2010

Automated measurement of quality of mucosa inspection for colonoscopy

Xuemin Liu
Iowa State University

Automated measurement of quality of mucosa inspection for colonoscopy

by

Xuemin Liu

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Co-majors: Computer Science; Human Computer Interaction

Program of Study Committee:
Wallapak Tavanapong, Co-major Professor
Johnny Wong, Co-major Professor
Leslie Miller

Iowa State University

Ames, Iowa

2010

TABLE OF CONTENTS

ABSTRACT

With 655,000 deaths worldwide per year, colorectal cancer it is the third most common form of cancer and the third leading cause of cancer-related death in the Western world. Colonoscopy is currently the preferred screening modality for prevention of colorectal cancer, in which a tiny camera is inserted into the colon to look for early signs of colorectal cancer. A recent systematic review calculated a 22% miss rate for all colonoscopic neoplasia, being 2.1% for advanced lesions. This could be attributed to factors such as inadequate endoscope withdrawal time, poor range of motion of the endoscope, and general endoscopist experience. Therefore the demand for quality control for colonoscopic procedures is increasing, and many researchers have been taking efforts in this area. In this paper, we first presented a novel technique - Colon Center Axis Determination Technique for Non-dark Lumen Images, and the performance evaluation result demonstrates that this technique enables a more accurate view mode classification for all kind of images. Secondly, we proposed two novel approaches to help objectively measure the quality of colonoscopy. A set of objective metrics has been proposed, and preliminary analysis result shows the spiral number during whole procedure/withdrawal phase has a relatively strong positive association with the ground truth circumferential inspection score. The other approach is using association rule mining knowledge to determine patterns of colon inspection. The preliminary result demonstrates that endoscopists with good and relatively poor inspection skill have different inspection patterns, and thus using patterns to assess colonoscopy quality would be anther feasible and promising method.

1. INTRODUCTION

Colorectal cancer, also called colon cancer or large bowel cancer, includes cancerous growths in the colon, rectum and appendix. With 655,000 deaths worldwide per year, it is the third most common form of cancer and the third leading cause of cancer-related deaths in the Western world [1]. Colorectal cancers are malignant tumors that develop in the colon and rectum. The survival rate is higher if the cancer is found and treated early before metastasis to lymph nodes or other organs occurs. Colonoscopy is currently the preferred screening modality for prevention of colorectal cancer. The procedure can help doctors diagnose unexplained changes in bowel habits, abdominal pain, bleeding from the anus, and weight loss. Over 14 millions colonoscopic procedures are performed annually in the US alone [2], and this number is rising.

During colonoscopy, an endoscope, which is a flexible tube having a tiny video camera with a wide-angle lens at the tip, is inserted into the rectum via the anus. Then it is advanced gradually into the cecum (the most proximal part of the colon) or the terminal ileum. This phase is called an insertion phase. The withdrawal phase follows, where the endoscope is gradually withdrawn. Careful mucosa inspection and diagnostic or therapeutic interventions such as biopsy, polyp removal, etc., are performed during the withdrawal phase. During the whole procedure, the camera generates a video signal of the interior of the human colon, which is displayed on a monitor for real-time analysis by the physician.

Although colonoscopy has made a significant contribution to the decline in the number of colorectal cancer-related deaths, a miss rate for polypoid lesions occur with colonoscopy, presumably resulting in a reduced preventive efficacy of the procedure [3, 4]. A recent systematic review calculated a 22% miss rate for all neoplasia, being 2.1% for advanced lesions [5]. Moreover, virtual colonoscopy studies estimated a colonoscopy miss rate ranging from 12% to 17% [6].

For this reason, issues regarding the quality of colonoscopy, such as cecum intubation rate and bowel cleansing, have been extensively addressed in the literature [7, 8]. A further issue, namely, the withdrawal time, has been recently associated with an increased detection of advanced neoplasia, with a withdrawal period lasting at least 6 min being recommended [9]. A number of indirect markers of quality have been proposed in 2006 by American Society of Gastroenterology and American College of Gastroenterology, including duration of the withdrawal phase and the average number of polyps detected per screening colonoscopy [10]. Post-procedure manual analysis of procedure quality is both time-consuming and subjective since the domain expert needs to review the entire video of a procedure. This limitation motivates researchers to develop automated methods that derive various objective quality metrics that can be compared among endoscopists [11-15]. Other automated analyses included polyp detection [16], appendiceal orifice image detection [17], 3D reconstruction of the colon surface for surgical planning [18, 19], image-guided automated colonoscopy [20], and 3D reconstruction of a colon structure [DongHo] from a 2D colonoscopy image.

Oh et al. presented the first set of objective quality metrics: time spent during the insertion phase, time spent during the withdrawal phase, etc. They proposed a number of algorithms that analyze a colonoscopy video to generate these objective metrics [11]. Liu et al. proposed another metric aiming to specifically measure quality of inspection of the colon mucosa (i.e., looking at off-axial or lateral wall mucosa or inspection behind mucosal folds) and a technique that computes the metric [15] from analysis of a colonoscopy video. The technique first determines the view mode of a clear colonoscopic image (either *lumen view* for an image which the colon lumen appears in the image or *wall view* for the others). The view mode detection is based on the application of decision tree and support vector machine classifiers on a set of features that captures darkness, shape, and location of the colon lumen. Next, the colon axis center is estimated and the metric is derived from the location of the colon axis center. Their technique correctly classifies lumen view images with relatively dark lumen but fails for the other lumen view images. Furthermore, the authors did not show validation

results of the proposed metric against the ground truth score given by the domain experts.

To address the drawbacks of Liu's work, in this thesis, we made the following contributions. First, we proposed a new method utilizing parallel curves of nested colon folds to estimate the location of the colon axis center. We apply the technique on images classified as wall images by Liu's technique to identify additional non-dark lumen images. We call our new technique *"PCurve"* to signify the utilization of parallel curves. Our performance evaluation demonstrates that *PCurve* improves sensitivity of view mode detection. Second, we proposed new metrics to indicate quality of mucosa inspection and algorithms that compute these metrics. Third, we conducted correlation analysis of the new metrics and Liu's proposed metric against the mucosa visualization ground truth score given by four domain experts. Our correlation analysis shows a relatively strong positive correlation between the ground truth score and one of our metrics – the spiral number. As a result, this metric is promising for indicating the quality of mucosa inspection of a colonoscopic procedure. Forth, we investigated another approach using association rule mining to discover patterns of good and poor quality mucosa inspection. To the best of our knowledge, there are no existing methods in the literature. The preliminary result demonstrates different inspection patterns between endoscopists with good and relatively poor inspection skills.

The remainder of this thesis is organized as follows. In Chapter 2, related work is provided. Chapter 3 presents PCurve and evaluation results in detail. Chapter 4 describes our proposed metrics and the result of correlation study of these metrics and the manual scores. We present the association rule mining method and its preliminary result in Chapter 5. Lastly, we conclude this paper and provide the description of our future work in Chapter 6.

2. RELATED WORK

2.1 View Mode and View Direction

To estimate the view mode for a given colonoscopic image, we rely on the presence of the lumen in the image. In the work [21], the researchers defined a *lumen view* as a clear (non-blurred) frame in which the distant colon lumen is seen. That means that the line of view is along the longitudinal axis of the colon proximal to the endoscope tip. If the distant lumen is central in the image, the view is axial; if the distant lumen is in the periphery of the image, the view is off-axial. A lumen view provides a global inspection and includes more distant examination in which more than one side of the colonic wall is seen. A clear frame without the distant colon lumen is called a *wall view*. The wall view most often occurs as a result of a close inspection of the lateral colon wall. Both lumen (global inspection) and wall (close inspection) views are important, and thus should be present in a good colon examination. Once the correct classification of the view mode is made, we can derive different quality metrics such as (i) the ratio of close inspection to global inspection [21], (ii) the duration of a sequence of wall views or a sequence of lumen views, or (iii) the interleaving pattern between close inspection and global inspection. Later, Liu et al. defined *view direction* as a direction toward the distant, most proximal colon lumen [13]. They suggested that the side of the colon that is inspected is estimated to be 180 degrees opposite from the view direction of the colon lumen.

2.1.1 View Mode Detection Technique

The most related research efforts are in the area of microrobotic endoscopy [22, 23, 24, 25]. These efforts focus on the following problem: given an endoscopic image with the lumen, identify the lumen boundary. Khan [23] proposed to use an N-level quadtree-based pyramid structure to find the most homogenous large dark region. A region-growing scheme [22] based on an inter-pixel grayscale

difference was proposed to automatically detect luminal borders. Kumar et al. [24] proposed a global thresholding technique with differential region-growing to segment the lumen region. Tian et al. proposed APT-Iris that utilizes the relative darkness of the lumen [25]. The work in [22, 23, 24, 25] do not discuss how to determine whether the lumen is seen in the image or not. Unlike the aforementioned techniques, “Grayscale Shape-based View Mode Classification (GSVM)” technique [21] determines whether the lumen is seen in the image or not. The technique employs the relative darkness of the lumen coupled with the following facts. First, more than one bilateral convex colon wall is seen around the colon lumen. Second, the intensity difference between consecutive colon walls is small.

Later, Liu et al. stated several drawbacks of GSVM including not utilizing useful chrominance information of the pixels and misclassification due to the adaptive threshold methods on pixel intensity alone [15]. They proposed View Mode Detection Technique which consists of three major steps: first, lumen pixel classification uses a decision-tree classifier to classify each pixel in an image into either a *lumen pixel* or a *wall pixel*, and outputs an intermediate image called “red-green image”; second, feature extraction extracts seven images features including number of lumen pixels identified by the lumen pixel classifier, area of the largest foreground object in a reconstructed image, distance of the centroid of the largest cluster from the image boundary, etc., from the intermediate and original images; third, image classification uses a decision-tree classifier to determine whether the image is a wall view or lumen view. According to the evaluation result, the proposed algorithms perform well. However, it is worth noticing that this technique would misclassify lumen images where there is no dark lumen region as wall images. Later, a new technique solving this problem is proposed in this paper.

2.1.2 View Direction Annotation

In [13], View Direction Annotation has been proposed. Each of the detected lumen images is annotated with an arrow indicating the corresponding view direction. One preferable arrow representation method was discussed in [13]: the arrow is drawn from the center of the image toward the centroid of the distant lumen. Domain experts indicate that the arrow head should point approximately to the lumen, and they recommend that the position of the arrow head should consider both the most distant part (darkest lumen area) and the first few folds close to the lumen. Therefore, the position of the arrow head is calculated as the weighted centroid of the detected lumen pixels.

The View Direction Annotation is very useful as it plays an important role for determining the quality of the colon wall examination.

2.2 Objective Measurement of Colonoscopic Quality

As mentioned in the Introduction section, there is a significant miss rate for the detection of even large polyps and cancers. The miss rate varies among endoscopists and is closely related to their skill sets. The reasons for development of colorectal cancer not prevented by colonoscopy include truly missed lesions and failure to recognize, to adequately treat or to arrange appropriated follow-up for advanced adenomas [35]. Even though the demand for quality control for colonoscopic procedures is increasing, neither a manual protocol nor an automated system is available to produce detailed measurements of quality of colonoscopy during routine clinical practice. In general, the global quality of a colonoscopic procedure can be evaluated in terms of time of the withdrawal phase and thoroughness of inspection of the colon mucosa.

Quadrant Histogram Coverage (QCH) Technique

Experienced endoscopists have stated that as much as possible of the mucosa (circumferential

inspection of all sides or 360 degrees inspection) of the colon wall should be examined throughout the withdrawal phase. Liu et al. are the first to propose a method called Quadrant Coverage Histogram (QCH) that provides a quantitative measure of this desirable inspection pattern [13]. QCH counts the number of inspected quadrants of the colon wall in a given duration (time window) averaged over the span of inspection time (whole procedure or withdrawal phase only). Given a time window, the method works as follows.

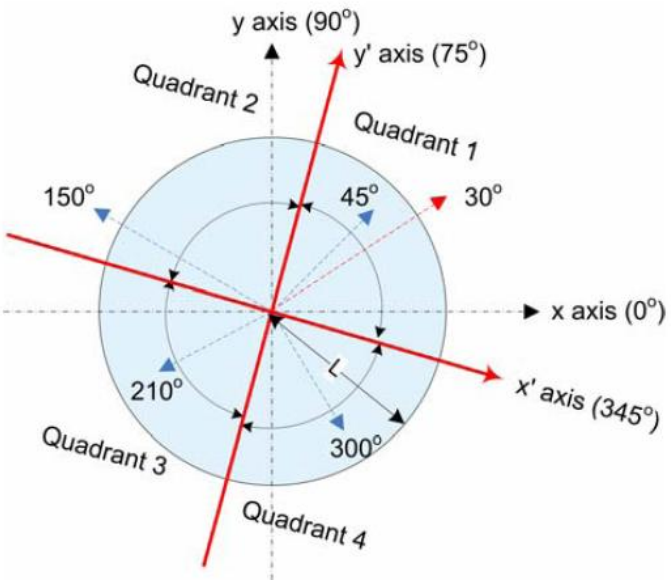


Fig. 1. Illustration for the calculation of QCH. The length of time window is 5 second. Red-dashed arrow points to the view direction of the frame at the center of the time window and blue-dashed arrows indicate the view direction of frames within the time window of the current frame. Two red-solid axes form four new quadrants. [Courtesy of Liu et al.].

