Available online at www.sciencedirect.com
ScienceDirect
Medical Image Analysis 11 (2007) 110–127
MEDICAL IMAGE ANALYSIS
www.elsevier.com/locate/media

Informative frame classification for endoscopy video

JungHwan Oh [a,*], Sae Hwang [b], JeongKyu Lee [b,1], Wallapak Tavanapong [c], Johnny Wong [c], Piet C. de Groen [d]

[a] Department of Computer Science and Engineering, University of North Texas, P.O. Box 311366, NTRP F274, Denton, TX 76203, USA
[b] Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76019, USA
[c] Computer Science Department, Iowa State University, Ames, IA 50011, USA
[d] Mayo Clinic College of Medicine, Rochester, MN 55905, USA

Received 2 May 2005; received in revised form 6 October 2006; accepted 27 October 2006
Available online 27 February 2007

Abstract

Advances in video technology allow inspection, diagnosis and treatment of the inside of the human body without or with very small scars. Flexible endoscopes are used to inspect the esophagus, stomach, small bowel, colon, and airways, whereas rigid endoscopes are used for a variety of minimal invasive surgeries (i.e., laparoscopy, arthroscopy, endoscopic neurosurgery). These endoscopes come in various sizes, but all have a tiny video camera at the tip. During an endoscopic procedure, the tiny video camera generates a video signal of the interior of the human organ, which is displayed on a monitor for real-time analysis by the physician. However, many out-of-focus frames are present in endoscopy videos because current endoscopes are equipped with a single, wide-angle lens that cannot be focused. We need to distinguish the out-of-focus frames from the in-focus frames to utilize the information of the out-of-focus and/or the in-focus frames for further automatic or semi-automatic computer-aided diagnosis (CAD). This classification can reduce the number of images to be viewed by a physician and to be analyzed by a CAD system. We call an out-of-focus frame a non-informative frame and an in-focus frame an informative frame. The out-of-focus frames have characteristics that are different from those of in-focus frames. In this paper, we propose two new techniques (edge-based and clustering-based) to classify video frames into two classes, informative and non-informative frames. However, because intensive specular reflections reduce the accuracy of the classification we also propose a specular reflection detection technique, and use the detected specular reflection information to increase the accuracy of informative frame classification. Our experimental studies indicate that precision, sensitivity, specificity, and accuracy for the specular reflection detection technique and the two informative frame classification techniques are greater than 90% and 95%, respectively.

Keywords: Endoscopy; Colonoscopy; Clustering; Texture; Frame classification; Specular reflection detection

1. Introduction

Advances in video technology are being incorporated into today's healthcare practice. Various types of endoscopes are used for colonoscopy, upper gastrointestinal endoscopy, enteroscopy, bronchoscopy, cystoscopy, laparoscopy, wireless capsule endoscopy and some minimal invasive surgeries (i.e., video endoscopic neurosurgery). These endoscopes come in various sizes, but all have a tiny video camera at the tip of the endoscopes. During an endoscopic procedure, the tiny video camera generates a video signal of the interior of the human organ, which is displayed on a monitor for real-time analysis by the physician. Endoscopy of the colon or colonoscopy is currently the accepted "gold standard" technique for prevention and early detection of colorectal cancer. In the US, colorectal cancer is the second leading cause of all cancer deaths behind lung cancer (Society, 2005). Colonoscopy allows for the inspection of the entire colon: a flexible endoscope

* Corresponding author.
E-mail address: jhoh@cse.unt.edu (J. Oh).
[1] Present address: Department of Computer Science and Engineering, University of Bridgeport, Bridgeport, CT 06604, USA.

1361-8415/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.media.2006.10.003

(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 (Meyerhardt and Mayer, 2005; Phee and Ng, 1998; Sucar and Gillies, 1990; Khessal and Hwa, 2000; Dario and Lencioni, 1997).

There are many out-of-focus frames in colonoscopy videos since current endoscopes are equipped with a single, wide-angle lens that cannot be focused. We define an *out-of-focus* frame as a *non-informative frame* (Fig. 1) and an *in-focus* frame as an *informative frame* (Fig. 2). The non-informative frames are usually generated due to two main reasons: too-close (or too-far) focus into (from) the mucosa of colon, for example by rapidly moving through the intracolonic space (Fig. 1a and b), or foreign substances (i.e., stool, cleansing agent, air bubbles, etc.) covering camera lens (Fig. 1c and d). We call the procedure that distinguishes informative frames from non-informative frames *Informative Frame Classification for Endoscopy Video* in this paper. We propose two new techniques to distinguish informative frames from non-informative frames based on the detected edges, and discrete Fourier transform (DFT) with clustering, respectively. The edge-based approach is relatively simple and easy to implement, but sensitive to the selected threshold values. The DFT with clustering approach addresses the drawbacks of the edge-based approach, and provides more robust and accurate results.

However, most informative and non-informative frames have some over-reflected areas as seen in the white areas of Figs. 3 and 4. These areas are called *specular reflections* (or *highlights*) (Klinker et al., 1990). The color of every pixel from an object can be described as a linear combination of the object color and its reflection. The object color is a diffuse reflection from the body of the material, and the specular reflection is a stronger reflection (a brighter spot) in one viewing direction from the object surface. The specular reflection is readily apparent on shiny surfaces, which disturb the distinction of informative frames from non-informative ones because these areas can be interpreted as informative contents. Therefore, we need to reduce the effect of the specular reflection to increase the performance of our informative frame classification techniques. For this purpose, we propose using a new technique considering multiple thresholding and outlier detection to determine specular reflections in each frame.

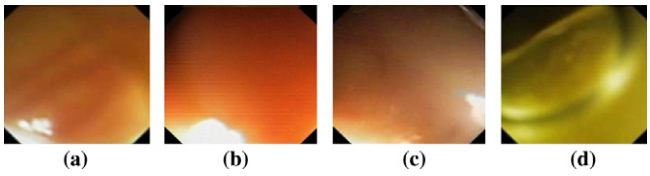


Fig. 1. Examples of non-informative frames.

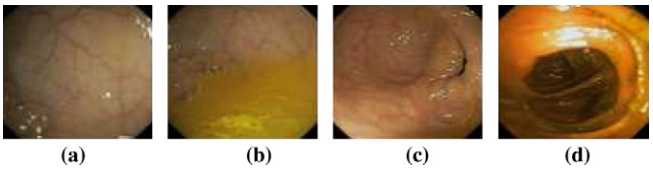


Fig. 2. Examples of informative frames.

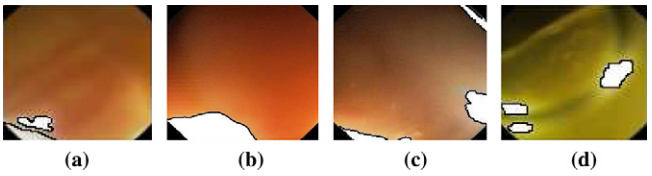


Fig. 3. Specular reflections of non-informative frames in Fig. 1.

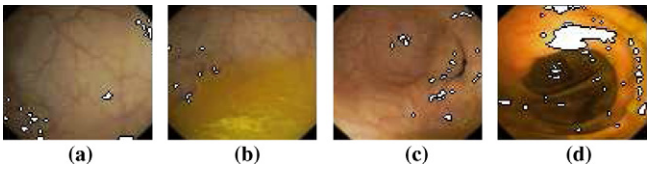


Fig. 4. Specular reflections of informative frames in Fig. 2.

The output of endoscopy video frame classification provides information (i.e., frames that are informative) that will be used for further automatic or semi-automatic computer-aided diagnosis (CAD). It can reduce the number of images to be viewed by a physician and to be analyzed by a CAD system.

The contribution of our proposed techniques can be summarized as follows.

- Typically, a reference image is required to decide the quality (i.e., informative and non-informative) of an image. However, reference images are not available for a specific patient (each patient and each colon is unique). We propose two techniques that are able to evaluate the quality of an image without a reference image.
- Since we do not use any domain knowledge of the video, the proposed technique is domain independent. Hence, it can be used for other medical videos such as upper gastrointestinal endoscopy, enteroscopy, bronchoscopy, cystoscopy, and laparoscopy.
- Specular reflections may considerably disturb human examination and computer-aided analysis so we propose a new technique to detect these with very high accuracy.

The remainder of this paper is organized as follows. First, two techniques for endoscopy video frame classification (Edge-based and Clustering-based), are introduced in Sections 2 and 3, respectively. We explain our specular reflection detection technique in Section 4 to increase the accuracy of the two techniques for endoscopy video frame classification. The performance study is reported in Section 5. Finally, Section 6 presents some concluding remarks.

2. Edge-based frame classification

There are existing techniques (Kundur and Hatzinakos, 1996; Ayers and Dainty, 1988; McCallum, 1990; Bates and Jiang, 1991; Nakagaki and Katsaggelos, 2003; Pai and Bovik, 2001; Giannakis and Heath, 2000) 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 as already stated we only have test images. In this section, we propose a technique to distinguish informative frames from non-informative ones based on a property of isolated edge pixels.

We detect the edges from each frame using Canny Edge Detector (Canny, 1986). Canny Edge Detector first smoothes an image to eliminate noises based on the Gaussian model. Then, it tracks along the local maxima of the gradient magnitudes (edge strengths) of an image, and sets to zero all pixels that are not actually the local maxima, which is known as non-maximal suppression. These two processes generate a single thin line for each edge when an image contains clear edge information. But they generate many isolated pixels when an image does not contain any clear edge information. Examples of the edge detection results are shown in Fig. 5, in which Fig. 5b and c are the images generated from applying the Canny Edge Detector on the image in Fig. 5a. Fig. 5f and g shows images generated from the image in Fig. 5e. The parameters for the edge detector to generate images (b) and (f) are the same, but different from those used to generate images (c) and (g). As shown in this figure, the edge lines of the non-informative images are blurry, but those of the informative images are clear regardless of the parameters used. The blurry lines occur due to discontinuity of the edge pixels constituting a line as seen in Fig. 5d and h. Hence, to distinguish the blurry lines from the clear ones, we defined two terms, *number of isolated pixels* (*IP*) and *isolated pixel ratio* (IPR) for a frame as follows. An *IP* is a number of isolated edge pixels (edge pixels that are not connected to any other edge pixels) in a frame. We computed IPR as the percentage of the number of isolated edge pixels to the total number of edge pixels in the frame:

$$\mathrm{IPR} = \frac{\text{Number of isolated pixels (IPs)}}{\text{Total number of pixels}} \times 100\ (\%) \qquad (1)$$

The frame with the value of IPR greater than a certain threshold is declared a non-informative frame. Otherwise, the frame is considered an informative frame. However, there are some ambiguous images that can be either informative or non-informative according to the threshold value as seen in Fig. 6a. This is because some images may have some parts that are blurry and other parts that are clear. For instance, in a tangential view along the mucosa, only some parts of the image are clear. To handle these ambiguous images and optimize overall accuracy of frame classification, we propose a two-step approach.

