Endoscope Distortion Correction Does Not (Easily) Improve Mucosa-Based Classification of Celiac Disease*

Jutta Hämmerle-Uhl[1], Yvonne Höller[1], Andreas Uhl[1], and Andreas Vécsei[2]

[1] Department of Computer Sciences
University of Salzburg, Austria
[2] St.Anna Children's Hospital, Dept. Pediatrics
Medical University, Vienna
andreas.uhl@sbg.ac.at

Abstract. Distortion correction is applied to endoscopic duodenal imagery to improve automated classification of celiac disease affected mucosa patches. In a set of six edge- and shape-related feature extraction techniques, only a single one is able to consistently benefit from distortion correction, while for others, even a decrease of classification accuracy is observed. Different types of distortion correction do not lead to significantly different behaviour in the observed application scenario.

Keywords: endoscope distortion correction, shape- and edge-based features, celiac disease, automated classification.

1 Introduction

Computer-aided decision support systems relying on automated analysis of endoscopic imagery receive increasing attention [1].

A specific type of degradation, present in all endoscopic images, is a barrel-type distortion. This type of degradation is caused by the wide-angle (fish eye) nature of the optics used in endoscopes.

The aim of correcting this distortion in endoscopy is manifold. Barrel type distortion is claimed to affect diagnosis [2], since it introduces nonlinear changes in the image, due to which the outer areas of the image look significantly smaller than their actual size. Therefore, the estimation of area or perimeter of observed lesions can be significantly incorrect depending on the position in the image [3]. In a recent study [4] it has been demonstrated, that in classification of celiac disease based on duodenal images, in fact misclassification cases can be related to the extent of barrel distortion of the texture patches involved in classification. Using the same image material, the impact of distortion correction on classification accuracy has been investigated [5], [4]. In these studies it turned out that most feature extraction methods considered failed to take advantage of applying distortion correction as a pre-processing step to the endoscopic images, resulting in an even decreased classification accuracy. It has been suspected that these unexpected results might be due to the (i) (too) simple distortion correction technique applied. The only feature extraction techniques exhibiting improved classification

* This work has been partially supported by the Austrian Science Fund project no. 24366.

N. Ayache et al. (Eds.): MICCAI 2012, Part III, LNCS 7512, pp. 574–581, 2012.

when applied to distortion corrected images were based on edges and geometrical features [5]. Therefore, it was also speculated that in general, (ii) edge and shape-related feature types would be able to benefit from distortion correction.

In this work we focus on those two conjectures (i) and (ii) stated as conclusions after result analysis in the mentioned studies. First, we employ the more recent parameter-free distortion correction approach of Hartley and Kang [6]. Second, we use a set of features related to edge and shape information instead of the mostly texture-oriented descriptors in [5,4]. Further contributions of this work are the usage of a more realistic evaluation protocol for classification assessment (leave-one-patient-out (LOPO) cross validation) and the application of a richer set of classifiers to avoid bias due to the use of a single classifier.

The manuscript is structured as follows. Section 2 explains the background of applying duodenal mucosa texture classification for diagnosis and staging of celiac disease and describes the image database used with the corresponding histological ground truth. In Section 3, we describe the experimental setup by first explaining the distortion correction techniques and their respective application to our image test and subsequently, by reviewing the feature extraction (and classification) techniques employed. Section 4 presents and discusses experimental results and in Section 5 we finally conclude this work.

