**Superpixel-based classification of gastric chromoendoscopy images**

Davide Boschettoa, Enrico Grisan\*b

a IMT Institute for Advanced Studies Lucca, Italy;
b Department of Information Engineering, University of Padova, Padova, Italy;

Abstract

Chromoendoscopy (CH) is a gastroenterology imaging modality that involves the staining of tissues with methylene blue, which reacts with the internal walls of the gastrointestinal tract, improving the visual contrast in mucosal surfaces and thus enhancing a doctor’s ability to screen precancerous lesions or early cancer. This technique helps identify areas that can be targeted for biopsy or treatment and in this work we will focus on gastric cancer detection. Gastric chromoendoscopy for cancer detection has several taxonomies available, one of which classifies CH images into three classes (normal, metaplasia, dysplasia) based on color, shape and regularity of pit patterns.

Computer-assisted diagnosis is desirable to help us improve the reliability of the tissue classification and abnormalities detection. However, traditional computer vision methodologies, mainly segmentation, do not translate well to the specific visual characteristics of a gastroenterology imaging scenario. We propose the exploitation of a first unsupervised segmentation via superpixel, which groups pixels into perceptually meaningful atomic regions, used to replace the rigid structure of the pixel grid. For each superpixel, a set of features is extracted and then fed to a random forest based classifier, which computes a model used to predict the class of each superpixel. The average general accuracy of our model is 92.05% in the pixel domain (86.62% in the superpixel domain), while detection accuracies on the normal and abnormal class are respectively 85.71% and 95%. Eventually, the whole image class can be predicted image through a majority vote on each superpixel's predicted class.

**Keywords:** chromoendoscopy, gastric cancer, superpixel, classification

# INTRODUCTION

Gastroenterology (GE) imaging is an essential technique for abnormalities identification in the gastrointestinal (GI) tract, given the fact that it is minimally invasive and relatively painless. It is a powerful diagnostic tool for critical pathologies, being useful especially for early stages disease detection, such as Crohn’s disease, Barrett’s esophagus, adenomas and celiac disease. It offers physicians the chance to extract relevant clinical information such as the presence of lesions or polyps that are typically confined to a specific region of an image. Digital endoscopes created a new field in the last decade in medical endoscopy (Computer-assisted decision support systems, or CADSS). These can be useful in improving the clinical experts’ ability to screen the GI tract by providing a second opinion, improving the accuracy of medical diagnosis and keeping the expert at pace with technology. Most traditional CADSSs rely on manual segmentation for their purpose to be met. Segmentation is required because of the nature of the procedure: the physician is in fact interested in inspecting a clinically relevant region in an image, and to encode this information segmentation is required (see Fig. 1). Only these regions are the ones that convey clinically relevant information that can be useful for diagnosis, while the rest of the image can be discarded. The heterogeneous nature of GE images make this process a hard task, given the high variability among camera conditions and imaging sites, even in the same imaging session. Because of this, manual ground truth can be used as a starting point in a CADSS. A fast automatic way to classify images based on the manual ground truth could automatically give suggestions to medical experts and point out trends that might be missed even with a trained naked eye.

# MATERIALS

The chromoendoscopy imaging (CH) data used have been made freely available [1], and were originally collected using an Olympus GIF-H180 endoscope at the Portuguese Institute of Oncology (IPO) Porto, Portugal during routine clinical work as described in [2-4]. Briefly, the endoscopic videos were recorded during routine endoscopic examinations, obtaining approximately 4 hours of video (360 000 images) from 28 patients. An expert reviewer selected the clearest shot related to each particular observation of tissue of interest. A total of 176 images were initially selected and then provided to two physicians for an independent analysis, and annotated to identify clinically relevant patches in the image (region of interest—ROI), classification of group, classification of subgroup [2,4] and confidence level of the physician for group assignment: a high-quality annotated dataset of 176 images was defined. The relevant patches of each image were classified into two groups based on dye captation and the appearance of their pit-pattern. When the images have a regular gastric mucosal patterns and no change in color after staining with methylene blue is observed, they are classified as belonging to Group I. If the mucosa are regular and tissues are stained in blue, they belong to Group II. If neither a clear pattern was noticeable nor a change in color was observed, the image belongs to Group III. For the purpose of the present work, group I images are considered normal and group II cases are considered metaplasia and dysplasia lesions (respectively), therefore labeled as abnormal.

