Discriminative Subtree Selection for NBI Endoscopic Image Labeling

Tsubasa Hirakawa[1], Toru Tamaki[1], Takio Kurita[1], Bisser Raytchev[1], Kazufumi Kaneda[1], Chaohui Wang[2], Laurent Najman[2], Tetsushi Koide[3], Shigeto Yoshida[4], Hiroshi Mieno[4], Shinji Tanaka[5]

[1] Hiroshima University, Japan
[2] Laboratoire d'Informatique Gaspard-Monge, Université Paris-Est, France
[3] Research Institute for Nanodevice and Bio Systems (RNBS), Hiroshima University, Japan
[4] Hiroshima General Hospital of West Japan Railway Company, Japan
[5] Hiroshima University Hospital, Japan

Abstract. In this paper, we propose a novel method for image labeling of colorectal Narrow Band Imaging (NBI) endoscopic images based on a tree of shapes. Labeling results could be obtained by simply classifying histogram features of all nodes in a tree of shapes, however, satisfactory results are difficult to obtain because histogram features of small nodes are not enough discriminative. To obtain discriminative subtrees, we propose a method that optimally selects discriminative subtrees. We model an objective function that includes the parameters of a classifier and a threshold to select subtrees. Then labeling is done by mapping the classification results of nodes of the subtrees to those corresponding image regions. Experimental results on a dataset of 63 NBI endoscopic images show that the proposed method performs qualitatively and quantitatively much better than existing methods.

1 Introduction

Colorectal cancer has been one of the major cause of cancer death in many advanced countries [1]. For early detection of colorectal cancer, colorectal endoscopy (colonoscopy) with Narrow-Band Imaging (NBI) system is widely used, where endoscopists observe the condition of tumor displayed on a screen. However, because of intra/inter-observer variability [2–4], the visual inspection of a tumor depends on the subjectivity and experience of endoscopists. Therefore, developing a computer-aided diagnosis system that provides objective measure of tumor to endoscopists would be greatly helpful [5]. To develop such a computer-aided system, Tamaki et al. [6] have proposed a recognition system to classify NBI endoscopic image patches into three-types (types A, B, and C3) based on the NBI magnification findings [4, 7] (see Figure 1). Furthermore, this system has been extended to a frame-wise recognition that classifies the center patch of each endoscopic video frame and shows classification results on a monitor by a frame-by-frame manner [8]. Although these systems have achieved high

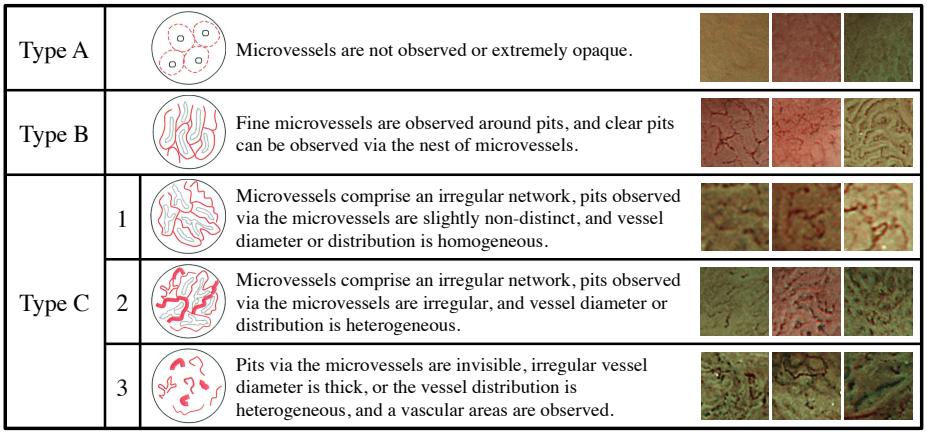

Fig. 1. NBI magnification findings [7].

recognition rate and have been confirmed its medical significance, an important limitation lies in the fact that they can only a part of images of the frame. For instance, in case that a tumor is not in the center of the frame or multiple tumors exist in the frame, these systems cannot provide appropriate objective measures. Therefore, recognizing an entire endoscopic image would be a further assistance for endoscopists during examinations, and could be used to train inexperienced endoscopists.

In this paper, we aim to assign labels each pixel in an entire NBI endoscopic image. Also for the same purpose, previously we proposed an image labeling method for endoscopic images that uses a posterior probability obtained from an SVM classifier trained with a Markov Random Field (MRF) [9], but the obtained results were not satisfactory enough. One reason lies in the large variation of the texture caused by geometrical and illumination changes. Colorectal polyps and intestinal walls are not flat but undulating (wave-like or spherical shapes). Furthermore, endoscopic images have high contrast textures due to the lighting condition of the endoscope. In such a circumstance, recognition methods would fail because texture descriptors such as BoVW, Gabor, wavelet, and LBPs become unstable to be computed. Another reason is the lack of spatial consistency of MRF framework. In general, object shapes and boundaries are roughly modeled by the pairwise term of an MRF model with edges in the image. However NBI endoscopic images used in our task often do not have clear boundaries between categories and therefore it would be difficult to model the edge information by the MRF.

Towards a robust texture representation, Xia et al. [10] proposed a texture descriptor, shape-based invariant texture analysis (SITA), based on a tree of shapes [11]. SITA consists of histograms of texture features computed from all nodes in a tree of shapes. Thanks to the hierarchical structure of a tree of shapes, SITA has the invariance to local geometric and radiometric changes. In

classification and retrieval experiments with texture image datasets, SITA was shown to achieve a better performance.

