

# Automatic Classification of Digestive Organs in Wireless Capsule Endoscopy Videos

Jeongkyu Lee<sup>1</sup>, JungHwan Oh<sup>2</sup>, Subodh Kumar Shah<sup>1</sup>, Xiaohui Yuan<sup>2</sup>, Shou Jiang Tang<sup>3</sup>

<sup>1</sup>Dept. Comp. Sci. & Eng.  
University of Bridgeport  
Bridgeport, CT 06604

{jelee,subodhs}@bridgeport.edu

<sup>2</sup>Dept. Comp. Sci. & Eng.  
University of North Texas  
Denton, TX 76203

{jhoh,xyuan}@cse.unt.edu

<sup>3</sup>Division of Digestive Diseases  
UTSW Medical Center  
Dallas, TX 75390

shou-jiang.tang@utsouthwestern.edu

## ABSTRACT

Wireless Capsule Endoscopy (WCE) allows a physician to examine the entire small intestine without any surgical operation. With the miniaturization of wireless and camera technologies the ability comes to view the entire gestational track with little effort. Although WCE is a technical breakthrough that allows us to access the entire intestine without surgery, it is reported that a medical clinician spends one or two hours to assess a WCE video. It limits the number of examinations possible, and incur considerable amount of costs. To reduce the assessment time, it is critical to develop a technique to automatically discriminate digestive organs such as esophagus, stomach, duodenum, small intestine, and colon. In this paper, we propose a novel technique to segment a WCE video into the distinctive organs based on color change pattern analysis. The basic idea is that the each digestive organ has different patterns of intestinal contractions that are quantified as the features. We present the experimental results that demonstrate the effectiveness of the proposed method.

## Categories and Subject Descriptors

I.2.10 [Artificial Intelligence]: Vision and Scene Understanding—*Video analysis*; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—*Time-varying imagery*

## General Terms

Algorithms, Experimentation, Software

## Keywords

Wireless capsule endoscopy, event boundary detection, event hierarchy, energy function, high frequency content function

## 1. INTRODUCTION

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SAC'07 March 11-15, 2007, Seoul, Korea

Copyright 2007 ACM 1-59593-480-4 /07/0003 ...\$5.00.

A human digestive system consists of a series of several different organs including the esophagus, stomach, duodenum, small intestine, colon and terminal ileum. Standard endoscopy has been playing a very important role as a diagnostic tool for the digestive track. For example, various endoscopies such as colonoscopy, upper gastrointestinal endoscopy, push enteroscopy and intraoperative enteroscopy have been used for the visualization of digestive system. However, all methods mentioned above are limited in viewing small intestine. It is hard to reach with instruments passed by either the mouth or the anus because it is located between the stomach and the large bowel. To address the problem, Wireless Capsule Endoscopy (WCE) was first proposed in 2000, which integrates wireless transmission with image and video technology [13, 2, 1, 12, 14, 8]. After FDA approved in 2002, it is now used as one of the important tools to examine small intestine.

WCE allows a physician to examine the entire small intestine non-invasively. WCE uses a small capsule, 11 mm in diameter and 25 mm in length (see Figure 1 (a)). The front end of the capsule has an optical dome where white light emitting diodes (LEDs) illuminate the luminal surface of the gut, and a micro camera sends images via wireless transmission to a receiver worn by a patient (see Figure 1 (b)). Another part of the capsule contains a small battery that can last up to 8 hours. The patient swallows a small capsule, usually after an overnight fast. As the capsule moves through the gastrointestinal tract, images are transmitted by the digital radio frequency communication channel to a data recorder (see Figure 1 (c)). This data are transferred to a computer for interpretation by the specialists. The capsule is then passed in the patient's stool and discarded.

