Proceedings of the 2008 IEEE
International Conference on Robotics and Biomimetics
Bangkok, Thailand, February 21 - 26, 2009

Evaluating the Control of the Adaptive Display Rate for Video Capsule Endoscopy Diagnosis

Hai Vu, Ryusuke Sagawa, Yasushi Yagi
The Institute of Scientific and Industrial Research, Osaka University
{*vhai, sagawa, yagi*}*@am.sanken.osaka-u.ac.jp*

Tomio Echigo
Osaka Electro-Communication University
echigo@isc.osakac.ac.jp

Masatsugu Shiba, Kazuhide Higuchi, Tetsuo Arakawa
Graduate School of Medicine, Osaka City University
{*shiba, khiguchi, arakawat*}*@med.osaka-cu.ac.jp*

Keiko Yagi
Kobe Pharmaceutical University
k-yagi@kobepharma-u.ac.jp

***Abstract*— The excessively long reviewing times for diagnosis of capsule endoscopy present a clinical problem. A new method, "adaptive speed", which automatically controls display rates of the video capsule endoscopy images, was proposed to address the problem. In this paper, we investigate the effectiveness of this method versus a standard-view using the existing system. The main activities of examining doctors during a series of evaluations using both systems are recorded. For comparisons, logged actions are analyzed to show three criteria: 1. Diagnostic time, 2. Ability to capture abnormal regions, and 3. Operability for the examining doctors. We conclude that adaptive speed reduces examination time by ten minutes from that of the existing system, while the number of abnormalities found are similar. As well, examining doctors need less effort because of the systems efficient operability.**

***Index Terms*— Adaptive video rate control, Capsule endoscopy, Diagnostic time, Logged actions based analysis.**

I. Introduction

Capsule Endoscopy (CE) [1], [2] utilizes a swallowable endoscopic device that is propelled by peristalsis through the GastroIntestinal (GI) tract. In a typical examination, the capsule takes approximately 7 - 8 hours to go through the GI tract for acquisition of images at a rate of two frames per second. The sequence thus has around 57,000 images that can be used for diagnoses. With such a large number of images, review and interpretation of video capsule endoscopy can be time consuming and present a heavy time load for physicians [3].

To reduce diagnostic time, some viewing modes are provided in the RAPID Reader [4], a CE annotation software developed by the capsule manufacturer. For example, with *dual-view*, two consecutive frames are simultaneously displayed; *quad-view* reshapes four consecutive images into one. *Automatic-view* combines successive similar images to display representative frames; *quick-view* mode allows a fast preview by showing only highlight images. Mitigating against reducing diagnostic time by using these techniques is that some clinical images, including abnormalities, may only be seen in a single or just a few frames [3]. These are not easily identifiable in the *quad-view* mode because images are distorted and may not even be seen if that image is skipped. Different from these techniques, in [5] we proposed a new method, named as adaptive speed, for automatically controlling the display rate of the CE sequence. Adaptive speed utilizes image features extractions and classification techniques to calculate delay time between consecutive frames. The main advantages are that diagnostic time can be reduced while images are displayed in their original form without skipping any. However, the preliminary validations in [5] included only ninety-minutes sequences and the method did not undergo extensive clinical trails.

Thus, in this paper, we investigate the effectiveness of the adaptive speed method versus the standard viewing (*dual-view* + *automatic-view*) of the RAPID Reader through clinical evaluations of examining doctors. Forty-eight evaluations were conducted of four doctors who examined six full sequences using both systems. The procedure for the evaluations was prepared so that operations on the systems were as similar as possible. The main activities of the examining doctors during the evaluations were recorded. Clinical validations were presented with three criteria utilizing analysis based on the logged actions. The results showed that average diagnostic time using the adaptive speed system was 32.5 ± 7 minutes, this time is ten minutes less than evaluations implemented using the RAPID Reader. The adaptive speed system also required less effort, while the number of abnormalities found under both systems was similar. These results should convince doctors that they can safely use this approach and still obtain reduced diagnostic times by utilizing techniques of computer vision.

