

Superpixel-based automatic segmentation of villi in confocal endomicroscopy

D. Boschetto^{1,2}, H. Mirzaei³, R. W. L. Leong³, E. Grisan²

Abstract—Confocal Laser Endomicroscopy (CLE) is a technique permitting on-site microscopy of the gastrointestinal mucosa after the application of a fluorescent agent, allowing the evaluation of mucosa alterations. These are used as features by skilled technicians to stage the severity of multiple diseases, celiac disease or irritable bowel syndrome among the others. We present an automatic method for villi detection from confocal endoscopy images, whose appearance changes with mucosal alterations. Superpixel segmentation, a well-known technique originating from computer vision, is used to identify and cluster together pixels belonging to uniform regions. Each image in the dataset is analyzed in a multiscale fashion (scale 1, 0.5 and 0.25). From each superpixel, 37 features are extracted at multiple image scales. Each superpixel is classified using a random forest, and a post-processing step is performed to refine the final output. Results in the test set (70 images, 30870 superpixels) show 85.87% accuracy, 92.88% sensitivity, 76.99% specificity in the superpixel space, and 86.36% of accuracy and 87.44% Dice score in the pixel domain.

Index Terms—SLIC superpixel, confocal endomicroscopy, automatic identification, segmentation, classification

I. INTRODUCTION

The mucosa of the gastrointestinal tract represents the main barrier between the inner body and the external world. A layer of cells runs from the esophagus to the rectum, playing a key role in preventing access to environmental hostile factors that could cause inflammation. In particular, the intestinal epithelium is the largest mucosal surface, regulating the transit of macromolecules [1], [2]. This barrier is formed by a double layer of lipid cells, offering strong resistance to water soluble constituents. The junction between epithelial cells is a region in which inter-cellular junctions (tight junctions) are formed, to regulate the constituents' flow. This junction's permeability is dynamic, according to dietary state, humoral or neural signals and inflammatory mediators, among others. If pathological conditions ensue, the permeability is increased and a loss of epithelial integrity is suffered. Impaired epithelial barrier function is present in irritable bowel disease (IBD), irritable bowel syndrome (IBS), Crohn's disease and ulcerative colitis.

Features have been introduced in the literature to assess

intestinal permeability [3], [4]. In particular, for IIP, three features have been defined: Cell Drop-out (CDO), shedding of an enterocyte into the luminal space; Cell Junction Enhancement (CJE), a fluorescein build-up between two epithelial cells representing impaired tight-junction proteins before breakage of the final basal junction, and FL: a fluorescein plume entering the lumen representing loss of apposition between two adjacent cells.

Another important disease affecting the GI tract is celiac disease (CD), an immune-mediated enteropathy affecting genetically susceptible persons triggered by exposure to gluten and similar proteins. It is one of the most frequent enteropathies and is a hidden epidemic, since most of the celiac patients will remain undiagnosed during their life. Exposure to gluten causes variable damage to small bowel mucosa: mild damage include cases with increased number of intraepithelial lymphocytes and the presence of Crypt Hyperplasia (CH), while severe forms of the lesions involve various degrees of endoscopically relevant lesions such as villous Atrophy (VA) [5]. Overall sensitivity and positive predictive values of VA and CH are poor even when zoom endoscopy is used [6], implying that these two alterations of the mucosa are not easily recognized during endoscopy. Thus, in everyday practice, the identification of CD is made on the basis of a positive diagnostic intestinal biopsy and of the concomitant presence of a positive celiac serology [7]. The gold standard in the diagnosis of CD is the demonstration of VA in duodenal biopsies, a feature extensively investigated in the medical community [8], [9]. Image processing methods as well as quantitative computational methods are highly needed, required and recommended from the community for the characterization of the small intestinal mucosa in suspected and known CD patients [10].

Confocal Laser Endomicroscopy (CLE) is a technique permitting on-site microscopy of the gastrointestinal mucosa after the application of a fluorescent agent, allowing the evaluation of mucosal alterations [11]. Images originating from CLE, as Fig. 1 shows, are very informative about the status of the mucosa: depending on the region under analysis and each patient's health status, all the previously introduced features can possibly be discerned, but require manual labeling, skilled technicians and lengthy times to be quantified and scored. All these features rely on the evaluation of shape and texture of mucosal villi. In CLE images, though, villi can exhibit smooth and fuzzy borders among (and between) villi and inter-villous space, and vessels can be found in inter-villous space. In severe CD stages, a possible collapse of all villi into a uniform mucosa that is depleted

¹D. Boschetto is with IMT Institute for Advanced Studies Lucca, 55100 Lucca, Italy and with the Department of Information Engineering of the University of Padova, 35131 Padova, Italy.

