

Research Article

A diagnosis support system for capsule endoscopy

Y. Yagi^{*1}, H. Vu¹, T. Echigo², R. Sagawa¹, K. Yagi³, M. Shiba⁴, K. Higuchi⁴, T. Arakawa⁴

¹ The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Osaka, 567-0047, Japan, Fax: ++81-6-6877-4375, e-mail: yagi@am.sanken.osaka-u.ac.jp,

² 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 Osaka City University Graduate School of Medicine, 1-4-3 Asahimachi, Osaka, 545-8585, Japan

Received and accepted 12 July 2006

Abstract. The diagnostic time required for a full, 8-hour video capsule endoscopy is usually between 45 and 120 min. The aim of this work is to evaluate the diagnostic time required when applying a method that adaptively controls the image display rate. The advantage of the method is that the sequence can be played at high speed in stable smooth sequences to save time and then decreased at sequences where there are sudden rough changes, in order to assess suspicious findings detail. In this paper, this method is examined under real conditions: 10 sequences were independently evaluated by 4 medical doctors. The methods of evaluation include: 1) the time required for reading a sequence, 2) the percentage of abnormal regions accurately found, and 3) the manipulations of the evaluating physicians. The results indicate that the proposed method reduces diagnostic time to around $10 \pm 1.5\%$ length of the sequence and is of valuable assistance to medical doctors.

Key words: Wireless capsule endoscopy; Adaptive image display; Diagnostic time

1. Introduction

The investigation and visualization of the whole of the small bowel has been greatly improved by the great breakthrough known as the wireless capsule endoscope (WCE). One clinical product has become widely used is the M2ATM (Iddan et al., 2000; Adler, 2003; ASGE, 2002), invented by Given Imaging Ltd (Yoqneam, Israel). The video sequence of a typical examination has around 57,000 frames that can be used for interpretation in diagnostic procedures. With such a huge

number of frames, examination by physicians is not an easy task, and it is also time consuming. Using the Rapid^{TM2} application (developed by Given Imaging Ltd), controlling the display of images is adjustable, allowing manual changes between 1 and 25 frames per second to be made. The reports (Hadathi et al., 2005; Swain, 2004; Medical Advisory Secretariat, 2003) showed that the average time to evaluate a video ranged from 76 ± 30 min. In later versions, Rapid^{TM3} and Rapid^{TM4}, a variable viewing speed function and a “quick view”/“omit frames” mode was added. Keuchel et al. (2006) investigated precise reading times using these two versions. Using the automatic varying speed mode of Rapid^{TM3} yielded a reading time of 52.4 ± 16.4 min, while a time of 36.9 ± 13.4 min was obtained using the “omit frames” mode of Rapid^{TM4}. These experiments were carried out by two gastroenterologists on 40 sequences. However, the report also notes that using the “omit frames” mode in Rapid^{TM4} requires additional evaluation in order to confirm that there was no loss of relevant findings. Unfortunately, no technical description of these features is currently known to exist.

Obviously, reducing the diagnostic time of a wireless capsule video endoscopy is still a major obstacle. A technique is proposed in our work (Vu et al., 2006) for efficiently controlling the frame rate of the display of images. Sequences are played at high speed in continuous frames that are stable in order to save time, and the speed is decreased at areas where there are rough changes so that any suspicious findings can be more conveniently ascertained. In other words, the main idea is to keep the frame rate of the image display proportionally linear with the difference between two consecutive frames. This technique, therefore, can intuitively assist physicians by automatically changing speed when reading a sequence.

