SVM-MRF segmentation of colorectal NBI endoscopic images

Tsubasa Hirakawa, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda[1],
Tetsushi Koide[2], Yoko Kominami, Shigeto Yoshida, Shinji Tanaka[3]

***Abstract*— In this paper we investigate a method for segmentation of colorectal Narrow Band Imaging (NBI) endoscopic images with Support Vector Machine (SVM) and Markov Random Field (MRF). SVM classifiers recognize each square patch of an NBI image and output posterior probabilities that represent how likely the given patch falls into a certain label. To prevent the spatial inconsistency between adjacent patches and encourage segmented regions to have smoother shapes, MRF is introduced by using the posterior outputs of SVMs as a unary term of MRF energy function. Segmentation results of 1191 NBI images are evaluated in experiments in which SVMs were trained with 480 trimmed NBI images and the MRF energy was minimized by an $\alpha - \beta$ swap Graph Cut.**

I. Introduction

This paper proposes a novel method of segmentation problem of colorectal cancer and polyps in images taken by a Narrow Band Imaging (NBI) zoom-videoendoscope. Since colorectal cancer has been a leading cause of cancer death in many advanced countries [1]. Colorectal endoscopy (an examination of the colon with an endoscope) is widely used for early detection of colorectal cancer, where colorectal tumors are visually inspected with an NBI zoom-videoendoscope. Since the visual inspection depends on the subjective experience of medical doctors, it is important to develop a computer-aided diagnosis system to evaluate quantitatively the condition of a tumor. To achive this objective, we have developed a recognition system [2] for classifying NBI images of colorectal tumors into three types (A, B and C3) based on the NBI magnification findings [3] (see Figure 1). This system achieved a recognition rate of 96% on a dataset of 908 NBI images, each of which was collected during an NBI colonoscopy examination and trimmed to a rectangle.

In this paper, we extend our previous work [2] for NBI image classification to segmentation: trying to find which part of the NBI image falls into one of three types of colorectal tumors. The previously developed system was based on a bag-of-visual word (BoVW) framework with densely sampled SIFT descriptors (see Figure 2). Each training image was transformed into a histogram of visual words, then classified by a Support Vector Machine (SVM) classifier with a linear kernel. To build a segmentation method on top of our previous classification system, we combine SVM classifiers with a Markov Random Field (MRF) minimization framework; first we divide an NBI image into a number of small square patches, then classify each patch by SVM classifiers that are separately trained, and finally make the classification results for the patches spatially consistent by minimizing an MRF energy function. In section IV, segmentation results obtained by the developed system are evaluated on a dataset of colorectal NBI endoscopic images.

[1] Graduate School of Engineering, Hiroshima University, Japan
[2] Research Institute for Nanodevice and Bio Systems (RNBS), Hiroshima University, Japan
[3] Graduate School of Biomedical Science, Hiroshima University, Japan

Type A		Microvessels are not observed or extremely opaque.
Type B		Fine microvessels are observed around pits, and clear pits can be observed via the nest of microvessels.
Type C	1	Microvessels comprise an irregular network, pits observed via the microvessels are slightly non-distinct, and vessel diameter or distribution is homogeneous.
	2	Microvessels comprise an irregular network, pits observed via the microvessels are irregular, and vessel diameter or distribution is heterogeneous.
	3	Pits via the microvessels are invisible, irregular vessel diameter is thick, or the vessel distribution is heterogeneous, and a vascular areas are observed.

Fig. 1. NBI magnification findings [3]. C1 and C2 are not used in this work.