First, a view direction for each lumen view is estimated according to the method described in section 2.1.2. A view direction arrow consists of two important parts: the angle and the length of the arrow. The arrow length should be at least a *laterality threshold* for the corresponding view angle to have enough coverage to be accounted for in QCH. Taking the angle of the view direction of the frame at the center of the time window as the reference angle, we define two perpendicular axes forming four new quadrants. Each of the two perpendicular axes is 45^{o} away from the reference angle (x' and y'

axes in Fig. 1). Second, a Quadrant Coverage Histogram (QCH) that maps view directions of the lumen views in the time window to the number of quadrants (1-4) of the colon wall that have been inspected in that time window. Third, they compute an average quadrant coverage score over N lumen views from QCH, which can be used as a key metric to evaluate the quality of circumferential inspection of colonoscopy. Liu et al. proposed to use plots, which show the average coverage score during the withdrawal phase given different values of the length (T) of the time window and laterality threshold L, to obtain the inspection time to complete all the four quadrants of colon wall. Fig. 2 is an example plot, where the endoscopist spent around 10 and 40 seconds to inspect all the quadrants with L=0 and L=0.66, respectively.

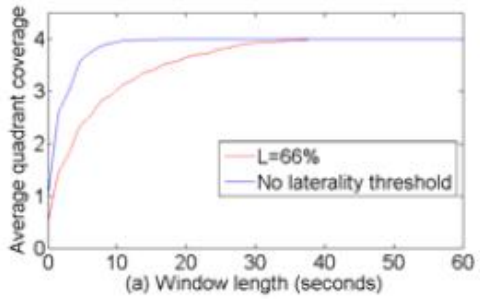
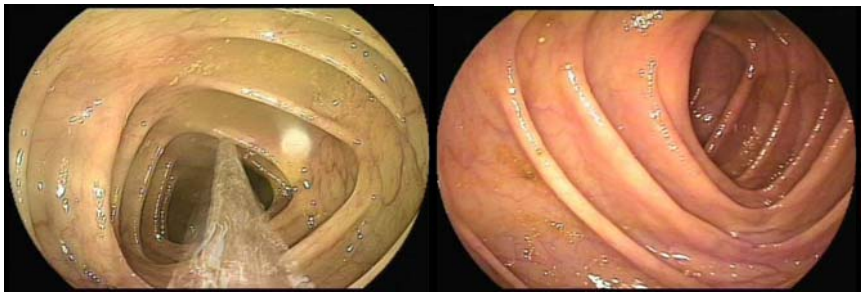


Fig. 2 Plot of average quadrant coverage scores of one video, red curves show the scores for the laterality threshold L=66% of the maximum radius; blue curves show the scores when the laterality threshold is not considered.

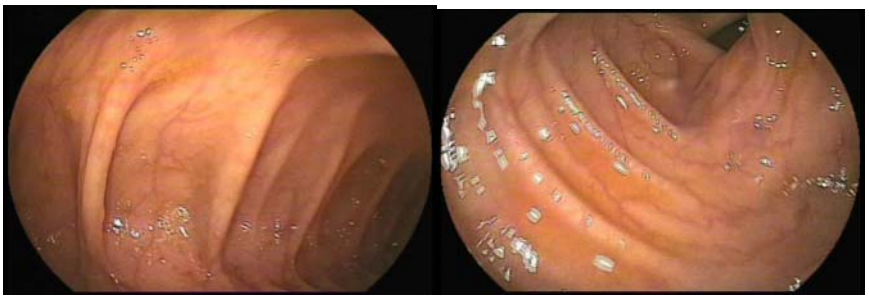
If all the four quadrants are inspected in a short time period, the average coverage score would approach the value of four quickly. The implication is that the endoscopist may on average inspect all quadrants of one colon fold. This pattern is more desirable than when more time is taken to complete the four quadrants, which may imply that the inspection covers more than one fold. Hence, some areas of some folds are not seen. For performance evaluation, they ran their QCH technique on three raw videos and the results shows that the endoscopist completed inspection of the four quadrants on average every 20 seconds.

3. PCURVE TECHNIQUE FOR NON-DARK LUMEN IMAGES

Liu's method misclassifies true non-dark lumen images as wall images. The number of misclassified non-dark lumen images varies depending on how the endoscopist positions the camera. Adding non-dark lumen pixels for training of the decision tree classifier improves sensitivity but reducing specificity. We, therefore, propose the PCurve technique that utilizes quasi-parallel edges/curves of nested colon folds to estimate the location of the colon center axis. We use the term "quasi" to indicate that these curves are not strictly parallel due to the complex nature of the colon and colon distortion. Based on our observation, we categorize non-dark lumen images into two classes: 1) "M-Class" having at least two groups of quasi-parallel folds as in Fig. 3(a) and (b) and 2) "S-Class" having a single group of quasi-parallel folds as in Fig. 3(c) and (d). For M-Class, we can use the intersection of the perpendicular bisectors of any two groups of the quasi-parallel folds as the location of the colon center axis. For S-Class, a different method is needed. We apply PCurve only on images classified as wall images by Liu's method to recall true non-dark lumen images missed by our previous method (see the flow chart in figure 4). The detected location of the colon center axis is used to derive quality metrics described in Chapter 4.



(a) **(b)**



(c) **(d)**

Fig 3. Non-dark lumen image examples: images on first row have two groups of parallel folds; images on second row have one group of parallel folds

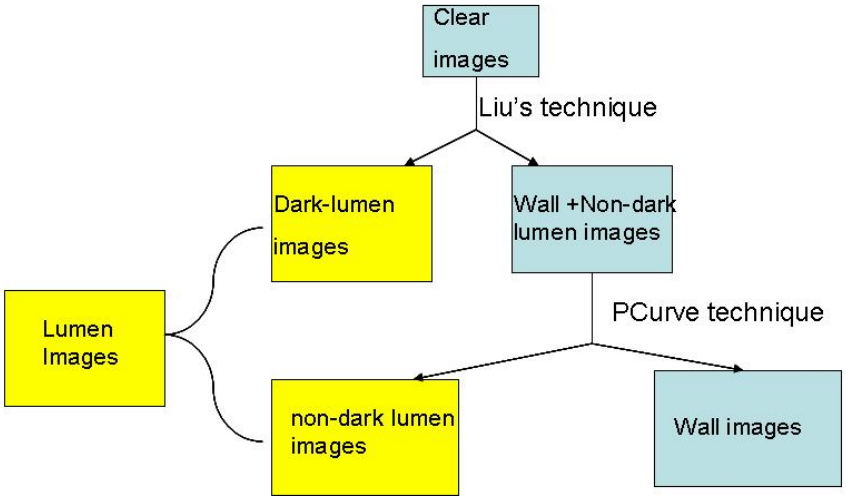


Fig. 4 Flow chart illustrating the composition of colonoscopy images and the result of Liu's and PCurve techniques

3.1 PROPOSED APPROACH

PCurve consists of three major steps: 1) pre-processing, 2) determination whether an image is a wall image or an S-class lumen image or an M-Class lumen image, and 3) estimation of the colon center axis location. For non-dark lumen images of each class, we accordingly develop 'PCurve for M-Class' and 'PCurve for S-Class' to determine the colon center axis location. Various thresholds in the technique are derived from training data and are summarized in Table 2.

1) Pre-processing discards an edge with the number of edge pixels outside a pre-determined range observed for most colon fold edges. Next, we cut the remaining edges having corners (e.g., edges of

a triangle in Fig. 5 (a)) at each corner into smaller edge segments to facilitate detection of quasi-parallel folds. Last, we remove small branches of these edges.

2) We calculate the number of groups of parallel edges in the image to be between zero and two inclusive. The first step is to find an edge pair in which the edges in the pair are quasi-parallel. Checking all possible combinations of any two edges whether they are quasi-parallel is time consuming. To reduce the number of combinations, we first compute the coarse orientation of each edge by fitting a line on the edge using the built-in polyfit Matlab function. Next, we calculate the angle value of the fitted line using the vertical axis as the reference. We use K-means clustering algorithm with K=2 to divide the edges into two groups based on the angle values. In each group, we perform a detailed check of all possible pair combinations whether edges in each pair are quasi-parallel to each other as follows.

Given a pair of edges, we project each edge on to each other by locating on the other edge the points closest (in the sense of Euclidean distance) to the endpoints of the edge being considered. For instance, consider the edge pair AB and CD (Fig. 5). The closest points of A and B on edge CD are A' and B', respectively. The closest points of C and D on edge AB are C' and D', respectively. Hereafter, we use the projected edges (e.g., A'B' and C'D') instead of the original edges to decide whether the two original edges in the pair are quasi-parallel. We discard short projected edges (the edge length smaller than a threshold L). We cut each remaining projected edge pair (e.g., A'B' and C'D') into an equal number of small edge segments such that those of the same projection are of equal length. Fig. 5(d) shows a number of small edge segment pairs between the two projected edges. Next, we determine whether edge segments in each of these pairs are quasi-parallel, i.e., they have the angle difference within a degree threshold A. If at least T percent of all pairs is quasi-parallel, we claim the entire edge pair quasi-parallel (e.g., AB and CD are quasi-parallel). The number of edge segments is a tradeoff between accuracy and execution time. We chose the number

of edge segments so that every edge segment has a length of 5-10 pixels. We selected L=40, A=20, and T=70 based on experiments with our training dataset.

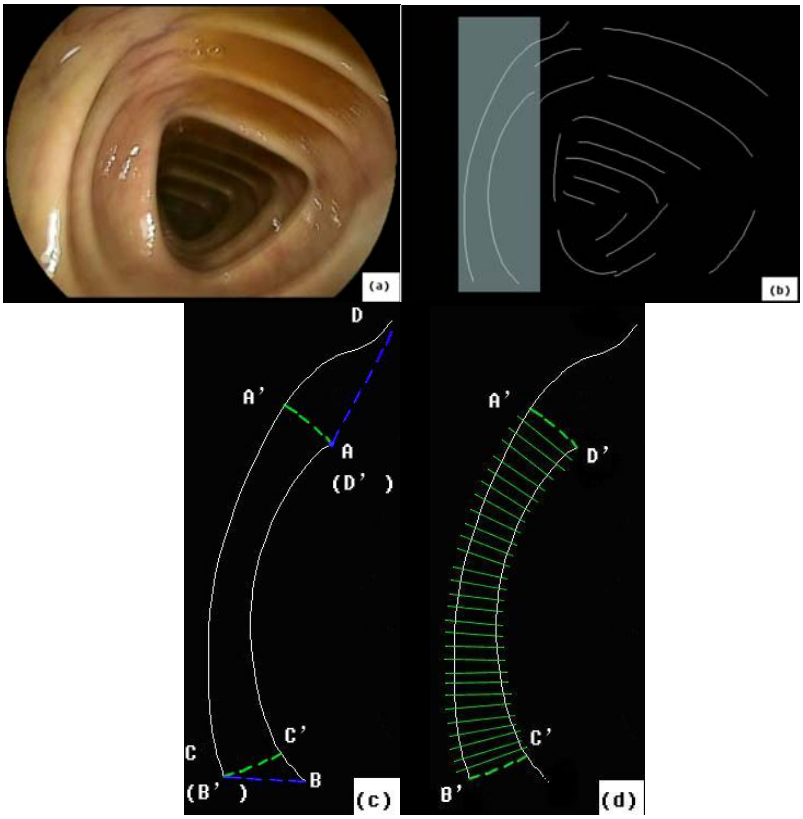


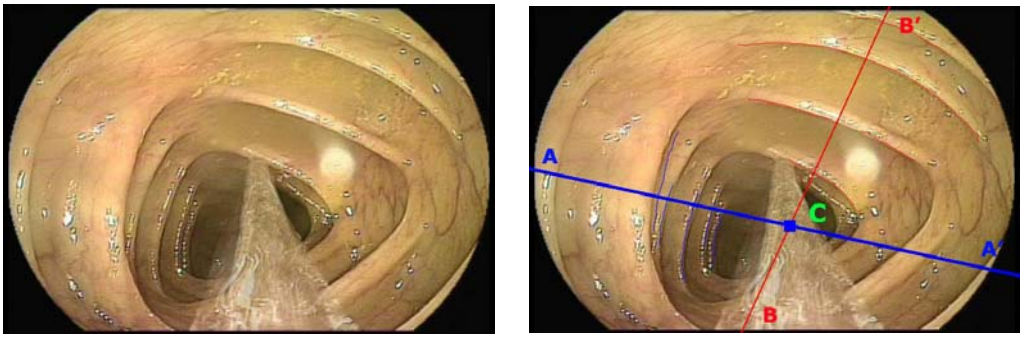
Fig. 5 Quasi-parallel edge detection (a) a non-dark colon lumen image; (b) corresponding qualified edges; (c) a pair of candidate quasi-parallel curves marked by blue semi-transparent rectangle in b; (d) the pair of candidate edges cut into pairs of small edge segments between green bars

3) Based on our observation, some wall images have only one group of two parallel edges (e.g., images in the cecum around the appendiceal orifice or two parallel edges of the same very protruding fold). Lumen images typically have more than two parallel edges in a group. Therefore, if we detect only one group of two parallel edges in an image, we classify the image as a wall image. Otherwise, the image is considered a lumen image and is further classified. Then we define the angle of a group of quasi-parallel curves as the average of the angle values of all the curves in the group. There are two angle differences between two different groups of quasi-parallel curves, and their summation is equal to 180 degrees. If the smaller angle difference is less than 35 degrees, we

consider them in the same group. Finally, we assign lumen images with one group of parallel curves into S-Class and images with at least two groups of parallel curves into M-Class.

