The Effect of Endoscopic Lens Distortion Correction on Physicians' Diagnosis Performance

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

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

Abstract In endoscopic images, significant barrel-type distortions are introduced in case of deploying wide-angle lenses. In this work, the effect of lens distortion correction on the human experts' classification performance is investigated. Especially, the discrimination between healthy patients and patients suffering from celiac disease is examined. Furthermore, the classification results of human experts are compared with those achieved with state-of-the-art computer aided decision support systems. This paper considers a two- as well as a four-classes case. Finally we come to the conclusion, that distortion correction does not improve the human experts' classification performance.

1 Introduction

Celiac disease is an autoimmune disorder that affects the small bowel in genetically predisposed individuals after introduction of gluten containing food. Characteristic for this disease is an inflammatory reaction in the mucosa of the small bowel caused by a dysregulated immune response triggered by ingested gluten proteins of certain cereals.During the course of the disease the mucosa loses its absorptive villi and hyperplasia of the enteric crypts occurs, leading to a diminished ability to absorb food. According to a study [1], the prevalence of the disease in the USA in not-at-risk groups was 1:133.

Endoscopy with biopsy is currently considered the gold standard for the diagnosis of celiac disease. However, in the past several techniques for enhanced celiac disease classification avoiding biopsies have been investigated, such as an immersion technique [2], chromoendoscopy [3] and zoom endoscopy [4]. In this paper, focus is on barrel-type distortions which are introduced by the wide-angle optics deployed in endoscopes. These distortions can be undone, by first estimating the distortion function and then applying the inverse of this function

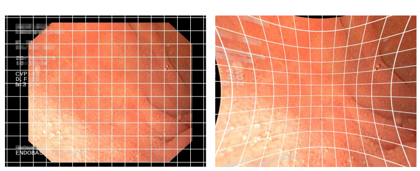


Figure 1. This figure shows an original endoscopy image (left) and the distortion corrected version (right). The grid is artificially added for a better visualization of the distortion correction.

to the distorted images (see Fig. 1). However, during distortion correction (DC) new issues arise as the images must be stretched especially in peripheral regions and thereby due to the lack of data points, these image regions are blurred. In a recent study [5], the effect of barrel-type distortions and distortion correction on the classification accuracy of computer aided celiac disease diagnosis has been investigated. Whereas in most cases the distortion correction leads to decreased classification accuracies, for some special cases stable improvements are observed.

In this work, we investigate the effect of barrel-type distortion correction on the classification performance of human experts. We separately analyze the effect on accuracy (in case of 2 and 4 classes), specificity and sensitivity. Furthermore, we compare the achieved rates with those achieved by computer based methods.

This paper is organized as follows: In Sect. 2, the objectives and the experimental setup of this study are explained. In Sect. 3, the experimental results are presented. These results are discussed in Sect. 4.

2 Materials and Methods

The key issue of this paper is whether the correction of barrel-type distortions has a positive (or negative) impact on the classification performances of medical experts. We separately consider accuracy, specificity and sensitivity in the two-classes case as well as the accuracy in a four-classes case.

In case of the original barrel-type distorted images, peripheral regions appear distinctly smaller than central regions. This might be an issue during classification. However, a distortion correction, which is able to retrieve the geometrical correctness not necessarily enhances the classification rates. Medical experts are used to the wide-angle optics and potentially compensate the lens distortion mentally. Moreover, due to their practice they might be able to additionally compensate perspective distortions by considering the image properties. This process might be compromised in case of distortion correction. Moreover, as the images must be stretched during distortion correction, peripheral image regions are significantly blurred (due to a lack of data points). Especially this issue turned out to compromise some computer aided decision support techniques [5]. To estimate the effect of distortion correction on the human experts' classification performance, a first experiment (stage 1) is based on the complete endoscopic images. As the medical doctors are accustomed to these images, the straightforward investigation is to provide original and distortion corrected complete images which have to be manually classified.

Whereas human experts usually deal with these complete images, computer aided decision support systems are optimized for 128×128 pixel patches [6]. To get a ground-truth database, experts have manually extracted these patches. In order to be able to compare classification results of physicians with those of computer based methods, another experiment (stage 2) is based on such manually extracted patch images. Considering such image patches, the prior knowledge of the location in the complete image is removed and thereby it is no longer possible to mentally compensate the distortions.

Finally, we would like to know if the physicians are aware of the fact that the images suffer from lens distortions. Therefore, we provided 25 synthetically distorted checkerboard patterns (see Fig. 2). The physicians should estimate the barrel-type distortions of typical endoscopes used for celiac disease diagnosis.

2.1 Experimental Setup

The image test set used contains images of the duodenal bulb taken during duodenoscopies at the St. Anna Children's hospital using pediatric gastroscopes. First, we chose image patches which turned out to be hard to classify by computer based methods (to get larger differences between the classification performances). Then we added complete images from which these patches have been extracted.