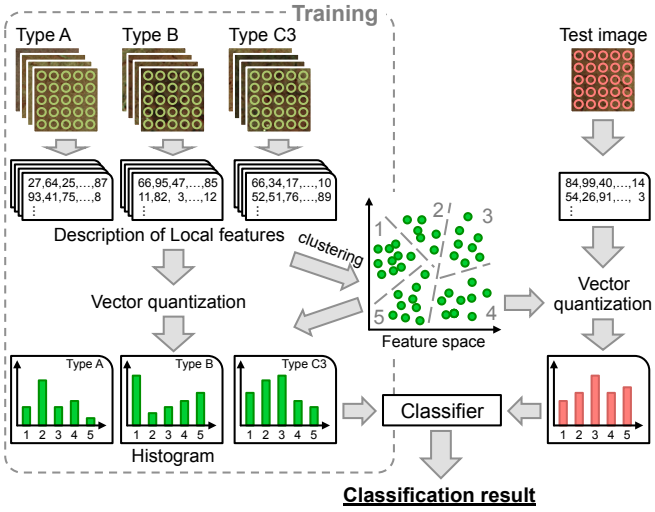


Fig. 2. Outline of the BoVW approach.

II. Related work

The combined use of SVM and MRF has been already explored in the literature and some relevant work is reviewed here.

Wu et al. [4] proposed a prior feature SVM-MRF, a

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

segmentation method for 3D Magnetic Resonance images. They trained SVMs using pixel locations in addition to the voxel intensity and the given label of the point, then used the posteriors (outputs) of the SVMs for the pixels in the unary term of the MRF energy, which was minimized by the Iterated Conditional Modes (ICM). Wang and Manjunath [5] proposed an SVM-MRF framework for image retrieval based on semantic segmentation of images. For each block, SVMs output conditional probability which is used as the unary term of the MRF energy. They used a causal greedy algorithm to minimize the energy. Moser and Serpico [6], [7] used binary SVMs with a particular kernel to define a local decision rule of MRF with updates by ICM. Hoefel and Elkan [8] proposed a two-stage SVM/CRF for sequence (hence chain) classification. Discretized scores of SVMs, which are trained separately, are used as a term in the CRF energy function, and the weights of the terms are learnt.

Most relevant to ours is the work of Fulkerson et al. [9]. They trained SVMs by bag-of-feature histograms, and then scores by SVMs for each pixel are used as confidence values for localizing objects. Later, they extended the approach by incorporating spatial context with Conditional Random Field (CRF) [10]. SVMs are trained with bag-of-feature histograms of super pixels, and then probability outputs by SVMs for each superpixel are used as a unary term of CRF energy. To handle boundaries of objects, neighboring histograms of a superpixel are merged, then more accurate object boundaries are inferred by the CRF energy minimization with α-expansion. The pairwise term uses color difference and edge length between superpixels. The parameter learnt as a CRF model is a single weight between unary and pairwise terms.

Our objective is to segment an NBI image into three types (A, B and C3) of the NBI magnification findings (Fig. 1). This segmentation problem is very different from the targets of the existing works mentioned above, or other works on colon polyp segmentation [11], [12], [13], [14]. First, we can not use the location and rotation of colorectal tumors in NBI images because those are taken by an endoscopy that arbitrarily moves around in the colon. Therefore, a location prior [4] in an image is not useful. Second, there is no clear borderline between the types; instead, visual appearance of texture of the mucosal surface changes gradually from one type to another. Edge information is usually useful for segmentation of polyps [11], [12], [13], [14], however we do not expect so in our task. In the next section, we introduce an SVM-MRF combination similar to [10], but without edge information and superpixels.

III. SVM-MRF SEGMENTATION

Here we describe a segmentation method using SVM and MRF. First, we divide an NBI image into square patches P_i $(i = 1, \ldots, n)$, and each patch corresponds to a site (or node) of a grid of MRF. Adjacent patches may overlap depending on the size of patches and the spacing of the MRF grid.

Each site i has a label x_i taking values of A, B, or C3 corresponding to the three types of the NBI magnification findings. We denote all the labels collectively as $\boldsymbol{x} = (x_1, \ldots, x_n)$. The bag-of-visual word histogram of the patch is denoted as y_i, and collectively $\boldsymbol{y} = (y_1, \ldots, y_n)$.

