Feature extraction from images of endoscopic large intestine

Masasi HIROTA †   Toru TAMAKI †   Kazuhumi KANEDA †   Shigeto YOSIDA ‡   and   Shinji TANAKA ‡

† : Hiroshima University
hirota@eml.hirohsima-u.ac.jp
tamaki@hirohsima-u.ac.jp
kin@hirohsima-u.ac.jp
‡ :Hiroshima University Hospital

Abstract   In this paper, we propose feature extraction methods from two types of images of endoscopic large intestine taken by a colonoscopy for diagnosis of colon cancer. Today, there are two observation methods. One is staining surface of large intestine. The other is colonoscopy using Narrow Band Imaging (NBI) system, a new feature of endoscope. We describe extraction methods of features for each observation method so that the features may be used to estimate colon cancer staging from an observed image.Pit pattern is a texture that appears on the surface of stained intestine and they are categorized and used for diagnosis. Thus, we extract pits from an endoscope image to analyze patterns. First, color edge of the image is extracted, then watershed segmentation is applied. In the result, pits are roughly extracted. NBI system can observe vascular structure under the surface of large intestine. The vascular structure can be used to estimate cancer staging. A vascular area is roughly extracted by adaptive binarization, then the fine shape of vascular area is extracted by the level set method

1   INTRODUCTION

According to "VITAL STATICS OF JAPAN HEALTH (2005)"[1], about 40,000 people die from colon cancer. It follows lung cancer (60,000) and stomach cancer (50,000). An endoscope have been a major way for cancer test but only 6.2% of small clinics whereas 59.4% of hospitals have. It is because the endoscope is difficult to use, and diagnosis from endoscopic image requires skill. Therefore, a diagnosis system that automatically estimates cancer staging from an endoscopic image by medical nonspecialists is expected. Providing quantitative measurements that does not depend on doctor's subject is also important.

Today, there are two observation methods. One

is staining surface of large intestine and another is colonoscopy using NBI system. Even if a same part is observed, two different images are taken. Therefore, the final goal is to estimatie cancer staging by these images. Extracting features from these images is needed as a preliminary step for the goal.

A texture called *pit pattern* is observed in a stained surface image and it is basis of diagnosis. A part, which is suspected to be a cancer, is stained to observe fine tissue in colonoscopy. By staining, color difference between pit and nonpit area. Therefore, contour of a pit is emphasized by color edge extraction. Then pit area is extracted by watershed segmentation (section.3).

A vascular area is observed in NBI system

colonoscopy. There is relationship between vascular structure and cancer. A vascular area is shown as black area (low pixel value area) in NBI image. Therefore, a vascular area is roughly extracted by adaptive binarization. Then the fine shape of vascular area is extracted by level set method based on this result (section.4).

In the following, two different observation methods are described in Section 2 and pit pattern extraction from a stained surface image is described in Section 3. In Section 4, vascular area extraction from NBI image is described. We conclude the paper and future work is described in Section 5.

2 Observational methods

2.1 Staining surface of large intestine

This observational method is usually done at diagnosis today. First, pit pattern is emphasized by staining a surface of large intestine. Then a doctor make a diagnosis by observing the emphasized pit pattern. Pit pattern, observed on stained surface, is categorized into five stages shown as Fig.1, and procedure that estimates organizational diagnosis by pit pattern is established. Therefore we can determine a tissue is tumor or not, if the pit pattern is staged. But the diagnosis depends on a doctor's subject.

2.2 Observation by NBI system

Narrow Band Imaging (NBI) system is a new feature of an endoscope developed in these years. Outline of NBI system is shown in Fig.3. A light from a xenon light source is spectral filtered by a RGB filter of a rolling turret. A light with narrow band in short wavelength makes black in hemoglobin in NBI images. Therefore, NBI system

I		Round pit (normal pit)
II		Asteroid pit
IIIs		Tubular or round pit that is smaller than the normal pit
III$_L$		Tubular or round pit that is larger than the normal pit
IV		Dendritic or gyruslike pit
Vi		Irregular arrangement and sizes of III$_L$, III$_s$, IV type pit pattern
V$_N$		Amorphous or nonstructural pit pattern

Fig.1：A schematic diagram of pit pattern classification[4]
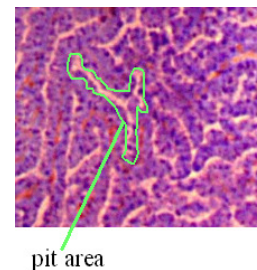

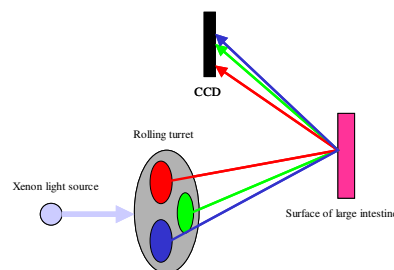
pit area

Fig.2: Stained endoscopic image

Fig.3: NBI system

can observe vascular structure. The advantage of NBI system is observation without stain process.

