Vol.2009-CVIM-167 No.5
2009/6/9

Video Capsule Endoscopy Analysis For Diagnostic Assistance

Hai Vu,[†1] Tomio Echigo,[†2] Ryusuke Sagawa,[†1] Keiko Yagi ,[†3]
Masatsugu Shiba,[†4] Kazuhide Higuchi,[†4] Tetsuo Arakawa[†4]
and Yasushi Yagi[†1]

Video capsule endoscopy (VCE) represents a significant advance in examinations of digestive diseases by providing a non-invasive method to view the small bowel. In addition, VCE provides a valuable source for visualizing the intestinal contractions, which are mainly events for intestinal motility assessment. However, the advantages of VCE diagnosis technique are facing with the time consuming for reading video sequence as well as challenging to detect the intestinal contractions. In this paper, we present our works to approach these motivations through techniques of VCE analysis. VCE interpretations could be implemented by analyzing spatial and temporal features. First, several image features such as color, edges, and motion displacement are extracted. Then their temporal analyses are presented in several ways to adapt with different tasks. Two applications utilizing this framework are developed. In the first application, we propose a new method to reduce diagnostic time under the constraint that all original images should be displayed to an examining doctor without skipping frames. To realize such a system, delay time for drawing images between frames is controlled in adaptive rate, according to the states of capturing images. Several techniques for the state classification, delay time calculation, and log-based analysis are deployed in this application. In the second application, we develop a three-stage procedure for the intestinal contraction detections. Based on the characteristics of contractile patterns, the possible contractions can be investigated using essential images features extracted from VCE such as changes in edge of the intestinal folds and by evaluating similarities features in consecutive frames. Then true contractions are determined through spatial analysis of directional information. To exclude as many non-contractions as possible, we consider about information of contractions frequencies along capsule transit time. Both the quality and quantity indices are analyzed in experiments for performance evaluations.

1. Introduction

Visualization of the small bowel in human has remained limitations of invasive and non-complete diagnosis for several decades. Conventional investigations involve the introduction of a probe into the patient's gastrointestinal (GI) tract. These techniques are highly invasive, and can cause significant patient discomfort, due to the long distances to be examined and the loop configuration of the small bowel. Recently, video capsule endoscopy (VCE[18)]) is a new diagnostic technology which for the first time allows to implement non-invasive examinations throughout small bowel regions. VCE diagnosis is particularly successful in finding causes of GI bleeding of obscure origin, Crohn's disease, and suspected tumors of the small bowel that are difficult to carry out by conventional endoscopic techniques.

In this paper, we present our research for video capsule endoscopy diagnostic assistance. VCE diagnosis involves reading a large number of frames, that is under extreme and careful examinations by examining doctors. Reducing diagnostic time is the main challenge for bright scenario of the VCE diagnosis. In addition, in term of GI physiology, VCE presents a valuable image source for visualizing the intestinal contractions. Recognizing the intestinal contractions from VCE provides a non-invasive method for the GI motility assessment because relevant information such as the contraction position, their frequency institutively reflect intestinal motility patterns during capsule trajectory. With such objectives, VCE interpretations are main factors to develop intelligent systems and/or computer-assisted tools. In this

†1 The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Osaka, 567-0047, Japan
†2 Dept. of Engineering Informatics, Osaka Electro-Communication University, 18-8 Hatsu-cho, Osaka, 572-8530, Japan
†3 Kobe Pharmaceutical University, 4-19-1 Motoyamakita-machi, Hyogo, 658-8558, Japan
†4 Graduate School of Medicine, Osaka City University 1-4-3 Asahimachi, Osaka, 545-8585, Japan

work, we tackle utilizing spatial and temporal image features extracted from VCE. These features can be investigated under pattern recognition schemes. Results are highly interesting in the diagnostic assistance because of effectively reducing examination time as well as automatically recognizing intestinal contractile patterns.

2. Related works

The sections below briefly introduce methods in order to suggest computer-assisted applications for VCE diagnostic assistance. Based on the purposes of these works, we categorize research into four groups: image display supporting, digestive organ segmentation, the intestinal motility assessments, and abnormal regions detections.

- Image display supporting: Reading and interpretation VCE are very important tasks in clinical examinations for examining doctors. However, the tasks become more difficult because of some specific features of the capsule endoscopic images, e.g, luminal regions is common, angle of view is non-adjustable, or finding regions are sometimes only noted on single frames, etc. Several works proposed techniques to assist examining doctors such as: image enhancement by BaoLi et al.[7]; image distortion corrections in works of C.Hu et al.[6]; supporting a glanced view through 2-D map generated from an image sequence by P.M.Szczypinski et al. in[32),33)]; or automatically controlling image display by Vu et al. in[37)]. Baopu Li et al.[7)] is aimed at improving in the quality of a given image.

- Digestive organ segmentation: In human body, the GI tract includes digestive organs such as esophagus, stomach, small intestine and colon regions. Finding the points in the video can be difficult and time-consuming, even for an experienced viewer, e.g., images from the stomach and intestine regions around the pylorus appear visually highly similar. Approaching this issue, a series works of Combra et al. in[8),9),24)], Lee et al. in[20)], Mackiewicz et al. in[22),23)] proposed several methods for automatically segmenting the digestive organs. Institutively, results would reduce the amount of time taken by a doctor to examine a VCE sequence.

- Abnormal regions detections: supporting abnormal regions detections seemly is the most complicated in field of the computer assistance for VCE diagnosis because of diverse appearances of the suspicious regions and passively moving of the capsule device. This topic is firstly presented in works of Boulougoura et al. in[4)]. Recently,[3)] also reports a new scheme for detecting four types abnormal regions. There is a point sharing that combination features of texture and color is effective to discriminate normal and abnormal images. Results in[3),4)] implied promising results for detecting abnormal regions from VCE sequence. However, in fact these works utilize a limited database in their experiments.[4)] uses only 73 clinical images (including 33 abnormal and 38 normal images), and[3)] use 75 images (including 41 normal and 34 normal images).

- Intestinal motility assessments: because VCE is a passive device, it is ingested and propelled by peristaltic waves through the GI tract, the image sequence captured in VCE transit time institutively provide valuable image source for visualizing intestinal motility. Utilizing VCE for intestinal motility assessments was carried out in[11),27),28),36)] by a research team in Computer Vision Center, Universitat Autonoma de Barcelona, Spain. In their works, they claimed in[36)] an important issue for the intestinal contraction detections from VCE is that the prevalence of the intestinal contractions in an image sequence is very low, showing a ratio about 1:50. With this imbalanced issue, we are facing a large number of false positives, or a large number of non-contraction could be labeled as contractions at the output results of classifiers.

3. Video Capsule Analysis Techniques

3.1 Spatial-temporal combinations for VCE interpretations

The capsule endoscopic image features include a standard size of 256 x 256 pixels, 8 bit per channel in RGB color space. Common appearance of a capsule endoscopic image is shown in Fig. 1, in which available information appears in a circle with radius 224 pixels. Fig. 1 shows common parts of the GI tube that consist of intestinal wall regions, intestinal folds and darkness regions of intestinal lumen. These appearances are differences in texture, shape, color, and strongly depend on the digestive or-

gans along transit of the capsule device. For example, gastric regions usually appear with fold shapes of the stomach wall, colon images present large surface without villus-based texture. Because small bowel is the most important region in VCE diagnosis, the feature extractions in this section focused on appearances of the images in this region.

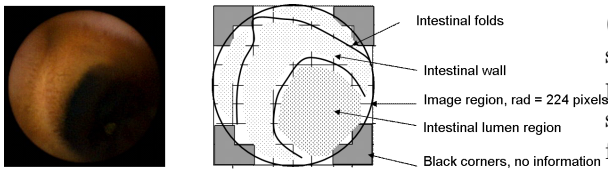


Fig. 1 A common appearance of a small intestinal image. Left panel: an original image. Right panel: a schematic view with corresponding appearance of the original image

In our opinion, the image features in a still image and their temporal changes play important role to interpret VCE sequence. Because a VCE sequence actually involves random images, that are captured under natural movements of the capsule device. The states of capturing images depend on the natural characteristics of intestinal motility patterns. Therefore it is difficult to suggest a visual scene or a specific object from VCE sequence. This led us to a general framework for VCE analysis including two levels: the image feature extractions and their temporal analysis, as shown in Fig. 2 - left panel. Based on appearance of images in VCE sequence, if we assume that an image feature χ_i varying along capsule transit time can be expressed by a function $f(\chi_i, t)$, then analysis of this function in temporal dimension can provide meaningful data, e.g., to measure state of image acquisition or to describe constituent circle of the intestinal contractions. Utilizing spatial image features χ_i in temporal dimension could be seen in the adaptive speed technique for a feature vector combining color similarity and motion displacement between two consecutive frames, or changes in edge of the intestinal folds of several adjacent frames.