3.1.1 PCurve for M-Class

When we detect two or more than two groups of quasi-parallel folds in an image, we apply PCurve technique for M-Class to find out the position of its approximate colon center axis. The basic idea is: the intersection point of two perpendicular bisectors of two groups of quasi-parallel folds would be in the close neighborhood of colon center axis. Let's consider the following example for more detailed information. In Fig 6a, there is an example of non-dark lumen image with three groups of quasi-parallel folds, but one group of quasi-parallel folds is occluded by the water used for cleansing colon inner wall. Therefore, two groups of quasi-parallel folds are detected (see Fig. 6b), and one is marked by a set of blue curves along quasi-parallel edges in the first group (group I) on the left side, and the other is marked by a set of red curves along quasi-parallel edges in the second group (group II) on the top-right side. The blue line AA' and the red line BB' are the perpendicular bisectors for curves in group I and II, respectively. The intersection point C of these two lines lies correctly in the neighborhood of colon center axis.



(a) **(b)**

Fig 6. (a) Non-dark lumen image with three groups of quasi-parallel folds; (b) colon center axis determination technique intermediate and final result

To find the perpendicular bisector line, we need to know at least the slope (S) of the line and the coordinate of one point (P) on the line. Given slope values (s1, s1, …, sn) of n quasi-parallel curves in a group, we determine the slope S of the perpendicular bisector line for a group of quasi-parallel curves using Equation (1).

$$S = -1/\tan(\sum_{i=1}^{n} \arctan(s_i)/n) \quad (1)$$

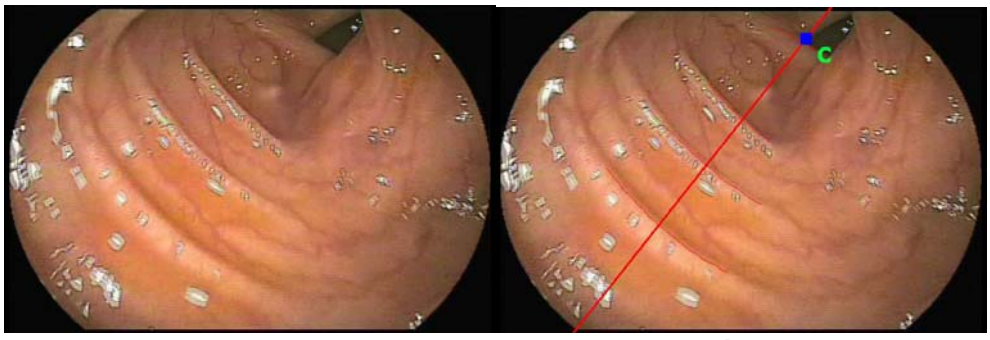
The coordinate of the point P on the perpendicular bisector is chosen as follows. We first draw a line (line L) to be

quasi-parallel to all the curves in its group (with slope value equal to -1/S) and across the center of the image. Then, we project the center point of each curve in the group onto L. At last, given the coordinates of all projected center points (xi, yi) on L and the length of the corresponding curve li, where i=1, 2, …, n, we calculate the coordinate (X,Y) of P as the weighted center of the projected points using Equation (2). The intersection point of the two perpendicular bisectors is marked as the location of the colon center axis.

$$X = \sum_{i=1}^{n} x_i \bullet l_i / \sum_{i=1}^{n} l_i \ , \quad Y = \sum_{i=1}^{n} y_i \bullet l_i / \sum_{i=1}^{n} l_i \qquad (2)$$

3.1.2 PCurve for S-Class

When what we can detect is only one group of quasi-parallel curves in a non-dark lumen image, we cannot find any intersection point like what we do in 3.1.2. Therefore, we resort to intensity change along the perpendicular bisector of quasi-parallel curves. Fortunately, we find some rules based on the two following observations: 1) there is a significant intensity change before and after entering the neighborhood of colon center axis; 2) the intensity of the spot around the colon center axis is lower than the average intensity of the image. With the help of such rules, we can find the approximate colon center axis for such kind of images. In Fig. 7a, you can see a non-dark lumen image with one group of quasi-parallel curves. Fig. 7b shows you the intermediate and final result of colon center axis determination technique: one group of quasi-parallel folds is detected, and it is marked by a set of red curves along quasi-parallel edges. The thick red line is the perpendicular bisector for curves, and the blue dot marked by letter C is the detected colon center axis.

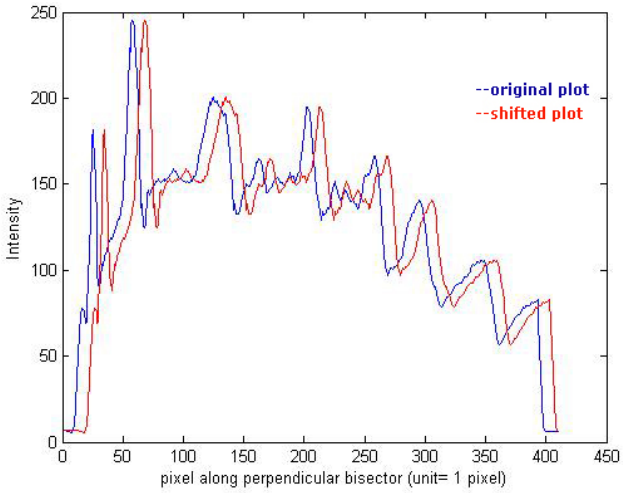


(a) **(b)**

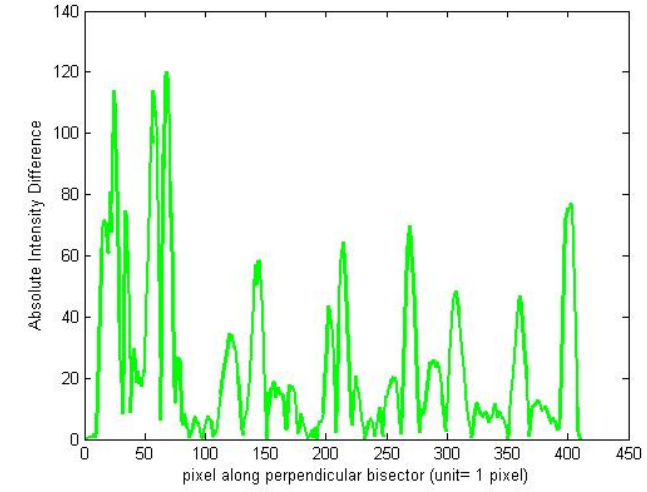
Fig 7. (a) Non-dark lumen image with one group of quasi-parallel folds; (b) colon center axis determination technique intermediate and final result

Now let's consider the algorithm of this technique. After we determine the perpendicular bisector, we can obtain

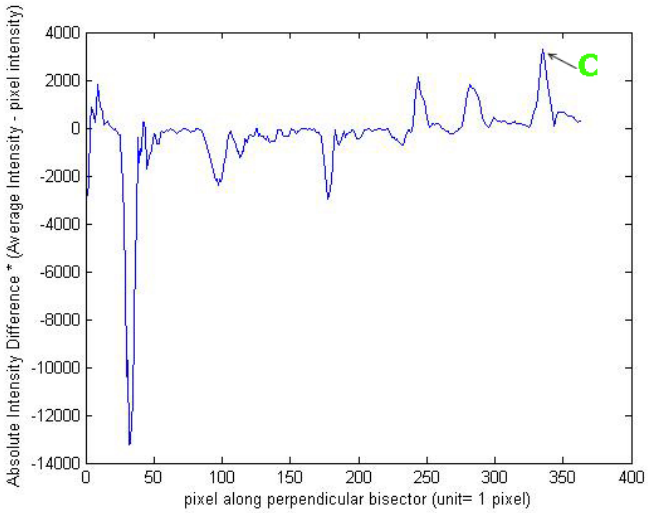
an array of pixel intensity values along the perpendicular bisector in the corresponding gray-scale image. Here, for finding the relatively larger intensity change, we first plot the intensity array, and then generate another one by shifting the plot by a distance of ten pixels, and last make the absolute difference I_{diff} between the original and shifted plots. The larger the absolute difference value I_{diff} is, the larger the intensity change is. For the image in Fig 7, the corresponding original and shifted plots along the thick red line (from bottom-left corner to top-right corner) is shown in Fig. 8a, and the absolute intensity difference is plotted in Fig. 8b. We can make use of such plots to reflect the first observation, and find out several candidate spots to focus on in later step. In the consideration of second observation, we calculate the average intensity value of the intensity array, and name it as I, and then make difference between average intensity I and pixel intensity I_P for each pixel P. If the pixel intensity value is lower than I, i.e. the pixel looks darker than the average brightness of the image, the difference gives us a positive value; whereas the difference is a negative value if the pixel intensity value is larger than I. After this step, we draw another plot which shows the values of $I_{diff} * (I-I_P)$ for each pixel along the perpendicular bisector, see Fig. 8c. $I_{diff} * (I-I_P)$ combines the impact of both intensity change and actual intensity.



(a)



(b)



(c)

Fig 8. (a) the original and shifted intensity plot; (b) the plot of absolute intensity difference Idiff; (c) the plot of Idiff * (I-IP)

Now what we are interested is the pixel with Idiff * (I-IP) value above zero and relatively larger than others'. For determining candidate spots, we set a threshold which is a quarter of maximum value of Idiff * (I-IP). Any pixel with the Idiff * (I-IP) value bigger than this threshold may become our candidate spots. Besides, we find all candidate pixels satisfy all the following criteria: 1) its intensity is lower than some threshold (190); 2) the corresponding position of the pixel is in the area with endoscope signals (no black corners as in Fig. 7(a)). We simply use 18 pixels from the image border in either *x* or *y* direction as an estimate. Among all these candidates, we select the pixel with the lowest intensity as the colon center axis. Therefore, in the above example the spot C in Fig. 8c is our detected approximate colon center axis.

3.2 Experimental Results

Videos in our experiments were selected from routine colonoscopy screening performed by several endoscopists using Fujinon endoscopes. No patient identifiable information is included in these videos. One video file contains a single colonoscopic procedure. The video format is MPEG-2 with the image resolution of 720x480 pixels. We created a test bed of close to 3,000 images selected from eight videos listed in Table 1. For test sets I and III, images were extracted at one frame per second. For test set II, images were extracted at five frames per second. Then we use our blurry frame detection software [26] to obtain only clear images and applied Liu's method [15] on them to obtain the test bed. We selected only S-Class lumen images to create test set I. Test set II includes only M-Class lumen images. The reason for using a higher image extraction rate for test set II is because each of the videos does not have many M-Class images. Test set III consists of only wall view images.

TABLE 1: DETAILS OF GROUND TRUTH IMAGE DATASETS

Video ID	0293	0295	0297	0300	0304	0305	0307	0315	Total
Testing Set I	172	47	74	131	76	58	63	332	953
Testing Set II	143	-	143	569	-	-	-	143	998
Testing Set III	-	196	-	-	489	315	-	-	1000

We implemented PCurve in Matlab. Major functions are listed in the first column of table 2. Threshold and parameter values for the software were chosen experimentally from a separate training data set that does not overlap with the test data sets. We summarize all the parameters that used in our technique and the purpose of their usage in the second and third columns of table 2. The two parameters brightnessThresh and cannyVector (with superscript number [2] in the table) are likely to be retrained for different endoscope brands - Olympus, as the images captured by Olympus camera have different light condition and brightness.