II. Adaptive Speed Technique

The adaptive speed technique is approached from observations that the states of the image acquisitions depend on the motility patterns in the GI tract. A CE sequence can be displayed at high speed during a stationary state to save time, and this speed is then decreased during rough changing

states to enable easier viewing images. A framework for the technique was proposed in [5] with a series of steps. First, cues information for disparity between consecutive frames, including color similarity and motion displacements, are extracted. Then a decision tree utilizes these features to classify the states of the image acquisitions. For each classified state, parametric functions calculate the delay time between adjacent frames with a constraint so that speed changes between the states do not occur abruptly. For details of these steps, please see [5]. An example of distribution of the delay time calculated by the adaptive speed technique of a sequence is plotted in Fig. 1. The delay time spreads in range from 30 ms/frame to 150 ms/frame, corresponding with the disparity of images varying from stationary to suddenly changing.

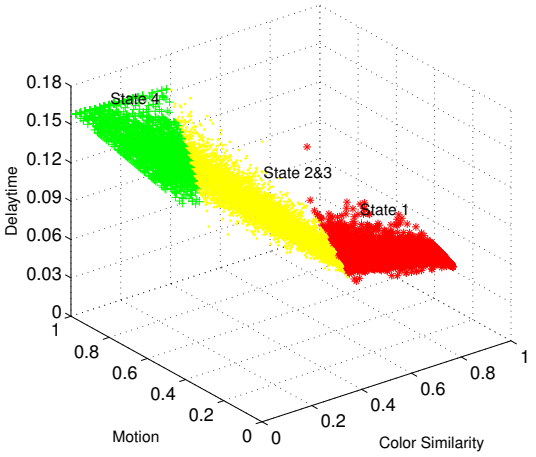

Fig. 1: Distribution of the delay time calculated from the motion displacement and similarity features of a sequence.

To demonstrate the effectiveness of the adaptive speed technique, two cases when physicians examine the sequence under this mode are discussed in Fig. 2. Corresponding with the variations of delay time, display rate (in the *dual-view* mode) of this sequence spreads in range from a minimum speed of 12 fps to a maximum one of 60 fps (Fig. 2(b)). For frames at position [A] in Fig. 2(b), their delay time and some representative frames are shown in detail in Fig. 2(c). Compared with playing the sequence at a constant frame rate (assumed as 13 fps), the images are displayed for twice the constant value (160 ms, compared with 77 ms). With a longer delay time, the frames under Fig. 2(c) are clearer for analysis of examining doctors. Contrarily, the frame rates at position [B] are increased. The values of the delay times and typical frames at this position are shown in Fig. 2(a). Obviously, the frames are similar. The delay time in this case is smaller than four times if the sequence is played at a fixed speed (around 20 ms, compared with 77 ms).

The demonstrations in Fig. 2 show that the two-fold effectiveness of the proposed method is more convenient and requires less attention by examining doctors. However, to confirm performance and present more convincing validated solutions, the adaptive speed technique needed to undergo clinical trials. These are to address the issues of the subjectivity of reducing diagnostic times such as how differences in capture rate, and in sensitivity for abnormal regions when a same video CE is observed using the existing system. Therefore, a series of clinical evaluations to compare two systems was prepared as described below.

III. Conducting Evaluations

To ensure that the conditions for the evaluations of both systems were as similar as possible, a GUI application (named as *P system*) was developed for the adaptive speed technique so that normal diagnostic functions were available; such as the capture of abnormal regions, the manual adjustment of viewing speeds, changes in viewing display, as well as functions for navigating and verifying suspicious regions. RAPID Reader application Version 4 (the *G system*) is downloadable at [4]. Both systems were installed on a same PC having a Pentium IV 3.2 GHz, 1GB RAM.