Step 1. We classify frames into three categories: informative frames, non-informative frames and ambiguous frames using two very obvious thresholds for IPR, which are called the upper-threshold (TH_U) and the lower-threshold (TH_L). In other words, if an IPR of

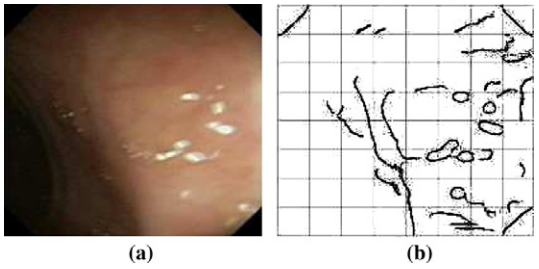


Fig. 6. (a) Ambiguous frame and (b) edge detected from (a) with 64 Blocks.

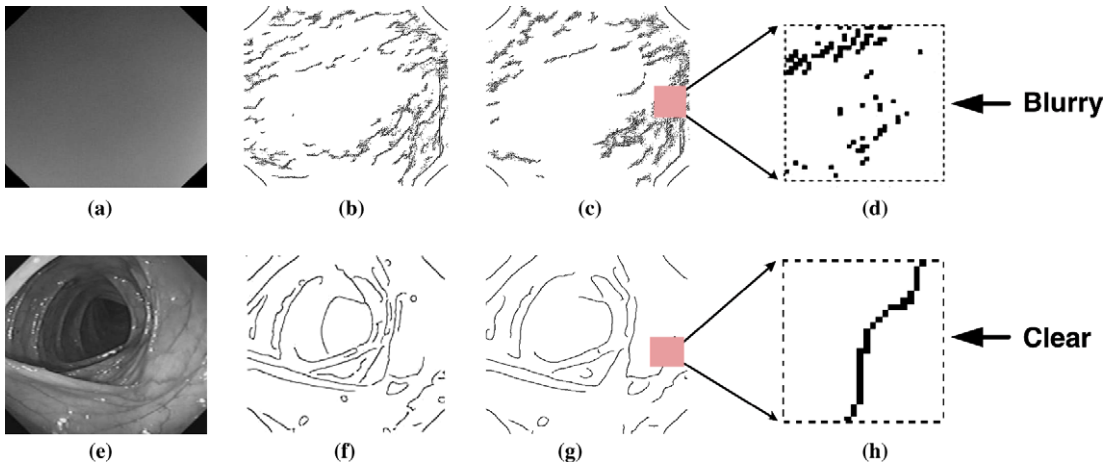


Fig. 5. (a) Non-informative image, (b) and (c) edges detected from (a), (d) details of blurry edge, (e) informative image, (f) and (g) edges detected from (e), and (h) details of clear edge.

an image is larger than the upper-threshold value (TH_U), the image is classified as non-informative. If an IPR of an image is smaller than the lower-threshold value (TH_L), the image is classified as informative. If an IPR of an image is between upper-threshold and lower-threshold, the image is classified as ambiguous, and we proceed to Step 2.

Step 2. An ambiguous frame is divided into a number (64 in our case) of blocks as seen in Fig. 6b. First, each block is classified as empty or non-empty block. An empty block has no pixels. A non-empty block is classified into a clear or blurry block again. For block classification, we use only the lower-threshold value. If a frame has more informative blocks than non-informative ones, then it is classified as an informative frame.

3. Discrete Fourier transform and clustering-based frame classification

The edge-based informative frame classification algorithm shows good performance results (more details are shown in Section 5). However, there is a major drawback in this approach, which is that the performance of our edge-based technique is susceptible to the appropriate values of various parameters (i.e., sigma, high, low, etc.) in the edge detection algorithm, and the upper and lower thresholds in Step 1 and Step 2 of Section 2. To address this, we investigate a new approach based on discrete Fourier transform (DFT), texture analysis, and data clustering. Fig. 7 shows the framework of the proposed algorithm.

3.1. Feature extraction

The basic idea used to detect informative frames comes from discrete Fourier transform (DFT) and texture analysis of their frequency spectrums. The process of DFT for a

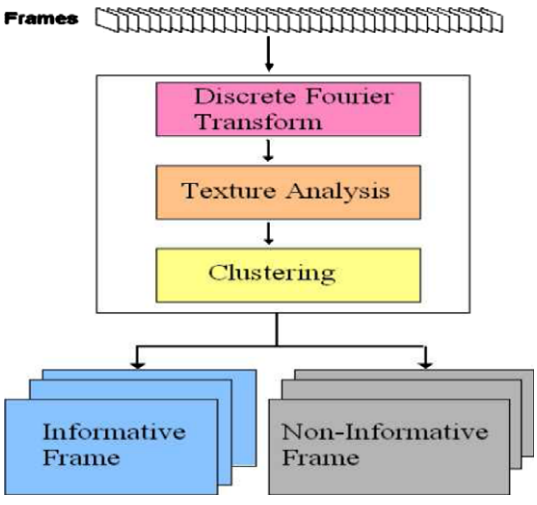


Fig. 7. Framework of informative and non-informative frame classification.

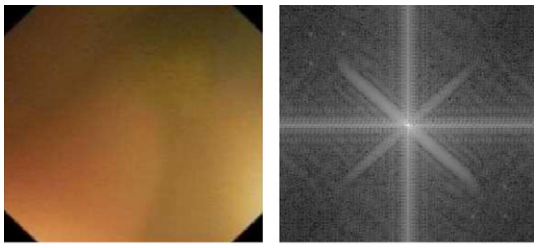
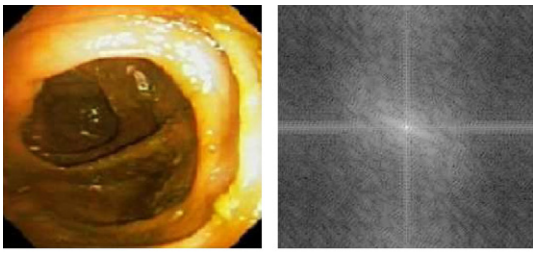


Fig. 8. (a) Non-informative frame and (b) frequency spectrum of (a).

2D image is that first, an image such as Fig. 8a or Fig. 9a is converted into the grayscale image, then the grayscale image is transformed using the Fourier Transform (Ramirez, 1985; Walker, 1996; Sid-Ahmed, 1995; Gonzalez, 2002; Sonka, 1999). The frequency spectrum, 2D plot of the magnitude of the Fourier Transform, is constructed using the coefficients of the Fourier Transform of a grayscale image. The frequency spectrum shows the frequency distribution of an image (Fig. 8b or Fig. 9b). Based on the contents of the image, the frequency spectrums generate different patterns. It is usually impossible to make direct associations between specific components of an image and its transform. However, some general statements can be made about the relationship between the frequency components of the Fourier transform and spatial characteristics of an image. Typically, high frequencies hold the information of fluctuations of edges and boundaries, and low frequencies correspond to the slowly varying components of an image. The non-informative frame (Fig. 8a) has no clear object information except the four strong edges at the corners of an image running approximately at $\pm 45°$ so its Fourier spectrum (Fig. 8b) shows prominent components along the $\pm 45°$ directions that correspond to the four corners of an image. Compared to the non-informative frame, the informative frame (Fig. 9a) has a lot of clear edge information so its spectrum (Fig. 9b) of the informative frame does not show prominent components along the $\pm 45°$ directions because it has a wider range of bandwidths from low to high frequencies.

3.2. Texture analysis

The texture analysis is applied on the frequency spectrum image, which is a 2D plot of the magnitude, in order to find the pattern difference between the informative and the non-informative frames. The most well-known statistical approach toward texture analysis is the gray level co-occurrence matrix (GLCM) (Haralick et al., 1973; Shuttleworth et al., 2002; Bevk and Kononenko, 2002; Felipe et al., 2003; Weszka et al., 1976; Conners and Harlow, 1980; Hall-Beyer, 2000). The co-occurrence matrix contains the elements that are the counts of the number of pixel pairs for specific brightness levels, when separated by some distance (or displacement) at some relative inclination. To construct the co-occurrence matrix for this texture



(a) (b)

Fig. 9. (a) Informative Frame and (b) frequency spectrum of (a).

analysis, we set up a window (matrix) of size equal to the size of the frequency spectrum image itself, a displacement to 1, and a relative inclination to 0. The original investigation into the texture features based on the co-occurrence matrix was pioneered by Haralick et al. (1973). They defined 14 texture features. However, only some features among 14 texture features are in wide use in many applications (Weszka et al., 1976; Conners and Harlow, 1980). For our experiments, seven texture features (Entropy, Contrast, Correlation, Homogeneity, Dissimilarity, Angular Second Moment, and Energy) are extracted as follows (Hall-Beyer, 2000):

$$\text{Angular Second Moment (ASM)}: \quad \sum_i \sum_j P(i,j)^2 \tag{2}$$

$$\text{Contrast}: \quad \sum_i \sum_j (i,j)^2 \cdot P(i,j) \tag{3}$$

$$\text{Correlation}: \quad \sum_i \sum_j \frac{(i-\mu_x)\cdot(j-\mu_y)\cdot P(i,j)}{\sigma_x \sigma_y} \tag{4}$$

$$\text{Dissimilarity}: \quad \sum_i \sum_j P(i,j)\cdot |i-j| \tag{5}$$

$$\text{Entropy}: \quad \sum_i \sum_j P(i,j)\cdot \log P(i,j) \tag{6}$$

$$\text{Energy}: \quad \sqrt{ASM} \tag{7}$$

$$\text{Uniformity}: \quad \sum_i \sum_j \frac{P(i,j)}{1-|i-j|} \tag{8}$$

where $P(i,j)$ is the probability of a certain value in the co-occurrence matrix, $\mu_x = \sum_i \sum_j i \cdot P(i,j)$, $\mu_y = \sum_i \sum_j j \cdot P(i,j)$, $\sigma_x = \sqrt{\sum_i \sum_j (i-\mu_x)^2 \cdot P(i,j)}$ and $\sigma_y = \sqrt{\sum_i \sum_j (j-\mu_y)^2 \cdot P(i,j)}$. The extracted seven texture features are used to distinguish the informative from the non-informative frames in the colonoscopy video using K-means clustering algorithm.

3.3. Clustering-based informative frame classification

The K-means method is commonly used partitioning method (Han and Kamber, 2001; Witten and Frank, 2000; van Zyl and Cloete, 2003; Xu and Liao, 1998; Chen et al., 1998; Bhangale et al., 2000). The K-means method clusters data objects into K subsets using a certain distance function, where data objects in the same cluster are similar to one another but data objects in other clusters are dissimilar. Fig. 10 describes the K-means clustering algorithm when a data object (X_i) consists of p dimensional features (i.e., $X_i = \{x_i^1, x_i^2, \ldots, x_i^p\}$).

For our purpose, it is natural to set up the initial number of clusters to 2 ($k = 2$) and cluster the frames into two groups. One represents the informative frame group, and the other represents the non-informative frame group. We call this approach a one-step K-means clustering scheme. Even though the one-step K-means clustering scheme distinguishes the informative frame from the non-informative frame very well, we investigate whether a larger number of initial clusters (k) can further increase its overall accuracy. There are frames in which some parts are clear, but other parts are blurry. As before, we call these frames *ambiguous* frames. Figs. 11–13 show three types of frames (Non-informative, Informative and Ambiguous).