We define the following MRF energy function in terms of the posterior probability:

$$f(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left(\sum_i A(x_i, y_i) + \sum_{j \in N_i} I(x_i, x_j)\right). \quad (1)$$

Here $A(x_i, y_i)$ is a unary term that represents the inconsistency of the patch label x_i to data y_i. In this study, we use the posterior probability outputs $\log P(x_i|y_i)$ of SVM classifiers that are learnt with a separate training set of NBI images. Then we use the posterior as the unary term as follows:

$$A(x_i, y_i) = -\log P(x_i|y_i). \quad (2)$$

The second term $I(x_i, x_j)$ in the MRF energy is called an interaction term which describes the spatial inconsistency between x_i and its neighbors x_j, where N_i is a neighbor of site i. Here we define the interaction term as follows:

$$I(x_i, x_j) = \begin{cases} -\log p, & x_i = x_j \\ -\log \frac{1-p}{2}, & \text{otherwise} \end{cases}, \quad (3)$$

where $p \in (0, 1)$ is a probability that site i and its neighbor j take the same label. Because the segmentation problem here has three labels (A, B, and C3), there are two cases where site i and its neighbor j take different labels. Therefore the probability $1 - p$ is halved.

The MRF energy function (1) is minimized by α–β swap Graph Cuts [15] in order to obtain an MAP estimate of labels $\boldsymbol{x}$.

IV. EXPERIMENTAL RESULTS

We have 1671 NBI images (Type A: 504, Type B: 847, and Type C3: 320), which were collected during NBI colonoscopy examination. Each image was trimmed from an original NBI image to a rectangle as a training image representing typical microvessel structure appearance, and was labeled by medical doctors and endoscopists. We selected 160 images from each type for a training set; totally 480 training images are used for training an SVM classifier. For segmentation, the remaining 1191 NBI images (Type A: 344, Type B: 687, Type C3: 160) were used. In each NBI image, an MRF grid of spacing 10 pixels is constructed. A site of MRF corresponds to a square patch of the size of 120 pixels, in which SIFT descriptors are extracted at each 5 pixels with fixed scales of 5 and 7 pixels [2].

Segmentation methods are usually evaluated by ground truth labels of an entire image. This means that doctors need to paint a lot of NBI images but this is impractical. Instead, we use the above-mentioned trimmed regions from the original NBI images for training the SVM classifiers, because the labeled rectangle in the original NBI image is a good indicator how good the segmentation is. As shown in Figure 3, we evaluate a segmentation result by the ratio of areas inside a labeled rectangle whose estimated labels are correct. Using the estimated labels in the labeled region,

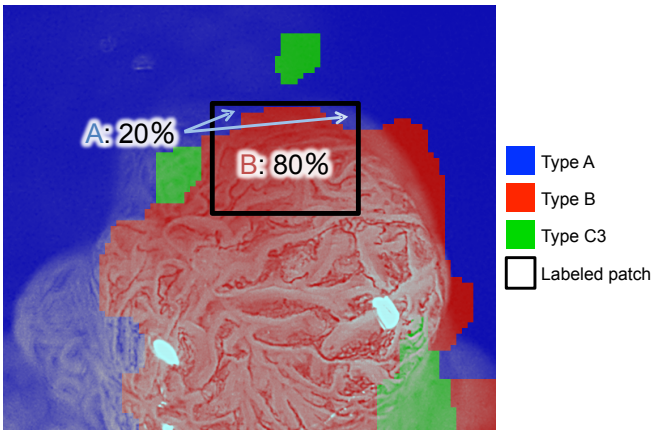


Fig. 3. Evaluation procedure (best viewed in color). Performance is evaluated by the ratio of areas inside a rectangle whose estimated labels are correct. In this example, the ground truth of the rectangle is Type B and 80% of the area inside the rectangle is correctly labeled.

correct rate, precision rate and recall rate are calculated from a confusion matrix [16].

