

Contraction Detection in Small Bowel from an Image Sequence of Wireless Capsule Endoscopy

Hai Vu¹, Tomio Echigo², Ryusuke Sagawa¹, Keiko Yagi³, Masatsugu Shiba⁴,
Kazuhide Higuchi⁴, Tetsuo Arakawa⁴, and Yasushi Yagi¹

¹ The Institute of Scientific and Industrial Research, Osaka University

² Osaka Electro-Communication University

³ Kobe Pharmaceutical University

⁴ Graduate School of Medicine, Osaka City University

Abstract. This paper describes a method for automatic detection of contractions in the small bowel through analyzing Wireless Capsule Endoscopic images. Based on the characteristics of contraction images, a coherent procedure that includes analyzes of the temporal and spatial features is proposed. For temporal features, the image sequence is examined to detect candidate contractions through the changing number of edges and an evaluation of similarities between the frames of each possible contraction to eliminate cases of low probability. For spatial features, descriptions of the directions at the edge pixels are used to determine contractions utilizing a classification method. The experimental results show the effectiveness of our method that can detect a total of 83% of cases. Thus, this is a feasible method for developing tools to assist in diagnostic procedures in the small bowel.

1 Introduction

Information about contractions in the small bowel, in particular the number of contractions, distributions and their visualization along the Gastrointestinal (GI) tract usually reveals clinical pathologies in diagnostic procedures. For example, weak and disorganized contractions are stated as bacterial overgrowth disease; dysfunctions or the absence of contractions in a long duration is present in some patients with functional dyspepsia [1]. The current techniques to measure motility patterns in the small bowel are well tolerated but still invasive tests. Recently, a new clinical device using a system known as Wireless Capsule Endoscopy (WCE) [2,3] has become widely used. It allows examinations in the small bowel to be both more intensive yet comfortable for patients. Using these image data as the source for recognition of contractions presents intuitively opportunities for developing tools to assist in diagnostic procedures.

In a typical examination, the WCE goes through the GI tract in 7-8 hours capturing images at 2 fps, with around 20,000 to 30,000 frames of intestinal images [3]. Visualization of the contractions is rendered in consecutive frames showing shrinkage of intestinal folds. Particularly, at center of the contractions, a wrinkle pattern that presents a state of occlusion of the intestinal lumen has

strong and sharp edges of wrinkles toward to a center point. Thus, contractions are highly recognizable in a WCE image sequence from the combination of the spatial and temporal features.

Addressing the issue of the detection of contractions is still at an initial state. To recognize wrinkle patterns, extracted features from a structured tensor of the intestinal images are mapped into a star-wise pattern in [4]; or in [5] the features of "steerable filters" are coded into radial shape patterns. On the other hand, along a time dimension the contractions are determined by examining 5 wrinkle frames within a window of a ± 5 frames neighborhood (in [5]) by an evaluation of intestinal lumen features (solidity, sharpness, deepness) of each image in an interval of 9 continuous frames (in [6]). Accordingly, contractions are generated in the form of a fixed number of frames by using features of separate frames.

Different from these systems, we examine continuous changes of features over time and evaluated the relationship between frames and other frames to detect possible contractions. Then, the spatial features are analyzed by describing the structural pattern to determine contractions utilizing a classification method. Following this strategy, the detection of contractions is more reasonable because the duration of the contractions is not constrained but varies along the small bowel. Implementations are undertaken in a coherent procedure that includes three stages to automatically detect the contractions. The experimental results imply that the method provides a useful way to recognize contractions.

The rest of this paper is organized as follows: Sec. 2 describes the characteristics of the contractions. Sec. 3 explains the techniques for implementation of a three-stage procedure. In Sec. 4, experimental results and discussions are presented. Finally, Sec. 5 concludes the work and suggests further research.

2 Characteristics of Contractions in a WCE Image Sequence

In terms of physiology, contractions are produced by the shrinkage of circular muscle layers. A cycle of contractions begins at the widest state of the intestinal lumen and rapidly occludes the lumen area until reaching the strongest shrinkage state of the inner wall, then these folds relax at an ending state. Fig. 1 marks contractions in an image sequence that is made up of 60 frames.