2 Classification of Duodenal Texture for Celiac Disease Diagnosis

Celiac disease, commonly known as gluten intolerance, is a complex autoimmune disorder that affects the small bowel in genetically predisposed individuals of all age groups after introduction of food containing gluten. Endoscopy with biopsy is currently considered the gold standard for the diagnosis of celiac disease. During standard upper endoscopy at least four duodenal biopsies are taken. Microscopic changes within these specimen are then classified in a histological analysis according to the Marsh classification. The modified Marsh classification [7] distinguishes between classes Marsh-0 to Marsh-3, with subclasses Marsh-3a, Marsh-3b, and Marsh-3c, resulting in a total number of six classes. An automated system identifying areas affected by celiac disease in the duodenum can help to improve biopsy reliability (by indicating areas eventually affected by celiac disease), can aid to improve less invasive diagnosis techniques avoiding biopsies, and can reduce the costs of interpreting video material captured during capsule endoscopy [7]. Prior approaches dealing with the computer-aided diagnosis of celiac disease using endoscopic still images images include feature extraction based on Local Binary Pattern based operators, band-pass type Fourier filters, histogram and wavelet-transform based features, as well as smoothness/sharpness measures [7]. Techniques involving temporal information computed from video-capsule endoscopy have been also described [8].

The image test set used in this work (see [7] for example images) stems from three pediatric gastroscopes without magnification, types GIF-Q165 and GIF-N180, Olympus, with two of the first type and one of the latter type, respectively. The patients presented in the pediatric Department because of celiac-like symptoms. Diagnostic evaluation was indicated because of dyspeptic symptoms, positive celiac serology,

anaemia, malabsorption syndromes, inflammatory bowel disease, and gastrointestinal bleeding. For the endoscopy, the modified immersion technique was applied. This procedure is based on the instillation of water into the duodenal lumen for better visibility of the villi. Then, the tip of the gastroscope is inserted into the water in order to take images of meaningful regions. The images were taken from the Duodenal Bulb and the Pars Descendens. Most importantly, these regions differ by their geometric properties. Thus, it is necessary to treat these image sets separately.

From the acquired images, an experienced endoscopist extracted 128×128 pixels patches significant for diagnosis. The images and patients were pre-classified by the diagnostic outcome of the biopsy of the significant region at the hospital into the modified Marsh classification as shown in Table 1.

Table 1. Number of images/patients in the data sets and Marsh-classes

Data-Set	**Marsh-0**	**Marsh-3a**	**Marsh-3b**	**Marsh-3c**
Bulbus	163/60	47/8	54/8	23/8
Pars Descencens	141/72	47/10	60/8	72/12

3 Experimental Study

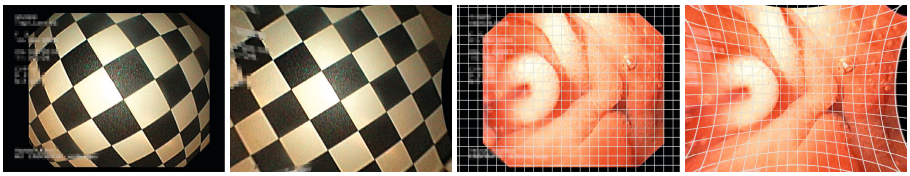
3.1 Distortion Correction

The following stages are applied to the entire endoscopic images before the extraction of the texture patches used for classification. Each colour image has been transformed into a grayscale image with the usual conversion formula and subsequently, the MATLAB built-in function for histogram equalisation (flat histogram) has been applied.

We use a planar checkerboard pattern (with points on a known grid) for distortion calibration (see Fig. 1.a). Fig. 1.b shows an example of a distortion corrected calibration pattern. The first distortion correction technique applies the MATLAB software developed by J.-Y. Bouguet[1]. For each gastroscope, 10 images were chosen to extract calibration points and distorted points. The algorithm to extract the grid corners of a checkerboard requires clicking on the four extreme corners of the rectangular checkerboard pattern. The calibration images that were used in this study contain only parts of a checkerboard, so that for each image some sensible area of the checkerboard was decided to contain the most extreme corners. Since the distortion in this study is quite significant, providing a manual estimation for radial distortion was required for all images. After computing the intrinsic and extrinsic camera parameters, the undistort-tool as provided in the tool-box is used (which is also used for computing the centres of the texture patches as mentioned below).

