Measuring Objective Quality of Colonoscopy

JungHwan Oh*, Sae Hwang, Yu Cao, Wallapak Tavanapong, Danyu Liu, Johnny Wong, and Piet C. de Groen

Abstract—Advances in video technology are being incorporated into today’s healthcare practices. Colonoscopy is regarded as one of the most important diagnostic tools for colorectal cancer. Indeed, colonoscopy has contributed to a decline in the number of colorectal-cancer-related deaths. Although colonoscopy has become the preferred screening modality for prevention of colorectal cancer, recent data suggest that there is a significant miss rate for the detection of large polyps and cancers, and methods to investigate why this occurs are needed. To address this problem, we present a new computer-based method that analyzes a digitized video file of a colonoscopic procedure and produces a number of metrics that likely reflect the quality of the procedure. The method consists of a set of novel image-processing algorithms designed to address new technical challenges due to uncommon characteristics of videos captured during colonoscopy. As these measurements can be obtained automatically, our method enables future quality control in large-scale day-to-day medical practice, which is currently not feasible. In addition, our method can be adapted to other endoscopic procedures such as upper gastrointestinal endoscopy, endoscopy, and bronchoscopy. Last but not least, our method may be useful to assess progress during colonoscopy training.

Index Terms—Colonoscopy, quality measurement metrics, therapeutic and diagnostic operation detection, video segmentation.

I. INTRODUCTION

COLORECTAL cancer is the second leading cause of all cancer deaths behind lung cancer in the United States [9]. 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 allows for inspection of the entire colon and provides the ability to perform a number of therapeutic operations such as polyp removal during a single procedure. A colonoscopic procedure consists of two phases: insertion phase and withdrawal phase. During the insertion phase, a flexible endoscope (a flexible tube with a tiny video camera at the tip) is advanced under direct vision via the anus into the rectum, and then gradually into the most proximal part of the colon or the terminal ileum. In the withdrawal phase, the endoscope is gradually withdrawn [5], [11]. The main purpose of the insertion phase is to reach the end of the colon, whereas in the withdrawal phase, careful inspection of all visible mucosa, tissue sampling, polyp removal, etc., are performed.

As mentioned before, there is a significant miss rate for the detection of even large polyps and cancers [7], [10]. The miss rate varies among endoscopists and may be related to the experience of the endoscopist and the location of the lesion in the colon, but no study explaining the cause of the miss rate is available. Even though the demand for quality control for colonoscopic procedures is increasing, no manual protocol nor automated system is available to produce detailed measurements of quality of colonoscopy. 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. Current American Society for Gastrointestinal Endoscopy (ASGE) and American College of Gastroenterology (ACG) guidelines suggest that on average, the withdrawal phase during a screening colonoscopic procedure should last a minimum of 6 min, and the visualization of cecum anatomical landmarks such as the appendiceal orifice and the ileocecal valve should be documented [12]. This paper presents new objective metrics that reflect quality of colonoscopy as well as algorithms that analyze a colonoscopy video to generate these objective metrics. In order to measure the metrics, we follow the process in Fig. 1.

First, we detect and discard out-of-focus (blurred) frames from the videos. Current endoscopes are equipped with a single, wide-angle lens, and do not have typical video camera functions such as optical zoom (zoom-in, zoom-out) or auto focusing. Because of the limitation of endoscopes, a significant number of out-of-focus (blurred) frames are included in colonoscopy videos. The out-of-focus (blurred) frames do not hold any useful information. We define an out-of-focus (blurred) frame as a noninformative frame (Fig. 2, left image) and an in-focus (clear) frame as an informative frame (Fig. 2, right image). We call the procedure to detect noninformative frames from the endoscopy videos as frame classification.

Second, we estimate the boundary between insertion and withdrawal phases based on camera motions. The insertion phase does not always consist of continuous forward camera...
motions. Similarly, the withdrawal phase does not always consist of continuous backward camera motions. This is due to the fact that back-and-forth movements allow “shortening” of the endoscope during insertion (removal of loops) and optimize views of interesting regions such as polyps, cancers, terminal ileum, crowfoot with appendix, ilo-cecal valve, etc., during withdrawal. Hence, either phase has an arbitrary number, and combination of forward and backward camera motions while the dominant camera motions of insertion and withdrawal phases are forward and backward, respectively.

