



## Detection of contractions in adaptive transit time of the small bowel from wireless capsule endoscopy videos

Hai Vu<sup>a,\*</sup>, Tomio Echigo<sup>b</sup>, Ryusuke Sagawa<sup>a</sup>, Keiko Yagi<sup>c</sup>, Masatsugu Shiba<sup>d</sup>, Kazuhide Higuchi<sup>d</sup>, Tetsuo Arakawa<sup>d</sup>, Yasushi Yagi<sup>a</sup>

<sup>a</sup>Department of Intelligent Media, The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Osaka 567-0047, Japan

<sup>b</sup>Department of Engineering Informatics, Osaka Electro-Communication University, 18-8 Hatsu-cho, Osaka 572-8530, Japan

<sup>c</sup>Kobe Pharmaceutical University, 4-19-1 Motoyamakita-machi, Hyogo 658-8558, Japan

<sup>d</sup>Graduate School of Medicine, Osaka City University, 1-4-3 Asahimachi, Osaka 545-8585, Japan

### ARTICLE INFO

#### Article history:

Received 17 June 2007

Accepted 22 October 2008

#### Keywords:

Gastrointestinal motility

Intestinal contraction detection

Wireless capsule endoscopic images

### ABSTRACT

Recognizing intestinal contractions from wireless capsule endoscopy (WCE) image sequences provides a non-invasive method of measurement, and suggests a solution to the problems of traditional techniques for assessing intestinal motility. Based on the characteristics of contractile patterns and information on their frequencies, the contractions can be investigated using essential image features extracted from WCE videos. In this study, we proposed a coherent three-stage procedure using temporal and spatial features. The possible contractions are recognized by changes in the edge structure of the intestinal folds in Stage 1 and evaluating similarity features in consecutive frames in Stage 2. In order to take account of the properties of contraction frequency, we consider that the possible contractions are located within windows including consecutive frames. The size of these contraction windows is adjusted according to the passage of the WCE. These procedures aim to exclude as many non-contractions as possible. True contractions are determined through spatial analysis of directional information in Stage 3. Using the proposed method, 81% of true contractions are detected with a 37% false alarm rate for evaluations in the experiments. The overall performance of this method is better than that of previous methods, in terms of both the quality and quantity indices. The results suggest feasible data for further clinical applications.

© 2008 Elsevier Ltd. All rights reserved.

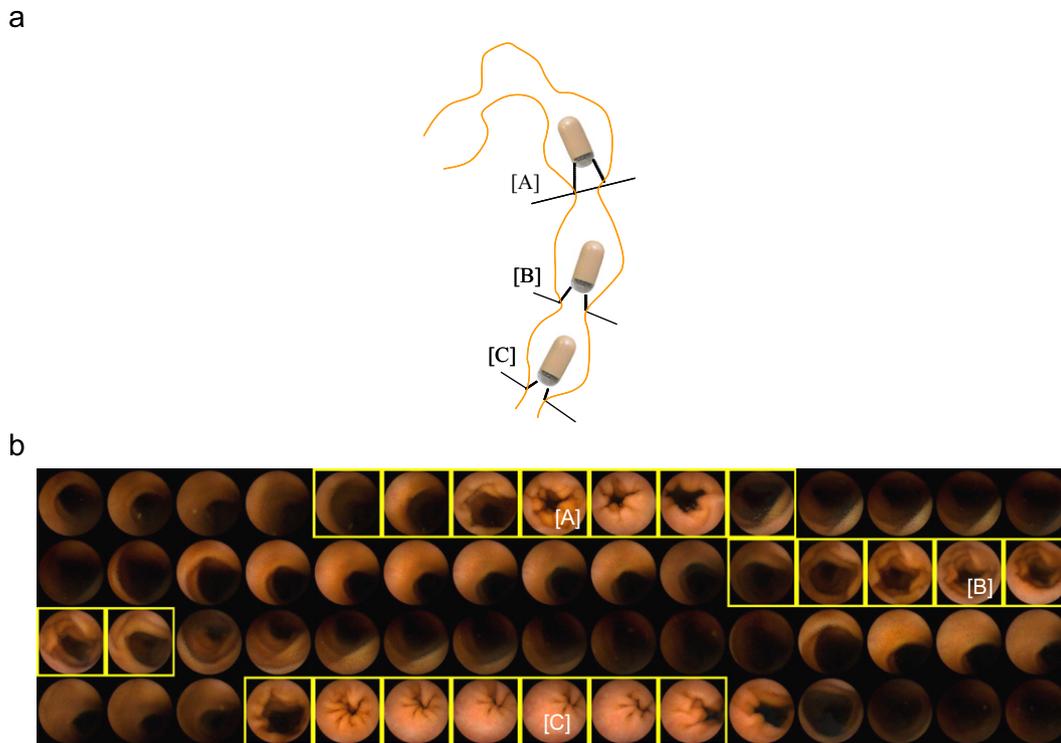
### 1. Introduction

The characteristics of intestinal motility in the human small bowel have been the subject of decades of exhaustive research by physiologists (e.g. [1–5]), because relevant information in terms of the number, frequency, and distribution of intestinal contractions can indicate the presence of different malfunctions. Weak and disorganized contractions are associated with bacterial overgrowth, intestinal obstruction or paralytic ileus [1], while dysfunctions in, or absence of, contractions over a long period can present as functional dyspepsia [2]. Manometry is currently the favored technique for measuring intestinal motility [2]. This involves the introduction of a probe into the patient's gastrointestinal (GI) tract, and is therefore highly invasive, and can cause significant patient discomfort, due to the long distances to be examined and the loop configuration of the small bowel.

Wireless capsule endoscopy (WCE) [6] has recently been introduced as a non-invasive means of inspection in GI tract. This technique is especially successful in finding bleeding regions, Crohn's disease, and suspected tumors of the small bowel [7,8]. The WCE utilizes a swallowable endoscopic device (11 × 26 mm, 3.7 g, in [9]) that is ingested and propelled by natural peristalsis through the GI tract. In a typical examination, WCE takes approximately 7–8 h to pass through the GI tract, capturing images at a rate of 2 fps. The sequence thus generates about 57,000 images that can then be analyzed. Although this technique was not originally designed for the assessment of intestinal motility, WCE image sequences reflect intestinal activity during the transit time of the capsule. They present a useful source information of visualizing intestinal contractions. However, manually analyzing the intestinal events from the WCE videos is a tedious and time-consuming task for physicians. Thus, the main purpose of this study was to use computer vision techniques to automatically detect and quantify intestinal contractions. Because of its non-invasive nature and the minimal attention needed by physicians, it is a promising method which overcomes the drawbacks of traditional techniques.

\* Corresponding author. Tel.: +81 66879 8422; fax: +81 66877 4375.

E-mail address: [vhai@am.sanken.osaka-u.ac.jp](mailto:vhai@am.sanken.osaka-u.ac.jp) (H. Vu).



