2013 IEEE 10th International Symposium on Biomedical Imaging:
From Nano to Macro
San Francisco, CA, USA, April 7-11, 2013

CROHN'S DISEASE TISSUE SEGMENTATION FROM ABDOMINAL MRI USING SEMANTIC INFORMATION AND GRAPH CUTS

Dwarikanath Mahapatra[1,*], *Peter J. Schüffler*[1], *Jeroen A.W. Tielbeek*[2], *Franciscus M. Vos*[2,3], *Joachim M. Buhmann*[1]

[1]Department of Computer Science, ETH Zurich, Switzerland.
[2] Department of Radiology, Academic Medical Center, The Netherlands.
[3] Quantitative Imaging Group, Delft University of Technology, The Netherlands.
[*]*dwarikanath.mahapatra@inf.ethz.ch*

ABSTRACT

We propose a graph cut based method to segment regions in abdominal magnetic resonance (MR) images affected with Crohn's disease (CD). Intensity, texture, curvature and context information are used with random forest (RF) classifiers to calculate probability maps for graph cut segmentation. The RF classifiers also provide semantic information used to design a novel smoothness cost. Experimental results on 26 real patient data shows our method accurately segments the diseased areas. Inclusion of semantic information significantly improves segmentation accuracy and its importance is reflected in quantitative measures and visual results.

Index Terms— Crohn Disease, graph cut, semantic information, Random forests, context, MRI.

1. INTRODUCTION

Crohn's disease (CD) is an autoimmune disease that can affect any part of the gastrointestinal tract and leads to abdominal pain, weight loss, diarrhea and vomiting. Current diagnosis involves colonoscopy which is interventional and painful. This has motivated efforts to explore computational approaches for detecting and quantifying extent of CD [1, 2] in magnetic resonance imaging (MRI) data. Building on our work in [3], we propose a graph cut based segmentation approach that uses learned semantic information (from Random forest (RF) classifiers) to automatically segment CD regions in abdominal MRI.

Semantic segmentation assigns each image pixel to a object class and has invoked interest in the medical imaging community because of its suitability in detecting diseased anatomical structures [4, 5]. Automatic voxel-wise labeling of medical images is a major challenge because of the similar intensity of different organs, varying resolutions and complex anatomy. Therefore machine learning approaches have gained significance in such problems. Iglesias et al. [4] use active learning to choose samples that improve the classifier's accuracy and segment upto 9 anatomical structures from human computed tomography (CT) scans. Montillo et al. [5] use decision forests for semantic segmentation of multiple organs from CT images. A machine learning approach for detecting colorectal cancer is taken in [6].

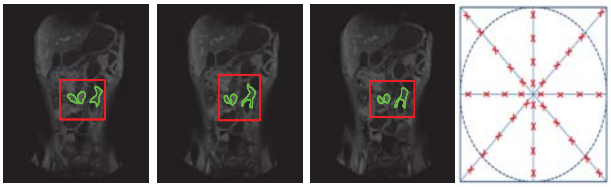
In this paper we propose to segment CD affected tissues using graph cuts and RF classifiers. RF classifiers use image features to calculate likelihood of each voxel being diseased, normal or background tissue. They also provide semantic information for designing a novel smoothness cost. This paper makes the following contributions: 1) in addition to the low level features described in [3], we propose a novel context feature; 2) novel smoothness measures (based on semantic information derived from RF classifiers) are used in a Markov random field (MRF) cost function which is then optimized using graph cuts. We describe our method in Section 2, present results in Section 3 and conclude with Section 4.

2. METHODS

Our method consists of identifying a smaller volume of interest (VOI) from the original test volume, generating probability maps for each voxel within the VOI and segmenting the diseased tissues.

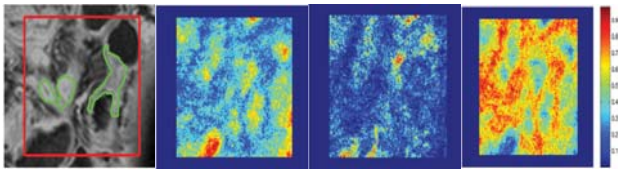
2.1. Identifying a volume of interest

First we identify those volume slices that contain part of the bowel using a learned RF classifier. In the 'bowel' slices small image patches are again classified for the presence of bowel tissues, using a second RF classifier. In [3] we showed the importance of higher order statistics of image features in discriminating between tissues of different classes. We use the same principle in identifying a VOI. Each volume slice is divided into non-overlapping 30×30 image patches. For each patch we calculate the mean, variance, skewness and kurtosis of intensity, texture and curvature values. Texture maps are obtained using oriented Gabor filters at angles of



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

Fig. 1. Localized ROIS and sample locations for context information. (a)-(c) show three slices with localized ROIs in red and manual annotations in green; (d) sample locations for deriving context information.