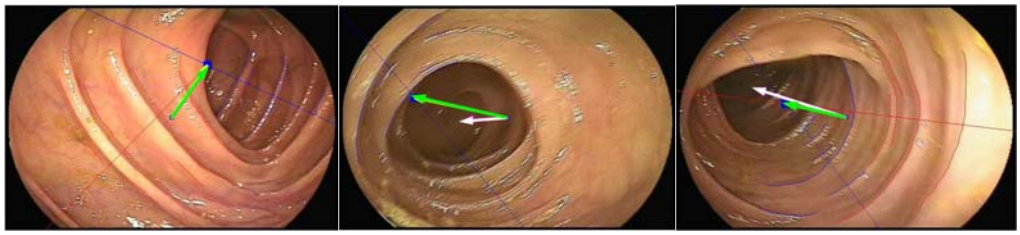
TABLE 2: PARAMETER TABLE

FUNCTION	PARAMETER	PURPOSE
cutTuringBound.m	segmentPixelNum=10 [1]	Making the curve segment small enough for calculating its slope value
cutTuringBound.m	derivativeThresh=15 [1] derivativeUpperbound=60 [1]	Upper and lower thresholds for helping determine whether there is a big turning in the curve, trained by a set of around 10 images

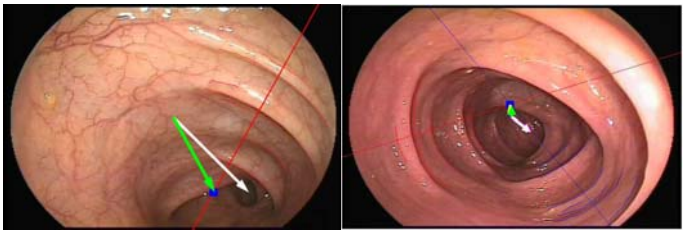
TABLE 2: PARAMETER TABLE (continued)

FUNCTION	PARAMETER	PURPOSE
detectGroupParallelCurve.m	minimalParallelCurves=3[1]	Helping determine whether there is a quasi-parallel curve group in the image
detectParallelCurve.m	cluster=2 [1]	A parameter of K-mean classifier used for grouping all the edges into two groups in terms of their approximate angle values, in the consideration of the fact that we only need at most two groups of quasi-parallel folds to determine colon axis center
detectTwoCurves.m	angleDiffThresh=20 [1]	Help determine quasi-parallel edge segments
detectTwoCurves.m	outlierNumThresh=30% [1]	Help determine two quasi-parallel folds
detectTwoCurves.m	curveSegmentLowerThresh=40 [1]	Help determine whether the projected curve is qualified one for later comparison
determineColonCenterAxis.m	threshTooClose2Border= 18[1]	Threshold used for determine whether the colon center axis is too close to the image border
determineColonCenterAxis.m	angleThresh=35[1]	Angle threshold helping determine whether two groups of parallel curves should be considered as one
determineColonCenterAxis.m	boundaryLengthLowerThresh=30[1]	Help delete small noise curves
findIntensityDrop.m	shift=10 [1]	Threshold for number of pixels that need to be shifted to generate shifted plot of intensity
findIntensityDrop.m	tolerant=5 [1]	Help cancel the influence caused by shifting
findIntensityDrop.m	ratio=0.25 [1]	Helping determine the candidate pixels for colon center axis
findIntensityDrop.m	brightnessThresh=190 [2]	Help filter out unqualified non-dark lumen images, trained by a set of wall and non-dark lumen images
getEdges.m	cannyVector=[0.02, 0.10] [2]	Help remove as much as noise edges, and meanwhile get most of the major boundaries when applying canny edge operator, trained by a set of images

For performance evaluation, we superimpose a “view direction” arrow on the original image to create the corresponding arrow-annotated image with the arrow head at the colon center axis and the arrow tail at the center of the image. The average time taken per image for determining the colon center axis on the test machine (Windows 2003 SP2 on Intel Xeon dual Quad-Core 1.86GHz with 4 GB RAM) is 3.43 seconds. Each annotated image was manually evaluated by one trained staff and verified by one experienced endoscopist. For the first two test image sets, we gave each image a score between A (best quality) to D ratings (worst quality). Fig. 9 shows images and corresponding scores. Table 3 shows that PCurve is effective for 92.45% (A and B ratings) of S-Class images and 90.18% (A and B ratings) for M-Class images.



A B (too long) B (too short)



C D

Fig. 9 Example of categories; white arrows are the ground truth.Green arrows are annotated by our program.

TABLE 3: EFFECTIVENESS OF PCURVE FOR S-CLASS AND M-CLASS IMAGES

Evaluated Technique Category	PCurve for S-Class % Correct	PCurve for M-Class (%) % Correct
A	74.92	73.95
B	17.52	16.23
C	4.30	6.61
D	3.25	3.21

For the third test set, only 8.4% (84/1000) of wall view images are incorrectly detected as non-dark lumen images.

The drawbacks of PCurve are as follows. When there is significant distortion (i.e, a bend or corner distal to the position of the camera) inside the colon, the distal colon center axis is not close to the intersection of

perpendicular bisectors of the two groups of parallel folds. Furthermore, only partial edges of the folds are detected. One possible solution is to give different weights to different segments along the parallel folds according to the intensity.

4. OBJECTIVE MEASUREMENT OF COLONOSCOPY QUALITY USING METRICS

Experienced endoscopists have stated that it is desirable to inspect all quadrants of the colon wall while the endoscope is gradually withdrawn. In this chapter, we describe a set of proposed metrics and present the result of the study on the correlation between these metrics and the circumferential inspection quality marked by domain experts on a large data set of colonoscopy videos by different endoscopists.

4.1 PROPOSED APPROACH

4.1.1 Proposed Metrics

Based on experienced endoscopists' suggestion, we proposed several metrics and discuss the rationale for including these metrics and the calculation methods.

4.1.1.1 Modified QCH Score

The tradition Quadrant Coverage Histogram and QCH score proposed by Liu et al. only make use of clear lumen view images of colonoscopic videos, discarding all the blurry and wall view images. However, blurry and wall view images are other important indicators of quality in some cases. When an endoscopist withdraws the instrument too fast, and does not perform careful inspection, blurry images would be generated. Although there are other situations (e.g. when the endoscopist uses water jet to cleanse stool in the colon) where blurry images could appear, we propose to include blurry images into our metric calculation. Besides, a long sequence of wall view images would sometimes tell us that the camera tip stays close to the colon wall during a relatively long period of time due to the absence of endoscopist. It may reduce the inspection quality under such a circumstance, and inclusion of wall view images would be also reasonable and helpful. Therefore, based on the traditional QCH we come up with a more elaborate metric - the modified QCH score, and the only difference between the two is that we count blurry and wall view images in the metric calculation. Specifically, we know that there is no way to

determine the view direction of wall and blurry images as neither of them has information about the relative position of lumen. Therefore the modified QCH maps their view directions to the number of 0 in the time window, which means none of the quadrants of the colon wall have been inspected on the corresponding moment. At last, we compute a modified average quadrant coverage score over all the images including wall and blurry ones from modified QCH.

4.1.1.2 Spiral Number of Withdrawal Phase

Experienced endoscopists have stated that as much as possible of the mucosa (circumferential inspection of all sides or 360 degrees inspection) of the colon wall should be examined throughout the withdrawal phase. Fig. 10 (A-B) shows a desirable withdrawal inspection pattern in which the lumen is seen in the camera field of view. The colon center axis is at the center of the lumen (green area). Fig. 10 (C-D) shows a different inspection pattern where one side of the colon wall is seen each time without the center of the lumen. Depending on the configuration of the anatomy and the location in the colon a combination of these two types of inspection may be required to see all mucosa.

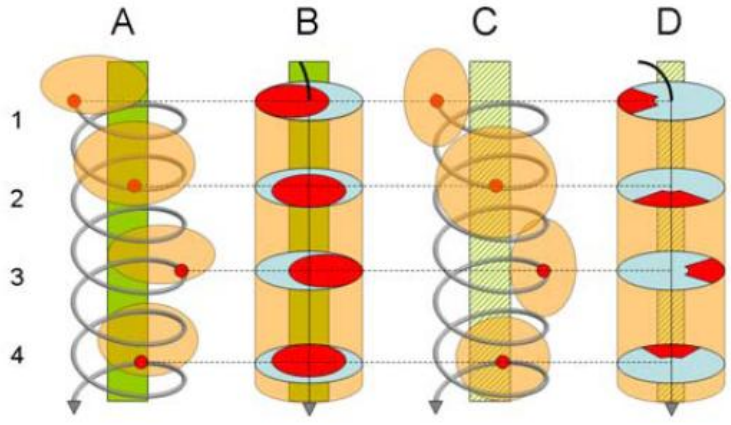


Fig. 10 Circumferential withdrawal inspection pattern: A and C show the spiral shape of the movement of the tip of the endoscope. B and D show the visual field in red. A and B reflect views in which the distant proximal colon is visible; C and D reflect views in which the distant proximal colon is absent and only colon wall is seen. [Courtesy of Dr. de Groen]

We define a ‘spiral’ as a completion of inspection of four different quadrants of the colon wall. We measure the number of spirals considering only the lumen images. Using the center of the image as the reference, the

quadrants are fixed for all images: top-left quadrant (Q1), top-right quadrant (Q2), bottom right quadrant (Q3), and bottom left quadrant (Q4). The more ‘spirals’, the more likely a high-quality inspection of the colon. The side of the colon that is inspected is estimated to be 180 degrees opposite from the view direction of the colon lumen. For instance, Fig. 11(a-d) shows a series of inspection of the quadrants: Q1, Q2, Q4, and Q3. Since endoscopists may have individual inspection preference, e.g. clockwise or counter-clockwise order, we do not consider the order of these numbers as important. As long as all four quadrants are inspected, one spiral is counted. Using this definition of ‘spiral’, we define ‘spiral number’ as the number of spirals performed during a given duration of a procedure. We can automatically measure ‘spiral number’ of a withdrawal phase. This spiral number metric provides an indication of the quality of circumferential inspection. It becomes obvious that the four sequential images form a ‘spiral’ in Figure 11.

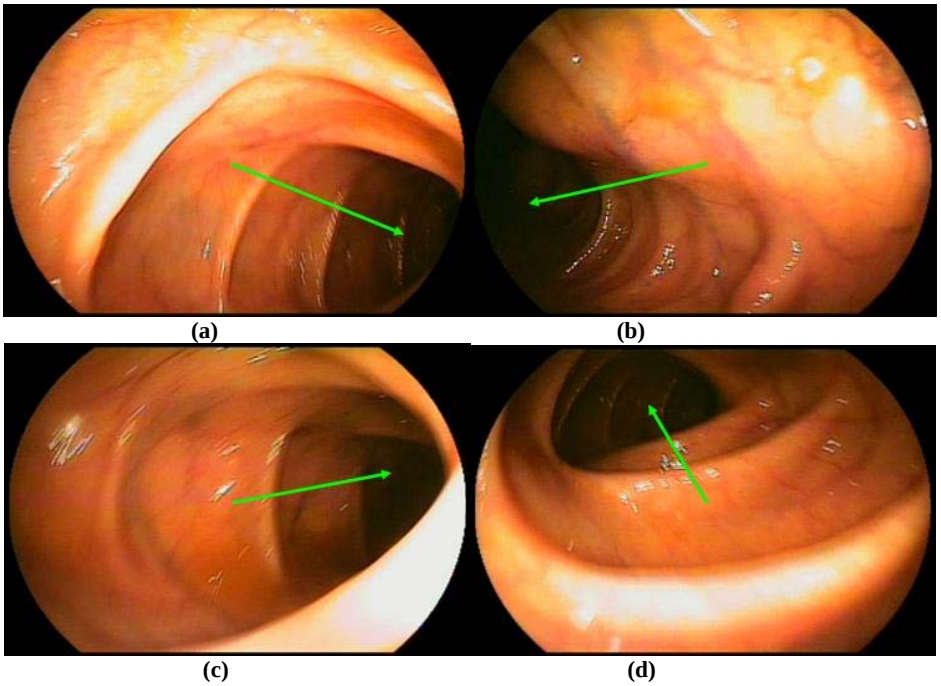


Figure 11 four sequential images form a ‘spiral’

4.1.1.3 Average Spiral Number of Withdrawal Phase

In our study, we notice that different endoscopists spend different average amounts of time to finish one ‘spiral’. Under some circumstances, it may also take an endoscopist different lengths of time period to complete ‘spirals’, and Figure 12 shows this situation. If we extract one frame per second from videos, the completion of the 1st

spiral takes 4 seconds, whereas it takes 13 seconds to finish the 2nd spiral.