**Fig. 1.** (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.

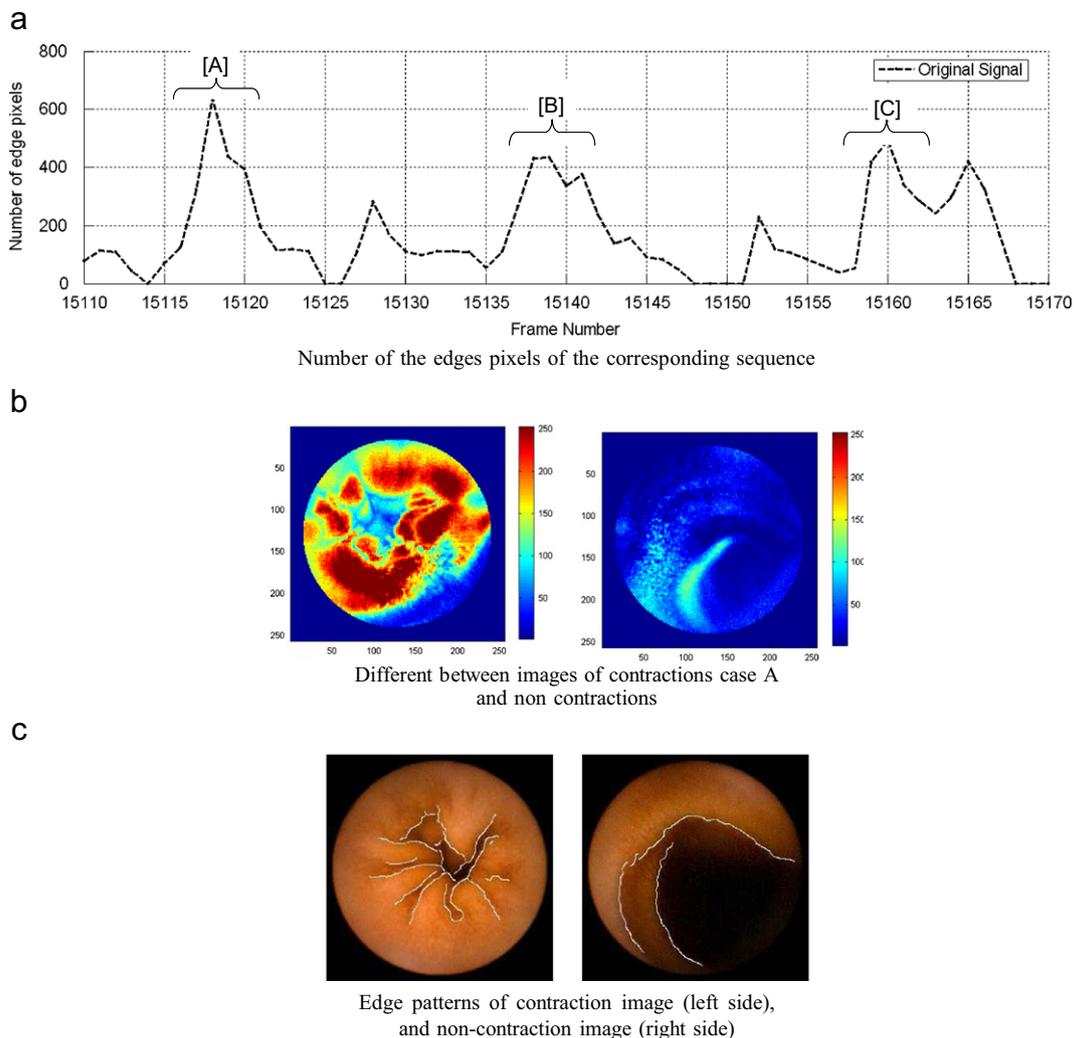
As observed in a WCE image sequence, a cycle of contraction begins at the widest state of the intestinal lumen, then proceeds to occlude the lumen, before reaching the most extreme state of shrinkage with extensive intestinal wrinkling. The intestinal folds then relax again at the end of the contraction. A schematic view of contractions along the small bowel is illustrated in Fig. 1a and some patterns of contraction cycles in a sequence including 60 continuous frames are shown in Fig. 1b. As shown in these figures, variations in the image 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 stated by [1,5], 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 WCE image sequences.

Recognizing intestinal contractions in WCE videos is still at an early stage of development. In order to describe the intestinal structure at the most extreme stage of a contraction cycle, previous studies focused on extracting the skeleton of intestinal folds (wrinkles) and/or the intestinal lumen regions. These features were then used to describe directional information which was encoded into symmetrical patterns, such as linear radial patterns in [10], or the star-wise patterns in [11]. On the other hand, considering temporal patterns of contractions in an image sequence, [12] claimed an imbalance problem between the numbers of contractile and non-contractile frames. The prevalence of contractile frames is very small, i.e. between 1:50 and 1:100. To address this problem, in [10] a contraction cycle is defined if there are at least five wrinkle frames in a sequence of 10 consecutive images. Spyridonos et al. [13] identified possible contractions by sharp variations in the gray-level intensity within a fixed sequence of nine frames. In this preliminary study, the

natural characteristics of intestinal contractions, including their variable length, could be recognized [14]. Although the properties of contractions frequency as described above imply a various contraction length along small bowel, no attempt has been made in any of these studies to investigate them in term of the temporal patterns. Using these properties to aid the recognition of contractions thus should help to improve detection rate.

In this scenario, we have extended our earlier work in [14] to take account of the frequency gradient of intestinal contractions throughout the passage of the WCE. Contraction detection is therefore refined in terms of GI physiology. The contractions are considered as dynamic events in WCE 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 contractions, 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. Our results showed an overall detection rate of 81% for true contractions (sensitivity) and a 37% false alarm rate (FAR). These recognition rates are better than those produced by [10,13], and show a significant reduction in FAR compared with [14]. These results suggest that this technique is reliable for further clinical applications.

The rest of this paper is organized as follows: Section 2 describes the methodologies of the proposed method. Section 3 explains



**Fig. 2.** Formulating problems for contraction detections from an image sequence. (a) Number of the edges pixels of the corresponding sequence. (b) Different between images of contractions case A and non-contractions. (c) Edge patterns of contraction image (left side), and non-contraction image (right side).

detail implementations of a three-stage procedure. In Section 4, experiments and discussions of the results are presented. Finally, Section 5 concludes the work and suggests further research.

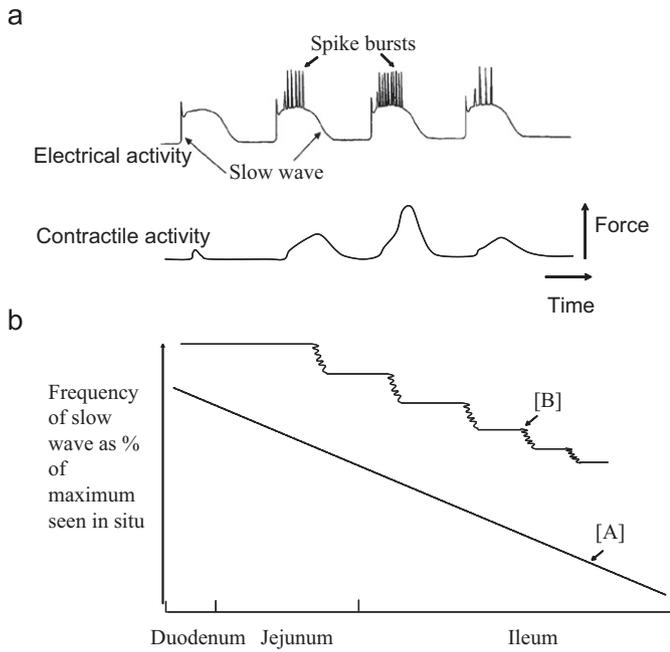
## 2. Methodology

### 2.1. Characteristics of intestinal contraction frequencies

From a physiological point of view, intestinal motility is manifested by electrical activity in the cell membranes in the intestinal wall, under myogenic, neural and chemical control [5]. The electrical activity can be divided into two distinct signals: the slow wave (referred to as the electrical control activity—ECA) and spike bursts (referred to as electrical response activity—ERA). Spike bursts occur during depolarization plateaus of the slow wave signals. They are associated with the muscular contractions required to produce intestinal contractions (Fig. 1a). Fig. 3a illustrates the relationship between the slow wave and the appearance of the spike burst signal. As shown, spike bursts are superimposed on the slow wave signal and are therefore phase-locked to the slow wave frequency. They can only occur at the crest of a slow wave, and the maximal frequency of contractile activity at any given site in the intestine is thus directly related to the slow wave frequency in that region. As stated

by [1,5], the slow wave frequency in the small bowel decreases in a series of steps from proximal to distal regions, as shown in Fig. 3b (for detail, refer to [1, Chapter 7] or [5, Chapter 6]). In humans, the maximal slow wave frequency in the duodenum is 12 cycles/min, but is only 3–4 cycles/min at the end of the small bowel.