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

Fig. 2. Probability maps for ROI voxels. (a) Cropped showing ROI (red) and diseased annotations (green). Probability maps for (b) background; (c) normal; and (d) diseased (with colormap).

$\{0°, 45°, 90°, 135°\}$ and scale of 1 and 0.5. The feature vectors of each patch is concatenated into one big signature vector for the whole image, which is classified as either 'bowel' or 'non-bowel' slice.

For a 'bowel' slice each 30×30 patch is again classified as 'bowel' or 'non-bowel' patch indicating the presence or absence of bowel tissues. A collection of these patches makes up the region of interest (ROI) in a slice, and a collection of ROIs over adjacent slices makes up the VOI. For training on 'bowel' slices we make use of the slices which have annotations, while 'non-bowel' slices are visually identified. Similarly, bowel and non-bowel patches can be identified from the annotations. This preprocessing step is essential to reduce the computation time. Figures 1 (a)-(c) show three consecutive slices from a patient in which the ROI is shown in red, and the manually annotated diseased regions shown in green. This indicates that the VOI localizes the desired region with a good degree of accuracy.

2.2. Probability maps using RF classifiers

A probabilistic classifier like RF [7] is used to obtain class probability values for each voxel in the VOI. RFs are being increasingly used in medical applications [8, 9] because of their ability to handle large datasets, multi-class classification and interpret learned knowledge. In this section we describe different image features and our approach to extract semantic information from the learned RF classifier.

Image Features: For every voxel features are extracted over a 30×30 neighborhood. The mean and variance of intensity, texture and $3D$ mean-curvature values are used. Texture maps are obtained using oriented Gaussian filters at angles of $\{0°, 45°, 90°, 135°\}$ and two scales $(1, 0.5)$. These set of values give a 20 dimensional feature vector - 2 each from intensity and curvature, and 2 each from the 8 texture maps (2 scales and 4 orientations).

Relative Context: Since the relative arrangement of organs is constant (except for missing organs) one organ can provide contextual information about many others through relative distance and orientation. Thus context is particularly important for medical images and has been used for medical image segmentation [10, 11, 12]. Referring to Figs. 1 (a)-(c) we observe that apart from the bowel other organs like the liver are also visible. Similarly in other slices, kidneys, spine and spleen are also present to provide context information. We aim to exploit this relationship for higher segmentation accuracy.

Since contextual information depends on relative orientation and distance we sample regions at fixed positions from a pixel. Figure 1 (d) shows an illustration of the sampling scheme where the circle center is the pixel in question and the sampled points are identified by a red 'X'. At each point corresponding to a 'X' we extract a $3 \times 3 \times 3$ region and calculate the mean intensity, texture and curvature values. The texture values were derived from the texture maps obtained at $90°$ orientation and scale 1. The 'X's are located at distances of $3, 8, 15, 22$ pixels from the center, and the angle between consecutive rays is $45°$. The values from the 32 regions are concatenated into a 96 dimensional feature vector. The final feature vector has 116 values.

Equal number of samples from all classes in the training images are used to train a RF classifier. The trained classifier outputs probability maps for every voxel within the VOI of the test image. Each voxel has 3 values corresponding to the respective probabilities of belonging to 3 classes, namely, diseased, normal and background. Figure 2 (a) shows the cropped ROI corresponding to the image in Fig. 1 (b). Figures 2 (b)-(d) show the probability maps, respectively, for background, normal bowel and diseased bowel. The diseased area in Fig 2 (a) shows a corresponding high probability value in Fig 2 (d).

2.3. Graph Cut Segmentation

The segmentation is obtained by optimizing a second order MRF energy function which is written as

$$E(L) = \sum_{s \in P} D(L_s) + \lambda \sum_{s \in P} \sum_{t \in N_s} V(L_s, L_t), \qquad (1)$$

where P denotes the set of pixels; L_s is the label of pixel s and N_s is the neighborhood of s. The cost function is opti-

mized using graph cuts [13]. λ is a weight that determines the relative contribution of penalty cost (D) and smoothness cost (V). $D(L_s)$ is given by

$$D(L_s) = -\log\left(Pr(L_s) + \epsilon\right), \tag{2}$$

where Pr is the likelihood (or probabilities) previously obtained using RF classifiers and $\epsilon = 0.00001$ is a very small value to ensure that the cost is a real number.

