7-2015

Cable Footprint History: Spatio-Temporal Technique for Instrument Detection in Gastrointestinal Endoscopic Procedures

Chuanhai Zhang
Iowa State University, czhang89@iastate.edu

Wallapak Tavanapong
Iowa State University, tavanapo@iastate.edu

Johnny S. Wong
Iowa State University, wong@iastate.edu

Piet C. de Groen
Mayo Clinic College of Medicine

JungHwan Oh
University of North Texas

Cable Footprint History: Spatio-Temporal Technique for Instrument Detection in Gastrointestinal Endoscopic Procedures

Abstract
We propose a new fast spatio-temporal technique that detects an operation scene---a video segment corresponding to a single purpose diagnosis action or a single purpose therapeutic action. The technique utilizes (1) color contrast of the cable region and the background, (2) the new area-based coordinate system to compute spatial features, and (3) the history of locations of detected cables of the instrument in a video to discard false regions. The proposed technique and software are useful for (1) automatic documentation of diagnostic or therapeutic operations at the end of the procedure, (2) a second review for causes of complications due to these operations, and (3) a building block for an effective content-based retrieval system to facilitate endoscopic research and education. On 38 fulllength colonoscopy and upper endoscopy video files with different cable colors and four different insertion directions of the instruments, the average percentage of false positive duration is small at 2.23%. The average percentage of true positive duration is high at 92.9%. The average analysis time per image is 13 milliseconds on an inexpensive off-the-shelf PC.

Keywords
Colonoscopy, Polypectomy, Instruments, Image Analysis, Image Features

Disciplines
Analytical, Diagnostic and Therapeutic Techniques and Equipment | Computer Engineering | Computer Sciences | Gastroenterology

Comments
This is from IPCV'15 - The 19th International Conference on Image Processing, Computer Vision, & Pattern Recognition, July 2015, pp.308-314. Posted with permission.

Cable Footprint History: Spatio-Temporal Technique for Instrument Detection in Gastrointestinal Endoscopic Procedures

Chuanhai Zhang[1], Wallapak Tavanapong[1], Johnny Wong[1], Piet C. de Groen[2], and JungHwan Oh[3]
[1]Dept. of Computer Science, Iowa State University, Ames, IA, USA
[2]Mayo Clinic College of Medicine, Mayo Clinic, Rochester, MN, USA
[3]Dept. of Computer Science and Engineering, University of North Texas, Denton, TX, USA

Abstract - *We propose a new fast spatio-temporal technique that detects an operation scene---a video segment corresponding to a single purpose diagnosis action or a single purpose therapeutic action. The technique utilizes (1) color contrast of the cable region and the background, (2) the new area-based coordinate system to compute spatial features, and (3) the history of locations of detected cables of the instrument in a video to discard false regions. The proposed technique and software are useful for (1) automatic documentation of diagnostic or therapeutic operations at the end of the procedure, (2) a second review for causes of complications due to these operations, and (3) a building block for an effective content-based retrieval system to facilitate endoscopic research and education. On 38 full-length colonoscopy and upper endoscopy video files with different cable colors and four different insertion directions of the instruments, the average percentage of false positive duration is small at 2.23%. The average percentage of true positive duration is high at 92.9%. The average analysis time per image is 13 milliseconds on an inexpensive off-the-shelf PC.*

Keywords: Colonoscopy, Polypectomy, Instruments, Image Analysis, Image Features

1 Introduction

Gastrointestinal endoscopy is a procedure using an endoscope to diagnose or treat a condition in the digestive system. For instance, colonoscopy enables inspection of the inside of the human colon and diagnostic and therapeutic operations. Colonoscopy is currently the gold standard for colorectal cancer screening. Upper Endoscopy (EGD) is the procedure for inspection of the stomach. In the US, Colorectal cancer is the second leading cause of cancer-related deaths behind lung cancer [1], causing about 50,000 annual deaths. Colorectal cancer and stomach cancer are the third and the fifth most common cancer in the world [2].