Based on these formulations for contractions, if an existing  $f(x)$  signal reflects contraction activity in a WCE image sequence, potential contractions are located where this signal is a triangular shape. From the theory of signal processing, these positions can be located within windows including  $w$  frames by convolving this signal with kernel functions. Moreover, these kernels can be configured so that  $w$  can be adjusted during WCE transit through the small bowel. As a result, the length of candidate contractions can be adjusted to take account of the reduction in contraction frequencies. For example, at a capture rate of 2 fps, the maximal contraction rate in humans (around 12 cycles/min) suggests that the contractions visualized in WCE images should last for approximately five frames in the proximal regions of the small bowel. This value increases during WCE transit to adapt to the reducing frequency of contractions in distal regions. The section below explains the selection of image features so that the detection of patterns of contractions along  $f(x)$  signal can be implemented as well as determining true contractions.



**Fig. 3.** (a) Relations of slow wave and spike burst signals. Contractions are associated with the spike bursts that are superimposed on the slow wave signal [5]. (b) Slow wave frequency gradient in the small bowel. [A]: Linear gradient along small bowel. [B]: The frequency falls in a series of steps [1].

## 2.2. Characteristics of image features for the contraction detections

A WCE image sequence consists of frames of  $256 \times 256$  pixels and 24-bit color in RGB color space. The image features include a  $140^\circ$  field of view, 1–30 mm depth of view, and 1:8 magnification (referring to [9] for technical specifications). Similar to images captured by conventional endoscopy devices, WCE images usually show homogeneous regions inside the GI tract. To characterize the meaning of the WCE images, previous studies have selected various image features, depending on their aims. For example, to segment digestive organs in the GI tract, color features were favored in the works of Mackiewicz et al. [15,16] and Coimbra et al. [17,18]. For calculating disparity between images, Glukhovskiy et al. [19] used the average intensity of pixels or pixel clusters. To control display rates of WCE videos, Vu et al. [20] used combinations of color similarity and motion displacement features.

For detecting contractions, the edges of intestinal wrinkles are important cues, because shrinkage of the intestinal muscle is always associated with changes in the edge features. The result of previous works [10,11] also revealed that the edges of intestinal folds were discriminative features for the recognition of contractile patterns. Therefore, edge features were used in this study in order to construct a  $f(x)$  signal to identify potential contractions. Fig. 2a shows the edge signals from 60 consecutive frames in Fig. 1b. During contractions, the number of edge pixels in the image increases rapidly, and then decreases again. This pattern shows local peaks that correspond with the shrinkage of the intestinal folds in contraction cycles.

In addition to edge features, contraction events could be also inferred from sharp variations between consecutive frames, because the contractile activity spans adjacent frames with sudden shrinkages of the intestinal lumen. In order to make use of this feature, an evaluation of the distribution of similar regions in consecutive frames, called the similarity pattern, can be applied. An example of the maximum dissimilarities between pixels in adjacent contraction frames is plotted as in Fig. 2b—left panel. This discriminates them from non-contraction cases, as shown in Fig. 2b—right panel. As a

comparison, if the similarity pattern of a candidate contraction includes large regions of high similarity, it suggests it is not a true contraction.

The results of the above analysis of image features in consecutive frames, or temporal analysis, allow the identification of possible contractions. The patterns at the most extreme stage in the contraction cycle, as shown in Fig. 1b, when the intestinal wall strongly and sharply contracts the muscle towards a center point, are the most clearly observed. The structure of the image at this stage is therefore used to identify true contractions. Fig. 2c shows different patterns in contractile frame and non-contraction frame. The directions of intestinal wrinkles plot a structure with dominant orientations toward to a center point of the intestinal lumen in contraction cases (left panel); whereas this pattern is spread in a same direction in non-contraction cases (right panel). These discriminative features suggest that contractions can be described using a fixed paradigm or a statistical model of directional information. For example, in [10], descriptors code directional information in radial patterns, while in [11] a star-wise pattern was used. In this study, a statistical model is used to determine the pattern, because it can take account of natural characteristics of contractions that do not fit into specific categories.

The above analysis shows that contractions are highly recognizable by both spatial and temporal features. The addition of features reflecting the contraction frequency gradient, as described in Section 2.1, suggests a means of improving the recognition rate. Thus, if  $\Omega$  is a set of contractions in an image sequence, an instance including frames  $F_i \dots F_j$  can be considered to be located within a window such that the size of the windows are adjustable during WCE transit. They are true contractions if conditions below are satisfied:

$$\Omega = \left\{ F_i \dots F_j \mid \begin{array}{l} \text{edges}(F_i \dots F_j) \text{ is a maximal local peaks} \\ \text{lower distribution of } \text{similarity}(F_i \dots F_j) \\ \text{edges direction of the peak in } (F_i \dots F_j) \\ \approx \text{a contractile pattern} \end{array} \right. \quad (1)$$

Following this definition, a coherent procedure for implementations of the proposed method is described below.

## 3. Detection of contractions using a three-step procedure

### 3.1. 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:

$$f(x) = \sum_i \delta \quad \text{with} \quad \begin{cases} \delta = 1 & \text{if } i \text{ is an edge pixel} \\ \delta = 0 & \text{otherwise} \end{cases} \quad (2)$$

To extract edge features, some edge detection algorithms such as the Laplace of Gaussian and Canny edge detectors [21] can be implemented. Experiments show that the Canny method makes a trade-off between performance and computation times. To reduce the noise due to uninformative edges, edge pixels are counted in a region where most of the edges appear. The size of the region ( $192 \times 192$  pixels) is large enough to ensure that no important edges are lost. The position of the region is determined using a raster scan procedure. The signal  $f(x)$  is normalized in the range of  $[0, 1]$  and smoothed by a Gaussian function to remove noise.

Based on the properties of contractions as described in Section 2, 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. In particular, the morphological opening can suggest the location of positive peaks that are narrower than

a structural element. The opening operator consists of applying a dilation of an erosion of the smoothed signal  $f(x)$  by the flat structuring element  $B$  as defined by:

- Erosion operator  $\varepsilon_B(f(x))$ :

$$\varepsilon_B(f(x)) = \inf_{t \in B} [f(x-t)] = \min\{f(x+t) - B(t) | t \in B\}$$

- Dilation operator  $\delta_B(f(x))$ :

$$\delta_B(f(x)) = \sup_{t \in B} [f(x-t)] = \max\{f(x+t) + B(t) | t \in B\}$$

- Opening operator  $O(f(x))$ :

$$O(f(x)) = \delta_B[\varepsilon_B(f(x))] \quad (3)$$

A procedure to look for a number of consecutive frames within the opening signal that exceed the structuring element ( $B$ ) duration is implemented. As a result of this process, potential contractions are located within windows of length  $w$  (in number of frames). To satisfy the descriptions in Section 2.1,  $w$  can be adjusted in the context of changing values of  $\sigma$  of the Gaussian function and  $length$  of the structuring element  $B$ . A note that the  $\sigma$  is converted from full width at half maximum values (with  $FWHM = 2\sqrt{2 \ln 2} \sigma$ ). Meaning of  $FWHM$  values is to control number of nearest neighbor frames for the smoothing function. These values are set along the WCE transit time as described below:

---

```
# define Nstep K
# define nIncreasing k
% startingpoint: the frame number at the starting point of small
bowl
% endpoint: the frame number at the ending point of small
bowl
% sigma_0: initial value of the sigma of Gaussian function
% length_0: initial value of the length of structuring element B
nDuration = IntegerRound((endpoint - startingpoint)/K)
for i = 0 to K - 1
{
startingpoint = (i - 1) * Duration + startingpoint
endingpoint = startingpoint + nDuration - 1
if frame_number in [startingpoint..endingpoint]
{
sigma = sigma_0 + i * nIncreasing
length = length_0 + i * nIncreasing
}
call detect_maximal_local_peaks(sigma, length)
}
```

---

In our implementations, we defined values of  $Nstep$  ( $K = 4$ ), in which each step increased by  $nIncreasing$  ( $k = 2$ ). The positions of  $startingpoint$  and  $endingpoint$  of the small bowel are determined by the physicians. As a result, window size  $w$  is adjustable during the course of WCE transit, as shown in Fig. 4. For example, in the proximal regions, the contractions usually span five frames (corresponding with 2.5s), whereas in the distal regions, this value is larger, at approximately 11–13 frames. To avoid small variations, the distance between maximum and minimum values of local peaks must be above a certain level. The results of this procedure assign labels to starting and ending frames for each contraction. Fig. 5a shows an example of the processing of a sequence including 100 consecutive frames. Figs. 5b and c show the effects of different configurations on changing window sizes assigned for the same edge signal at nearly end of the small bowel. With configurations designed to obtain larger window sizes, the FAR is reduced. We will discuss the results from different positions in the small bowel in Section 4.

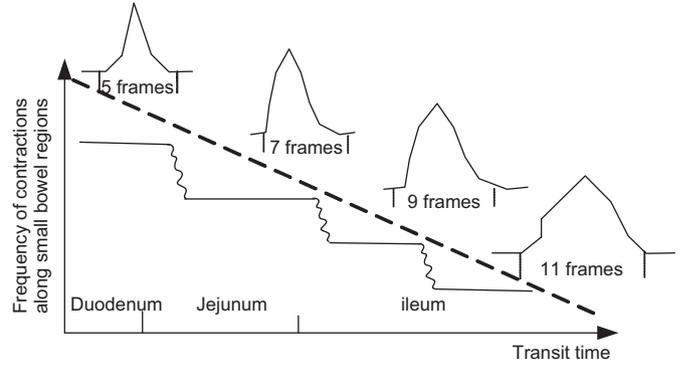


Fig. 4. Various window sizes during transit time of the WCE. The figure of contractions frequency gradient along small bowel is adapted from [1].

### 3.2. Stage 2—evaluation of the similarities between consecutive frames for eliminating non-contractions

The results in Fig. 5a show two instances of local maximal peaks: [A] representing a true contraction (positive case), whereas [B] is a false result (negative case). As described in Section 2.2 and shown in Fig. 3b, 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 [22]. 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 as in the procedure below. Assuming that a possible contraction includes  $w = N$  frames, feature vectors were constructed as in the configuration in Fig. 6a. 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 distance  $sim$  of a block  $t$  between two adjacent frames  $j$  and  $j+1$  is defined by

$$sim_{j,j+1}^t = \sum_{m=1}^{N_{bins}} |H_j^t(m) - H_{j+1}^t(m)|$$

$$\text{where } H(m) = \frac{1}{XY} \sum_{x \in X} \sum_{y \in Y} \sigma$$

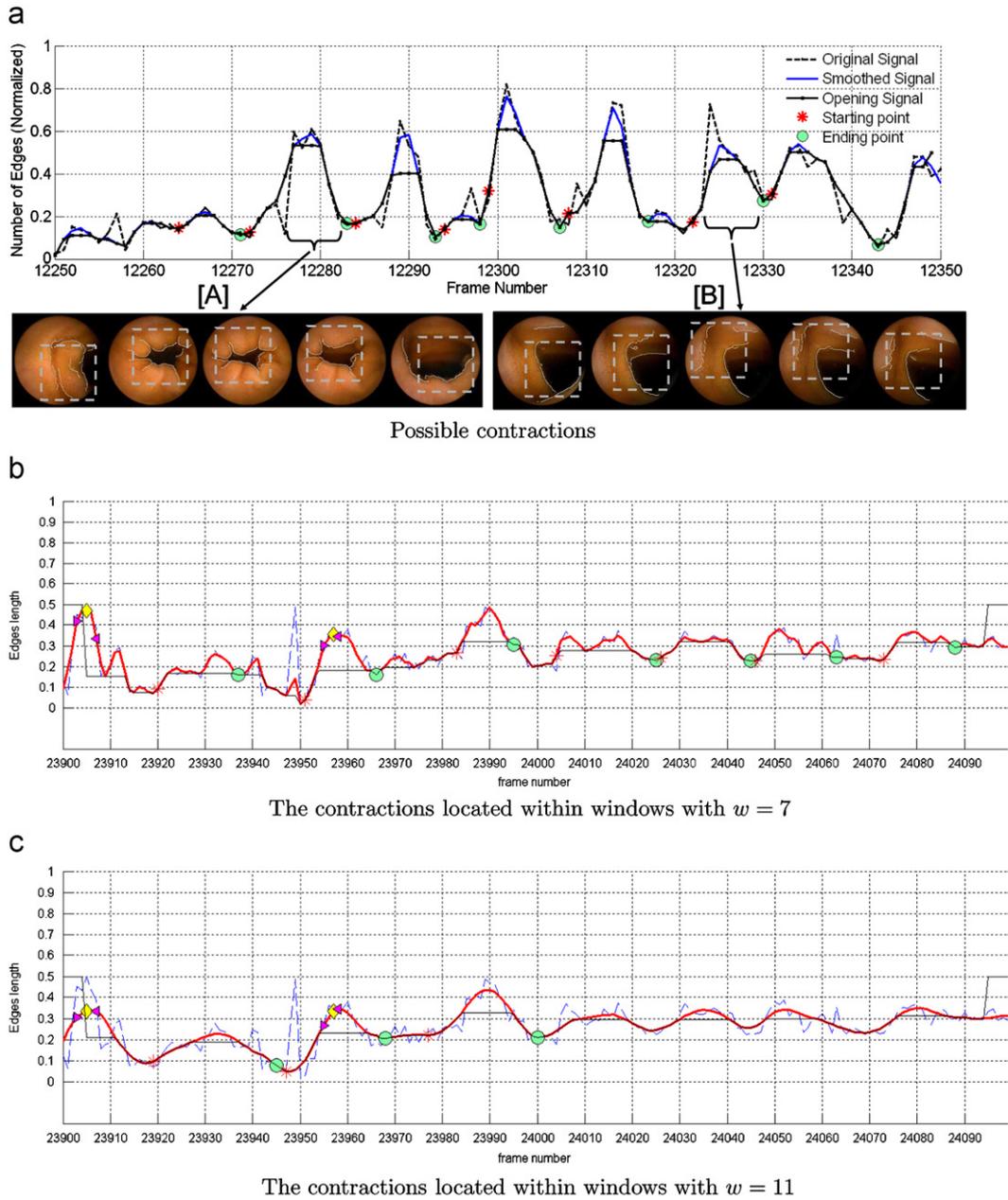
$$\text{with } \begin{cases} \sigma = 1 & \text{if } \text{Round}\left(\frac{I(x,y)}{N_{bins}}\right) = m \\ \sigma = 0 & \text{otherwise} \end{cases} \quad (4)$$