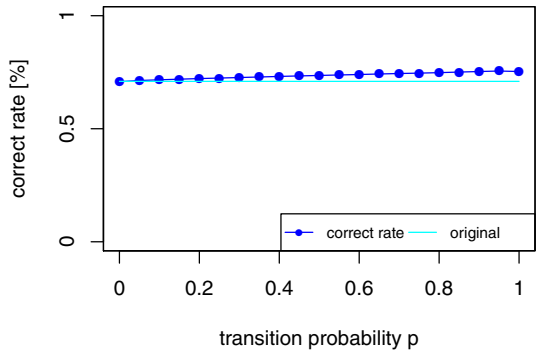
Figure 4 shows the pchange in performance of the segmentation results as the probability p takes values of $0, 0.05, \ldots, 0.95, 0.99$. We can see that the correct rate improves as p becomes large; the smoother the labels of the adjacent patches, the better the result.

Figures 5 to 7 show examples of segmentation results for each NBI type. In all cases, resulting regions become smoother and large as p becomes large, and particularly a good segmentation result is obtained in Figure 6.

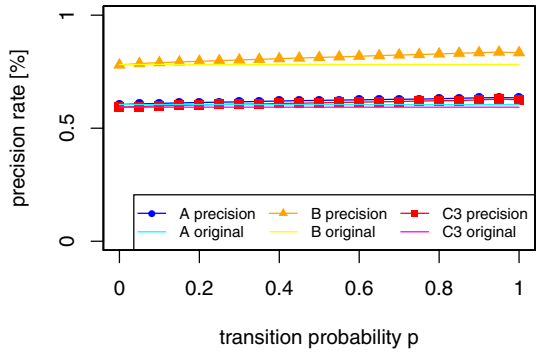
Note that the part around highlights due to the reflection of light in Fig. 5 is classified as Type B while the true label is Type A. This may be caused by the strong edge of the highlight; many edges can be seen in Type B images while textures in Type A images are rather smooth. These effect will be investigated in future for improving the segmentation results.

V. Conclusions

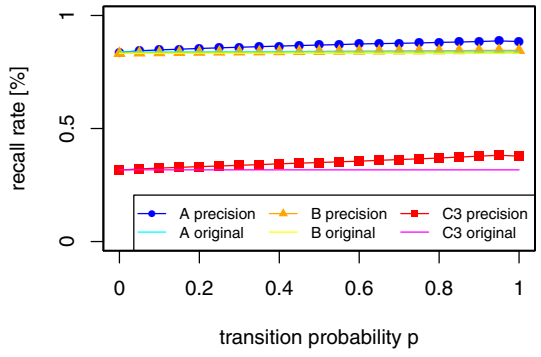
In this paper, we presented a segmentation method based on a SVM-MRF combination. Currently the parameter p, the probability that two adjacent patches takes the same label, is left for operators to tune how much the segmentation result is sensitive to noise. Future work includes automatic adjustment of the parameters, and segmentation of NBI endoscopic video sequences.



(a) correct rate



(b) precision rate



(c) recall rate

Fig. 4. Performance of segmentation for different values of p (best viewed in color) in terms of correct, precision and recall rates for each type. "Original" means results obtained when MRF is not used and each patch is independently classified by an SVM.

References

[1] Health statistics and informatics Department, World Health Organization, "Global Burden of Disease: 2004 update." http://www.who.int/evidence/bod, 2008.

[2] T. Tamaki, J. Yoshimuta, M. Kawakami, B. Raytchev, K. Kaneda, S. Yoshida, Y. Takemura, K. Onji, R. Miyaki, and S. Tanaka, "Computer-aided colorectal tumor classification in nbi endoscopy using local features," *Medical Image Analysis*, vol. 17, no. 1, pp. 78–100, 2013.

