Labeling colorectal NBI zoom-videoendoscope image sequences with MRF and SVM

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

***Abstract*— In this paper, we propose a sequence labeling method by using SVM posterior probabilities with a Markov Random Field (MRF) model for colorectal Narrow Band Imaging (NBI) zoom-videoendoscope. Classifying each frame of a video sequence by SVM classifiers independently leads to an output sequence which is unstable and hard to understand by endoscopists. To make it more stable and readable, we use an MRF model to label the sequence of posterior probabilities. In addition, we introduce class asymmetry for the NBI images in order to keep and enhance frames where there is a possibility that cancers might have been detected. Experimental results with NBI video sequences demonstrate that the proposed MRF model with class asymmetry performs much better than a model without asymmetry.**

I. Introduction

Colorectal cancer has been one of the major causes of cancer death all around the world [1], [2]. An early detection of colorectal cancer through colorectal endoscopy, or *colonoscopy*, is important and widely used in hospitals as a standard medical procedure. During colonoscopy, the lesions of colorectal tumors on the colon surface (Figure 1) are visually inspected by a Narrow Band Imaging (NBI) zoom-videoendoscope. By using the visual appearance of colorectal tumors in endoscopic images, histological diagnosis is presumed based on classification schemes for NBI magnification findings. A diagnosis by visual inspection, however, is affected by the skill and familiarity of each endoscopists. Hence, a computer-aided system for supporting the visual inspection would be of great help for colonoscopy, due to the large number of images of colorectal tumors which should be classified in a periodical medical examination for detecting cancer in its early stage.

Tamaki et al. [3], [4] proposed an NBI endoscopic image recognition system with Bag–of–Visual Words and Support Vector Machine (SVM) for a 3-class classification problem based on the *NBI magnification findings* [5], [6]. It divides the microvessel structures of mucosal surface in an NBI image into types A, B, and C (see Figure 1). In type A, microvessels are not observed, or slightly observed but opaque with very low contrast (typical images are shown in the top row of Figure 2). In type B, fine microvessels are visible around clearly observed pits (the middle row of Figure 2). Type C is divided into three subtypes C1, C2, and C3 according to detailed texture. In type C3, which exhibits the most irregular texture, pits are almost invisible because of the irregularity of tumors, and microvessels are irregular and thick, or heterogeneously distorted (the bottom row of Figure 2). Based on this NBI magnification findings, Tamaki et al. [3], [4] used a dataset of 908 NBI images (Type A: 359, Type B: 462, and Type C3: 87) to learn and classify rectangular patches trimmed from NBI endoscopic images, and achieved 96% recognition rate with 10–fold cross validation.

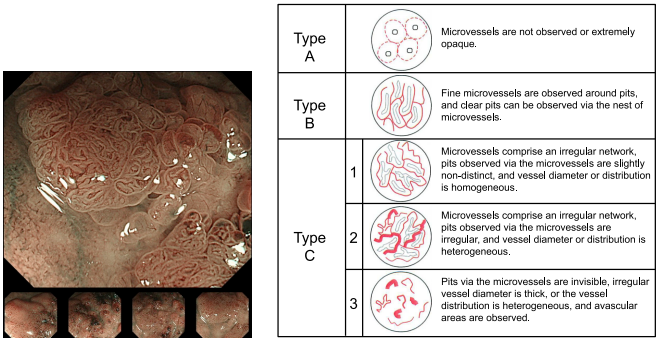


Fig. 1. (left) A screen shot of an actual NBI zoom-videoendoscopiy. (right) NBI magnification findings [5].

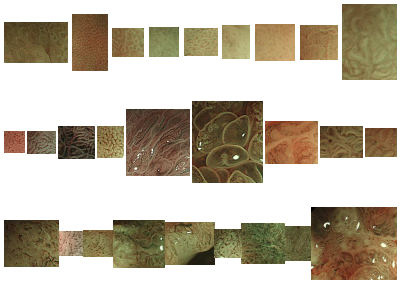


Fig. 2. Examples of NBI images of types A (top row), B (middle row), and C3 (bottom row).

This enables us to develop a real time recognition/assessment system during endoscopic examination of the colon, which is expected to be of a great help for colonoscopy [7]. A simple way is feeding an NBI videoendoscopic image sequence frame by frame to the recognition system: A rectangular window at the center of each frame is classified, and the output of each frame can be either of a label or posterior probabilities of three classes in order to show how the estimated label is confident.