Inspired by the work of Xia et al. [10], we propose here a novel image labeling method for texture images using a tree of shapes. The basic idea is to compute histograms of texture features, such as SITA, at every node. Histograms of nodes are then classified to assign labels to the corresponding pixels. However, histograms of smaller nodes close to leaf would be less informative for classification. Therefore, our method aims to find subtrees having nodes discriminative enough for classification. We then introduce a threshold for selecting discriminative subtrees and formulate a joint optimization problem of estimating the threshold and training a classifier.

The rest of this paper is organized as follows. Section 2 reviews related medical and morphological work. Then, a tree of shapes and the SITA textures descriptor are briefly introduced in Section 3. We formulate the problem in Section 4. Section 5 shows some experimental results with an NBI endoscopic image dataset. Finally, we conclude this paper in Section 6.

2 Related work

Polyp segmentation is a well studied task in endoscopic image analysis. Gross et al. [12] proposed a segmentation method of NBI colorectal polyps using the Canny edge detector and Non-Linear Diffusion Filtering (NLDF), which is the first attempt for polyp segmentation of NBI endoscopic images. Ganz et al. [13] proposed Shape-UCM, an extension of gPb-OWT-UCM [14] for segmentation of polyps in NBI endoscopic images. Shape-UCM solves a scale selection problem of gPb-OWT-UCM by introducing a prior about the shape of polyps. Collins et al. [15] proposed a method using Conditional Random Field (CRF) with Deformable Parts Model (DPM) and a response of positive Laplacian of Gaussian filter (negative responses clipped to zero). Some other methods using watershed and region merging [16] or GrabCut [17] have proposed.

A popular approach for polyp segmentation is the use of active contours. Breier et al. [18] proposed a method for localizing colorectal polyps in NBI endoscopic images. They assumed that polyps appear as convex objects in a image, and introduced active rays to obtain a smoother contour. Figueiredo et al. [19] proposed a segmentation method for assessing the aberrant crypt foci captured by an endoscope. They used variational level sets and active contours without the edge model of Chan and Vase [20].

All of the above existing methods try to find contours between polyps and non-polyp intestine walls. Instead, we aim to assign labels conditions of cancer to pixels. To avoid confusion, we mention the term *segmentation* for finding contours and *labeling* for assigning pixel labels like as our task. The most similar task to ours is one conducted by Nosrati et al. [21]. To increase a surgeon's visibility in an endoscopic view, they label visible objects (such as tumors, organs, and arteries) in an endoscopic video frame. Their approach is based on trans-

ferring 3D data into 2D images. The pose and deformation of objects estimated from preoperative 3D data are aligned into 2D images.

In the field of computer vision, image labeling is a well studied task and many methods have been proposed including ones using CNN features. Farabet et al. [22] proposed a labeling method for scene parsing. Their approach assigns estimated labels to pixels, and then refines the results using superpixels, CRF, and optimal-purity cover on a segmentation tree. Long et al. [23] used a fully convolutional network trained in a end-to-end manner, and some more methods have been proposed [24–26]. Although these methods have achieved high accuracies, these need a large amount of training samples, which is impractical to medical image analysis; in general, it is difficult to collect a large amount of medical data in a short period of time. If fact, we have only a dataset of 63 NBI endoscopic images (described in Section 5). In contrast, our method trains classifiers with histogram features from thousands of nodes in a tree of shapes, hence requires relatively few training images.

In the field of mathematical morphology, a hierarchical representation, so-called morphological tree, is a popular framework, and a number of hierarchical trees have been proposed such as min/max-trees [27, 28], binary partition trees [29], minimum spanning forests [30], and tree of shapes [11]. Morphological trees have been applied to various images. One of the most popular application is biomedical imaging [31–33].

Meanwhile, Xia et al. [10] focused on the natural scale-space structure and invariance for contrast change of tree of shapes, they proposed a texture descriptor based on tree of shapes (details are described in Section 3.2). To the best of our knowledge, this is the first attempt to make texture descriptor from tree of shapes. Then, another texture descriptors based on tree of shapes have been proposed. Liu et al. [34] introduced a bag-of-words model of the branches in a tree of shapes and represented co-occurrence patterns of shapes. He et al. [35] adopted a basic idea of LBP, and proposed a texture descriptor. It divides a concentric circle of a shape into fan-based regions and computes the ratio of occlusion of a shape for each region. Histogram is computed from these ratios. However, these works handle only texture patch classification and retrieval and ignore multiple textures in a single image.

3 Tree of shapes and texture feature

In this section, we briefly introduce the definition of the tree of shapes and the SITA histogram features.

3.1 Tree of shapes

A tree of shapes [11] is an efficient image representation in a self-dual way. Given an image $u : \mathbb{R}^2 \to \mathbb{R}$, the upper and lower level sets of u are defined as $\chi_\lambda(u) = \{x \in \mathbb{R}^2 | u(x) \geq \lambda\}$ and $\chi^\lambda(u) = \{x \in \mathbb{R}^2 | u(x) < \lambda\}$ respectively, where $\lambda \in \mathbb{R}$.

