Quantitative analysis and development of a computer-aided system for identification of regular pit patterns of colorectal lesions

Yoshito Takemura, MD,[1] Shigeto Yoshida, MD,[2] Shinji Tanaka, MD,[2]

Keiichi Onji, MD,[1] Shiro Oka, MD,[2] Toru Tamaki, PhD,[3] Kazufumi Kaneda, PhD,[3]

Masaharu Yoshihara, MD,[4] Kazuaki Chayama, MD[1]

[1]Department of Medicine and Molecular Science, Graduate School of Biomedical Sciences, Hiroshima University, Hiroshima, Japan

[2]Department of Endoscopy, Hiroshima University Hospital, Hiroshima, Japan

[3]Department of Information Engineering, Graduate School of Engineering, Hiroshima University, Hiroshima, Japan

[4]Department of Health Service Center, Hiroshima University, Hiroshima, Japan

The work described herein was performed at the Department of Endoscopy, Hiroshima University Hospital, Hiroshima, Japan.

Running title: Computer-aided identification of the pit patterns of colorectal lesions

Address for correspondence and reprint requests:

Shigeto Yoshida, MD, Department of Endoscopy, Hiroshima University Hospital, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8551, Japan

Tel: +81-82-257-5538

Fax: +81-82-257-5538

E-mail: yoshida7@hiroshima-u.ac.jp

Dr. Takemura 10.023RR

ABSTRACT

Background: Because pit pattern classification of colorectal lesions is clinically useful in determining treatment options for colorectal tumors but requires extensive training, we developed a computerized system to automatically quantify, and thus classify, pit patterns depicted on magnifying endoscopy images.

Objective: To evaluate the utility and limitations of our automated pit pattern classification system.

Design: Retrospective study.

Setting: Department of endoscopy, university hospital.

Main Outcome Measurement: Performance of our automated computer-based system for classification of pit patterns on magnifying endoscopic images in comparison to classification by diagnosis of the 134 regular pit pattern images by an endoscopist.

Results: For types I and II pit patterns, the results of discriminant analysis were in complete agreement with the endoscopic diagnoses. For type IIIL, out of 30 cases, 29 (96.7%) were diagnosed as type IIIL and 1 as type IV. For type IV, 29 out of 30 cases (96.7%) were diagnosed as type IV pit pattern. Overall accuracy of our computerized recognition system was 132/134 (98.5%).

Conclusions: Our system is best characterized as semi-automated but a step toward development of a fully automated system to assist in the diagnosis of colorectal lesions based upon classification of pit patterns.

Word count: 195

Key words: Computer-aided recognition; pit pattern

INTRODUCTION

Magnifying endoscopy permits detailed visualization of the surface of the gastrointestinal tract and thus allows examination of the pit pattern (shape of the openings of colorectal crypts) of colorectal tumors (Figure 1).[1] Pit pattern classification has been shown to aid in the differentiation of non-neoplastic and neoplastic colorectal lesions[2] and thus may be able to guide therapeutic decisions. Magnifying endoscopy with pit pattern recognition can be done during routine colonoscopy with indigo carmine dye spraying or crystal violet staining with little added expense or time. We have been interested in developing a software program that can analyze pit patterns quantitatively and thus be used with magnifying endoscopy to diagnose colorectal tumors. We created a custom software program for this very purpose and herein describe its development and an experimental study in which we tested its clinical utility and limitations.

METHODS

Image analysis software

We developed a custom software program (HuPAS ver. 1.3) that can outline various regular pits identified on digitized endoscopic images. HuPAS is designed to mark the color edges of a pit outline based on differences in color tone between the pit outline and the background; it then automatically extracts the identified pits. The software identifies regional segmentation using a watershed algorithm and combines integrated regions with excessive segmentation.[3]

Endoscopic procedure

The HuPAS image analysis software was tested on magnifying endoscopy images obtained from patients who had undergone diagnostic endoscopic study at Hiroshima University

Hospital. An Olympus CF-H260AZI magnifying colonoscope (Olympus Co., Tokyo, Japan) was used, which provides magnification up to ×70 (optical magnification) on a 19-inch monitor. After performing white light endoscopy, we examined the lesion at maximum magnification with crystal violet staining. Thereafter, images were digitized and stored on an Olympus EICP-D HDTV recorder (1,440×1,080 pixels). Informed consent was obtained from patients and/or family members for endoscopic examination.