Applying this simple strategy is, however, usually not appropriate for observations by endoscopists. Figure 3 shows an example of recognition for an NBI image sequence.

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

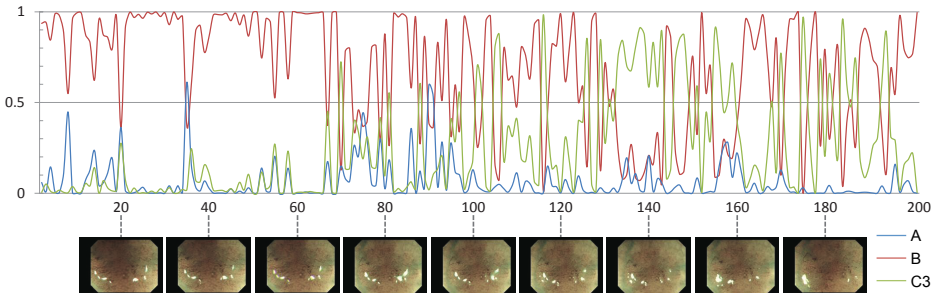


Fig. 3. An example output for an NBI image sequence. Posterior probabilities of three classes are shown in different colors.

Because each frame is processed independently, the posterior probabilities from an SVM classifier, the output of the system, are highly unstable. The corresponding labels for each frame are also unstable because those are labels with the highest probability at each frame. This result should be temporally smoothed and stabilized in order to support the diagnosis by endoscopists during the endoscopy examination.

A straightforward way to handle this problem might be just to apply smoothing methods, such as moving average, spline interpolation, or a Kalman filter. But temporal smoothing of posterior probability "curves" is inadequate because there is no probabilistic interpretation, and usually the smoothed result needs to be re-normalized at each frame to sum up to 1, and often values outside of the range $[0, 1]$ are needed to be cropped.

Instead of smoothing, we try to label a sequence of posterior probabilities in a way that avoids frequent label changes. In this paper, we propose a sequence labeling method by using the posterior probabilities obtained from an SVM as a data penalty term in a Markov Random Field (MRF) framework.

II. Related work

Several methods have been proposed which use SVM posterior probabilities for MRFs. Bae et al. [8], [9] proposed an extended MRF with SVM for segmentation of mouse brain images. They used voxel values and locations of Magnetic Resonance Image (MRI) volumes as feature vectors, and trained SVM and MRF simultaneously in order to learn voxel location as a prior information. This approach is not applicable to our task because NBI image sequences do not have any structure as time series. One of the key concepts in our approach is to use an SVM classifier, trained separately from an MRF. This allows us the flexibility to replace, additionally train, and evaluate the classifiers separately from the MRF inference.

Hosaka et al. [10] proposed a method for matting with SVM training of colors of foreground and background regions specified as few strokes. This method separates the SVM training and MRF inference, however, the training step is oversimplified because it needs to be fast for its interactivity but not so accurate. In contrast, our approach can guarantee high performance of classification because of completely separated training classifiers and adjusting parameters. Note that classifiers other than SVM can be combined with MRF. For example, Lee et al. [11] used outputs of AdaBoost with MRF for image segmentation.

Another key concept in our approach is to introduce class asymmetry to the smoothness term of the MRF, which is specific to the NBI images. The NBI magnification findings we use in this paper have three classes: Type A (corresponding to hyperplasias, not cancers), B (adenomas, not cancers) and C3 (carcinomas or cancers). Type C3 must not to be smoothed out, therefore, we propose a smoothness term in MRF which is able to enhance type C3 detection.

III. MRF with SVM posterior probabilities

Let $\boldsymbol{x} = \{x_1, \ldots, x_i, \ldots, x_n\}$ be a sequence of estimated labels, where $x_i \in \{\mathrm{A}, \mathrm{B}, \mathrm{C3}\}$ is a label of an image at time i, and n the length of the sequence. Denote a video sequence as $\boldsymbol{y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i, \ldots, \boldsymbol{y}_n\}$ where $\boldsymbol{y}_i$ is a feature vector of an image at time i.

The posterior probability of $\boldsymbol{x}$ is represented by an MRF model as:

$$f(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left(\sum_i A(x_i, \boldsymbol{y}_i) + \sum_{(i,j)\in\mathcal{N}} I(x_i, x_j)\right), \quad (1)$$

where $A(x_i, \boldsymbol{y}_j)$ is a data term given by $A(x_i, \boldsymbol{y}_i) = \log P(x_i|\boldsymbol{y}_i)$, where $P(x_i|\boldsymbol{y}_i)$ is the posterior probability obtained from an SVM at time i.