VCE is suffered from non-object and non-scene characteristics, therefore we do not make efforts to detect these relationships in VCE analysis. Instead of that, temporal analysis is to show relationships between images, then results can be used in pattern recognition schemes. These relationships are taken into account in different way, depending on task of the application. For instance, the right panel in Fig.2 presents how the framework can be utilized to determine states of image acquisitions. As shown, the image features are extracted to make measurement of disparity of two consecutive frames. Through a classifier, the states (presented in same color) are determined. Besides evaluations of spatial features along temporal dimension, the motion displacement (a simple case is motion between two consecutive frames) is taken into account in our works because the motion feature directly carries a lot of information about spatio-temporal relationships between image objects.

3.2 Color information

The use of color histograms is a promising way of quickly indexing a large number of frames, such are found in a VCE sequence. In our implementations, capsule endoscopic image is divided into small blocks and a histogram is computed for each block. The color histogram method[30] is applied to each block by dividing R, G, B components into a number of bins $N_{bins} = 16$. The distance of the local histograms is computed from the L1 distance, $D_{blk}(i)$:

$$\sum_{k=1}^{N_{bins}} (|H^n_{R,k} - H^{n+1}_{R,k}| + |H^n_{G,k} - H^{n+1}_{G,k}| + |H^n_{B,k} - H^{n+1}_{B,k}|) \quad (1)$$

where H is the histogram of each color component for block i and between frames $< n, n+1 >$.

Block matching between frames $< n, n+1 >$ is decided using a selected threshold value. The accumulation of matching blocks reveals overall similarity between two frames:

$$Sim(n) = \frac{1}{N_{blk}} \sum_{i=1}^{N_{blk}} sim_{blk}(i)$$
$$\text{With} \begin{cases} sim_{blk}(i) = 1 \text{ if } D_{blk}(i) \leq T_{blk} \\ sim_{blk}(i) = 0 \text{ otherwise} \end{cases} \quad (2)$$

Block size value was decided heuristically through experiments with various block size values. Fig. 3 shows results with block size values increasing from 4x4 pixels to 64x64 pixels. The gray scale presents block differenc-

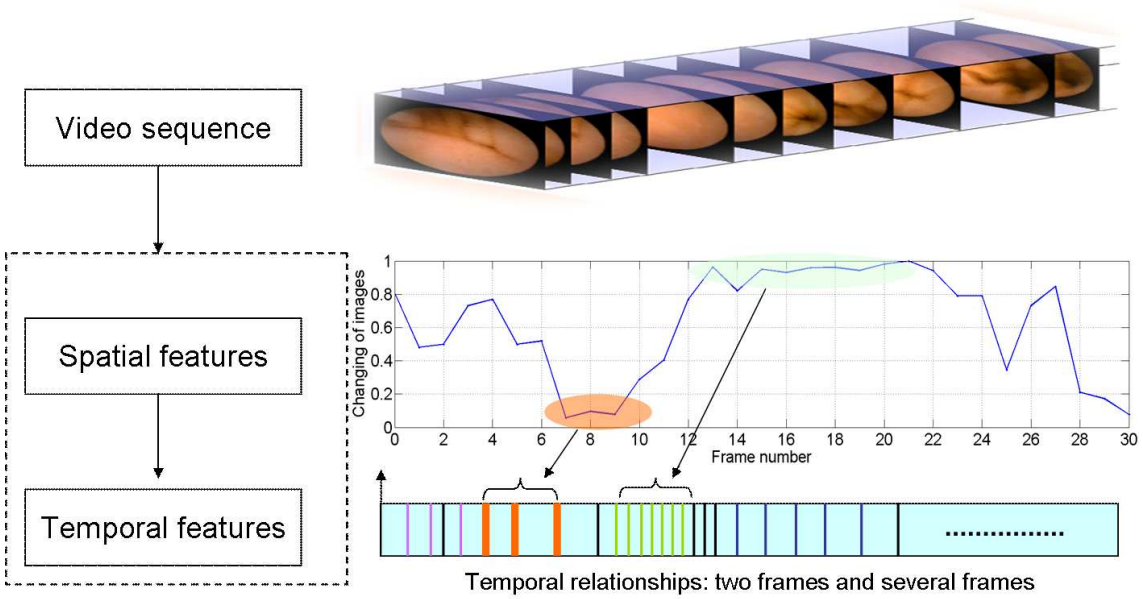


Fig. 2 The proposed framework of VCE analysis

ing with a brighter intensity showing a larger change. With a small block size, image differences show sensitivity to the changes, whereas a too large block size can lose the changes in important regions. For a reasonable selection, the image is divided into $N_{blk} = 64$ blocks with a predetermined 32×32 pixels block size.

3.3 Motion displacement

While the color and edge of the intestinal folds give us interpretations of appearances in a still image, the motion feature, here is displacement between two consecutive frames, carries a lot of information about spatial-temporal relationships between them. Particularly for purpose of the controlling image display application, the motion displacement is an important feature because it is a very strong cue for human vision. In this study, the Kanade-Lucas-Tomasi (KLT) algorithm was utilized to estimate the motion displacement. This method showed reliable results in[2] that emphasized the accuracy and density of measurements for real image sequences. As well it has been reported to be successfully applied to conventional endoscopic images[29),38)]. This algorithm is a feature-tracking procedure developed for video by Tomasi and Kanade[35)]. It is based on earlier work by Lucas and Kanade[21)]. Extensions of the KLT algorithm[25)] include support for a framework of a multi-resolution scheme[1)] and constraints of affine transformation[26)]. Motion is usually represented by set trajectories of the matching points of local features. Fig. 4 shows the motion fields for some frames in a sequence that includes 16 continuous frames (upper panel). The results of frames 1 to 6 and 8 to 14 show that motion estimations are clear and realizable (as shown in Fig. 4(a) and Fig. 4(c). At position (b) (frames 6 and 7) and (d) (frames 14 and 15) the results of the motion fields are a mess (as shown in Fig. 4(b) and Fig. 4(d)).

3.4 Edge of the intestinal folds extraction

Among image features that can be used to describe parts of the GI tube image, edge of the intestinal folds plays an important factor because it can precisely sketch intestinal muscles. To extract edge of the intestinal folds, we first deploy LoG and Canny detectors for the small intestinal images which are extracted from several regions in the small bowel. Based on observations in experiments, the Canny method makes a trade-off between performance and computation time. In our implementations, three threshold values of the Canny detector are predetermined: σ of the smoothing function, lower threshold values T_{low} and upper threshold values T_{high}. A selection of these values influence