Each of the 11 medical experts has classified 104 complete images and 144 image patches (see Table 1). In each stage, the images are randomly permuted, separately for each physician. For each original image, the database additionally contains a corrected version. The group of human experts involved in this study consist of 9 pediatrics and 2 adult gastroenterologists. All of them are experts in the field of celiac disease diagnosis.

The condition of the mucosa should be estimated according to the modified Marsh classification scheme [7]. The experts had to assign one of the four labels Marsh-0 (healthy), Marsh-3A, Marsh-3B and Marsh-3C. As this four-classes (4C) case is highly difficult, we additionally evaluate the two-classes (2C) case. Therefore, the classes Marsh-3A, Marsh-3B and Marsh-3C are simply identified as Marsh-3. 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.

For computer aided classification, we use the Shape Curvature Histogram [8] as feature extraction technique. Due to its compact dimensionality, this feature is optimally suited for the small datasets. In order to avoid overfitting, the parameters as proposed in [8] are utilized. For classification, we utilize the k-nearest neighbor classifier in combination with leave-one-patient-out cross validation. The classification accuracies for k reaching from 1 to 12 (12 corresponds to the minimal count of images per class) are averaged, to get stable and sound rates. For lens distortion correction, we deployed the method introduced in [9] (in combination with bi-linear interpolation), which is based on the division model [10]. This quite simple method proved to effectively rectify the geometrical image properties and is appropriate for a following feature extraction (as shown in [5]).

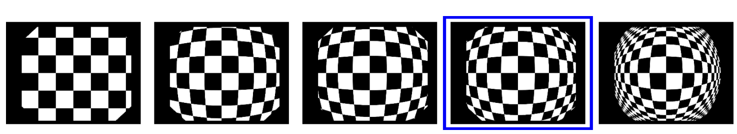


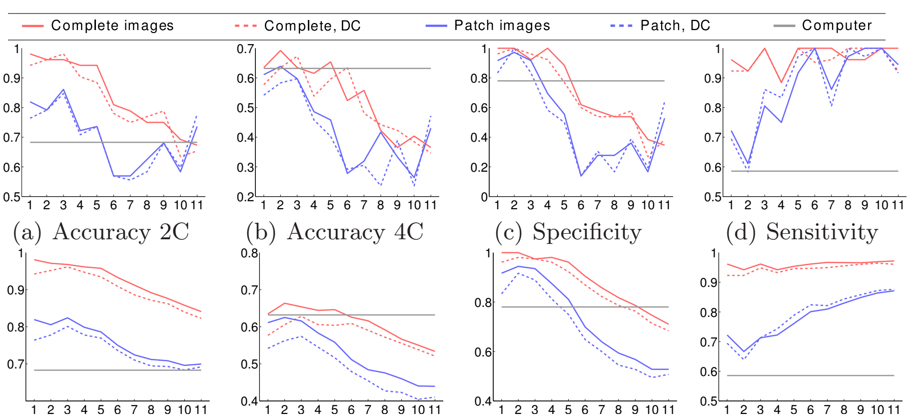
Figure 2. Synthetically distorted checkerboard patterns. The framed image represents barrel-type distortions, typical for the used endoscopes.

Table 1. The test image databases used for the experiments.

Stage	1: Complete images	2: Image patches
Image size (pixels)	768×576 / 528×522	128×128
Number images: Marsh-0	26	36
Number images: Marsh-3 (3A, 3B, 3C)	26 (8, 8, 10)	36 (12, 12, 12)

3 Results

Figure 3 shows the 2C accuracy ((a) and (e)), the 4C accuracy ((b) and (f)), the (2C) specificity ((c) and (g)) and the (2C) sensitivity ((d) and (h)). The top sub-figures show the rates for each expert (x-axis). The physicians are ordered by their 2C accuracy (without DC). The bottom sub-figures show the cumulative average rates. To get these rates, for each x-axis value n we average the rates of the (ordered) experts from 1 to n. Thereby, we get mean rates of the best n physicians (where n corresponds to the x-axis value). We decided for this visualization strategy in order to get more stable and sound results. The gray horizontal line indicates the rate achieved with the computer based method (and the image patches). In the following we especially consider the accumulative rates (right sub-figures). Regarding these plots, we notice that the distortion correction (dashed lines) in general compromises the classification process. No matter if considering the 2C or the 4C case or the specificity, the dashed lines are always below the solid ones. Only the sensitivity in case of patch classification, seems to be similarly high of even slightly above in case of distortion correction.



(a) Accuracy 2C (b) Accuracy 4C (c) Specificity (d) Sensitivity

(e) Accuracy 2C, cumulative mean (f) Accuracy 4C, cumulative mean (g) Specificity, cumulative mean (h) Sensitivity, cumulative mean

Figure 3. These plots show the 2C and 4C accuracies as well as the specificities and the sensitivities. Whereas the top row ((a) - (d)) provides the rates for each individual expert, the bottom row ((e) - (h)) shows the cumulative rates.