Analogous to the edge-based method, we next develop a two-step K-means clustering scheme to distinguish the informative frames from non-informative frames. In the first clustering step, we set the initial number of clusters to 3 ($k = 3$) in order to cluster frames into three groups:

K-Means Algorithm

(i) Initialization - randomly choose K points $C_1, C_2, \cdots, C_k$ as initial centroids, where an initial centroid(C_j) is $C_j = \{c_j^1, c_j^2, \cdots, c_j^p\}$

(ii) Repeat

(a) For i=1 to n (n = the total number of data objects)

compute distance $d_j(X_i) = \sqrt{(x_i^1 - c_j^1)^2 + (x_i^2 - c_j^2)^2 + \cdots + (x_i^p - c_j^p)^2}$

assign X_i to cluster D_{j*} where $j* = \min(d_1(X_i), d_2(X_i), \cdots, d_k(X_i))$

(b) Compute the new centroid (C_j) for each cluster $D_j, j = 1, 2, \cdots, k$

$$C_j = \frac{1}{|D_j|} \sum_{X_i \in D_j} X_i = \left(\frac{\sum x_i^1}{|D_j|}, \frac{\sum x_i^2}{|D_j|}, \ldots, \frac{\sum x_i^p}{|D_j|} \right)$$

where $|D_j|$ is the number of data objects in cluster D_j

(c) Exit, if the centroids no longer move

Fig. 10. K-means clustering algorithm.

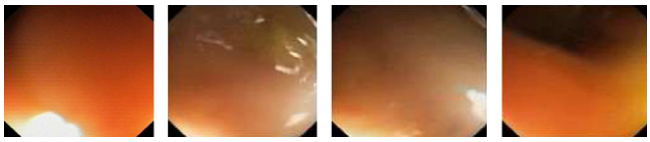


Fig. 11. Examples of non-informative frames.

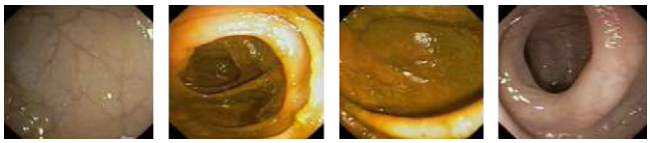


Fig. 12. Examples of informative frames.

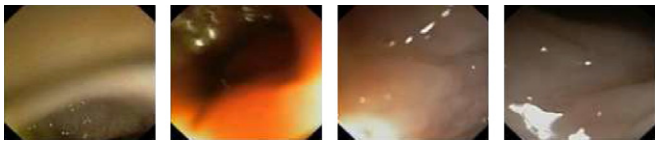


Fig. 13. Examples of ambiguous frames.

informative frames, non-informative frames, and ambiguous frames. The frames detected as ambiguous from the first step are used in the next clustering step. In the second clustering step, we set up the number of clusters to 2 ($k = 2$) in order to further divide the ambiguous frames into two groups that consist of informative frames and non-informative frames. Finally, all frames are clustered into two groups, either the informative frame or the non-informative frame groups. Our experiment results show that the two-step K-means clustering scheme is better than the one-step K-means clustering scheme.

4. Specular reflection detection

The specular reflection may considerably disturb human examination and computer-aided image processing techniques such as edge detection and texture analysis. When regarding medical images, especially endoscopic images, the problem is even worse because light source and viewing direction are almost identical; thereby, wet mucosa surfaces perpendicular to the viewing direction show the specular reflection. The specular reflection disturbs the distinction of informative frames from non-informative ones because these areas can be interpreted as informative contents. A model for separating specular reflectance from diffuse reflection is proposed for di-electric inhomogeneous material, the so-called di-chromatic reflectance model (Shafer, 1985). Algorithms based on this model have been applied to detect specular reflection for biological material (Taxt, 1994; Beach, 2002; C.D.S. and L.Z.K., 2003). However, human tissue dose not fit exactly into the di-chromatic reflectance model. In recent years, color gradients have been proposed to detect specular reflection (Gevers and Stokman, 2000). Vogt et al. (2002) utilized a simple thresholds method in Hue–Saturation–Value (HSV) color space to detect specular reflections. They converted an image in RGB color space to an image in HSV color space where $H \in [0, 359]$, $S \in [0, 255]$ and $V \in [0, 255]$. Two different thresholds were used for two different data sets. The thresholds for the gall sequence were $0 \leqslant H \leqslant 359$, $0 \leqslant S \leqslant 20$ and $0 \leqslant V \leqslant 200$. The thresholds for the thoracic cavity were $0 \leqslant H \leqslant 359$, $0 \leqslant S \leqslant 40$ and $0 \leqslant V \leqslant 200$. However, the detection accuracies are very sensitive to the thresholds because they used only one set of thresholds. Furthermore, it is difficult to find the optimal threshold values. In this section, we introduce our specular reflection detection algorithm using multiple thresholds adaptively, which is less sensitive to the thresholds, and very accurate.

The pixels in specular reflection do not always have absolute brighter colors than those in non-specular reflection; i.e. some pixels in specular-reflection areas are lower-intensity (darker) than non-specular reflection areas. Fig. 14a is an original color image, (b) is its gray level image, and (c) is a 3D plot of (b). As seen in this figure, the specular reflection pixel indicated by light blue color with the dotted line has lower RGB values (Fig. 14a), and lower Intensity value (Fig. 14b) than the non-specular reflection pixel indicated by red color with the solid line, so it is difficult to detect exact specular-reflection areas using one global threshold. Fig. 14 shows that a specular-reflection area is relatively brighter when compared with its surrounding area. We define two different specular-reflection areas, *Absolute Bright Area* and *Relative Bright Area*, and

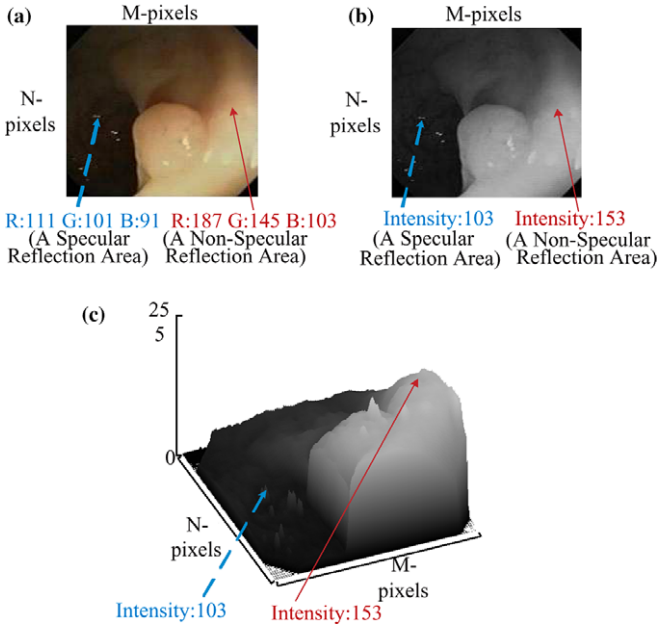


Fig. 14. (a) Color pixel values of specular reflections (light blue with dotted line) and non-specular reflections (dark red with solid line), (b) intensity values of specular reflections (light blue with dotted line) and non-specular reflections (dark red with solid line), and (c) 3D surface of plot (b). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

propose a specular reflection area detection technique using multi-thresholds. An absolute bright area is defined as an area with absolutely bright pixels. If any pixel is brighter than a certain threshold value, the pixel is considered as a specular reflection pixel. Absolute bright pixels usually appear in a larger area. Relative bright area is defined as an area with relatively brighter pixels. The relative bright area is decided by outlier detection. Using this property, we propose a three-step specular reflection detection technique as follows:

Step 1. First, we convert the color space of an input frame from RGB to HSV (Hue, Saturation and Value) because a frame in HSV space is less sensitive to noise. The ranges of saturation (S) and value (V) are between 0 and 1 and the range of hue (H) is between 0 and 360 (Gonzalez, 2002).

Step 2. Absolute bright areas are detected by two thresholds TH_s and TH_v for saturation (S) and value (V), respectively, as follows. We only consider saturation and value to detect bright areas since hue representing the purity of colors is rarely related to the brightness.

$$\begin{aligned}&\text{if } S(i) < TH_s \quad \text{and} \quad V(i) > TH_v\\ &\text{then pixel } i \text{ is in Absolute Bright Area}\\ &\text{otherwise it is in Non-absolute Bright Area}\end{aligned} \tag{9}$$

where $S(i)$ and $V(i)$ are the saturation and the value of pixel i, respectively. TH_s and TH_v are the thresholds for saturation (S) and value (V), respectively. In our experiments, absolute bright areas can be detected where the saturation (S) is lower than 0.35 and the value (V) is higher than 0.75. Fig. 15 shows an example of absolute bright areas detected using the above thresholds.

Step 3. To find a relative bright area, we segment an image into several regions; each of which consists of the similar color and texture. The good segmentation results could be evaluated based on the three criteria: (1) each image contains a set of approximately homogeneous color-texture regions, (2) the colors between two neighboring regions are distinguishable, and (3) the segmentation results are robust against the parameters of the algorithm. Based on the above three criteria, we chose JSEG (Deng and Manjunath, 2001) over the others (i.e., clustering image segmentation technique (Comaniciu and Meer, 1997), morphological watershed-based region growing (Shafarenko et al., 1997), energy diffusion (Ma and Manjunath, 1997), graph partitioning (Shi and Malik, 2000) and Blobworld (Carson et al., 2002)) since JSEG performs better on our image set. Even though direct clustering methods in color space also provide good results, the clustering method is very sensitive to the number of clusters. Besides, JSEG considers not only color information but also texture information of segment images which makes the method more resistant to noise. JSEG consists of two independent steps: color quantization and spatial segmentation. In the first step, colors in the image are quantized to several representative classes that can be used to differentiate regions in the image. The pixels are then replaced by their corresponding color class labels to form a class-map of the image. High and low values of color class labels in the class-map correspond to possible boundaries and interiors of color-texture regions. A region growing method is then used to segment the image based on the multi-scale class-maps. The segmentation result is mainly determined by the parameter value related color quantization. We experimentally determine the appropriate parameter values for JSEG such that a desirable segmentation result is obtained. The region segmentation result of our sample image can be seen in Fig. 16. A relative bright area is the area that has relatively higher intensity in a local region. The problem of detecting relative bright areas can be replaced by the problem of detecting outlier pixels (pixels distinguishably brighter than the neighboring pixels) in each segmented region. We use the statistic box-plot method (Tukey, 1977) to detect outlier pixels. A box plot is a graph that is useful for analyzing very large data sets such as identifying outliers and comparing distributions. As seen in Fig. 17b, a box plot summarizes the data to five numbers: median, upper quartile ($Q3$), lower quartile ($Q1$), upper outlier boundary, and lower outlier boundary. A median is found by listing the data values in an increasing order, then finding the center value. If there is an even number of data values, a median becomes an average of two center values. This median value is indicated by

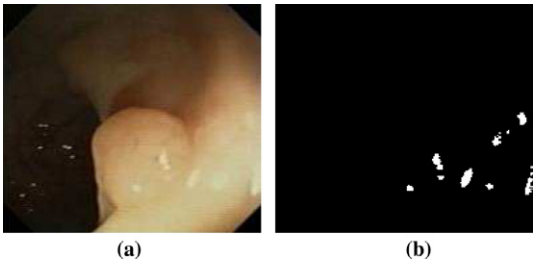


Fig. 15. (a) Original image and (b) Absolute Bright Area map of (a).