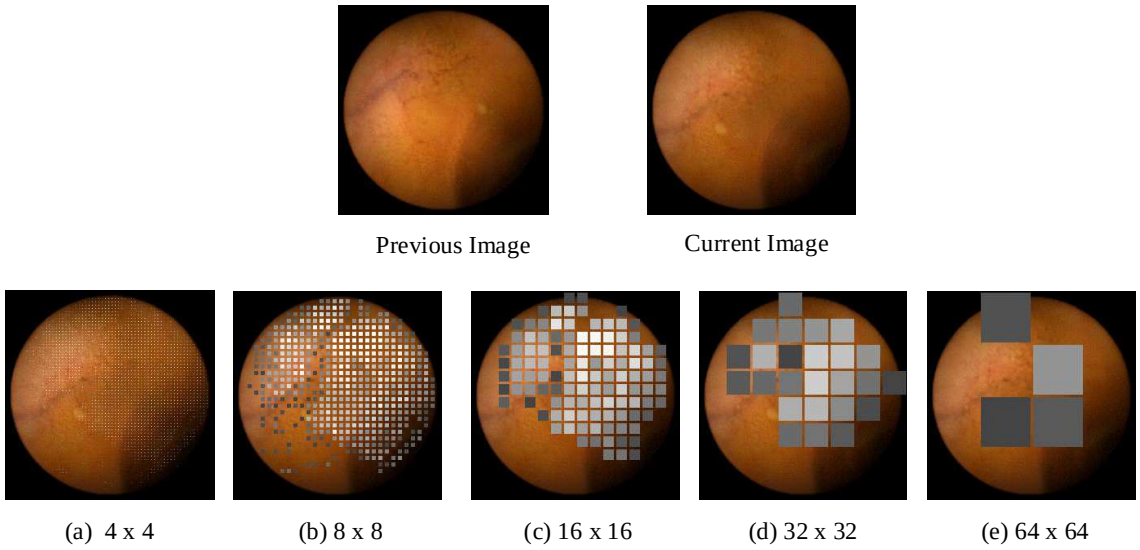


Fig. 3 Color similarity in different block size values

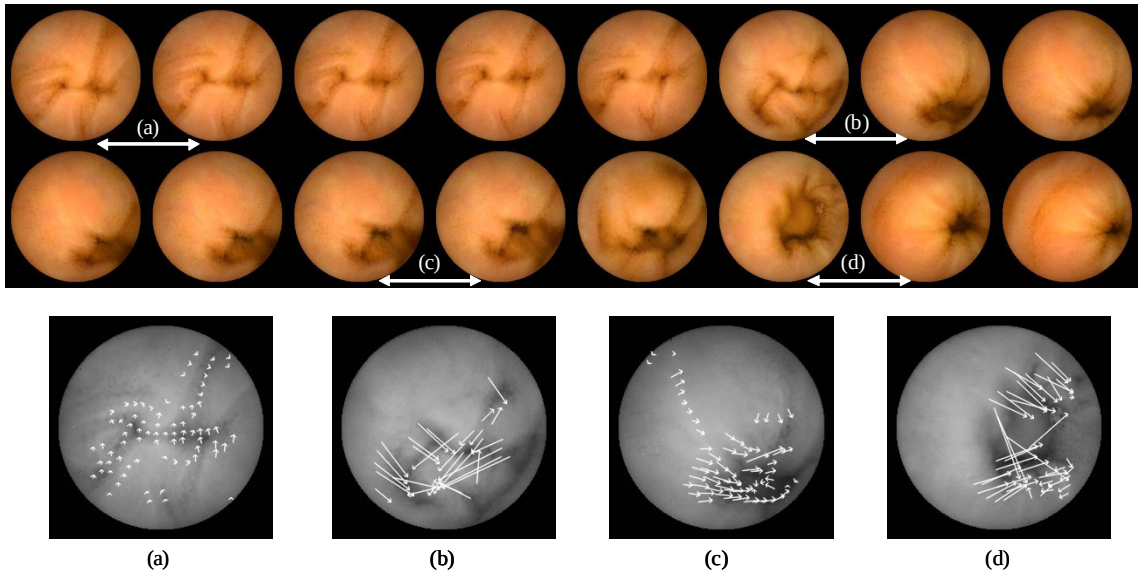


Fig. 4 A continuous image sequence of 16 frames (upper panel). Results of motion estimations at some positions are illustrated (bottom panel). At (a) and (c), the results of the motion fields are reliable, while at (b) and (d) the motion fields are not confident.

to results of the edge detections. In our implementations, a pair of thresholds $T_{low} = 0.3$ and $T_{high} = 0.85$ are fixed, then a range of σ is adjustable. Fig. 5 demonstrates the effect of using the same T_{low} and T_{high} over a range of *sigma* values. With *sigma* small, many noises can appear in the detection results, whereas with *sigma* too large, important edges can be lost. Throughout experiments, a reasonable value $\sigma = 2.5$ is determined because it could obtain acceptable results.

Besides true edge of the intestinal folds, the detection results can consist of edges from other burdens in capsule endoscopic images. To elim-

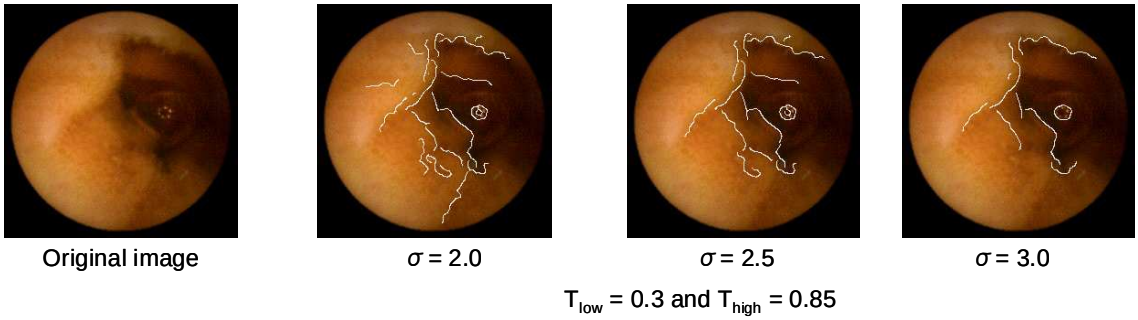


Fig. 5 Edge of intestinal folds detected by Canny edge detector over a range of σ: $\sigma = 2.0$, $\sigma = 2.5$ and $\sigma = 3.0$. The thresholds are fixed at $T_{low} = 0.3$ and $T_{high} = 0.85$.

inate non-edge of the intestinal folds, several techniques are investigated on the edge detections. Finally, we select a safe and simple technique to determine the true edge of the intestinal folds. This method is based on observations that the intestinal folds, particularly in contractile image patterns, usually appear in a concentrated region. Thus, edge pixels are counted in a region where is most of the edges appear. The size of the region 192 x 192 pixels is large enough to ensure that no important edges are lost.

4. Controlling the display of VCE for reducing diagnostic time

In a typical examination, the capsule takes approximately 7 - 8 hours to go through the GI tract for acquisition of images at a rate of two frames per second. The sequence thus has around 57,000 images that can be used for diagnoses. With such a large number of images, review and interpretation of video capsule endoscopy can be time consuming and present a heavy time load for physicians[31].

To reduce diagnostic time, some viewing modes are provided in the RAPID Reader[13], a CE annotation software developed by the capsule manufacturer. For example, with *dual-view*, two consecutive frames are simultaneously displayed; *quad-view* reshapes four consecutive images into one. *Automatic-view* combines successive similar images to display representative frames; *quick-view* mode allows a fast preview by showing only highlight images. Mitigating against reducing diagnostic time by using these techniques is that some clinical images, including abnormalities, may only be seen in a single or just a few frames[31]. These are not easily identifiable in the *quad-view* mode because images are distorted and may not even be seen if that image is skipped. Different from these techniques, we proposed a new method, named as adaptive speed, for automatically controlling the display rate of the CE sequence based on states of image acquisition. Adaptive speed utilizes image features extractions to classify states and to calculate delay time between consecutive frames. The main advantages are that diagnostic time can be reduced while images are displayed in their original form without skipping any frame.

4.1 States of image acquisition conditions: a classification scheme

Studies in the field of gastrointestinal motility reveal an idea for classifications into states of changes between two consecutive frames that correspond to the conditions of image acquisitions. Here, four states of image acquisitions can be defined. For convenience, the four states corresponding to changes in contractions in the small bowel are presented in Fig. 6(a)-(d). *State 1*: Images are captured in a stationary condition. This state appears when the GI motility is in a stable phase. Thus, the position of capsule remains almost still. *State 2*: The capsule device captures images when it moves with just gradual transitions and there is no change in the viewing direction. *State 3*: Images are captured when the capsule undergoes larger movements. The strong contractions that sweep or mix the contents are considered to cause this state. *State 4*: This state occurs when there are brief bursts of contractions or giant migrating contractions. This type of

contraction makes the capsule suddenly change direction and move.

With natural characteristics of GI motility, the states classification task is faced with the problem that a reasonable performance can only be achieved by using of a very large design set for proper training; probably much larger than the number of frames available. Such a difficulty can be overcome based on the above descriptions of the states in which the color similarity is the most discriminating feature for separating global changes (e.g., stationary states (*State 1*) vs. abrupt changes (*State 4*)), while motion displacement is clearly used for discriminating small adjustments (e.g., stationary states (*State 1*) vs. gradually change (*State 2*)). With such discriminations of feature subsets, a "divide and conquer" principle, or a decision tree classifier, is usually applied. For classifying an unknown pattern into a class in successive stages, a decision function at a certain stage can perform rather well by using the discriminating feature[10]. Therefore, a decision tree as shown in Fig. 7 is proposed.