During the insertion phase of an endoscopic procedure, a flexible endoscope (with a tiny video camera at the tip) is advanced under direct vision (via the anus for colonoscopy) and (via the mouth for EGD). The video camera generates video of the internal mucosa of the organ. The video data are displayed on a monitor for real-time analysis by the endoscopist. During the withdrawal phase, the endoscope is gradually withdrawn with careful examination of the mucosa. Necessary biopsy and therapeutic operations (e.g., polypectomy) are performed. During these operations, an instrument is inserted via a working channel of the endoscope through the shaft. A variety of instruments (e.g., Fig. 1(a-c), cytology brushes, and sclerotherapy needles) can be used. Within a single procedure, the head and the cable of the instrument typically appears in the field of view (FoV) of the camera from the same approximate direction and location in the image. We call them *insertion direction* and *insertion location,* respectively. For instance, in Fig. 1(d-e), the instrument is in the lower right corner. In Fig. 1(f), the instrument is from the lower left corner.

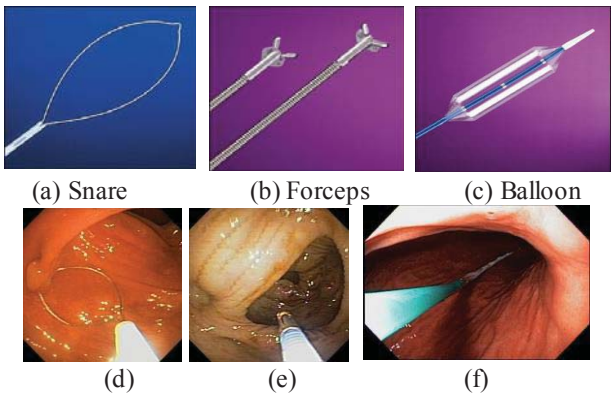


Fig. 1. (a-c) Instruments; (d-f) Appearance of a snare, a biopsy forceps, and an inflatable balloon in endoscopic images

Previously, we defined an *operation shot* in an endoscopic video as a segment of visual data that corresponds to a diagnostic or therapeutic operation [3]. For instance, we count each biopsy as one operation. We proposed an algorithm that detects these shots automatically [3] based on key visual properties of the cable of the instrument. However, the technique is very slow, making it impractical for automatic documentation at the end of the procedure to help

saving physician's reporting time. Furthermore, it is not uncommon that multiple biopsies are performed consecutively on the same or nearby colon mucosa.

Our contribution in this paper is as follows. First, we define an *operation scene* as *a video segment corresponding to a single purpose diagnosis action or a single purpose therapeutic action*. For instance, a series of consecutive biopsies is considered in the same scene with the biopsy purpose. Second, we propose a new, fast spatio-temporal technique for detection of operation scenes. The crux of the technique is (1) the detection of the footprint of the cable of the instrument utilizing the contrast of the cable from the background, (2) a new area-based coordinate system to compute spatial features, and (3) the history of detected footprints of the cable to automatically detect the insertion direction which is effective for eliminating false positives.

The highlight of our findings is as follows. On our data set of 38 full-length Gastrointestinal (GI) endoscopic video files covering a total of 20 hour worth of video data with different cable colors and four different insertion directions, the average percentage of false positive duration is very small at 2.23%. The average percentage of true positive duration is high at 92.9%. On 5 additional full-length GI video files without any operations, the technique does not report any false operations. Most significantly, the average analysis time per image was small, about 13 milliseconds (*ms*) on a PC with 3.4 GHz Intel® Xeon® and 16GB of RAM on our data sets. The analysis time is less than the time interval between two consecutive frames. Reporting detected operation scenes right at the end of the procedure is achievable without slowing down video capturing or other processing.

The remainder of the paper is organized as follows. Section 2 discusses related work on video segmentation techniques for endoscopic videos. We present our proposed technique in Section 3 and discuss the evaluation method and results in Section 4. We give the conclusion and the description of the future work in Section 5.

2 Related work

We limit our discussion of related work in optical endoscopic procedures. We exclude those of wireless capsule endoscopy (WCE) since therapeutic intervention with instruments during WCE procedures is not possible.

2.1 Endoscope hardware and endoscopic images

Within a single GI endoscopic procedure, instruments appear in the same general area in the video because the instruments are inserted through the same working channel from the top of the endoscope. The cable of the instrument has similar tubular shape regardless of various instrument types. The cable appears from one of the borders of the image. The cable may have different colors (e.g., green, blue, orange, and red), but it usually has some parts with high contrast. While the cable has a higher chance to be detected reliably, some operations do not have any cable present in the image because only the head of the instrument appears. A biopsy is typically very short between 2-4 seconds. Multiple biopsies in nearby mucosa are not uncommon.