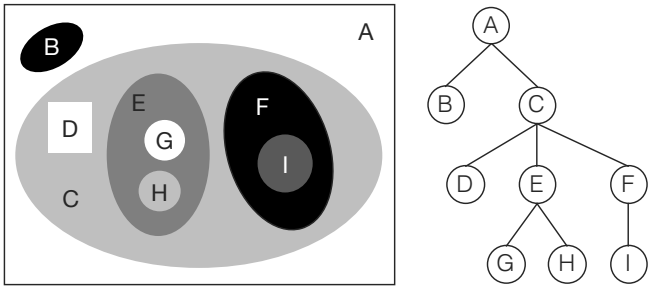

Fig. 2. An example of a synthetic image (left) and corresponding tree of shapes (right). Alphabets denote the correspondence between blobs and tree nodes.

From these level sets, we can obtain tree structures $\mathcal{T}_{\geq}(u)$ and $\mathcal{T}_{<}(u)$ that consist of connected components of upper and lower level sets: $\mathcal{T}_{\geq}(u) = \{\Gamma \in \mathcal{CC}(\chi_\lambda(u))\}_\lambda$ and $\mathcal{T}_{<}(u) = \{\Gamma \in \mathcal{CC}(\chi^\lambda(u))\}_\lambda$ where $\mathcal{CC}$ is an operator giving a set of connected components.

Furthermore, we define a set of upper shapes $\mathcal{S}_{\geq}(u)$ and lower shapes $\mathcal{S}_{<}(u)$. These sets are obtained by the cavity-filling (saturation) of components of $\mathcal{T}_{\geq}(u)$ and $\mathcal{T}_{<}(u)$. A *tree of shapes* of u is defined as the set of all shapes defined as $\mathcal{G}(u) = \mathcal{S}_{\geq}(u) \cup \mathcal{S}_{<}(u)$.

Thanks to the nesting property of level sets, the tree of shapes forms a hierarchical structure. Figure 2 shows an example of a tree of shapes. Given an image u whose image size is A. Let $T = \{V, E\}$ be a tree of shapes where V is a set of nodes, E a set of edges, $n_j \in V$ be nodes in the tree of shapes. We define parent and children nodes of n_j as $Pa(n_j) = \{n_k | (n_j, n_k) \in E, a_j < a_k\}$ and $Ch(n_j) = \{n_k | (n_j, n_k) \in E, a_j > a_k\}$ respectively, where a_j is area of n_j.

3.2 Shape-based invariant texture analysis

Xia et al. [10] proposed a texture descriptor based on the tree of shapes, *Shape-based Invariant Texture Analysis* (SITA). It consists of four shape features of the blob corresponding to a node.

Let s_j be a blob of n_j. The $(p+q)$-th order central moment μ_{qp} of s_j is defined by

$$\mu_{pq}(s_j) = \int\int_{s_j} (x_j - \bar{x}_j)^p (y_j - \bar{y}_j)^q dx_j dy_j, \tag{1}$$

where $(\bar{x}_j, \bar{y}_j)$ are the center of mass of s_j.

The normalized $(p+q)$-th order moments are

$$\eta_{pq}(s_j) = \frac{\mu_{pq}(s_j)}{\mu_{00}(s_j)^{(p+q+2)/2}}. \tag{2}$$

Then, two eigenvalues $\lambda_{1j}, \lambda_{2j}$ $(\lambda_{1j} \geq \lambda_{2j})$ of the normalized inertia matrix are computed as

$$\epsilon_j = \frac{\lambda_{2j}}{\lambda_{1j}} \tag{3}$$

and

$$\kappa_j = \frac{1}{4\pi\sqrt{\lambda_{1j}\lambda_{2j}}}, \tag{4}$$

where ϵ_j is elongation and κ_j is compactness. These are two shape features of a blob.

The third feature $\alpha(s_j)$ is computed from blob sizes and the parent-children relationship defined as

$$\alpha(s_j) = \frac{\mu_{00}(s_j)}{\sum_{s_k \in Pa^M(s_j)} \mu_{00}(s_k)/M}, \tag{5}$$

where $Pa^M(s_j) = \{s_m, \forall m \in (1, \ldots, M)\}$ is a set of M-th ancestor blobs. This feature is the ratio of blob sizes between s_j and the ancestor blobs, which is called a scale ratio. According to [10], we set $M = 3$ in our method.

The last feature is a normalized gray value computed for each pixel x in the image u as follows

$$\gamma(x) = \frac{u(x) - mean_{s(x)}(u)}{\sqrt{var_{s(x)}(u)}}, \tag{6}$$

where $s(x)$ is the smallest blob containing x. $mean_{s(x)}(u)$ and $var_{s(x)}(u)$ are mean and variance of pixel values over $s(x)$.

These are computed on every nodes (hence blobs) in the tree of shapes. The first three features are computed at all nodes, and the last feature is computed for all pixels in the image. Histograms of each feature are then constructed. These histograms are concatenated to form the SITA texture descriptor of the image, which is invariant to local geometric and radiometric changes because of the hierarchical structure of the tree of shapes.

4 Proposed method

In this section, we develop a method for selecting discriminative subtrees in a tree of shapes by using SITA at each node. Figure 3 shows an overview of the proposed method. In the training phase, the histogram features of all nodes are used for train a classifier and estimate a size threshold. In the labeling phase, histogram features of only nodes whose sizes are larger than the estimated threshold are classified to assign labels to the nodes, then map to the corresponding blobs. Hereafter, we introduce the details of the proposed method.