[3] H. Kanao, S. Tanaka, S. Oka, M. Hirata, S. Yoshida, and K. Chayama, "Narrow-band imaging magnification predicts the histology and invasion depth of colorectal tumors," *Gastrointestinal Endoscopy*, vol. 69, no. 3, Part 2, pp. 631–636, 2009.

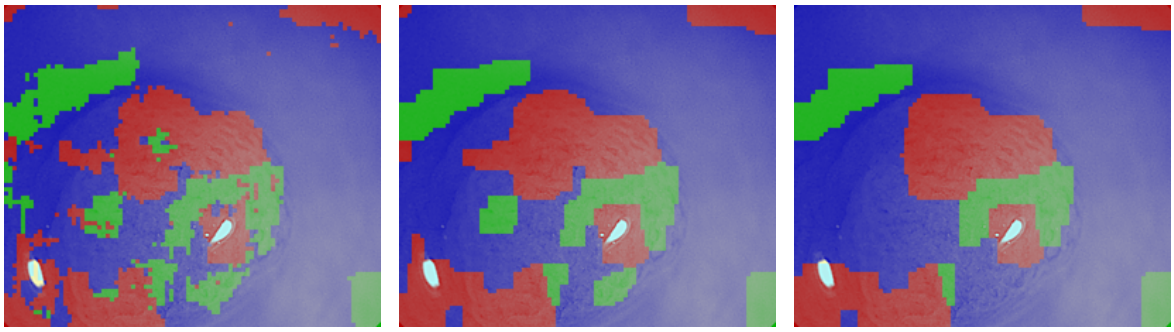
[4] T. Wu, M. H. Bae, M. Zhang, R. Pan, and A. Badea, "A prior feature svm-mrf based method for mouse brain segmentation," *NeuroImage*, vol. 59, no. 3, pp. 2298 – 2306, 2012.

[5] L. Wang and B. Manjunath, "A semantic representation for image retrieval," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, vol. 2, pp. II–523–6 vol.3, Sept 2003.

[6] G. Moser and S. Serpico, "Contextual remote-sensing image classification by support vector machines and markov random fields," in *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, pp. 3728–3731, July 2010.

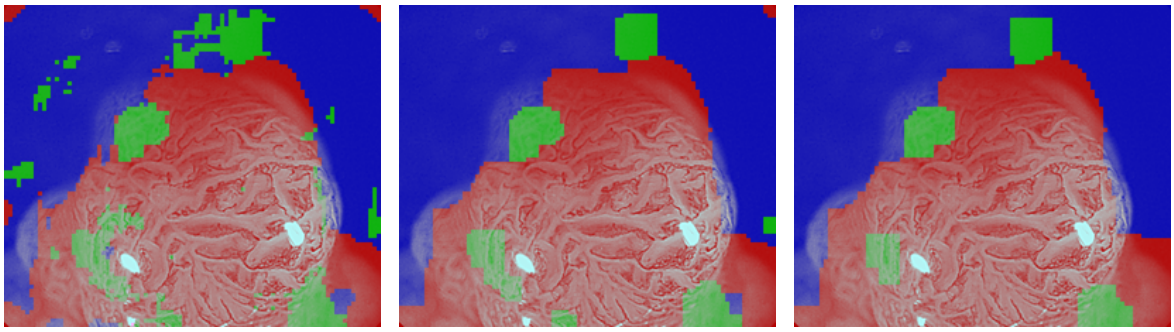
[7] G. Moser and S. Serpico, "Combining support vector machines and markov random fields in an integrated framework for contextual image classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 51, pp. 2734–2752, May 2013.