The above reports also imply that the evaluation of techniques, which assist doctors reduce time consumption, are urgently required. Therefore, in this study, methods for evaluating the capacity of the proposed technique though ex-

* Corresponding author

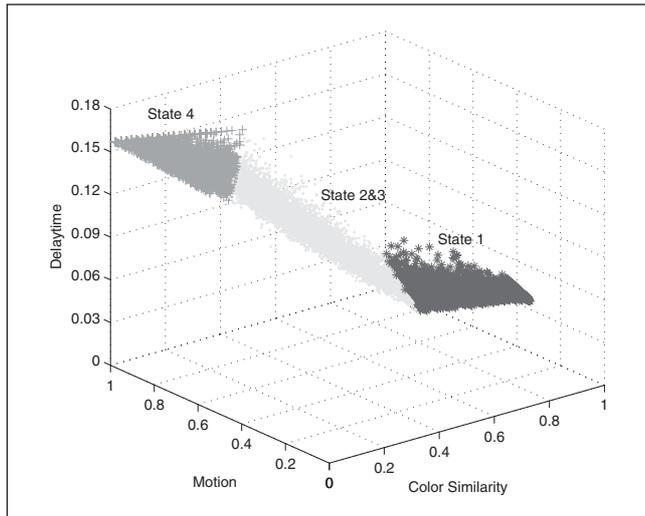


Fig. 1. Distribution of delay time and corresponding features of color similarity and motion displacement a sequence using the proposed method.

periments are presented. Four medical doctors implemented diagnostic procedures on 10 sequences by a GUI system, which was developed to demonstrate the proposed method. The results of the evaluations are analyzed from three aspects: the average diagnostic time, the percentage of abnormal regions found, and the manipulations of the physicians regarding the control of the image display in the second stage. The results of the experiments provide evidence in support of the method introduced for reducing diagnostic time and the ability to assist medical doctors in capturing abnormal regions.

2. The method of adaptively controlling image display

A physician's observations when viewing a sequence are impacted by movements of the capsule. Therefore it is important to select set features that can adequately present the changing states of image acquisition and a method to calculate the time delay between two consecutive frames. Below, we describe a summary the technique to address this problem. For further details, please consult in Vu et al. (2006).

Based on the characteristics of the capsule, endoscopic images are generally covered by many homogeneous regions of villi and remain congruent between sequential frames; with movement being controlled though natural peristalsis throughout the entire GI tract. The two features of color similarity and motion displacement are selected and extracted to measure changes between two consecutive frames. There are four different states of image acquisition, and they are described below along with a scheme for controlling the image display for each state:

- State 1: Both the small bowel and the capsule are stationary. This state appears when the contractions of the small bowel are in a stable phase, which makes the capsule remain almost still. As for the control of display images in this state, when continuous frames are exactly the same, the frames play at high frame rate to save time.

- State 2: Movements of the small intestine are small. This state consists of consecutive frames of gradual transitions. The state corresponds with moments where the impact of contraction are not so large. The display images in this state are controlled at medium speed.
- State 3: The small intestine has larger movements. A phase of irregular contraction impacts is characteristic of this state. Doctors examining this state would therefore need to view display images at slower speeds in order to place more of a focus on the changing regions. Therefore images would be displayed longer.
- State 4: The small intestine has abrupt changes causing capsule movements. This state is reached when the contraction phase is at its strongest point. The interval time between two frames increases to the maximum to enable viewing to be as easy as possible.

An example of the distribution of delay time and corresponding color similarity and motion displacement of a sequence is shown in Figure 1. The area of distribution of State 1 shows that delay time is round 30ms/frame. At States 2 and 3, the distributions are sloped and proportionally linear with the features of motion and color similarity. At State 4 (abrupt changes) the delay time is around 150ms/frame. Thus it can be seen that the range of delay times is from 30ms/frame to 150ms/frame, providing evidence of the impact of the proposed method (compared with the sequence play method at a fixed frame rate (i.e., a typical rate is 13 fps), in which the delay time between two frames is a constant value: 77 ms).

The definitions of states and calculation of the corresponding delay times are dependent on the parameters of the method. The optimal values are decided from suggestions of medical doctors through heuristic experiments. In addition, the method also supports a variety of skill levels in order to assist medical doctors of different specializations in their understanding of capsule endoscopic images.