2.2 Methods for endoscopic video segmentation

We proposed algorithms for blurry frame detection and shot segmentation of colonoscopic videos by color difference in 2004 [4]. Frames with similar color histograms in RGB color space are grouped in the same shot after blurry frames are discarded. These shots do not correspond to a biopsy or therapeutic operation. We proposed a method to segment shots based on camera motion into forward-camera-motion and backward-camera-motion shots [5]. Cumulative camera motion over time is used to find the frame separating the insertion and withdrawal phases. We later proposed a faster technique using motion vector templates [6].

We introduced algorithms for detection of operation shots [3, 7]. Hessian matrix and hierarchical clustering are used in [3] for detection of the insertion direction. This initial step is accurate, but is very slow. The detected insertion direction is used to discard regions outside the area of the detected direction. Moment invariants and Fourier shape descriptors are used in [3] and [7], respectively, for matching the detected regions with the cable template regions. The average processing time per frame with 390×370 pixel resolution once the insertion direction was identified was about 7*s* (seconds) on a PC with 3.40 GHz Pentium(R) 4 and 1GB of RAM. JSEG took 6*s* for segmentation of images into regions. The average true positive fraction and false positive fraction are 0.94 and 0.10, respectively. Only 3% of the actual shot boundaries was missed with 7 false shots on 25 videos of full-length colonoscopic procedures.

Shot segmentation by significant motion changes was proposed [9]. The method uses Kaneda-Lucas Tomasi (KLT) tracking on feature points in nine non-overlapping areas of the images. Motion within each area is calculated and smoothed over the same area in a temporal window. For each frame, the standard deviation *s* of the motion of all the areas is computed. The high value of *s* over a threshold signifies an object movement. Camera movement is determined with the high mean motion and low standard deviation. A candidate motion shot boundary is detected when the change in the standard deviation *s* over a time window is at peak. The detected boundaries are then further refined. The reported average recall and precision are same at 0.86. The processing time was not reported. The technique was not specifically designed to detect instruments in colonoscopy or EGD where instruments appear infrequently or not at all.

In 2004, we introduced a method for detection of endoscopic scenes: rectum, sigmoid, descending colon, transverse colon, ascending colon, and cecum scenes, each corresponding to different sections of the colon [8]. In 2014, a method to detect endoscopic scenes, each defined as one stable feature track of a tissue on the organ surface, was

proposed [10]. This method uses an optical flow enhanced with forward-backward tracking of SIFT features. Proposed Scale-Invariant Distance and Rotation Invariant Angle of pairwise relationships between landmarks in the same frame are used as features to compute a likelihood score to include a subsequent image in the same scene. Scene segmentation was performed on the fly. Average precision between 0.74 and 0.99 and maximum recall between 0.40 and 0.84 of endoscopic scenes were reported on four videos from Olympus Narrow Band Imaging endoscopes.

3 Cable footprint history technique

3.1 Preprocessing

We sample t images per second from the input video, forming what we called *reduced colonoscopy video*. The smaller the value of t, the larger the reduction in the processing time, but the larger the difference between the actual and detected scene boundaries. To reduce processing time, we sub-sample and classify pixels in each input image I_i of the reduced video into two sets: I_{ND}--- non-dark pixel set using Equation (1) where T_c is a constant in the range of 0 and 1 and I_D---dark pixel set for all the pixels excluded from I_{ND}.

$$I_{ND} = \{p |\ p\ is\ a\ pixel\ of\ I_i\ with\ all\ its\ normalized\ \ R,G,B\ values > T_c\ \} \qquad (1)$$

Next, for each image, we compute the median values of the non-dark pixel values in *CIE LUV* color space denoted as M_{LUV}. For each pixel p in the image, we compute the Euclidean distance $d(p, M_{LUV})$ [14] between the pixel values in *CIE LUV* color space of p and M_{LUV}. *LUV* is one of the uniform color spaces for which the distance function maps to perceptual distance well. We separate the foreground and the background using Equation (2) where T_F is a contrast threshold.

$$d(p, M_{LUV}) > T_F,\ \ d(p, M_{LUV}) = \begin{cases} \sqrt{\frac{(p - M_{LUV})^2}{3}} & if\ p \in I_{ND} \\ 0 & if\ p \in I_D \end{cases} \qquad (2)$$

We perform an erosion with a disk structuring element and remove regions that are too large or too small. Let R_i represent the i-th remaining connected component. We justify these threshold values based on experiments discussed in Section 4. This step replaces the time-consuming image segmentation method used in the previous algorithms [3,7]. We did not use the well-known Otsu method [14] to obtain a dynamic threshold value for each image because it wrongly considered cable regions as background for images with strong light reflection with higher contrast than cable regions in our training sets.