We extend basic notions of the tree of shapes defined for a single image to that for a set of images. Let $\{u_i\}_{i=1}^N$ be a set of images and A_i be the image size of i-th image u_i. A tree of shapes of image u_i is defined as $T_i = \{V_i, E_i\}$,

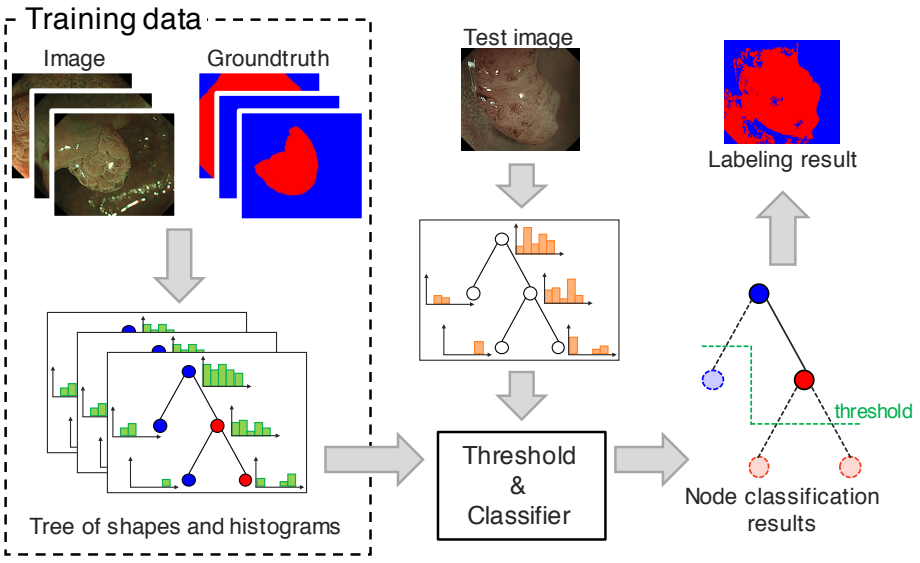

Fig. 3. Overview of the proposed method.

where V_i is a set of nodes and E_i a set of edges. Each node $n_{ij} \in V_i$ has the corresponding label $y_{ij} \in \{-1, 1\}$ and the area a_{ij} of the corresponding blob.

Herein, we explain the details of the histogram feature used in our method. In the work of Xia et al., a SITA descriptor is computed as a histogram feature for a given image. This means that the SITA is computed at the *root* node of the tree by aggregating feature from all descendant nodes. In contrast, our method constructs the histogram features at *all* nodes of the tree. Let $\boldsymbol{g}(n_{ij})$ be a histogram computed by the features from node n_{ij} only. Then the total histogram $\boldsymbol{h}(n_{ij})$ of node n_{ij} is computed recursively as

$$\boldsymbol{h}(n_{ij}) = \boldsymbol{g}(n_{ij}) + \sum_{n_{ik} \in Ch(n_{ij})} \boldsymbol{h}(n_{ik}), \tag{7}$$

then normalized to have a unit L1 norm. This means that $\boldsymbol{h}(n_{ij})$ is computed in a bottom-up manner, i.e., the computation is done from leaf nodes to the root node. For the sake of simplicity, we denote $\boldsymbol{h}(n_{ij})$ as $\boldsymbol{h}_{ij}$.

As we mentioned, discriminative subtrees useful for labeling are expected to exist in the tree of shapes. Figure 4 shows examples of labeling results with different subtrees. If we use smaller, less discriminative subtrees (such as leaf nodes), labeling fails; Using subtrees with large size nodes (e.g. the root node), labeling wouldn't be satisfactory because such nodes correspond to a large part of the image. Therefore, we introduce a threshold θ to node sizes for selecting discriminative subtrees, and define the objective function to estimate θ and

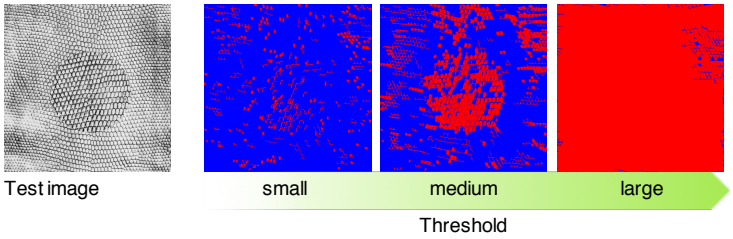

Fig. 4. Examples of labeling results using subtrees with different node sizes. Subtrees used for labeling are decided by an estimated threshold.

classifier parameters as follows:

$$E(\theta, \boldsymbol{w}, b) = \|\boldsymbol{w}\|^2 + \frac{1}{NM_i} \sum_i^N \sum_j^{M_i} m_{ij}(\theta)\ \ell\left(y_{ij}(\boldsymbol{w}^T \boldsymbol{h}_{ij} + b)\right) + \lambda \frac{\theta}{\bar{A}}, \quad (8)$$

where $\boldsymbol{w}$ and b are the weight and bias of an SVM classifier, respectively. The first term denotes a regularizer for the weight. The third term $\lambda \frac{\theta}{\bar{A}}$ is a regularizer for the threshold θ, where λ is a scale parameter and $\bar{A} = \frac{1}{N} \sum_i^N A_i$ is the mean size of N training images. $\ell(\cdot)$ is the hinge loss function of the SVM classifier.

