Quality Based Information Fusion in Fully Automatized Celiac Disease Diagnosis

Michael Gadermayr[1], Andreas Uhl[1], and Andreas Vécsei[2]

[1] Department of Computer Sciences, University of Salzburg, Austria
[2] St. Anna Children's Hospital, Department of Pediatrics, Medical University Vienna, Austria

Abstract. Up to now, for most endoscopical computer aided celiac disease diagnosis approaches, image regions showing discriminative features have to be manually extracted by the physicians, prior to their automatized classification. This is obligatory to get idealistic and reliable data which is free from strong image degradations. On the one hand such a human interaction during endoscopy is subjective, expensive and tedious, but on the other hand state-of-the-art fully automatized selection corresponds to decreased classification accuracies compared to experienced human experts. In this work, a fully automatized approach is introduced which exploits the availability of a significant number of subimages within one original endoscopic image. A weighted decision-level and a weighted feature-level fusion method are introduced and investigated with respect to the achieved classification accuracies. The outcomes are compared with simple decision-level and feature-level fusion methods and the manual and the automatized patch selection. Finally, we show that the proposed feature-level fusion method outperforms all other automatized methods and comes close to manual patch selection.

1 Introduction

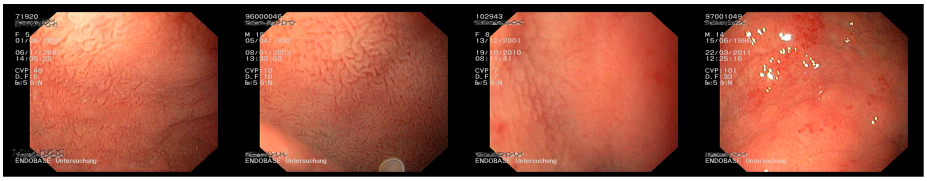
Celiac disease [15] is a disorder affecting the small bowel. After introduction of gluten containing food, the disease leads to an inflammatory reaction in the mucosa of the small intestine caused by a dysregulated immune response triggered by ingested gluten proteins of certain cereals. During the course the disease, the mucosa loses its absorptive villi completely and hyperplasia of the enteric crypts occurs, leading to a strongly diminished ability to absorb nutrients. The overall prevalence [5] of the disease in the USA is about 1:133. Figure 1 shows example images, captured during endoscopy.

For most computer aided celiac disease diagnosis approaches [3, 4, 11, 21], reliable image regions (e.g. patches with a size of 128×128 pixels) showing discriminative features have to be identified prior to the automatized classification. This must be done to get idealistic data which is free from strong image degradations as in case of strong degradations the classification accuracy of the decision support system decreases [9, 10]. The identification of reliable regions could be done manually [11, 21] on the one hand or by means of a computer based method

[9] on the other hand. Other approach for detection of "informative" frames [2, 1], do not directly focus on a succeeding computer aided diagnosis and are certainly not optimized for celiac disease diagnosis. Although the manual method seems to be beneficial if done by experienced medical doctors [9], there are two incentives to use a computer based selection method: Firstly, a human interaction during endoscopy is time consuming and tedious, which probably leads to a diminished acceptance of the physicians. Apart from that, especially in case of physicians which are inexperienced, inattentive or just unfamiliar with the (new) decision support system, a weak selection automatically leads to weak classification accuracies [9].

The reason for the decreased classification accuracies in case of randomly or weakly selected patches (or if using the complete images) is the vulnerability of image classification methods to various types of degradations which are prevalent in endoscopic images [10]. It could be shown that image degradations definitely affect the feature extraction and consequently lead to a reduced classification accuracy. Such degradations are blur, noise, a lack of contrast and reflections caused by the light of the endoscope.