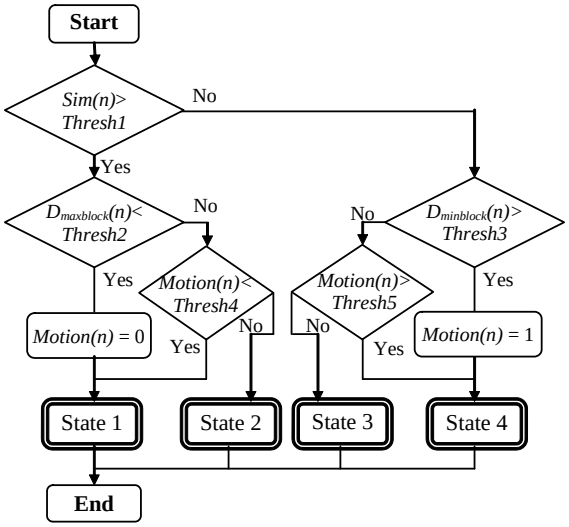


Fig. 7 A decision tree for classifying states.

A combination of threshold values of the decision tree, named as a *parameter set*. The optimal *parameter set* was decided through an empirical study. The idea of this task is that we establish a series of *parameter sets* to enable an exhaustive search among the predetermined candidates to ascertain a reasonable decision tree.

4.2 Calculating delay time

In[14], Glukhovsky et al. introduced a framework for controlling the in vivo camera capture and display rate. After evaluating differences of the multiplicity of frames, they suggested an empirical database or a look-up table so that the display rate is varied accordingly. However, they leave unresolved the method needed to develop this type of database, look-up table, or a specific mathematical function. In our work, if the delay time between two consecutive frames is denoted by D_t, we express the correlating function between D_t and the disparity of images by:

$$D_t = \Theta(f(.), \xi_{skill}, \xi_{system}) \tag{3}$$

where $f(.)$ is a function to estimate perceptual differences between frames by color similarity and motion displacement. The function $f(.)$ can be evaluated by adopting a method that queries the similarity/dissimilarity of images in a CBIR system. Given a query, the overall similarity/dissimilarity between the query and an image in a database is obtained from a combination of individual features $S(f_i)$ as below:

$$f(.) = \sum_i w_i S(f_i) \tag{4}$$

where the coefficients w_i are the weight of the features.

The coefficient ξ_{skill} indicates if a physician is accustomed to viewing such sequences, this is called the skill coefficient. This coefficient is treated differently for each state. The coefficient ξ_{system} is also added to (3) to ensure that the delay time function is adaptive to various display system platforms. A delay time D_t between frames $< n, n+1 >$ can be computed by one of the parametric functions below:

- For *State 1*:

$D_t = A_1(1 - Sim(n)) + A_2 Motion(n) + \xi_{system}$

- For *State 2* and *State 3*:

$$D_t = [B(1 - Sim(n)) + (1 - B)Motion(n)]\xi_{skill} \tag{5}$$

- For *State 4*:

$D_t = D_1(1 - Sim(n)) + D_2 Motion(n) + \xi_{skill}$

where $Sim(n)$ and $Motion(n)$ are calculated color similarity and motion displacement, respectively. The coefficients $< A_1, A_2, B, D_1, D_2 >$ are multiplied by monotone r and the weights of the selected features.

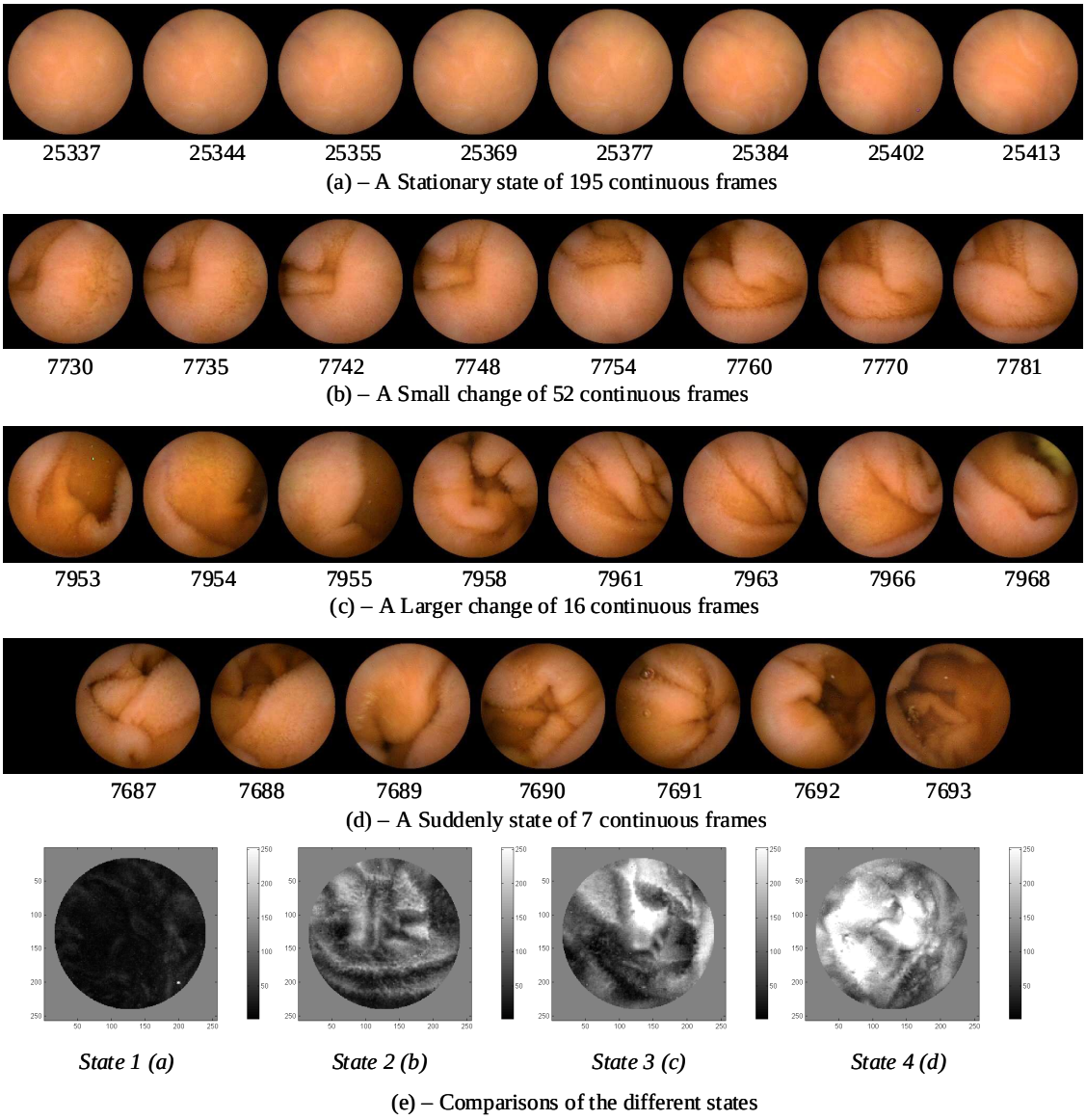


Fig. 6 (a-d) States of image acquisition. (e) A comparative differences between images for corresponding states. Pixel values at (i, j) in each sub-figure is plotted by maximum values from differencing of adjacent images $< t, k >$ shown in (a)-(d). The image differencing is calculated by $g_{i,j}^{t,k} = \sum_{R,G,B} |f_{i,j}^{t} - f_{i,j}^{k}|$. The gray scale bar presents image differencing with a brighter intensity showing a larger change.

In term of variability in delay time values, (5) separately defines the functions for each state, while the classification scheme suggests that a principle of continuity exists between states. Thus, a constraint that ensures "jumping" between states occurs smoothly must exist. This constraint intuitively creates ties between parameters A_1, A_2, B, D_1 and D_2 in (5). Fig. 8 shows distributions of delay time D_t and selected features of a full sequence with smooth changing between states. The delay time values spread in a range from 30 ms/frame to 150 ms/frame, corresponding to the disparity of images varying between stationary and suddenly changing. For comparing image display when the sequence is played at a fixed frame rate (i.e.,

13 fps or a delay time with a constant value of 77 ms), the proposed method allows physicians to flexibility review the VCE sequence.