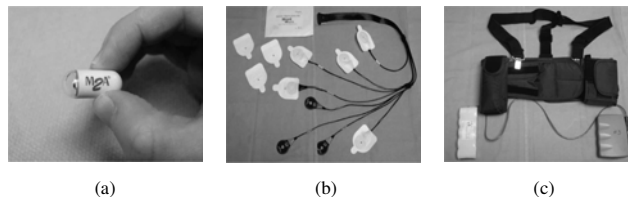


Figure 1: Wireless Capsule Endoscopy Equipments: (a) the capsule, (b) the 8-lead antenna array, and (c) receiver/recorder unit and battery.

Although WCE is a technical breakthrough that allows to access the entire intestine without surgery, it is reported

that the interpretation time takes 1 or 2 hours. This is a heavy load for the specialist, i.e. the gastroenterologist. It limits the number of examinations possible and incur considerable amount of costs. To reduce the assessment time, Berens et al. proposed the technique for automatically discriminating stomach, intestine, and colon by using the Discrete Cosine Transform (DCT) and Principal Component Analysis (PCA) classifiers [1]. However, the technique can detect only two boundaries of stomach/intestine and intestine/colon, which is not enough to assist the specialist. In [14], ROC curves (receiving operating characteristic) analysis are used for the classification of contraction and non-contraction images in WCE videos. However, it is not based on the contents of WCE video.

In this paper, we propose a novel algorithm for event boundary detection in Wireless Capsule Endoscopy videos based on an energy function. The basic idea is that each digestive organ such as esophagus, stomach, duodenum, small intestine, and colon, has different patterns of intestinal contractions. These patterns have been widely used in the analysis of biogenic signals, such as Electrogastrogram [7] or Electrocardiogram [10]. We first characterize the contractions of WCE video using Energy function in a frequency domain. Then, we segment WCE video into events by using a high frequency content (HFC) function. The detected event boundaries indicate either entrance of the next organ or unusual events in the same organ, such as intestinal juices, bleedings, and unusual capsule movements. We classify the segmented events into higher level events that represent digestive organs. The classification result is represented by a tree structure, which is called an event hierarchy of WCE.

The remainder of this paper is organized as follows. Feature extraction and event boundary detection in WCE video are discussed in Section 2. Section 3 presents the proposed technique for building event hierarchy of WCE video. In Section 4, we discuss our experimental results. Finally, Section 5 presents some concluding remarks.

## 2. EVENT BOUNDARY DETECTION IN WCE VIDEOS

In this section, we first discuss the intestinal contractions of which patterns are the most important discriminator of digestive organs. In order to characterize the contractions, we extract energy-based feature in frequency domain from WCE images, and then detect event boundaries by using a high frequency content (HFC) function.

### 2.1 Intestinal Contractions

Contractions are one of the important motility patterns in bowel movements, which is also the basic activity throughout entire gastrointestinal tract. Since the intestinal contraction is a very good pathological indicator, it is widely used for the diagnosis of many gastrointestinal diseases. One of examples using contractions is Electrogastrogram [7] (EGG). EGG is a non-invasive recording of the electrical activity of stomach. Slow waves are generated from the activity of the gastrointestinal wall surface, or gastrointestinal contractions.

WCE videos can record continuous activities in digestive system such as intestinal contractions, and visualize them easily. Figure 2 shows a number of examples of intestinal contractions taken by WCE. Figure 2 (a) and (b) are se-

quences of frames captured from stomach and small bowel, respectively. As seen in the figure, the motility patterns are different since each digestive organ has different types of movements and functionalities. We will characterize intestinal contractions to find events in WCE videos in the following subsections.

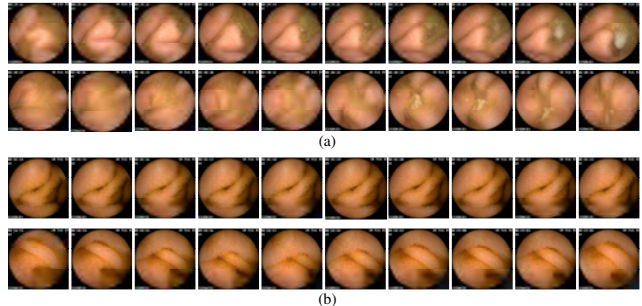


Figure 2: Sequence of Frames of WCE Video: (a) stomach and (b) small bowel.