Image processing

Pit region extraction

From each magnified endoscopic image recorded at maximum optical magnification (Figure 2a), a 250×250-pixel region was cut out as a region of interest (ROI) (Figure 2b). The cut-out image was processed automatically by using the custom HuPAS ver. 1.3 software to outline various pits identified on the digitized image (Figure 2c). The computer operator (S.Y.) removed non-pit regions and/or joined excessively segmented pit regions manually if necessary using Adobe Photoshop (Adobe Systems Inc. San Jose, CA) (Figure 3).

Quantification of pit features

Using ImageJ software (National Institutes of Health, Bethesda, MD), we quantified the following six shape descriptors for each pit on the extracted images: area, perimeter, major and minor axes of the best fit ellipse, circularity (represented as $4\pi(\text{area}/\text{perimeter}^2)$), wherein $1.0 =$ a precise circle that becomes an elongated polygon as it approaches 0.0), and Feret's diameter (longest distance between any two points within the selected frame) (Figure 4). These six shape descriptors were chosen for quantification because statistical analysis (Kruskal-Wallis test) revealed that differences in the values of these descriptors between types of regular pit patterns (Kudo and Tsuruta classes sI, sII, sIIIL, sIIIs, sIV) were significant.

Quantitative analysis of regular pit patterns

Feature extraction and quantification of the regular pit

Colorectal magnifying endoscopy was performed in 72 cases at Hiroshima University Hospital between June 2007 and September 2007. Images of regular pits were obtained from the magnified endoscopic images (type I: 20 cases; type II: 10 cases; type IIIL: 10 cases; type IIIs: 2 cases; and type IV: 30 cases) as a set of training images. After extracting the pit region of the ROI using HuPAS, an endoscopist (Y.T.) selected separate images of regular pits based on the morphology of the opening of the colorectal crypt and classified each pit, according to the Kudo and Tsuruta classification system (Figure 3b), as type sI (round), type sII (asteroid), type sIIIL (larger than type sI, ranging from tubular to round), type sIIIs (smaller than type sI, ranging from tubular to round), or type sIV (dendritic or gyrus-like). Based on the image processing described above, quantitative features were defined for each pit type.

Computer-aided identification of regular pit patterns

A set of validation images was gathered from among images obtained in other cases examined at Hiroshima University Hospital. We excluded lesions that were not suitable for evaluation (exclusion criteria: out-of-focus images, images that showed insufficient staining, images that were blurred). A total of 134 regular pit pattern images (type I: 32 cases; type II: 43 cases; type IIIL: 29 cases; and type IV: 30 cases) were sequentially obtained and comprised the validation set. After the separate pit types were determined based on the imaging processing described above, a discriminant analysis using JMP statistical software (SAS Institute Inc. Cary, NC) was conducted by referring to the quantitative characteristics of the set of training images for automated recognition of pit patterns on the endoscopic images. Discriminant analysis is used for estimating the population to which sample data belong when the population in which the sample data reside is unknown. Thus, we first obtained quantitative characteristics for each pit type as reference data and then conducted discriminant analysis for each pit based on these data. Together with the subsequent step of weighing the proportion of each pit within the respective images, we defined these steps as the two steps required to

definitively identify the patterns of the pits identified on endoscopic images. In addition, the pit

patterns on the validation images were classified according to the Kudo and Tsuruta system by

the same endoscopist (Y.T.), who was blinded to the computer-aided results.

135

Statistical analysis

Values are reported as mean ± SD. Differences in the six quantitative features (shape

descriptors) between the various pit patterns were analyzed by Kruskal-Wallis test, with

significance accepted at $P < 0.05$.

140

RESULTS

Quantification of the regular pit patterns

Values for each of the six shape descriptors are shown per pit pattern in Table 1. For

145 each of the six features, differences in values between the five regular pit patterns were

significant.

Automated identification of regular pit patterns

Performance of the automated computer-aided system for pit pattern classification of

150 colorectal lesions is shown relative to endoscopy findings in Table 2. Overall accuracy of the

automated computer-aided system for identification of regular pit patterns was 132/134

(98.5%).

DISCUSSION

155

To our knowledge, there are no reports of computerized quantitative analysis of pit

patterns of the colorectal mucosal surface. Computerized quantification of the pit pattern of a

colorectal lesion should allow for objective diagnosis, avoiding subjectivity and eliminating the need for extensive training in evaluating pit patterns.

160   We developed a software program, known as HuPAS, that can be used to outline and characterize various pits in the colonic mucosa on endoscopically obtained images. We found that values for six shape descriptors differed significantly between regular separate pit patterns (types sI-sIV), so these shape descriptors became the basis for our quantitative analysis.