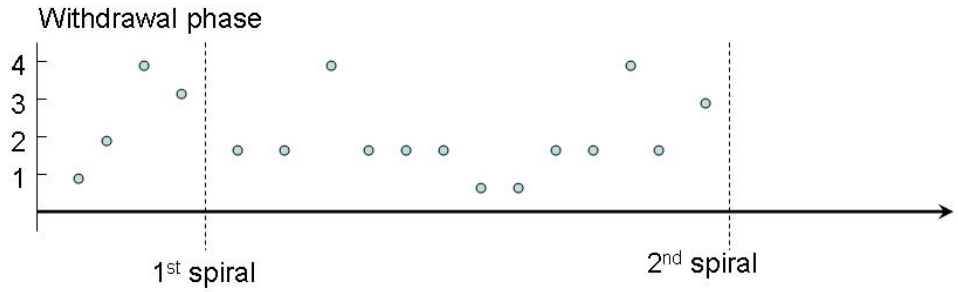


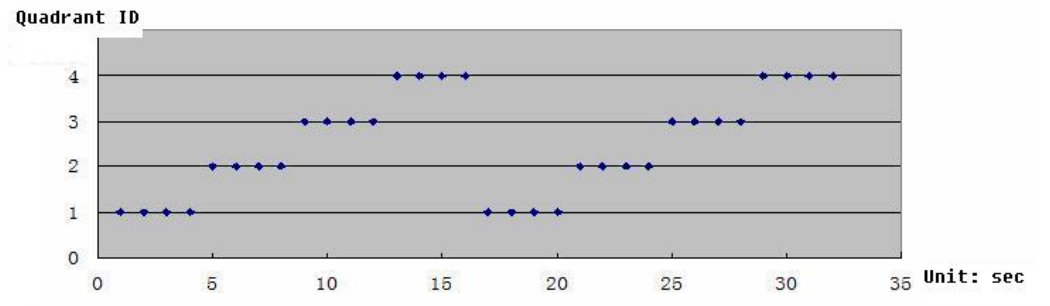
Figure 12 endoscopist spends different time amounts to complete different 'spirals'

Therefore, time might be another important factor to help assess endoscopists' procedure skill, and we'd like to take it into account. This is why we propose another useful metric – average spiral number of withdrawal phase. The formula of this metric is: the number of spirals during withdrawal phase/ withdrawal duration.

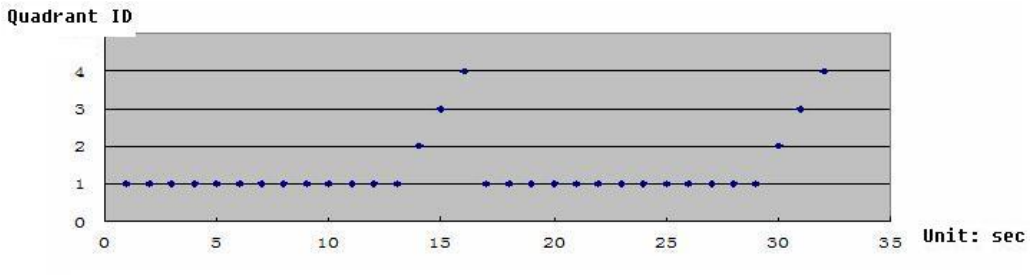
We will use it as another useful metric to explore objective method to measure inspection quality.

4.1.1.4 Evenly Distribution Score

Besides QCH score and spiral number of withdrawal phase, we find another metric – evenly distribution score also plays an important role in our study. Look at Figure 13 where x-axis represents time and y-axis indicates the inspection quadrant, according to our previous discussion you can easily tell that there are two spirals within 35 seconds in each cases, and the average spiral number is the same too.



(a) Spending the same time amount inspecting each quadrant



(b) Spending most of the time inspecting the first quadrant
Figure 13 two different inspection cases

However, you may intuitively feel that there is some difference between the two patterns. In fact, in the first case the endoscopist spends the same time amount to inspect each quadrant of colon wall; whereas in the latter case it takes another endoscopist most of the time to inspect first quadrant and only a very small amount of time to inspect the rest of quadrants. According to experienced endoscopists' view, it is more desirable that endoscopists spend equal amount of time to inspect each quadrant of colon wall, which means (a) is a much better inspection pattern than (b). Therefore, we want to include evenly distribution score into our study to reflection this fact. The metric can help measure how evenly an endoscopist distributes time among each colon wall quadrant when performing inspection.

The definition of evenly distribution score is provided by an experienced endoscopist. He defines it as the average of the two shortest quadrant times. To obtain the metric, we first need to calculate an average cycle time, i.e. the average time within which the endoscopist can complete the inspection of all four quadrants. Then we fix the length of time window with the average cycle time, and count the times spent inspecting each quadrant within each time window. For each time window, we determine the two shortest quadrant times, and calculate their average. At the end, making the average among all the time windows would render us the evenly distribution metric. For example, in figure 13(b), the average cycle time is 16 seconds, and within the first 16 seconds the endoscopist spent 13 seconds to inspect 1st quadrant, and 1 second to inspect each of the rest quadrants. Therefore, the average of the two shortest quadrant times is 1 for the first time window. Same thing happens in the second time window. Thus it is easy to know that the evenly distribution score for the case in figure 13(b) is 1. Also, it is worth noticing that this definition can avoid the skew caused by one significant outlier.

4.1.2 Proposed Analysis Method

In statistics, the Pearson correlation coefficient is a measure of the correlation (linear dependence) between two variables X and Y, giving a value between +1 and -1 inclusive. It is widely used in the sciences as a measure of the strength of linear dependence between two variables. Position correlation means that when one variable increases, the other tends to increase; whereas negative one means that when one variable increases, the other tends to decrease. When a correlation coefficient is close to +1 (or –1), it means that there is a strong correlation. For example, a correlation r = 0.7 may be considered strong. However, the closer a correlation coefficient gets to 0, the weaker the relationship is.

After having determined the above metrics, we used Pearson correlation analysis method to explore the correlation between those metrics and the circumferential inspection quality.

4.1.2.1 Required Assumptions of Pearson Correlation Analysis

Before applying this method to our study, we need to first make sure that the required assumptions for Pearson correlation analysis are satisfied. These assumptions include linearity, normality, independence, and constant variance.

As mentioned earlier, this correlation coefficient measures a linear relationship. That is, the relationship between the two variables measures how close the two measurements form a straight line when plotted on an x-y chart. Therefore, it is important that data be graphed before the correlation is interpreted. To check whether this assumption is satisfied, we can plot predictor values on the x-axis versus the residuals on the y-axis. If the data meet the assumption of linearity, the points would be symmetrically distributed around the horizontal line (as in Fig. 14) and the plot should not form a U-shaped line or an up-side down U.

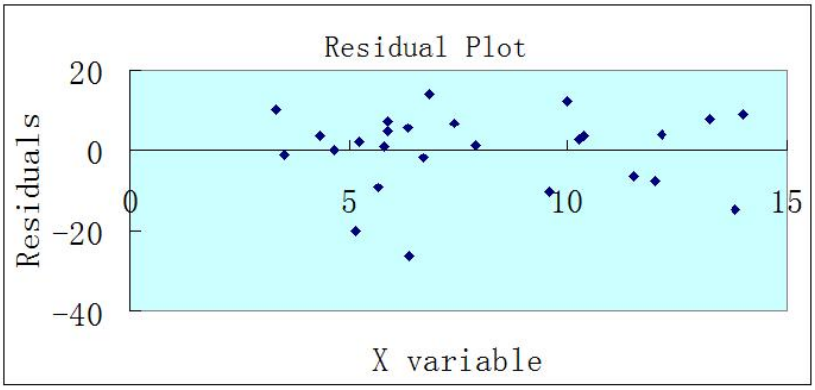


Fig. 14 Residual plot when the assumption of linearity is satisfied

The best test for normality is a normal probability plot of the residuals. If the residuals are normally distributed with mean 0, the points on this plot should fall close to the diagonal line (e.g. the red-dashed line in Fig. 15).

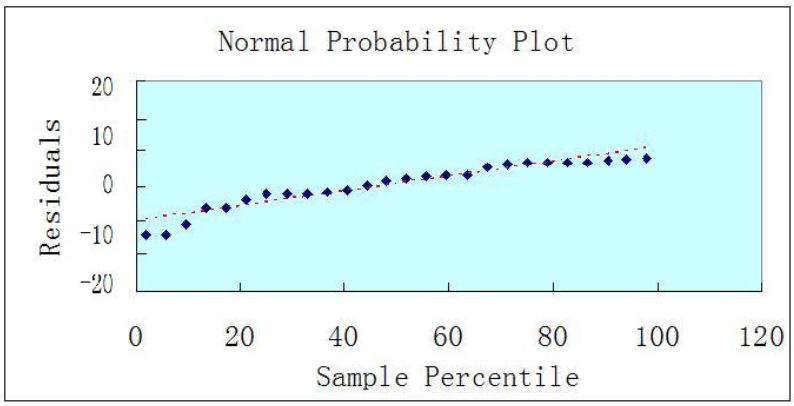


Fig. 15 Normal probability plot of residuals when the assumption of normality is satisfied

Pearson correlation analysis assumes that the data points are independent of each other, meaning that the value of one data point does not depend on what the value of any other data point is. The most common violation of this assumption is in time series data, where some Y variable has been measured at different times. However, in this study, what we have are not time sequence date, therefore the independence assumption is satisfied here.

As for constant variance, we look at the residuals versus predictor values plot. If the residuals are not getting larger as a function of the predicted value, we may not have a problem of non-constant variance. In the case of figure 14, it is obvious that the assumption of constant variance is satisfied.

4.1.2.2 Pearson Correlation Analysis

Pearson's correlation coefficient between two variables is defined as the covariance of the two variables divided by the product of their standard deviations. After making sure that all the assumptions are satisfied in the study, we can use the following formula to calculate the Pearson correlation coefficient:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

Y_i is the ground truth score (0-100) of quality of the visualization technique for video *i* given by the domain experts. Hence, the score reflects not only the circumferential inspection quality but also the quality of examination of flexures, rectal valves, and the ileocecal valve. X_i is the value of one metric for video *i* depending on which metric is being correlated with the ground truth score. $\overline{X}$ and $\overline{Y}$ are the means of X_i and Y_i, respectively.

4.2 EXPERIMENT RESULT

All of the twenty one videos in our experiments are of procedures performed by more than one endoscopist. They were captured during colonoscopic procedures using Fujinon scopes. No patient identifiable information is available in these videos. The video format is MPEG-2 with the image resolution of 480×720 pixels. All the selected videos are of good or excellent bowel preparation condition (i.e. colon wall has been clean carefully and there is little stool inside the colon) rated by four domain experts. The videos also have no biopsy or therapeutic operations. This is to prevent inclusion of other patterns such as colon cleaning, polyp removal, biopsy from being measured as circumferential inspection patterns. The experienced endoscopists provided us the ground truth inspection quality scores, which reflect not only the circumferential inspection quality but also the quality of examination including flexures, rectal valves, and the ileocecal valve. Here, we need to mention that what we need in our research is only the ground truth score for circumferential inspection quality. However, before we obtain more fine-tuned scores from this experienced endoscopist, we treat the average inspection quality scores for each video as ground truth visualization score (Y variable) in our early-stage analysis.

For obtaining the metrics, we first extracted one frame per second from each video. In order to locate the colon

center axis of each frame, we used our previous technique [13, 15] supplemented with the proposed PCurve technique handling non-dark lumen images. Then, we calculated the new quality metrics (modified QCH scores, spiral number, average spiral number and evenly distribution score) and Liu's metric (QCH score) [13] for the whole procedure and the withdrawal phase only. We determined the start of the withdrawal phase for each video using the average maximum intubation frame provided by the same domain experts. Table 4 presents the result of all the useful metrics calculated for whole procedure and Table 5 shows the result calculated for the withdrawal phase. The traditional and modified QCH scores in the two tables can be calculated by using the plot of average quadrant coverage scores introduced in section 2.2 of Chapter 2. They indicate the inspection time to complete 3.5 quadrants of colon wall with laterality threshold equal to 0.66. The reason why we use the inspection time for 3.5 quadrants as our metrics here is because some videos are too short in length to obtain the inspection time for four quadrants.