²E. Grisan is with the Department of Information Engineering of the University of Padova, 35131 Padova, Italy.

³H. Mirzaei and R. W. L. Leong are with Gastroenterology and Liver Services, Sydney South West Area Health Service, Bankstown Hospital, Faculty of Medicine, The University of New South Wales, Sydney, Australia. email: davide.boschetto@imtlucca.it, mirzaei.hadis@yahoo.com, rupertleong@hotmail.com, enrico.grisan@dei.unipd.it

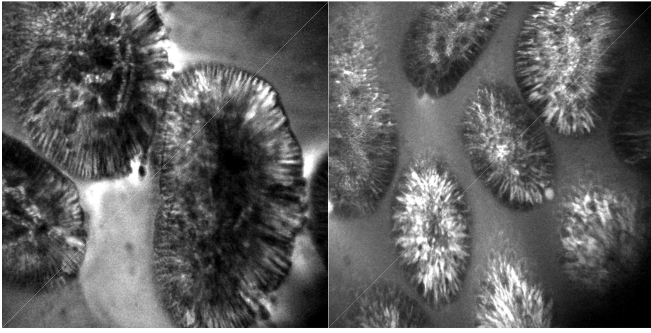


Fig. 1. Two images from the dataset, showing very heterogeneous structures and illumination.

of villi can be observed. With IBS/IBD, fluorescein leakage can occur in the lumen. Other than this, crypts or impaired junctions can prevent accurate generalization with standard image processing methods.

We present an automatic method for villi detection from confocal endoscopy images. An automatic way to identify such villi in CLE images might accelerate the adoption of quantitative processes to evaluate features and score the severity of various diseases in a faster and more robust way. This work is an improvement on our previous work [12], which based its roots in morphological processing for achieving semi-automatic villi identification for feature extraction, but suffering when tested on highly heterogeneous datasets. Villi identification is not a trivial process, given that these structures are highly textured and present high variability in appearance, shape and dimension.

II. MATERIALS

In this study, 155 confocal images were obtained from a previous clinical trial conducted at the Gastroenterology and Liver Services of the Bankstown-Lidcombe Hospital (Sydney, Australia) [3]. Each patient underwent a confocal gastroscopy (Pentax EC-3870FK, Pentax, Tokyo, Japan) under conscious sedation and with an intra-venous aliquots of fluorescein sodium and topical acriflavine hydrochloride to enhance images. Each image represent a mucosal region of 0.5×0.5 mm, with an in-plane resolution of $2 \text{ pixel}/\mu\text{m}$, resulting in images of 1024×1024 pixels. As Fig. 1 shows, images conveying very heterogeneous information have been selected for this study, for generalization purposes. Among the three CLE features (fluorescein leakage, cell drop-out and cell junction enhancement), each image of the dataset exhibit only one feature. In total, we have 29 CDO images, 65 CJE and 61 FL images. A random selection of 70 images has been used for training (training set). Another random selection of 15 of the remaining images were used to tune the post-processing analysis, a step that will be detailed in the following section. The remaining 70 images were used for testing the method’s performance. In order to provide a ground truth, all images have been manually analyzed by an expert, providing an outline of each visible villus in the image.

III. METHODS

The first step in the proposed method aims at the construction of a rough segmentation identifying a candidate region of the image with the highest possibility of being part of a villous fold. This is performed by processing the image with a computer vision technique called superpixel segmentation, in particular using the SLIC implementation [13]. The purpose of this process is to create clusters of spatially connected pixels exhibiting similar texture. Each of the superpixels is then analyzed, and 37 features are extracted from each of them, to be fed to a classifier. A multiscale analysis is performed, by computing and analyzing three versions of the original image (original size, plus two rescaled versions by a factor of $1/2$ and $1/4$ respectively), bringing the total size of the feature vector for each superpixel to 111. The classification step is performed with an ensemble of 50 decision trees, trained on 70 random images from the dataset. A post-processing refinement step of the computed prediction is performed to improve the accuracy, tuned on a sub-sample of the image dataset (15 images, referred to as “tuning set”) to maximize ground truth adherence and prediction accuracy. The algorithm is then tested on the remaining 70 images.