### 2.2 Feature Extraction

To characterize the contractions in WCE videos, we tested several different features. Many works have tackled the problem of the characterization of intestinal contractions in capsule endoscopy using various features [1, 12, 14]. In [14], 34 features are extracted to describe the contractions along 9 frames: 9 mean intensities, 9 hole sizes, 9 global contrasts, 6 correlations among frame sequences, and 1 variance of intensities. However, the approach requires high computations to extract many features from videos, and pre-processing to test the data, which prevents on-line processing.

We mainly focus on color features in order to pursue on-line processing. Since colors are the only feature values captured by a camera directly, it can reflect the activities of digestive system effectively such as contractions and juices. In other words, intestinal movements, i.e. contractions, could change the color values. Among the various color domains, we select HSI color space. The reason we choose HSI is that hue, saturation and intensity are most robust components for video and image processing [3].

First, we convert the RGB color space into the HSI color space for every frame in WCE video. Since the intestinal contractions that are periodic movements affect all components in the color space, we can use either any components among hue, saturation and intensity, or any combinations of components in HSI. In this paper, we use mean of whole pixel values in a frame. Figure 3 shows the average values of HSI for the first 5000 frames of a sample WCE video.

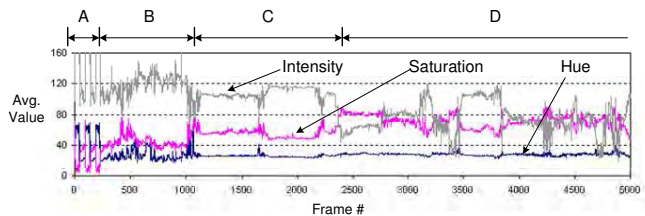


Figure 3: Average values of HSI colors for the first 5000 frames of a sample WCE video.

The blue, pink, and gray lines in the figure indicate the average values of hue, saturation, and intensity of each frame in a video, respectively. We marked the actual sections of digestive system. For example, ‘A’ indicates esophagus, ‘B’ indicates stomach, and ‘C’ indicates duodenum. Also, ‘D’ is the part of small bowel. It is clear that each part of WCE video, i.e. ‘A’ to ‘D’, has different pattern of color sequence values. We refer the sequence of color values as *color signal* of WCE video because it has the same properties as a signal such as frequency and wave length. Contractions that are periodically occurred in digestive tube distort the color values, which make the signal of color values in WCE video. The characteristics of color signal in WCE video can be summarized as follows:

- Unlike EGG that is an electrical activity from the gastrointestinal wall surface as well as the contractile activity of the smooth muscles, the color signal is caused by only the contraction of digestive movements, and
- When a capsule enters the next digestive organ, the corresponding color signal has a short-term change that is the suddenness of the signal change and the increase in energy.

The aforementioned characteristics will be used to define detection function of WCE video events in the following subsection. In this paper we choose the color signal generated from intensity value (gray signal in Figure 3) of HSI color domain for the efficiency of processing.

### 2.3 Event Boundary Detection

A shot is very useful processing unit in many video applications such as video segmentation, video indexing, and video annotation. However, it cannot be applied into non-produced videos such as video surveillance systems and medical videos, since they are taken without any stops or pauses. To address it, we define an *event* of WCE videos as follows:

**Definition 1.** *An event of WCE videos is a sequence of continuous frames that include the same semantic contents.*

The examples of events in WCE videos are each digestive system (i.e. esophagus, stomach, duodenum, small intestine, and colon), and anomaly (i.e. bleeding and juices). Since one event has the same semantic contents including the similar digestive movements, the color signal in a single event has the similar pattern. Therefore, we can find the event boundaries of WCE videos based on a signal processing.