3. Methods for performance evaluation

The examination data and evaluation procedures were implemented with the assistance of medical doctors in the Osaka City University Medical School, Japan. To ensure unbiased evaluations, the experiments were set up under conditions of normal diagnostic procedures. A GUI system was developed on desktop PCs to present the proposed method, which also supports common functions such as the capturing of abnormal regions, the changing of display modes, the adjusting of skill levels, and functions to navigate and scanning/browsing frame-by-frame.

Ten sequences from patients were selected and extracted with a length of 1.5h to save time for the physicians' evaluations. The delay times of the extracted sequences were calculated by the optimal values.

The evaluations were carried out by four medical doctors (referred to in this paper as MD. A, MD. B, MD. C, and MD. D), and their skill levels ranged from junior to senior in the field of diagnostic endoscopic imagery. The medical doctors were asked to independently find and capture suspicious regions. The time codes of abnormal regions as well

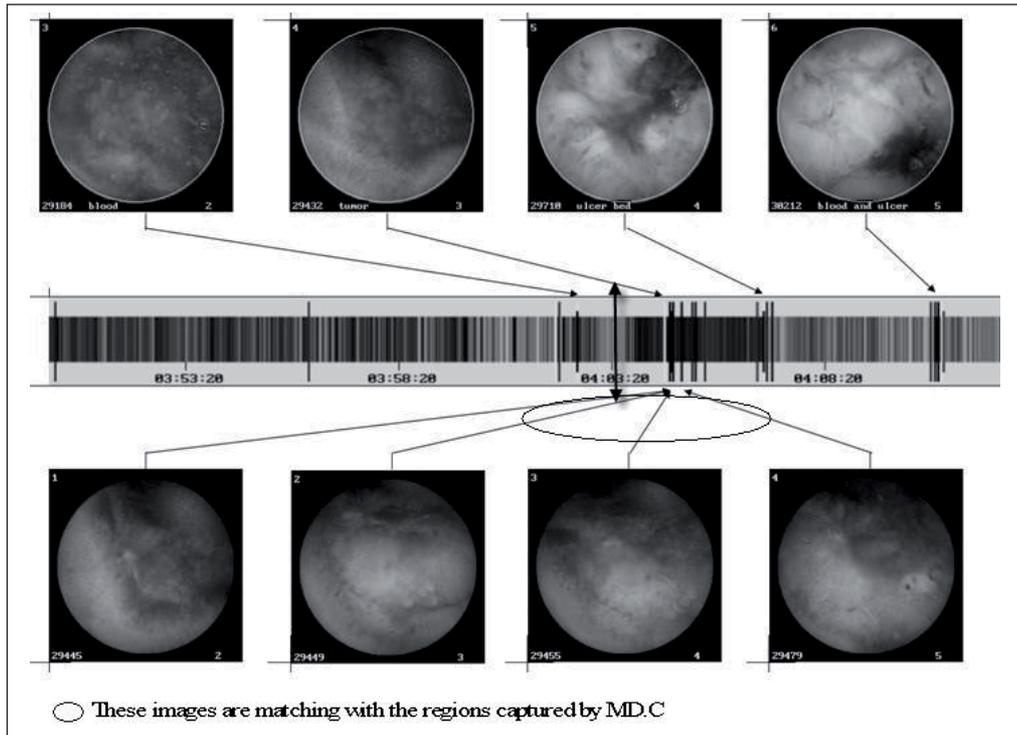


Fig. 2. The method to verify abnormal regions is captured by imaging. The images captured by the MD. (bottom row) under assessment in compared to the baseline, time ground data, of MD.C.(top row)

as the events/activities of the medical doctors during the diagnostic procedures were logged. For assessment of the capacity and performance, these data then were analyzed and inspected as described below.

3.1 Average measure of the diagnostic time

To explore in detail the diagnostic time for each evaluation section, the time code data at the moment of each start/stop action were analyzed. In addition, frame-by-frame scanning to finding abnormal regions was also inspected. The diagnostic time is the total of the following two components:

- Playing time: the total duration that each MD's played the sequences continuously, without actions such as jumping, scanning, or frame navigation.
- Scanning/Browsing time: the total time for browsing or frame-by-frame scanning to verify abnormal regions.