Thus, the feature vector  $\chi$  of the possible contraction is defined by

$$\chi = \{sim_{0,1}^t, \dots, sim_{N-1,N}^t, pos_x, pos_y\} \quad (5)$$

In (5), besides the similarity of blocks, the two features  $pos_x$  and  $pos_y$  identify a position of the block on the images. After discarding 16 pixels in each dimension that belong black corner regions, visible regions of WCE images are 240 pixels. In our implementation, values of  $N_{bins} = 16$  and the number of blocks  $N_{blocks} = 144$  (or a  $20 \times 20$  pixels size for each block) are predetermined. For a mixture of  $K$  Gaussians, a random variable  $\chi$  presents a probability to a Gaussian component  $k$  by

$$f_k(\chi|\theta) = \alpha_k \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} \exp \left\{ -\frac{1}{2} (\chi - \mu_k)^T \Sigma_k^{-1} (\chi - \mu_k) \right\} \quad (6)$$



**Fig. 5.** (a) 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. (b) and (c) show the results with different window sizes for edge signals almost at the end of the small bowel. The manual contractions are marked by diamonds (the strong stages) and rectangles (starting and ending) in (b) and (c).

where the parameter set  $\theta = \{\alpha_i, \mu_k, \Sigma_k\}_{k=1}^K$  consists of  $\alpha_k > 0, \sum_{k=1}^K \alpha_k = 1$  and  $\mu_k \in R^d$  and  $\Sigma_k$  is a  $[d \times d]$  positive definite matrix (in this case,  $d = N - 1$ ).

Given a set of feature vectors  $\chi_1, \dots, \chi_n$  a maximum likelihood (ML) criterion is used for training of the data to derive the parameters set  $\theta$ , yielding

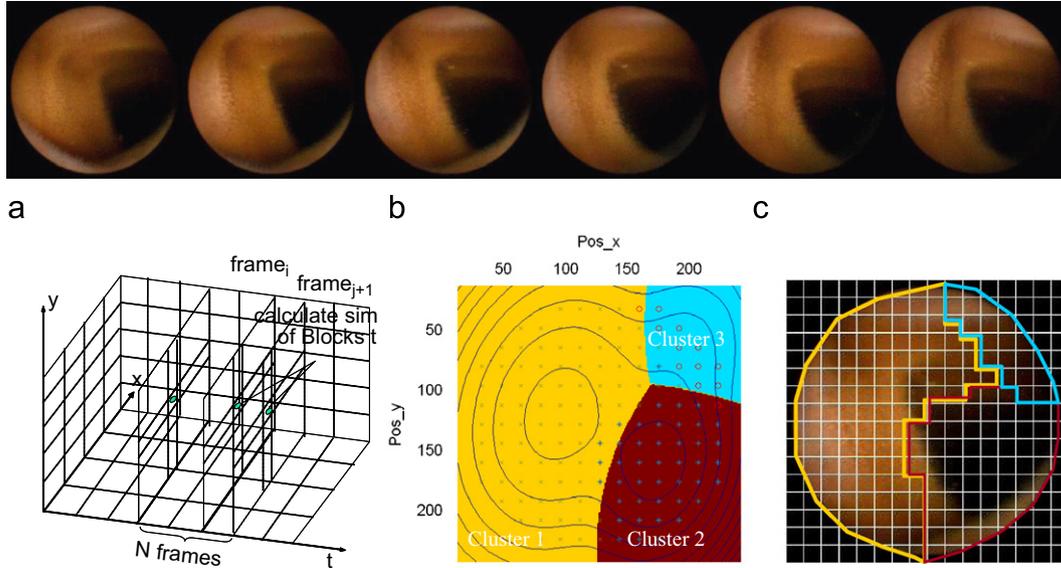
$$\theta_{ML} = \operatorname{argmax}_{\theta} f(\chi_1, \dots, \chi_n | \theta) \quad (7)$$

The EM algorithm [23] is an iterative method to obtain  $\theta_{ML}$ . The iterative updating process is based on a random initialization and repeated until it reaches a pre-defined number of iterations. The parameter set  $\theta_{ML}$  then provides probabilities following (6) to assign a feature vector  $\chi$  to a cluster by using a maximum a posteriori (MAP) principle. The MIXMOD library [24] is used for this scheme.

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. These would therefore be classed as non-contractions. A procedure for implementation is described below:

- Step 1: Select the largest clusters (the clusters with a high proportion; in our implementation, with a predefined  $K=3$ , the two largest clusters are selected).
- Step 2: Calculate mean values in these clusters.
- Step 3: If these mean values are smaller than a threshold, this candidate becomes a non-contraction case.

Figs. 6b and c show the results of this process to eliminate negative cases of contractions [B] in Fig. 5a.



**Fig. 6.** (a) Configuration to get the feature vectors. (b) Results of clustering similarity data (with  $K=3$ ) of the negative case (marked as [B]) in Fig. 5a, the ratio of cluster 1 is 61% and of cluster 2 is 30%. (c) Overlap the results on middle frames. Means of similarity data are 0.6 for cluster 1 and 0.53 for cluster 2.

### 3.3. Stage 3—detect true contractions through spatial features

The structures of contractile patterns vary because of their natural characteristics. Not all of the wrinkle directions are isotropic and the patterns are not always purely symmetrical. Moreover, the pattern in the occluded intestinal lumen is sometimes quite ambiguous to define. Therefore, unlike [10,11], we described the structure of the contractile image in terms of a statistical model. Features derived from this model were then utilized to classify true contractions.

From observations in Section 2.2, the orientation of the edges of intestinal wrinkles during the strongest stage of contraction was a powerful feature for discriminating between contractions and non-contractions. In order to characterize these patterns, an edge direction histogram was used. As shown in [25], 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.

Based on the edge detector in Section 3.1, the gradient directions of edge pixels were extracted. For each edge pixel  $p$ , its gradient vector is defined as

$$D(p) = \{dx, dy\}$$

where  $dx$  and  $dy$  are estimated from a gradient operator in vertical and horizontal directions, respectively. The amplitude and direction of gradient vectors are

$$Amp(p) = |dx| + |dy| \quad \text{and} \quad \theta(p) = \arctan\left(\frac{dy}{dx}\right) \quad (8)$$

To express directions, a polar histogram  $H$  was constructed, assuming that the direction range from  $0^\circ$  to  $360^\circ$  is divided into  $K$  bins (predefined with  $K=256$ ,  $\Delta\theta = 360/K = 1.4^\circ$ ):

$$H(\alpha_i) = \frac{N(\alpha_i)}{SN}$$

$$\text{where } N(\alpha_i) = \sum_{p \in \Theta} \log(Amp(p)) \quad \text{and} \quad SN = \sum_{i=1}^K N(\alpha_i),$$

$$\Theta = \left\{ p \mid \alpha_i - \frac{\Delta\theta}{2} \leq \theta(p) < \alpha_i + \frac{\Delta\theta}{2} \right\} \quad (9)$$

Figs. 7a and 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 ( $0-180^\circ$ ) are then divided into 16 bins of a histogram  $D$  in a Cartesian system, as shown in the left panel in Figs. 7a and b. Based on the signal of the histogram  $D$ , a simple  $K$ -nearest-neighbors ( $K$ -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