In this work we exploit the availability of numerous small subimages (patches) in each original endoscopic image. This availability of large data not just allows to select the best patch per image, as done in previous work [9], but also facilitates a redundant processing (i.e. feature extraction and classification) of these multiple images aiming at improving the overall accuracy. In order to generate one final decision for each image, these redundant threads have to be fused. This can be done on varying levels [19], such as feature-, score- or decision-level. Whereas in celiac disease diagnosis, information fusion has not been investigated so far, in biometric systems considerable improvements with these techniques could be obtained by considering multiple modalities [19], multiple instances of one modality [20] or multiple processing techniques [17]. As the simple fusion methods do not lead to improved accuracies, we utilize patch quality measures [9] to introduce a weighting. Based on this weighting, we propose a weighted decision-level as well as a weighted feature-level fusion method. The training set in each scenario investigated in this work consists of manually extracted idealistic patches. It should be mentioned that this manual stage can be done beforehand



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

Fig. 1: Endoscopic images of healthy mucosa (1a and 1b) clearly showing the villi structure and images of diseased mucosa (1c and 1d).

by experts and does not require any interaction during medical treatment. The ground-truth, which has been determined by histological examination of biopsies, is available for each original image and can be directly taken for all patches extracted from the respective image.

The paper is organized as follows: In Sect. 2 the two fusion approaches are introduced and the quality measures are outlined. In Sect. 3 the experimental results are analyzed and discussed. Finally, Sect. 4 concludes this paper.

2 Fusion Methods

Before introducing the two weighted fusion methods, sensible quality measures have to be defined. In the following we use the same metrics as utilized in previous work [9] to identify one single "best" patch, as these methods seem to be appropriate for our problem definition.

2.1 Quality Measures

- The first measure addresses the problem of a too low illumination. As such a weak illumination generally corresponds to images with a low average gray value, we propose a quality measure being based on the mean of the pixel intensities
$$q_A(P) = \frac{1}{|Z|} \cdot \sum_{z \in Z} P(z) \;, \tag{1}$$
where Z comprises the coordinates of the image patch P.
- The next measure is utilized to detect image regions lacking from any significant gray value differences. Such image patches can be identified by measuring the contrast which is defined by
$$q_C(P) = \sum_{i,j \in K} |i - j| \cdot p(i,j) \;, \tag{2}$$
where K comprises all gray values in P and $p(i,j)$ stands for the probability of these two gray values to be present in a certain image neighborhood in P. In order to focus on real contrast rather than on noise, for this neighborhood we use a quite large offset of four pixels in vertical and in horizontal direction and average these two values.
- The next measure is based on a blur metric b [14]. For computing this metric, first in one direction the edges are identified by extracting all local minima and maxima. Finally the ratio between the lengths and the pixel differences of the edges is computed which indicates the blur level. As all of our images suffer from more or less significant sensor noise, the patches are previously denoised using a Gaussian filter G_2 with $\sigma = 2$.
$$q_B(P) = -b(P * G_2) \;. \tag{3}$$

– To detect noisy image patches, we sum up the differences between the original image and a denoised version of the same image

$$q_N(P) = \sum_{z \in Z} |P - G_1 * P| \; . \tag{4}$$

The denoised image is achieved by filtering the original image with a Gaussian G_1 with $\sigma = 1$.

– Finally, we need a measure to address the problem of reflections and extremely high illuminations. These regions can be detected quite easily by considering the maximum gray values.

$$q_I(P) = \begin{cases} 1, & \text{if } max(P) < T \\ 0, & \text{otherwise} \; . \end{cases} \tag{5}$$

T is set to 245 (eight bit gray scale), which turned out to be appropriate for separating extremely bright regions (by manual inspection of a set of training images).

As shown in [9], one single quality measure is unable to represent the "quality" of a patch with respect to the classification performance. Therefore, we do not focus on single measures but instead utilize a weighted combination of the proposed quality measures.

2.2 Weighted Fusion

Let Q be a matrix containing the row vectors $(q_A, q_C, q_B, q_N, q_I)$ of each patch of one original image and let W be a properly chosen column vector containing a weight for each quality measure. Then the column vector $Q \cdot W$ is the weighted summed overall quality measure as used in the original work on patch selection [9]. In this previous work, the row with the maximum value of this product is evaluated and the corresponding image is used for feature extraction and classification. Classification in this context refers to the discrimination between images showing healthy and diseased mucosa. In the current work, by computing the element-wise exponentiation of $Q \cdot W$ with the properly chosen exponent k, the ratio between the impact of high and low quality patches can be adjusted ($^\circ$ denotes the element-wise matrix exponentiation which corresponds to the repeated Hadamard matrix product). In case of setting k to zero, the quality measures and the weights are ignored and each image finally has the same impact. The thereby achieved fusion methods (unweighted decision-level (DLF) and unweighted feature-level fusion (FLF)) are compared with the weight based methods in the experimental section. If assigning a large value to k, the methods converge to the patch selection strategy as small values are thereby suppressed. In the following two subsections we show how the quality vector $(Q \cdot W)^{\circ k}$ can be used in patch fusion. For the experiments, W and k are evaluated during exhaustive search based on a separate data set.