3.2 Spatial feature extraction

A number of shape features have been proposed [11] with varying degrees of computational complexity. Instead of using invariant moments [12] as in [3] or Fourier shape descriptors [7] to represent region shape, we introduce a new method based on the domain knowledge to calculate a new Cartesian coordinate system to derive region shape features.

By examining the cable regions from 17 endoscopic videos in our training video set, we further refine possible insertion directions into twelve general triangular areas as shown in Fig. 2(a). For each area, two of the borders are denoted by the two-headed arrow in the figure. The third border is a portion of the image border intersecting the first two borders. We define a new Cartesian coordinate system for each corresponding Area k as shown in Fig. 2(b-d). For instance, for $Area_{(2,5,8,11)}$, we choose X' perpendicular to the diagonal line and Y' perpendicular to X' as shown in Fig. 2(d). We derive four spatial features to represent the cable shape as follows.

First, we assign a cable region R_i to Area k denoted by the triangle $Area_k$ if more than half of the pixels of the region are in $Area_k$ as shown in Equation (3). The $|x|$ denotes the number of pixels in region x.

$$Area\ number\ = \{k | \frac{|R_i \cap Area_k|}{|R_i|} > \frac{1}{2}\} \qquad (3)$$

Next, we calculate the features using the area-based coordinate system. Equation (4) shows the eccentricity defined as the ratio of the cable region height in the longest extension of R_i along the Y' axis to the width (the longest extension of R_i along X') of a region R_i.

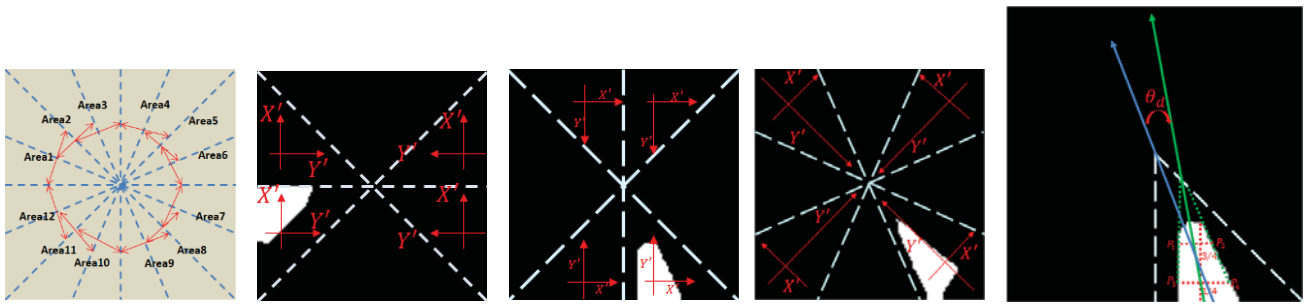
$$eccentricity = \frac{height}{width} = \frac{\max_{x'} \sum_{y'} R_i(x', y')}{\max_{y'} \sum_{x'} R_i(x', y')} \qquad (4)$$

Next, we calculate the orientation of the region. To get a reliable orientation, we only use the middle part of the region. We define two lines: P_3P_4 at the one-fourth height of the region and P_1P_2 at the third-fourth height of the region as shown in Fig. 2(e). The orientation vector of the region is calculated using Equation (5).

$$orientation\ vector\ = \frac{\overrightarrow{P_3P_1} + \overrightarrow{P_4P_2}}{2} \qquad (5)$$

$$d = \min_{(x',y') \in R_i} \{d(R_i(x', y'), border\ of\ Area_k)\} \qquad (6)$$

$$S = \frac{\sum_x \sum_y R_i(x,y)}{Number\ of\ all\ pixels\ in\ the\ image} \qquad (7)$$



(a) (b) for Areas 1,6,7,12 (c) for Areas 3,4,9,10 (d) for Areas 2,5,8,11 (e)

Fig. 2. (a) 12 general insertion areas; (b-d) new Cartesian coordinates; (e) insertion direction calculation

Now, we compute the angle difference θ_d between the region orientation vector and the angle bisector of $Area_k$. See an example θ_d in Fig. 2(e). We calculate the distance d from a region R_i to the border of its $Area_k$ defined in Equation (6) and the normalized area s of the region in Equation (7).