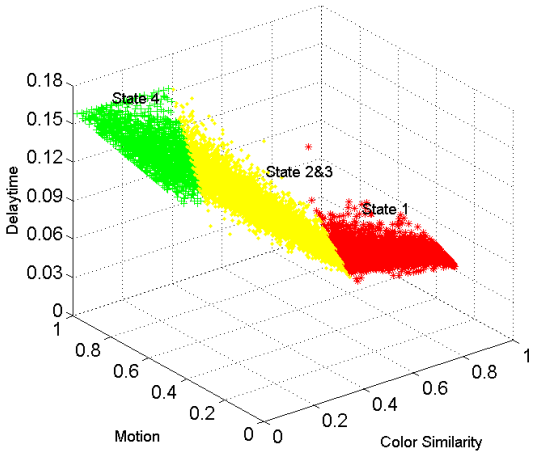


Fig. 8 Distribution of the delay time calculated from the motion displacement and similarity features of a sequence.

4.3 Experimental results

The proposed method is compared with RAPID Version 4 (the *G system*, downloadable at[12]). To ensure that the conditions for the evaluation of both systems were as similar as possible, a GUI application (called *P system*) was developed for the proposed method so that normal diagnostic functions such as the capture of abnormal regions, the manual adjustment of viewing speeds and changes in viewing display, as well as functions for navigating and verifying suspicious regions were available. Both systems were installed on a same PC with a Pentium IV 3.2 GHz, and 2 GB RAM.

We prepared six full sequences of patient data. The evaluations were implemented on both systems by the same four physicians from the Graduate School of Medicine, Osaka City University. Thus, forty-eight evaluations were conducted. To facilitate unbiased evaluations, the order of the evaluations of a certain sequence were established so that the number of anterior/first evaluations on each system was equal. The physicians were asked to independently find and capture suspicious regions.

The main activities of the physicians as they used the two systems were recorded. These included: [*play* → *stop*], *browsing/scanning* frames to examine suspicious regions, *jumping* frames, *changing manually display speed* and *capturing* abnormal regions. Fig. 9(a) shows an example of the logged activities of physician *MD. A* for Seq. #3 under the two systems. From logs expressed in this figure, the logged action based analysis is described below to compare the performances of two systems through three criteria; diagnostic time, abnormal regions captured, and system operability.

Average diagnostic time by sequence is shown in Fig. 10. From this figure, the diagnostic time on the *P system* was seen as reduced for all six sequences. The average diagnostic time for the *P system* was 32.5±7 minutes and it was 42.4 ± 9 minutes for the *G system*. The diagnostic times using the proposed system were significantly reduced for most evaluations (approximately 16 min. for *MD. A*, 6 min. for *MD. B*, and 14 min. for *MD. C*). The diagnostic time of *MD. D* was equal in both systems.

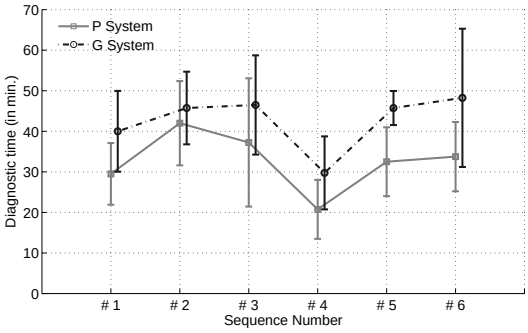


Fig. 10 Average diagnostic time by sequences

5. Detection of the intestinal contractions from VCE

5.1 A schematic view of intestinal contractions from VCE

The characteristics of intestinal motility in the human small bowel has been the subject of decades of exhaustive research by physiologists (e.g.,[5),16),17),19),34)]), because relevant information in terms of the number, frequency, and distribution of intestinal contractions can indicate the presence of different malfunctions. For example, weak and disorganized contractions are associated with bacterial overgrowth, intestinal

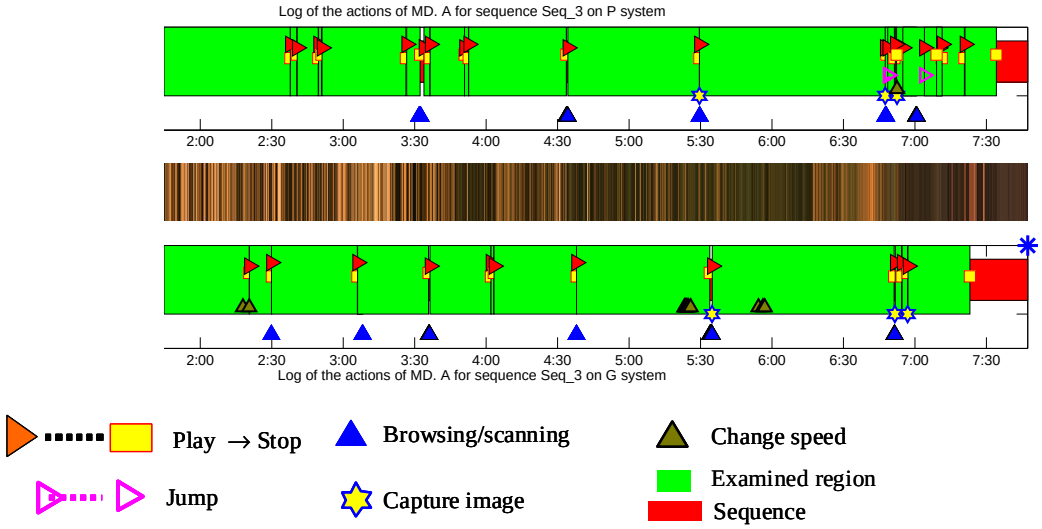


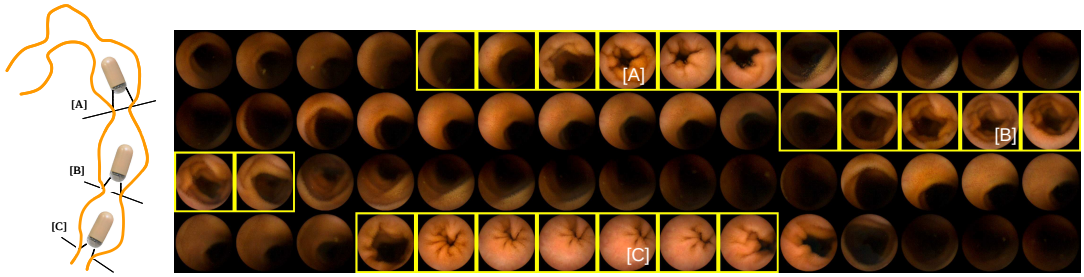
Fig. 9 Logged actions of *MD. A* for Seq. #3. The upper panel shows activities under the *P system*, the lower panel shows activities under *G system*. Same abnormal regions captured on both system are indicated by boxes

obstruction or paralytic ileus[16], while dysfunctions in, or absence of, contractions over a long period can present as functional dyspepsia[17].

VCE technique has recently been introduced as a non-invasive means of inspection in GI tract. Although this technique was not originally designed for the assessment of intestinal motility, VCE image sequences reflect intestinal activity during the transit time of the capsule. They present a useful source information of visualizing intestinal contractions. As observed in a VCE image sequence, a cycle of contraction begins at the widest state of the intestinal lumen, then proceeds to occlude the lumen, before reaching the most extreme state of shrinkage with extensive intestinal wrinkling. The intestinal folds then relax again at the end of the contraction. A schematic view of contractions along the small bowel is illustrated in Fig. 11(a) and some patterns of contraction cycles in a sequence including 60 continuous frames are shown in Fig. 11(b). As shown in these figures, variations in the images features throughout the sequence of consecutive frames suggest a possible contraction cycle. The frames showing images at the most extreme stage of a contraction cycle are the easiest to recognize, because they are associated with shrinkage of the circular muscle layers. Moreover, in terms of the physiology, the duration of intestinal contractions varies along the small bowel; a well-recognized pattern of intestinal contractions, as described above, is that the maximal rate of contractions decreases in a series of steps from proximal to distal regions. These characteristics obviously provide useful information for detecting contractions in VCE image sequences.