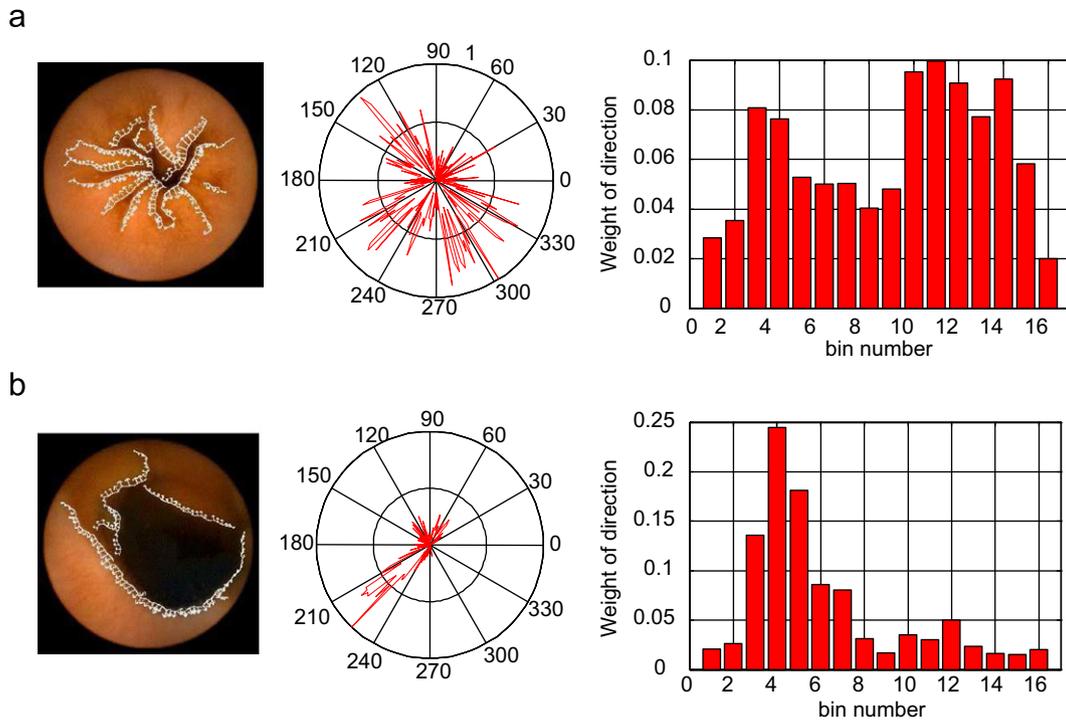
$$corre(x,y) = \frac{\delta_{xy} + C}{\delta_x \delta_y + C} \quad (10)$$

where  $\delta_x$  and  $\delta_y$  are the standard deviations of the feature vectors  $D_x$  and  $D_y$ , respectively;  $\delta_{xy}$  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). Given a query concerning the classification of a contraction, we found  $K$  cases in the training set closest to the query frames. A decision of whether or not it was a true contraction was based on the majority vote of the  $K$  objects found.

## 4. Experimental results and discussion

### 4.1. Experimental results

The experimental data were obtained from volunteers, supported by the Graduate School of Medicine, Osaka City University, Japan. Six WCE image sequences were obtained from different positions of the small bowel. The length of each sequence was 10 min. 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. Table 1



**Fig. 7.** 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.

**Table 1**  
The data for experiments

Seq. no.	Pos. in small bowel (min)	# of contractions
Seq_1	35	20
Seq_2	96	30
Seq_3	180	16
Seq_4	98	48
Seq_5	30	46
Seq_6	171	33

shows detailed information on the experimental sequences used for the evaluations.

According to the proposed method, the procedures for each stage were set up and implemented by C++ programs. An example demonstrates the effectiveness of each stage shown Fig. 8. The stages involved: mark possible contractions in Stage 1; remove redundant information in Stage 2 and; determine true contractions in Stage 3. 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:

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

$$\text{False Alarm Rate}(\text{FAR}) = \frac{FP}{TP + FP} \quad (11)$$

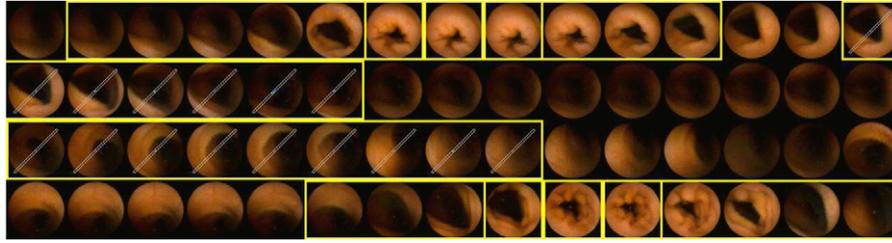
The effectiveness of adapting window sizes along the small bowel in the Stage 1 was evaluated by comparing this technique with the results using a fixed window size. The window sizes were assigned based on WCE transit through the small bowel, as in the procedure

described in Stage 1 (Section 3.1). Detailed data for each sequence are shown in Table 2.

As shown in Table 2, 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 convulsive 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 Stages 2 and 3, the overall performance of the proposed method is shown in Table 3. 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.

The above results suggest that contractions were successfully detected through a combination of temporal and spatial features. These features were effective for identifying contractions of variable lengths, instead of just fixed length ones, such as in the studies by [10] or [13]. This takes account of the natural characteristics of intestinal contractions. Moreover, spatial analysis using a statistical model can adapt to the ambiguous patterns of contractions (such as in Fig. 1b) rather than assuming a fixed, symmetrical pattern in the previous works. To compare qualitative indices between previous studies and the proposed method, although contractions in [10,13] were determined using different experimental data, average values for Sens and FAR were used, and are shown in Fig. 9. Note that the values of the Sens and FAR in Fig. 9 for methods [10,13] were re-computed from data supported in these studies, to match with the definitions in (11).

Using the proposed method, contractions could be detected throughout the full length of the small bowel, as shown in Fig. 10.



**Fig. 8.** An illustration of the effectiveness of the method. Possible contractions are marked inside rectangle boxes. The redundant cases are removed after evaluating the similarity between frames (marked by slanting), and the contractions are recognized as positive cases by using the classification method (marked in square boxes).

**Table 2**

Results of the stage one with fixed and variable window sizes

Seq.	Ground-truth data	Fixed window size <sup>a</sup>				Various window sizes <sup>b</sup>				Window size <i>w</i>
		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	

<sup>a</sup>Fix window size: *sigma* of Gaussian and *length* of element *B* (in the Stage 1) are set fixed values along full WCE transit time.

<sup>b</sup>Various window size: The configurations as described in Section 3.1 are deployed so that window size *w* can be adjusted during WCE transit. Correspondence values of *w* are shown in next column.

**Table 3**

Recognition rates of the experimental sequences using the proposed method

Seq. no.	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 contraction signals detected by variable window sizes were closer with the data detected manually. With window size adjusted during WCE transit, *FAR* was significantly reduced compared with that determined with a fixed window size. The performance of the proposed method was therefore more robust and reliable.

#### 4.2. Discussions

The computations of the three-stage procedure were performed off-line on a PC Pentium IV 3.2 GHz, 1 GB RAM. The average computation time for six sequences in the experiments was approximately 6 min. The longest time (4 min) was devoted to edge extraction in Stage 1. The computation time for Stage 2, approximately 90 s, was significantly reduced due to the elimination of non-contraction frames by Stage 1. Once the K-NN was trained in Stage 3, the classification results were produced in 30 s. In fact that computation time depended on the WCE transit time in the small bowel regions; e.g. with a data set of 100 patients, 80% small bowel transit time was approximately 200–250 min. With this duration of transit, the proposed procedures could span from 2 to 2.5 h. To get better performance achievements, a pre-computed edge extractions

(Stage 1) could be done with a back-end PC, and developing parallel algorithms within Stages 2 and 3 due to non-dependency characteristics of the features extractions procedures.