$I(x_i, x_j)$ is a smoothness term (or a pairwise interaction term) and $\mathcal{N}$ is a neighborhood. If the asymmetry of the classes is not considered, the smoothness term can be written in a simple form by using a single transition probability $p \in [0, 1]$ as follows:

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

Note that p controls the degree of smoothness. Usually the value of p is set to be close to 1 so that the adjacent frames tend to the same label.

IV. Smoothness term with class asymmetry

The MRF model described above allows labeling without frequent label changes, however, it may smooth out (e.g., eliminate) frames detected as type C3. To keep and enhance those frames, we introduce a smoothness term that takes

TABLE I

TRANSITION PROBABILITIES WITH CLASS ASYMMETRY.

$x_i = \mathrm{A}$		x_j		
		A	B	C3
x_h	A	p_1	p_2	$2(1-p_4)/3$
	B	p_2	$(1-p_1)/2$	$(1-p_4)/3$
	C3	$2(1-p_4)/3$	$(1-p_4)/3$	$(1-p_3)/2$

$x_i = \mathrm{B}$		x_j		
		A	B	C3
x_h	A	$(1-p_1)/2$	p_2	$(1-p_4)/3$
	B	p_2	p_1	$2(1-p_4)/3$
	C3	$(1-p_4)/3$	$2(1-p_4)/3$	$(1-p_3)/2$

$x_i = \mathrm{C3}$		x_j		
		A	B	C3
x_h	A	$(1-p_1)/2$	$1-2p_1$	p_4
	B	$1-2p_1$	$(1-p_1)/2$	p_4
	C3	p_4	p_4	p_3

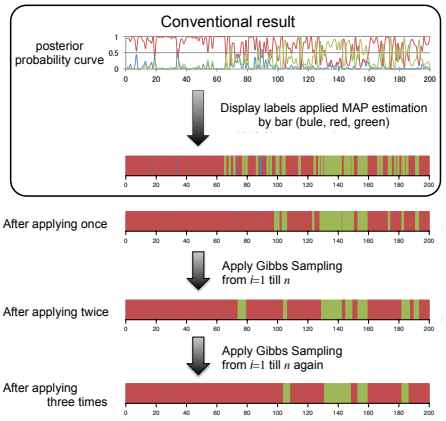


Fig. 4. Gibbs sampling. Labels are sampled sequentially until convergence (burn-in).

the class asymmetry into account. Instead of a *pairwise* smoothness potential which involves x_i and x_j, we use a *triplet* smoothness potential with x_h, x_i, and x_j. In addition, the transition probability to type C3 is set to be large relative to other classes. In this case, the MRF model is given as follows:

$$f(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left(\sum_i A(x_i, y_i) + \sum_{(h,i,j)\in\mathcal{N}} I(x_h, x_i, x_j)\right). \quad (3)$$

Now the smoothness term is much more complicated and involves four parameters. The complete definition is shown in Table I. Here, there are four parameters that define the transition probability.

- p_1 defines the probability in the case where the three labels are the same and either type A or B (i.e., $x_h = x_i = x_j = \mathrm{A}$ or B),
- p_2 when two labels are the same and either type A or B ($x_i = x_j = \mathrm{A}$ or B, or $x_i = x_h = \mathrm{A}$ or B),
- p_3 when the three labels are type C3 ($x_h = x_i = x_j = \mathrm{C3}$),
- p_4 when two labels are the same and type C3 ($x_i = x_j = \mathrm{C3}$, or $x_i = x_h = \mathrm{C3}$).

V. SEQUENCE LABELING BY GIBBS SAMPLING

Dynamic Programming (DP) can be used to obtain the MAP estimate of the MRF model (1), however, it can not be used for the MRF model with asymmetry (3). Therefore, we employ Gibbs sampling to sample $\boldsymbol{x}$ sequentially by using the following distribution:

$$f(x_i|x_{-i}, \boldsymbol{y}) \propto \exp\big(A(x_i, y_i) + I(x_{i-2}, x_{i-1}, x_i) + I(x_{i-1}, x_i, x_{i+1}) + I(x_i, x_{i+1}, x_{i+2})\big), \quad (4)$$

where $x_{-i} = \{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\}$.