Thus, the main difference between this method and other methods (Medical Advisory Secretariat, 2003; Hadathi et al., 2005; Keuchel et al., 2006) is that the reading time details are inspected by two separate components, and this helps one to better understand not only the time for viewing a sequence but also the time used seeking abnormal regions.

3.2 Matching abnormal regions captured

In the experiment, the medical doctors were asked to capture abnormal regions independently. Then, knowing the degree

to which the abnormal region capture precised were accurate and complete would allow the performance of the method to be evaluated, a routine for checking the relevant findings was therefore implemented after the evaluations of the medical doctors.

Previous studies (Medical Advisory Secretariat, 2003; Hadathi et al., 2005) showed only the total diagnostic time without information regarding the verification of abnormal regions detected. To check abnormal regions, Keuchel et al. (2006) compared the results of two medical doctors, and in cases of discrepancies a third gastroenterologist made the final decision.

In our research, the regions captured by MD. C, who has much experience at examining endoscopic images, were used as the true ground data. The results of the other doctors were compared with this data. Relevant matching was observed and checked by MD. C. As illustrated in Figure 2, showing a part of an example sequence, the positions of the captured frame are marked on a bar code. The top row shows the results captured by MD. C, and the bottom row shows those captured by the others. The matching data were marked using the results of both paper print outs and the GUI system.

3.3 Analyzing the manipulations of medical doctors in the diagnostic procedure

Besides analyzing the average diagnostic times and rates of the relevant findings, one additional aspect of the proposed method was considered in order to verify whether the method can be comfortably used by medical doctors of a variety of skill levels. The activities of the medical doctors during the

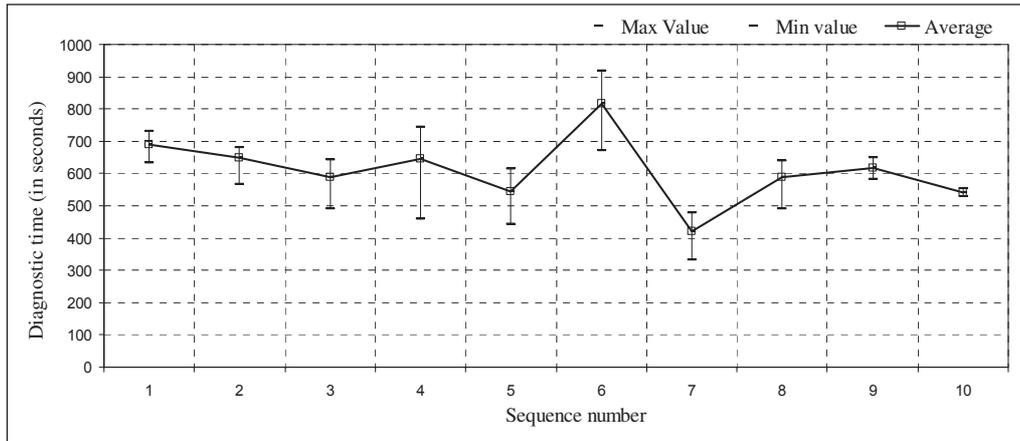


Fig. 3. Result for the analyzing of diagnostic time by sequences.