As showing of the marked frames, the edge features are important cues for detecting contraction events. The number of edge pixels in the images increases rapidly and then decreases, resulting in a strong peak at the contractions. Moreover, because of the movement of the WCE, the changes between successive frames at contraction positions is more discriminative than others.

On the other hand, the center of the contractions presents a muscular structure, as shown in Fig. 2. These patterns include strong edges of the intestinal wrinkles when the inner walls are folded. The muscular tone is toward a center point and produces a state of occlusion of the intestinal lumen. The structure of wrinkle patterns thus can be expressed by directional information.

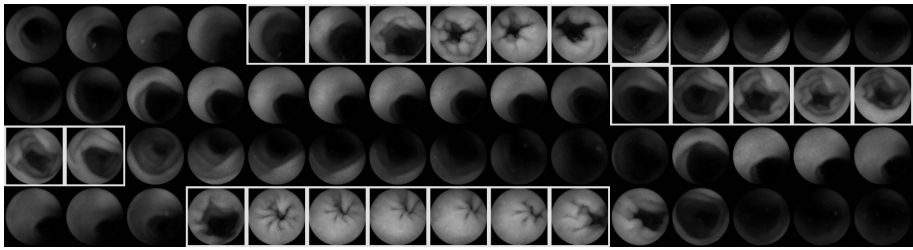


Fig. 1. Some contractions are marked in boxes of 60 continuous frames (from left to right and top to bottom)

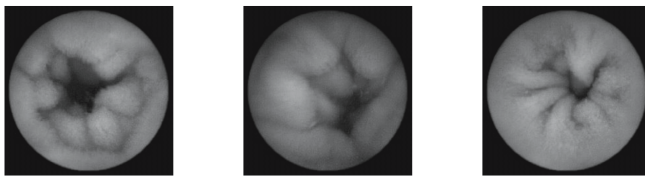


Fig. 2. Some examples of the contractile patterns

For these characteristics, detecting maximal local peaks along the edge signal and evaluating the similarity of successive frames allows for the detection of events, which are considered as possible contractions. These results are highly recognizable for final decisions based on analyzing the directions at the edge pixels. A coherent procedure for implementation is described below.

3 Contractions Detection by Temporal and Spatial Features

3.1 Edges Extractions to Detect Possible Contractions

To detect possible contractions, denote $f(x)$ as a function of the edge number:

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

The Canny edge detector [7] is used to extract edges from an image. Edge pixels are counted in a region where most of the edges appear. The size of the region (192x192 pixels) is large enough to ensure that no important edges are lost (with image resolution is 256x256 pixels). The signal $f(x)$ is normalized in the range of $[0, 1]$ and smoothed by a Gaussian function to remove noise.

Possible contractions are located where $f(x)$ is in a triangular shape. However, not all signals perfectly present this type of pattern because of the length and the strength during the contractions. Thus, a morphological opening operator is applied to create a simpler graph than the original signal. The opening signal suggests locations of positive peaks that are thinner than a structural element.

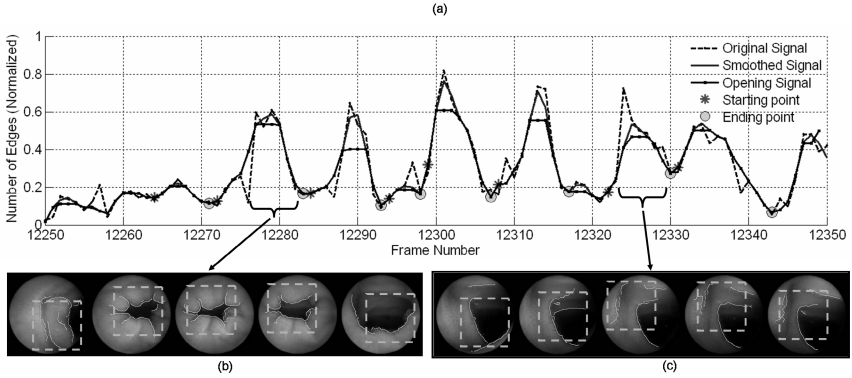


Fig. 3. (a) Possible contractions are marked on an original signal with starting (star points) and ending (circle points) frames. (b) A positive case; (c) a negative case for a contraction. The most edges regions are also marked in rectangles in (b) and (c)

A procedure to look for a number of consecutive frames within an opening signal that exceeds the structuring element duration is implemented. Fig. 3 shows the positions of contractions detected in a sequence of 100 consecutive frames.