      We also analyzed the accuracy of the automated computerized identification of the pit 165 patterns in reference to endoscopic diagnosis. The overall diagnostic accuracy of the computer-aided diagnosis based on automated calculation of the pit area was 98.5%. Thus, the custom software and computer-aided diagnosis algorithm together approached the diagnostic ability of the trained endoscopist.

      Our automated system is limited in that some non-pit regions are extracted with the pit regions and 170 some pit regions are excessively segmented. This is because some non-pit regions differ in color tone from the background and some pit regions are of low contrast. Thus, it was necessary to remove these non-pit regions and/or join excessively segmented pit regions using Adobe Photoshop in 11% of cases. This procedure does not make the process subjective, but it does add a manual step. In addition, endoscopic images that were out-of-focus, that showed insufficient staining, or that were blurred could not be 175 evaluated. It might be possible to overcome these limitations by adding another algorithm. Unfortunately, our computer analysis takes several minutes, so the results are not available during colonoscopic examination. Improvements are needed that will allow real-time computerized evaluation of the pit patterns of colorectal lesions. If rapid, accurate differentiation between neoplastic and non-neoplastic polyps can be made by magnification endoscopy with computer-aided diagnosis, this technique could 180 reduce the number of polypectomies required and also reduce complications.

      We succeeded in developing a computerized system for automated recognition of regular pit patterns on magnified endoscopy images. With its limitations, our system is perhaps best characterized as semi-automated but a step toward development of a fully automated system to assist in the diagnosis of colorectal lesions. We anticipate development of a fully

185 automated system that will recognize both regular and irregular pit patterns and will meet the rigors of blinded prospective evaluation comparing the results of computerized analysis against pathologic classification as the gold standard.

REFERENCES

190

1.   Tanaka S, Kaltenbach T, Chayama K, et al. High-magnification colonoscopy (with videos). Gastrointest Endosc 2006;64:604-13.

2.   Fu KI, Sano Y, Fujii T, et al. Chromoendoscopy using indigo carmine dye spraying with magnifying observation is the most reliable method for differential diagnosis between non-neoplastic and

195   neoplastic colorectal lesions: a prospective study. Endoscopy 2004;36:1089-93.

3.   Hirota M, Tamaki T, Kaneda K, et al. Feature extraction from images of endoscopic large intestine. Proceedings of FCV2008; The 14th Korea-Japan Joint Workshop on Frontiers of Computer Vision 2008;01:94-9.

200

Figure Legends

Figure 1. Classification of pit patterns of colorectal lesions.

Figure 2. Pits are outlined on the magnified endoscopic images with the use of our custom HuPAS ver.1.3 software. a: Observation and recording of the stained (crystal violet) image at maximum optical magnification (x70). b: A region of interest (ROI) measuring 250×250 pixels is cut out for analysis. c: Example of a pit region automatically extracted by HuPAS.

Figure 3. Extracted images of pit outlines within the region of interest. a: Original image of a pit region automatically extracted by HuPAS. b: The image generated by HuPAS required some Adobe Photoshop editing. The separate pit images (sI, sII, sIIIs, sIIIL, sIV) (arrows) were classified according to the Kudo and Tsuruta criteria.

Figure 4. Diagram of the six shape descriptors used for quantitative analysis of pit patterns. (1) area, (2) perimeter, (3) major axis of the best fit ellipse, (4) minor axis of the best fit ellipse, (5) circularity: $4\pi(area/perimeter^2)$, wherein1.0 = precise circle that becomes an elongated polygon as it approaches 0.0, and (6) Feret's diameter, which is the longest distance between any two points within the selected frame.

Type I		Round pit (normal pit)		non-neoplastic
Type II		Asteroid pit		non-neoplastic
Type IIIs		Tubular or round pit that is smaller than the normal pit (Type I)		neoplastic
Type IIIL		Tubular or round pit that is larger than the normal pit (Type I)		neoplastic
Type IV		Dendritic or gyrus-like pit		neoplastic
Type VI		Irregular arrangement and sizes of IIIL, IIIs, IV type pit pattern		neoplastic
Type VN		Loss or decrease of pits with an amorphous structure		neoplastic
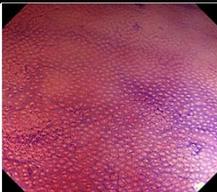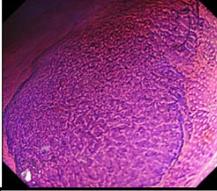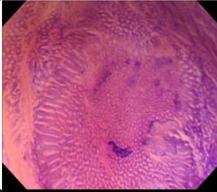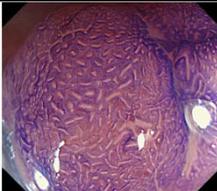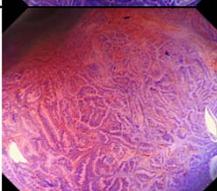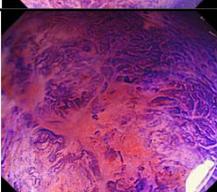
Figure 1. Classification of pit patterns of colorectal lesions. Figure 1, Takemura et al