We prepared six full sequences of patient data. The evaluations were implemented on both systems by four physicians (named as *MD. A*, *MD. B*, *MD. C* and *MD. D*) at the Graduate School of Medicine, Osaka City University, Japan. Thus, forty-eight evaluations were conducted. For unbiased evaluations of the examining doctors, the order of the evaluations of a certain sequence were established so that the number of anterior/first evaluations on each system was equal. The examining doctors were asked to independently find and capture suspicious regions.

The main activities of the examining doctors were recorded under the two systems. These included: $[play \rightarrow stop]$, $browsing/scanning$ frames to examine suspicious regions, $jumping$ frames, *changing manually display speed* and $capturing$ abnormal regions. The *P system* is programmed to record logs of the activities of the examining doctors to a database. To monitor their actions when using the *G system*, we develop a utility that captures the screen when the computer mouse is activated. Interpretation of these logs was implemented by manually reading the captured images. Fig. 3 shows an example of the logged activities of the examining doctors *MD. A*, *MD. B* and *MD. C* for Seq_3 under the two systems. Obviously, the logs expressed in this figure allow investigating clinical issues such as diagnostic time, abnormal regions captured and system operability.

IV. Logged actions Based Analysis

A. Diagnostic time

The examining doctors were asked to fill in evaluation forms when they started and finished an evaluation. Di-