2.3.1. Semantic Information for Smoothness Cost

V ensures a spatially smooth solution by penalizing discontinuities. We formulate the smoothness cost by using semantic information from the trained RF classifier. The RF classifier returns a measure of the importance of each dimension in the feature vector to the classification task. Inspite of the multiple dimensional feature vector, the features are of three types - intensity, texture and curvature (context information is a combination of the three). By aggregating the importance values of each feature type and normalizing them we obtain their relative importance in the classification task. This provides the necessary semantic information for voxel classification. Let the normalized weight of the different features be w_I (intensity), w_T (texture) and w_C (curvature), where $w_I + w_T + w_C = 1$. The smoothness cost V is given by

$$V(L_s, L_t) = \left\{ \begin{array}{ll} w_I V_I + w_T V_T + w_C V_C, & L_s \neq L_t, \\ 0 & L_s = L_t. \end{array} \right\} \tag{3}$$

where V_I, V_T, V_C are the individual contributions to the smoothness by intensity, texture and curvature. V_I is defined as

$$V_I(L_s, L_t) = e^{-\frac{(I_s - I_t)^2}{2\sigma^2}} \cdot \frac{1}{\|s - t\|}, \tag{4}$$

I is the intensity and σ is the intensity variance over N_s (i.e., the 8 neighbors). V_T and V_C are similarly defined using texture and curvature instead of intensity. After training $w_I = 0.23, w_T = 0.33, w_C = 0.44$. To choose the value of λ we adopt the following steps. We choose a small subset of the training data consisting of 10 patient volumes, and perform segmentation using our method but with λ varying from 0 to 1 in steps of 0.001. The maximum average segmentation accuracy using Dice Metric (DM) was obtained for $\lambda = 0.02$ and we set $\lambda = 0.02$.

3. EXPERIMENTAL RESULTS

3D T1-weighted spoiled gradient echo sequence (SPGE) images were acquired from 26 patients (including the 10 used to calculate λ) in supine position using a 3-T MR imaging unit (Intera, Philips Healthcare). The spatial resolution of the images was 1.02 mm $\times$ 1.02 mm$\times$ 2 mm, and the acquired volume dimension was $400 \times 400 \times 100$ voxels. We adopt a leave-one-out approach where our classifier is trained on 25 patients and the 26th patient is used for evaluating our segmentation algorithm. Thus we report performance on the average of 26 runs.

	Our	Our_{nC}	Our_{nV}	$Our_{nV_{TC}}$
DM	85.8	81.2	83.9	79.4
HD	3.2	7.2	4.3	10.2
		Our_{nV_I}	Our_{nV_T}	Our_{nV_C}
DM		84.6	83.4	83.1
HD		3.9	5.0	5.4

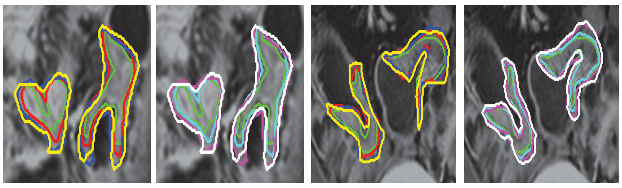
Table 1. Quantitative measures for segmentation accuracy. DM- Dice Metric in $\%$ and HD is Hausdorff distance mm

As part of evaluation of we present segmentation results of the following methods: 1) Our - our proposed method; 2) Our_{nC} - Our without context information from images to train the RF classifier; 3) Our_{nV} - Our without semantic context in V; $w_I = w_T = w_C = 0.33$; 4) Our_{nV_I} - Our with $w_I = 0$; 5) Our_{nV_T} - Our with $w_T = 0$; 6) Our_{nV_C} - Our with $w_C = 0$; 7) $Our_{nV_{TC}}$ - Our with $w_I = 1, w_T = 0, w_C = 0$ which is a conventional smoothness cost based on intensity features. For $4, 5, 6$ above the weights are normalized by the sum of the two values. We are unable to report comparisons with other methods because there dont exist methods specific to CD segmentation. Quantitative evaluation of segmentation performance is given in terms of Dice Metric (DM) and Hausdorff distance (HD) measures.

Table 1 summarizes the performance of all the above methods. Our performs the best among all methods followed by Our_{nV_I} and Our_{nV}. $Our_{nV_{TC}}$ gives the worst performance because it imposes smoothness constraints using only intensity information. Note that if we totally eliminate smoothness we get isolated clusters of different labels which is undesirable. From the results in Table 1 we draw the following conclusions. First, comparing the results of Our and Our_{nC} we observe that context features are integral to our method's and their exclusion leads to maximum degradation in performance.

Second, semantic information is also important. $p < 0.025$ from t-tests between Our and Our_{nV} supports this observation. Our_{nV} weights intensity, texture and curvature equally in the smoothness cost, thereby neglecting the semantic information learned from the trained classifier. This lowers the DM and increases the HD compared to Our, although the magnitude is less than Our_{nC}. Particularly for the bowel it is difficult to distinguish different tissues from only the intensity values in MR images. Learned semantic information helps to distinguish between neighboring voxels from different categories.