3.2 Evaluations of Similarity for Eliminating Non Contractions

Figure 3 shows two cases of possible contractions in which sharp variations between consecutive frames in the positive case (3b) are opposed to the high similarity in the negative case (3c). Shrinkage of the intestinal folds in the positive cases make most regions that have low similarity, whereas the negative cases imply a high similarity of homogeneous regions spanning consecutive frames. These regions can be grouped into clusters by grouping similarity feature space then evaluating similarity to discard redundant cases. Therefore, an unsupervised clustering method that adopted from works in [8] are applied. Based on an observation that a block of image pixels is more likely to belong to a certain cluster if it is located near the cluster centroid. It is reasonable to assume that the similarity of blocks and their positions is represented by a Gaussian distribution, and a set of regions are generated by a mixture of Gaussians.

First, N frames of a possible contraction are divided into N_{blocks} (Fig. 4a) and an intensity histogram H (with N_{bins}) for each block is calculated. The similarity sim of block t between frames j and $j+1$ are:

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

$$With H(m) = \sum_{x,y \in block} \begin{cases} 1 & \text{if } IntegerRound(\frac{I(x,y)}{N_{bins}}) = m \\ 0 & \text{Otherwise} \end{cases} \quad (2)$$

A feature vector is notated by: $\chi = \{sim_{0,1}^t, \dots, sim_{N-1,N}^t, pos_x, pos_y\}$, including the similarity of blocks and their positions pos_x and pos_y . For a mixture

of K Gaussians, the random variable χ presents a probability for a Gaussian component k by:

$$f_k(\chi|\theta) = \alpha_k \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} \exp\left\{-\frac{1}{2}(\chi - \mu_k)^T \Sigma_k^{-1}(\chi - \mu_k)\right\} \quad (3)$$

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 Σ_k is a $[d \times d]$ positive definite matrix (in this case, $d = N - 1$).

Given a set of feature vectors $\chi_1, \dots, \chi_{Nblocks}$, a Maximum Likelihood (ML) criterion is used to train the data to derive a parameters set θ , yielding:

$$\theta_{ML} = \arg \max_{\theta} f(\chi_1, \dots, \chi_{Nblocks}|\theta) \quad (4)$$

The EM algorithm [9] is an iterative method to obtain θ_{ML} . The parameter set θ_{ML} then provides probabilities following Eq. 3 to assign a feature vector χ to a cluster using a Maximum A Posteriori (MAP) principle. The MIXMOD library [10] is used for this implementation. The results of clustering are then assessed through examining similarity data of the largest clusters. If these regions include high similarity values, it implies a low probability of a true contraction and so is decided as being non contractions. For example, Fig. 4 shows the results for a negative case of contractions in Fig. 3c. With number of clusters $K = 3$, $Nblocks = 144$ and $Nbins = 16$ are preselected in order to obtain a trade-off between number of clusters regions and computational time, two largest clusters 1 and 2 include 60% and 28% total blocks. The average of similarity in these regions are 0.63 and 0.52, respectively. This result show large homogenous regions along the frames that are reasons to assign this case as one of non contractions.

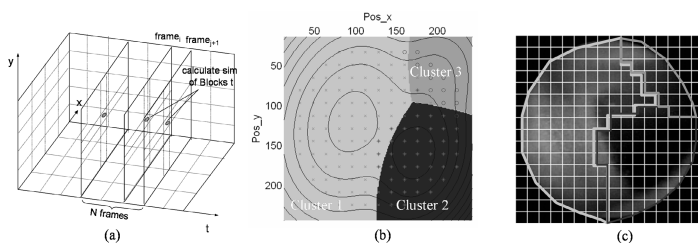


Fig. 4. (a) A configuration to obtain feature vectors. (b) Results of clustering similarity data using GMM for the possible contraction in Fig. 3(c). (c) Borders of the clusters are superimposed on middle frames.

3.3 Detect True Contractions Through Spatial Features

As descriptions of the wrinkle patterns, orientations distribution of edge pixels appears to be a powerful feature for discriminating between contractions and non contractions. For the natural characteristics of contractions, not all of the directions of wrinkles are isotropic and these patterns are not always purely symmetric. Thus, we describe the structure of an image by using an edge direction

histogram that seems well able to deal with more general cases of contractile patterns. The frame that has the maximum edge number of each possible contraction is selected for this procedure.