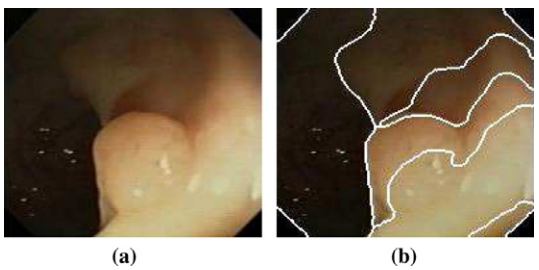


Fig. 16. (a) Original image and (b) region segmentation result of (a) using JSEG.

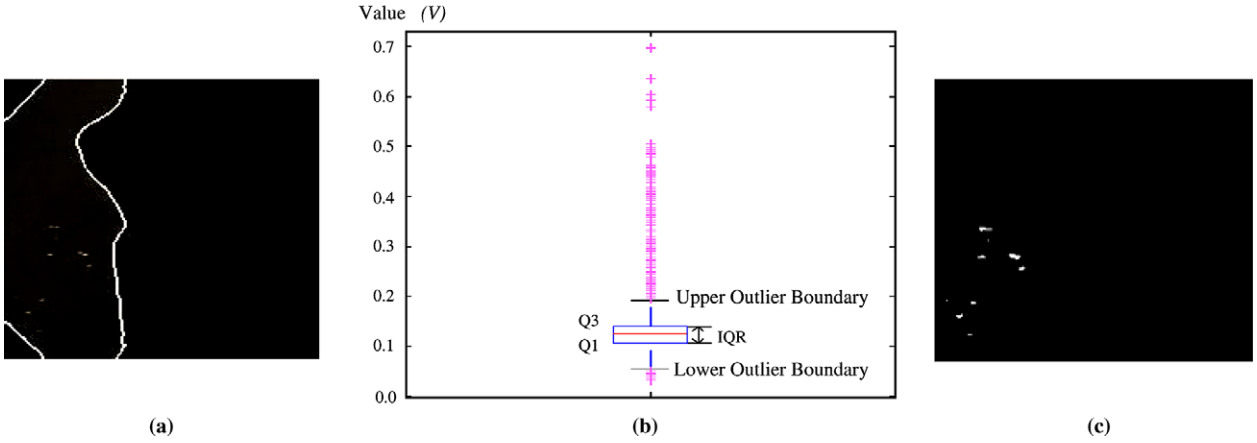


Fig. 17. (a) One segmented region of Fig. 16b, (b) box plot of values (V) of (a), and (c) Relative Bright Area map of (a).

the interior line of the box. The lower quartile ($Q1$) is the median of the lower half of the data divided by the overall median. The lower quartile value forms the bottom line of the box. The upper quartile ($Q3$) is the median of the upper half of the data divided by the overall median. The upper quartile value forms the top line of the box. The difference between the upper quartile ($Q3$) and the lower quartile ($Q1$) is called the interquartile range or IQR. To find outliers, we first define the lower outlier boundary and upper outlier boundary; the lower outlier boundary = the lower quartile ($Q1$) − 1.5*IQR, and the upper outlier boundary = the upper quartile ($Q3$) + 1.5*IQR. These lower and upper outlier boundaries form the ends of the whiskers of the graph, and any data values falling outside these boundaries are considered outliers. The outlier pixels (pixels distinguishably brighter than the neighboring pixels) in each segmented region are determined as follows:

$$\text{if } S(i,k) < \text{TH}_s \quad \text{and} \quad V(i,k) > \text{TH\_v}(k)_{\text{outlier}} \text{ and pixel } i \text{ in region } k \text{ is in Non-absolute Bright Area then it is in Relative Bright Area} \tag{10}$$

where $S(i,k)$ and $V(i,,k)$ are the saturation and the value of pixel i in region k, respectively. TH_s is the threshold of saturation, and $\text{TH\_v}(k)_{\text{outlier}}$ is the upper outlier boundary of region k and it is defined as follows.

$$\text{TH\_v}(k)_{\text{outlier}} = Q3(k) + 1.5 * \text{IQR}(k) \tag{11}$$

where $Q1(k)$ is the 25th percentile of value (V) for region k, $Q3(k)$ is the 75th percentile of value (V) for region k and $\text{IQR}(k) = Q3(k) - Q1(k)$.

An example of a relative bright area is shown in Fig. 17, in which Fig. 17a shows one of the segmented regions of Fig. 16b. Fig. 17b is the box plot of values (V) of the region in Fig. 17a. The crosses highlighted with the light pink color in Fig. 17b represent the outlier pixels, and Fig. 17c is the relative bright area map corresponding to the pixels above the upper outlier boundary in Fig. 17b. Total specular reflections (Fig. 18d) can be obtained by combining *Absolute Bright Area* (Fig. 18b) and *Relative Bright Area* together (Fig. 18c).

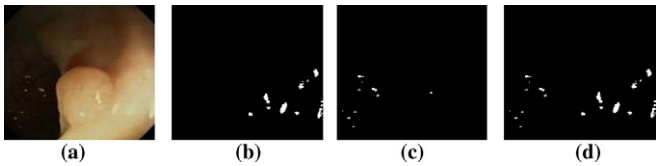


Fig. 18. (a) Original image, (b) Absolute Bright Area map of (a), (c) Relative Bright Area map of (a), and (d) total specular reflection of (a).

By discarding the detected specular-reflection areas, we can increase the performance of the informative frame classification techniques which were introduced in Sections 3 and 4. We will present our experimental results in Section 5 showing how our specular reflection algorithm is applied to the informative frame classification techniques, and how much it can increase their accuracies.

5. Experimental results

Our experiments assess the performances of the three proposed techniques for specular reflection detection, edge-based and clustering-based frame classification. To verify the effectiveness of our proposed algorithms, four traditional performance metrics (Han and Kamber, 2001) such as precision, sensitivity (recall), specificity, and accuracy, are measured in our experiments. Those four performance metrics are described as follows.

	Predicted as positive	Predicted as negative
Actually positive	TP	FN
Actually negative	FP	TN

$$\text{Precision} = \frac{TP}{TP+FP}, \quad \text{Sensitivity} = \frac{TN}{FP+TN}$$
$$\text{Specificity} = \frac{TP}{TP+FN}, \quad \text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

We note that the resolutions of the original images are 391×375 and 571×451. However, odd lines (or even lines) in both horizontal and vertical directions are removed, and the images are resized from 391×375 to 195×187 and 571×451 to 285×225 to reduce degradation by interlacing.

In our experiments, three different data sets were used. First, to test the performances of the two frame classification techniques, edge-based technique and clustering-based technique, 923 frames extracted from two different colonoscopy videos were used in Sections 5.1, 5.2, 5.4 and 5.5. As discussed in Section 2, the edge-based frame classification technique requires two threshold values: the upper threshold and the lower threshold. The study of selecting the two thresholds was performed using 4000 frames, which is discussed in the beginning of Section 5.1. At last, 70 frames were used to test the performance of the specular reflection detection as reported in Section 5.3.

5.1. Evaluation of edge-based frame classification

To distinguish informative frames from non-informative frames using the proposed edge-based method, we need to decide the upper-threshold (TH_U) and the lower-threshold (TH_L) values as mentioned in Section 2. We examined two sample data sets, each of which contains 2000 frames, to determine the thresholds. The size of the frames in the first set is 285×225 pixels and that of the second set is 195×187 pixels. Each frame of the data sets is classified into one of the three categories (informative frame, non-informative frame and ambiguous frame) manually based on the quality of the images. The results of this manual classification for the two sample data sets can be seen in Table 1 as follows.

The IPR value for each frame in the two data sets is computed. The Minimum, Maximum, Average and Median values of IPR for each category of the data sets are shown in Tables 2 and 3. For illustration purpose, the distribution of the IPR values of 2000 frames is attached at the bottom of Tables 2 and 3. As seen in Tables 2 and 3, most of the informative frames have low IPR values such that the average IPR of informative frames is around 1%, and the maximum IPR of informative frames is less than 5%. In contrast, the ambiguous frames and the non-informative frames have higher IPR values such that the average IPR of non-informative frames is around 6–7%, and the average IPR of ambiguous frames is around 3–5%.

Fig. 19 shows the accumulated ratios of the number of informative frames, non-informative frames and ambiguous

Table 1
Manual classification of two sample data sets

	Set 1 (285×225)	Set 2 (195×187)
# of informative frames	1479	1157
# of non-informative frames	258	646
# of ambiguous frames	263	197
Total	2000	2000

Table 2
Statistics of data set 1 (285×225)

IPR(%)	Informative (IPR)	Non-informative (IPR)	Ambiguous (IPR)
Minimum	0.016	1.725	0.460
Maximum	4.926	10.451	9.155
Average	0.849	7.291	4.615
Median	0.541	7.455	4.387

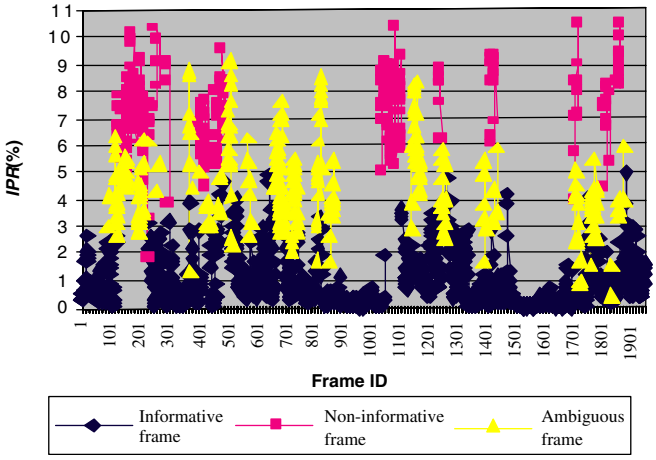
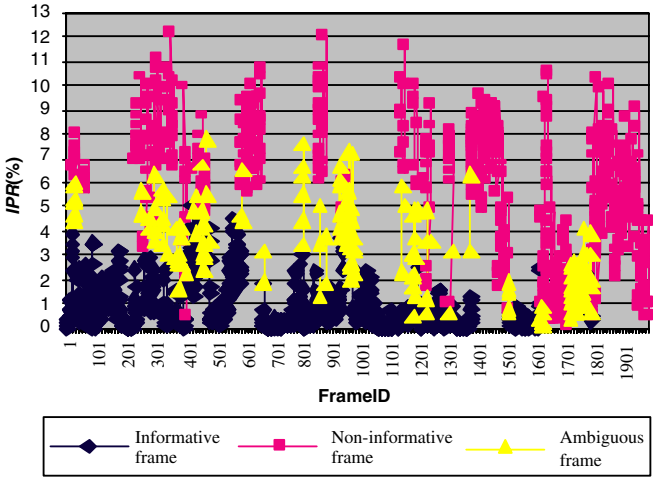


Table 3
Statistics of data set 2 (195×187)

IPR(%)	Informative (IPR)	Non-informative (IPR)	Ambiguous (IPR)
Minimum	0.000	0.222	0.133
Maximum	4.930	12.130	7.821
Average	0.753	5.982	3.137
Median	0.401	6.538	3.000



