SMOOTHING POSTERIOR PROBABILITIES WITH A PARTICLE FILTER OF DIRICHLET DISTRIBUTION FOR STABILIZING COLORECTAL NBI ENDOSCOPY RECOGNITION

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

Hiroshima University, Japan

ABSTRACT

This paper proposes a method for smoothing the posterior probabilities obtained from classification results of time series input. We deal with this problem as a filtering problem with Dirichlet distribution and develop a particle filtering for this task. As a practical example of smoothing, we apply the proposed method to stabilizing NBI endoscopy recognition results over time. Experimental results demonstrate that our approach can effectively smooth highly unstable probability curves.

***Index Terms*—** time series filtering, Dirichlet distribution, particle filtering, colorectal NBI endoscopy

1. 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 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. [4] proposed an NBI endoscopy 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. Based on this NBI magnification findings, a 96% recognition rate was achieved with 10–fold cross validation.

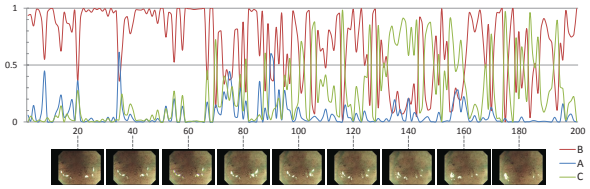


Fig. 1. An example output for an NBI image sequence. Posterior probabilities of three classes are shown in different colors. Horizontal axis: frame number. Vertical axis: posterior probabilities obtained from SVM.

This enables us to develop a real-time recognition and assessment system during endoscopic examination of the colon, which is expected to be of a great help for colonoscopy [7]. One 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 at each frame can be either a class label or the posterior probabilities of each class (showing the confidence in the estimated class labels).

This strategy is, however, not appropriate for observations by endoscopists. Figure 1 shows an example of recognition results for an NBI image sequence. Because each frame is processed independently, the posterior probabilities obtained by the classifier are highly unstable. The corresponding labels for each frame (Figure 2) are also unstable because those are simple MAP (maximum a posteriori) estimates at each frame. The output of the system therefore should be temporally smoothed and stabilized in order for endoscopists to easily understand the status of tumors of interest.

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 usually the smoothed result needs to be re-normalized at each frame to sum up to 1, which leads to inconsistency between frames as there is no theoretical and probabilistic background. Hirakawa et al. [8] proposed a probabilistic approach to this problem by using a sequential labeling within a Markov Ran-

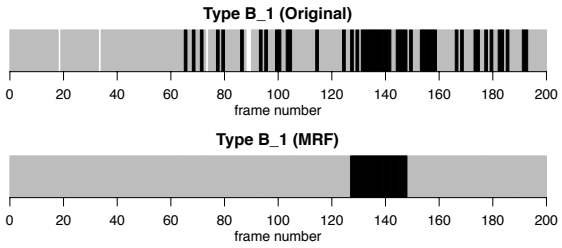


Fig. 2. Labeling result for an NBI video sequence of a type B polyp. Type A labels are shown in white, B in gray, C3 in black. (Original): posterior probability curves. (MRF): labeling result.

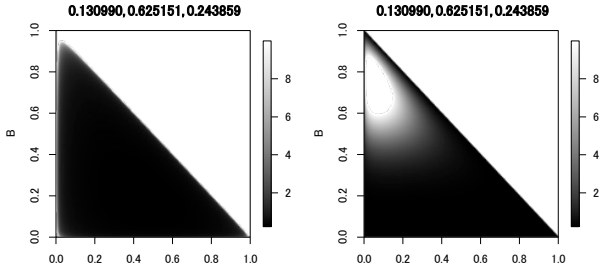
dom Field (MRF) framework [9]. Figure 2 shows the labeling results corresponding to the posterior probabilities shown in Figure 1. The frequent changes of labels in the top of Figure 2 are suppressed by the MRF labeling shown at the bottom. Labels are however less informative than the original probability curves as details are lost, hence not as useful to help endoscopists make a decision.