$m_{ij}(\theta)$ represents the sample weight for $\boldsymbol{h}_{ij}$. In our method, we need to select the threshold θ to find discriminative subtrees. In other words, we have to use histograms of nodes whose area is larger than θ, otherwise ignore. Therefore, we define $m_{ij}(\theta)$ as a step function;

$$m_{ij}(\theta) = \begin{cases} 0 & if\ a_{ij} < \theta \\ 1 & otherwise \end{cases}. \quad (9)$$

4.1 Optimization

Given a training data, we need to estimate $\hat{\theta}, \hat{\boldsymbol{w}}, \hat{b}$ by minimizing the cost function;

$$\hat{\theta}, \hat{\boldsymbol{w}}, \hat{b} = \underset{\theta, \boldsymbol{w}, b}{\operatorname{argmin}} E\left(\theta, \boldsymbol{w}, b\right). \quad (10)$$

However, it is difficult to estimate all parameters at once. Therefore, we use block-coordinate decent to solve this optimization; estimate the threshold θ and the classifier parameters $\boldsymbol{w}, b$ iteratively. For estimating θ, we solve

$$\theta_k = \underset{\theta}{\operatorname{argmin}}\, E(\theta, \boldsymbol{w}_{k-1}, b_{k-1}). \quad (11)$$

This problem is non-convex because θ depends on histograms $\boldsymbol{h}_{ij}$. However, we experimentally confirmed that the cost function is rather smooth and have a

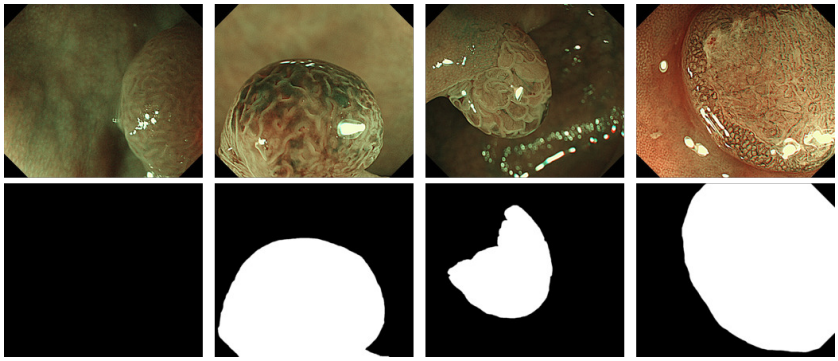

Fig. 5. Examples of images in the NBI endoscopic image dataset. Upper row shows NBI images and bottom row shows corresponding masks. White color of the mask represents foreground and black represents background. The left-most image is a negative sample which doesn't have any foreground region.

single minimum (details are discussed in Section 5.1). For estimating $\boldsymbol{w}$ and b, we solve

$$\boldsymbol{w}_{k+1}, b_{k+1} = \underset{\boldsymbol{w},b}{\operatorname{argmin}} E(\theta_k, \boldsymbol{w}, b). \tag{12}$$

This is an SVM formulation with sample weights, which is convex. For a large number of training samples, it would be difficult to obtain a nonlinear SVM problem within a practical time, therefore we solve the SVM in a primal domain by using the primal solver of LIBLINEAR [36].

We stop the alternation when θ converges with the termination criterion of

$$|\theta_k - \theta_{k-1}| < \epsilon. \tag{13}$$

4.2 Labeling procedure

After training phase, we label a test image as follows. First, we classify histograms $\boldsymbol{h}_{ij}$ of nodes n_{ij} if $a_{ij} \geq \hat{\theta}$, that is, the node are larger than the threshold, then assign the estimated labels to the nodes n_{ij}. For smaller nodes, we assign the label of their parent node. This procedure is done from the root node down to the leave nodes. Once labels are assigned to every nodes, labeling results are obtained by mapping labels of nodes into the corresponding blobs.

5 Experimental results

We have prepared a dataset of 63 NBI endoscopic images. Example images in the dataset are shown in Figure 5. Sizes of images are 1,000 × 870 pixels. There are two label categories (foreground and background) based on the NBI magnification findings (see Figure 1). Foreground regions correspond to polyps of types

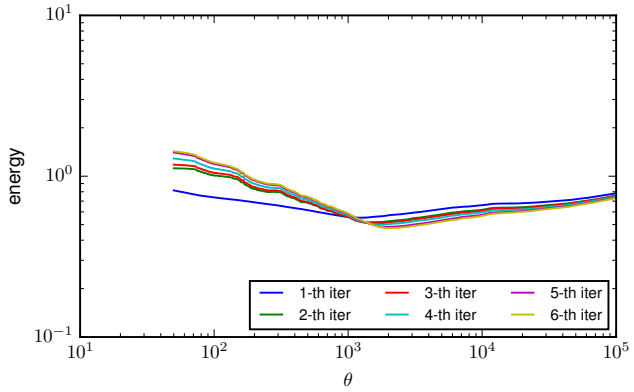

Fig. 6. Energies at each iteration. Horizontal axis shows θ value and vertical axis shows energy. Colorized curves are energies of each iteration. Initial value of $\theta_0 = 1000$ is used, $\lambda = 1.0$ is fixed.

B and C, and background regions are others (type A polyps, normal intestinal walls, and uninformative dark regions). Among 63 images, 20 images are negative samples which don't contain any foreground regions; the left-most image in Figure 5 captures only a hyperplastic polyp (i.e. benign tumor and non-cancer, hence Type A) labeled as background. A tree of shapes created from an NBI endoscopic image contains a large number of nodes. The average number of nodes from images in the dataset is 24,070. We randomly divided the dataset into half for training and test. We set parameters λ as 1.0 and initial value of threshold θ_0 as 1000.