TABLE 4: SIX METRICS CALCULATED DURING WHOLE PROCEDURE

video ID	ground truth score	procedure duration(sec)	traditional QCH score (sec)	modified QCH score (sec)	Spiral No.	Average Spiral No. (/min)	evenly distribution score (sec)
11.mpg	91.75	1455	6	17	35	1.44	9.56
14.mpg	60.25	965	8	23	20	1.24	13.83
15.mpg	55	521	6	35	7	0.81	3.5
17.mpg	42.5	739	9	36	10	0.81	5.67
20.mpg	92.75	1312	6	20	35	1.6	6.7
27.mpg	62	941	11	29	19	1.21	6.33
28.mpg	64	758	8	23	16	1.27	11.5
29.mpg	58.75	1150	5	95	11	0.57	5.17
31.mpg	37.5	749	20	41	8	0.64	12
33.mpg	62.5	664	5	23	17	1.54	4.67
34.mpg	84.25	1224	5	35	22	1.08	5.88
35.mpg	68.75	746	6	35	14	1.13	4.33
37.mpg	86.25	1097	11	35	12	0.66	14
39.mpg	47.25	999	8	66	14	0.84	6.38
40.mpg	98.75	1477	9	27	28	1.14	7.4
41.mpg	93.5	1319	8	26	30	1.36	10.36
46.mpg	75.5	1141	6	97	14	0.74	10
49.mpg	67.25	530	8	26	11	1.25	6.83
52.mpg	81	797	5	29	13	0.98	5.8

TABLE 4: SIX METRICS CALCULATED DURING WHOLE PROCEDURE (continued)

video ID	ground truth score	procedure duration(sec)	traditional QCH score (sec)	modified QCH score (sec)	Spiral No.	Average Spiral No. (/min)	evenly distribution score (sec)
54.mpg	59.75	891	5	115	8	0.54	3.33
56.mpg	82	1227	5	74	13	0.64	5.88

TABLE 5: SIX METRICS CALCULATED DURING WITHDRAWAL PHASE

video ID	ground truth score	withdrawal duration(sec)	traditional QCH score (sec)	modified QCH score (sec)	Spiral No.	Average Spiral No. (/min)	evenly distribution score (sec)
11.mpg	91.75	502	8	17	12	1.4343	8.3333
14.mpg	60.25	355	11	23	5	0.8451	8
15.mpg	55	147	6	33	2	0.8163	1.5
17.mpg	42.5	330	14	32	3	0.5455	3
20.mpg	92.75	930	8	17	29	1.871	6.6875
27.mpg	62	479	14	27	10	1.2526	2.5
28.mpg	64	277	14	26	4	0.8664	6
29.mpg	58.75	611	8	115	4	0.3928	5
31.mpg	37.5	122	6	36	2	0.9836	2.5
33.mpg	62.5	333	5	23	8	1.4414	5
34.mpg	84.25	489	5	33	10	1.227	5.75
35.mpg	68.75	281	8	24	6	1.2811	4
37.mpg	86.25	425	12	41	4	0.5647	8
39.mpg	47.25	187	6	23	4	1.2834	4
40.mpg	98.75	853	11	20	17	1.1958	5.375
41.mpg	93.5	542	8	29	13	1.4391	6
46.mpg	75.5	461	6	38	6	0.7809	5.1667
49.mpg	67.25	279	12	41	6	1.2903	9.5
52.mpg	81	298	5	15	8	1.6107	6.25
54.mpg	59.75	339	5	69	3	0.531	3
56.mpg	82	361	5	44	4	0.6648	4.75

TABLE 6: PEARSON CORRELATION ANALYSIS RESULT BETWEEN MANUAL SCORE AND EACH METRIC DURING WHOLE PROCEDURE

Metrics	traditional QCH score (sec)	modified QCH score (sec)	Spiral No.	Average Spiral No. (/min)	evenly distribution score (sec)
Pearson Correlation	-0.378	-0.251	0.719	0.411	0.139

TABLE 7: CORRELATION ANALYSIS RESULT BETWEEN MANUAL SCORE AND EACH METRIC DURING WITHDRAWAL PHASE

Metrics	traditional QCH score (sec)	modified QCH score (sec)	Spiral No.	Average Spiral No. (/min)	evenly distribution score (sec)
Pearson Correlation	-0.059	-0.254	0.679	0.421	0.565

For doing the Pearson correlation analysis, we first check whether the required assumptions are satisfied in the

study. Since the assumptions are the same as those for linear regression analysis and Excel has regression analysis in Analysis Tools, we used Excel to do the assumption checking. We list all the Pearson analysis result between ground truth score and each metric during whole procedure and withdrawal phase only in table 6 and 7. The Excel result shows that the assumptions are satisfied when we use ground truth score as Y variable and each metric as X variable for both table 4 and 5. The Pearson correlation coefficients are relatively low between the ground truth score and the metrics including traditional and modified QCH scores, average spiral number and evenly distribution score, whereas the coefficient is around 0.68 for withdrawal phase; and 0.72 for whole procedure when considering the metric of the spiral numbers (see table 6 and 7). We consider the coefficients of 0.68 and 0.72 as high. Therefore, we conclude that the number of spirals is a marker of circumferential inspection quality and that we indeed can use this metric to provide as an estimates of mucosa inspection quality.

4.3 CONCLUSION

In this study, we have first proposed several useful metrics including modified QCH score, spiral number, average spiral number and evenly distribution score. Then we proposed an approach to help measure colonoscopic circumferential inspection quality by analyzing Pearson correlation between ground truth scores given by four experienced endoscopists and each individual metric. Given colonoscopic videos with good or excellent preparation condition, we calculated the values for all the proposed metrics, and the Pearson correlation analysis result shows that there is relatively strong positive correlation between the number of spirals performed during whole procedure or withdrawal phase and ground truth scores.

Besides it is worth noticing that, as we mentioned in the Experiment Result section, the ground truth scores given by the experienced endoscopist reflect not only the circumferential inspection quality but also the quality of examination including flexures, rectal valves, and etc. Therefore, it may somehow negatively influence our analysis and might reduce the accuracy of the correlation analysis. One solution is to obtain ground truth scores only reflecting circumferential inspection quality from experienced endoscopists, and then to use these fine-tuned scores to conduct further research.

5. OBJECTIVE MEASUREMENT OF COLONOSCOPY QUALITY USING ASSOCIATION RULE MINING

In Chapter 5, we investigate another approach to determine patterns of colon inspection using association rule mining. In this chapter, we provide the background knowledge related to association rule mining and discuss our application rule mining to find interesting inspection patterns in colonoscopy videos.

5.1 BACKGROUND KNOWLEDGE

Association rule mining is a popular and well researched method for discovering interesting relations between variables in large databases. Piatetsky-Shapiro [27] describes analyzing and presenting strong rules discovered in databases using different measures. Based on the concept of strong rules, Agrawal et al. [28] introduced association rules for discovering regularities between products in large scale transaction data recorded by point-of-sale (POS) systems in supermarkets. For example, the rule {onions, potatoes}→{beef} found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, he or she is likely to also buy beef. Such information can be used as the basis for decisions about marketing activities such as, e.g., promotional pricing or product placements. In addition to the above example from market basket analysis, association rules are employed today in many application areas including Web usage mining, intrusion detection, and bioinformatics. Two key performance metrics of association rule mining are *support* and *confidence*. The *support* supp(X) of an itemset X is defined as the proportion of transactions in the data set which contain the itemset; and the *confidence* of a rule is defined conf(X→Y) = supp(X U Y)/supp(X). Many algorithms for generating association rules were presented over time [27-34]. Apriori [29] is the best-known algorithm to mine association rules. It uses a breadth-first search strategy to count the support of itemsets and uses a candidate generation function which exploits the downward closure property of support.

5.2 PROPOSED APPROACH

We hypothesize that a physician has two different inspection patterns: one to search for abnormalities and the

other to treat the found abnormalities, for instance, to perform polypectomy on polyps. In this work, we focus on mining patterns of searching for abnormalities during colonoscopy. The data pre-processing step is used to convert raw colonoscopic video data to the format suitable for association rule mining. We use Weka' association rule mining that implements Apriori [29] algorithm to find interesting association rules.

Data Pre-processing

First, we manually extract frames (at the rate of 1 frame per second) of the withdrawal phase of a colonoscopic video according to the ground truth position given by a physician. Using the view mode estimation method in [15], we filter out all the wall view images. Then, we apply our colon axis center determination technique to get the colon center axis information (x, y coordinates) for each clear lumen view frame. Two examples are provided in Figure 16, where an arrow is pointing at the location of colon axis center for each lumen view image.

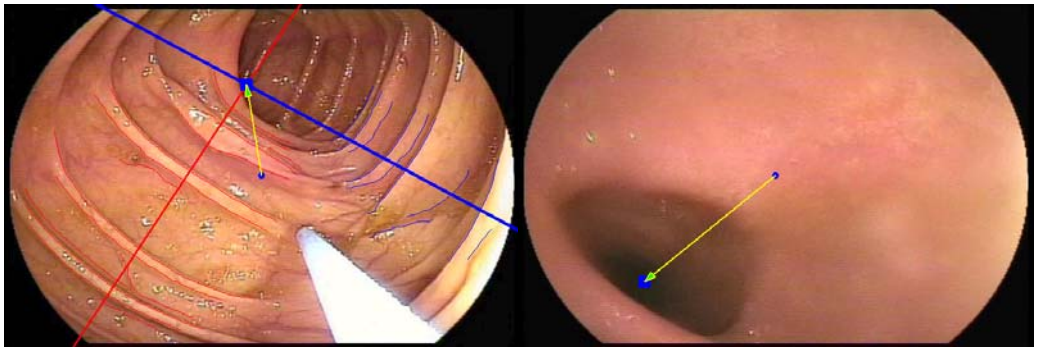


Fig. 16 Lumen view images with arrows pointing out the colon axis center

We propose two methods to convert raw data into an appropriate format for association rule analysis.

I. The quadrant method

We divide each frame into four fixed quadrants (see Figure 17), and then assign each frame one number according the location of colon center axis in the frame. For instance, if the colon axis center falls into the bottom left quadrant, we assign a number of 4 to this frame. After this conversion process, we have a string of numbers with each corresponding to a frame in the sequence of lumen view images (e.g.21114433333333222221222333…..).

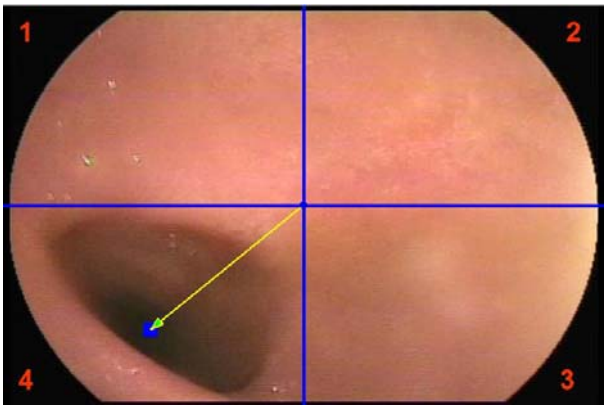


Figure 17 Lumen view image with 4 fixed quadrants

Next, we simplify the data that we obtained from previous steps. Noting that the length of consecutive identical numbers in the string tells us how long a physician spends examining the same area. However, we don't want to focus on such time information in this research; instead we care about the position information of examined area. Therefore, we use a single number to represent consecutive identical numbers, e.g. a string of 1222222223 would become 123 in our dataset.

After we obtained a string without consecutive identical numbers, we use a sliding window with fixed length to generate number segments. In this study, we choose the window length to be 3 as we want to start with simple cases. If the analysis result is satisfactory, we would try longer window length later. Therefore, a string of numbers, say 12324132, would become several number segments 123, 232, 324, 241, 413, 132 after we use the sliding window of length 3.

At last, we group numbered segments with same medical interpretation in the same group. We think the three following numbered segments 123, 231, and 312 have the same medical interpretation, since a physician would inspect the same area of colon wall by following these patterns and they are all either clockwise or counter-clockwise. Therefore, in this research we replace patterns by the one having the same medical interpretation and the smallest beginning number. In the case above, we would replace 231 and 312 by 123 in our dataset. Table 8 and 9 list all the possible patterns, where the patterns on the same row have the same medical interpretation. The patterns in Table 8 have two identical numbers; whereas the patterns in Table 9 have there different numbers.

Table 8 POSSIBLE PATTERNS OF TWO IDENTICAL NUMBERS WITH SAME MEDICAL INTERPRETATION

121	212
131	313
141	414
232	323
242	424
343	434

Table 9 POSSIBLE PATTERNS OF THREE DIFFERENT NUMBERS WITH SAME MEDICAL INTERPRETATION

123	231	312
124	241	412
132	321	213
134	341	413
142	421	214
143	431	314
234	342	423
243	432	324

II. The 12-zone method

We divide each quadrant that we defined in the first method into three different zones (see Figure 18). According to the opinion of an expert colonoscopist, the zones within the purple band are the desirable inspection zones. The reason is that a physician can inspect as many as three quarters of colon wall if the corresponding colon center axis falls into the this band. For instance, in Figure 19(a), the center axis is the intersection point of the two blue line segments, and it falls into the purple band. Under such circumstance, the top left, bottom right and bottom left areas of the colon wall can be inspected. The zones in the blue area are the acceptable one because in such cases only one quarter of the colon wall is under inspection (in Figure 19(b), only the top left area is inspected). The undesirable zones are in the yellow area, which is around the center of the frame. If the colon center axis falls into this area, it means the physician doesn't closely inspect colon wall with intention (see Figure 19(c)). To sum up, using the second method we end up with having 12 zones, and frames with colon center axis falling into the purple or blue area are desirable.