2.3 Weighted Decision-Level Fusion (W-DLF)

The first method based on the computed quality vector $(Q \cdot W)^{\circ k}$ operates on the decision level. That is, for each patch in an original image, first the classifier's decision is computed by means of traditional feature extraction and classification. All decisions for one original image are stored in the row vector D, where 1 stands for a positive and -1 stands for a negative decision. By computing

$$D_f = sgn(D \cdot (Q \cdot W)^{\circ k}) \,, \tag{6}$$

the single decisions are multiplied with the corresponding weights (image qualities), summed up and finally thresholded using the sign function sgn. We have to content with the rather simple sum rule, as more elaborate decision-level fusion approaches like the Behavior-Knowledge Space [18] or Decision Templates [12] are developed for fusing different classifiers and not different input data.

2.4 Weighted Feature-Level Fusion (W-FLF)

In opposite to W-DLF, W-FLF operates on the feature level. This implies that the features are fused prior to the classification step. In this approach the classification step that corresponds to a loss of information is postponed and applied to the fused features, which could be a benefit compared to the simpler decision-level fusion. The fused feature vector F_f which is used for classification is calculated by

$$F_f = F \cdot \frac{(Q \cdot W)^{\circ k}}{||(Q \cdot W)^{\circ k}||} \,, \tag{7}$$

where F is a matrix containing the feature vectors (columns) for each patch. The quality vector $(Q \cdot W)^{\circ k}$ is normalized to ensure that the sum of all contributions is one. The column vector F_f contains the element-wise weighted sum of all feature vectors and can be directly given to the classifier. F_f could be interpreted as a weighted average feature vector. We pursue this strategy, as it intuitively allows a weighting of the individual features, which cannot be achieved easily in case of a feature concatenation. The averaging theoretically requires that the decision boundaries are linear as otherwise the averaging of two features of one class could lead to an averaged descriptor located in the subspace of the other class. However, in the experiments we do not restrict to linear classification. To investigate the impact of the decision boundary on our approach, the utilized features are individually analyzed with respect to this problem in Sect. 3 with variable classifier adjustments.

2.5 Runtime Analysis

The major steps, as far as computational effort is concerned, consist of quality measurement (consisting of five single measures) and feature extraction. Whereas in the fused approach the quality measures as well as the features must be computed for each patch, in case of patch selection [9] the feature must be

computed only for the best patch. The overall computation time[3] for all quality measures on 128×128 pixel gray value patches is 37 milliseconds (ms) (q_A: 1 ms, q_C: 16 ms, q_B: 1 ms, q_N: 1 ms, q_I: 18 ms). The computation time for the features ranges from 6 to 142 ms (6 ms (LBP), 6 ms (ELBP), 13 ms (SCH), 142 ms (MFS), 2 ms (FPS)). For example in case of fusion based classification with LBP or ELBP and extracting 16 patches per original image, for each original image the computation time would be about 688 ms where 592 of them are consumed for quality measurement and only 96 are used for feature extraction. In case of patch selection based classification, it would take 598 (592+6) ms which is not significantly faster. Thus, we claim that the small additional computational effort is justified if the fusion leads to increased accuracies.

3 Experiments

3.1 Experimental Setup

The image test set used contains images of the duodenal bulb and the pars descendens taken during duodenoscopies at the St. Anna Children's hospital using pediatric gastroscopes (with a resolution of 768×576 (Olympus GIF Q165) and 528×522 pixels (GIF N180), respectively).

