trans Geosci Remote Sens.* 1995;33(2):293–304.
8. Pietikainen M, Hadid A, Zhao G, et al. *Computer Vision Using Local Binary Patterns.* vol. 40 of *Computational Imaging and Vision.* Springer, London; 2011.
9. Platt J. Sequential minimal optimization: a fast algorithm for training support vector machines. Microsoft Research; 1998.
10. Breiman L. Random forests. *Mach Learn.* 2001; p. 5–32.