Comparing between Our_{nV_I}, Our_{nV_T} and Our_{nV_C} highlights the relative importance of different features. Curvature and texture features are deemed most important in the classification, although intensity also plays a significant role. $t-$tests between the results gives $p < 0.035$ for all cases in-



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

Fig. 3. Segmentation results for two patients: (a)-(b) Patient 16; (c)-(d) Patient 26. The manual annotations are shown in green with other colours showing the following annotations: red-Our,blue-Our_{nV},yellow-Our_{nC},magenta-Our_{nV_I},cyan-Our_{nV_T},white-Our_{nV_C}.

dicating that none of the features can be discarded without a drop in performance, and their combination gives the best performance.

Figure 3 shows segmentation results on Patient 16 and Patient 26 with the diseased regions cropped for clarity. We show results for Our(red),Our_{nV}(blue), Our_{nC}(yellow), Our_{nV_C}(magenta), Our_{nV_T}(cyan), Our_{nV_C}(white). These two cases were particularly challenging because of multiple diseased regions. Using our method we are able to achieve an accurate segmentation (DM$= 0.88$ and $HD = 3.5$ mm). Although only a single slice is shown the measures are for the whole volume. Our_{nV_C} shows the maximum over segmentation, i.e., many healthy tissues are labeled as diseased because of the absence of context information. The segmentation of the different methods is consistent with the values observed in Table 1. The average segmentation time for a volume was 155 seconds on a system running MATLAB on a Core2 quad core 2.66 GHz CPU having 4 GB RAM.

4. CONCLUSION

We have presented a novel way to segment regions affected by Crohn's disease in abdominal MR volumes. We develop a novel context feature that incorporates relative context information between different tissues. Intensity, texture, curvature and context are used to generate probability maps whose negative log-likelihood are the penalty costs in a graph cut segmentation framework. Semantic information about the contribution of each feature is obtained by training random forest classifiers annotated images, and this is used to design a novel smoothness cost. Experimental results on 26 patient datasets show that inclusion of semantic information and context features plays a very important role in accurate segmentation of the diseased region.

Acknowledgement. This research was partly funded from the European Community's Seventh Framework Programme (FP7/2007-2013): the VIGOR++ Project (grant agreement number 270379).

5. REFERENCES

[1] J. Rimola, S. Rodriguez, O. garcia Bosch, and et al., "Magnetic resonance for assessment of disease activity and severity in ileocolonic Crohn's disease.," *Gut*, vol. 58, pp. 1113–1120, 2009.

[2] F.M. Vos et al., "Computational modeling for assessment of ibd: to be or not to be?," in *Proc. IEEE EMBC*, 2012, pp. 3974–3977.

[3] D. Mahapatra, P. Schueffler, J. Tielbeek, J.M. Buhmann, and F.M. Vos., "A supervised learning based approach to detect crohn's disease in abdominal mr volumes," in *Proc. MICCAI-ABD*, 2012, pp. 97–106.

[4] J.E. Iglesias, E. Konukoglu, A. Montillo, Z. Tu, and A. Criminisi, "Combining generative and discriminative models for semantic segmentation of ct scans via active learning," in *IPMI*, 2011, pp. 25–36.

[5] A. Montillo, J. Shotton, J. Winn, J.E. Iglesias, D. Metaxas, and A. Criminisi, "Entangled decision forests and their application for semantic segmentation of ct images," in *IPMI*, 2011, pp. 184–196.

[6] M. Bhushan, J.A. Schnabel, L. Risser, M.P. Heinrich, J.M.Brady, and M. Jenkinson, "Motion correction and parameter estimation in dcemri sequences: Application to colorectal cancer.," in *MICCAI*, 2011, pp. 476–483.

[7] L. Breiman, "Random forests.," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[8] Z. Yi, A. Criminsi, J. Shotton, and A. Blake, "Discriminative semantic segmentation of brain tissue in MR images," in *MICCAI*, 2009, pp. 558–565.

[9] E. Geremia, B.H. Menze, O. Clatz, E. Konukgolu, A. Criminsi, and N. Ayache, "Spatial decision forests for MS lesion segmentation in multichannel MR images," in *MICCAI*, 2010, pp. 111–118.

[10] Z. Tu and X. Bai, "Auto-context and its application to high-level vision tasks and 3d brain image segmentation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1744 – 1757, 2010.

[11] D. Mahapatra and Y. Sun, "Integrating segmentation information for improved mrf-based elastic image registration.," *IEEE Trans. Imag. Proc.*, vol. 21, no. 1, pp. 170–183, 2012.

[12] D. Mahapatra, "Cardiac lv and rv segmentation using mutual context information.," in *Proc. MICCAI-MLMI*, 2012, pp. 201–209.

[13] Y. Boykov and O. Veksler, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, pp. 1222–1239, 2001.