We design the event detection method by recognizing two signal properties associated with a short-term change, which is the suddenness of the signal change, and the increase in energy [11]. In addition, we choose a *frequency domain method* since it is able to reveal not only changes in overall energy, but also the energy concentration in frequency [11]. The frequency location of energy is very important since the sudden changes in the signal cause phase discontinuities. In the frequency spectrum, this appears as high frequency energy. We define the energy function of color signal,  $E$  as the sum of the magnitude squared of each frequency bin in the specified range. The energy function of the  $i^{th}$  frame of WCE video,  $E_i$  is defined as:

$$E_i = \sum_{k=2}^{\frac{N}{2}+1} (|X_i(k)|^2) \quad (1)$$

where  $N$  is the FFT (fast fourier transforms) array length, and  $X_i(k)$  is the  $k^{th}$  bin of the FFT. In Equation (1),  $\frac{N}{2} + 1$  indicates the frequency  $\frac{F_S}{2}$  where  $F_S$  is the sample rate. The function to measure high frequency content is defined as a weighted energy function, which is linearly increased toward the higher frequencies. The high frequency content function of the  $i^{th}$  frame of WCE video,  $HFC_i$  is defined as:

$$HFC_i = \sum_{k=2}^{\frac{N}{2}+1} (|X_i(k)|^2 \cdot k) \quad (2)$$

where  $k$  is a weight of the energy to increase the higher frequencies. In Equation (1) and (2), we ignore the lowest two bins in order to avoid unwanted bias from low frequency components. Now, we can define the condition for event detection in WCE videos by combining the results of Equation (1) and (2). The condition for event detection in WCE videos are as follows:

$$\frac{HFC_i}{HFC_{i-1}} \cdot \frac{HFC_i}{E_i} > T_{event} \quad (3)$$

where  $T_{event}$  is a empirical threshold value. If  $E_i$  and  $HFC_i$  in the current frame satisfy the condition in Equation (3), an event boundary is detected in between the  $i^{th}$  and  $(i-1)^{th}$  frames. Otherwise, both of the frames are in the same event. To avoid the potential divided-by-zero error the denominators in Equation (3), i.e.  $HFC_{i-1}$  and  $E_i$  have a minimum value of one. The detection function in Equation (3) is the product of the rise in high frequency energy between the two frames and the normalized high frequency content for the current frame. Some sample results of the event detection function are shown in Figure 4. Figure 4 (a) is an original color signal for frame # 0 to frame # 5000 in a sample WCE video, which is around 42 minutes long. The energy and HFC for each frame are plotted in Figure 4 (b). In Figure 4 (c), the event detection function in Equation (3) finds 14 events from the color signal.

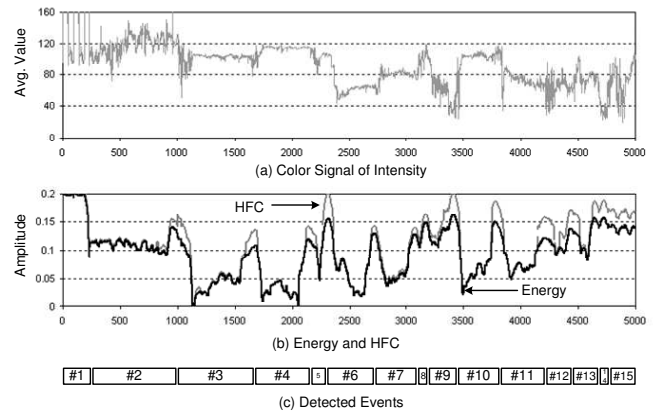


Figure 4: Results of event detection for WCE.