Figure 4 shows an overview of Gibbs sampling. In the figure, the labels at each frame are shown in color (blue: type A, red: type B, green: type C3). The original posterior probability is used as an initial value for each frame by taking the label that gives the maximum probability (i.e., the MAP estimate for each frame).

VI. EXPERIMENTAL RESULTS

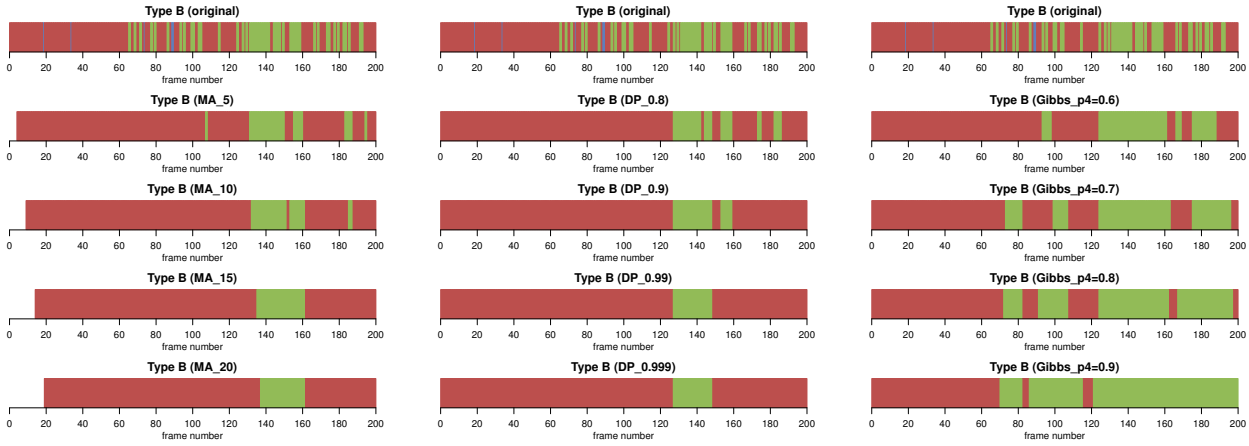
To evaluate the proposed method, we used an NBI video sequence of length 200 frames, where large enough polyps were captured in each image. For the MRF (1) without class asymmetry, we changed the transition probability as $p = 0.8, 0.9, 0.99, 0.999$. For the MRF (3) with class asymmetry, we fixed probabilities $p_1 = 0.99, p_2 = 0.45, p_3 = 0.9999$ while we changed p_4 as $0.6, 0.7, 0.8, 0.9$. For comparison, we also apply a simple moving average (MA) with different lengths of smoothing.

Figure 5 shows the results of MA and MRF with and without class asymmetry. Without class asymmetry, the MRF performs quite similar to MA. This result is not good for observing frames detected as type C3 (green) because most of the images between frames $70 \sim 110$ and $170 \sim 190$ are detected as type C3 in the original sequence but those are smoothed out. However, with class asymmetry, the MRF keeps those type C3 frames while frequent label changes that occur in the original sequence are reasonably well suppressed. Of course the result is sensitive to the transition probabilities $p_1, \ldots, p_4$, however, enhancing type C3 frames is always achieved.

Figure 6 shows labeling results for other NBI video sequences of type A and B polyps. In Figure 6(a), some frames were detected as type C3 (around frames $15 \sim 40, 70, 125$), and the MRF without class asymmetry smoothed those frames out (the middle row of Figure 6). The proposed MRF with class asymmetry succeeds to correctly retain and enhance those frames (the bottom row of Figure 6).