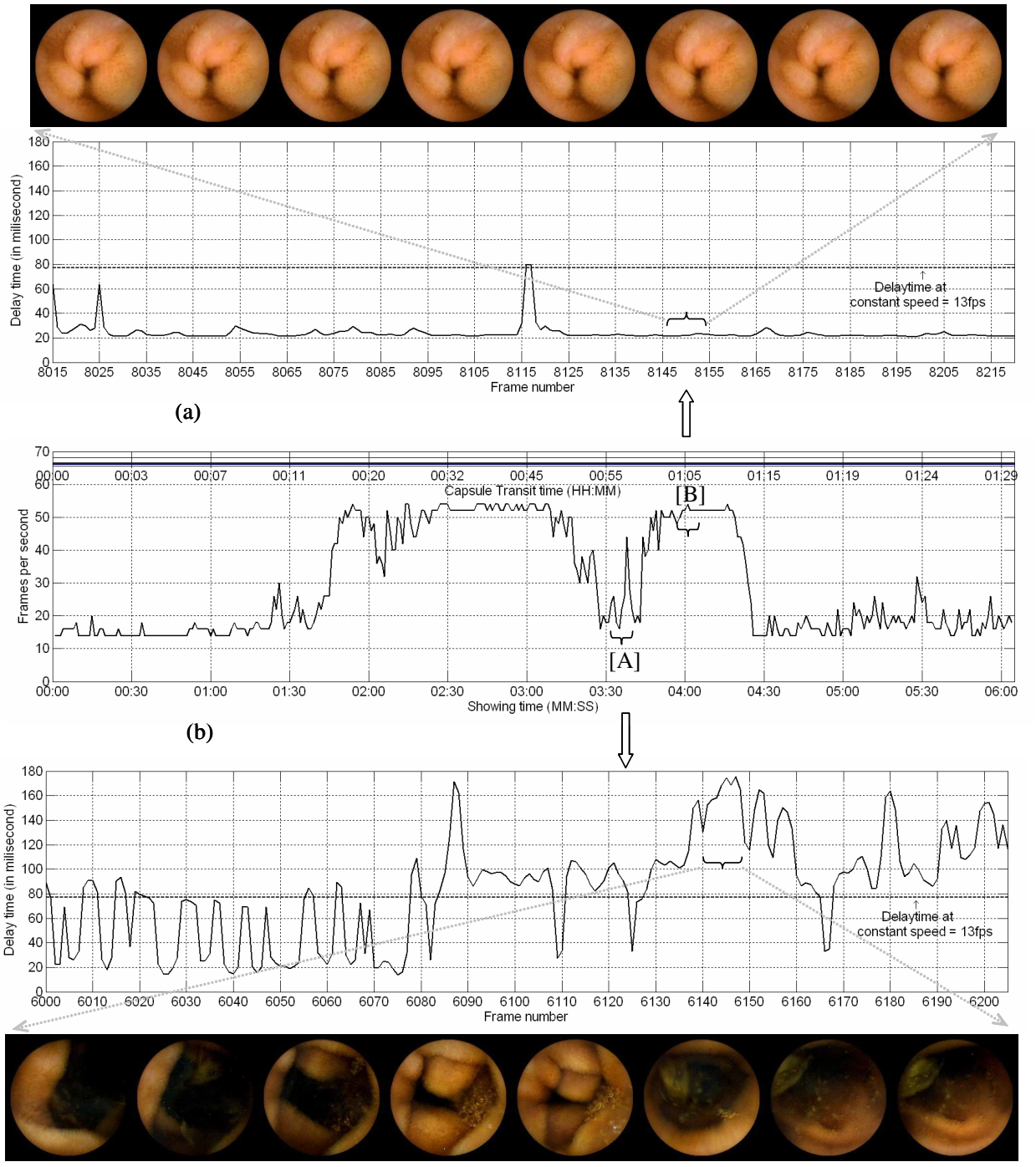

Fig. 2: (b) Varying frame rates of an example sequence. (a) Delay time at positions [B], the sequence play at high speed with some continuous frames are displayed by its upper row. (c) Delay time at [A], the sequence plays at high speed with some continuous frames are displayed by its lower row.

agnostic times were calculated from this data. The durations of activities such as continuously $[play \rightarrow stop]$, $browsing/scanning$ frames, and $jumping$ frames were summated by investigating the captured logs under both