Generally, there is relationship between vascular structure and cancer staging.

3 Pit pattern extraction from stained intestine surface

3.1 Color edge extraction

2

1/4	2/4	1/4
0	0	0
-1/4	-2/4	-1/4

1/4	0	-1/4
2/4	0	-2/4
1/4	0	-1/4

2/4	1/4	0
1/4	0	-1/4
0	-1/4	-2/4

0	1/4	2/4
-1/4	0	1/4
-2/4	-1/4	0

Fig.4: Robinson operators

There is a large color difference between pits and around areas because the large intestine surface is stained. Therefore, first we extracte color edge .

Each pixel value of endoscopic image is converted RGB to L*a*b and robinson operators (shown in Fig.4) is applied. Let ΔL, Δa, Δb be values of filtered by L*, a*, b*. The magnitude of edge is calculated by the following equation.

$$\Delta Lab(x) = \sqrt{(\Delta L(x))^2 + (\Delta a(x))^2 + (\Delta b(x))^2} \quad (1)$$

The color edge value of a pixel x is the largest value of filtered pixel value by the four filters.

3.2 Watershed method

It is difficult to extract pit's contour from the extraction result of color edge by simple binarization because the measure of color edge is not uniform over the pit's contour. Calculated color edge value is extremum in pit's contour because of color changes around pit's contour in the original image. Therefore, pit's contour can be extracted using watershed method but as a result, a pit area is oversegmented. We combine area merging and watershed segmentation as shown in Fig.5. Specifically, average L*a*b* value of each water area is calculated as watershed segmentation processes. Then color difference is calculated using Eq.(1) when two water areas are overlapped. If the color difference is less than a threshold value, watershed is not made and merge these areas.
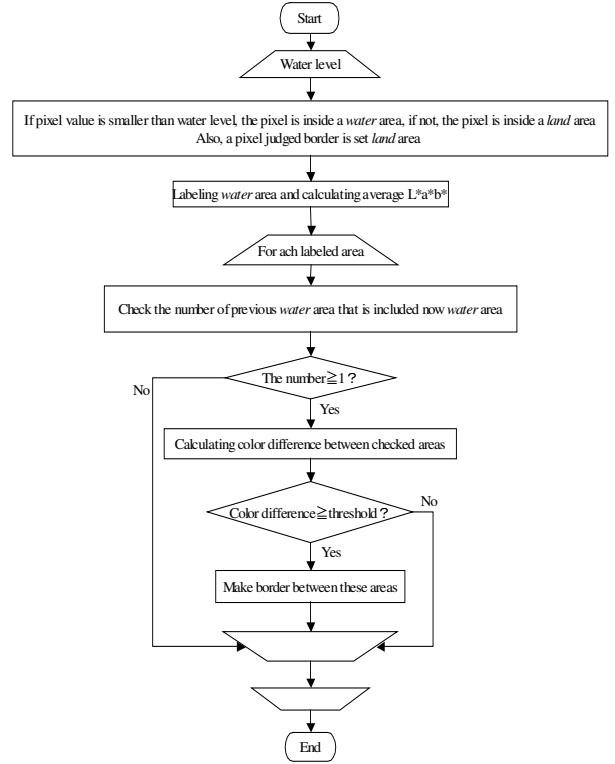
Start

Water level

If pixel value is smaller than water level, the pixel is inside a *water* area, if not, the pixel is inside a *land* area. Also, a pixel judged border is set *land* area

Labeling *water* area and calculating average L*a*b*

For ach labeled area

Check the number of previous *water* area that is included now *water* area

The number ≧ 1 ?   No   Yes

Calculating color difference between checked areas

Color difference ≧ threshold ?   No   Yes

Make border between these areas

End

Fig.5: Flow chart of area merging

3.3 Extraction results

An extraction result is shown in Fig.6. A size of target image is 250 by 250 pixel (actual size is about 1 by 1 mm). Pit patterns are well extracted from high contrast image as shown in Fig.6(d), but they are not well extracted from low contrast image as shown in Fig.6(c). This is due to the low color difference between pit area and nonpit area, and some pit areas are processed as nonpit areas. Also, even if extraction result is well, nonpit area is extracted too. Therefore, the solution of these problems is future work.