Barreto et al. [9] have found the parameter-free approach of Hartley and Kang [6] being better suited for endoscopic imagery as compared to Bouguet's approach – therefore, we have developed a corresponding MATLAB implementation of their technique. After a manual extraction of calibration points and their adjustment using Bouguet's

[1] `http://www.vision.caltech.edu/bouguetj/calib_doc/`



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

Fig. 1. Distortion correction applied to checkerboard (taken with Olympus GIF-Q165) and to entire endoscopic image (Bouguet distortion correction applied)

tool, the implementation of Peter Kovesi[2] was applied to all images to calculate the fundamental matrix. The reminder of the algorithm is implemented as described in the paper, partially using MATLAB built-in functions for e.g. optimisation ("fmincon").

Since after distortion correction the squared texture patches using for classification do no longer correspond to squares (see Fig. 1.c and 1.d) these cannot be used immediately for subsequent classification (most feature extraction techniques implicitly assume at least a rectangularly shaped texture patch). Therefore we apply the following technique to generate square-shaped texture from distortion corrected image material: Based on the original (distorted) endoscopic images, we record the coordinates of the centre of the extracted 128×128 pixels. Subsequently, distortion correction is applied to the entire original images and the recorded centre coordinates are mapped into the distortion corrected image. Using these coordinates, a 128×128 pixels texture square is extracted from the distortion corrected image which is then used for classification.

3.2 Feature Extraction and Classification

To be able to assess the impact of distortion correction techniques on the classification accuracy, we use a set of different feature extraction techniques. Contrasting to earlier studies, emphasis is given to edge- and shape-related strategies.

Fractal Dimension: Boxcounting [10]: A texture signature is computed from binary images obtained from original images using different thresholds and application of the box-counting fractal dimension on each thresholded image. In our implementation, all gray-level thresholds from 50 to 175 were used to generate binary images, for box-counting, Moisy's tool[3] is used with box-sizes from 2^2 - 2^{32}. For the final signature, the mean and standard deviation over the values for the box sizes are used for each threshold value.

Locally Invariant Fractal Features [11]: Local fractal dimension (also termed local density) is computed for each pixel in an image after applying the MR8 filterbank. For each class, the 8-dimensional local density vectors of all training images are aggregated and subjected to k-means clustering resulting in cluster centres termed textons. For an image to be classified, local density vectors are computed and each one is labelled with

[2] www.csse.uwa.edu.au/~pk/research/matlabfns/
[3] www.fast.u-psud.fr/~moisy/ml/

the texton that is closest to it. The frequency histograms of the texton occurrences are used as feature descriptors. Custom MATLAB code is developed for this approach.

Gray-Level Co-occurrence Matrix (GLCM [12]): The GLCM is defined over an image as the distribution of co-occurring values at a given offset Δx and Δy for a $n \times m$ image as follows:

$$C_{\Delta x,\Delta y}(i,j) = \sum_{p=1}^{n}\sum_{q=1}^{m} \begin{cases} 1, & if\ I(p,q) = i\ and\ I(p+\Delta x, q+\Delta y) = j \\ 0, & otherwise \end{cases} \quad (1)$$

For classification, the Haralick features contrast, correlation, energy, and homogeneity are used for offset-values 1,2,4, and 8 in 4 directions (vertical, horizontal, diagonal 45° and 135°). Calculations are performed with the MATLAB built-in function "graycomatrix".

Edge Co-occurrence Matrix (ECM [13]): For this approach (custom MATLAB implementation), a Sobel edge detection approach using the Robinson compass masks is applied to the images. Subsequently, a co-occurrence matrix is constructed using 1,2,3, and 4 as distances in 8 directions. Again, the Haralick features computed from this matrix as mentioned above are used as feature vector.

Edge Orientation Histogram (EH [14]): The EH is one of three MPEG-7 texture descriptors. The EH requires dividing the image into 4×4 sub-images, where each sub-image is sub-divided again into blocks of typically 4×4 pixels. The EH finds vertical, horizontal, diagonal and non-directional edges. This makes the EH specifically well suited for natural images with non-uniform edge distribution. Each image block is filtered to obtain the most prominent edge in the block. If the block is monotone, no edge is counted. As a consequence, a histogram with 5 bins can be computed over all the image blocks in each of the 16 sub-images. Thus, this results in an 80 bin histogram which is computed using the implementation of OConaire[4].