## 3. AUTOMATIC GENERATION OF EVENT HIERARCHY

In the previous section, event boundaries of WCE video are detected by the energy-based detection function. However, it is possible that a single event can be divided into several events because of local maxima of the color signal and

threshold value. In order to find exact boundaries of every digestive organs, we need to merge these events into a single event. For the merge, we apply bottom-up approach to build a tree. We refer to the constructed tree as an event hierarchy of WCE video. Using the energy function in Equation (1), we determine the correlation between two events,  $Event_i$  and  $Event_j$ . The correlation between  $Event_i$  and  $Event_j$ ,  $Corr(Event_i, Event_j)$ , can be determined as follows:

$$Corr(Event_i, Event_j) = \begin{cases} true & \text{if } \sum_{k=2}^{\frac{N}{2}+1} (|X_i(k) - X_j(k)|^2) > T_{corr}, \\ false & \text{otherwise.} \end{cases} \quad (4)$$

where  $X_i(k)$  and  $X_j(k)$  are the  $k^{th}$  bins of the FFT from  $Event_i$  and  $Event_j$ , respectively.  $T_{corr}$  is an empirical threshold value. When  $Event_i$  and  $Event_j$  have the same pattern of color signal, two events are related to each other. Otherwise, they are not related.  $Corr()$  is used for constructing an event hierarchy. Event hierarchy is a tree structure represented as nodes and levels. Each node in the hierarchy is called an event node labeled as  $EN_i^m$ , where the subscript denotes an event, and the superscript indicates a level of the node in the hierarchy. An event hierarchy is based on a scene tree in [9], and the procedure is as follows:

1. Create an event node  $EN_i^0$  for each  $Event_i$
2. Set  $i \leftarrow 2$ .
3. Apply an event correlation  $Corr()$  to check if  $Event_i$  is similar to  $Event_{i-1}, \dots, Event_1$  in descending order. The comparison stops when a related event, say  $Event_j$ , is identified. If no related shot is found, we create a new empty node, connect it as a parent node to  $EN_i^0$ , and proceed to step 5.
4. We connect all event nodes,  $EN_i^0$  through  $EN_j^0$ , to a parent node of  $EN_j^0$ .
5. If more events, we set  $i \leftarrow i + 1$ , and go to step 3.
6. For each event node at upper level, we select them as an boundaries of digestive organs.

Figure 5 illustrates an example of the event hierarchy construction. We consider a WCE video with seven events as shown in bottom of Figure 5. Let  $Event_0$  and  $Event_2$  be correlated, and  $Event_3, Event_5$ , and  $Event_6$  be correlated, respectively. We have two event nodes in upper level,  $EN_0^1$  and  $EN_3^1$ , which indicates that the video has two different organs. The dotted line in Figure 5 shows its boundaries.  $Event_1$  and  $Event_4$  might be some anomaly, such as bleedings or juices.

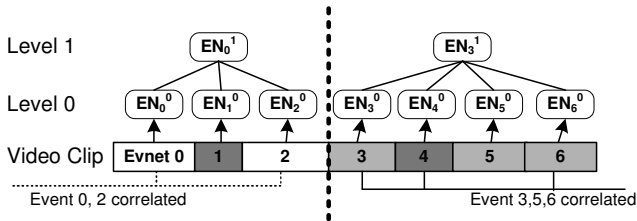


Figure 5: Example of constructed event hierarchy.

## 4. EXPERIMENTAL RESULTS

To assess the proposed methods, we have performed the experiments with ten wireless capsule endoscopy videos. The data set was prepared in the following way. First, the specialist (Dr. Tang) examined the ten videos, and annotated them for the following information: entering the next digestive organ, such as the entering stomach, duodenum (small intestine), ileum, and cecum (large intestine), bleedings in intestines, and other anomalies. Second, we take off non-informative part of each image, which is usually black area outside of circle. For the evaluation metrics, we employ ‘recall’ and ‘precision’, which come from Information Retrieval [4]. We evaluate the performance of the proposed schemes by demonstrating that:

- The proposed event boundary detection technique using Energy and High Frequency Contents (HFC) function can detect and classify accurately transitions of events in WCE videos.
- The proposed event hierarchy can provide the boundaries of digestive organs each of which has different types of intestinal contraction.