Figure 2

a

b

c

Figure 3

sType I

sType II

sType IIIL

sType IV

a   b
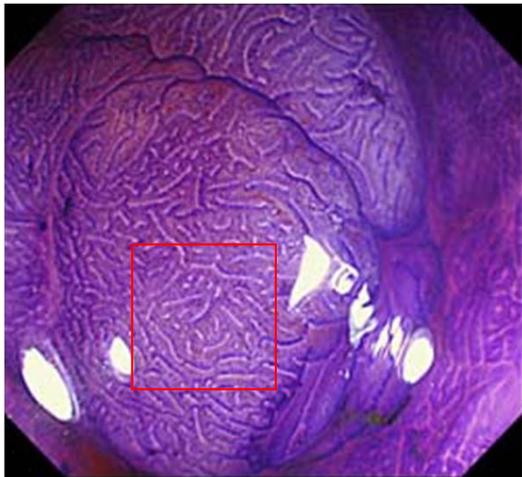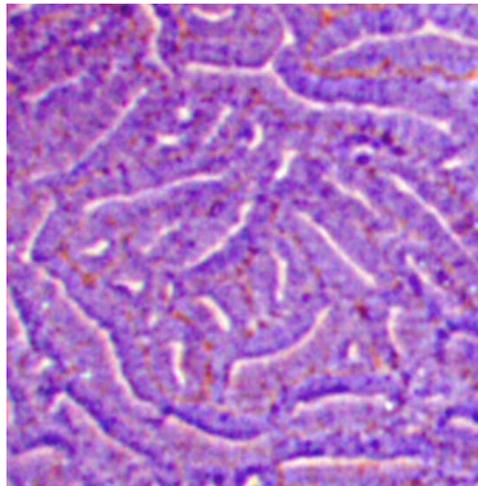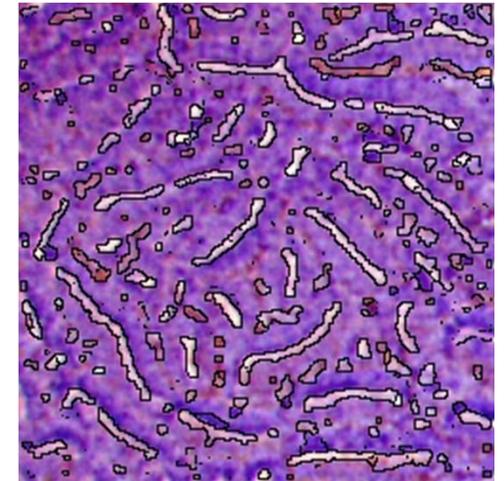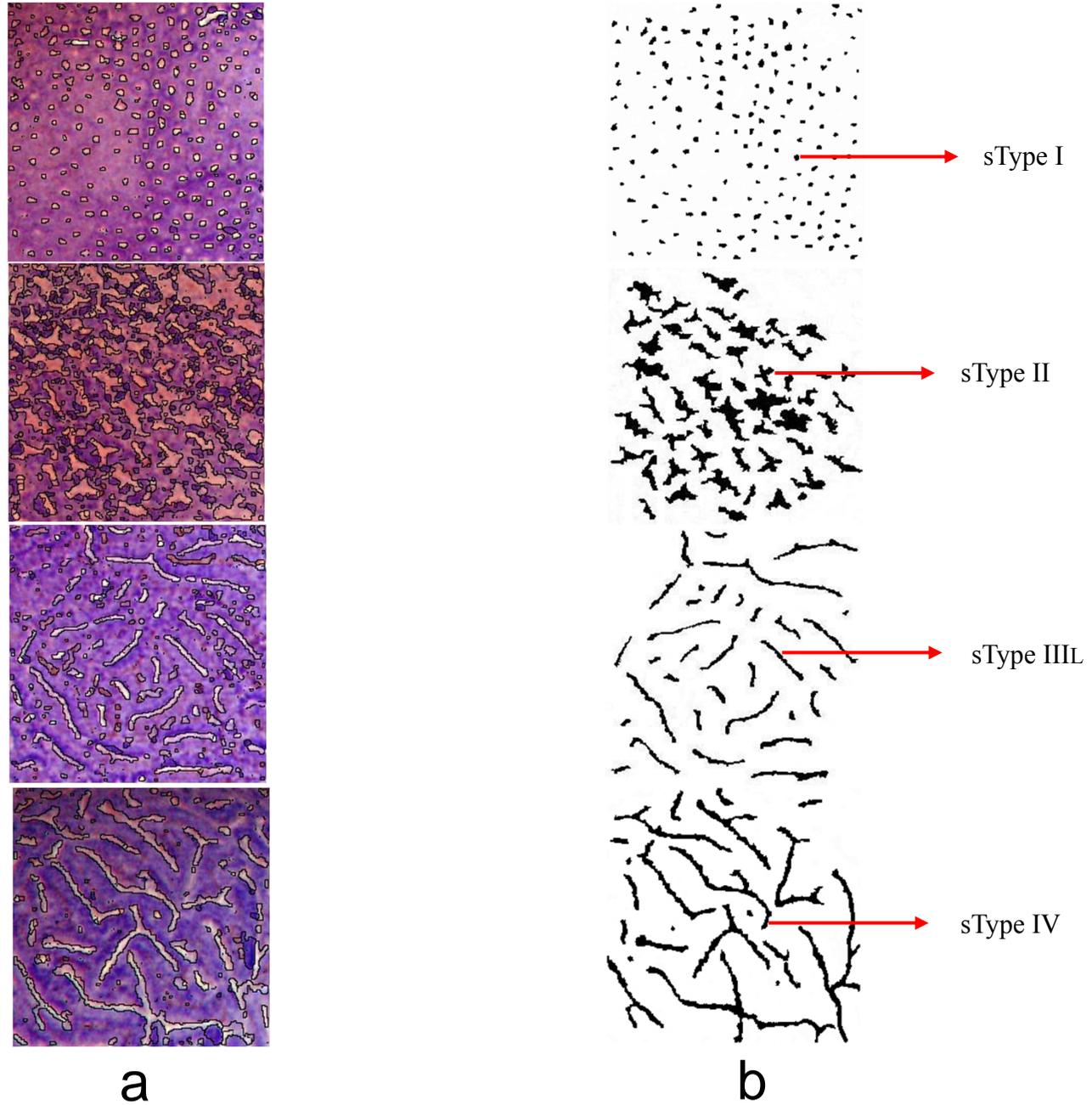
Figure 4