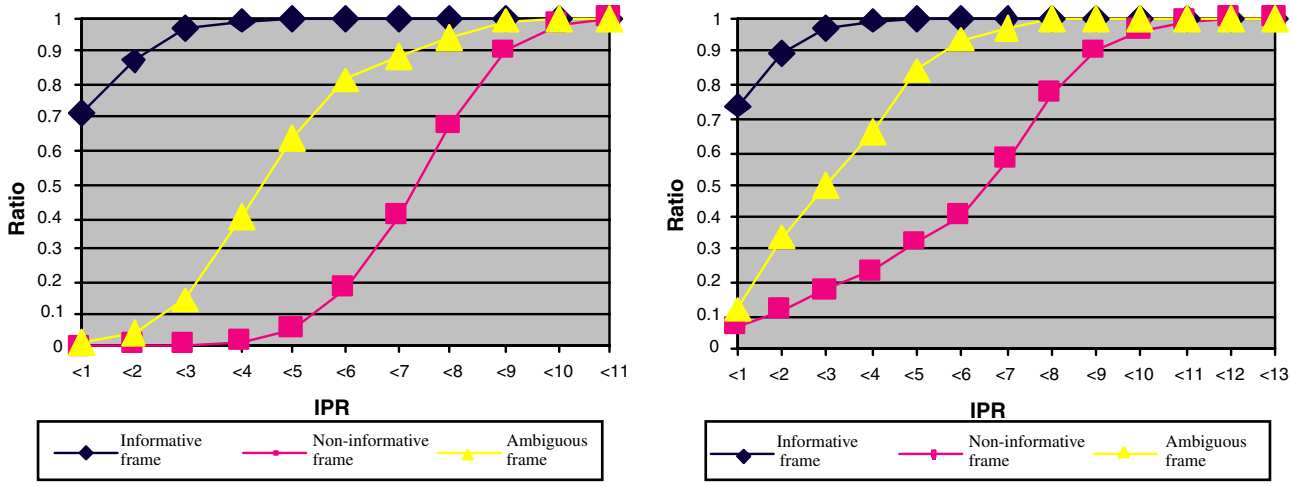


Fig. 19. Accumulated ratios of informative frames, non-informative frame and ambiguous frames for data set 1 (left) and data set 2 (right).

frames of each data set based on IPR values. As shown in this figure, the IPR values of all informative frames are less than 5%. However, the IPR values of all non-informative and ambiguous frames are distributed over a wide range (from less than 1% to more than 12%). Therefore, we select the two threshold values as follows.

– The candidates for the lower-threshold (TH_L) value should be less than 5% because all informative frames have the IPR values less than 5%. The intuitive criterion for the TH_L is that the portion of detected informative frames by the selected TH_L should be greater than that of the detected non-informative and ambiguous frames. This comparison can be done by computing the difference between the ratio of the number of informative frames and the ratio of the number of non-informative and ambiguous frames. The difference (D^{IPR}) for an IPR value, i, is calculated as follows: $D_i^{IPR} = CR_i - (BR_i + AR_i)$, where CR_i is the ratio of the number of informative frames, BR_i is the ratio of the number of non-informative frames, and AR_i is the ratio of the number of ambiguous frames at IPR i. The subtraction works here since each value is a ratio which is not an absolute but a relative value. The results for the IPR 1%, 2%, 3%, 4% and 5% are illustrated in Table 4. In our experiment, IPR 1%, 2% and 3% are selected as TH_L values since the differences (D^{IPR}) of the three are much larger than those of the others.

Table 4
Results of differential calculation for low threshold

IPR	D^{IPR} of set 1	D^{IPR} of set 2	Average D^{IPR}
<1	0.699	0.554	**0.6265**
<2	0.831	0.442	**0.6365**
<3	0.825	0.299	**0.5620**
<4	0.576	0.097	**0.3365**
<5	0.314	−0.170	**0.0720**

– The candidates for the upper-threshold (TH_U) value should be selected greater than or equal to 5% because all informative frames have the IPR values less than 5%. Since we already determined the lower-threshold (TH_L) values as 1%, 2%, or 3%, we ran experiments with different pairs of TH_U and TH_L values such as 5, 6, 7, and 8 for TH_U and 1, 2 and 3 for TH_L to determine the optimal TH_U value. The results are shown in Fig. 20. As seen in the figure, there is little change in the number of frames detected as informative even if TH_U values are changing from 5 to 8. For example, in the first graph, about 1450 and 1330 frames are detected as informative frames when TH_L is 1 for the sample data sets 1 and 2, respectively, irrespective of TH_U values, which are ranged from 5 to 8. In the second graph, about 1600 and 1440 frames are detected as informative frames when TH_L is 2, and about 1680 and 1520 frames are detected as informative frames when TH_L is 3 in the third graph for the sample data sets 1 and 2, respectively.

Using a set of threshold values determined above (1, 2 and 3 for TH_L, and 5, 6, 7, and 8 for TH_U), we have run our edge-based informative frame detection algorithm. The overall results for the precision and the recall are summarized in Table 5 compared with several combinations of the low-threshold (TH_L) from 1 to 3 and the upper-threshold (TH_U) from 5 to 8. The 'Average' in Table 5 is an average value of the precision and sensitivity. As seen in the table, the results are very good, and the accuracy does not vary much with the threshold values.

We applied our edge-based technique to the two colonoscopy video test sets. The actual video frame rate of our colonoscopy video is 30 frames per second. However, we extracted frames at the rate of 1 frame per second because the evaluation is performed on individual frames so the extraction rate does not become a performance degrading factor. The total length of videos in our test set is about

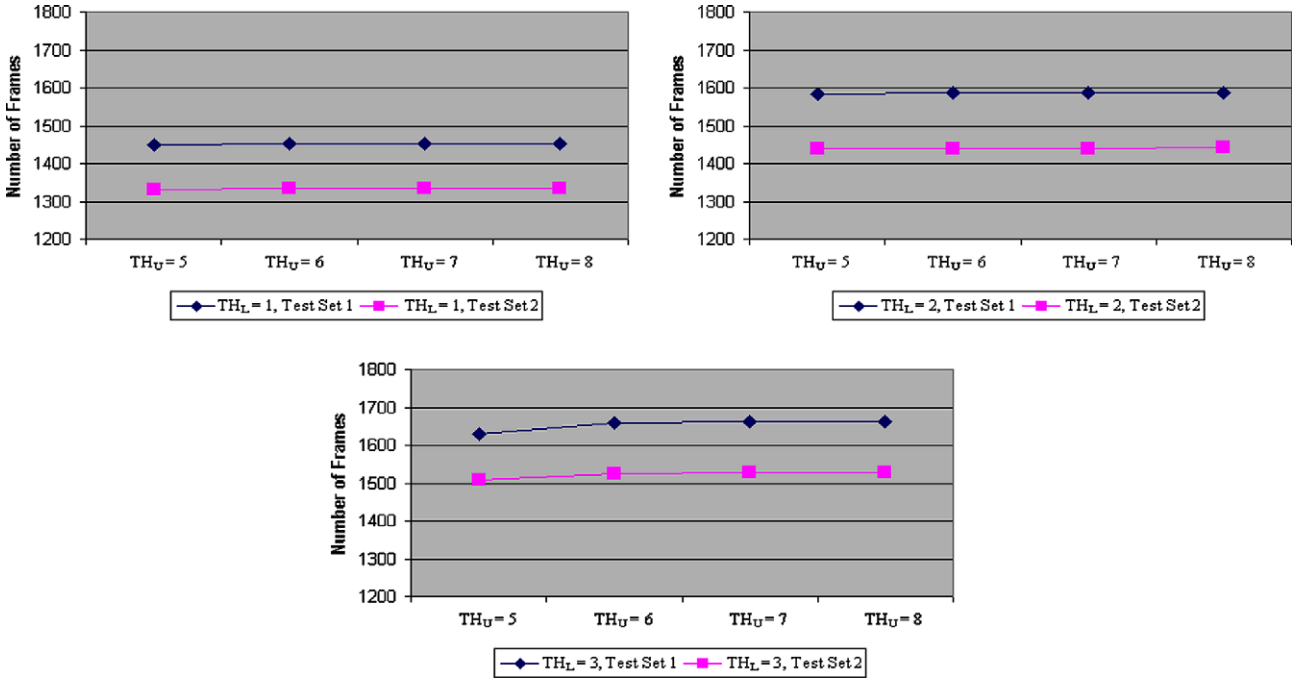


Fig. 20. Detected informative frames based on different pairs of thresholds.

Table 5
Precision and sensitivity based on several combinations of thresholds

Thresholds	Test set 1			Test set 2		
	Precision	Sensitivity	Average	Precision	Sensitivity	Average
$TH_L = 1$, $TH_U = 5$	1.000	0.936	0.968	0.916	0.965	0.940
$TH_L = 2$, $TH_U = 5$	0.979	1.000	0.989	0.898	0.996	0.947
$TH_L = 3$, $TH_U = 5$	0.949	1.000	0.974	0.869	1.000	0.934
$TH_L = 1$, $TH_U = 6$	1.000	0.936	0.968	0.915	0.965	0.940
$TH_L = 2$, $TH_U = 6$	0.976	1.000	0.988	0.897	0.996	0.946
$TH_L = 3$, $TH_U = 6$	0.934	1.000	0.967	0.859	1.000	0.929
$TH_L = 1$, $TH_U = 7$	1.000	0.936	0.968	0.915	0.965	0.940
$TH_L = 2$, $TH_U = 7$	0.976	1.000	0.988	0.897	0.996	0.946
$TH_L = 3$, $TH_U = 7$	0.932	1.000	0.966	0.857	1.000	0.928
$TH_L = 1$, $TH_U = 8$	1.000	0.936	0.968	0.915	0.965	0.940
$TH_L = 2$, $TH_U = 8$	0.975	1.000	0.987	0.897	0.996	0.947
$TH_L = 3$, $TH_U = 8$	0.930	1.000	0.966	0.856	1.000	0.928

15 min and the test set consists of 923 frames. There are two different resolutions (285×225 and 195×187pixels) in our videos. The details about our test video set can be found in Table 6.

Fig. 21 shows the experimental results of our edge-based frame classification technique. The results indicate the proposed technique is acceptable achieving over 88% for four different performance metrics (i.e. precision, sensitivity, specificity, and accuracy).

Table 6
Test set of videos

Video ID	Video length (min)	Total # of frames	Resolution
Colon-1	10	627	285×225
Colon-2	5	296	195×187
Total	15	923	

5.2. Evaluation of clustering-based frame classification

Next, we studied the performance of each of the seven texture features and compared the performance of the one-step and the two-step clustering schemes. The data set used in this section is the same test video (two colonoscopies) set described in Table 6. First, we examined the individual performance of each of the seven texture features to see if there is a dominant texture feature distinguishing informative frames from non-informative frames. We also present the performance of all seven features used together. Fig. 22 shows each performance metric

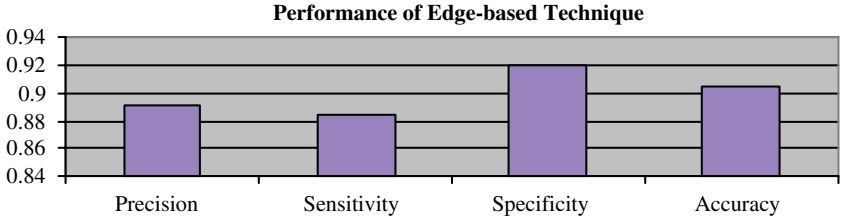


Fig. 21. Performance of edge-based technique.

