Dealing with Intra-Class and Intra-Image Variations in Automatic Celiac Disease Diagnosis

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

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

Abstract Computer aided celiac disease diagnosis is based on endoscopic images showing the villi structure in regions of the small bowel. Especially unavoidably variable illuminations and varying viewing angles of the individual villi are a source for high intra-class as well as intra-image variations in the image domain. We clarify that common texture descriptors are unable to compensate such a high degree of variance, which is supposed to be a crucial problem in computer aided diagnosis. In this work, a straight-forward split and merge approach is presented which facilitates the final classification task by reducing the intra-image variance and simultaneously enlarging the training set. Using different well known feature extraction techniques as well as two classifiers, it can be shown that the overall classification accuracies can be increased consistently. Additionally, the proposed approach is compared to the related but more complex bag-of-visual-words method.

1 Introduction

Typically an image texture is associated with a regular, periodic pattern (e.g. as shown in Fig. 1). However, in real work applications, textured images are often quite non-periodic as well as inhomogeneous. This is especially true in case of computer aided celiac disease diagnosis which relies on images of the mucosa of a part of the small bowel, taken during endoscopy (see Fig. 2).

This inhomogeneity could be due to variations in the acquisition conditions. In recent years high effort has been made on developing texture descriptors which are invariant to certain properties such as illumination, scale, affine transformations and rotations. However, in case of endoscopic images the variations often cannot be effectively described. For example, a slightly different viewing

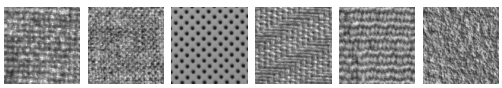


Figure 1. Periodic texture patches from Kylberg texture database [1]

(a) Marsh-0: Healthy villi structure (b) Marsh-3: Affected mucosa

Figure 2. Endoscopic images showing the mucosa of the small bowel.

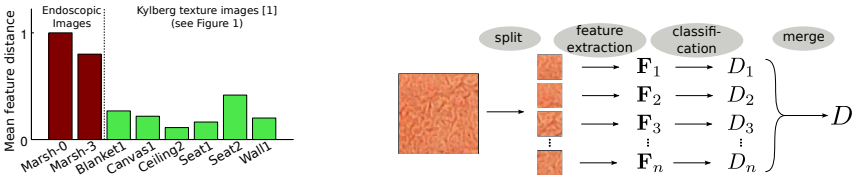
angle leads to a different illumination which might cause totally different image properties. Furthermore, as villi present a three dimensional flexible structure, a varying viewing angle cannot be modeled approximatively for instance by means of an affine transformation.

In case of our image data, image variations occur between different images of one class, which is referred to as intra-class variation. However, even within one image the degree of self similarity is often quite low (which is referred to as intra-image variation). This can be seen in Fig. 2, showing images of healthy (Marsh-0) and diseased patients (Marsh-3).

Visually it is obvious that the degree of regularity as well as periodicity in case of the endoscopic images is significantly lower compared to the regular patterns in Fig. 1. In Fig. 3(a), this is quantified by computing the average distances of extracted features from the upper-left quarter and the lower-right quarter (64×64 pixels) of the same patch (128×128 pixels). Especially, Local Binary Patterns [2] (the exact setup is given in Sect. 2.1) are used for feature extraction in combination with the squared Euclidean distance. However, with other image descriptors a similar output is generated. It can be seen, that the endoscopic images not only visually present a high degree of intra-image distances. The variations are even preserved if switching to feature domain.

Previous work on computer aided celiac disease diagnosis relies on feature extraction from 128×128 pixel patches [3] or even on 576×576 pixel images [4]. Features which have been declared to be invariant to scale, rotations as well as affine transformations have been investigated in previous work [3,5]. However, although some of them seem to be beneficial in synthetic scenarios (e.g. if training is based on idealistic and evaluation is based on transformed images), for real world applications highly straight-forward methods such as Local Binary Patterns and derivatives [2,6,7] often outperform these more elaborated techniques. This is supposed to be due to the fact that often distinctiveness has to be sacrificed for a higher degree of invariance, which has been particularly discussed in previous work [8].

In this work, we propose a split and merge approach which decreases the degree of intra-image variations by splitting the images into several smaller sub-images. After feature extraction and classification of the sub-images, the decisions are merged in order to get one final decision. Experiments with high performing feature extraction techniques and two well known and commonly used classifiers show that in case of most configurations an improvement is achieved.



(a) Intra-image feature distances in case of the Local Binary Pattern descriptor.

(b) Schematization of the sequences of the split and merge methodology.

Figure 3.