5.2 A three-step procedure for the intestinal contraction detection

We consider the contractions as dynamic events in VCE videos and are recognized using both spatial and temporal information. The temporal features provide the potential to detect contractions through changes in the edges pixels of the intestinal folds (edge signal), and by evaluating the degree of similarity between successive frames. In the context of signal processing, the positions of possible contractions can be located within windows including consecutive frames by convoluting the edge signal with kernel functions. Relevant configurations of these kernels are established such that varying the size of the windows reflects the contraction frequency gradient, which gradually reduces in a series of steps along the small bowel. Based on results detecting possible contrac-



(a) (b)

Fig. 11 (a) A schematic view of contractions along small bowel. (b) Some patterns of contraction cycles in a sequence including 60 continuous frames (left to right, top to bottom). The [A] - [C] are frames at most extreme stage of the contraction cycle.

tions, the spatial features are analyzed in order to identify true contractions through a classifier method. With this approach, the results of temporal analysis can be tuned to prune as many non-contraction frames as possible, and thus the detection of contractions throughout the whole small bowel is improved. Therefore, we formulate the intestinal contractions based on features below.

Stage 1 - Edge extractions to detect possible contractions

The $f(x)$ (with x is frame number) is denoted as a function of edge pixels number of intestinal folds:

$$f(x) = \sum_i \delta \text{ with } \begin{cases} \delta = 1 \text{ if i is an edge pixel} \\ \delta = 0 \text{ otherwise} \end{cases} \quad (6)$$

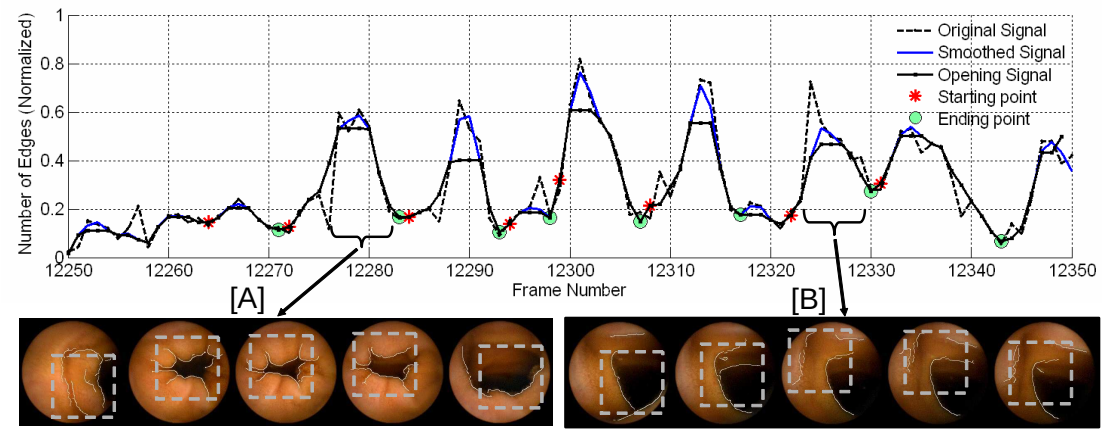
The technique that we used to extract edges of intestinal folds. The signal $f(x)$ is normalized in the range of $[0, 1]$ and smoothed by a Gaussian function to remove noise. The possible contractions are located where $f(x)$ is in the form of a triangle. These positions can be detected by locating local peaks. However, not all the signals present a perfect triangular pattern; this depends on the length and the strength of the contractions. Thus, a mathematical morphology method is applied to create a simpler graph than the original signal. As a result of this process, potential contractions are located within windows of length w (in number of frames). Fig. 12(a) shows an example of the processing of a sequence including 100 consecutive frames.

Stage 2 - Evaluation of the similarities between consecutive frames for eliminating non-contractions

The results in Fig. 12(a) show two instances of local maximal peaks, [A] representing a true contraction (positive case), whereas [B] is a false result (negative case). If a candidate contraction includes large regions of high similarity, it is not a true contraction. Thus, to eliminate the negative cases, evaluation of the similarities between frames was implemented. These evaluations were based on results using an unsupervised clustering method adopted from[15]. Because of the observation that each homogeneous region in consecutive frames was represented by a Gaussian distribution, the set of regions was represented by a Gaussian Mixture Model (GMM). The similarities were extracted and clustered. Assuming that a possible contraction includes $w = N$ frames, feature vectors were constructed as in the configuration in Fig. 13(a). Frames were first divided into N_{blocks} with a $X \times Y$ pixels size and an intensity histogram H including N_{bins} for each block was calculated. The results of the clustering process are then assessed to discard redundant cases. If the largest clusters include the high similarity values, it implies that almost all frames of a candidate contraction are similar, suggesting a low probability of it being a positive contraction. Figure 13(b) and (c) shows the results of this process to eliminate negative cases of contractions [B] in Fig. 12(a).

Stage 3 - Detect true contractions through spatial features

The orientation of the edges of intestinal wrinkles during the strongest stage of contraction was a powerful feature for discriminating



(a) Possible contractions

Fig. 12 Possible contractions are marked on an original signal with starting (asterisks) and ending (circles) frames. [A] is a positive contraction; [B] is a non-contraction.

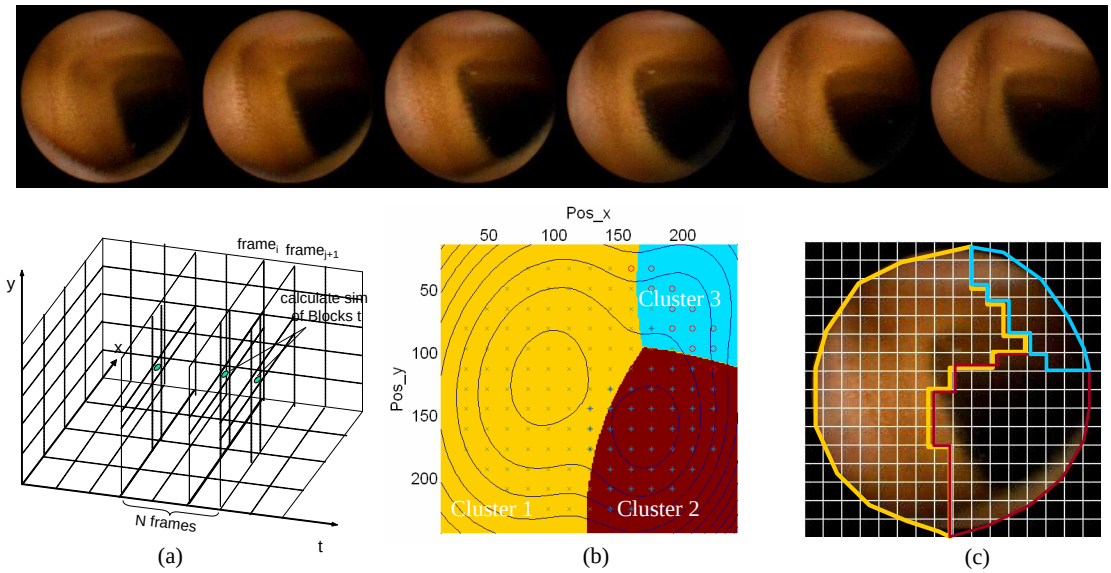


Fig. 13 Configuration to get the feature vectors. (b) Results of clustering similarity data (with K=3) of the negative case (marked as [B]) in Fig. 12(a), the ratio of cluster 1 is 61% and of cluster 2 is 30%. (c) Overlap the results on middle frames. Mean of similarity data are 0.6 for cluster 1 and 0.53 for cluster 2

between contractions and non-contractions. In order to characterize these patterns, an edge direction histogram was used. As shown in[39], this model is robust enough to compare structural similarities between images. The frame with the maximum edge number, after Stage 2, was selected as representing the strongest stage of contraction.