To generate the ground-truth, the condition of the mucosal areas covered by the images was determined by histological examination of biopsies from the corresponding regions. Severity of villous atrophy was classified according to the modified Marsh classification as proposed in [15]. Although it is possible to distinguish between the several stages of the disease, we only aim in distinguishing between images of patients with (Marsh-3) and without the disease (Marsh-0), because this 2-classes case is more relevant in practice [21]. Another incentive for preferring the 2-classes case is that the distinction between the different stages of the disease is considerably subjective even as far as the histological examination is concerned [22]. Thereby, the ground-truth and furthermore the evaluation in a multi-classes case would be less reliable.

Our experiments are based on a balanced database containing 612 idealistic patches (i.e. patches 306 per class) which are used for classifier training and 172 original images that are used for evaluation. From each original image, 16 non-overlapping 128×128 pixel patches are automatically extracted and furthermore used for fused classification. The patch size is chosen in order to be able to compare the results with the manual extraction that is done by a highly experienced endoscopist. The original images are captured during endoscopies from 72 different patients. To allow an efficient parameter estimation, this database (consisting of 612 idealistic and 172 original images) is divided into two equally sized sets (DB 1 and DB 2). Each of them contains 306 idealistic images (patches) which are used for training as well as 86 original images. In case of multiple images of

[3] Runtime tests are executed on an Intel i5 architecture with 3.1 MHz. All functions are implemented in MATLAB 2013a

one patient, we had to ensure that they end up in the same set. The weight vector W as well as the exponent k are evaluated during exhaustive search, based on the opposing data set as follows: In order to evaluate the accuracies based on the original images of DB 1, the idealistic images of DB 2 are utilized for training and the original images of DB 2 are used for parameter estimation. The same procedure is applied (vice versa) to evaluate DB 2. Thereby a strict separation between training set, test set and evaluation set is achieved. The search space for each element of W is between 0.0 and 1.0 with a step-size of 0.33 and k is within $\{2^{-1}, 2^0, 2^1, ..., 2^6, 2^7\}$. The parameters of the quality measures are taken from the previous work on patch selection [9] as described in Sect. 2.

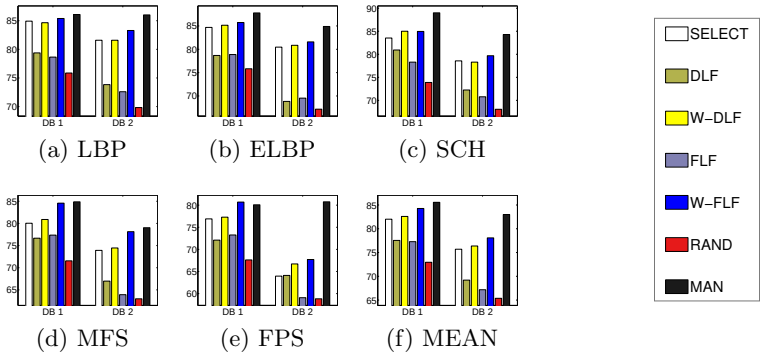
We perform two different experiments. Experiment A corresponds to the natural fusion of patches extracted from one distinct original image. Experiment B should show if the accuracy improvements are limited by the correlations within one original image. Such correlations are quite natural, as degradations like blur or noise often do not occur only in a small region, but sometimes even compromise a whole image. Therefore, in this experiment the patches of each patient are randomly interchanged across the images leading to virtual images consisting of patches from the same patient, but from different original endoscopic images. This is done as the patches from the new virtual images are supposed to be less correlated and the used database does not contain enough patients to fuse all patches from one patient.

For classification the k-nearest neighbor classifier is used. We utilize this simple classifier in order to focus on the effect of different settings rather than on achieving the highest overall classification rates. For the first experiments (A and B), the rates achieved with odd k values reaching from 1 to 31 are averaged, to get highly stable results rather than to get the highest possible rates. In an analysis (Fig. 4) we investigate the impact of different k values.