2 Material and Methods: Split and Merge Approach

Most highly distinctive texture feature extractors such as Local Binary Patterns and derivatives are based on the assumption that a textured image is homogeneous, which allows a simple global collection of data in histograms. We try to retrieve an increased degree of homogeneity by splitting the original image into several non-overlapping sub-images of equal size and aspect ratio. Theoretically the method could also handle overlapping sub-images, however, in our case this did not significantly improve the overall results. Afterwords, for each sub-image, a texture feature extraction method is performed individually. Based on these computed feature vectors, classification is just as done individually for each sub-image. Classifier training is performed as well based on the sub-images. The final decision for an image is obtained by majority voting of the decisions of all sub-images. Fig. 3(b) schematizes this methodology. The proposed method attends the following potentially beneficial effects:

- By focusing on smaller regions in a textured image, the degree of intra-image variances is reduced as the neighborhood is limited.
- The training stage is also based on the smaller sub-images, which means that the number of images in the training set is multiplied by the split factor. This is especially valuable in case of small training sets.
- A set of decisions is available for acquiring the final decision for one image. This redundancy can be used to increase the accuracy as well as to give a statement on the certainty of the overall decision.

Using the bag-of-visual-words approach [9], in a similar way small sub-images are extracted to introduce a higher degree of invariance. The major difference is that in case of this elaborated technique one final feature is computed per image and not per sub-image.

2.1 Experimental Setup

The image testset used for experimentation contains images of the duodenal bulb and the pars descendens taken during endoscopies at the St. Anna Children's Hospital using pediatric gastroscopes (Olympus GIF N180 and Q165). Prior to processing, all images are converted to gray scale images as the additional use of

color information did not lead to continuous improvements. In a preprocessing step, texture patches with a fixed size of 128×128 pixels have been manually extracted. These patches are split into several smaller sub-images. We consider splits into four, nine and 16 equally sized square sub-images. In case of the four and the 16 sub-images split, another (overlapping) patch in the center is extracted to avoid ties during majority voting. To get the ground truth for the texture patches, the condition of the mucosal areas covered by the images has been determined by histological examination of biopsies from corresponding regions. The severity of the villous atrophy has been classified according to the modified Marsh classification [10]. Although it is possible to distinguish between different stages of the disease, we aim in distinguishing between images of patients with (Marsh-3) and without the disease (Marsh-0), as this two classes case is most relevant in practice. Our experiments are based on a data set containing 612 images (306 Marsh-0 and 306 Marsh-3 images) from 171 patients [3]. All overall accuracies computed are based on the mean accuracy of 32 random splits. One distinct split divides the data set into an approximately balanced training (50 %) and evaluation set (50 %), restricting images of one patient to be in the same set to avoid any bias.

To extensively study the effect of the proposed approach on the overall classification accuracy, six different feature extraction techniques which turned out to be appropriate for celiac disease classification are investigated. The chosen parameters turned out to be optimally suited in earlier experiments.

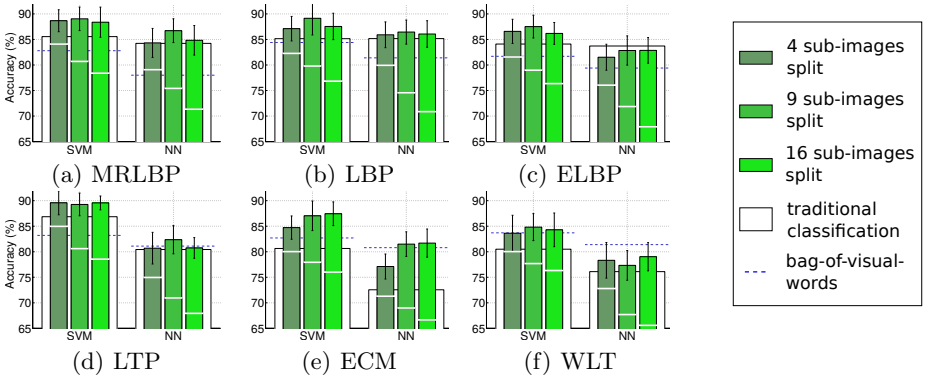
- Local Binary Patterns (LBP) [2]: LBP is deployed with eight circularly neighboring samples and a radius of two pixels.
- Multi-resolution Local Binary Patterns (MRLBP) [2]: This feature vector consists of the concatenation of an LBP vector with a neighborhood radius of one and a radius of two pixels.
- Extended Local Binary Patterns (ELBP) [6]: This edge-based derivative of LBP is as well used with eight neighbors and a radius of two pixels.
- Local ternary patterns [7] (LTP): LTP is used with a radius of two, eight neighbors and a threshold of three.
- Edge Co-occurrence Matrix [11] (ECM): The ECM is achieved by computing the gray-level co-occurrence matrix of the edge-orientation within a specified displacement. In the experiments the matrices for a displacement of one and two pixels are concatenated.
- Wavelet Variance [12] (WAV): After computing a four level two dimensional wavelet packets decomposition, the variance in each sub band is calculated. The final feature vector consists of the concatenation of these values.

For feature discrimination, we deploy the linear support vector classifier [13] (SVM) and the (highly non-linear) nearest neighbor classifier (NN) which are both commonly utilized. The results of the split and merge approach are compared to the traditional classification (i.e. feature extraction is based on original images) and the bag-of-visual-words approach which is similar to our new approach as the descriptors are also extracted from small sub-images. For this, the cluster count has been fixed to seven and a sub-image size of 32×32 pixels has

been chosen, which turned out to be optimal and corresponds to the 16 sub-images split in our novel approach. Additionally, to get a higher data density, 49 (instead of 16) sub-images have been extracted in an overlapping manner.