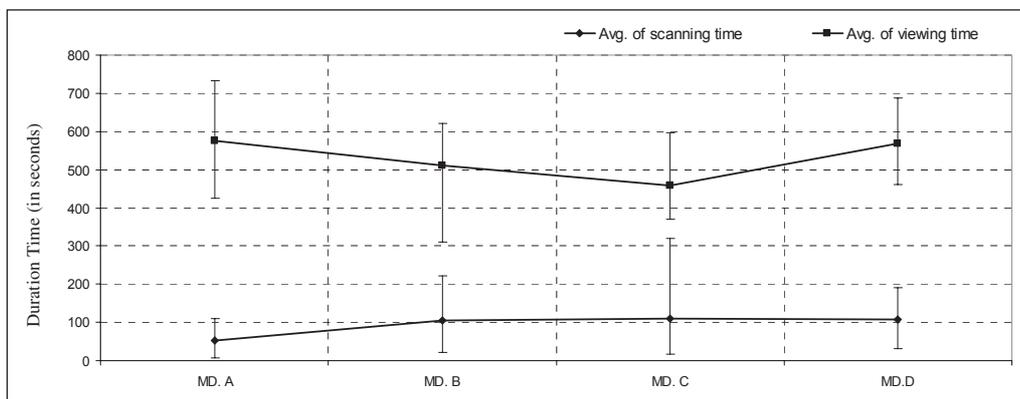


Fig. 4. Result for the analyzing diagnostic time by medical doctors.

diagnostic procedure were inspected, focusing on the actions related to different skill levels.

Comments made by the medical doctors and a review by Douglas G. A. et al. (2003) also show that the presentation of capsule endoscopic images presents some challenges with regard to the interpretation of the findings because the doctors must learn to visualize the bowel in a new way. Therefore, multi-skill level is supported to adapt to the different skill levels of physicians. The experimental system was set up 7 levels, with assume that level 1 supports junior skill and level 7 is for medical doctors with senior skill in diagnostic endoscopic images.

The actions that most closely demonstrated different skill levels were analyzed and inspected, and the approximate skill level of each medical doctor was evaluated. Moreover, such data related to the skill levels, i.e., the number of actions performed in a diagnostic session, also indicate whether the methods used to calculate the delay time are adaptive and if they require any actions for the adjustment of speed.

3.4 Statistical analysis

All measurements of diagnostic time in this study were recorded as mean ± standard deviation, rounded off to 0.5, in seconds; with the results of the counting actions given in integer values.

4. Results

4.1 Average diagnostic times

As explained in section 3.1, the diagnostic time includes the viewing time and the scanning time required for finding suspicious regions. The results of the data analysis showed that the average viewing time was 8:48 ± 1:06 min/sequence while the average scanning time was 1:20 ± 0:45 min/sequence. The mean ratio of viewing time/scanning time was 7.15 ± 3.3, which implies that the variations between the viewing and scanning time sequences were quite large. Figure 3 shows the average diagnostic time for each sequence, with the mean value being approximately 10:08 ± 1:44 min/sequence.

Figure 4 shows the average value of time consumed for each medical doctor. The data in this figure imply that medical doctors may treat the evaluation of the sequences differently. For example, MD. C used the least time for viewing, worked with a high skill level, and spend more time inspecting abnormal regions, which is shown by the scanning time value. Meanwhile, MD. A had the highest view time due to playing the sequences longer at the low levels and lowest scanning time, including using actions such as start/stop many times instead of scanning for suspicious regions. This data suggests that while MD. C was used to being presented with capsule endoscopic images, MD. A had some challenges to understanding the images.

No*	MD.C	MD.A		MD.B		MD.D	
		Found	Lost	Found	Lost	Found	Lost
1	8	7	1	6	2	5	3
2	4	3	1	3	1	3	1
3	0	0	0	0	0	0	0
4	11	6	5	4	7	6	5
5	3	3	0	2	1	3	0
6	5	3	2	4	1	3	2
7	2	0	2	1	1	-	-
8	1	1	0	1	0	1	0
9	6	5	1	5	1	-	-
10	2	1	1	1	1	-	-
Σ	42	29	13	27	15	21	11
% match		69		64		66	
Σ abnormal		127		101		94	