Third, we segment a colonoscopy video based on the camera motions such as forward and backward, which are called oral direction and anal direction, respectively. We define a camera motion shot as a sequence of consecutive frames with a single direction of camera motion. A camera motion shot can be either an oral shot, which represents the camera motion from the anus to the terminal ileum (forward camera motion), or an anal shot, which represents the camera motion from the terminal ileum to the anus (backward camera motion). By accumulating the values of camera motions in the oral and anal shots within an entire video, and finding a peak value, we can locate the end of the insertion phase and compute four metrics for quality evaluation.

Fourth, we identify an operation shot—a segment of visual data in a colonoscopy video that corresponds to a diagnostic or therapeutic operation. We add the time for each operation shot during the withdrawal phase and use this value to derive a fifth metric, which is the clear, operation-free withdrawal time. Our contributions are summarized as follows.

1) We are the first to investigate an automatic measurement method that generates a number of objective metrics to evaluate the quality of colonoscopic procedures. Our metrics are developed based on ASGE and ACG criteria and our domain expertise. No prior research has investigated objective quality measurement methods for colonoscopy or other endoscopic procedures.

2) To calculate the proposed quality metrics, we address a number of technical challenges. Based on the analysis of camera motions, we introduce new video segmentation techniques to segment a colonoscopy video into oral and anal shots, and the insertion and withdrawal phases. We evaluate our proposed techniques on real colonoscopy videos. Our experimental results show that these techniques are promising.

The remainder of this paper is organized as follows. Background of our previous techniques relevant to this paper is discussed in Section II. The video segmentation technique based on camera motion estimation is presented in Section III. In Section IV, we present the formal definitions of the quality metrics. We discuss our experimental results in Section V. Finally, Section VI presents some concluding remarks.

II. BACKGROUND ON PREVIOUS WORK

We summarize the techniques that we have developed that are relevant to this paper.

A. Frame Classification

Endoscopy videos including colonoscopy videos contain many noninformative frames due to three reasons. The first reason is the frequent shifts of camera while moving along the organ of interest (motion blur). The second reason is that current endoscopes are equipped with a single, wide-angle lens that cannot be focused (out-of-focus blur). The third reason is material (e.g., diluted stool, water) partially blocking light from reaching the lens (interference blur). These noninformative frames do not contain useful information and a significant number of noninformative frames are included in this type of videos [8]. There are existing techniques [3], [6], [13] to handle out-of-focus images using image restoration. However, these existing techniques are not applicable to endoscopy video frames because these techniques need a reference image to compute the quality of the test image, and we only have test images. We previously proposed a frame classification technique that does not require a reference image. The technique is briefly summarized shortly. More details can be found in [8]. Each image is classified as either informative or noninformative. The frame classification is based on the different patterns of the frequency spectrums of informative and noninformative frames. For instance, the noninformative frame [Fig. 3(a)] does not contain clear object information except the four strong edges at the corners of the image running approximately at 45° relative to the sides. Its Fourier spectrum [Fig. 3(b)] shows prominent components along the 45° directions that correspond to the four corners of the image. Compared to the noninformative frame, the informative frame [Fig. 3(c)] has a lot of clear edge information so that its spectrum [Fig. 3(d)] does not show prominent components along the 45° directions because it has a wider range of bandwidth from low to high frequencies. Texture analysis is then applied on the frequency spectrum image, and seven well-known texture features, entropy, contrast, correlation, homogeneity, dissimilarity, angular second moment, and energy, are extracted for each image. We utilize a two-step K-means method to classify frames into two groups (i.e., the informative frame group and the noninformative frame group) based on the texture features of these frames.

B. Diagnostic and Therapeutic Shot Detection

Recall that an operation shot is defined as a segment of visual data in a colonoscopy video that corresponds to a diagnostic or therapeutic operation. We previously proposed a technique that detects an operation shot based on template matching of shape features of the cable of diagnostic or therapeutic instruments [1]. The challenges include the following. First, cables
come in different colors and materials. Second, cables appear very bright in many frames. Third, depending on the type and brand of the endoscopy instrument, a cable can appear from the edge of the image, anywhere within the 360° of a circle. Fourth, the appearance of a cable in a frame varies from one frame to another during an operation. Our technique overcoming these challenges works as follows. First, each image is enhanced by smoothing small light-reflection areas and then segmented into a number of regions. Some parts of a cable appear very bright, causing the cable to be segmented into more than one region. We combine falsely segmented regions into one potential cable region based on texture and region locations. Next, candidate regions are matched with predefined templates of cables. Finally, we group a sequence of images of which some have a cable detected into the same operation shot according to our predefined rules.