Spatial Size Distributions (SSD [15]): The difference of the autocorrelation for a given image and the autocorrelation of the same image after applying a morphological opening is computed. To be able to capture texture properties of different sizes, these computations are performed using scaled versions of the structure element used. The results obtained for the different structure element sizes are then summed up and normalised by the square of the sum over all grayscale values within the image. The result is a cumulative distribution function. The probability density associated with this cumulative distribution is called a spatial size distribution. The features which are then used to classify textures are obtained by computing first-order and second-order moments of the probability density.

While the shape variants of the structuring element have been chosen in accordance to the original suggestion, the number of scales of the structuring element and the

[4] clickdamage.com/sourcecode/index.php

number of different support disc sizes have been reduced to 4 and 2 in our MATLAB implementation, respectively, to limit the very demanding computations.

Classification: To avoid over-fitting phenomena the leave-one-patient-out cross-validation protocol is used to estimate classification accuracy. In each validation run, MATLAB built in classifiers are used: Discriminant analysis with a diagonal quadratic function, knn classification (with Euclidean distance metric and $k = 1$), and classification with SVM using a linear or a quadratic kernel function, respectively. Classification is performed separately for the two topographical regions of the duodenum. From the resulting classes that are assigned to the images, the performance of the classification was evaluated by calculating sensitivity, specificity and overall accuracy.

3.3 Experimental Results

In Tables 2 - 4 we display classification results. Configurations where results obtained from distortion corrected material are superior in terms of overall accuracy are marked in bold face.

Table 2. Classification performance for fractal dimension-based feature extraction

Fractal dimension: Boxcounting					
class.	**region**	**images**	**Sens.**	**Spec.**	**Acc.**
diag. quad.	**Bulbus**	**distorted**	.51	.66	.59
		Hartley	.58	.82	**.72**
		Bouguet	.46	.86	**.69**
	Pars Desc	**distorted**	.46	.52	.49
		Hartley	.51	.65	**.57**
		Bouguet	.66	.57	**.62**
knn	**Bulbus**	**distorted**	.40	.75	.60
		Hartley	.48	.79	**.66**
		Bouguet	.85	.65	**.65**
	Pars Desc	**distorted**	.47	.57	.52
		Hartley	.50	.57	**.53**
		Bouguet	.45	.59	.51
SVM-linear	**Bulbus**	**distorted**	.38	.72	.57
		Hartley	.52	.79	**.67**
		Bouguet	.53	.82	**.69**
	Pars Desc	**distorted**	.57	.43	.51
		Hartley	.63	.50	**.57**
		Bouguet	.59	.45	**.53**

Locally invariant fractal features					
class.	**region**	**images**	**Sens.**	**Spec.**	**Acc.**
diag. quad.	**Bulbus**	**distorted**	.35	.91	.67
		Hartley	.30	.93	.66
		Bouguet	.19	.96	.62
	Pars Desc	**distorted**	.64	.72	.68
		Hartley	.65	.64	.65
		Bouguet	.47	.70	.57
knn	**Bulbus**	**distorted**	.73	.65	.69
		Hartley	.71	.54	.61
		Bouguet	.56	.82	**.71**
	Pars Desc	**distorted**	.73	.51	.63
		Hartley	.56	.58	.57
		Bouguet	.55	.51	.53
SVM linear	**Bulbus**	**distorted**	.73	.90	.83
		Hartley	.59	.85	.74
		Bouguet	.70	.88	.80
	Pars Desc	**distorted**	.73	.68	.71
		Hartley	.72	.62	.68
		Bouguet	.78	.68	**.74**

The only feature extraction technique where distortion corrected images consistently lead to better results is the “fractal dimension: boxcounting” technique. Also, for edge co-occurrence features we notice a few improvements (e.g. for knn classification and Bouguet distortion correction applied to Pars Descendens images also for the other two classifiers). For the other four feature extraction techniques improvements are sParse and in the majority of configurations result degradations are observed.