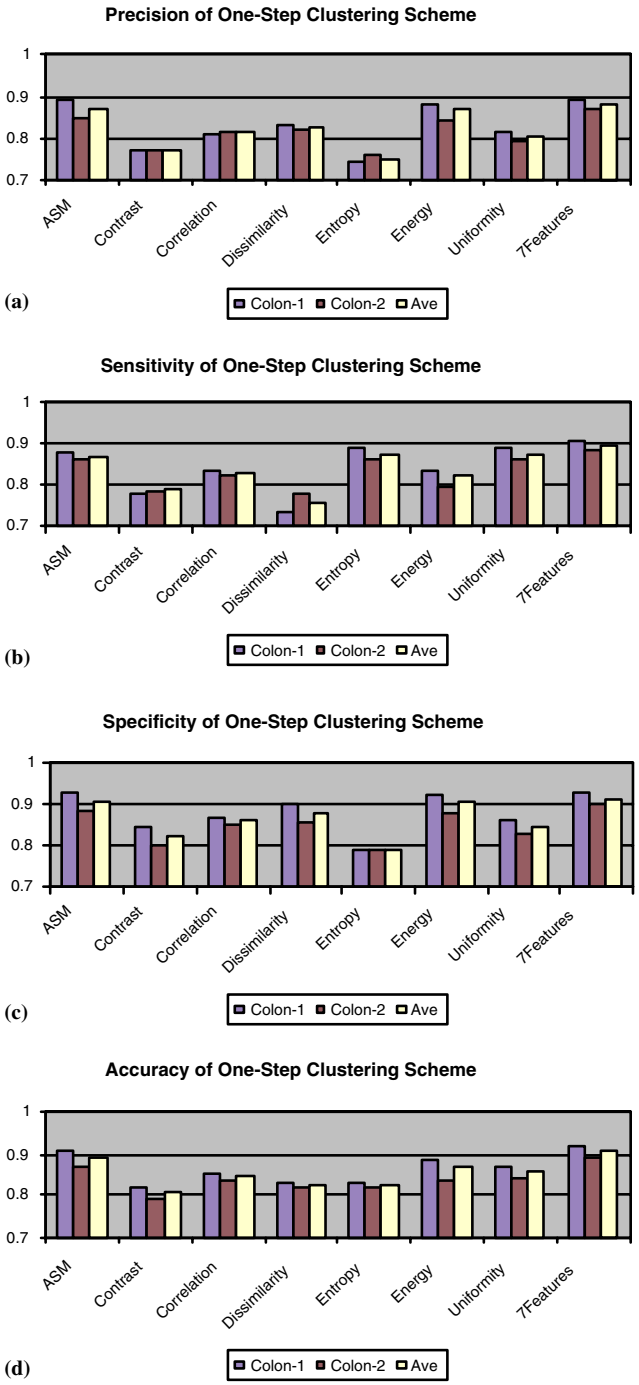


Fig. 22. Effectiveness of different texture features on performance of one step clustering scheme: (a) precision, (b) sensitivity, (c) specificity, and (d) accuracy.

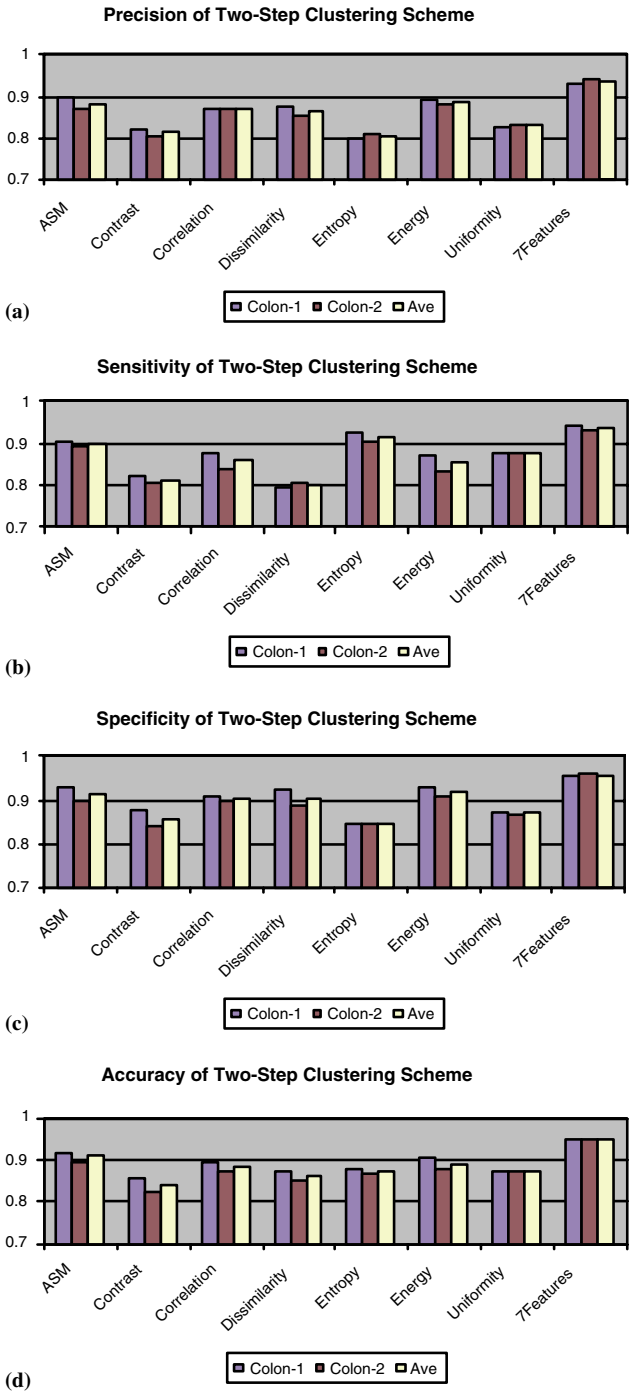


Fig. 23. Effectiveness of different texture features on performance of two step clustering scheme: (a) precision, (b) sensitivity, (c) specificity, and (d) accuracy.

of the one-step clustering scheme and Fig. 23 shows each performance metric of the two-step clustering scheme. The labels in the x-coordinate represent the name of texture features and the label of '7 Features' means that all seven features are used together. 'Colon-1' and 'Colon-2' in the legend indicate the video ID, and 'Ave' in the legend means the average performance metrics of two colonoscopy videos. Figs. 22 and 23 show that the performance of all seven features used together is better than performances of individual texture features for both the one-step and the two-step clustering schemes. We note that the two-step clustering scheme provides better results than the one-step clustering scheme, and the combination of all seven features optimizes the results.

5.3. Evaluation of specular reflection detection

For this experiment, 70 frames were selected as a test set from three colonoscopy videos. This test set consists of 35 informative frames and 35 non-informative frames. The details about our test set are described in Table 7. We examined the performances by comparing the specular reflections extracted manually with those extracted by our method. Fig. 24a is an original image with the detail of a specular reflection area. Fig. 24b is the specular reflection detection of (a) by the proposed method, in which the detected areas are highlighted with green color. Fig. 24c is the manual specular reflection detection of (a). The regions highlighted in green color in Fig. 24c represent the specular-reflection areas detected by both the proposed method and the manual procedure. And, the regions in red color, which are in the edges of the green color regions represent the specular-reflection areas missed by the proposed method. These errors happen due to the selected threshold value. Table 8 compares the performance metrics obtained by our proposed technique with the performance metrics obtained by the simple thresholds method presented in Vogt et al. (2002) on the pixel level. Using the simple thresholds method we obtained the results as shown in Table 8 with the saturation (S) lower than 0.45 and the value (V) higher than 0.70; these were the best thresholds for our data set. Table 8 also shows that the proposed specular reflection detection technique generates better results achieving over 92% for four different performance metrics and showing increases of the specular reflection detection performance by a 11.6%, 2.6%, 0.8% and 0.9% for the values of precision, sensitivity, specificity and accuracy, respectively, when compared with the simple thresholds method.

Table 7
Manual classification of test set

Class of frame	# of frames	# of specular pixels	Frame size
Informative	35	13949	195×187
Non-informative	35	24966	195×187

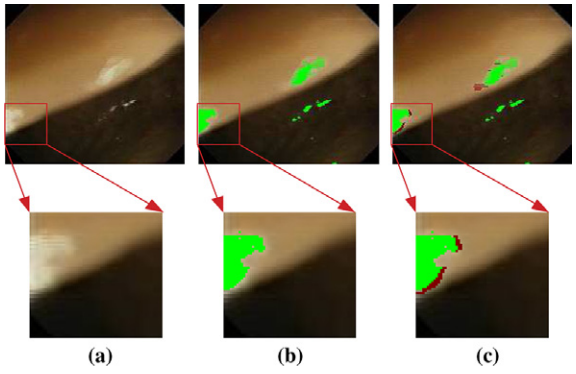


Fig. 24. (a) Original image and details of specular reflection region, (b) specular reflection of (a) detected by proposed algorithm and its details, and (c) specular reflection of (a) identified manually and its details.

Table 8
Performance of specular reflection detection

	Precision	Sensitivity	Specificity	Accuracy
Proposed technique	0.9242	0.9698	0.9958	0.9945
Simple thresholds	0.8085	0.9436	0.9882	0.9859

5.4. Performance enhancement of edge-based technique using specular reflection

The edge-based informative frame detection algorithm shows good performance results. However, as mentioned earlier the edge-based approach is affected by the specular reflections, which may cause incorrect detections. For instance, as seen in Fig. 25, specular reflections in non-informative frame can be detected as object boundaries causing the frame to be misclassified into an informative frame.

To prevent incorrect classification of non-informative frames due to specular reflections, we have used specular reflection information to eliminate the pixels in specular reflections from computation by the edge detector algorithm. Canny Edge Detector consists of four components; Gaussian smoothing, finding zero crossing using the derivative of Gaussian, non-maximal suppression and hysteresis thresholding. The gradient map is generated at the finding zero crossing step. By assigning the values of the gradient map corresponding to the specular pixels with zeros, we eliminate the pixels in specular reflections from computation by Canny Edge Detector. As seen in Fig. 26, first,

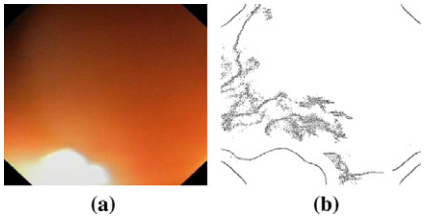


Fig. 25. (a) Non-informative image with specular reflections and (b) edges detected from (a).

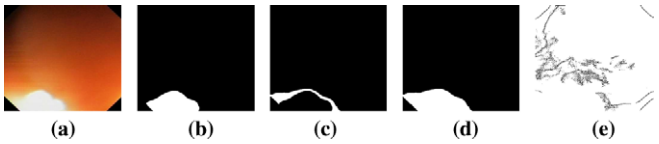


Fig. 26. (a) Non-informative image with specular reflections, (b) Absolute Bright Area of (a), (c) Relative Bright Area of (a), (d) specular reflection map of (a), and (e) improved edges of (a) using specular reflection information.