Table 1. Result for the verification abnormal regions captured

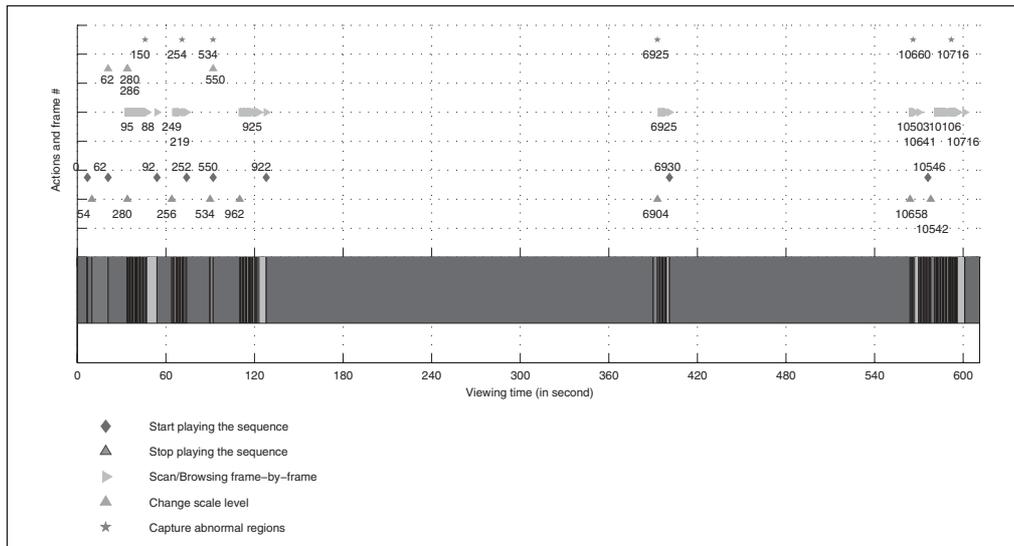


Fig. 5. An example showing the manipulations of medical doctors during a diagnostic procedure.

The resulting average diagnostic time of the extracted sequence with a length of 1.5h implies that when using the proposed method the time consumed is around $10 \pm 1.5\%$ of the length of the sequence. The predicted value, therefore, when applying this method to a full sequence, will bring the average of the total diagnostic time to around 39.2 ± 8.5 min.

4.2 Percentage of matching captured abnormal regions

Using the method for checking abnormal regions discussed in section 3.2, details of the results of the relevant findings are shown in Table 1. The numbers of matching regions are shown in the “Found” column, while numbers in the “Lost” column refer to no relevant findings.

The total number of abnormal regions captured by MD. C was 42 regions. The numbers of abnormalities present differed with each sequence. For some sequences, such as sequences 1, 2, 5, 6, and 9, there were from 3 to 8 abnormal

regions, and thus the rate of matching in these sequences was high. Sequence 4, however, included 11 regions, and as it was the sequence with the maximum number of abnormalities present it had a lower rate of matching for three of the MDs. Overall, the average matching rate of the abnormal regions was 69% for MD. A, 64% for MD. B, and 66% for MD. D. These results imply that although finding suspicious regions depends on others factors, such as one’s personal judgment and skills, the concentration of the physicians as well as the number of abnormalities present, the proposed method produces acceptable rates of capture of relevant findings.

4.3 Result of analyzing the activities of medical doctors

Figure 5 shows the detailed results of the analysis of the actions of the medical doctors. Their normal activities include: start/stop viewing, changing the skill level, scan/browsing frame-by-frame, and the capturing of abnormal regions.

This figure also illustrates the interpretation capacity of a medical doctor with a type of images. For example, at the beginning of a sequence (from frame 0 to 1,000), the actions of stop/start is repeated sometimes and followed by scan/browsing actions. There are 3 regions to be captured in this part, which correspond to the scan/browsing regions, and the same meaning with actions in the last section (from frame 10658 to the end) include: 2 frames captured among 3 regions while browsing. These actions imply that the proposed method is adequate to allow medical doctors to capture abnormal regions at full capacity.