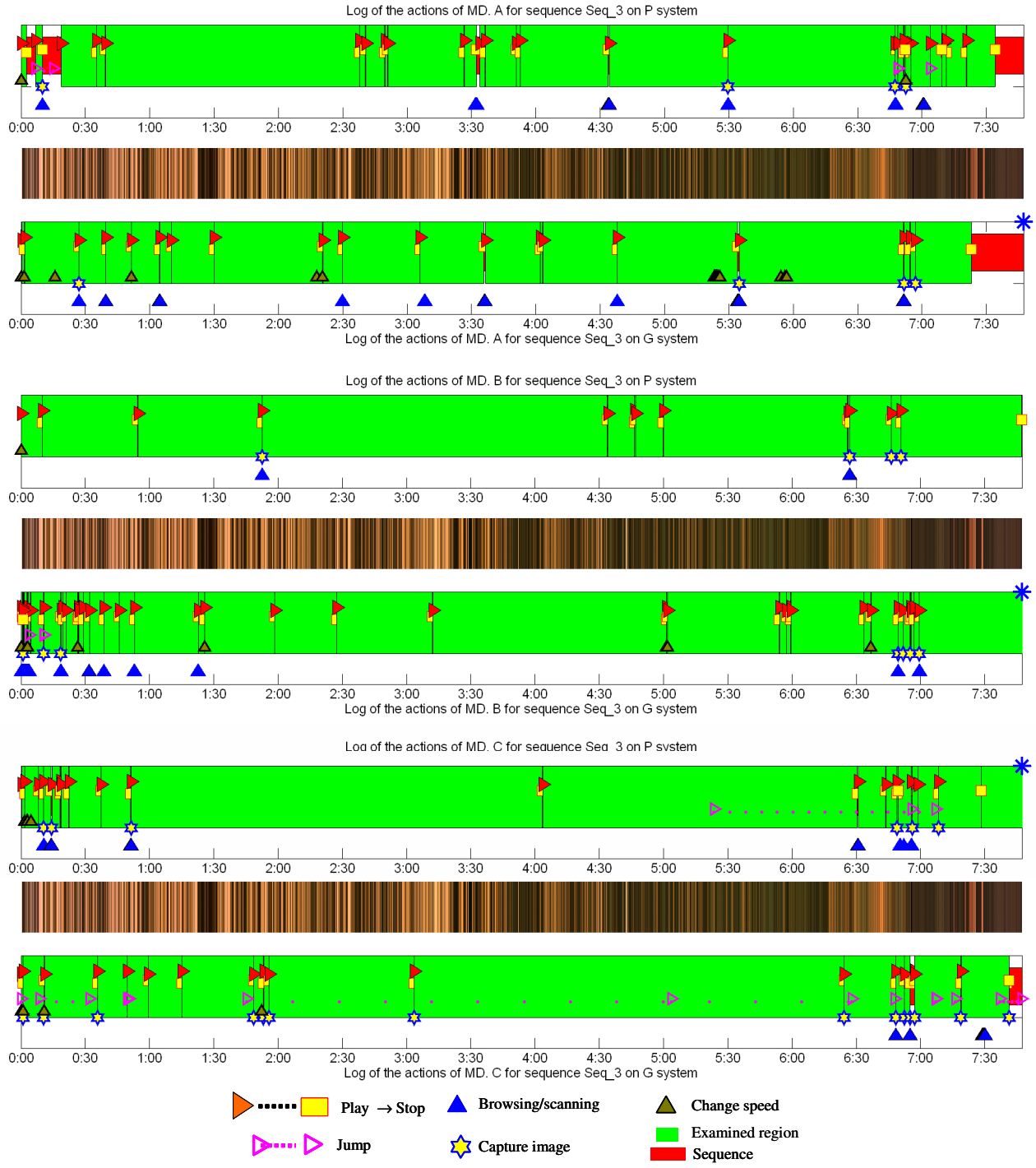

Fig. 3: Logged actions of the examining doctors for same CE sequence. For each evaluation, the upper panel shows activities under the *P system*, the lower panel shows operations under *G system*. The middle bar is the average colors of the images sequence.

systems. These data confirmed the diagnostic times noted by the examining doctors.

Figure 4 compares the diagnostic times of the examining doctors for each sequence. The first evaluation on the cor-

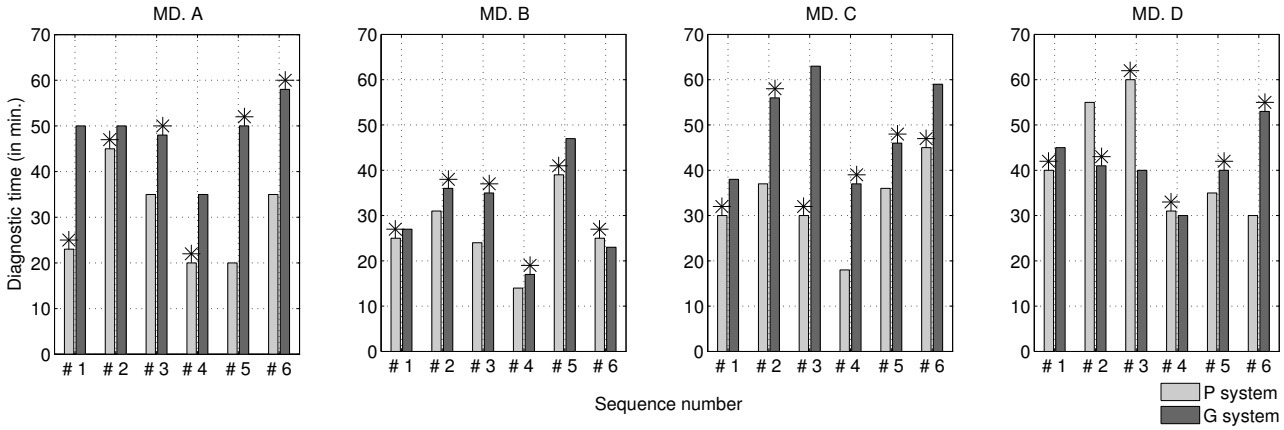

Fig. 4: Diagnostic times of the examining doctors under the two systems. Asterisks mark the first evaluation of the corresponding sequence.