We used two methods for comparison. One is to simply classify histograms of nodes in a tree of shapes and assign labels to pixels, which is corresponding to $m_{ij}(\theta) = 1$ in Eq. (8). This is a simple application of SITA for every nodes and is an obvious extension, while our proposed method is not. In the following experiments, we refer this method as conventional method. The other is a patch based segmentation method using MRF and posterior probabilities obtained from a trained SVM classifier [9]. For training SVM, we used 1,608 NBI endoscopic image patches (type A: 484, types B and C3: 1,124) trimmed and labeled by endsocpists. In this method, densely sampled SIFT features are extracted from these patches and converted as BoVW histograms. BoVW histograms are then used for training an SVM classifier. Small square patches corresponding to each site of the MRF grid are classified to obtain posterior probabilities used as the MRF data term. The MRF energy is minimized by Graph Cut for obtaining labeling results.

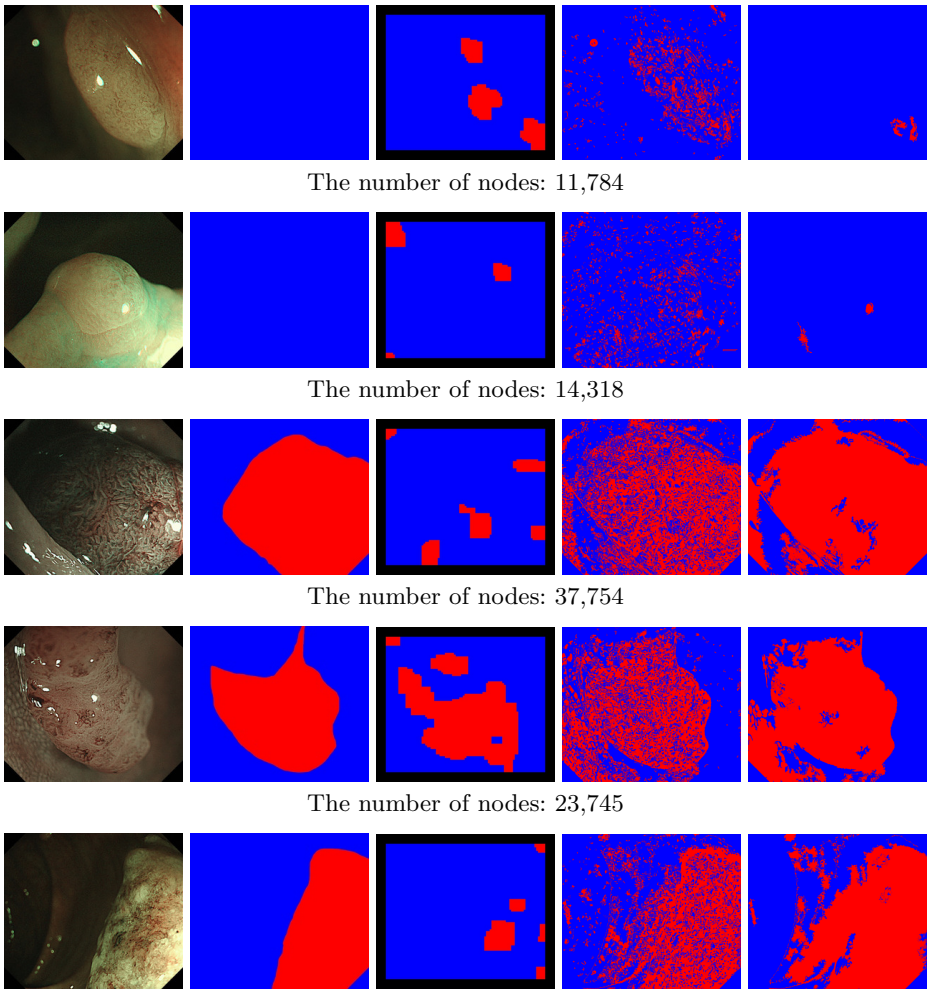

Fig. 7. Labeling results. From left to right: test image, ground truth, labeling result of SVM-MRF [9], conventional, and proposed. The number of nodes in the trees of shapes created from test images are shown below the images. Red color represents foreground and blue background. Black color of SVM-MRF results represents unlabeled region due to the boundary effect.

5.1 Labeling results of NBI endoscopic images

Figure 6 shows the cost function values over different threshold at each iteration in the training phase. We can see that the minimum of the cost function become smaller and threshold θ converges.

Figure 7 shows labeling results. As we mentioned above, we used the half of dataset (31 images) are used for training. The total number of nodes for training

Table 1. Dice coefficients of labeling results.

Method	dice coefficient
SVM-MRF [9]	0.555
conventional	0.522 ± 0.056
proposed	0.633 ± 0.041

is 747,937 and the primal solver for SVM training is necessary. The numbers of nodes of each test images are also shown in Figure 7. In SVM-MRF segmentation, labeling results are poor because the accuracy of MRF-based approaches highly depends on the data term. In other words, failures by the SVM classifier have the large impact on the poor accuracy. The conventional method provides cluttered labeling results because it classifies even small nodes. For instance, in the first two rows shows that the results of the conventional method provide small foreground regions. Meanwhile the proposed method can suppress the cluttered labels by selecting discriminative subtrees. In the middle and last two rows, foreground shapes of the proposed results are similar to the ground truth.

For quantitative evaluation, we used the dice coefficient [37]. Table 1 shows dice coefficients of each method. For conventional and proposed methods, we tested the procedures mentioned above repeatedly ten times and for the SVM-MRF method we tested only once. Note that the dice coefficient is calculated only for samples containing foreground. We can see that the proposed method outperforms the other two methods because using discriminative subtrees suppresses cluttered labels.