In this paper, we develop a particle filter [9] with Dirichlet distribution [10] for smoothing the temporal posterior probabilities. In contrast to the labeling, which is a MAP estimate of a sequence, our Dirichlet particle filter can provide richer information as well as the original probability curves because at each frame probabilities are retained. At each frame, an observation (i.e., posterior probability of $N = 3$ categories in our case) is obtained from an SVM classifier. The state to be estimated is modeled by a Dirichlet distribution. We carefully designed a state transition and likelihood by taking into account the task under consideration. Then overall filtering is implemented with particle filtering, an online Bayesian sequential filtering.

There is one closely related work to our approach: Mochihashi et al. [11] proposed multinominal particle filtering to estimate topic shifts in the context of natural language processing. To detect the topic shift in a document, latent topics are modeled by a multinominal distribution, which is generated by a Dirichlet distribution when a topic shift happens. This approach however is not applicable to estimate an internal state of posterior probabilities.

2. MODELING WITH THE DIRICHLET DISTRIBUTION

We are given observations $\boldsymbol{y}_t$ obtained at time t from a classifier as a result of an N class classification problem. In our case, $\boldsymbol{y}_t$ is a discrete posterior probability over $N = 3$ classes provided by an SVM classifier, and it can be seen as an N-dimensional vector whose L_1 norm is 1: $\|\boldsymbol{y}_t\|_1 = 1$. Let us denote $\boldsymbol{y}_{1:t} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t\}$ for a series of observations obtained before time t.



(a) $\boldsymbol{x}_{t-1}$ (b) $\theta_1 \boldsymbol{x}_{t-1}$

Fig. 3. State transitions modeled by Dirichlet distributions (a) $\mathrm{Dir}_{\boldsymbol{x}_t}[\boldsymbol{x}_{t-1}]$ and (b) $\mathrm{Dir}_{\boldsymbol{x}_t}[\theta_1 \boldsymbol{x}_{t-1}]$ where $\boldsymbol{x}_t = (0.13, 0.63, 0.24)$ and $\theta_1 = 10$.

In our approach, the state of the system $\boldsymbol{x}_t \in R^N$ at time t represents a true but unknown posterior probability at time t and also $\|\boldsymbol{x}_t\|_1 = 1$. Based on observations $\boldsymbol{y}_{1:t}$, we want to estimate the internal state $\boldsymbol{x}_t$ as a posterior $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$. To avoid confusion, hereafter we use the term "posterior" only to refer to $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$, which is a continuous probability distribution of $\boldsymbol{x}_t$. A vector $\boldsymbol{y}_t$ is called an observation while it is actually a discrete "posterior" obtained from an classifier at each time.

In Bayesian filtering, such as Kalman filtering or particle filtering, the posterior is estimated in two steps: a prediction step computes a prior $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})$ at time t based on observations before time $t-1$, and a filtering step calculates a posterior $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ at time t according to the prior and likelihood. These steps involve a state transition and a likelihood which are probability distributions over $\boldsymbol{x}_t$ or $\boldsymbol{y}_t$. In this paper, we use the Dirichlet distribution to represent those probability distributions. The Dirichlet distribution of order N with parameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N),\ \alpha_i > 0$ is given by

$$\mathrm{Dir}_{\boldsymbol{x}}[\boldsymbol{\alpha}] = \frac{\Gamma(\sum_{i=1}^N \alpha_i)}{\prod_{i=1}^N \Gamma(\alpha_i)} \prod_{i=1}^N x_i^{\alpha_i - 1}, \tag{1}$$

where Γ is the gamma function, $\boldsymbol{x} = (x_1, \ldots, x_N)$, and $\|\boldsymbol{x}\|_1 = 1$.

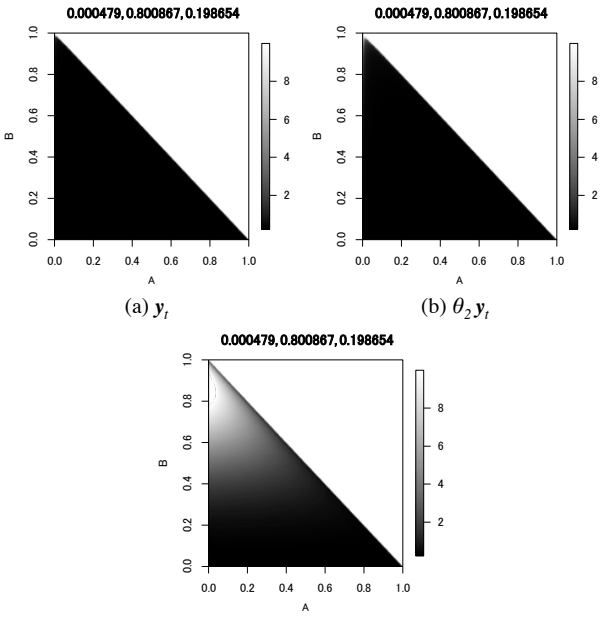
2.1. Prediction

The prediction needs to compute the following integral:

$$p\left(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}\right) = \int p\left(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}\right) p\left(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1}\right) d\boldsymbol{x}_{t-1}, \tag{2}$$

where $p\left(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}\right)$ is the state transition probability distribution between time $t-1$ and t. Since observations are expected to be highly unstable as shown in Figure 1, we impose a constraint on $\boldsymbol{x}_t$ be very close to $\boldsymbol{x}_{t-1}$. In other words, we make the probability density $p\left(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}\right)$ have a peak around $\boldsymbol{x}_{t-1}$.

To this end, we represent the state transition with the Dirichlet distribution with the parameter being a function of



(a) $\boldsymbol{y}_t$ (b) $\theta_2 \boldsymbol{y}_t$

(c) $\theta_2 \boldsymbol{y}_t + 1$

Fig. 4. Examples of likelihoods modeled by Dirichlet distributions (a) $\mathrm{Dir}_{\boldsymbol{x}_t}[\boldsymbol{y}_t]$, (b) $\mathrm{Dir}_{\boldsymbol{x}_t}[\theta_2 \boldsymbol{y}_t]$ ($\theta_2 = 100$), and (c) $\mathrm{Dir}_{\boldsymbol{x}_t}[\theta_2 \boldsymbol{y}_t + 1]$ ($\theta_2 = 3$), where $\boldsymbol{x}_t = (0.000479, 0.800867, 0.198654)$.

$\boldsymbol{x}_{t-1}$:

$$p\left(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, \theta_1\right) = \mathrm{Dir}_{\boldsymbol{x}_t}[\boldsymbol{\alpha}_1\left(\boldsymbol{x}_{t-1}, \theta_1\right)], \tag{3}$$

where the scalar θ_1 is an additional parameter.

In this paper, we propose to use the function $\boldsymbol{\alpha}_1\left(\boldsymbol{x}_{t-1}, \theta_1\right) = \theta_1 \boldsymbol{x}_{t-1}$ as the parameter of the state transition Dirichlet distribution. Figure 3 shows two examples of Dirichlet distributions of order $N = 3$. Each location inside the bottom-left triangle represents $\boldsymbol{x}_{t-1}$ and gray values represents probabilities (the higher the brighter). Fo $\alpha_i < 1$, the density has higher probabilities near the vertices but no peak where desired, while the density has a peak around the point $\boldsymbol{\alpha}_1 / \|\boldsymbol{\alpha}_1\|_1$ when $\alpha_i > 1$. The peak becomes steeper as the norm $\|\boldsymbol{\alpha}_1\|_1$ increases. Therefore, we use $\boldsymbol{x}_{t-1}$ to set the peak location of the distribution, and θ_1 to control the steepness of the peak. A typical choice for the value of θ_1 is between 10 and 100.

2.2. Filtering

The filtering step at time t computes the following posterior probability:

$$p\left(\boldsymbol{x}_t | \boldsymbol{y}_{1:t}\right) \propto p\left(\boldsymbol{y}_t | \boldsymbol{x}_t\right) p\left(\boldsymbol{x}_t | \boldsymbol{y}_{1:t-1}\right), \tag{4}$$

where $p\left(\boldsymbol{y}_t | \boldsymbol{x}_t\right)$ is the likelihood. We represent this likelihood of the observation $\boldsymbol{y}_t$ for given $\boldsymbol{x}_t$ with the Dirichlet distribution of $\boldsymbol{x}_t$ with the parameter $\boldsymbol{\alpha}_2$ as a function of $\boldsymbol{y}_t$

$$p\left(\boldsymbol{y}_t | \boldsymbol{x}_t, \theta_2\right) = \mathrm{Dir}_{\boldsymbol{x}_t}[\boldsymbol{\alpha}_2\left(\boldsymbol{y}_t, \theta_2\right)], \tag{5}$$

where θ_2 is an additional parameter to control the shape of the density. In other words, we regard the probability distribution of $\boldsymbol{x}_t$ as the likelihood of $\boldsymbol{y}_t$.