3.3 Feature classification

We investigated the effectiveness of the classification of non-cable/ cable regions using Support Vector Machine (SVM) with the linear kernel and the radial-basis-function kernel as well as J48 Decision Tree on the spatial feature vector ($\theta_d, d, s,$ and *eccentricity*). The parameter values of each type of classifiers were optimized to achieve the best performance as described in Section 4.

3.4 Cable footprint history

In Equation (8), we compute the weight w_i of the pixel at the coordinate (*x*,*y*) on image *i* using the corresponding binary image B_i where only pixels of the detected regions by the classifier have the values of one and the rest have zeros.

$$w_1(x,y) = B_1(x,y)$$
$$w_i(x,y) = B_i(x,y) * (w_{i-1}(x,y) + 1) \quad (8)$$

In other words, the weight of a pixel depends on whether it is part of the detected cable region in the current frame multiplied by one plus the weight of the corresponding pixel in the previous frame. The implication of this recursive equation is that the weight of this *x*-*y* location increases when it is part of the detected cable regions in consecutive frames. The weight is reset to zero whenever this location is not part of any detected cable regions. Other positive constant positive values instead of 1 can be used. We chose 1 for simplicity.

We compute the cable footprint history for each frame *i* from the first frame to the last frame of the video using Equation (9). Fig. 3(a-d) shows cable regions of the same video. Fig. 3(e) shows the cable footprint history of the last frame of the entire video. The brightest region marks the most common pixel locations of detected cable regions in the video.

$$H_1(x,y) = w_1(x,y)$$
$$H_i(x,y) = H_{i-1}(x,y) + w_i(x,y) \quad (9)$$

We use a binary threshold T_H to segment the cable footprint history of the last frame *t* of the video into a set of connected components. Let HR_j represent the *j*-th connected component in the set. We choose the brightest connected component $R_\#$ as the insertion area of this video as illustrated in Equation (10).

$$R_\# = \{HR_j \mid \max_{(x,y)\in HR_j} H_t(x,y) = \max_{(x,y)\in I_t} H_t(x,y)\} \quad (10)$$

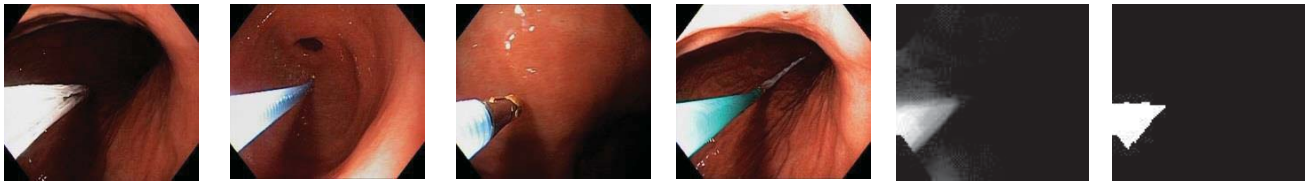
After locating the insertion direction for the video, we discard all candidate regions that do not intersect with $R_\#$ since they are not likely a true cable region. Finally, we assign each frame either 0 or 1. A frame is assigned a 0 if it does not have any remaining cable candidate region. Otherwise, we label it as 1.

3.5 Operation scene detection

This step utilizes temporal information and domain knowledge to identify operation scenes. This step accepts *L*, a sequence of 0 and 1 from the previous step, as input and outputs the frame numbers indicating the boundaries of the detected operation scenes.

3.5.1 Eliminate falsely detected cable images

This step corrects the misclassification results. We initialize the output sequence *L** with zeros. We slide a window of *W* frames on *L* from the beginning to the end of *L* one digit at a time. Each time, we compute the sum of all the



(a) (b) (c) (d) (e) (f)

Fig. 3. (a)-(d) Cable regions in one colonoscopy video; (e) cable footprint history of this video; (f) detected cable region

numbers under the sliding window. When the sum is equal to W (i.e., all the frames under the window are cable images), we copy all the numbers under the sliding window in L to L*. We set the window size W to $t/2$ where t is the temporal sampling rate in frames per second used in the pre-processing step. This window size covers frames within half a second since we observe that true cable frames typically appear consecutively more than half a second.