In our experiments, we measured the true positive fraction (TPF) defined as the ratio of the total number of correctly detected images as part of actual operation shots (true positives) to the total number of images of actual operation shots, and the false positive fraction (FPF) defined as the ratio of incorrectly detected images as part of operation shots (false positives) to the total number of images of actual operation shots. Averages of true positive and false positive fractions on 25 colonoscopy videos are 0.94 and 0.10, respectively. More details can be found in [1].

III. VIDEO SEGMENTATION BASED ON CAMERA MOTION ESTIMATION

In this section, we present the video segmentation technique based on camera motions (especially, forward and backward camera movements) in compressed moving picture experts group (MPEG) videos since the provided colonoscopy videos are in MPEG format. Based on the estimated camera motions, the video is segmented into oral and anal shots, and insertion and withdrawal phases, as defined previously.

A. Camera Motion Estimation

In this section, we discuss the camera motion estimation in compressed MPEG video. Each frame in an MPEG video consists of a number of macroblocks, and each macroblock may be associated with a motion vector (mv). These motion vectors are already included in MPEG-encoded videos, so we extract the motion vectors directly from the MPEG-encoded colonoscopy video. A motion vector (mv) is defined as the displacement \((d_x,d_y)\) between a macroblock of the current frame and the best matching macroblock in the previous frame within a predefined search radius. A motion vector does not represent the true camera motion if the actual corresponding macroblock of a macroblock in the current frame is placed outside the search radius. Therefore, these outliers need to be discarded because the outliers do not describe the true camera motion in a frame. We apply the outlier detection algorithm proposed in [4]. This method consists of two main steps named smooth change and neighborhood. A motion vector (mv) is declared as an outlier if both the steps declare it as an outlier (see the examples in Fig. 4). All detected outliers motion vectors are then removed. The two steps for outlier detection are explained as follows.

1) Smooth change: The central \(mv\) is compared to each average of four pairs of opposite neighbors. If the distance between the average \(mv\) of each pair and the central \(mv\) is less than a certain threshold, it is considered a supporting pair. In Fig. 4(a), pairs 1 and 3 are supporting pairs, so the number of supporting pairs is 2. If the number of supporting pairs is below a threshold, the central \(mv\) is declared an outlier. For our experiments, we use a value 3 for the threshold of supporting pairs.

2) Neighborhood: A neighborhood motion vector supports the central \(mv\) if it lies within a tolerance angle [see Fig. 4(b)]. If the number of supporting vectors is below a threshold, then the central \(mv\) is declared as an outlier. For our experiments, we use a value 4 for the threshold of supporting vectors.

For camera motion estimation, we employed the affine model for the following reasons, realizing that there are more accurate models of higher order [2], [14]. First, the affine model is known to be more resilient to noise and sparse motion vector fields. Second, it can represent all basic types of camera motions. As seen in Fig. 5, six camera motions can be defined as follows.
All six camera motions can be expressed in the affine model as follows:

\[
\begin{pmatrix}
  d_x \\
  d_y 
\end{pmatrix} = \begin{pmatrix}
  a_1 \frac{d_{\text{dolly}}}{z} & b_1 \frac{d_{\text{dolly}}}{z} \\
  -b_2 \frac{d_{\text{dolly}}}{z} & a_2 \frac{d_{\text{dolly}}}{z} 
\end{pmatrix} \begin{pmatrix}
  x \\
  y 
\end{pmatrix} + \begin{pmatrix}
  e_{\text{pan}} + \frac{e_{\text{track}}}{z} \\
  d_{\text{tilt}} + \frac{d_{\text{boom}}}{z} 
\end{pmatrix}
\]

where \((d_x, d_y)\) is the motion vector of a macroblock located at position \((x, y)\) of each frame, \(z\) is the depth of the real world, \(a_1, a_2, b_1, b_2\) are related scalar coefficients. We note that there is no coefficient related with the zooming camera motion because the endoscopes do not have optical zoom-in and zoom-out functions. Given the motion vectors, we calculate these coefficients using the least-square fitting method. Even though the coefficient values are all related to camera motions, we can obtain reliable camera motions such as dolly camera motion (DCM), rolling camera motion (RCM), horizontal camera motion (HCM = panning + tracking), and vertical camera motion (VCM = tilting + booming) as follows:

\[
\text{DCM} = \frac{1}{2} \left( \frac{a_1 d_{\text{dolly}}}{z} + \frac{a_2 d_{\text{dolly}}}{z} \right) \quad \text{HCM} = e_{\text{pan}} + \frac{e_{\text{track}}}{z}
\]

\[
\text{RCM} = \frac{1}{2} (b_1 d_{\text{dolly}} + b_2 d_{\text{dolly}}) \quad \text{VCM} = d_{\text{tilt}} + \frac{d_{\text{boom}}}{z}
\]

Among the four camera motions, the dolly camera motion (DCM) will mainly be examined because the dolly camera motion is directly related to the forward and backward camera movements. The positive DCM value means forward, oral movement, and the negative DCM value means backward, anal movement.