Although the extracted results include nonpit areas, we select manually only pit areas to see the relation between staging and pit features. Calculated features are area, principal axis length, boundary length.
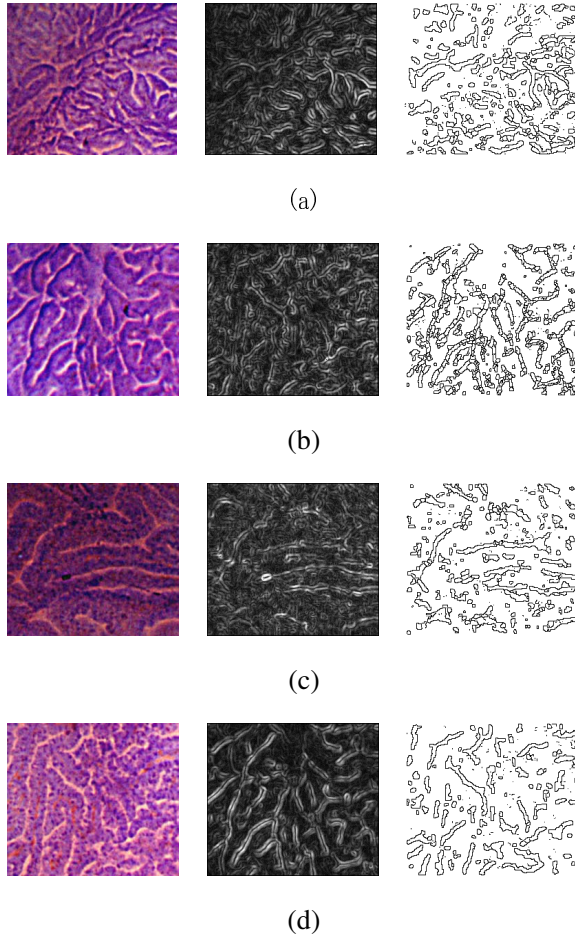
(a)

(b)

(c)

(d)

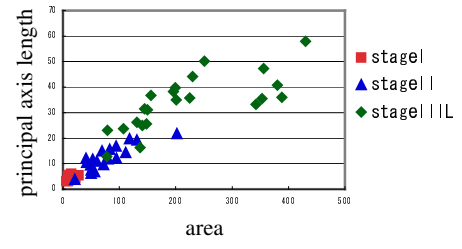Fig.6: left: Target image.　middle: Color edge extracted result.　right: Extracted result.

(a) III$_L$　(b)III$_L$　(c)IV　(d)Vi

(a)

(b)

Fig.7: Calculated result
(a) area - principal axis length
(b) area - boundary length

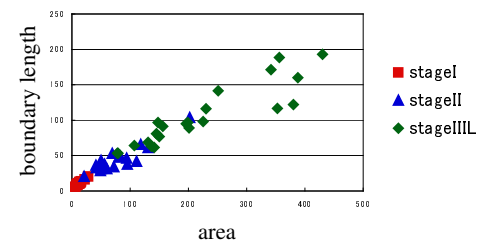Area: The number of pixels of an area.

Principal axis length: Maximum length of an area along the principal axis.

Boundary length: The number of pixels of boundary calculated by contour tracking.

Relations between these features and stages are shown as Fig.7(a) and (b). In the result, cluster of each stage (only three stages) is looked. Therefore, if pit is selected automatically, cancer staging can be estimated.

4　Vascular extraction from NBI image

4.1 Adaptively binarization

Vascular area is shown as black area (low pixel value area) in NBI image. Therefore, vascular area is extracted by binarization.

Green component of NBI image (Green image) and original NBI image have similar contrast as shown in Fig.8. Therefore, vascular area is extracted from Green image by binarization. But brightness of Green image is not uniform over the image. Therefore, adaptive binarization shown in the following is applied.

(1)　*N by N* average filter is applied.

(2)　A pixel value *X* is calculated by Eq.(2).

$$X = \begin{cases} x_2 - x_1 & (x_2 \geq x_1) \\ 0 & (x_2 < x_1) \end{cases} \tag{2}$$

4

Where x_1 is pixel value before filter is applied, x_2 is pixel value after applied.

(3)   If X is larger than threshold, the pixel is in a vascular area.

Result is shown in Fig.9. The size of the target image shown in Fig.8(a) is 340 by 290 pixel (actual size is about 1.5 by 1.3 mm).

Vascular area is roughly extracted as shown in Fig.9(a)~(c), but problem is that one vascular area is segmented into several areas. We apply active contour model to improve segmentation.