3.5.2 Locate cable scene boundaries

Like in our previous techniques [3,7], we scan L^* from the beginning to the end. We first determine a sequence S of consecutive frames from L^* with all the following properties.

1. The sequence S starts and ends with a 1, followed by at least $K * t$ consecutive 0s. In other words, the first and the last frames in S are cable images. The value of K should be the maximum temporal distance between consecutive cable shots of the same scene learned from training.
2. The sequence S must have the ratio between the total number of 1s (cable images) and the length of S greater than a threshold r_1.
3. The sequence S lasts at least 2 seconds based on a consultation with our endoscopist and our observation. A biopsy is typically short about 2-4 seconds. A scene can have multiple sequences.

If the temporal distance between two consecutive sequence S is less than T_t seconds, we group them in the same operation scene. We select the values of r_1 and T_t based on experiments discussed in Section 4.

4 Performance evaluation

The software for operation scene detection was implemented in Matlab. Weka was used for selecting the best classifiers for cable region classification. All the experiments were conducted on a PC with 3.4 GHz Intel® Xeon® and 16GB of RAM.

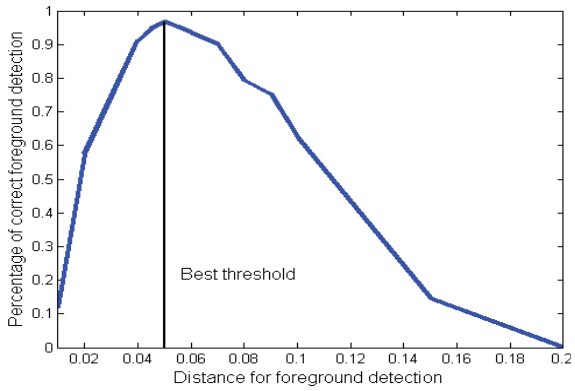


Fig. 4. Sensitivity analysis of the color contrast threshold values tested on cable images in the image

4.1 Data sets and parameter values

Table 1 shows the description of the data sets used in this study. The training data set consists of 17 endoscopic videos that are not in any of the test data sets. All the videos are of full length procedures captured in MPEG2 format. The first and last frames of each ground truth operation scene have cable images in them. The image set was for evaluation of classification effectiveness of cable regions using new features and existing features. The video sets I, II, and III do not overlap. The testing video sets were used to evaluate the effectiveness of the cable region detection and operation scene detection.

Table 1. Description about the data sets

Name	Number	Purpose
Training set		
Image set	1,000 cable images and 1,100 non-cable images	For cable region detection evaluation
Training video set	17 videos	For temporal parameters such as t, K, and T_t
Testing set		
Video set I	38 videos	Strong contrast between cable colors (blue, green, white)and the background
Video set II	18 videos	Weak contrast between cable colors (orange and red) and the background
Video set III	5 videos	No operation scenes

Table 2. Parameters and values used in experiments

Parameter	Value
Sampling rate t fps (frames per second)	6
Image resolution after spatial subsampling (pixels x pixels)	112x112
Dark pixel threshold T_c for preprocessing	0.3
Color contrast threshold T_F for preprocessing	0.05
Disk structuring element for erosion	5
Range of acceptable region size R_S in pixels	50<R_S<image area /14
Threshold T_H for determining true cable area	0.5
Ratio of cable frames in an operation shot (r_1)	0.1

The parameter values in Table 2 were determined from the training data in Table 1. We chose the temporal sampling rate of 6 *fps* to minimize the minimum distance between the true scene boundary and the detected scene boundary to 16 *ms*. For spatial subsampling rate, any higher rate, resulting in a smaller image resolution does not provide good classification result though it reduces the processing time. For the optimal color contrast threshold value, we plotted the percentage of correct foreground (cable region) detection with different threshold values using the cable images in the image

set. The plot in Fig. 4 shows that the color contrast threshold of 0.05 gives the highest correct foreground detection result. We observed that many cable images in an operation scene are difficult to detect for several reasons such as blurry images, strong light reflected regions with tubular shape, use of dye color, and too small cable regions. Therefore, we set the ratio of cable frames in an operation shot, r_1 to a small value of $\frac{1}{10}$.