For each edge pixel p , its gradient vector is defined as: $D(p) = \{dx, dy\}$. The amplitude and direction of gradient vectors are:

$$Amp(p) = |dx| + |dy| \text{ and } \theta(p) = \arctan\left(\frac{dy}{dx}\right) \quad (5)$$

To express directional features, a polar histogram H is built with the assumption that the directions range from 0 to 360° and are divided into K bins (predefined with $K = 256$, $\Delta\theta = 360/K = 1.4^\circ$):

$$H(\alpha_i) = \frac{N(\alpha_i)}{SN} \text{ where } N(\alpha_i) = \sum_{p \in \Theta} \log(Amp(p)) \text{ and } SN = \sum_{i=1}^K N(\alpha_i) \quad (6)$$

$$\Theta = \{p | \alpha_i - \frac{\Delta\theta}{2} \leq \theta(p) < \alpha_i + \frac{\Delta\theta}{2}\}$$

Figure 5 shows the polar histograms H of non contraction (5a) and contraction (5b) cases. The patterns of the polar histogram show the directions are spread every way in the contraction case, whereas in the non contraction case, the polar histogram is distributed in only a dominant direction.

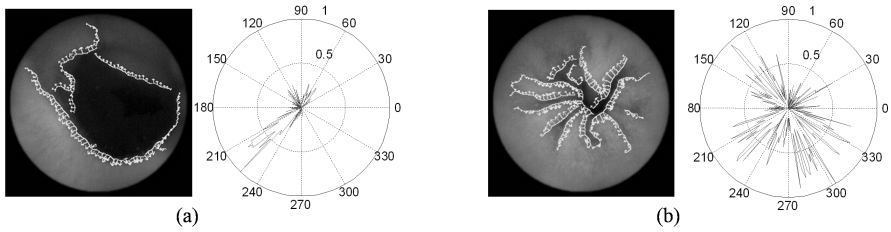


Fig. 5. Direction histogram of a non contraction (a) and contraction (b). Left side shows the original frames with gradient direction at edge pixels, right side is a polar histogram

Based on the signal of H , a simple K-Nearest-Neighbours classifier is used to decide the contraction pattern. The structural similarity between two feature vectors Hx and Hy is estimated by calculating the correlation coefficient $corre(x, y)$, that is:

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

where δ_x and δ_y are standard deviations of Hx and Hy ; δ_{xy} is the covariance of vectors and C is a small constant to avoid the denominator being zero. The K-NN classifier trained with a data set which includes 1000 frames has been labeled manually as non contraction and contraction cases.

4 Experimental Results

The experimental data were supported by the Graduate School of Medicine, Osaka City University, Japan. Six sequences were extracted from different parts of the WCE image sequences in the small bowel. The length of each sequence is 10 minutes. For each sequence, to get ground truth data, manual detections were implemented by medical doctors. The positions at starting and ending frames and the strongest position of each contraction are also marked.

According to the proposed method, the procedures are set up and implemented by C++ programs on a PC Pentium IV 3.2 GHz, 1 GB RAM. Fig. 6 shows the results of the method for an example of 60 continuous frames (from left to right and top to bottom). To evaluate the performance of the proposed method, data as below are calculated for each sequence:

- 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)

Using these data, two criteria for the evaluation are:

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

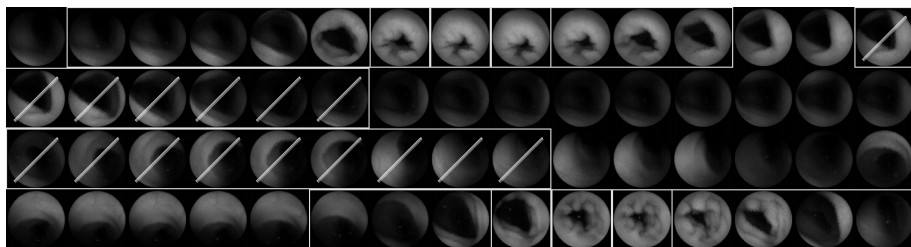


Fig. 6. An illustration the effectiveness of the method. Possible contractions are marked inside rectangle boxes. The redundant cases are removed after evaluating the similarities between frames (marked by slanting). Contractions are recognized as positive cases after utilizing the classification method (marked in square boxes).