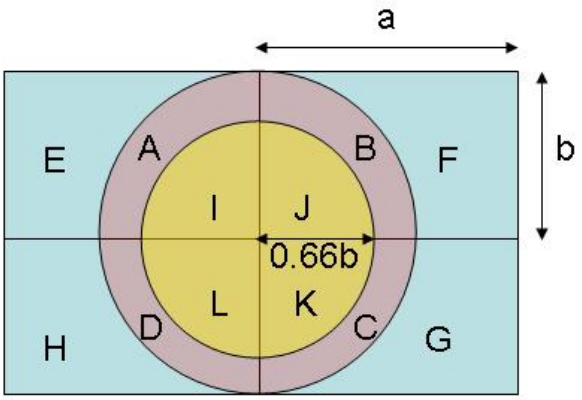
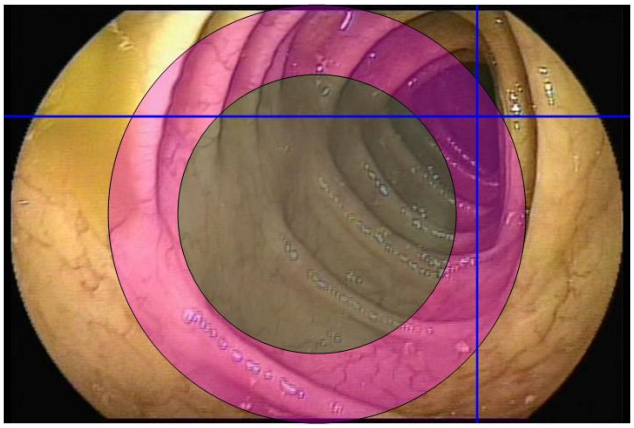
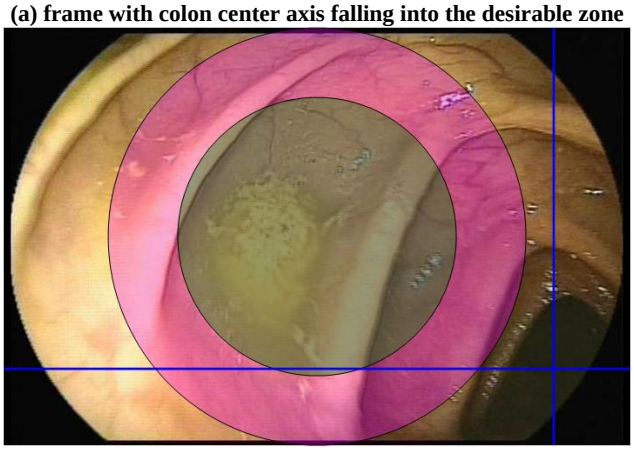


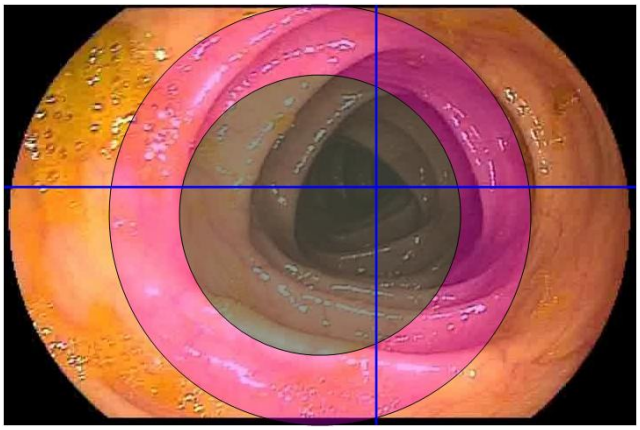
Figure 18 Twelve quadrants obtained by the 12 zone method



(a) frame with colon center axis falling into the desirable zone



(b) frame with colon center axis falling into the acceptable zone



(c) frame with colon center axis falling into the undesirable zone

Figure 19 colon center axes fall into different inspection zones

5.3 EXPERIMENT RESULT

We first take three videos of colonoscopic procedures performed by an experienced physician having good skill of inspection (video ID: capture0011, capture0295, capture0300), and one video by another physician with relatively poor skill (video ID: 6.mpg). According to the two methods described in Section 3, we pre-process the raw data and obtain appropriate datasets for these four videos. Later, we use the software Weka's associate rule learning to mine the inspection patterns for each of the physicians. In this software, we can set several parameters including the values of minimum support, the minimum confidence.

With the quadrant method we obtained the raw result presented in Appendix I – RAW RESULT WITH QUADRANT METHOD (page 42) directly from Weka for the four selected videos, and Appendix II – RAW RESULT WITH 12-ZONE METHOD (page 44) shows the raw result for the three selected videos after applying the 12-zone method.

Table 10 is the summary of the analysis result for the four selected videos by using the quadrant method. In the "Major inspection pattern" column of the table, "First location" shows the quadrant location of colon center axis in one frame; while "Second location" is the quadrant location of colon center axis in the consecutive frame.

Table 10 summary of the analysis result with the quadrant method

Video ID	Major inspection pattern	support	confidence	inspection quality
Capture0011	Second location =2 => First location =1	29.79%	100%	Good
Capture0295	Second location =2 => First location =1	30.77%	100%	Good
Capture0300	Second location =2 => First location =1	22.52%	100%	Good
6.mpg	Second location =2 => First location =1	27.78%	100%	poor

The analysis result shows the fact that the selected videos of colonoscopic procedures performed by physicians with good or poor inspection skill have the same major inspection pattern, which is {Second inspection=2} → {First inspection=1}. Therefore, the quadrant method cannot help differentiate colonoscopic procedures with good and poor inspection quality by using inspection pattern.

Table 11 is the analysis result for the three selected videos by using the 12 zone pre-processing method. The meanings of “First location” and “Second location” remain the same as in Table 10, except that it is the zone location of colon center axis. The reason why we didn’t list the result of video Capture0300 is because we didn’t mine any rule with minimum support equal to 5%.

Table 11 summary of the analysis result with 12 zone pre-processing method

Video ID	Major inspection pattern	support	confidence	Video quality
Capture0295	Second location =E => First location =A	8.99%	73%	Good
	Second location =F => First location =B	7.87%	66%	
Capture0011	Second location =G => First location =E	10.16%	59%	Good
	Second location =E => First location =A	8,56%	59%	
6.mpg	First location =I => second location =J	12.82%	83%	poor
	First location =C => second location =K	10.26%	80%	

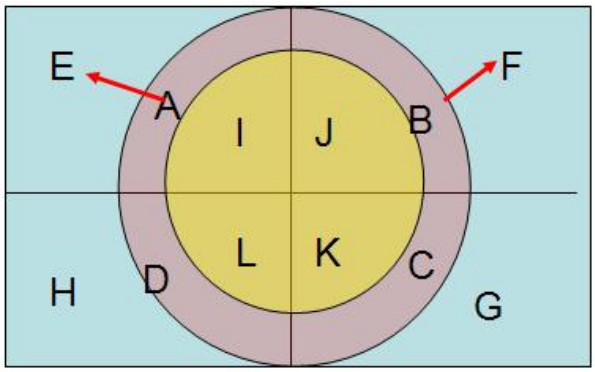


Figure 20 major inspection pattern of video capture0295

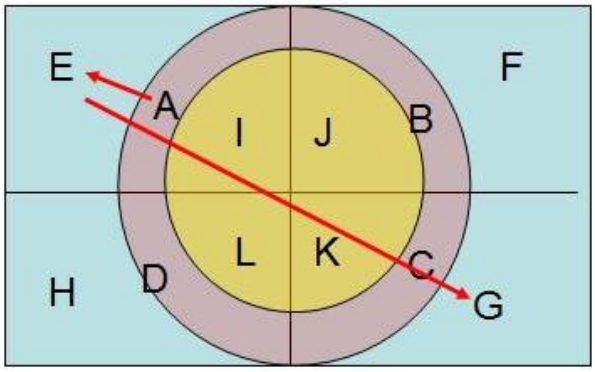


Figure 21 major inspection pattern of video capture0011

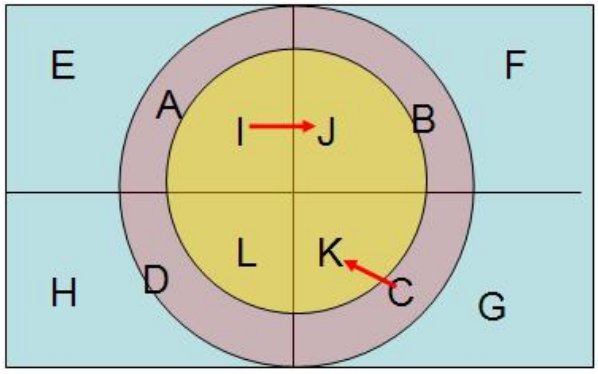


Figure 22 major inspection pattern of video 6.mpg

The analysis result in Table 11 and in Fig. 20, 21 & 22 shows the fact that the selected videos of colonoscopic procedures performed by physicians with good or poor inspection skill have different major inspection patterns. For the videos capture0295 and capture0011 by experienced physician, the two major patterns have a common feature: in one and consecutive frames the colon center axes fall into the desirable and/or acceptable quadrants; whereas with the video 6.mpg by physician with poor inspection skill the two major patterns have a common

feature: in one and consecutive frame the colon center axes fall into the acceptable and/or undesirable quadrants. This analysis result coincides with the medical interpretation given by experienced colonoscopists. Therefore, the 12 zone method can help differentiate colonoscopic procedures with good and poor inspection quality by using inspection pattern.

5.4 CONCLUSION

In this study, we've proposed a technique to help measure colonoscopic procedures' inspection quality by applying data mining technique. Given colonoscopic videos without physician intervention, we focus on mining inspection patterns of looking for abnormalities during the withdrawal phase. For doing so, we came up with two different data pre-processing methods. After we obtained appropriate datasets for four different videos including three videos of colonoscopic procedures performed by an experienced physician with good inspection skill and one video by another physician with relatively poor inspection skill, we use Weka's association rule learning tool to mine inspection patterns for the two physicians. The analysis result has shown that the 12-zone method can help differentiate colonoscopic procedures with good inspection quality from ones with poor inspection quality, however the quadrant method cannot. Another conclusion is: colonoscopic videos with high inspection quality have a common feature: in one and consecutive frame the colon center axes fall into the desirable and/or acceptable quadrants; whereas the common feature of videos with low inspection quality is: in one and consecutive frame the colon center axes fall into the acceptable and/or undesirable quadrants.

6. CONCLUSION AND FUTURE WORK

We have presented our novel technique for determining colon center axis of non-dark lumen images. Also, two approaches to help objectively measure the quality of colonoscopy have been discussed in this paper. The preliminary result on real colonoscopy videos demonstrates that using the spiral number during whole procedure/withdrawal phase and inspection patterns to assess colonoscopy quality would be feasible and promising methods. With the initial research step complete, some future work is planned. First, an important limitation of our colon center axis determination technique is the assumption that there is not much distortion inside the colon and the colon center axis is always located around the intersection of two groups of parallel folds. However, this is not always true and in some cases severe distortion degrades the accuracy of the technique. Therefore, one of our future tasks will remove this limitation through giving different weights to different segments along the parallel folds according to the intensity. Second, as we mentioned in chapter 4, the ground truth scores given by the experienced endoscopists do not exactly reflect circumferential inspection quality, and this somehow negatively influences our analysis and reduces the result's accuracy. Thus, another future work is to adjust our correlation analysis after obtaining fine-grained ground truth scores. Third, we would run our pattern-mining technique on more colonoscopic videos in order to obtain more reliable inspection patterns for endoscopists with good/poor colonoscopic skills.