3.2 Feature Extraction Techniques

For the experimental analysis we deploy the following feature extraction techniques which proved to the adequate for celiac disease classification in previous work [6]:

- Local Binary Patterns [16] (LBP):
  The commonly used Local Binary Patterns describe a texture by computing the joint distribution of binarized intensity differences within a certain neighborhood. This widely used feature extraction technique is used with eight neighbors and a radius (i.e. the distance to the neighbors) of two pixels.
- Extended Local Binary Patterns [13] (ELBP):
  ELBP is an edge based derivative of Local Binary Patterns. As LBP it is used with eight neighbors and a radius of two pixels.
- Fourier Power Spectra Rings [8] (FPSR):
  To get this descriptor, first the Fourier power spectra of the image patches are computed, in a way that the low frequencies are in the image center. Afterwards, a ring with a fixed inner and outer radius is extracted and the



(a) LBP (b) ELBP (c) SCH

(d) MFS (e) FPS (f) MEAN

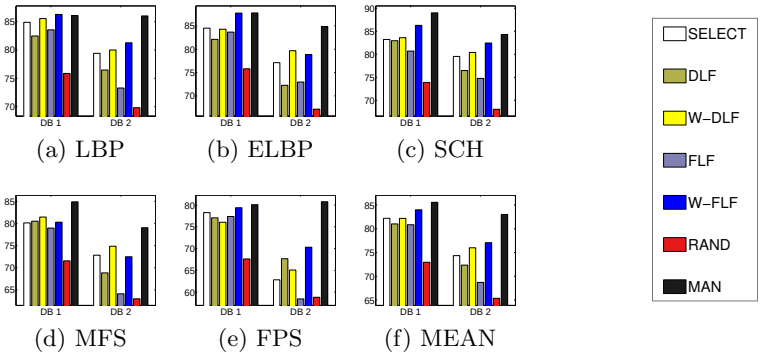
Fig. 2: Experiment A: These plots show the overall classification accuracies achieved with patch selection (SELECT), decision-level fusion (DLF and W-DLF), feature-level fusion (FLF and W-FLF), a random patch selection (RAND) and the manual patch selection (MAN).

median of the values in this ring are calculated. For our experiments we use an inner radius of seven and an outer radius of eight pixels, which turned out to be suitable in previous work [8].

– Shape Curvature Histogram [7] (SCH):
  SCH is a shape feature, especially developed for celiac disease diagnosis. After detection of significant locations, a histogram collects the occurrences of the contour curvature values in these regions. As in the original work, we consider a histogram bin count of eight.
– Multi-Fractal Spectrum [23] (MFS):
  The local fractal dimension is computed for each pixel using three different types of measures for computing the local density. The feature vector is built by concatenation of these fractal dimensions.

3.3 Results and Discussion

In Fig. 2 the overall classification accuracies achieved with patch selection (SELECT), unweighted and weighted decision-level fusion (DLF and W-DLF), unweighted and weighted feature-level fusion (FLF and W-FLF), a random patch selection (RAND) and the patch selection based on the human experts (MAN) are shown for experiment A. It can be seen that the unweighted feature-level fusion method FLF as well as the unweighted decision-level method DLF are unable to compete with the single patch selection in case of any feature. The rates obtained with the manual selection are totally out of reach. However, these methods, which are not based on any (obviously highly important) quality measure, are at least able to outreach the random selection. Considering the weight based methods we recognize that especially the weighted feature-level based method W-FLF is able to outperform the single patch selection method SELECT in case of all features and all databases with differing extent. Considering



(a) LBP (b) ELBP (c) SCH

(d) MFS (e) FPS (f) MEAN

Fig. 3: Experiment B: These plots show the accuracies with the same strategies as in Fig. 2. In opposite to experiment A, the patches of one patient are randomly interchanged.

MFS, LBP and ELBP, the accuracies of the manual patch selection can be virtually reached. Quite high differences are observed in case of SCH which does not strongly benefit from fusion. A quite interesting aspect is the difference between the two weighted fusion techniques. In almost each case, the feature based W-FLF corresponds to the higher accuracy compared to W-DLF. Obviously the early fusion prior to the (information reducing) classification has a positive impact on the final discriminative power.

In Fig. 3 the results of experiment B, which is based on randomly interchanged patches across images of the same patient, are shown. As in this experiment not only patches extracted from the same image are fused, but patches from different images, we expected that in this scenario more significant improvement could be obtained in case of information fusion. Actually, on average (see Fig. 3f) the rates with the weight based fusion methods are similar. Interestingly, it can be observed that the expected accuracy increase holds off in case of the methods DLF and FLF, not being based on weighting. However, as they are unable to outperform the weight based ones, this observation has no practical relevance.