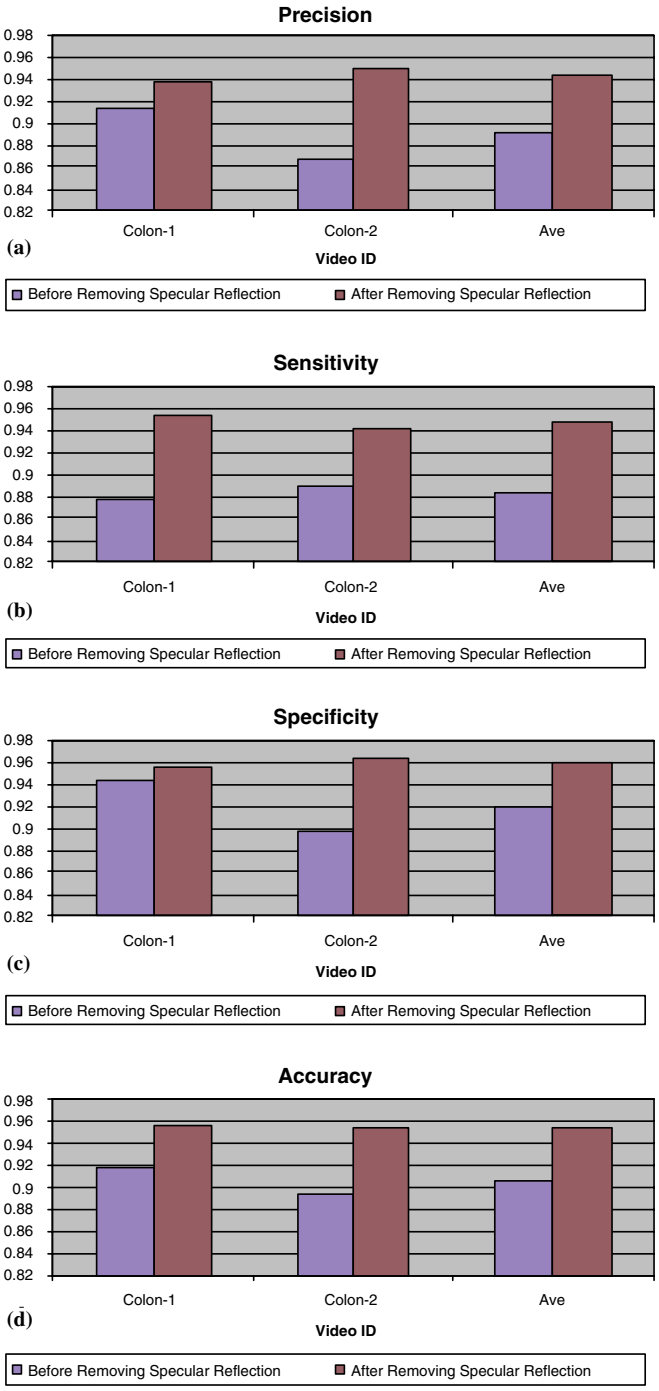


Fig. 27. Comparison of performance of edge-based technique based on consideration of specular reflections: (a) precision, (b) sensitivity, (c) specificity, and (d) accuracy.

we obtain the specular reflection map using the specular reflection detection technique. Fig. 26b is the absolute specular reflection map, Fig. 26c is the relative specular reflection map and Fig. 26d is the specular reflection map obtained by combining the absolute specular reflection and the relative specular reflection. After that, the pixels detected as specular reflections are not considered when we apply Canny Edge Detector to the original image. Fig. 26e shows an edge detection result of Fig. 26a, where the edges by specular reflections are not included. When we evaluate this (Fig. 26e), it can be classified as non-informative since it only includes isolated edge pixels.

Fig. 27 shows the experimental results of our edge-based frame classification technique considering and not considering specular reflection information for the comparison purpose. The data set used in this experiment is the same test video (two colonoscopies) set described in Table 6. The label of 'Before Removing Specular reflections' means that the edge-based frame classification is applied to original frames, and the label of 'After Removing Specular reflections' means that the pixels in specular reflections are discarded when the edge-based frame classification is applied. The 'Ave' of Video ID indicates the average of each performance metric. The results indicate that removing the pixels in specular reflections from the computation of Canny Edge Detector (after removing specular reflections) generates better results achieving over 94% for four different performance metrics (i.e., precision, sensitivity, specificity, and accuracy) and show an increase of the informative frame classification performance of about 5% for each. The main reason for the improvements is that specular reflections generate many non-isolated pixels: a non-informative frame with many specular reflections is misclassified as an informative frame. By removing the specular reflection, non-information frames are correctly classified as non-informative.

5.5. Performance enhancement of clustering-based technique using specular reflection

Next, we performed the experimental study to see how our specular reflection detection technique can increase the performance of the clustering-based frame classification scheme. We use specular reflection information to alter the pixel information in specular reflections prior to the computation of discrete Fourier transform (DFT) by replacing the specular pixels with the average pixel value of all the pixels on the boundary of the specular area. Fig. 28a is the original image and Fig. 28b is the specular reflection map obtained by the specular reflection detection technique. Fig. 28c is the specular free image in which the specular pixels are replaced with the average pixel value of all the pixels on the boundary of the specular area. Fig. 28d and e are the frequency spectrums obtained from Fig. 28a considering and not considering the detected specular reflection information, respectively. Fig. 28e shows how the exclusion of the pixels in specular reflections using

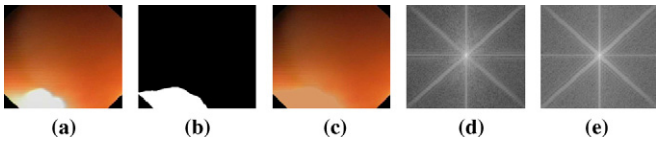


Fig. 28. (a) Non-informative image with specular reflections, (b) specular reflection Map of (a), (c) specular free image of (a), (d) frequency spectrum of (a) and (e) frequency spectrum of (c).

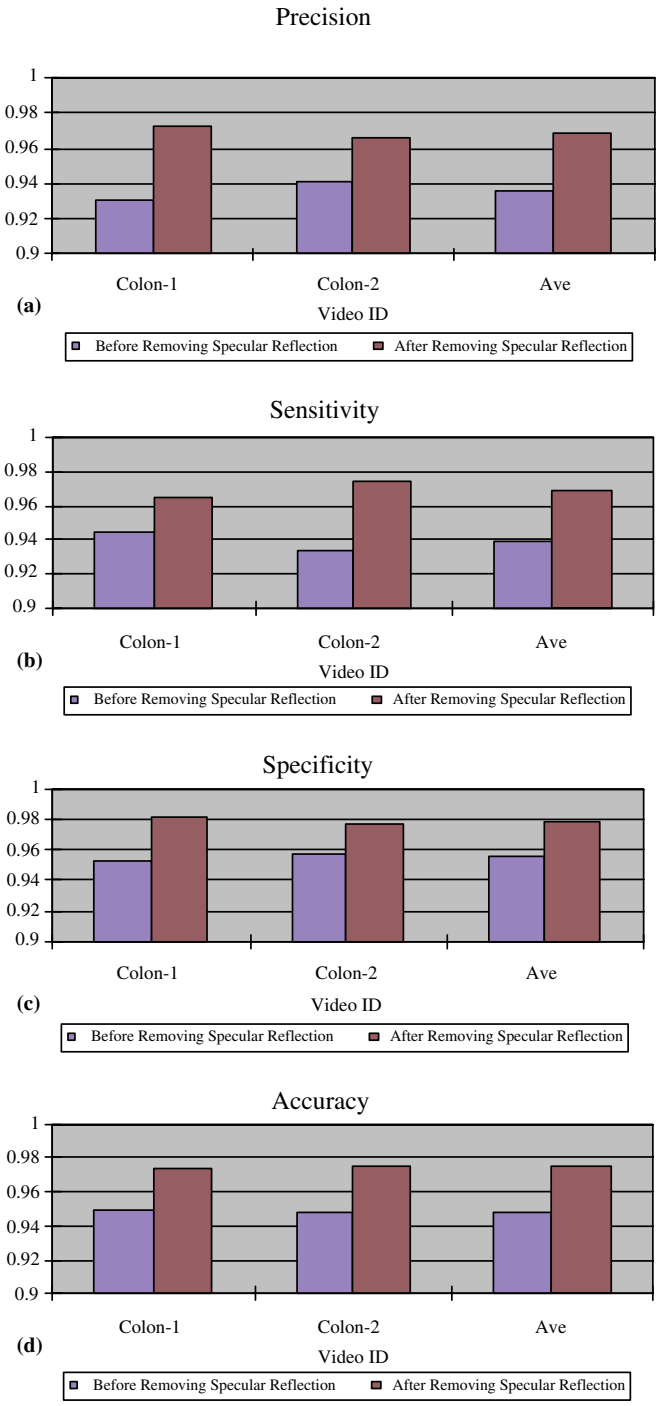


Fig. 29. Comparison of performance of clustering-based technique based on consideration of specular reflections: (a) precision, (b) sensitivity, (c) specificity and (d) accuracy.

the specular reflection free image can improve the informative frame classification because the frequency spectrum of Fig. 28e shows clearer prominent components along the $\pm 45°$ directions than the frequency spectrum of Fig. 28d.

Fig. 29 shows the experimental results of our clustering-based frame classification technique considering and not considering the specular reflection information for comparison purposes. The label of 'Before Removing Specular reflections' means that clustering-based frame classification is applied to original frames, and the label of 'After Removing Specular reflections' means that the pixels in specular reflections are replaced with boundary pixels when the clustering-based frame classification is applied. The 'Ave' of Video ID indicates the average of each performance metric. The results indicate that altering the pixels in specular reflections prior to the computation of discrete Fourier transform (DFT) (After Removing Specular reflections) generates better results than not altering the pixels in specular reflections (Before Removing Specular reflections). The results indicate that the two-step clustering-based frame classification scheme taking into account the specular reflection information gives the best results achieving over 96% for four different performance metrics (i.e., precision, sensitivity, specificity, and accuracy) and show an increase of the informative frame classification performance by about 3% for the four parameters.

5.6. Comparison study between edge-based and clustering-based classification techniques

Finally, we compared between the edge-based method and the clustering-based method in terms of the frame classification performance and the computational complexity. Table 9 shows the average of performances of 'Before Removing Specular reflections' and 'After Removing Specular reflections' of our two frame classification techniques. Overall, the clustering-based technique generates better performance results than the edge-based technique and the performance of both the edge-based technique and the clustering-based technique increases when corrections for the specular-reflection areas are incorporated into the computation. We note that the edge-based technique is more affected by correction of specular reflection than the clustering-based technique: for the edge-based technique the metrics improved about 5% on average. The edge-based technique with specular reflection area correction gives slightly better results than the clustering-based technique without correcting of specular-reflection areas. We achieve the best results (over 96.5% for all four performance metrics) using the clustering-based technique with correcting of specular-reflection areas prior to computation. In addition, the clustering-based technique has the advantage that it does not require selection of the optimal thresholds which are required for the edge-based technique.