The proposed method outperforms the others in both the qualitative and quantitative evaluations. However, we need to discuss failure labeling results. Some failure examples are shown in Figure 8. In the case of top row, the image is almost labeled as foreground. A possible reason is that the used histogram is simply constructed from four low level features, which might be too few to be discriminative enough. Therefore, using richer texture features is included in our future work. The proposed method labels as background inside of the foreground region in the middle row. In our method, subtrees are selected by one threshold, but optimal thresholds may be different for different images, which is a limitation of the proposed method. Results of bottom row provide small foreground labels, which correspond to specular reflections (highlight) and the surrounding regions. Because highlights are large area nodes, texture features extracted from highlights may affect classification results, and dealing with highlights is also one of our future work.

About the computational time, our python implementation of the proposed method takes about 600 seconds for training and about 70 seconds for labeling an image. Although we handle ten thousand and more samples, our method can be trained within a practical time.

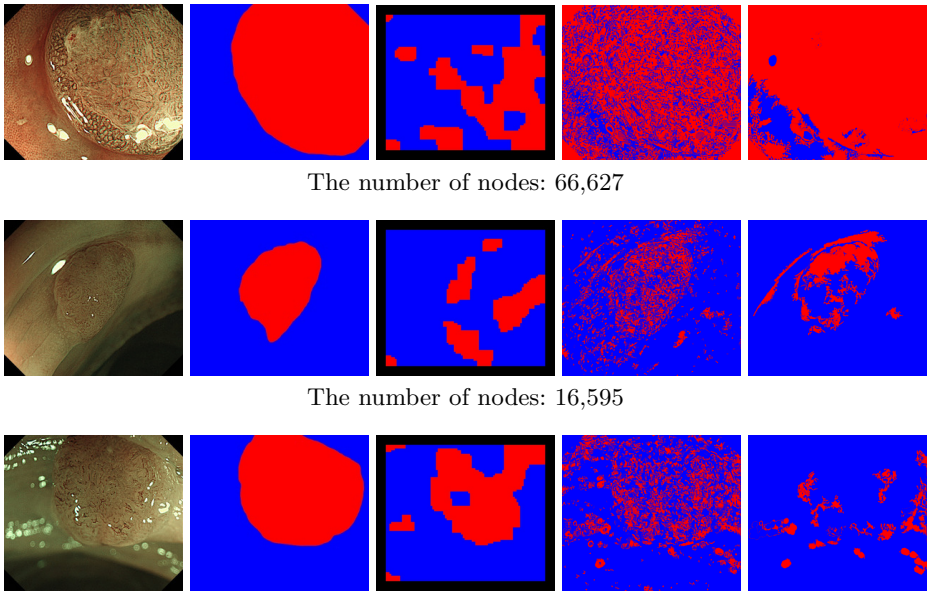

Fig. 8. Some failure examples. From left to right: test image, ground truth, labeling result of SVM-MRF [9], conventional, and proposed. The number of nodes in a tree of shapes created from the test image is shown in the bottom of images. Red color represents foreground and blue background. Black color of SVM-MRF results represents unlabeled region due to boundary effect.

6 Conclusion

In this paper, we proposed an image labeling method for NBI endoscopic images using a tree of shapes and histogram features derived from the tree structure. Our method selects optimal discriminative subtrees for tree node classification. This is formulated as a simultaneous optimization problem for estimating the threshold and classifier parameters and is solved via iterative block-coordinate decent. Then, we label images using the estimated parameters and the tree of shapes, by classifying each node from the root node to leave nodes, and mapping classification results into pixels. Experimental results on NBI endoscopic images show that the proposed method outperforms conventional methods and provides more reliable results.

Our future work includes improving the sample weights, extending to a multi-class problem, and seeking a more effective way of the labeling procedure using the hierarchical structure.

Acknowledgement

This work was supported in part by JSPS KAKENHI grants numbers JP14J00223 and JP26280015.

References

1. Cancer Research UK: Bowel cancer statistics. http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bowel/ (Accessed: 7 August 2016) (2015)
2. Meining, A., Rösch, T., Kiesslich, R., Muders, M., Sax, F., Heldwein, W.: Inter- and intra-observer variability of magnification chromoendoscopy for detecting specialized intestinal metaplasia at the gastroesophageal junction. Endoscopy **36** (2004) 160—164
3. Mayinger, B., Oezturk, Y., Stolte, M., Faller, G., Benninger, J., Schwab, D., Maiss, J., Hahn, E.G., Muehldorfer, S.: Evaluation of sensitivity and inter- and intra-observer variability in the detection of intestinal metaplasia and dysplasia in barrett's esophagus with enhanced magnification endoscopy. Scand J Gastroenterol **41** (2006) 349–56
4. 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. Scand J Gastroenterol **45** (2010) 1084–92
5. Takemura, Y., Yoshida, S., Tanaka, S., Kawase, R., Onji, K., Oka, S., Tamaki, T., Raytchev, B., Kaneda, K., Yoshihara, M., Chayama, K.: Computer-aided system for predicting the histology of colorectal tumors by using narrow-band imaging magnifying colonoscopy (with video). Gastrointest Endosc **75** (2012) 179–85
6. Tamaki, T., Yoshimuta, J., Kawakami, M., Raytchev, B., Kaneda, K., Yoshida, S., Takemura, Y., Onji, K., Miyaki, R., Tanaka, S.: Computer-aided colorectal tumor classification in NBI endoscopy using local features. Medical Image Analysis **17** (2013) 78 – 100
7. 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** (2009) 631 – 636
8. Kominami, Y., Yoshida, S., Tanaka, S., Sanomura, Y., Hirakawa, T., Raytchev, B., Tamaki, T., Koide, T., Kaneda, K., Chayama, K.: Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. Gastrointestinal Endoscopy **83** (2016) 643 – 649
9. Hirakawa, T., Tamaki, T., Raytchev, B., Kaneda, K., Koide, T., Kominami, Y., Yoshida, S., Tanaka, S.: Svm-mrf segmentation of colorectal nbi endoscopic images. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. (2014) 4739–4742
10. Xia, G.S., Delon, J., Gousseau, Y.: Shape-based invariant texture indexing. International Journal of Computer Vision **88** (2010) 382–403
11. Monasse, P., Guichard, F.: Fast computation of a contrast-invariant image representation. IEEE Transactions on Image Processing **9** (2000) 860–872
12. Gross, S., Kennel, M., Stehle, T., Wulff, J., Tischendorf, J., Trautwein, C., Aach, T. In: Polyp Segmentation in NBI Colonoscopy. Springer Berlin Heidelberg, Berlin, Heidelberg (2009) 252–256