### B. Video Segmentation

Using the dolly camera motion (DCM), we can easily segment a colonoscopy video into a number of shots, each of which consists of the frames with the same direction of camera motion (oral shot or anal shot). However, the insertion phase does not always consist of continuous forward camera motions. Similarly, the withdrawal phase does not always consist of continuous backward camera motions. Even though all frames do not have the same directional camera movements in each phase, the insertion phase consists of a large number of oral shots (oral direction frames) and a small number of anal shots (anal direction frames), while the withdrawal phase consists of a large number of anal shots (anal direction frames) and a small number of oral shots (oral direction frames). Using these characteristics of colonoscopy videos, we propose the following video segmentation method.

1. **Frame classification step** removes noninformative frames.
2. **Camera motion estimation step** calculates four camera motions (DCM, RCM, HCM, and VCM) for the informative frames.
3. **Unreliable DCM value filtration step** filters out unreliable DCM values as follows.
   a) If two consecutive frames are very different (an “abrupt change”), e.g., due to the MPEG frame, it may consist of a small number of macroblocks with motion vectors and a large number of macroblocks without motion vectors. The estimated camera motions using this MPEG frame may not be correct. We remove this type of errors by assigning DCM = 0 if the number of motion vectors is less than a certain threshold value \(mv < TH_d\). In our experiments, we use a value 10 for \(TH_d\).
   b) The DCM tends to have an incorrect value when other camera motions such as HCM or VCM have larger values compared with DCM. To reduce this type of errors, we assign DCM = 0 if the ratio of the magnitudes of HCM and VCM to DCM \((\sqrt{HCM^2 + VCM^2})/DCM\) is larger than a certain threshold \((TH_d)\). In our experiments, we use a value 1500 for \(TH_d\).
   c) Temporal information is utilized to filter out incorrect DCMs. It is highly likely that any oral or anal shots have more than two frames (we are using 30 frames/s rate videos). Therefore, we assign DCM = 0 if the number of consecutive frames with the same direction is less than a certain threshold \((TH_d)\). In our experiments, we use a value 2 for \(TH_d\).
4. **Shot boundary detection step** detects shot boundaries of a colonoscopy video based on camera movements. As seen in Fig. 6, a colonoscopy video can be decomposed into a number of pieces \((P_1, P_2, \ldots, P_i, \ldots)\) by the noninformative frame filtration. Each piece consists of a number of frames with three different kinds of DCM values: frames with positive DCM values, frames with negative DCM values, and frames with DCM = 0. Based on the DCM values of frames in a piece, there are two different shots as follows.
   a) **Oral shot**: A sequence of frames with positive DCM.
   b) **Anal shot**: A set of frames with negative DCM.

We note that a DCM value of frame with DCM = 0 follows the value of previous frame. Let us assume that the current frame is a frame with DCM = 0. If the DCM value of the previous frame is positive, the DCM value of the current frame becomes positive; otherwise, it becomes negative. Frames with the different directions of camera
motions (frames with positive DCM values and frames with negative DCM values) cannot be in the same shot.

5) **Phase boundary detection step** detects the boundary between the insertion phase and the withdrawal phase using the accumulated DCM. When we add up all DCM values, the accumulated DCM value will increase until the end of the insertion phase because most of the frames in the insertion phase have forward movements (i.e., positive DCM values). However, the accumulated DCM value will decrease during the withdrawal phase because most of the frames in the withdrawal phase have the backward movement (i.e., negative DCM values). For this reason, the boundary frame between the insertion phase and the withdrawal phase has the highest accumulated DCM values.

Fig. 6 shows an example of video segmentation obtained using our shot and phase segmentation method.

**IV. QUALITY METRICS**

In this section, we discuss five metrics proposed to evaluate the quality of colonoscopy. The information to calculate the quality metrics is obtained from our video segmentation techniques and the technique for operation shot detection described in the previous sections. We note that the superscript of each metric represents the phases I and W (i.e., I for insertion phase and W for withdrawal phase) unless defined otherwise.