[8] G. Hoefel and C. Elkan, "Learning a two-stage svm/crf sequence classifier," in *Proceedings of the 17th ACM Conference on Information*



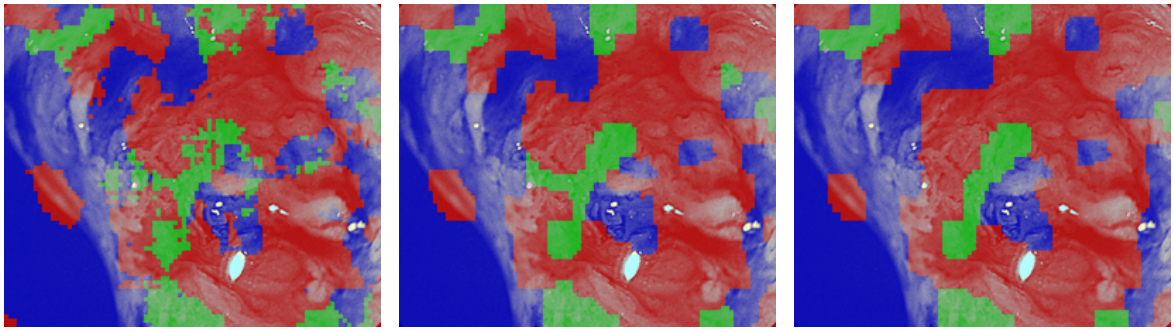
(a) $p = 0.00$ (b) $p = 0.50$ (c) $p = 0.99$

Fig. 5. Segmentation for NBI images of Type A for different values of p (best viewed in color). Blue color represents Type A, red Type B, and green Type C3.



(a) $p = 0.00$ (b) $p = 0.50$ (c) $p = 0.99$

Fig. 6. Segmentation for NBI images of Type B for different values of p (best viewed in color).



(a) $p = 0.00$ (b) $p = 0.50$ (c) $p = 0.99$

Fig. 7. Segmentation for NBI images of Type C3 for different values of p (best viewed in color).

and Knowledge Management, CIKM '08, (New York, NY, USA), pp. 271–278, ACM, 2008.

[9] B. Fulkerson, A. Vedaldi, and S. Soatto, "Localizing objects with smart dictionaries," in *Computer Vision – ECCV 2008* (D. Forsyth, P. Torr, and A. Zisserman, eds.), vol. 5302 of *Lecture Notes in Computer Science*, pp. 179–192, Springer Berlin Heidelberg, 2008.

[10] B. Fulkerson, A. Vedaldi, and S. Soatto, "Class segmentation and object localization with superpixel neighborhoods," in *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 670–677, Sept 2009.

[11] M. Breier, S. Gross, and A. Behrens, "Chan-vese-segmentation of polyps in colonoscopic image data," in *Proceedings of the 15th International Student Conference on Electrical Engineering POSTER 2011*, (Prague, Czech Republic), May 12 2011.

[12] M. Breier, S. Gross, A. Behrens, T. Stehle, and T. Aach, "Active contours for localizing polyps in colonoscopic NBI image data," in *Proc. of Medical Imaging 2011: Computer-Aided Diagnosis*, pp. 79632M–79632M–10, 2011.

[13] J. Bernal, F. J. Sánchez, and F. Vilariño, "A region segmentation method for colonoscopy images using a model of polyp appearance," in *IbPRIA* (J. Vitrià, J. M. R. Sanches, and M. Hernández, eds.), vol. 6669 of *Lecture Notes in Computer Science*, pp. 134–142, Springer, 2011.

[14] S. Gross, M. Kennel, T. Stehle, J. Wulff, J. Tischendorf, C. Trautwein, and T. Aach, "Polyp segmentation in nbi colonoscopy," in *Bildverarbeitung für die Medizin 2009* (H.-P. Meinzer, T. Deserno, H. Handels, and T. Tolxdorff, eds.), Informatik aktuell, pp. 252–256, Springer Berlin Heidelberg, 2009.

[15] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222–1239, 2001.

[16] F. Provost and R. Kohavi, "On applied research in machine learning," in *Machine learning*, pp. 127–132, 1998.