A. Superpixel segmentation

As a pre-processing step for each image, all greyscale values were normalized between 0 and 256, and a median filter was then applied to reduce noise. Segmentation via superpixel is then performed by grouping pixels into perceptually meaningful atomic regions, used to replace the rigid structure of the pixel grid. Many computer vision algorithms use superpixels as their building blocks [14], [15], given their straightforwardness and the ease of their implementation. A commonly used superpixel implementation is the Simple Linear Iterative Clustering (SLIC) [13]: this implementation, based on k-means clustering, is fast to compute, memory efficient, simple to use, and outputs superpixels that adhere well to image boundaries. SLIC implementation clusters pixels of the image to efficiently generate compact and nearly uniform superpixels, imposing a degree of spatial regularization to extracted regions. Two typical images from the dataset, with superpixels superimposed, are shown in Fig. 2. This step has been implemented with MATLAB R2015b, using an implementation of SLIC superpixels by *vfeat* [16]. This technique only requires two parameters to set: the desired size of each superpixel N and a regularization parameter λ , that tweaks the smoothness of their contours. Once the superpixel segmentation is obtained, the manual ground truth is transformed in superpixel space, as Fig. 3 illustrates. Each region of this image (corresponding to each computed superpixel) is labeled as part of a villous fold if, for that superpixel, the ratio among villous-labeled pixels and background-labeled pixels is greater than a threshold R , whose value is computed as explained in Sec. IV.

B. Feature Extraction

For each image in the dataset, three different scales are analyzed for feature extraction: the image at the original

scale, and two rescaled versions of it by factor of 1/2 and 1/4, respectively. In this way, a multiscale analysis of each image is performed, to improve robustness of the classification and to avoid possible errors due to texture similarities at the original scale. A total of 111 features are extracted for the multiscale analysis of each superpixel S , 37 for each image scale:

- Mean intensity μ_S and standard deviation σ_S : greyscale intensity variations can be useful features to differentiate among villous folds (i.e., foreground) and mucus (i.e., background);
- Contrast C_S , Energy E_S and Homogeneity H_S from the Gray Level Co-Occurrence Matrix (GLCM): GLCM is a statistical method of examining texture considering the spatial relationship of pixels. It calculates how often pairs of pixels with specified values and spatial locations occur in an image, building a 8×8 occurrence matrix. Extracting statistical measures from this matrix provide information about the specific texture. From this analysis, contrast (local variations in the GLCM), energy (sum of squared elements in GLCM) and homogeneity (how close the distribution of the elements in the GLCM is to its diagonal values) measures have been included in the feature set;
- Histogram of Local Binary Patterns [17] with 32 bins, $hLBP_S$. Local Binary Patterns (LBP) are one of the most descriptive features in the field of texture classification, and are commonly used in computer vision. They permit the creation of features able to identify different textures in an image. In this work, for each pixel of the image, an 8-bit word is created by comparing its value with the ones in its 8-neighborhood. Iteratively, starting from a fixed direction, if the central pixel has a grayscale value greater than its neighbor a 1 is encoded, a 0 otherwise. A histogram (32 bins) is computed for the LBP in each superpixel, expressing the spectrum of the texture of the selected portion of the image, and the result is added to the feature vector.

C. Classification with random forests

For each superpixel, the probability of it being part of a villus fold is computed as the score of a binary random forest

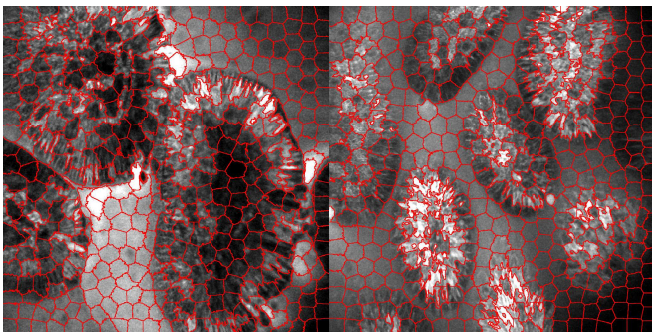


Fig. 2. Two images extracted from the dataset, with SLIC Superpixels superimposed.

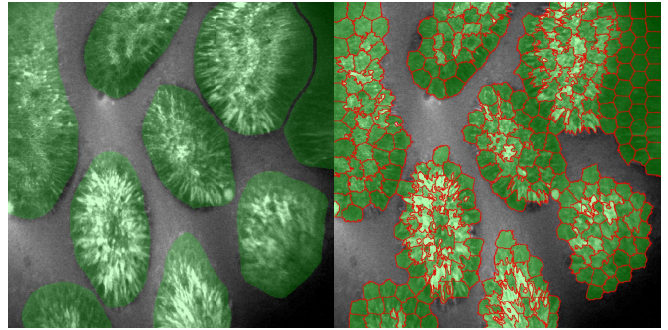


Fig. 3. From manual ground truth (left) to superpixel-based ground truth (right).

classifier using 50 classification trees. The training process has been performed using 30870 superpixels belonging to the 70 images from the training set.