Here we choose the function $\boldsymbol{\alpha}_2\left(\boldsymbol{y}_t, \theta_2\right) = \theta_2 \boldsymbol{y}_t + 1$, which unlike the state transition, has an additional bias term. Figure 4 shows three examples of the Dirichlet distributions to demonstrate the effect of the bias term. A simple scaling function even with a large factor, as shown in Figure 4(b), may not provide a reasonable Dirichlet distribution with a broad peak around $\boldsymbol{y}_t$. The reason is that sometimes elements in the observation vector $\boldsymbol{y}_t$ are very small and hence the distribution does not have a peak as desired. To alleviate this problem, we simply employ a bias in addition to the scaling. A small bias like just one is enough to make the distribution have a peak. A typical choice for θ_2 is between 3 to 10.

Algorithm 1 Dirichlet Particle Filter.

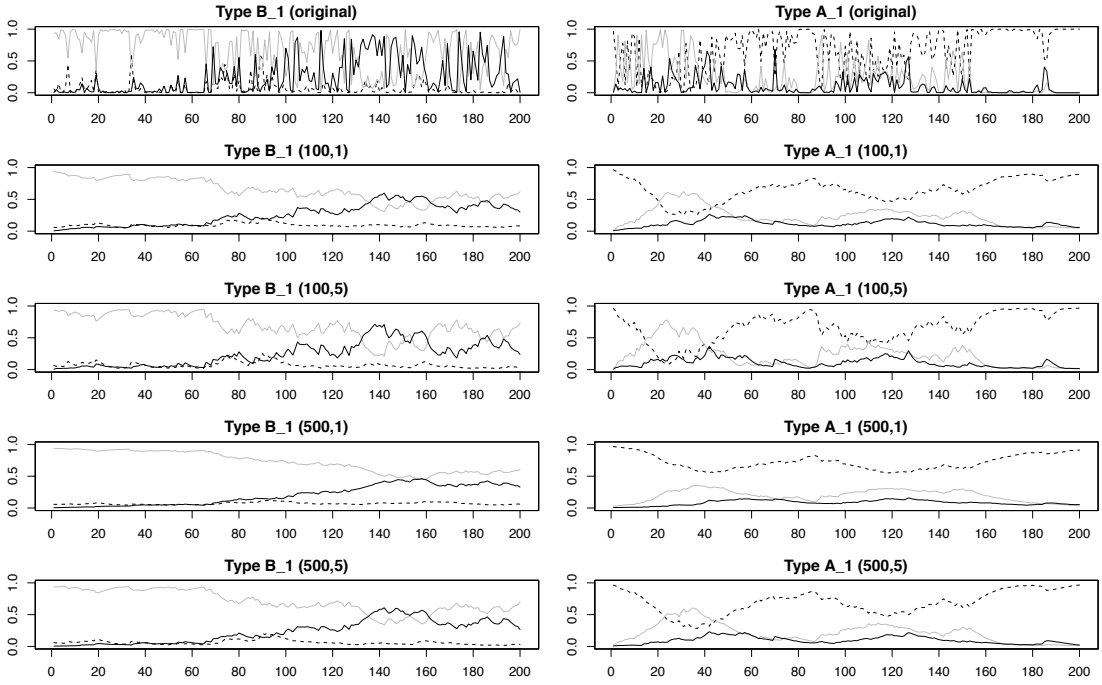
Generate K samples $\{\boldsymbol{s}_{0|0}^{(i)}\}_{i=1}^{K}$ from to an initial distribution $\mathrm{Dir}_{\boldsymbol{x}_1}[\boldsymbol{\alpha}_2(\boldsymbol{y}_1, \theta_2)]$.
for Frame number $t = 1 \ldots T$ **do**
  for $i = 1 \ldots K$ **do**
    Draw a sample $\boldsymbol{s}_{t|t-1}^{(i)} \sim \mathrm{Dir}_{\boldsymbol{x}_t}[\boldsymbol{\alpha}(\boldsymbol{s}_{t-1|t-1}^{(i)}, \theta_1)]$.
  end for
  for $i = 1 \ldots K$ **do**
    Compute a weight $\pi_t^{(i)} = \mathrm{Dir}_{x_t = \boldsymbol{s}_{t|t-1}^{(i)}}[\boldsymbol{\alpha}_2(\boldsymbol{y}_t, \theta_2)]$.
  end for
  Sample K times as $\{\boldsymbol{s}_{t|t}^{(i)}\}_{i=1}^{K}$ from $\{\boldsymbol{s}_{t|t-1}^{(i)}\}_{i=1}^{K}$ with replacement according to the weights $\pi_t^{(i)}$.
  Compute a MAP estimate $\hat{\boldsymbol{x}}_t$ at time t.
end for