Quantitative characteristics for six examined items

- Area
- Perimeter
- Major Fit Ellipse
- Minor Fit Ellipse
- Circularity: $4\pi(\text{area}/\text{perimeter}^2)$ (1.0=precise circle; becomes an elongated polygon as it approaches 0.0)
- Feret's Diameter: longest distance between any two points within the selected frame.

Table 1. Quantitative analysis of regular pit patterns

Pit pattern	Number of pits	Area	Perimeter	Major fit ellipse	Minor fit ellipse	Circularity	Feret's diameter
Type sI	453	44.0±24.5	26.1±8.3	9.3±2.9	5.7±1.6	0.78±0.12	10.4±3.1
Type sII	210	195±134	101.0±58.3	23.4±8.0	9.9±4.5	0.29±0.15	32.1±13.6
Type sIIIL	268	226±155	118.7±65.8	40.2±18.5	6.7±2.3	0.24±0.13	49.2±26.5
Type sIIIS	26	5.4±3.4	7.93±3.00	3.16±1.09	2.03±0.67	0.93±0.12	3.61±1.14
Type sIV	268	749±455	354±203	56.3±18.9	17.4±8.9	0.097±0.067	109.1±42.5
		P＜0.0001	P＜0.0001	P＜0.0001	P＜0.0001	P＜0.0001	P＜0.0001

Data are mean±SD.

P values were obtained by Kruskal-Wallis test, which was used to analyze between-pattern differences in the values of each of the six features.

Table 2. Performance of the semi-automated CAD algorithm for pit pattern classification of colorectal lesions

Endoscopic diagnosis	Classification using the CAD software				
	Type I	Type II	Type III$_L$	Type IV	Total
Type I	32 (100)	0	0	0	32/32 (100)
Type II	0	43 (100)	0	0	43/43 (100)
Type III$_L$	0	0	28 (96.6)	1 (3.4)	28/29 (96.6)
Type IV	0	0	1 (3.3)	29 (96.7)	29/30 (96.7)

Data are number (percentage) of lesions.

Overall accuracy: 132/134 (98.5%)

CAD : computer-aided diagnosis