Our experiments are performed on an Intel Pentium IV 3.0 GHz CPU and 960 MB memory computer with Java 2 SDK 1.4.2 and JMF 2.1e.

### 4.1 Performance of the Event Boundary Detection

To assess the performance of our energy-based event boundary detection algorithm (EG-EBD), we compare it with a histogram based comparison technique (HS-EBD). Among many existing techniques based on color histogram, we choose a selective HSI histogram comparison algorithm proposed in [6, 5], since it outperforms both histogram (gray level on global and local) and pixel differences approaches for the temporal video segmentation. The details of our dataset, i.e. wireless capsule endoscopy videos, and event boundary detection results are given in Table 1. In this table, the numbers in the third column indicate the numbers of annotated information by the specialist, such as location of digestive organs, and active bleedings. We use ‘Recall’ ( $H_r$ ) and ‘Precision’ ( $H_p$ ) mentioned above to verify the performance of those techniques. The higher recall indicates a higher capacity of detecting correct events, while the higher precision indicates a higher capacity of avoiding false matches. In medical domains, recall is more important than precision since it is not desirable to loose any true positive.

The results given in Table 1 show that the averages of  $H_r$  and  $H_p$  using EG-EBD are 0.76 and 0.51, respectively. The recall of EG-EBD is two-times better than that of HS-EBD. For the precision, that is three-times better than HS-EBD.

### 4.2 Event Hierarchy

In order to evaluate the performance of the event hierarchy for WCE videos, we compare the results of the upper level in the event hierarchy with the annotated information provided by a specialist. Since the upper level of event hierarchy is built by merging the correlated events, it eventually represents the digestive tract. Table 2 shows the results of digestive organs using event hierarchy. We focus on four different events: ‘entering stomach’, ‘entering duodenal’ (i.e. small intestine), ‘entering ileum’, and ‘entering cecum’ (i.e.


**Table 1: Results of Event Boundary Detection**

Video No.	Video Name	Actual # of Events	EG-EBD		HS-EBD	
			Recall ( $H_r$ )	Precision ( $H_p$ )	Recall ( $H_r$ )	Precision ( $H_p$ )
1	PLT	17	0.82	0.50	0.35	0.09
2	LZ	26	0.81	0.49	0.38	0.12
3	WD1	23	0.70	0.43	0.35	0.11
4	CS	28	0.61	0.49	0.43	0.17
5	WD2	25	0.84	0.62	0.28	0.08
6	TG	93	0.74	0.52	0.16	0.15
7	JB	52	0.73	0.56	0.35	0.22
8	GT	50	0.84	0.56	0.28	0.19
9	JM	37	0.76	0.52	0.32	0.18
10	ME	13	0.77	0.32	0.46	0.10
		364	0.76	0.51	0.30	0.14

large intestine). The ‘O’ and ‘X’ in the cells indicate the correct detection, and false detection, respectively. The colored cell means that no annotated information is provided by the specialist. We allow 5 seconds for the error margin. As seen in table, the proposed scheme can detect the most of stomach and duodenum. However, the accuracies of ileum and cecum are less than that of other parts of digestive systems. Because of the limitation of battery, the pictures taken in large intestine have low quality. However, since a doctor uses WCE to examine small intestine rather than large intestine, this is not a critical issue.

**Table 2: Results of digestive organs detection using Event Hierarchy**

Video No.	Video Name	Entering Stomach	Small Intestine		Large Intestine
			Entering Duodenal	Entering Ileum	Entering Cecum
1	PLT	O	O	*	O
2	LZ	O	O	*	O
3	WD1	O	O	*	X
4	CS	O	O	X	O
5	WD2	O	O	*	O
6	TG	O	O	*	X
7	JB	O	O	O	*
8	EE	O	O	*	*
9	JM	O	X	*	O
10	ME	O	O	X	O

\*  No annotated information are provided by a specialist.