4.2 Extraction by Level Set Method

We use Level Set as an active contour method to merge vascular areas. Suitable speed function is needed to extract a vascular area. We use a speed function F as shown Eq.(5).

$$F = \frac{F_1}{F_2} \qquad (5)$$

$$F_1 = \begin{cases} -1 & (I(x, y) < T) \\ 1 & (I(x, y) \geq T) \end{cases} \qquad (6)$$

$$F_2 = 1 + \nabla S \otimes I(x, y) \qquad (7)$$

$I(x,y)$ is pixel value at (x,y). If F_1 is smaller than threshold value T, F_1 is -1, and if not, F_1 is 1. If the contour is inside a vascular area, the contour spreads, and if not, the contour shrivels. F_2 is a result of applying sobel filter, and small in edge area.

4.3 Discussions

The extracted result is shown in Fig.10(c). And the result by simple binarization is shown in Fig.10(b). These results are similar. Because F_1 is
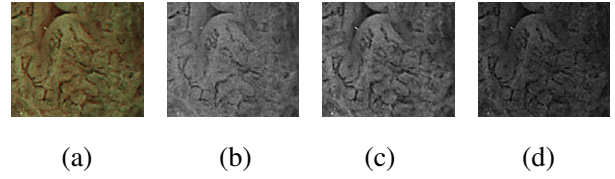

(a)   (b)   (c)   (d)

Fig.8: Target image (a) and each color elements (b) R, (c) G, (d) B

(a) N=20   (b) N=30   (c) N=40
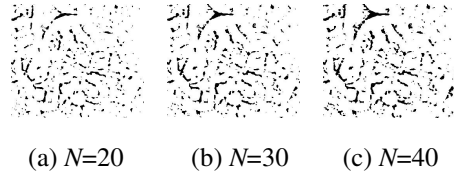
Fig.9: Processed result (t=20)

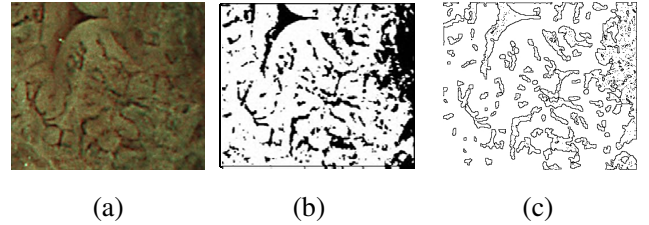(a)   (b)   (c)

Fig.10: Comparing binarized result and extracted result
(a)Target image   (b)Binarization   (c) Extracted result

set base on the threshold value. In fact, the threshold value of binarization as shown in Fig.10(b) and the threshold value T of F_1 of the extracted result as shown in Fig.10(c) are the same as 70. Setting more suitable speed function is future work.

5   Conclusions

We proposed feature extraction methods from two types of images of endoscopic large intestine taken by a colonoscopy for diagnosis of colon cancer. A pit pattern is extracted from a stained surface image but is not well extracted from low contrast image. Also even if extraction from high contrast image, there are problem that extracted

result has nonpit area is also extracted. But relation between staging and selected pit features is looked, so if pit is selected automatically, cancer staging can be estimated. A vascular area is extracted from NBI image by Level Set Method based on adaptive binarized result. In the result, vascular area is not well extracted because of unsuitable speed function. Setting more suitable speed function is future work.

References

[1] Statistics and Information Department, Minister's Secretariat, Ministry of Health, Labour and Welfare, " Vital Statistics of Japan Health Volume.3", Health and Welfare Statistics Association, 2005.

[2] Luc Vincent, Pierre Soille, "Watershed in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.13, No.6, pp.583-598, 1991

[3] Stanley Osher, Nikos Paragious, "Geometric Level Set Methods in Imaging, Vision, and Graphics", Springer-Verlag, 2003.

[4] Mayuko Hirata, Shinji Tanaka, Shiro Oka, Iwao Kaneko, Shigeto Yoshida, Masaharu Yoshihara, Kazuaki Chayama, "Magnifying endoscopy with narrow band imaging for diagnosis of colorectal tumors", Gastrointestinal Endoscopy, Vol.65, No.7, pp.988-995, 2007.

[5] Mayuko Hirata, Shinji Tanaka, Shiro Oka, Iwao Kaneko, Shigeto Yoshida, Masaharu Yoshihara, Kazuaki Chayama, "Evaluation of microvessels in colorectal tumors by narrow band imaging magnification", Gastrointestinal Endoscopy, Vol.66, No.5, pp.945-952, 2007.