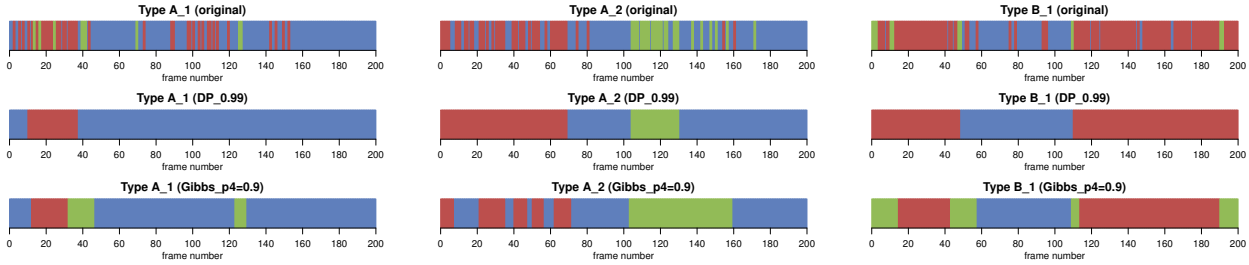
VII. CONCLUSIONS

In this paper, we proposed an MRF-based sequence labeling method by using SVM posterior probability. To suppress frequent changes of labels in a sequence, we introduced the class asymmetry of NBI images and employed Gibbs sampling. Our future work includes deciding transition probabilities, optimal selection of the smoothness term [12], and extension to spatial labeling for spatio-temporal segmentation of NBI video sequences.



(a) MA (b) without asymmetry (c) with asymmetry

Fig. 5. Comparison of labeling results for an NBI video sequence of Type B polyp. Type A labels are shown in blue, B in red, and C3 in green. Original: posterior probability curves. MA_N: moving average of length N. DP_x: labeling by DP without class asymmetry for $p = x$. Gibbs_p4=x: labeling by Gibbs sampling with class asymmetry for $p_4 = x$.



(a) Type A polyp 1 (b) Type A polyp 2 (c) Type B polyp

Fig. 6. Labeling results for three NBI video sequences by (middle) DP without class asymmetry and (bottom) Gibbs sampling with class asymmetry.

REFERENCES

[1] Health statistics and informatics Department, World Health Organization, Global Burden of Disease: 2004 update, 2008.

[2] Ministry of Health, Labor and Welfare, Japan, Vital Statistics in Japan — The latest trends —, 2009.

[3] Tamaki, T., Yoshimuta, J., Takeda, T., Raytchev, B., Kaneda, K., Yoshida, S., Takemura, Y., Tanaka, S., A system for colorectal tumor classification in magnifying endoscopic NBI images, ACCV 2010, pp. 452-463. 2010.

[4] Toru Tamaki, Junki Yoshimuta, Misato Kawakami, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Yoshito Takemura, Keiichi Onji, Rie Miyaki, Shinji Tanaka, Computer-aided colorectal tumor classification in NBI endoscopy using local features, Medical Image Analysis, Vol. 17, No. 1, pp. 78–100, 2013.

[5] Kanao, H., Tanaka, S., Oka, S., Hirata, M., Yoshida, S., Chayama, K., Narrow-band imaging magnification predicts the histology and invasion depth of colorectal tumors. Gastrointestinal Endoscopy 69, 631–636, 2009.

[6] Oba, S., Tanaka, S., Oka, S., Kanao, H., Yoshida, S., Shimamoto, F., Chayama, K., Characterization of colorectal tumors using narrow-band imaging magnification: combined diagnosis with both pit pattern and microvessel features. Scandinavian Journal of Gastroenterology 45, 1084–1092, 2010.

[7] Rex, D.K., Kahi, C., O'Brien, M., Levin, T., Pohl, H., Rastogi, A., Burgart, L., Imperiale, T., Ladabaum, U., Cohen, J., The american society for gastrointestinal endoscopy PIVI (Preservation and incorporation of valuable endoscopic innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps. Gastrointestinal Endoscopy 73, 419–422, 2011.

[8] Min Hyeok Bae, Rong Pan, Teresa Wu, Alexandra Badea, Automated segmentation of mouse brain images using extended MRF, NeuroImage, Vol. 46, No. 3, pp. 717–725, 2009.

[9] Teresa Wu, Min Hyeok Bae, Min Zhang, Rong Pan, Alexandra Badea, A prior feature SVM-MRF based method for mouse brain segmentation, NeuroImage, Vol. 59, No. 3, pp. 2298–2306, 2012.

[10] Tadaaki Hosaka, Takumi Kobayashi, Nobuyuki Otsu, Image matting based on local color discrimination by SVM, Pattern Recognition Letters, Vol. 30, No. 14, pp. 1253–1263, 2009.

[11] Sang Hak Lee, Hyung Il Koo, Nam Ik Cho, Image segmentation algorithms based on the machine learning of features, Pattern Recognition Letters, Vol. 31, No. 14, pp. 2325–2336, 2010.

[12] Yuko Maruyama, Kenji Yamanishi, Dynamic Model Selection with Its Applications to Computer Security, Proc. of 2004 IEEE International Workshop on Information Theory, 2004.