## 5. CONCLUSIONS

Finding events in Wireless Capsule Endoscopy videos such as entering the next digestive organs and detecting active bleedings, is a major concern when a gastroenterologist reviews the videos. It is reported that the reviewing time takes 1 or 2 hours. This is a heavy load for the specialist, which will limit the number of examinations, and incur considerable amount of costs. In this paper, we propose a novel algorithm for event boundary detection in WCE videos based on energy of contractions. We first characterize the contractions of WCE video using Energy function ( $E$ ) in a frequency domain. Then, we segment WCE video into the events by using a high frequency content ( $HFC$ ) function. The detected event boundaries indicate either entrance of the next organ or unusual events. We classify the segmented events into higher level events that represent digestive organs, called an event hierarchy of WCE video. Our

experimental results indicate that the recall and precision of the proposed event detection algorithm reach up to 0.76 and 0.51, respectively.

## 6. REFERENCES

- [1] J. Berens, M. Mackiewicz, and D. Bell. Stomach, intestine and colon tissue discriminators for wireless capsule endoscopy images. In *Proc. of SPIE Conference on Medical Imaging*, volume 5747, pages 283–290, Bellingham, WA, 2005.
- [2] G. Bresci, G. Parisi, M. Bertoni, T. Emanuele, and A. Capria. Video capsule endoscopy for evaluating obscure gastrointestinal bleeding and suspected small-bowel pathology. *J Gastroenterol*, 39(8):803–806, August 2004.
- [3] Y. Du, C.-I. Chang, and P. D. Thouin. Unsupervised approach to color video thresholding. *Optical Engineering*, 43(2):282–289, 2004.
- [4] W. B. Frakes and R. Baeza-Yates. *Information Retrieval - Data Structures and Algorithms*. Prentice Hall, Englewood Cliffs, 1992.
- [5] Y. Gong, H. Chua, and X. Guo. Image indexing and retrieval based on color histogram. In *Proc. of Int’l Conf. Multimedia Modeling*, pages 115–126, Singapore, Nov. 1995.
- [6] J. Hafner and et al. Efficient color histogram indexing for quadratic from distance function. In *IEEE Transaction on Pattern Analysis and Machine Intelligence*, pages 729–736, 1995.
- [7] P. Hubka, V. Rosik, J. Zdinak, M. Tysler, and I. Hulin. Independent Component Analysis of Electrogastrographic Signals. *MEASUREMENT SCIENCE REVIEW*, 5(2):21–24, 2005.
- [8] S. Hwang, J. Oh, J. Cox, S. J. Tang, and H. F. Tibbals. Blood detection in wireless capsule endoscopy using expectation maximization clustering. volume 6144. SPIE, 2006.
- [9] J. Lee, J.-H. Oh, and S. Hwang. Scenario based dynamic video abstractions using graph matching. In *ACM Multimedia*, pages 810–819, Singapore, November 2005.
- [10] M. Masaru Suzuki, S. Hori, T. Funabiki, T. Masaoka, R. S. Hirota, and N. Aikawa. Electrocardiogram Signal De-noising Using Multiple Auto-filtering. *Acad Emerg Med*, 13(5):S187, 2006.
- [11] P. Masri and A. Bateman. Improved modelling of attack transient in music analysis-resynthesis. University of Bristol., 1996.
- [12] P. Spyridonos, F. Vilarino, J. Vitria, F. Azpiroz, and P. Radeva. Anisotropic Feature Extraction from Endoluminal Images for Detection of Intestinal Contractions. In *Proceedings of the 9th MICCAI*, Copenhagen, Denmark, October 2006.
- [13] S. Tang, R. Jutabha, and D. Jensen. Push enteroscopy for recurrent gastrointestinal hemorrhage due to jejunal anastomotic varices: a case report and review of the literature. In *Endoscopy*, volume 34, pages 735–7, 2002.
- [14] F. Vilarino, L. I. Kuncheva, and P. Radeva. ROC curves and video analysis optimization in intestinal capsule endoscopy. *Pattern Recognition Letters*, 27:875–881, 2006.