3 Results and Discussion

In Fig. 4, the finally achieved overall classification accuracies are shown. The wide bars indicate the rates obtained with traditional classification without splitting. The narrow bars indicate the rates achieved with the novel split and merge method and the splitting into four (left bar), nine (center bar) and 16 sub-images (right bar). The dashed line represents the accuracy obtained with the Bag-of-visual-words approach. It can be seen that in each case, except for ELBP in combination with the NN classifier, the traditional classification accuracy (wide bars) can be increased using the split and merge approach. Especially the splitting into nine sub-images (represented by the center bar) mostly corresponds to remarkable improvements. These improvements are slightly more significant if the linear SVM is used. This is supposed to be due to the high flexibility of the non-linear NN classifier (compared to the linear SVM) that helps to compensate the intra-class variations using the traditional method. Considering the performance compared to the bag-of-visual-words approach it can be seen that the new method is highly effective in case of LBP-derivatives. Whereas the computationally more complex bag-of-visual-words approach is rarely ever able to outperform traditional classification, the novel method definitely is. The interruptions of the narrow bars indicate the accuracies achieved with the sub-images without a decision-level fusion. It can be seen that with an increasing size reduction, the error rates rise with varying extent. This effect is more distinct in case of the NN classifier. This shows us that, especially in case of the 16 sub-images split, the final decision fusion has a highly positive effect on the accuracies.



(a) MRLBP (b) LBP (c) ELBP

(d) LTP (e) ECM (f) WLT

Figure 4. Classification accuracies of traditional classification (wide bars), split and merge (narrow bars) and bag-of-visual-words based classification (dashed lines). The interruption of the narrow bars indicate the accuracies, achieved with small sub-images without the final decision level fusion.

3.1 Conclusion

We have proposed a straight-forward split and merge approach for improving the classification of images showing high intra-image and intra-class variations. The intra-image variations are reduced and simultaneously the training set is increased, which helps to handle intra-class variances. Using different well known feature extraction techniques as well as two classifiers, it has been shown that the overall classification accuracies can be increased consistently, although we have concentrated on feature extraction methods which are known to be effective in case of traditional classification. We suppose that other features that are optimized for small image data could lead to even higher overall accuracies.

References

1. Kylberg G. The Kylberg Texture Dataset v. 1.0. Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University, Uppsala, Sweden; 2011. 35. http://www.cb.uu.se/~gustaf/texture/.
2. Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on feature distributions. Pattern Recognition. 1996 Jan;29(1):51–59.
3. Hegenbart S, Uhl A, Vécsei A, Wimmer G. Scale invariant texture descriptors for classifying celiac disease. Medical Image Analysis. 2013;17(4):458–474.
4. Ciaccio EJ, Tennyson CA, Lewis SK, Krishnareddy S, Bhagat G, Green P. Distinguishing Patients with Celiac Disease by Quantitative Analysis of Videocapsule Endoscopy Images. Comp Meth and Programs in Biomed. 2010 Oct;100(1):39–48.
5. Hegenbart S, Uhl A. A Scale-Adaptive Extension to Methods based on LBP using Scale-Normalized Laplacian of Gaussian Extrema in Scale-Space. In: Proc. of the Intern. Conf. on Acoustics, Speech, and Signal Processing; 2014. p. 4352–4356.
6. Liao S, Zhu X, Lei Z, Zhang L, Li S. Learning Multi-scale Block Local Binary Patterns for Face Recognition. In: Advances in Biometrics; 2007. p. 828–837.
7. Tan X, Triggs B. Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions. In: Analysis and Modelling of Faces and Gestures. vol. 4778; 2007. p. 168–182.
8. Gadermayr M, Uhl A. Degradation Adaptive Texture Classification. In: Proc. of the Intern. Conf. on Image Processing 2014 (ICIP'14); 2014. .
9. Varma M, Zisserman A. Classifying Images of Materials: Achieving Viewpoint and Illumination Independence. In: Proc. of the 7th Europ. Conf. on Computer Vision (ECCV'02); 2002. p. 255–271.
10. Oberhuber G, Granditsch G, Vogelsang H. The histopathology of coeliac disease: time for a standardized report scheme for pathologists. Europ Journ of Gastroenterology and Hepatology. 1999 Nov;11:1185–1194.
11. Rautkorpi R, Iivarinen J. A Novel Shape Feature for Image Classification and Retrieval. In: Proc. of the Intern. Conf. on Image Analysis and Recognition (ICIAR'04); 2004. p. 753–760.
12. Garcia C, Zikos G, Tziritas G. Wavelet packet analysis for face recognition. Image and Vision Comp. 2000;18(4):289–297.
13. Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. LIBLINEAR: A Library for Large Linear Classification. Journ of Machine Learning Research. 2008;9:1871–1874.