responding system for a certain sequence is marked with an asterisk in these figures. The diagnostic time using the *P system* was significantly reduced from that of the *G system* for most evaluations (approximately 16 min. for *MD. A*, 6 min. for *MD. B*, and 14 min. for *MD. C*). The diagnostic time of *MD. D* was equal under both systems.

Average diagnostic times by sequence are shown in Fig. 5. From this figure, the diagnostic time for the *P system* was reduced from that of the *G system* for all six sequences. The average diagnostic time for the *P system* was 32.5±7 min. and 42.4 ± 9 min. for the *G system*. Applying the T-test to measuring the significance of any difference of the average values, we found that the diagnostic times using the *P system* showed a significant difference from evaluations implemented using the *G system*($t = 3.1, df = 47, p < 0.05$).

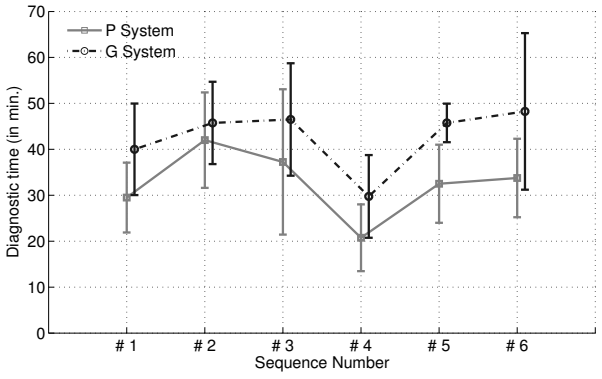

Fig. 5: Average diagnostic time by sequences

B. Ability to capture abnormal regions

The number of abnormalities present in any evaluation differed according to the doctor because each depend on factors such as the examining doctor's personal judgment, skill level and their concentration on the evaluation. Therefore, we took into account abnormal regions captured by the same doctor using the two systems. First, the abnormal regions χ of a sequence were considered by merging abnormal regions captured with both systems. The matching rate was the ratio between abnormal regions υ captured in a particular system and the χ abnormal regions, as below:

$$MatchingRate = \frac{\upsilon}{\chi}100(\%) \quad (1)$$

Table I shows the ratio of the evaluations by the examining doctors using both systems. The average value was 91% for the *P system*, approximating the matching rate on the *G system* (92%). The results implied that the proposed system had no limitations for capturing abnormal regions when the display frame rates were controlled under the adaptive speed technique.

C. Operability of the examining doctors

Operability criterion implies how a doctor can sensitively find abnormal regions. A suggestion for such assessment is that we can consider events that are *a stop* when continuously *playing* sequences, as being ones to verify or look for a suspicious region. Therefore, a ratio ($CaptureRate$) can measure this sensitivity by κ = the number of total $[play \rightarrow stop]$ actions and τ = the number of $capturing$ frames, as defined below:

$$CaptureRate = \frac{\tau}{\kappa}100(\%) \quad (2)$$

Table II shows the $CaptureRate$ for all of the evaluations; the $CaptureRate$ of three of the four examining doctors using the *P system* was clearly higher than when using the *G*

TABLE I: The *MatchingRate* values of evaluations on both systems (numerator is υ, denominator is χ of (1))