As especially the weight based feature-level fusion method in both experiments leads to increased accuracies, we expect that a fusion on the one hand across all patches in an image (derived from experiment A) and on the other hand across all images, captured during endoscopy of one distinct patient (derived from experiment B) could improve the rates from our experiments once again. Unfortunately, the data currently available is not large enough for such an experiment.

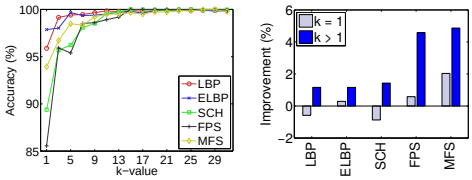
So far, we experimentally showed that the W-FLF approach is able to improve the classification accuracies of state-of-the-art patch selection, without regarding the theoretical issues in case of non-linear decision boundaries. Finally we investigate the impact of the classifier's decision boundary on the effectiveness of W-FLF. As stated in Sect. 2.4, the feature averaging theoretically requires that the decision boundaries are linear as otherwise the averaging of two features of

one class could lead to an averaged descriptor located in the subspace of the other class. To investigate how often an averaged feature of two correctly classified images would be incorrectly classified, now we consider all correctly classified images (from the idealistic patches data set). For each pair of these images, the average feature is computed and classified with varying settings (different k values). This is done as especially small k values correspond to highly non-linear decision boundaries, whereas with higher k values this effect is softened. Figure 4a shows that especially in combination with low dimensional features (FPS, SCH, MFS) and small k values (majorly for k=1), the feature averaging leads to decreased classification accuracies (as 100 % accuracy is expected in case of linear classification).

In Fig. 4b, the impact of a small k value (k=1) compared to the averaging (with k reaching from 3 to 31) on the improvement achieved with W-FWF compared to patch selection is shown. Apart from the classifier settings, the same setup as in experiment A is used. As expected, if k is set to one, the improvements of W-FLF are (especially in combination with the low dimensional features) considerably smaller or even negative, which is expected to be due to the highly non-linear decision boundaries. Therefore, we recommend to take care about the classifier choice using the proposed method for weighted feature-level fusion.

4 Conclusion

We have shown that information fusion in celiac disease classification is able to increase the classification accuracies with small additional computation costs compared to single patch selection. Whereas the simple decision-level and feature-level methods are not able to improve the performance, the introduced weight-based methods definitely are. Especially the feature-level fusion approach turned out to be most appropriate on average. Thereby, this paper showed that the measurement of image quality has a major impact not only in case of a single patch selection, but also in case of information fusion. Getting nearer to the classification rates of manual patch selection, this work brings us one step closer to fully automatized non-interactive celiac disease diagnosis.



(a) Analysis of the decision boundaries.

(b) Impact of varying k-values on the accuracies.

Fig. 4: Analysis of effects in weighted feature-level fusion.