2.3. Algorithm

The prediction and filtering steps are no more Gaussian and therefore need to be implemented by particle filtering. Algorithm 1 shows the steps of the proposed method. At each time step t, a MAP estimate $\hat{\boldsymbol{x}}_t$ is obtained and then visualized as a plot along with the input observation $\boldsymbol{y}_t$ in the experiments below.

3. EXPERIMENTAL RESULTS

We used a dataset of 907 trimmed NBI still images (Type A: 359, Type B: 462, Type C3: 87) to train an SVM classifiers [4]. To evaluate the proposed method, we used two NBI video sequences of length 200 frames, 640 $\times$ 480 pixels, and 30fps where large enough polyps were captured in each image.



(a) Type B polyp (b) Type A polyp

Fig. 5. Two smoothing results for NBI video sequences with three classes of Types A (dotted line), B (gray), and C3 (black). The horizontal axis shows the frame number, and the vertical axis is posterior probabilities. From top to bottom, input observation with no smoothing (original), and smoothing results for different parameter settings are shown (θ_1 and θ_2 are shown in that order).

The left column in Figure 5 shows smoothing results for the NBI sequence corresponding to Figures 2 and 1. The second to fifth rows are results obtained with parameters $\theta_1 = 100, 500$ and $\theta_2 = 1, 5$. The fourth row, $(\theta_1, \theta_2) = (500, 1)$, achieves the highest level of smoothing. In the right column of Figure 5 are shown results for another NBI sequence, and similar smoothing effect can be observed. We can see that generally larger values for θ_1 and smaller values for θ_2 can provide good smoothing results. Note that, we can get labeling results similar to the MRF approach by applying a MAP estimate to smoothing results.

4. CONCLUSION

In this paper, we have proposed a method for smoothing and stabilizing posterior probability curves with Dirichlet distributions implemented by particle filtering. In contrast to existing works, the proposed smoothing method is completely based on a probabilistic framework, and retains a good detail of the original sequences unlike sequence labeling methods.

The computational cost depends on K, the number of particles. Nevertheless, the method is very fast because the smoothing of 3-class posterior curves requires usually only a small number of particles.

Our future work includes the estimation of parameters θ_1 and θ_2 in an integrated Bayesian framework by using a set of NBI video sequences.

5. 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] Hiroyuki Kanao, Shinji Tanaka, Shiro Oka, Mayuko Hirata, Shigeto Yoshida, Kazuaki Chayama, "Narrow–band imaging magnification predicts the histology and invasion depth of colorectal tumors," Journal of Gastrointestinal Endoscopy, Vol. 69, No. 3, Part. 2, pp. 631-636, 2009.

[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] Tsubasa Hirakawa, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Tetsushi Koide, Shigeto Yoshida, Yoko Kominami, Taiji Matsuo, Rie Miyaki, Shinji Tanaka, "Labeling colorectal NBI zoom-videoendoscope image sequences with MRF and SVM," The 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013.

[9] Simon J.D. Prince, Computer Vision: Models, Learning, and Inference, Cambridge University Press, 2012.

[10] Christopher M. Bishop, Pattern Recognition and machine Learning, Springer, 2006.

[11] Daichi Mochihashi, Yuhi Matsumoto, "Context as Filtering," Advances in Neural Information Processing Systems 18, pp. 907-914, 2005.