There are absolutely no trends which justify the much more complicated parameter-free distortion correction as compared to Bouguets software. The tendency that Pars Descendens imagery is more difficult to classify can be confirmed with the results in this study. In most cases SVM classification delivers the best results, but there are also

Table 3. Classification performance for grayscale co-occurrence and SSD feature extraction

Gray scale Co-occurence features					
class.	**region**	**images**	**Sens.**	**Spec.**	**Acc.**
diag quad.	**Bulbus**	**distorted**	.70	.79	.75
		Hartley	.75	.76	**.76**
		Bouguet	.77	.81	**.79**
	Pars Desc	**distorted**	.74	.59	.68
		Hartley	.61	.72	.66
		Bouguet	.67	.67	.67
knn	**Bulbus**	**distorted**	.75	.79	.77
		Hartley	.62	.83	.74
		Bouguet	.64	.77	.71
	Pars Desc	**distorted**	.70	.68	.69
		Hartley	.68	.65	.67
		Bouguet	.68	.54	.62

SSD features					
class.	**region**	**images**	**Sens.**	**Spec.**	**Acc.**
diag. quad.	**Bulbus**	**distorted**	.53	.78	.67
		Hartley	.49	.79	.66
		Bouguet	.26	.91	.63
	Pars Desc	**distorted**	.57	.67	.62
		Hartley	.71	.54	**.63**
		Bouguet	.57	.71	**.63**
knn	**Bulbus**	**distorted**	.82	.83	**.83**
		Hartley	.73	.82	.78
		Bouguet	.75	.83	.79
	Pars Desc	**distorted**	.71	.67	.69
		Hartley	.71	.66	.69
		Bouguet	.65	.69	.67

Table 4. Classification performance for edge co-occurrence and edge orientation features

Edge Co-occurrence features					
class.	**region**	**images**	**Sens.**	**Spec.**	**Acc.**
diag quad.	**Bulbus**	**distorted**	.51	.67	.60
		Hartley	.59	.58	.58
		Bouguet	.51	.58	.55
	Pars Desc	**distorted**	.45	.67	.55
		Hartley	.49	.59	.52
		Bouguet	.53	.65	**.58**
knn	**Bulbus**	**distorted**	.39	.57	.49
		Hartley	.48	.55	**.52**
		Bouguet	.45	.54	**.50**
	Pars Desc	**distorted**	.53	.38	.46
		Hartley	.54	.48	**.51**
		Bouguet	.56	.48	**.53**
SVM linear	**Bulbus**	**distorted**	.68	.78	.74
		Hartley	.56	.72	.56
		Bouguet	.66	.76	.72
	Pars Desc	**distorted**	.64	.45	.55
		Hartley	.61	.40	.52
		Bouguet	.63	.49	**.57**

Edge orientation features					
class.	**region**	**images**	**Sens.**	**Spec.**	**Acc.**
diag. quad.	**Bulbus**	**distorted**	.69	.68	.68
		Hartley	.68	.72	**.70**
		Bouguet	.73	.72	**.72**
	Pars Desc	**distorted**	.49	.57	.53
		Hartley	.44	.55	.49
		Bouguet	.37	.57	.46
knn	**Bulbus**	**distorted**	.85	.45	.62
		Hartley	.72	.39	.53
		Bouguet	.81	.45	.57
	Pars Desc	**distorted**	.45	.52	.51
		Hartley	.48	.57	**.52**
		Bouguet	.58	.43	.51
SVM quadratic	**Bulbus**	**distorted**	.77	.62	.69
		Hartley	.63	.56	.59
		Bouguet	.63	.58	.60
	Pars Desc	**distorted**	.59	.55	.57
		Hartley	.46	.42	.44
		Bouguet	.62	.56	**.60**

a few exceptions. The best result with 0.83 accuracy is obtained with locally invariant fractal features and SSD on the distorted Bulbus image set using SVM classification and knn classification, respectively. Overall, classification accuracy is found to be lower for the considered set of feature descriptors as compared to transform-based or LBP-related methods.