For overall performance, the proposed method still resulted in a high number of contractions being missed in Seq\_5 and Seq\_6. Because of weak contractions (Seq\_6), or ambiguous patterns of the contractions (Seq\_5), the directional features of these sequences were less clear in frames at the distal regions of the small bowel. To overcome these issues, other features such as changes in the area of lumen regions, or the shape of the wrinkles could be added to the learning paradigm at Stage 3. The procedure in Stage 1 to vary the window size along the small bowel was only based on WCE transit time and more features such as movement and position of the WCE could be used to estimate window size. With these improvements, this method would be able to take account of individual patient's data.

The detection of contractions throughout a full-length sequence, as shown in Fig. 10, revealed characteristic motility patterns for an individual patient. However, data based on a larger group of patients are required to confirm the results. Patterns could then be compared between healthy individuals and those with various diseases, in order to determine differences in intestinal motility

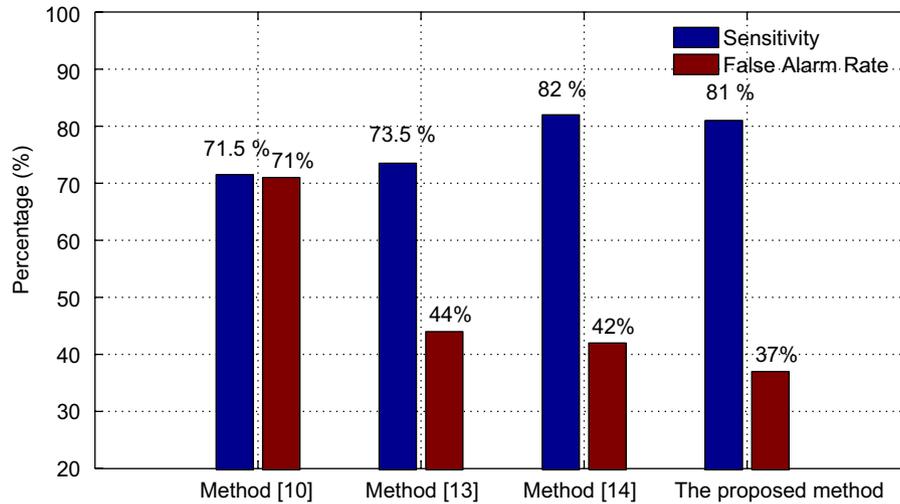


Fig. 9. Comparison of overall performance between the proposed method and previous methods.

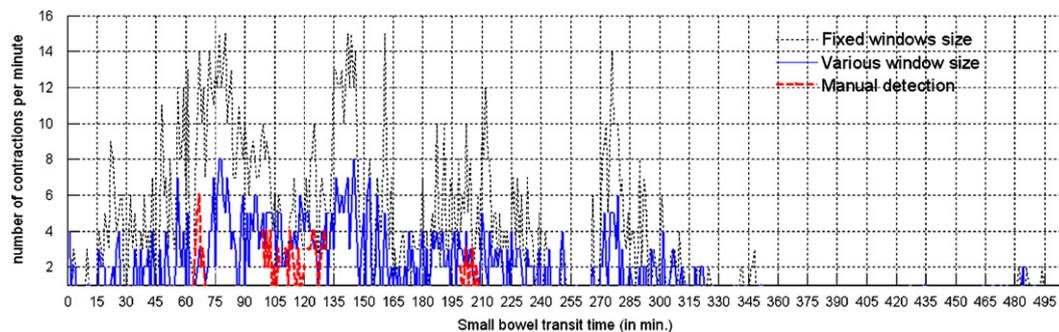


Fig. 10. Estimation number of contractions along GI tract of a full sequence.

characteristics. This could lead to future clinical applications, in which intestinal motility could be analyzed non-invasively with minimal attention from physicians.

## 5. Conclusion

This paper presented a method for the recognition of intestinal contractions during WCE of the small bowel, based on the efficient combinations of image features and the physiological properties of contraction frequencies. Contractions were successfully detected using a three-stage procedure: In Stage 1, potential contractions were located within windows including consecutive frames defined by edge signals. Relevant configurations were established such that the window sizes varied adaptively during the transit of the WCE. These features accounted for the characteristic physiological motility patterns. The possible contractions were then evaluated, based on the similarities between consecutive frames, allowing the negative cases to be discarded in Stage 2. For a final decision, in Stage 3, the spatial features of the possible contractions were classified using an edge direction histogram. The experimental results shown that 81% of true contractions were detected, with a 37% FAR. These experiments also confirmed that by considering the frequency patterns of the contractions, it was possible to eliminate many non-contraction cases throughout the WCE transit. Combinations of spatial and temporal features are therefore suitable, robust approach for the detection of contractions. The results suggest that this method could provide a non-invasive means of intestinal motility analysis with requiring minimal attention from physicians.

Some limitations of the proposed method have also been discussed, and suggestions for further research have been made. To ensure reliable results are obtained with different types of data, other features need to be included in the classification task. More information, such as WCE location and velocity, would allow more precise adjustment of window size. The development of a parallel algorithm would reduce the computation time. Examination of a larger sample of patients and further examination of the results of contraction detection could lead to valuable clinical applications.

## 6. Summary

Recognizing intestinal contractions from WCE image sequences is a non-invasive technique, which provides a solution to the problems of traditional techniques used for assessing intestinal motility assessment. Although not originally designed for this purpose, the image-capturing ability of WCE sequences reflect intestinal activity during the capsule transit time. These images provide a means of visualizing intestinal contractions. The aim of this study was to use computer vision techniques for the automatic detection and quantification of intestinal contractions, so providing physicians with a tool for assessing intestinal motility.

Based on the characteristics of contractile patterns and physiological properties relating to contraction frequency, we investigated image features extracted WCE image sequences, such as number of edge wrinkles, the similarity between frames, and directional edge information, to automatically recognize contractions during the transit of WCE through the small bowel. The method thus combined both temporal and spatial features. Contractions were detected

successfully using a three-stage procedure. Firstly, the possible contractions are recognized by changes in the edge structure of the intestinal folds (Stage 1) and evaluating similarities features in consecutive frames (Stage 2). In order to take account of the properties of contraction frequency, we consider that the possible contractions are located within windows including consecutive frames. The size of these contraction windows is adjusted according to the passage of the WCE. These procedures aim to exclude as many non-contractions as possible. True contractions are determined through spatial analysis of directional information in Stage 3.

Six sequences from different sections of the small bowel were used. This 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 reduce many non-contraction cases. 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. Some limitations of the method were discussed and further lines of research were suggested. To ensure reliable results with different types of data, other features need to be taken account of in our classification system. Further information, such as WCE location and velocity, would provide information to allow precise adjustment of window size, while the development of a parallel algorithm would improve performance time. The examination of data from more patients, as well as further analysis of the results, could lead to valuable clinical applications.

## 7. Conflict of interest statement

None declared.