Appendix I –RAW RESULT WITH QUADRANT METHOD

Capture0011:

```
Minimum support: 0.1 (14 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 9

Size of set of large itemsets L(2): 13

Size of set of large itemsets L(3): 4

Best rules found:

 1. third_inspection=1 48 ==> first_inspection=1 48    conf:(1)
 2. second_inspection=2 42 ==> first_inspection=1 42    conf:(1)
 3. second_inspection=4 third_inspection=1 18 ==> first_inspection=1 18    conf:(1)
 4. second_inspection=2 third_inspection=1 18 ==> first_inspection=1 18    conf:(1)
 5. second_inspection=2 third_inspection=4 16 ==> first_inspection=1 16    conf:(1)
 6. second_inspection=4 42 ==> first_inspection=1 34    conf:(0.81)
 7. first_inspection=2 26 ==> second_inspection=3 20    conf:(0.77)
 8. third_inspection=4 36 ==> first_inspection=1 25    conf:(0.69)
 9. third_inspection=2 39 ==> first_inspection=1 27    conf:(0.69)
10. second_inspection=3 57 ==> first_inspection=1 37    conf:(0.65)
```

Caputre0295:

```
Minimum support: 0.1 (18 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 9

Size of set of large itemsets L(2): 13

Size of set of large itemsets L(3): 3

Best rules found:

 1. second_inspection=2 56 ==> first_inspection=1 56    conf:(1)
 2. third_inspection=1 48 ==> first_inspection=1 48    conf:(1)
 3. second_inspection=2 third_inspection=1 25 ==> first_inspection=1 25    conf:(1)
 4. second_inspection=4 third_inspection=2 28 ==> first_inspection=1 22    conf:(0.79)
 5. third_inspection=4 31 ==> first_inspection=1 23    conf:(0.74)
 6. second_inspection=4 80 ==> first_inspection=1 59    conf:(0.74)
 7. third_inspection=3 53 ==> first_inspection=1 38    conf:(0.72)
 8. first_inspection=1 third_inspection=2 32 ==> second_inspection=4 22    conf:(0.69)
 9. third_inspection=3 53 ==> second_inspection=4 36    conf:(0.68)
10. third_inspection=2 50 ==> first_inspection=1 32    conf:(0.64)
```

Capture0300:

```
Minimum support: 0.15 (33 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 9

Size of set of large itemsets L(2): 10

Size of set of large itemsets L(3): 1

Best rules found:

 1. third_inspection=1 58 ==> first_inspection=1 58    conf:(1)
 2. second_inspection=2 50 ==> first_inspection=1 50    conf:(1)
 3. second_inspection=4 third_inspection=1 37 ==> first_inspection=1 37    conf:(1)
 4. third_inspection=3 72 ==> second_inspection=4 56    conf:(0.78)
 5. third_inspection=4 49 ==> first_inspection=1 34    conf:(0.69)
 6. third_inspection=4 49 ==> second_inspection=3 33    conf:(0.67)
 7. third_inspection=1 58 ==> first_inspection=1 second_inspection=4 37    conf:(0.64)
 8. first_inspection=1 third_inspection=1 58 ==> second_inspection=4 37    conf:(0.64)
 9. third_inspection=1 58 ==> second_inspection=4 37    conf:(0.64)
10. second_inspection=4 106 ==> first_inspection=1 66    conf:(0.62)
```

6.mpg (video with low quality):

```
Minimum support: 0.25 (4 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 8

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 1

Best rules found:

 1. third_inspection=2 5 ==> second_inspection=3 5    conf:(1)
 2. second_inspection=2 5 ==> first_inspection=1 5    conf:(1)
 3. first_inspection=1 third_inspection=2 4 ==> second_inspection=3 4    conf:(1)
 4. second_inspection=4 4 ==> third_inspection=3 4    conf:(1)
 5. third_inspection=1 4 ==> first_inspection=1 4    conf:(1)
 6. second_inspection=3 9 ==> first_inspection=1 8    conf:(0.89)
 7. third_inspection=2 5 ==> first_inspection=1 second_inspection=3 4    conf:(0.8)
 8. second_inspection=3 third_inspection=2 5 ==> first_inspection=1 4    conf:(0.8)
 9. third_inspection=2 5 ==> first_inspection=1 4    conf:(0.8)
10. third_inspection=3 6 ==> second_inspection=4 4    conf:(0.67)
```

Appendix II – RAW RESULT WITH 12-ZONE METHOD

Caputre0295 (video with good quality):

```
Apriori
=======

Minimum support: 0.05 (13 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 19

Generated sets of large itemsets:

Size of set of large itemsets L(1): 16

Size of set of large itemsets L(2): 4

Best rules found:

 1. second_inspection=E 33 ==> first_inspection=A 24    conf:(0.73)
 2. second_inspection=F 32 ==> first_inspection=B 21    conf:(0.66)
```

Capture0011(video with good quality):

```
Apriori
=======

Minimum support: 0.05 (9 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 19

Generated sets of large itemsets:

Size of set of large itemsets L(1): 15

Size of set of large itemsets L(2): 4

Best rules found:

 1. second_inspection=G 32 ==> first_inspection=E 19    conf:(0.59)
 2. second_inspection=E 27 ==> first_inspection=A 16    conf:(0.59)
```

6.mpg (video with low quality):

```
Apriori
=======

Minimum support: 0.1 (4 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 8

Size of set of large itemsets L(2): 3

Best rules found:

 1. first_inspection=I 6 ==> second_inspection=J 5    conf:(0.83)
 2. first_inspection=C 5 ==> second_inspection=K 4    conf:(0.8)
 3. second_inspection=J 10 ==> first_inspection=I 5    conf:(0.5)
 4. second_inspection=K 8 ==> first_inspection=C 4    conf:(0.5)
```

REFERENCE

[1] World Health Organization, "Cancer", February 2009, http://www.who.int/mediacentre/factsheets/fs297/en/

[2] S. Vijan, J. Inadomi, R. A. Hayward, T. P. Hofer, and A. M. Fendrick, "Projections of demand and capacity for colonoscopy related to increasing rates of colorectal cancer screening in the United States," Aliment Pharmacol Ther, vol. 20, pp. 507-515, 2004.

[3]. Hixson LJ, Fennerty MB, Sampliner RE, Garewal HS (1991) Prospective blinded trial of the colonoscopic miss-rate of large colorectal polyps. Gastrointest Endosc 37:125–127

[4]. RexDK, Cutler CS, LemmelGT, Rahmani EY, ClarkDW,HelperDJ, Lehman GA, Mark DG (1997) Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. Gastroenterology 112:24–28

[5]. van Rijn JC, Reitsma JB, Stoker J, Bossuyt PM, van deventer SJ, Dekker E (2006) Polyps miss rate determined by tandem colonoscopy: a systematic review. Am J Gastroenterol 101:343–350

[6]. Kim DH, Pickhardt PJ, Taylor AJ, Leung WK, Winter TC, Hinshaw JL, Hinshaw JL, Gopal DV, Reichelderfer M, Hsu RH, Pfau PR (2007) CT colonography versus colonoscopy for the detection of advanced neoplasia. N Engl J Med 357:1403–1412

[7]. Rex DK, Goodwine BW (2002) Method of colonoscopy in 42 consecutive patients presenting after prior incomplete colonoscopy. Am J Gastroenterol 97:1148–1151

[8]. Rex DK, Imperiale TF, Latinovich DR, Bratcher LL (2002) Impact of bowel preparation on efficiency and cost of colonoscopy. Am J Gastroenterol 97:1696–1700

[9]. Barclay RL, Vicari JJ, Doughty AS, Johanson JF, Greenlaw RL (2006) Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. N Engl J Med 355:2533–2541.

[10] D. K. Rex, J. L. Petrini, T. H. Baron, A. Chak, J. Cohen, S. E. Deal, B. Hoffman, B. C. Jacobson, K. Mergener, B. Pertersen, M. A. Safdi, D. O. Faigel, and I. M. Pike. "Quality indicators for colonoscopy". Gastrointestinal Endoscopy, vol. 63, pp.S16-S26, 2006.

[11]. J. Oh, S. Hwang, Y. Cao, W. Tavanapong, D. Liu, J. Wong and P. C. de Groen. “Measuring Objective Quality of Colonoscopy”. IEEE Transactions on Biomedical Engineering. (Accepted for publication at July 2008).

[12]. S. Hwang, J. Oh, W. Tavanapong, J. Wong, and P. C. de Groen. “Stool Detection in Colonoscopy Videos”. In Proc. of Int’l Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), Vancouver, British Columbia, Canada, August, pp. 3004-3007, 2008.

[13] D. Liu, Y. Cao. W. Tavanapong, J. Wong, J. Oh, and P. C. de Groen, "Quadrant Coverage Histogram: A New Method for Measuring Quality of Colonoscopy Procedures". In Proc. of IEEE Engineering in Medicine and Biology Society (EMBC), Lyon, France, pp. 3470-3473, August 2007.

[14] Y. Cao, D. Liu, W. Tavanapong, J. Wong, J. Oh, and P. C. de Groen. “Computer-aided Detection of Biopsy and Therapeutic Operations in Colonoscopy Videos”. IEEE Transactions on Biomedical Engineering, 54(7): 1268-1279, July 2007.

[15] Danyu Liu, Yu Cao, Wallapak Tavanapong, Johnny Wong, JungHwan Oh, and Piet C. de Groen, "Mining Colonoscopy Videos to Measure Quality of Colonoscopic Procedures," Proc. Of Fifth IASTED International Conference on Biomedical Engineering, Innsbruck, Austria, 2007, pp. 409-414.

[16] S. Hwang, J. Oh, W. Tavanapong, J. Wong, and P. C. de Groen. "Polyp Detection in Colonoscopy Video Using Elliptical Shape Feature". In Proc. of IEEE Int'l Conf. on Image Processing. pp. 465-468. San Antonio, Texas, September 2007.

[17] Y. Wang, W. Tavanapong, J. Wong, J. Oh, and P. C. de Groen. “Edge Cross-Section Features for Detection of Appendiceal Orifice Appearance in Colonoscopy Videos”. In Proc. of Int’l Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), Vancouver, British Columbia, Canada, August, 2008. pp. 3000-3003.

[18] A. Kaufman and J. Wang. “3D Surface Reconstruction from Endoscopic Videos”. Mathematics and Visualization, Springer Berlin Heidelberg, pp.61-74, 2007.

[19] D. Koppel, C. Chen, Y. Wang, H. Lee, J. Gu, A. Poirson, and R. Wolters. “Toward Automated Model Building from Video in Computer-Assisted Diagnoses in Colonoscopy”. In Proc. of SPIE, San Diego, CA, USA, Vol. 6509, 65091L, 2007.

[20] G. N. Khan, D. F. Gillies. “Vision Based Navigation System for an Endoscope”. Image and vision computing, Elsevier, Oxford, ROYAUME-UNI, Vol.14, pp.763-772, 1983.

[21] S. Hwang, J. Oh, J. Lee, Y. Cao, W. Tavanapong, D. Liu, J. Wong, and P. C. de Groen, Automatic measurement of quality metrics for colonoscopy videos, in Proceedings of the 13th annual ACM international conference on Multimedia, Hilton, Singapore, 2005, 912 - 921.

[22] P.A. Brathwaite, K.B. Chandran, D.D. McPherson, and E.L. Dove, "Lumen detection in human IVUS images using region-growing," Computers in Cardiology, pp. 37-40, 1996.

[23] G. N. Khan, "A highly parallel shade image segmentation method," Proc. of International Conference. on Parallel Processing for Computer Vision and Display, University of Leeds, 1988.

[24] S. Kumar, K. Vijayan Asari, and D. Radhakrishnan, "A New Technique for the Segmentation of Lumen from Endoscopic Images by Differential Region Growing," Proc. of 42nd Midwest Symposium on Circuits and Systems, New Mexico, 1999.

[25] H. Tian, T. Srikanthan, and K. Vijayan Asari, "Automatic segmentation algorithm for the extraction of lumen region and boundary from endoscopic images," Medical and Biological Engineering and Computing, vol. 39, pp. 8-14, 2001.

[26] J. Oh, S. Hwang, J. Lee, W. Tavanapong, P. C. de Groen, and J. Wong. “Informative Frame Classification for Endoscopy Video”. Journal of *Medical Image Analysis* 2007 Apr;11(2):110-27, Feb 2007.

[27] Piatetsky-Shapiro, G. (1991), Discovery, analysis, and presentation of strong rules, in G. Piatetsky-Shapiro & W. J. Frawley, eds, ‘Knowledge Discovery in Databases’, AAAI/MIT Press, Cambridge, MA.

[28] R. Agrawal; T. Imielinski; A. Swami: Mining Association Rules Between Sets of Items in Large Databases", SIGMOD Conference 1993: 207-216

[29] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules in large databases. In Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo, editors, Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September 1994.

[30] M. Houtsma and A. Swami. Set-oriented mining of association rules. Research Report RJ 9567, IBM Almaden Research Center, San Jose, California, October 1993.

[31] Fernando Berzal, Juan C. Cubero, Nicolas Marín, José-María Serrano, "TBAR: An efficient method for association rule mining in relational databases," Data & Knowledge Engineering, Vol. 37, No. 1, 2001, pp. 47-64.

[32] Sergey Brin, Rajeev Motwani, Jeffrey D. Ullman, Shalom Tsur, "Dynamic itemset counting and implication rules for market basket data", Proc. of the ACM SIGMOD Int’l Conf. on Management of Data, Tucson,AZ, USA, 1997.

[33] Toon Calders and Bart Goethals, "Mining All Non-Derivable Frequent Itemsets", Proc. of the 6th European Conf. on Principles of Data Mining and Knowledge Discovery, 2002, pp. 74-85.

[34] N.Pasquier, Y.Bastide, R.Taouil, and L.Lakhal, "Efficient mining of association rules using closed itemset lattices," Information Systems, Vol. 24, No. 1, 1999, pp. 25-46.

[35] Rohit Gupta, Brian N. Brownlow, Robert A. Domnick, Gavin Harewood, Michael Steinbach, Vipin Kumar, Piet C. de Groen, Colon Cancer Not Prevented By Colonoscopy, American College of Gastroenterology (ACG) Annual Meeting, 2008