To evaluate suitable delay times as well as degrees of comfort of the method, the actions of changing the skill level were analyzed. An examination of the activity data revealed that before selecting the most suitable level, the doctors usually tried to adjust the skill level at the beginning of a viewing sequence. After removing such actions at the starting points of the evaluations from the data, all subsequent actions for changing the skill level were counted.

The results of the data analysis showed that the level preferred by each doctor was different. MD. A usually used levels 3 and 4; MD. B and MD. D selected levels 4 and 5; and MD. C viewed the sequences at levels 5 and 6. Therefore, the data imply that having a "multi-skill level" ability allows medical doctors to adjust the method to fit their respective degrees of expertise.

The results the counting of skill-level changing actions in each evaluation were as follows: the percentage of sessions without any skill-changing actions was 50 %, that for sessions with 1–2 actions was 34 %, that for sessions with 3 to 5 actions was 6 %, and that for sessions with 5 or more actions was 3 %. These data show that a high number of evaluations do not require many speed adjustment manipulations, and that therefore the current method for calculating the delay time is suitable for such diagnostic procedures.

5. Discussion and conclusions

This study investigated the performance and effectiveness of a method for controlling adaptively the display rate of images in order to reduce the diagnostic time of video capsule endoscopy. The results of our experiments provided justification for the proposed method, indicating its value in assisting medical doctors in the following aspects. First, diagnostic time was approximately 10 min for a 1.5-hour sequence, i. e., the rate of reduced diagnostic time was $10 \pm 1.5\%$ of the

length of the captured sequence. Moreover, the proposed method allowed the physicians to find suspicious regions in full capacity as well as allowing them to keep required manipulations by the medical doctors during the examination procedure to a minimum.

Future work will consider positions on a sequence which candidates capture as abnormal regions. Because there are a huge number of normal frames, which is clearly understandable, these data can hopefully be further utilized based on artificial neural networks to classify abnormal regions. Such a tool could automatically determine which regions on sequences are suspicious, thereby further assist medical doctors in their navigation of suspicion sector and lead to reducing the diagnostic time.

The scheme of multi-skill level introduced in the proposed method also suggests the capacity to deploy other functions to assist medical doctors. For example, a learning system could be applied based on analyzing the factors that influence the changing of skill levels with reference to specific data. The results could then be used in the application of similar data to the automatic adjustment of skill levels if the abnormal region data is new to a different medical doctor. The impact of this work can be used to significantly assist the detection of disease.

References

- Adler, D. G., Gostout, C. J. (2003). Wireless Capsule Endoscopy – State of Art. *Hospital Physician*, pp. 14–22.
- ASGE, Technology Status Evaluation Report (2002). Wireless capsule endoscopy. *J. Gastrointestinal Endoscopy*, **56**, No. 5.
- Hadathi, M., Heine, G. D. W. *et al.* (2005). A Prospective study comparing Video Capsule Endoscope followed by Double Balloon Enteroscopy for suspected small bowel disease. *Proc. of Intl. Conf. Capsule Endoscope 2005 Abstract*, pp. 203.
- Iddan, G., Meron, G., Glukovsky, A. *et al.* (2000). Wireless Capsule Endoscope. *Nature* **405**, 417.
- Keuchel, M., Al-Harhi, S. *et al.* (2006). New Automatic mode of Rapid 4 Software reduces reading time for small bowel pillcam studies. *Proc. of Intl. Conf. Capsule Endoscope 2006 Abstract*, pp. 93.
- Medical Advisory Secretariat (2003). Wireless Capsule Endoscopy, Health Technology Literature Review. *Ontario Ministry of Health and Long-term Care*, Canada.
- Swain, P., Fritscher-Ravens, A. *et al.* (2004). Role of video endoscopy in managing small bowel disease. *J.GUT*, **53**, 1866–75.
- Vu, H., Echigo, T. *et al.* (2006). Adaptive Control of Video Display for Diagnostic Assistance by Analysis of Capsule Endoscopic Images. *Proc. of 18th International Conference on Pattern Recognition (Accepted)*.