Figure 14(a) and Fig. 14(b) show two examples of the polar histograms H of contraction and non-contraction cases, respectively. The patterns of the polar histogram imply that, in contraction cases, the directions are spread in every direction, whereas in non-contraction cases, the polar histogram is distributed in only the dominant direction. To reduce the dimensions of the histogram, and avoid losing important information, symmetric directions are combined. The directions $(0 - 180^o)$ are then divided into 16 bins of a histogram D in a Cartesian system, as shown in the left panel in Fig. 14(a) and Fig. 14(b). Based on the signal of the histogram D, a simple K-nearest-neighbours (K-NN) classifier was used to decide on the contraction pattern. In this learning model, the structural similarity between two feature vectors D_x and D_y was estimated by calculating the correlation coefficient $corre(x, y)$ according to:

$$corre(x, y) = \frac{\delta_{xy} + C}{\delta_x \delta_y + C} \quad (7)$$

where δ_x and δ_y are the standard deviations of the feature vectors D_x and D_y, respectively; δ_{xy} is the covariance of vectors and C is a small constant, to prevent the denominator from being zero. In our implementations, K-NN classifier trained with a data set including 1000 frames, which had been labeled manually, as non-contraction or contraction cases. The data set was established so that the number of contraction cases was equal to the number of non-contraction cases (500 cases each).

5.3 Experimental results

Six VCE image sequences were obtained from different positions of the small bowel. The length of each sequence was 10 minutes. Ground truth data for each sequence were obtained by manual examination by the endoscopist experts. The positions of the beginning, end, and the strongest stage of each contraction cycle were marked.

To evaluate the performance of the method, the data below were calculated after each stage:

- The number of true contractions detected (True positives - TP)
- The number of wrong contractions detected (False Positives - FP)
- The number of lost contractions (False Negatives - FN)

The performances were then evaluated using two criteria:

$$Sensitivity(Sens) = \frac{TP}{TP + FN} \text{ and}$$
$$FalseAlarmRate(FAR) = \frac{FP}{TP + FP} \quad (8)$$

The effectiveness of adapting window sizes along the small bowel in the Stage 1 was evaluated by comparing this technique with the results using a fixed window size. The window sizes were assigned based on the capsule transit through the small bowel. The FAR values obtained with adaptive changes in window size were better than those with a fixed window size, for all sequences. In particular, the yield was significantly reduced FAR for Seq_2, Seq_4 and Seq_6. However, with larger window sizes than those used for the fixed size, the detection of true contractions was reduced for Seq_4 and Seq_6. Contractions were lost in these cases because concussive contractions appeared only over a short time period (i.e. a few frames). This was known as a clustering contraction (in terms of physiology). Because at least one of the contractions in the cluster was still observable, they did not make a great impact.

Using the results of various window sizes for Stage 2 and Stage 3, the overall performance of the proposed method is shown in Table 2. The data in column 2 show the results of the proposed method after Stage 3. Among the contractions detected in ground truth data, a number of contractions which do not appear in the final results are shown in the next column. To compare qualitative indices between previous studies and the proposed method, although contractions in[11] and[27] were determined using different experimental data, average values for $Sens$ and FAR were used: 71.5% and 71% in[27]; 73.5% and 44% in[11], respectively . Note that the values of the $Sens$ and FAR for methods[11] and[27] were recomputed from data sup-

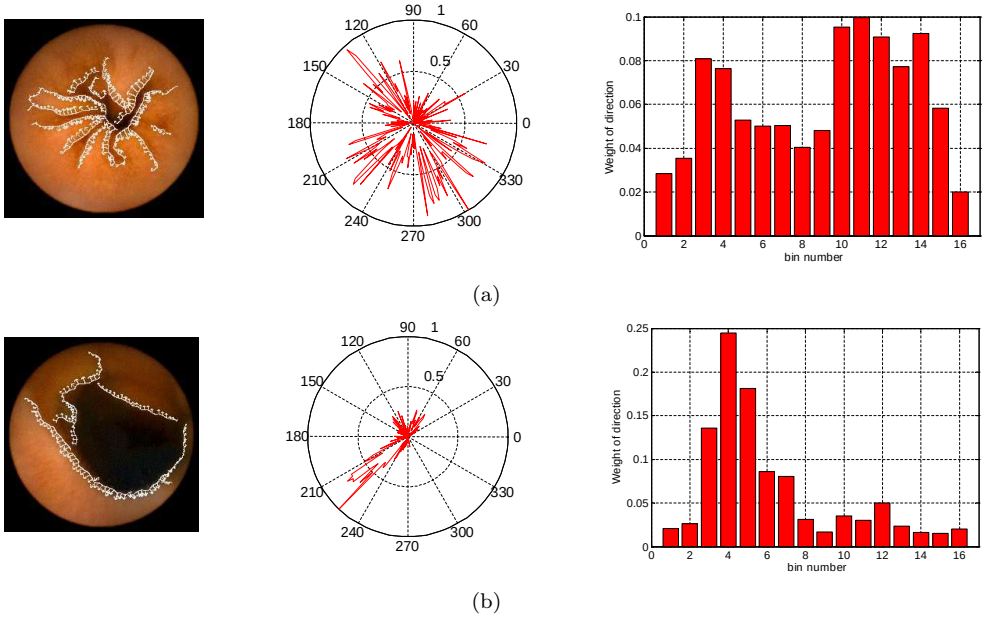


Fig. 14 Direction histogram of a contraction (a) and a non-contraction (b). Left side shows original frames with gradient direction at edge pixels, middle is a polar histogram and the right side is a Cartesian histogram in 16 bins

Seq.	Ground-truth data	Fixed window size				Various window sizes				window size w
		Number of Contraction Found	Number of Contraction Lost	*Sens*	*FAR*	Number of Contraction Found	Number of Contraction Lost	*Sens*	*FAR*	
Seq_1	20	88	1	95%	78%	88	1	95%	78%	5
Seq_2	30	76	1	97%	62%	62	2	93%	55%	7
Seq_3	16	92	1	94%	84%	36	3	81%	64%	11
Seq_4	48	124	1	98%	62%	57	6	88%	26%	9
Seq_5	46	109	4	91%	61%	66	5	89%	38%	7
Seq_6	33	78	1	97%	59%	44	4	88%	34%	11
	Average			94%	68%			89%	49%	

Table 1 Results of the stage one with fixed and variable window sizes

ported in these studies, to match with the definitions in (8).

6. Conclusions

The VCE analysis techniques were intensively studied and investigated to develop two applications for diagnostic assistance. Spatial and temporal image features are main suggestions for VCE interpretations. This led us to the framework including two levels: the image feature extractions and their temporal analysis. The techniques extracted the spatial image features such as color similarity, edge of the intestinal folds in still images as well as the spatial-temporal feature by extracting motion displacement of two consecutive frames.

Facing with the main challenge of reading VCE sequence that is time consuming and under careful examinations of examining doctors, we proposed the adaptive speed technique. This technique ensured that entire the video sequences are displayed in the original shape without skipping any frames; thereby enabling the inspection of all data. In experiments, the results confirm that the diagnostic time is reduced to around 32.5 ± 7 minutes per full sequence. Compared with a standard-view using the existing system, Rapid Reader Version 4,

Table 2 Recognition rates of the experimental sequences using the proposed method

Seq.#	Ground-truth data	Number of contractions detected	Number of true contractions lost	*Sens*	*FAR*
Seq_1	20	48	2	90%	63%
Seq_2	30	43	4	87%	40%
Seq_3	16	25	3	81%	48%
Seq_4	48	45	8	83%	11%
Seq_5	46	43	12	74%	21%
Seq_6	33	41	9	73%	41%
Average				81%	37%

the proposed method is 10 minutes less while the number of abnormalities found are similar under both systems. As well, the proposed system requires less effort because of its efficient operability. These results should convince physicians that the proposed technique can be safely used for routine clinical diagnoses.

Regarding automatic detection of the intestinal contractions in VCE sequence, the contractions were labeled successfully using a three-stage procedure. The proposed method detected 81% of the total contractions and the false alarm rate was 37%. The experimental results confirmed that, by taking account of the frequency of the contractions, it was possible to solve the imbalance problem of appearances of the contractions in VCE sequence. The combination of spatial and temporal features provided a workable, robust method for detecting contractions, which was both quantitatively and qualitatively superior to previous methods.