Table 3. Performance and metrics

Metric	Definition
Sensitivity (SE)	Ratio of correctly detected cable images to cable images in the ground truth
Specificity (SP)	Ratio of correctly detected non-cable images to non-cable images in the ground truth
Precision	Ratio of correctly detected cable images to detected cable images
Number of false scenes (#F)	Number of software detected scenes not overlapped with any operation scenes in the ground truth
Number of missed scenes (#M)	Number of operation scenes in the ground truth for which the software miss both boundaries
True Positive Fraction (TPF)	Ratio of correctly detected images as part of operation scenes in the ground truth (true positives) to images of operation scenes in the ground truth
False Positive Fraction (FPF)	Ratio of incorrectly detected images as part of operation scenes (false positives) to images of operation scenes in the ground truth

4.2 Performance metrics

Table 3 shows the performance metrics used in this study. These metrics are similar to the metrics defined on operation shots in the previous work [3]. Note that if one of the two boundaries of an operation scene is incorrect, the detected operation scene still captures part of an operation. In such cases, we did not treat the detected operation scene as a false or a missed scene, but used TPF and FPF to quantify the effectiveness. High TPF is desirable as the algorithm uncovers most frames in the scenes. Low FPF indicates that a small fraction of a detected operation scene is not part of an operation scene.

4.3 Results

We first evaluated the effectiveness of our proposed spatial features. We used the grid search method to find optimal SVM parameters [15] in this experiment. Incorporating the cable insertion direction into the spatial high sensitivity (SE) of 99.9% and specificity (SP) of 97.9% with the decision tree classifier. Since colonoscopy has many more non-cable images than cable images, the 97.9% specificity is still inadequate, causing a large number of false cable images. This is where our cable footprint history is helpful.

Table 4. Performance of 10-fold cross-validation classification of cable images on the image set I

Classifier		SE(%)	SP(%)
New features	SVM + Linear	97.9	94.2
	SVM + RBF	99.6	97.7
	Decision Tree	**99.9**	**97.9**
Existing Features	Invariant moments	62.0	55.3
	Fourier descriptors	67.3	55.2

Table 5. Effectiveness of cable image detection on the video sets I & II

Method	Precision
Using proposed spatial features only	0.45
Proposed spatial features + cable footprint history	0.91

Table 4 shows that Decision Tree outperforms the two SVMs for detecting cable regions using Weka[16]. We chose Decision Tree as the classifier. Our new features perform better than the two existing features used in previous work [3,7]. The angle difference of the insertion direction is quite effective in discarding false regions.

Table 5 shows that the cable footprint history significantly reduces false cable regions by about half compared to using our spatial features only, resulting in doubling the average precision. In this experiment, we did not apply the operation scene detection step described in Section 3.5.

We use the video set I and II to test the effectiveness of our entire operation scene segmentation. The duration of these videos ranges from $4.7m$ to $60.9m$, which illustrates that the algorithm works well. The cable footprint history detects the insertion area correctly in all the videos in both video test sets. The TPF and FPF for the video set I are 0.929 and 0.022, respectively. TPF of 0.929 is very high. The TPF and FPF for video set II with weak cable contrast are 0.813 and 0.027, respectively. In video set II, the TPF is lower and the FPF is higher as the weak color contrast makes it difficult to differentiate the cable region from the background. The weak color contrast mainly results from two cases. First, the cable color is orange or red, which has very weak color contrast with the colon mucosa background. Second, the cable is exposed to the strong white light so that it is mixed with the background. There are 8 false scenes because of bright tubular looking colon surfaces or colon folds. The average number of false scenes per video is very low, 0.14. There are 12

missed scenes which were not detected mainly in two cases. The average number of missed scenes per video is 0.21. On video set III without any operation scenes, the algorithm performs very well. It did not generate any false operation scenes. The average processing time per frame from video sets I and II using our algorithm is 13*ms*. This is a great improvement by at least 38 times when compared to over 500*ms* using the previous method to detect operation shots given a known insertion direction [3]. This great improvement results from the cable extracting method based on the color contrast feature which costs only linear time.

5 Conclusion and Futrue Work

We introduce a new fast operation scene detection technique. The technique is based on color contrast to separate cable regions, new Cartesian coordinates for computing spatial features, and the cable footprint history. Our study shows that the technique is effective and at least 38 times faster than the existing technique. This is because detailed image segmentation or complex foreground-background subtraction is not needed. The experiments on real endoscopic videos demonstrate that the proposed technique misses a small number of operation scenes and it generates a very small number of false video segments that do not correspond to diagnostic or therapeutic operations. The algorithm is operation equipment, endoscope brand and procedure independent.