|  |  |
| --- | --- |
| C:\Users\dvdbo\AppData\Local\Microsoft\Windows\INetCache\Content.Word\CHGastro_Normal_092.png(a) | C:\Users\dvdbo\AppData\Local\Microsoft\Windows\INetCache\Content.Word\CHGastro_Abnormal_055.png(b) |
| C:\Users\dvdbo\AppData\Local\Microsoft\Windows\INetCache\Content.Word\CHGastro_Normal_092_ROI.PNG(c) | C:\Users\dvdbo\AppData\Local\Microsoft\Windows\INetCache\Content.Word\CHGastro_Abnormal_055_ROI.PNG(d) |

Figure 1. Representative (a) normal image and (b) abnormal image from the chromoendoscopy dataset.
(c-d) Example of manually segmented region of interest for (a-b).

# methods

Similarly to the approach proposed in [5], the first step in the proposed method is performed by processing the image with a computer vision technique called superpixel segmentation, using the SLIC implementation [6, 7]. The purpose of this process is to create clusters of spatially connected pixels exhibiting similar texture.

Each of the superpixels is then analyzed, and 111 features are extracted from each of them, to be fed to a classifier.

The classification step is performed with an ensemble of random decision trees. Each image is finally labeled as Normal or Abnormal based on a majority vote among the predicted label of all its superpixels.

## 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, given their straightforwardness and the ease of their implementation.

A commonly used superpixel implementation is the Simple Linear Iterative Clustering (SLIC): 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.

This step has been implemented with MATLAB R2015b, using an implementation of SLIC superpixels by vlfeat [7]. This technique only requires two parameters to set: the desired size of each superpixel $N$ and a regularization parameter $λ$, that controls the smoothness of their contours. Each region of this image (corresponding to each computed superpixel $S$) is then analyzed separately for feature extraction.



Figure 2 A representative CH image of the dataset with the corresponding superpixel subdivision. The white lines correspond to superpixels boundaries

.

## Features Extraction

A total of 111 features are extracted from the analysis of each superpixel $S$, 37 for each image color plane:

* **Statistical:** Mean intensity $μ$ and standard deviation $σ$: greyscale intensity variations are the most basic difference among normal and abnormal tissue;
* **Co-occurrence**: Contrast $C\_{S}$, Energy $E\_{S}$ and Homogeneity $H\_{S}$ from the Gray Level Co-Occurrence Matrix (GLCM) [8]: 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 an $8x8$ 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;
* **Local Binary Patterns**: Histogram of Local Binary Patterns [9] 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 greyscale intensity 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 in the 8-bit word, a 0 otherwise. When a word has been assigned to each pixel, each word is translated to decimal (0-256). A histogram (32 bins) is then computed for the LBP of pixels in each superpixel $S$, expressing in such way the spectrum of the texture of the selected portion of the image. This is finally added to the feature vector.

## Classification with random forests

For each image, the probabilities that it presents normal or metaplastic classification are computed as the score of a binary random forest classifier [10]. Half of the image set has been used as training set, while the other half was used as test set. The random forest classifier if build by setting the number of trees to 50. Each set (both training and test sets) was composed by 88 images, divided between 28 images belonging to the ‘Normal’ class and 60 belonging to the ‘Abnormal’ class.

# EXPERIMENTS AND RESULTS

Superpixel parameters were set as $N = 90$ and $λ=0.05$ to obtain a large training dataset and reasonable-sized superpixels, each of them resulting well adherent to image borders.