D. Final refinement and results

After the classification step, binary prediction masks have been created according to the score assigned to each superpixel by the classifier. To discard isolated superpixels selected as villi, all connected regions smaller than P pixels have been excluded from the prediction masks, and holes in the binary masks were filled, to compensate for obvious false negatives in the classification step. The tuning of this final refinement process was based on the tuning set, excluded from both the training and the testing phase of the classifier. Accuracy, sensitivity and specificity of the classification step have been computed, both in superpixel and in pixel space, along with the Dice scores between the prediction masks and the superpixel based ground truth.

IV. RESULTS

Superpixel parameters were set as $N = 50$, $\lambda = 0.05$ to obtain a number of around 440 superpixels per image, each of them resulting well adherent to image borders. The value of $P = 15742$ was tuned by selecting the maximum area (smaller than $P_{lim} = 16900$ pixels, as a hard-coded safety value based on villi's sizes from the tuning set, corresponding to a patch of 130×130 pixels) among all false positive villi identified in the tuning set. The value of $R = 0.5$ was set by maximizing Dice correlation among the labeled ground truth in pixel space and the one in superpixel space in the images of the training set (Dice score between pixel-space GT and superpixel-space GT at $R = 0.5$ is 96.3%). To quantify the performance of our method, the Dice coefficient for each image is computed by comparing the prediction masks with the respective ground truth both in superpixel space and pixel space. Results are reported in Table I and Table II for superpixel-wise and pixel-wise analysis respectively. The proposed method has been tested on 70 images (a total of 336 villi), and reached an average general accuracy of 85.9%. Sensitivity (True Positive Rate, TPR) is 92.9%, while specificity (True Negative Rate, TNR) is 77.0%. Mean Dice values between each prediction mask and its ground truth in superpixel space is of 87.4%, and the pixel-wise total

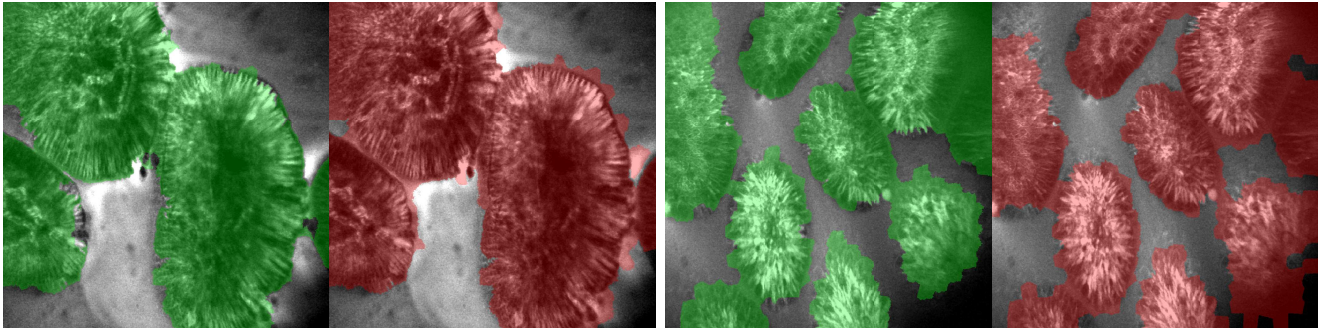


Fig. 4. (a-c) ground truth labeling and (b-d) villi estimated area superimposed on two different images from the dataset.

classification accuracy in pixel space is 86.4%. Sensitivity and specificity referring to the pixel domain are, respectively, 93.50% and 71.59%. Fig. 4 shows two images from the dataset and the true (a-c) and detected (b-d) villi for visual comparison, in superpixel space.

V. CONCLUSIONS

We have presented an automatic method to segment and detect villi in the epithelium of the gastrointestinal tract, using SLIC superpixel segmentation, local binary patterns and a random forest classifier. Automatic villi detection could lead to automatic quantitative analysis for staging and grading diseases such as celiac disease and irritable bowel disease. With the aid of the method presented in this work, experts will be able to quantitatively analyze images and compare patterns among pathological and non-pathological patients. The tools developed for this work will be refined and exploited in future studies for quantitative analysis and measurements using CLE in patients suffering from celiac disease or irritable bowel syndrome.

Accuracy _{SP}	TPR _{SP}	TNR _{SP}
85.9	92.9	77.0

TABLE I

TRUE POSITIVE RATE AND TRUE NEGATIVE RATE COMPUTED SUPERPIXEL PER SUPERPIXEL, COMPARING CLASSIFICATION AND GROUND TRUTH.