13. Ganz, M., Yang, X., Slabaugh, G.: Automatic segmentation of polyps in colonoscopic narrow-band imaging data. IEEE Transactions on Biomedical Engineering **59** (2012) 2144–2151
14. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **33** (2011) 898–916
15. Collins, T., Bartoli, A., Bourdel, N., Canis, M. In: Segmenting the Uterus in Monocular Laparoscopic Images without Manual Input. Springer International Publishing, Cham (2015) 181–189
16. Bernal, J., Sánchez, J., Vilariño, F. In: A Region Segmentation Method for Colonoscopy Images Using a Model of Polyp Appearance. Springer Berlin Heidelberg, Berlin, Heidelberg (2011) 134–142
17. Hegadi, R.S., Goudannavar, B.A.: Interactive segmentation of medical images using grabcut. International Journal of Machine Intelligence **3** (2011)
18. Breier, M., Gross, S., Behrens, A., Stehle, T., Aach, T.: Active contours for localizing polyps in colonoscopic nbi image data (2011)
19. Figueiredo, I.N., Figueiredo, P.N., Stadler, G., Ghattas, O., Araujo, A.: Variational image segmentation for endoscopic human colonic aberrant crypt foci. IEEE Transactions on Medical Imaging **29** (2010) 998–1011
20. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Transactions on Image Processing **10** (2001) 266–277
21. Nosrati, M.S., Amir-Khalili, A., Peyrat, J.M., Abinahed, J., Al-Alao, O., Al-Ansari, A., Abugharbieh, R., Hamarneh, G.: Endoscopic scene labelling and augmentation using intraoperative pulsatile motion and colour appearance cues with preoperative anatomical priors. International Journal of Computer Assisted Radiology and Surgery **11** (2016) 1409–1418
22. Farabet, C., Couprie, C., Najman, L., LeCun, Y.: Learning hierarchical features for scene labeling. IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013) 1915–1929
23. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2015) 3431–3440
24. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. (2014) 580–587
25. Liu, F., Lin, G., Shen, C.: CRF learning with CNN features for image segmentation. Pattern Recognition **48** (2015) 2983 – 2992 Discriminative Feature Learning from Big Data for Visual Recognition.
26. Noh, H., Hong, S., Han, B.: Learning deconvolution network for semantic segmentation. In: 2015 IEEE International Conference on Computer Vision (ICCV). (2015) 1520–1528
27. Jones, R.: Connected filtering and segmentation using component trees. Computer Vision and Image Understanding **75** (1999) 215 – 228
28. Najman, L., Couprie, M.: Building the component tree in quasi-linear time. IEEE Transactions on Image Processing **15** (2006) 3531–3539
29. Salembier, P., Garrido, L.: Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. IEEE Transactions on Image Processing **9** (2000) 561–576
30. Cousty, J., Najman, L. In: Incremental Algorithm for Hierarchical Minimum Spanning Forests and Saliency of Watershed Cuts. Springer Berlin Heidelberg, Berlin, Heidelberg (2011) 272–283

31. Xu, Y., Géraud, T., Najman, L. In: Two Applications of Shape-Based Morphology: Blood Vessels Segmentation and a Generalization of Constrained Connectivity. Springer Berlin Heidelberg, Berlin, Heidelberg (2013) 390–401
32. Dufour, A., Tankyevych, O., Naegel, B., Talbot, H., Ronse, C., Baruthio, J., Dokládal, P., Passat, N.: Filtering and segmentation of 3d angiographic data: Advances based on mathematical morphology. Medical Image Analysis **17** (2013) 147 – 164
33. Perret, B., Collet, C.: Connected image processing with multivariate attributes: An unsupervised markovian classification approach. Computer Vision and Image Understanding **133** (2015) 1 – 14
34. Liu, G., Xia, G.S., Yang, W., Zhang, L.: Texture analysis with shape co-occurrence patterns. In: Pattern Recognition (ICPR), 2014 22nd International Conference on. (2014) 1627–1632
35. He, C., Zhuo, T., Su, X., Tu, F., Chen, D.: Local topographic shape patterns for texture description. IEEE Signal Processing Letters **22** (2015) 871–875
36. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: LIBLINEAR: A library for large linear classification. Journal of Machine Learning Research **9** (2008) 1871–1874
37. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26** (1945) 297–302