Seq. No	*MD. A*		*MD. B*		*MD. C*		*MD. D*	
	P system	*G system*	*P system*	*G system*	*P system*	*G system*	*P system*	*G system*
# 1	2/3	3/3	3/3	2/3	2/2	2/2	2/4	3/4
# 2	3/3	3/3	4/5	5/5	4/5	5/5	5/6	5/6
# 3	4/4	4/4	3/4	3/4	5/6	6/6	7/7	7/7
# 4	2/3	2/3	1/2	2/2	3/3	3/3	5/5	4/5
# 5	5/5	4/5	3/4	3/4	5/5	4/5	5/6	6/6
# 6	5/5	5/5	6/6	6/6	8/8	8/8	2/2	2/2
Σ Reg. lost	1/23	2/23	4/24	3/24	2/29	1/29	4/30	3/30
$Avg.$	96%	91%	88%	92%	93%	96%	86%	90%
Average of *P system* = 91% and *G system* = 92%								

TABLE II: The *CaptureRate* values of evaluations on both systems (numerator is τ , denominator is κ of (2))

Seq. No	*MD. A*		*MD. B*		*MD. C*		*MD. D*	
	P system	*G system*	*P system*	*G system*	*P system*	*G system*	*P system*	*G system*
# 1	2/15	3/21	5/51	2/30	4/10	5/23	2/6	7/20
# 2	3/29	4/30	5/25	5/25	8/17	11/43	6/11	7/29
# 3	4/26	4/21	4/12	7/37	6/16	14/18	9/38	15/48
# 4	3/18	3/30	4/19	7/13	5/12	7/22	7/14	5/22
# 5	7/24	5/20	8/46	5/40	6/30	5/15	6/33	6/17
# 6	5/13	6/30	11/34	12/15	32/61	38/54	4/9	5/33
$Avg.$	21%	15%	24%	31%	44%	40%	32%	27%
Average of *P system* = 30% and *G system* = 28%								

Note is that τ is total frames captured. Because some captured frames are the same types from an abnormal region. This value is usually larger than the number of abnormal regions in Table I for each corresponding evaluation.

system. As shown, automatically adjustments of the display speed using the proposed technique are reasonable and it provides higher operability for finding abnormal regions.

V. Conclusion

This paper presented a method for evaluating the effectiveness of an adaptive speed system versus the existing standard-view system for video CE annotation. The evaluations focused on clinical validations. A method of analysis based on logged actions was proposed to address clinical problems such as diagnostic time, abnormal capture rate and operability. Analysis of the results confirmed that the evaluations under the adaptive speed system had higher operability and lowered the diagnostic time by ten minutes, while number of abnormalities found was similar under both systems. These results should convince doctors that they can safely use the adaptive speed technique for routine clinical diagnoses.

Using the logged actions of examining doctors, studies such as for regions with abnormalities direct to clinical applications or educational purposes. Skill levels of examining doctors also can be automatically evaluated or adjusted to their expertise. These works allow to effectively and quickly navigate interesting parts of a sequence and therefore they are suggested for future research.

References

[1] G. Iddan, G. Meron, A. Glukovsky, and P. Swain, "Wireless capsule endoscope," *Nature*, vol. 405, p. 417, 2000.

[2] American Society for Gastrointestinal Endoscopy - ASGE, "Technology status evaluation report wireless capsule endoscopy," *Gastrointestinal Endoscopy*, vol. 56, no. 5, pp. 1866–1875, Aug. 2002.

[3] P. Swain and A. Fritscher-Ravens, "Role of video endoscopy in managing small bowel disease," *GUT*, vol. 53, pp. 1866–1875, 2004.

[4] Given Imaging's Website (Oct, 2007), *http://www.givenimaging.com/en-us/HealthCareProfessionals/Products/Pages/Software.aspx*.

[5] H. Vu, T. Echigo, R. Sagawa, M. Shiba, K. Higuchi, T. Arakawa, and Y. Yagi, "Adaptive control of video display for diagnostic assistance by analysis of capsule endoscopic images," in *Proc. of the 18th Int. Conf. on Pattern Recognition*, vol. 3, Aug. 2006, pp. 980–983.