**Metric 1:** The insertion time (IT) can be measured as follows:

\[ \text{IT} = \frac{\text{NF}_I}{\text{FER}} \]

where \( \text{NF}_I \) represents the number of frames in the insertion phase and FER is frame extraction rate, which is the number of frames extracted for analysis per unit time from a video. FER is expressed in frames per second and we use 30 frames/s for our videos.

**Metric 2:** The withdrawal time (WT) is the duration of the withdrawal phase. We calculate WT as follows:

\[ \text{WT} = \frac{\text{NF}_W}{\text{FER}} \]

where \( \text{NF}_W \) represents the number of frames in the withdrawal phase.

**Metric 3:** Even though the duration of the withdrawal phase is long, the quality of the colonoscopic procedure may not be good if a lot of noninformative frames are present in the withdrawal phase. By adding up the duration of only the informative frames in the withdrawal phase, we can obtain the clear withdrawal time (CWT) and the clear withdrawal ratio (CWR) as follows:

\[ \text{CWT} = \frac{\text{NIF}_W}{\text{FER}} \hspace{1cm} \text{CWR} = \frac{\text{CWT}}{\text{WT}} \]

where \( \text{NIF}_W \) represents the number of informative frames in the withdrawal phase.

**Metric 4:** During the withdrawal phase, an endoscopist may move a camera back and forth to examine suspicious regions. This movement may be an indicator of quality as the endoscopist is trying to verify that an area is indeed free of lesions on multiple inspections. These movements can be estimated by measuring the number of camera motion changes in the withdrawal phase (NCMC), and the ratio of the number of camera motion changes to the clear withdrawal time (RCMC) as follows:

\[ \text{NCMC} = \frac{\text{NS}_W}{\text{FER}} \hspace{1cm} \text{RCMC} = \frac{\text{NCMC}}{\text{CWT}} \]

where \( \text{NS}_W \) represents the number of oral and anal shots in the withdrawal phase.

**Metric 5:** Clear, operation-free withdrawal time (COWT) measures the duration of the withdrawal phase without noninformative images and without diagnostic or therapeutic operations. This metric provides true inspection time without time spent for diagnostic or therapeutic operations as follows:

\[ \text{COWT} = \text{CWT} - \frac{\text{NOF}_W}{\text{FER}} \]

where \( \text{NOF}_W \) represents the number of frames belonging to detected diagnostic or therapeutic operations during the withdrawal phase using our operation shot detection technique introduced in Section II.

**V. EXPERIMENTAL RESULTS**

In this section, we report our experimental results using real colonoscopy videos.

A. Evaluations of Shot Segmentation and Phase Segmentation

In this section, we evaluate the effectiveness of our frame classification and shot segmentation techniques with real colonoscopy videos. These videos were randomly selected from a database of colonoscopy videos obtained during normal screening colonoscopy. The videos in our test set consist of 249,759 frames (30 frames/s) with a total time of 2 h and 18 min. The details are shown in Table I. The first column “ID” represents the unique ID number for each colonoscopy video, and the third column “Total Frames” represents the total number of frames for each video. Table II shows the performance of the shot segmentation technique based on camera motions. Column “S” represents the number of actual shot boundaries as manually determined. An actual shot boundary happens between two
consecutive frames in the transition from the forward camera motion to the backward camera motion and vice versa. Column “T” represents the number of detected shot boundaries and column “C” represents the number of correctly detected shot boundaries. The last two columns show the precision and recall for our shot segmentation techniques, which are very promising.

Using our video segmentation technique presented in Section III, we can detect the boundary between the end of the insertion phase and the start of the withdrawal phase, which, in the vast majority of cases, is characterized by the presence of terminal ileum, crowfoot with appendix, or ileo-cecal valve. Table III shows the performance of the phase detection technique. Column “APB (actual phase boundary)” represents the times of the actual phase boundaries as determined manually. Column “DPB (detected phase boundary)” represents the times of the detected phase boundaries by our algorithm. Column “Time Difference” represents the absolute time difference between the actual phase boundary and detected phase boundary. The maximum time difference is 28 s and the average time difference is 15 s, which are very promising.

**B. Evaluation of Quality Metrics**

We generated the quality metrics as presented in Section IV. Using the noninformative frame classification and video segmentation techniques, we obtained the following information from seven colonoscopy videos (Table IV). Column “NF” represents the number of frames; column “NIF” represents the number of informative frames; and column “NS” represents the number of shots. As before, the superscript of each column represents the phases I and W (i.e., I for insertion phase and W for withdrawal phase). The generated quality metrics from automatic analysis of videos in Table IV are shown in Table V. We discuss these results as follows.