References

1. Atasoy, S., Mateus, D., Lallemand, J., Meining, A., Yang, G.Z., Navab, N.: Endoscopic video manifolds. In: Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'10). Lecture Notes in Computer Science, vol. 6362, pp. 437–445 (2010)
2. Bashar, M., Kitasaka, T., Suenaga, Y., Mekada, Y., Mori, K.: Automatic detection of informative frames from wireless capsule endoscopy images. Medical Image Analysis 14(3), 449–470 (2010)
3. Ciaccio, E.J., Tennyson, C.A., Bhagat, G., Lewis, S.K., Green, P.H.R.: Classification of videocapsule endoscopy image patterns: comparative analysis between patients with celiac disease and normal individuals. BioMedical Engineering Online 9(1), 1–12 (2010)
4. Ciaccio, E.J., Tennyson, C.A., Lewis, S.K., Krishnareddy, S., Bhagat, G., Green, P.: Distinguishing patients with celiac disease by quantitative analysis of videocapsule endoscopy images. Computer Methods and Programs in Biomedicine 100(1), 39–48 (Oct 2010)
5. Fasano, A., Berti, I., Gerarduzzi, T., Not, T., Colletti, R.B., Drago, S., Elitsur, Y., Green, P.H.R., Guandalini, S., Hill, I.D., Pietzak, M., Ventura, A., Thorpe, M., Kryszak, D., Fornaroli, F., Wasserman, S.S., Murray, J.A., Horvath, K.: 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)
6. Gadermayr, M., Liedlgruber, M., Uhl, A., Vécsei, A.: Evaluation of different distortion correction methods and interpolation techniques for an automated classification of celiac disease. Computer Methods and Programs in Biomedicine 112(3), 694–712 (Dec 2013)
7. Gadermayr, M., Liedlgruber, M., Uhl, A., Vécsei, A.: Shape curvature histogram: A shape feature for celiac disease diagnosis. In: Medical Computer Vision. Large Data in Medical Imaging (Proceedings of the 3rd International MICCAI - MCV Workshop 2013). Springer LNCS, vol. 8331, pp. 175–184 (2014)
8. Gadermayr, M., Uhl, A., Vécsei, A.: Barrel-type distortion compensated fourier feature extraction. In: Proceedings of the 9th International Symposium on Visual Computing (ISVC'13). Springer LNCS, vol. 8033, pp. 50–59 (Jul 2013)
9. Gadermayr, M., Uhl, A., Vécsei, A.: Getting one step closer to fully automatized celiac disease diagnosis. In: IEEE International Conference on Image Processing Theory, Tools and Applications 2014 (IPTA'14) (Oct 2014), accepted
10. Hegenbart, S., Uhl, A., Vécsei, A.: Impact of endoscopic image degradations on lbp based features using one-class svm for classification of celiac disease. In: Proceedings of the 7th International Symposium on Image and Signal Processing and Analysis (ISPA'11). pp. 715–720. Dubrovnik, Croatia (Sep 2011)
11. Hegenbart, S., Uhl, A., Vécsei, A.: Impact of histogram subset selection on classification using multiscale LBP. In: Proceedings of Bildverarbeitung für die Medizin 2011 (BVM'11). pp. 359–363. Informatik aktuell, Lübeck, Germany (March 2011)
12. Kuncheva, L.I., Bezdek, J.C., Duin, R.P.: Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recognition 34(2), 299–314 (2001)
13. Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.: Learning multi-scale block local binary patterns for face recognition. In: Advances in Biometrics, pp. 828–837. Springer (2007)
14. Marziliano, P., Dufaux, F., Winkler, S., Ebrahimi, T., Sa, G.: A no-reference perceptual blur metric. In: IEEE International Conference on Image Processing (ICIP'02). pp. 57–60 (2002)

15. Oberhuber, G., Granditsch, G., Vogelsang, H.: The histopathology of coeliac disease: time for a standardized report scheme for pathologists. European Journal of Gastroenterology and Hepatology 11, 1185–1194 (Nov 1999)
16. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recognition 29(1), 51–59 (January 1996)
17. Prabhakar, S., Jain, A.K.: Decision-level fusion in fingerprint verification. Pattern Recognition 35(4), 861–874 (2002)
18. Raudys, v., Roli, F.: The behavior knowledge space fusion method: Analysis of generalization error and strategies for performance improvement. In: Proceedings of the 4th International Conference on Multiple Classifier Systems. pp. 55–64. MCS'03, Springer-Verlag (2003)
19. Ross, A., Jain, A.: Information fusion in biometrics. Pattern Recognition Letters 24(13), 2115–2125 (2003)
20. Uhl, A., Wild, P.: Single-sensor multi-instance fingerprint and eigenfinger recognition using (weighted) score combination methods. International Journal on Biometrics (Special Issue on Multimodal Biometric and Biometric Fusion) 1(4), 442–462 (2009)
21. 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 (Jun 2011)
22. Weile, B., Hansen, B.F., Hägerstrand, I., Hansen, J.P.H., Krasilnikoff, P.A.: Interobserver variation in diagnosing coeliac disease, a joint study by danish and swedish pathologists. APMIS 108(5), 380–384 (2000)
23. Xu, Y., Ji, H., Fermüller, C.: Viewpoint invariant texture description using fractal analysis. International Journal of Computer Vision 83(1), 85–100 (2009)