Accuracy _P	TPR _P	TNR _P	Dice _P
86.4	93.5	71.6	87.4

TABLE II

TRUE POSITIVE RATE AND TRUE NEGATIVE RATE COMPUTED PIXEL PER PIXEL, COMPARING PREDICTION MASKS AND GROUND TRUTH.

REFERENCES

- [1] MC. Arrieta, L. Bistriz, and JB. Meddings, "Alterations in intestinal permeability," *Gut*, vol. 55, no. 10, pp. 1512–1520, Oct 2006.
- [2] J. Visser, J. Rozing, A. Sapone, K. Lammers, and A. Fasano, "Tight junctions, intestinal permeability, and autoimmunity: celiac disease and type 1 diabetes paradigms," *Ann. N. Y. Acad. Sci.*, vol. 1165, pp. 195–205, May 2009.
- [3] J. Chang, M. Ip, M. Yang, B. Wong, T. Power, L. Lin, W. Xuan, TG. Phan, and R. Leong, "The learning curve, interobserver, and intraobserver agreement of endoscopic confocal laser endomicroscopy in the assessment of mucosal barrier defects," *Gastrointest. Endosc.*, Sep 2015.
- [4] R. Kiesslich, CA. Duckworth, D. Moussata, A. Gloeckner, LG. Lim, M. Goetz, DM. Pritchard, PR. Galle, MF. Neurath, and AJ. Watson, "Local barrier dysfunction identified by confocal laser endomicroscopy predicts relapse in inflammatory bowel disease," *Gut*, vol. 61, no. 8, pp. 1146–1153, Aug 2012.
- [5] C. Mulder, S. van Weyenberg, and M. Jacobs, "Celiac disease is not yet mainstream in endoscopy," *Endoscopy*, vol. 42, no. 3, pp. 218–9, 2010.
- [6] D. Dewar and P. Ciclitira, "Clinical features and diagnosis of celiac disease," *Gastroenterology*, vol. 128, no. Suppl1, pp. S19–S24, 2005.
- [7] A. Fasano and C. Catassi, "Current approaches to diagnosis and treatment of celiac disease: an evolving spectrum," *Gastroenterology*, vol. 120, no. 3, pp. 636–51, 2001.
- [8] EJ. Ciaccio, SK. Lewis, and PH. Green, "Detection of villous atrophy using endoscopic images for the diagnosis of celiac disease," *Digestive Diseases and Sciences*, vol. 58, no. 8, pp. 1167–9, 2013.
- [9] M. Bonamico, P. Mariani, E. Thanasi, M. Ferri, R. Nenna, C. Tiberti, B. Mora, MC. Mazzilli, and FM. Magliocca, "Patchy villous atrophy of the duodenum in childhood celiac disease," *J Pediatr Gastroenterol Nutr*, vol. 38, no. 2, pp. 204–7, 2004.
- [10] EJ. Ciaccio, G. Bhagat, SK. Lewis, and PH. Green, "Quantitative image analysis of celiac disease," *World J Gastroenterol*, vol. 21, no. 9, pp. 2577–81, 2015.
- [11] K. Venkatesh, A. Abou-Taleb, M. Cohen, C. Evans, S. Thomas, P. Oliver, C. Taylor, and M. Thomson, "Role of confocal endomicroscopy in the diagnosis of celiac disease," *J Pediatr Gastroenterol Nutr*, vol. 51, no. 3, pp. 274–9, 2010.
- [12] D. Boschetto, H. Mirzaei, R. Leong, and E. Grisan, "Semiautomatic detection of villi in confocal endoscopy for the evaluation of celiac disease," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2015, Engineering in Medicine and Biology Society (EMBC).
- [13] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans Pattern Anal Mach Intell*, vol. 34, no. 11, pp. 2274–2282, Nov 2012.
- [14] B. Fulkerson, A. Vedaldi, and S. Soatto, "Class segmentation and object localization with superpixel neighborhoods," 2009, pp. 670–677, Computer Vision, 2009 IEEE 12th International Conference on.
- [15] Y. Li, J. Sun, CK. Tang, and HY. Shum, "Lazy snapping," 2004, vol. 3, pp. 303–308, ACM Transactions on Graphics (SIGGRAPH).
- [16] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," <http://www.vlfeat.org/>, 2008.
- [17] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Gray scale and rotation invariant texture classification with local binary patterns..," 2000, In: Computer Vision, ECCV 2000 Proceedings, Lecture Notes in Computer Science 1842, Springer, 404 - 420.