Theoretical comparison of the computational complexity between the edge-based method and the clustering-based method is studied as follows. The edge-based method

Table 9
Comparison of average performance between edge-based technique and clustering-based technique

Metric	Edge-based technique		Clustering-based technique	
	Before removing specular reflections	After removing specular reflections	Before removing specular reflections	After removing specular reflections
Precision	0.89107	0.94356	0.93570	0.96886
Sensitivity	0.88342	0.94696	0.93910	0.96924
Specificity	0.92072	0.96053	0.95485	0.97859
Accuracy	0.90551	0.95489	0.94829	0.97480

consists of two main procedures; edge detection and isolated pixel ratio computation. Canny Edge Detector consists of four components; Gaussian smoothing, finding zero crossing using the derivative of Gaussian, non-maximal suppression and hysteresis thresholding. The complexity of Gaussian smoothing for each image is $O(\alpha N)$ for the size of Gaussian filter (α) and the size of image (N). The complexities of finding zero crossing using the derivative of Gaussian and non-maximal suppression are $O(N)$, and the complexity of hysteresis thresholding for each image is $O(\beta N)$ where β is decided depending on the parameter values. The complexity of the computation of isolated pixel ratio for each image is $O(N)$. When we consider all frames, the overall computational complexity of the edge-based method is $O(\chi N \cdot L)$ for $\chi = \max(\alpha, \beta)$ and the number of images (L). The clustering-based technique consists of three procedures; discrete Fourier transform (DFT), texture feature extraction based on the gray level co-occurrence matrix (GLCM) and K-mean clustering. The computational complexity of DFT for each image is $O(N^2)$. The complexity of the co-occurrence matrix construction is $O(\eta N)$ where η is the range of the intensity level (for instance 256), and the complexity of texture feature extraction for each image is $O(N)$. The complexity of K-mean clustering is $O(L \cdot F_k \cdot C_k \cdot T_k)$, where L is the number of frames, F_k is the number of features for distance measure, C_k is the number of clusters, and T_k is the number of iterations. The number of iterations varies depending on the data set. The overall computational complexity of the clustering-based method is $\max(O(N^2 \cdot L), O(L \cdot F_k \cdot C_k \cdot T_k))$ which is much bigger than the overall computational complexity of the edge-based method ($O(\chi N \cdot L)$) because χ is smaller than N. Therefore, the edge-based approach is a better candidate for cases where speed is more important than accuracy, and the clustering-based approach is better for the cases where the reverse is required.

6. Concluding remarks

In this paper, we propose two frame classification techniques, edge-based and clustering based techniques, and a specular reflection detection technique. The edge-based technique has two drawbacks. First, the edge-based technique is sensitive to specular reflections. To minimize this problem, we utilize a specular reflection detection algorithm and effectively eliminate the specular reflections as edges. The other problem is that the edge-based technique requires the computation of several threshold values, and once these have been determined it is not straightforward which ones to use. Here, we propose a new technique which addresses these two drawbacks using a combination of specular reflection detection, DFT, texture analysis, and clustering. The experimental results show that the specular reflection detection technique performs very well. Using the information obtained by our specular reflection detection technique, the edge-based frame classification technique can improve 5.2%, 6.3%, 3.9%, 4.9% and the clustering-based frame classification technique 3.3%, 3.0%, 2.3%, 2.6% for the value of precision, sensitivity, specificity and accuracy, respectively. Therefore, combined with the specular reflection detection the edge-based technique achieves on average of 95% in accuracy, and the clustering-based technique achieves on average of 97% in accuracy. However, as mentioned above, the clustering-based approach needs more computation time.

The classification of images based on their contents is an important step for computer-aid diagnosis applications as well as physicians in the colonoscopy videos. Our proposed technique distinguishing informative frames from non-informative frames can reduce the number of images to be viewed by physicians and to be analyzed by computer-aid image processing applications.

Are there disadvantages to removing images which we label non-informative? Do we remove information that may be of importance? Indeed, in theory it is possible that we may remove valuable information using our technique. However, the possibility of an important lesion being missed is very small in two reasons. First, as presented in Table 9, we can achieve very high specificity (over 97%) while also providing higher precision and sensitivity (over 96%) by removing the specular reflections, and using the clustering algorithms and texture analysis. Second, the video frame rate of our colonoscopy video is 30 frames per second, so any region is overlapped in a certain number of consecutive frames. Even if there is a missed frame showing an important lesion, the neighboring frames of a missed frame can show the missed lesion. Lengthy series of out-of-focus frames may indicate a specific colonic segment that is being traversed such as a flexure (colonoscope tip slides along mucosa), a low segment filled with cleansing fluid (fluid in descending colon with patient in left lat-

eral position), or a need for extensive washing of a dirty colon (continuous water irrigation). However, most of that information is not critical for medical management. What is important is the extent of clear vision that was achieved, and the time spent looking at good quality images without skipping large colon segments.

The technique presented here provides image quality evaluation without a reference image. This has as major advantage that the technique is domain-independent. Therefore, our method likely can be applied to a variety of other videos that lack a reference image. Indeed, we expect that our technique can be applied to analysis of videos captured from other endoscopic procedures such as upper gastrointestinal endoscopy, enteroscopy, cystoscopy, bronchoscopy, and laparoscopy.

Acknowledgements

One of the authors, Piet C. de Groen M.D of the Mayo Clinic College of Medicine provided the videos, his comments of the final evaluation of our experiments, and various supports in this work. This research is partially supported by the National Science Foundation Grants EIA-0216500, IIS-0513777, IIS-0513809, and IIS-0513582.

References

Ayers, G., Dainty, J., 1988. Iterative blind Deconvolution method and its applications. Optics Letters 13 (7), 547–549.

Bates, R., Jiang, H., 1991. Deconvolution – recovering the seemingly irrecoverable! International Trends in Optics, 423–437.

Beach, J., 2002. Spectral reflectance technique for retinal blood oxygen evaluation in humans. In: Proceedings of the Applied Imagery Pattern Recognition Workshop, October, pp. 117–123.

Bevk, M., Kononenko, I., 2002. A statistical approach to texture description of medical images: a preliminary study. In: Proceedings of the IEEE Symposium on Computer-Based Medical Systems, June, pp. 239–244.

Bhangale, T., Desai, U., Sharma, U., 2000. An unsupervised scheme for detection of microcalcifications on mammograms. In: Proceedings of the International Conference on Image Processing, pp. 184–187.

Canny, J., 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (6).

Carson, C., Belongie, S., Greenspan, H., Malik, J., 2002. Blobworld: image segmentation using expectation maximisation and its application to image querying. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 1026–1038.

C.D.S., L.Z.K., 2003. A novel approach to detect and correct specular reflectioned face region in color image. In: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, July, pp. 7–12.

Chen, C., Luo, J., Parker, K., 1998. Image segmentation via adaptive K-means clustering and knowledge-based morphological operations with biomedical applications. IEEE Transactions on Image Processing, 1673–1683.

Comaniciu, D., Meer, P., 1997. Robust analysis of feature spaces: color image segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, June, pp. 750–755.

Conners, R.W., Harlow, C.A., 1980. A theoretical comparison of texture algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 2 (3), 204–222.

Dario, P., Lencioni, M.C., 1997. A microrobotic system for colonoscopy. In: Proceedings of the IEEE International Conference on Robotic and Automation, Florence, Italy, pp. 1567–1572.

Deng, Y., Manjunath, B.S., 2001. Unsupervised segmentation of color-texture regions in images and video. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Felipe, J., Traina, A., Traina, C.J., 2003. Retrieval by content of medical images using texture for tissue identification. In: Proceedings of the IEEE Symposium on Computer-Based Medical Systems, June, pp. 175–180.

Gevers, T., Stokman, H.M.G., 2000. Classifying Color Transitions into Shadow-Geometry, Illumination Specular reflection or Material Edges. In: Proceedings of the IEEE International Conference on Image Processing, Vancouver, Canada, September, pp. 521–525.

Giannakis, G., Heath, R.J., 2000. Blind identification of multichannel FIR blurs and perfect image restoration. IEEE Transactions on Image Processing 9 (11), 1877–1896.

Gonzalez, R.C., 2002. Digital Image Processing. Prentice Hall.

Hall-Beyer, M., 2000. GLCM Texture: A Tutorial. National Council on Geographic Information and Analysis Remote Sensing Core Curriculum.

Han, J., Kamber, M., 2001. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers.

Haralick, R., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics, 610–621.

Khessal, N., Hwa, T., 2000. The development of an automatic robotic colonoscope. In: Proceedings of the TENCON 2000, September, pp. 71–76.

Klinker, G., Shafer, S., Kanade, T., 1990. A physical approach to color image understanding. International Journal of Computer Vision 4 (1), 7–38.

Kundur, D., Hatzinakos, D., 1996. Blind image deconvolution. IEEE Signal Processing Magazine 13 (3), 43–64.

Ma, W., Manjunath, B., 1997. Edge flow: a framework of boundary detection and image segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 744–749.

McCallum, B., 1990. Blind deconvolution by simulated annealing. Optics Communication 75 (2), 101–105.

Meyerhardt, J.A., Mayer, R.J., 2005. Therapy for colorectal cancer. New England Journal of Medicine 352 (5), 476–487.

Nakagaki, R., Katsaggelos, A., 2003. A VQ-based blind image restoration algorithm. IEEE Transactions on Image Processing 12 (9), 1044–1053.

Pai, H., Bovik, A., 2001. On eigenstructure-based direct multichannel blind image restoration. IEEE Transactions on Image Processing (October), 1434–1446.

Phee, S., Ng, W., 1998. Automatic of colonoscopy: visual control aspects. Medicine and Biology Magazine.

Ramirez, R.W., 1985. The FFT, Fundamentals and Concepts. Prentice-Hall.

Shafarenko, L., Petrou, M., Kittler, J., 1997. Automatic watershed segmentation of randomly textured color images. IEEE Transactions on Image Processing 6 (11), 1530–1544.

Shafer, S., 1985. Using color to separate reflection components. In: Color Research and Application, pp. 210–218.

Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8), 888–905.

Shuttleworth, J., Todman, A., Naguib, R., Newman, B., Bennett, M., 2002. Colour texture analysis using co-occurrence matrices for classification of colon cancer images. In: Proceedings of the Electrical and Computer Engineering 2002, May, pp. 1134–1139.

Sid-Ahmed, M.A., 1995. Image processing: Theory, Algorithms, and Architectures, New York, NY.

Society, A.C., 2005. Colorectal Cancer Facts and Figures. American Cancer Society Special Edition 2005, pp. 1–20.

Sonka, M., 1999. Image Processing, Analysis, and Machine Vision. PWS Pub.

Sucar, L.E., Gillies, D.F., 1990. Knowledge-based assistant for colonoscopy. In: Proceedings of the Third International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, June.

Taxt, T., 1994. Separation of the diffuse, specular and quasiperiodic signal components in medical ultrasound images. In: Proceedings of the IEEE Ultrasonics Symposium, November, pp. 1639–1644.

Tukey, J.W., 1977. Exploratory Data Analysis. Reading, Massachusetts.

van Zyl, J., Cloete, I., 2003. The influence of the number of clusters on randomly expanded data sets. In: Proceedings of the International Conference on Machine Learning and Cybernetics 2003, November, pp. 355–359.

Vogt, F., Paulus, D., N.H., 2002. Highlight Substitution in Light Fields. In: Proceedings of the IEEE International Conference on Image Processing, Rochester, USA, September, pp. 637–640.

Walker, J.S., 1996. Fast Fourier transforms. CRC Press.

Weszka, J.S., Dyer, C., Rosenfeld, A., 1976. A comparative study of texture measures for terrain classification. IEEE Transactions on Systems, Man, and Cybernetics 6 (4), 269–285.

Witten, I.H., Frank, L., 2000. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann Publishers.

Xu, H., Liao, M., 1998. Cluster-based texture matching for image retrieval. In: Proceedings of the International Conference on Image Processing, pp. 766–769.