The two first stages aim to effectively reduce non contractions with a minimum loss of true positives. The sensitivity of Stage 1 and Stage 2 are 96% and 92%, whereas the false positive rates are 68% and 52%, respectively. For an evaluation of the overall performance of the entire process, Table 1 shows detailed results of each sequence.

Comparisons of the average of the results with those reported in [5] and [6] are 71.5% and 73.5% for sensitivity, and 71% and 44% for false alarm rate, respectively. Obviously, with the proposed method, which combines both spatial and temporal features, the performance is more robust and thus more reliable.

Table 1. Results of the overall process for each sequence

Sequence	Manual Detection	Proposed method	True Pos.	Lost Rate	Sensitivity	FAR
Seq_1	20	56	19	5%	95%	66%
Seq_2	30	44	27	10%	90%	38%
Seq_3	16	25	13	19%	81%	48%
Seq_4	48	50	40	17%	83%	20%
Seq_5	46	60	35	24%	76%	41%
Seq_6	33	41	24	27%	73%	41%
Mean				17%	83%	42%

However, the loss rate is still high in Seq_5 and Seq_6. The reason being that the direction features are less effective for frames at the end of the small bowel because of the weak contractions (Seq_6) or that some contractions have ambiguous patterns (Seq_5). To overcome this issue, more features as changes in the darkness area, or variations of wrinkles patterns (ex., using linear radial patterns in [5]) along the time dimension can be added into the learning paradigm.

5 Conclusion

This paper presented a method to recognize contractions in the small bowel based on analyzing temporal and spatial features. Contractions were successfully detected through a coherent procedure. For temporal features, variations of edge features and evaluations of similarity data between interval frames were implemented to detect possible contractions. To detect a true contraction, the spatial features of the possible contractions were presented through descriptions of an edge direction histogram. From the experimental results, the overall performance implied that the proposed method could detect 83% of the total contractions. Thus, analyzing WCE image sequences by a combination of spatial and temporal features appears a useful way to characterize contractions in the small bowel. However, to ensure more reliable results with different types of data, in future work we need to consider and examine other features to be factored into the classification. In this way, the method proposed here will become a feasible method for developing tools to assist in diagnostic procedures.

References

1. Hansen, M.B.: Small intestinal manometry. *Physiological Research* 51, 541–556 (2002)
2. Iddan, G., Meron, G., Glukovsky, A., Swain, P.: Wireless capsule endoscope. *Nature* 405, 417 (2000)
3. Swain, P., Fritscher-Ravens, A.: Role of video endoscopy in managing small bowel disease. *GUT* 53, 1866–1875 (2004)

4. Spyridonos, P., Vilarino, F., Vitria, J., Azpiroz, F., Radeva, P.: Anisotropic feature extraction from endoluminal images for detection of intestinal contractions. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 161–168. Springer, Heidelberg (2006)
5. Vilarino, F., Spyridonos, P., Vitria, J., Azpiroz, F., Radeva, P.: Linear radial patterns characterization for automatic detection of tonic intestinal contractions. In: Martínez-Trinidad, J.F., Carrasco Ochoa, J.A., Kittler, J. (eds.) CIARP 2006. LNCS, vol. 4225, pp. 178–187. Springer, Heidelberg (2006)
6. Spyridonos, P., Vilarino, F., Vitria, J., Azpiroz, F., Radeva, P.: Identification of intestinal motility events of capsule endoscopy video analysis. In: Blanc-Talon, J., Philips, W., Popescu, D.C., Scheunders, P. (eds.) ACIVS 2005. LNCS, vol. 3708, Springer, Heidelberg (2005)
7. Canny, J.: A computational approach to edge detection. *IEEE T-PAMI* 8, 679–698 (1986)
8. Greenspan, H., Goldberger, J., Mayer, A.: A probabilistic framework for spatio-temporal video representation. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 461–475. Springer, Heidelberg (2002)
9. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the em algorithm. *Journal Royal Statistical Society B*, 39 (1), 1–38 (1997)
10. MIXMOD Ver. 2.0.1: (2007),
<http://www-math.univ-fcomte.fr/mixmod/index.php>