References

1) Anandan, P.: A Computation Framework and an Algorithm for the Measurement of Visual Motion, *International Journal of Computer Vision*, Vol.2, No.3, pp.283–310 (1989).
2) Barron, J., Fleet, D. and Beauchemin, S.: Performance of Optical Flow Techniques, *International Journal of Computer Vision*, Vol.12, No.1, pp.43–77 (1994).
3) Bonnel, J., Khademi, A., Krishnan, S. and Ioana, C.: Small bowel image classification using cross-co-occurrence matrices on wavelet domain, *Biomedical Signal Processing and Control* (2008). 10.1016/j.bspc.2008.07.002, in Press.
4) Boulougoura, M., Wadge, E., Kodogiannis, V. and Chowdrey, H.: Intelligent systems for computer-assisted clinical endoscopic image analysis, *in Proc. of the Int. Conf. on Biomedical Engineering*, pp.405–408 (2004).
5) Bronzino, J.D.: *The Biomedical Engineering Handbook, 3rd Edition*, CRC Press (2006).
6) C.Hu, M.Q.Meng, P.X.Liu and X.Wang: Image Distortion Correction for Wireless Capsule Endoscope, *in Proc. of the Int. Conf. on ICRA*, pp.4718–4723 (2004).
7) C.Hu, M.Q.Meng, P.X.Liu and X.Wang: Wireless Capsule Endoscopy Images enhancement using contrast driven forward and backward anisotropic diffusion, *in Proc. of the Int. Conf. on Image Processing*, pp. II 437–440 (2007).
8) Coimbra, M., Campos, P. and Cunha, J. P.S.: Topographic segmentation and trasit time estimation for endoscopic capsule exam, *in Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vol.II, pp.1164–7 (2006).
9) Cunha, J. P.S., Coimbra, M., Campos, P. and Soares, J.M.: Automated Topographic Segmentation and Transit Time Estimation in Endoscopic Capsule Exams, *IEEE Transaction on Medical Imaging*, Vol.27, pp.19–27 (2008).
10) de Sá, J.M.: *Pattern Recognition - Concepts, Method and Applications*, Springer (2001).
11) F.Vilarino, Spyridonos, P., Vitria, J., Azpiroz, F. and Radeva, P.: Linear Radial Patterns Characterization for Automatic Detection of Tonic Intestinal Contractions, *in Proc. of the CIARP*, LNCS, Vol.4225, pp.178–187 (2006).
12) Given Imaging Ltd.: RAPID Software Product, *http://www.givenimaging.com /en-us/HealthCareProfessionals /Products/Pages/Software.aspx* (2007).
13) Given Imaging Ltd.: Overview of Capsule Endoscopy, *http://www.givenimaging.com/en-us/Patients/Pages/pagePatient.aspx* (2008).
14) Glukhovsky, A., Meron, G., Adler, D. and Zinati, O.: System for controlling in vivo camera capture and display rate, *Patent number PCT WO 01/87377 A2* (2002).
15) Greenspan, H., Goldberger, J. and Mayer, A.: A Probabilistic Framework for Spatio-Temporal Video Representation, *in Proc. of the Int. Conf. on ECCV 2002*, pp.461–475 (2002).
16) Grundy, D.: *GastroIntestinal Motility - The Integration of Physiological Mechanisms*, MTP Press Limited, Hingham, MA (1985).

17) Hansen, M.B.: Small Intestinal Manometry, *Physiological Research*, Vol. 51, pp. 541–556 (2002).
18) Iddan, G., Meron, G., Glukovsky, A. and Swain, P.: Wireless Capsule Endoscope, *Nature*, Vol.405, p.417 (2000).
19) Imam, H., Sanmiguel, C., Larive, B., Yasser, B. and Soffer, E.: Study of intestinal flow by combined videofluoroscopy manometry, and multiple intraluminal impedance, *AJP- Gastrointestinal and Liver Physiology*, Vol.286, pp. 263–270 (2004).
20) Lee, J., Oh, J., Yuan, X. and Tang, S.: Automatic Classification of Digestive Organs in Wireless Capsule Endoscopy Videos, *in Proc. of the ACM Symposium on Applied Computing*, ACM SAC 2007 (2007).
21) Lucas, B. and T.Kanade: An Iterative Image Registration Technique with an Application to Stereo Vision, *in Proc. of the Int. Joint Conf. on Artificial Intelligence*, pp.674–679 (1981).
22) Mackiewicz, M., Berens, J. and Fisher, M.: Wireless Capsule Endoscopy video segmentation using Support Vector Classifiers and Hidden Markov Models, *in Proc. of the Int. Conf. on Medical Image Understanding and Analyses* (2006).
23) Mackiewicz, M., Berens, J., Fisher, M. and Bell, G.: Colour and Texture based GastroIntestinal tissue discrimination, *in Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vol.2, pp.597 – 600 (2006).
24) M.Coimbra and Cunha, J. P.S.: MPEG-7 visual descriptors - Contributions for automated feature extraction in Capsule Endoscopy, *IEEE Trans. Circuits and Systems for Video Technology*, Vol.16, No.5, pp.628–637 (2006).
25) S. Birchfield: KLT: Kanade-Lucas-Tomasi Feature Tracker, *http://www.ces.clemson.edu/stb/klt/* (2006).
26) Shi, J. and Tomasi, C.: Good Features to Track, *in Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp.593–600 (1994).
27) Spyridonos, P., F.Vilarino, Vitria, J., Azpiroz, F. and Radeva, P.: Identification of Intestinal Motility Events of Capsule Endoscopy Video Analysis, *in Proc. of the Int. Conf. on Advanced Concepts for Intelligent Vision Systems*, LNCS, Vol.3708, pp.531–537 (2005).
28) Spyridonos, P., F.Vilarino, Vitria, J., Azpiroz, F. and Radeva, P.: Anisotropic Feature Extraction from Endoluminal Images for Detection of Intestinal Contractions, *in Proc. of the Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, LNCS, Vol.4191, pp.161–168 (2006).
29) Suchit, P., Sagawa, R., Echigo, T. and Yagi, Y.: Deformable Registration for Generating Dissection Image of an Intestine from Annular Image Sequence, *in Proc. of the Int. Conf. on Computer Vision for Biomedical Image Applications*, LNCS, Vol.3765, pp.271–280 (2005).
30) Swain, M. and Ballard, D.: Color Indexing, *International Journal of Computer Vision*, Vol. 7, No.1, pp.11–32 (1991).
31) Swain, P. and Fritscher-Ravens, A.: Role of video endoscopy in managing small bowel disease, *GUT*, Vol.53, pp.1866–1875 (2004).
32) Szczypinski, P.M.: Selecting a motion estimation method for a model of deformable rings, *in Proc. of the Int. Conf. on Signals and Electronic Systems*, pp.297–300 (2006).
33) Szczypinski, P. M., Sriram, P. V. J., Sriram, R. D. and Reddy, D. N.: Model of Deformable Rings for Aiding the Wireless Capsule Endoscopy Video Interpretation and Reporting, *in Proc. of the Int. Conf. on CVG 2004*, pp. 167–172 (2006).
34) Thomas, E. A., Sjovall, H. and Bornstein, J.: Computational model of the migrating motor complex of the small intestine, *AJP- Gastrointestinal and Liver Physiology*, Vol.286, pp.564–572 (2004).
35) Tomasi, C. and Kanade, T.: Detection and Tracking of Point Features, Technical report (1991).
36) Vilarino, F., Kuncheva, L. and Radeva, P.: ROC curves and video analysis optimization in intestinal capsule endoscopy, *Pattern Recognition Letter*, Vol.27, No.8, pp.875–881 (2006).
37) Vu, H., Echigo, T., Sagawa, R., , Yagi, K., Shiba, M., Higuchi, K., Arakawa, T. and Yagi, Y.: Adaptive Control of Video Display for Diagnostic Assistance by Analysis of Capsule Endoscopic Images, *in Proc. of the Int. Conf. on Pattern Recognition*, pp.980–983 (2006).
38) Wu, C. H., Chen, Y. C., Liu, C. Y., Chang, C.C. and Sun, Y.N.: Automatic extraction and visualization of human inner structures from endoscopic image sequences, *in Proc. of the IS&T/SPIE*, Vol.5369, pp.464–473 (2004).
39) Zhou, W., Bovik, A., Sheikh, H. and Simoncelli, E.: Image Quality Assessment: From Error Measurement to Structural Similarity, *IEEE Transactions on Image Processing*, Vol.13, pp.600–613 (2004).