4 Conclusion

Distortion correction does not improve classification of celiac-disease related duodenal image material in many cases, even if edge- and shape-related feature descriptors are used. The role of interpolation as used in all distortion correction techniques needs to be investigated in more detail – especially in the corner regions of the images, where distortion correction is most crucial due to the strong barrel distortion, interpolation

artefacts are most severe due to the large extent of distances to be corrected. Furthermore, the validity of the conclusions found so far needs to be checked for other types of endoscopes (e.g. high-magnification or high-definition endoscopes) and other types of classification tasks (e.g. colon polyp classification, stomach mucosa classification etc.).

References

1. Liedlgruber, M., Uhl, A.: Computer-aided decision support systems for endoscopy in the gastrointestinal tract: A review. IEEE Reviews in Biomedical Engineering 4, 73–88 (2012)
2. Borcharrt, T., Conci, A., d'Ornellas, M.: A warping based approach to correct distortions in endoscopic images. In: Proceedings of the 22nd Brazilian Symposium on Computer Graphics and Image Processing (Sibgrapi 2009), Rio de Janeiro, Brazil (October 2009)
3. Asari, K.V., Kumar, S., Radhakrishnan, D.: A new approach for nonlinear distortion correction in endoscopic images based on least squares estimation. IEEE Transactions on Medical Imaging 18(4), 345–354 (1999)
4. Liedlgruber, M., Uhl, A., Vécsei, A.: Statistical analysis of the impact of distortion (correction) on an automated classification of celiac disease. In: Proceedings of the 17th International Conference on Digital Signal Processing (DSP 2011), Corfu, Greece (July 2011)
5. Gschwandtner, M., Liedlgruber, M., Uhl, A., Vécsei, A.: 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 2010), Corfu, Greece (November 2010)
6. Hartley, R., Kang, S.: Parameter-free radial distortion correction with center of distortion estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(8), 1309–1321 (2007)
7. Vécsei, A., Amann, G., Hegenbart, S., Liedlgruber, M., Uhl, A.: Automated marsh-like classification of celiac disease in children using an optimized local texture operator. Computers in Biology and Medicine 41(6), 313–325 (2011)
8. Ciaccio, E.J., Tennyson, C.A., Lewis, S.K., Krishnareddy, S., Bhagat, G., Green, P.H.: Distinguishing patients with celiac disease by quantitative analysis of videocapsule endoscopy images. Computer Methods and Programs in Biomedicine 100(1), 39–48 (2010)
9. Barreto, J., Swaminathan, R., Roquette, J.: Non parametric distortion correction in endoscopic medical images. In: Proceedings of the 3DTV Conference 2007, Kos, Greece, pp. 1–4 (2007)
10. Backes, A.R., Bruno, O.M.: A New Approach to Estimate Fractal Dimension of Texture Images. In: Elmoataz, A., Lezoray, O., Nouboud, F., Mammass, D. (eds.) ICISP 2008. LNCS, vol. 5099, pp. 136–143. Springer, Heidelberg (2008)
11. Varma, M., Garg, R.: Locally invariant fractal features for statistical texture classification. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV 2007), Rio de Janeiro, Brazil, pp. 1–8 (2007)
12. Haralick, R., Shanmugan, K., Dinstein, I.: Textual features for image classification. IEEE Transactions on Systems, Man, and Cybernetics SMC-3, 610–621 (1973)
13. Rautkorpi, R., Iivarinen, J.: A Novel Shape Feature for Image Classification and Retrieval. In: Campilho, A., Kamel, M. (eds.) ICIAR 2004. LNCS, vol. 3211, pp. 753–760. Springer, Heidelberg (2004)
14. Manjunath, B., Salembier, P., Sikora, T.: Introduction to MPEG-7: Multimedia Content Description Interface. Wiley & Sons (2002)
15. Ayala, G., Domingo, J.: Spatial size distributions: Applications to shape and texture analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(12), 1430–1442 (2001)