In our future work, we will improve the cable region extraction method to better separate the orange or red cables from the background. We will also design an online cable detection technique which outputs detected operation scene boundaries after each operation is complete. Such a method has the potential to be useful for analysis of quality of therapeutic operations where appropriate feedback is given for sub-optimal therapeutic operation

6 Acknowledgement

This work is partially supported by Iowa Regent Innovation Funds and EndoMetric Corp. Johnny Wong, Wallapak Tavanapong, and JungHwan Oh hold positions in EndoMetric Corporation, Ames, IA 50014, U.S.A, a for profit company that markets endoscopy-related software. De Groen is the medical advisor of EndoMetric.

7 References

[1] American Cancer Society. Colorectal Cancer Facts & Figures, 2014-2016.

[2] World Cancer Research Fund International. Colon cancer statistics and stomach cancer statistics. http://www.wcrf.org/int/data-specific-cancers.

[3] Y. Cao, D. Liu, W. Tavanapong, J. Wong, J. Oh, and P. C. de Groen. Computer-Aided Detection of Diagnostic and Therapeutic Operations in Colonoscopy Videos. IEEE Transactions On Biomedical Engineering, 54(7), July 2007.

[4] J. Oh, S. Hwang, W. Tavanapong, P. C. de Groen, and J. Wong. Blurry Frame Detection and Shot Segmentation for Colonoscopy Videos. In Proc. of IS&T/SPIE Conf. on Storage and Retrieval and Applications for Multimedia, pp. 531-542, San Jose, CA, USA, January 2004.

[5] S. Hwang, J. Oh, J. Lee, Y. Cao, W. Tavanapong, D. Liu, J. Wong, and P. C. de Groen. Automatic Measurement of Quality Metrics for Colonoscopy Videos. In Proc. of ACM Multimedia 2005, pp. 912-921, Singapore, November 2005.

[6] R. Nawarathna, J. Oh, J. Muthukudage, W. Tavanapong, J. Wong, and P. C. de Groen. Real-time Phase Boundary Detection for Colonoscopy Videos using Motion Vector Templates. In Springer Lecture Notes in Computer Science (MICCAI Workshop on Computational Abdominal Imaging, Computational and Clinical Applications). Vol. 7601, pp. 116-125, France, 2012.

[7] Y. Cao, D. Li, W. Tavanapong, J. Wong, J. Oh, and P. C. de Groen. Parsing and Browsing Tools for Colonoscopy videos. In Proc. of ACM Multimedia 2004, pp. 844-851, New York, NY, USA, October 2004.

[8] Y. Cao, W. Tavanapong, D. Li, J. Oh, P. C. de Groen, J. Wong. A Visual Model Approach for Parsing Colonoscopy Videos. In Proc. of Int'l Conf. on Image and Video Retrieval (LNCS 3115), pp. 160-169, Dublin, Ireland, July 2004.

[9] M. J. Primus, K. Schoeffmann, and L. Laszlo Böszörmenyi. Segmentation of Recorded Endoscopic Videos by Detecting Significant Motion Changes. In Proc. of CBMI 2013, June 2013.

[10] M. Ye, E. Johns, S. Giannarou, and G.-Z. Yang, Online Scene Association and Endoscopic Navigation, In Proc. of MICCAI 2014, Boston, MA, USA, Sept. 2014.

[11] Mingqiang Yang, Kidiyo Kpalma, Joseph Ronsin. A Survey of Shape Feature Extraction Techniques. Peng-Yeng Yin. *Pattern Recognition*, IN-TECH, pp.43-90, 2008.

[12] Invariant moments: N. Sebe and M. S. Lew, "Robust Shape Matching," In Proc. of IEEE International Conference on Image and Video Retrieval, London, UK, 2002, pp. 17-28.

[13] Poynton, Charles (2003). Digital Video and HDTV Algorithms and Interfaces, Morgan Kaufmann Publishers, page 226.

[14] Nobuyuki Otsu (1979). "A threshold selection method from gray-level histograms". IEEE Trans. Sys., Man., Cyber. 9 (1): 62–66.

[15] C.-W. Hsu, C.-C. Chang, C.-J. Lin, A Practical Guide to Support Vector Classification, Tech. Rep., Taipei, 2003.

[16] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.