4 Discussion

Obviously, diseased patients are slightly more likely to be detected in case of distortion corrected images. This diametric behavior might be due to the fact, that distortion correction leads to partly blurred images. Thereby the villi structure which indicates a healthy mucosa, might by hidden and thus false positives are provoked. However, to put it in a nutshell, distortion correction does not significantly enhance any of the classification properties and on average the classification accuracy drops.

Furthermore, some other quite interesting aspects can be deduced from the results. Considering image patches instead of complete images, the discriminative power in case of human experts significantly drops. Especially the sensitivity (see Fig. 3(d)) of the best experts falls from about 90 % to about 70 % and below. Obviously for the physicians it is crucial not just to see a small texture patch, but to have knowledge of the whole image.

Moreover we notice that the highest overall accuracies are achieved by those medical doctors with the highest specificities. However, the highest sensitivities are achieved by those with lower accuracies. Seemingly the less experienced doctors (with lower accuracies) tend to less false negatives but more false positives.

Finally we consider the rates achieved by the computer aided method. In the two-classes case, in combination with the image patches (which are also used by the computer based method), the best 6 experts are able to outperform this method (see 3(a)). In case of the accustomed classification of complete images, the human experts are significantly better. In this scenario, all but one experts (partially significantly) outperform the computer based approach. On average (see 3(e), rightmost values), the computer based approach is significantly below the complete-image-based and also slightly below the patch-based human expert's classification accuracy. Considering the four-classes case, the computer aided approach turns out to be much more competitive compared to the performances of the human experts. Only two experts achieve higher accuracies with complete images and only one of the experts is similarly accurate in case of patch images. On average, the computer based approach significantly outperforms the human experts. It is not sensible to compare the specificity and the sensitivity of the computer aided method, as the used classifier is optimized to achieve high accuracies (and a balanced ratio between sensitivities and specificities).

The distortion estimation experiment (see Fig. 2) reveals a quite interesting result. The majority of experts (7 persons) is not aware of the barrel-type distortions and chose the leftmost distortion-free image. The images chosen by the others (4 persons) are shown in Fig. 2 (second until fifth image). Each distorted pattern has been chosen once.

Obviously the majority of physicians are not aware of the barrel-type distortions introduced by the endoscopes. Seemingly this does not strongly affect their classification performance, as a distortion correction leads to worse results. However, in this work, we are not able to precisely estimate the effect of the lens distortions as the distortion correction leads to new inadequacies and there is nothing to prevent these distortions in the original endoscopic images.

4.1 Conclusion

This paper showed, that the human classification performance definitely does not benefit from barrel-type distortion correction. Quite the contrary, in almost all cases, the classification rates drop significantly. We also showed, that the computer based approach is highly competitive if considering the four-classes case. If considering the two-classes case, the human experts on average are more reliable. The human experts' performances significantly benefit from complete images instead of the smaller patches. Therefore, we encourage to develop a computer aided method, being based on the complete images to additionally exploit the (obviously important) larger image context.

Acknowledgment

This work is partially funded by the Austrian Science Fund (FWF) under Project No. 24366.

References

1. Fasano A, Berti I, Gerarduzzi T, Not T, Colletti RB, Drago S, et al. Prevalence of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study. Archives of internal medicine. 2003;163(3):286–92.
2. Cammarota G, Cesaro P, Martino A, et al. High accuracy and cost-effectiveness of a biopsy-avoiding endoscopic approach in diagnosing coeliac disease. Alimentary Pharmacology and Therapeutics. 2006;23(1):61–69.
3. Kiesslich R, Mergener K, Naumann C, et al. Value of chromoendoscopy and magnification endoscopy in the evaluation of duodenal abnormalities: a prospective, randomized comparison. Endoscopy. 2003;35(7):559–563.
4. Badreldin R, Barrett P, Wooff DA, et al. How good is zoom endoscopy for assessment of villous atrophy in coeliac disease? Endoscopy. 2005;37(10):994–998.
5. 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. 2013;112(3):694–712.
6. Hegenbart S, Kwitt R, Liedlgruber M, Uhl A, Vécsei A. 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); 2009. p. 718–723.
7. 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. 1999;11(10):1185–1194.
8. Gadermayr M, Liedlgruber M, Uhl A. Shape Curvature Histogram: A Shape Feature for Celiac Disease Diagnosis. In: Proceedings of the 3rd International Workshop on Medical Computer Vision (MICCAI-MCV); 2013. Accepted.
9. Melo R, Barreto JP, Falcao G. A new solution for camera calibration and real-time image distortion correction in medical endoscopy-initial technical evaluation. IEEE Trans Biomed Eng. 2012;59(3):634–44.
10. Fitzgibbon AW. Simultaneous linear estimation of multiple view geometry and lens distortion. In: Computer Vision and Pattern Recognition (CVPR); 2001. p. 125–132.