**Metric 1:** We computed the insertion time IT = NF^I/30 because our videos have a frame rate of 30 frames/s. The colonoscopy video Colon2 has a short insertion time (about 5 min) and colonoscopy video Colon4 has the longest insertion time (about 15 min). Fecal material was found during the insertion phase of colonoscopy video Colon4, making it more difficult for the endoscopist to reach the proximal end of the colon and explaining the higher IT value.

**Metric 2:** The withdrawal time (WT) is computed as WT = NF^W/30. Current ASGE and ACG guidelines suggests that, on average, the withdrawal time should last a minimum of 6 min. The details as listed in the third column of Table V show that, for these colonoscopies, that is essentially the case.

**Metric 3:** We measured the clear withdrawal time (WT = NIF^W/30) and the ratio of the clear withdrawal time to the withdrawal time (CWR = CWT/WT). The colonoscopy video Colon5 also has the shortest clear withdrawal time (i.e., 3 min and 57 s) among seven colonoscopy videos. However, this video has more clear frames than other videos because it...
has the highest ratio of the clear withdrawal time to the entire withdrawal time (CWR = 0.692). The details are found in the fourth and fifth columns of Table V.

**Metric 4:** We measure the number of the camera motion changes (NCMC = NS) and the ratio of the number of the camera motion changes to the clear withdrawal time (RCMC = NCMC/CWT). There are some regions in colonoscopy videos Colon2 and Colon7 that the endoscopist apparently cannot see well, so the endoscopist frequently moves the endoscope back and forth to examine these regions in order to get the best possible view. Colonoscopy video Colon7 has higher values of NCMC and RCMC than the other videos, so we can expect that the colonoscopy video Colon7 represents a colon that is different from the other two colons and may contain more angulations or haustrae, which require more efforts in order to achieve optimal mucosal inspection. The details are presented in the sixth and seventh columns of Table V.

**Metric 5:** We compute the clear, operation-free withdrawal time as COWT = CWT − NOF/W/30. Five of the videos (videos Colon1, Colon3, Colon5, Colon6, and Colon7) do not have any operation shot. Our algorithms generate one false shot for video Colon5. The reason for this is that some colon images in this video have a light reflection region whose shape is similar to the shape of the cable of an instrument. Our technique incorrectly declares these images as instrument images. For the rest of the videos, our algorithms found all operation shots and did not generate any false shots. The eighth column of Table V shows the COWT values for each video.

**VI. CONCLUDING REMARKS**

In this paper, we propose objective quality metrics from automated analysis of colonoscopy videos. These metrics may be used to evaluate the endoscopist’s skill and the quality of colonoscopic procedures. To compute these metrics automatically, we introduce a new colonoscopy video segmentation technique based on frame classification and camera motion. Our experiments on real colonoscopy videos demonstrate that our method indeed can be used to automatically measure a number of metrics from digitized colonoscopy video files. For cases where the endoscopist fails to reach the proximal end of the colon, we do identify the location of maximal extent.

We have considered combining metrics in a single quality score. This, however, is not possible, as several components make up the final quality of a colonoscopy. Some are patient-related (preparation of the colon), some are equipment-related (quality, color, and contrast of the image, automatic light adjustment characteristics of the light source), and many are endoscopist-related (maximal extent of insertion, withdrawal time, effort of inspection of the visible mucosa, instrumentation technique, appropriate selection of instruments for specific lesions). Therefore, a number of different quality measures will be needed to describe the quality of a procedure as complex as colonoscopy.

Our final goal is to test our quality control metrics in routine colonoscopy practice and determine which metrics predict a good long-term outcome. For instance, it seems likely that longer CWT and CWR are associated with a lower chance for missed polyps or cancers. Until now, studies aimed at assessing detailed components of the withdrawal phase were not possible as manual annotation was technically not feasible and financially not viable. At present our quality metrics are not generated in real-time; however, future implementation of some of our metrics in real-time is in development. Conversion from postprocedure to real-time analysis will not only provide quality measures and guide physicians, but may also directly benefit every patient being examined in the presence of our technology.

**ACKNOWLEDGMENT**

This work was done while S. Hwang was at the University of North Texas, Denton, and Y. Cao was at Iowa State University, Ames.

**REFERENCES**


Authors’ photographs and biographies not available at the time of publication.