The evaluation of the performance has been carried out computing classification accuracy of the segmentation either in superpixel space or in pixel space. In superpixel space, accuracy is the fraction of superpixels assigned to the correct class, with respect to the superpixels ground truth. This is obtained by computing the number of pixels manually annotated as normal or abnormal tissue, and assigning to each superpixel the class with major representation. In image space, the accuracy is computed as the fraction of pixels correctly assigned to the correct class.

The proposed method reached an average general accuracy of 92.05% in the image space, 86.62% in the superpixel space, respectively over 88 images and 1173 superpixels. A chart showing how accuracy depends on superpixel size is shown in Fig. 2. The smaller the superpixel, the longer the computation time in the feature extraction and training process. We settled for a superpixel size of 90, leading reasonable results with a relatively fast training and testing processing time. As an example, Fig. 1 shows the computational times for feature extraction with some superpixel sizes (on an Intel i7-3630QM @ 2.40 GHz with a Samsung 840 EVO SSD drive and 8 GB RAM).

When considering the classification of the whole image, considering only the superpixels within the manually annotated region of interest, detection accuracy of the normal class is 85.71%, while detection accuracy of the abnormal class is 95%. These two metrics in the superpixel domain are, respectively, 79.95% and 90.39%. An additional parameter that could be adjusted for increasing the performance is the ratio among normal and abnormal superpixels needed to classify one image into normal or abnormal. In Fig. 3, it is shown that by requiring less than half superpixels to classify an image as abnormal, general accuracy changes, along with normal and abnormal class accuracies. It is clear from the image that altering this ratio favours the accuracy in one of the two classes (the favoured one), but does not help in general accuracy. The only drawback of this method is if ROIs are so small that no superpixels fit inside them: in such cases, no superpixel from that image is used either for training or testing, leading to a failure in that particular image. However, the limitation could be tackled by reducing the superpixels size: for all computations with superpixel size greater than 70, all images gave contribution in the training process.



Figure 3. Average computation time for loading the two images (image and ROI) and for computing the superpixels and their features, with varying superpixel size. Times in second per image, obtained by averaging on the full computation on 88 images.



Figure 4. Effect of superpixel size parameter on the average (testing) accuracy. Solid lines represent accuracies computed at the pixel level, dotted lines at the superpixels level.

References

1. Analysis of Images to Detect Abnormalities in Endoscopy Challenge, IEEE-ISBI 2016, https://isbi-aida.grand-challenge.org/home/
2. M. Dinis-Ribeiro. Clinical, endoscopic and laboratorial assessment of patients with associated lesions to gastric adenocarcinoma. PhD thesis, Faculdade de Medicina da Universidade do Porto, 2005
3. F. Riaz, F. B. Silva, M. D. Ribeiro, and M. T. Coimbra. Invariant Gabor texture descriptors for classification of gastroenterology images. IEEE Trans Biomed Eng, 59(10):2893–2904, Oct 2012
4. F. Riaz, F. B. Silva, M. D. Ribeiro, and M. T. Coimbra. Impact of Visual Features on the Segmentation of Gastroenterology Images Using Normalized Cuts. IEEE Trans Biomed Eng, 60(5):1191–1201, May 2013
5. D. Boschetto, G. Gambaretto, E. Grisan, Automatic Classication of Endoscopic Images for Premalignant Conditions of the Esophagus, Proc. SPIE 9788, Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging, 978808 (March 29, 2016)
6. 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.
7. Fulkerson, B., Vedaldi, A., and Soatto, S., Class segmentation and object localization with superpixel neighborhoods," 670{677 (2009). Computer Vision, 2009 IEEE 12th International Conference on
8. A.Vedaldi and B. Fulkerson, “VLFeat: An open and portable library of computer vision algorithms,” MM’10, Proceedings of the 18th ACM international conference on Multimedia, 1469-1472, October 25–29, 2010, Firenze, Italy.
9. Haralick, R.M., and L.G. Shapiro. Computer and Robot Vision: Vol. 1, Addison-Wesley, 1992, p. 459.
10. T. Ojala, M. Pietikainen and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971 - 987, 2002
11. Breiman, L., Random forests," Machine Learning 45, 5:32 (2001)