## References

- [1] D. Grundy, *Gastrointestinal Motility—The Integration of Physiological Mechanisms*, MTP Press Limited, Hingham, MA, 1985.
- [2] M.B. Hansen, Small intestinal manometry, *Physiol. Res.* 51 (2002) 541–556.
- [3] E.A. Thomas, H. Sjøvall, J.C. Bornstein, Computational model of the migrating motor complex of the small intestine, *Am. J. Physiol. Gastrointest. Liver Physiol.* 286 (2004) 564–572.
- [4] H. Imam, C. Sanmiguel, B. Larive, B. Yasser, E. Soffer, Study of intestinal flow by combined videofluoroscopy manometry, and multiple intraluminal impedance, *Am. J. Physiol. Gastrointest. Liver Physiol.* 286 (2004) 263–270.
- [5] J.D. Bronzino, *The Biomedical Engineering Handbook*, third ed., CRC Press, Boca Raton, FL, 2006.
- [6] G. Iddan, G. Meron, A. Glukovsky, P. Swain, Wireless capsule endoscope, *Nature* 405 (2000) 417.
- [7] D.G. Adler, C.J. Gostout, Wireless capsule endoscopy—state of art, *Hosp. Physician* (2003) 14–22.
- [8] P. Swain, A. Fritscher-Ravens, Role of video endoscopy in managing small bowel disease, *GUT* 53 (2004) 1866–1875.
- [9] American Society for Gastrointestinal Endoscopy ASGE, Technology status evaluation report wireless capsule endoscopy, *Gastrointest. Endoscopy* 56 (5) (2002) 1866–1875.
- [10] F. Vilarino, P. Spyridonos, J. Vitria, F. Azpiroz, P. Radeva, Linear radial patterns characterization for automatic detection of tonic intestinal contractions, in: *Proceedings of the CIARP, Lecture Notes in Computer Science*, vol. 4225, Springer, Berlin, 2006, pp. 178–187.
- [11] P. Spyridonos, F. Vilarino, J. Vitria, F. Azpiroz, P. Radeva, Anisotropic feature extraction from endoluminal images for detection of intestinal contractions, in: *Proceedings of the MICCAI 2006, Lecture Notes in Computer Science*, vol. 4191, Springer, Berlin, 2006, pp. 161–168.
- [12] F. Vilarino, L. Kuncheva, P. Radeva, Roc curves and video analysis optimization in intestinal capsule endoscopy, *Pattern Recognition Lett.* 27 (8) (2006) 875–881.
- [13] P. Spyridonos, F. Vilarino, J. Vitria, F. Azpiroz, P. Radeva, Identification of intestinal motility events of capsule endoscopy video analysis, in: *Proceedings of the ACIVS 2005, Lecture Notes in Computer Science*, vol. 3708, Springer, Berlin, 2005.
- [14] H. Vu, T. Echigo, R. Sagawa, M. Shiba, K. Higuchi, T. Arakawa, Y. Yagi, Contraction detection in small bowel from an image sequence of wireless capsule endoscopy, in: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2007, Lecture Notes in Computer Science*, vol. 4791, Springer, Berlin, 2007, pp. 775–783.
- [15] M. Mackiewicz, J. Berens, M. Fisher, G.D. Bell, Colour and texture based gastrointestinal tissue discrimination, in: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, vol. 2, May 2006, pp. 597–600.
- [16] M. Mackiewicz, J. Berens, M. Fisher, Wireless capsule endoscopy video segmentation using support vector classifiers and hidden Markov models, in: *Proceedings of the International Conference on Medical Image Understanding and Analyses*, June 2006.
- [17] M. Coimbra, P. Campos, J.P. Silva Cunha, Topographic segmentation and transit time estimation for endoscopic capsule exam, in: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. II, May 2006, pp. 1164–1167.
- [18] M. Coimbra, J.P. Silva Cunha, Mpeg-7 visual descriptors—contributions for automated feature extraction in capsule endoscopy, *IEEE Trans. Circuits Syst. Video Technol.* 16 (5) (2006) 628–637.
- [19] A. Glukhovskiy, G. Meron, D. Adler, O. Zinati, System for controlling in vivo camera capture and display rate, Patent number PCT WO 01/87377 A2.
- [20] V. Hai, T. Echigo, R. Sagawa, M. Shiba, K. Higuchi, T. Arakawa, Y. Yagi, Adaptive control of video display for diagnostic assistance by analysis of capsule endoscopic images, in: *Proceedings of the 18th ICPR*, vol. 3, August 2006, pp. 980–983.
- [21] J. Canny, A computational approach to edge detection, *IEEE Trans. Pattern Anal. Mach. Intell.* 8 (1986) 679–698.
- [22] H. Greenspan, J. Goldberger, A. Mayer, A probabilistic framework for spatio-temporal video representation, in: *Proceedings of the ECCV 2002*, 2002, pp. 461–475.
- [23] A. Dempster, N. Laird, D. Rubin, Maximum likelihood from incomplete data via the em algorithm, *J. R. Stat. Soc. B* 39 (1) (1977) 1–38.
- [24] MIXMOD Ver. 2.0.1 (<http://www-math.univ-fcomte.fr/mixmod/index.php>), 2007.
- [25] W. Zhou, A. Bovik, H. Sheikh, E.P. Simoncelli, Image quality assessment: from error measurement to structural similarity, *IEEE Trans. Image Process.* 13 (2004) 600–613.

**Hai Vu** received the B.Sc. degree in Electronic and Telecommunication in 1999, and M.Sc. degree in Information Processing and Communication in 2002, both from Hanoi University of Technology, Vietnam. He is currently a Ph.D. student in the Graduate School of Information Science and Technology, Osaka University, Japan. His research interests are in computer vision, medical imaging.

**Tomio Echigo** received the B.Sc. and M.Sc. degrees in Electrical Engineering from the University of Osaka Prefecture, Osaka, Japan, in 1990 and 1982, respectively, and the Ph.D. degree in Engineering Science from Osaka University, Osaka, Japan, in 2003. He joined IBM Japan, Ltd., in 1982, where he was an Advisory Researcher at the Tokyo Research Laboratory, IBM Research, Kanagawa, Japan. From 2003 to 2006, he has been a Visiting Professor at Osaka University, Osaka, Japan. Since 2006, he has been with Osaka Electro-Communication University, Osaka, Japan, where he is currently Professor of the Department of Engineering Informatics. His research interests include image and video processing, medical imaging, and video summarization.

**Ryusuke Sagawa** is an Assistant Professor at the Institute of Scientific and Industrial Research, Osaka University, Osaka, Japan. He received a B.E. in Information Science from Kyoto University, Kyoto, Japan, in 1998. He received a M.E. in Information Engineering in 2000 and Ph.D. in Information and Communication Engineering from the University of Tokyo, Tokyo, Japan, in 2003. His primary research interests are computer vision, computer graphics and robotics (mainly geometrical modeling and visualization).

**Keiko Yagi** is a Lecturer at the Department of Clinical Pharmacy, Kobe Pharmaceutical University, Japan. She received a Ph.D. from Osaka City University. Her research interests are in the field of clinical pharmacology.

**Masatsugu Shiba** received a Doctor degree from Osaka City University Graduate Medical School, Japan, in 2002. Since 1995, he has been with Osaka City University Medical School Hospital, where he is an Assistant Professor of Gastroenterology in 2002. His research interests focus on gastrointestinal endoscopy and medical informatics.

**Kazuhide Higuchi** is a Professor at the Second Department of Internal Medicine, Osaka Medical College, Japan. He received M.D. degree 1982 and the Ph.D. degree in 1992, from Osaka City University. His interests are in the fields of gastroenterology, gastrointestinal endoscopy, capsule endoscopy, therapeutic endoscopy.

**Tetsuo Arakawa** is a Professor at the Department of Gastroenterology, Osaka City University Graduate School of Medicine, Japan. He received M.D. degree in 1975 and the D.M.Sc. degree in 1981, both from Osaka City University Medical School. He is a Director of Japanese Society of Senile Gastroenterology from 2001 and Japanese Gastroenterological Association from 2004.

**Yasushi Yagi** is a Professor at the Institute of Scientific and Industrial Research, Osaka University, Japan. He received B.E. and M.E. degrees in Control Engineering, 1983 and 1985, respectively, and the Ph.D. degree in 1991, from Osaka University. His interests are in the fields of computer vision, image processing and medical imaging.