33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

A Novel Method for Informative Frame Selection in Wireless Capsule Endoscopy Video

Yichen Fan, Max Q.-H. Meng, Baopu Li

Abstract—**Wireless capsule endoscopy (WCE) has been validated to be an important tool in the evaluation of gastrointestinal (GI) tract. Compared with traditional endoscope technologies, its non-invasiveness property meets with great favor of patients. However, from physician's point of view, WCE video suffers from low resolution, limited illumination, irregular movement, more importantly, imbalanced rate of abnormality which brings a lot of challenges for diagnosis. These challenges motivate us to devise an approach to guide the physicians to focus on the informative frames which could be convenient for review of the content of the GI tract. This paper presents a novel approach for automatic selection of the WCE video frames with lumen and coherent motility. We adopt lumen detection based on mean shift to provide robust and reliable selection of lumen. Together with the evaluation of coherent motility, we can provide a full and fast approach to select the informative frames for diagnosis. The experiments on real date are presented to show the performance of our proposed method.**

I. Introduction

Wireless capsule endoscopy has been gradually employed for examining the gastrointestinal (GI) tract [1] of a patient. A WCE measuring 11*26 mm is equipped with LEDs for illumination and an imaging sensor. After a patient swallows a WCE, it will capture images in a rate of two frames per second and simultaneously transmit them to a receiver wear on the patient's belt as it is traveling through the whole GI tract. Compared to the traditional endoscopy (TE) such as colonoscopy and virtual endoscopy, such a process is non-invasive and radiation-free. Additionally, WCE can reach the small intestine, where TEs are unable to reach. WCE has been demonstrated to be a significant tool in the diagnosis of bleeding, Crohn's disease, and Celiac disease in GI tract. New technologies aiming at helping physician to locate the position of a WCE has been developed with magnetic localization system [2].

Although WCE has a lot of advantages over the TEs, there are some shortcomings concerning the video date. First, over 50,000 images are taken for an exam [3] and physicians need to spend approximately 90 minutes on reviewing the whole WCE video. In addition, the examination process requires full concentration with proper training. Secondly, the field of view could be obstructed by the GI mucosa. Unlike the TE, the passive WCE moves by the peristalsis force of the GI tract. Thus the movement of the WCE cannot be controlled so as to offer any desired views. It is very challenging for physician to detect the structure abnormalities in WCE images without enough depth. Thirdly, when the WCE is passing through the contractions, the discontinuous movement may limit physicians' spatial cognition of the GI tract. Therefore, we are innovated to overcome the existing WCE challenges for facilitating diagnosis and increasing the detection accuracy especially for the structure abnormalities such as polyps in GI tract.

To date, various research efforts are concentrating on designing of computer aided diagnosis system for WCE images or video. For instance, several approaches are directing their interests on detection of WCE images with depth[5][6][7]. The authors in [6] consider the consistency appearance of lumen, the area of the largest dark blob in each frames has been extracted and evaluated by empirical selection of the threshold value. A more robust approach in [7] was proposed for automatic adjustment of the threshold. In [8], the authors extend lumen detection approach by introducing the local extreme of the image intensity by mean shift method. To better approximate the lumen boundary, region-growing method is adopted in [9]. Method in [10], detects the lumen in frequency domain based on template-matching in Fourier transform of the image. However, none of these methods consider the motion features of the WCE images. And the frames are selected individually may challenge physicians' perception and diagnosis.

The methods tackle the assessment of motility, especially the intestinal contractions in the WCE video based on textural, color [11] and blob features [12] together with machine learning approaches such as supported vector machine and neural networks. Although the result in [12] offers a promising performance, we should also pay attention to those frames with depth but not experiencing contraction are also informative frames for diagnosis.

Study on spatial visualization suggests that improving the interface appearance by reducing the number of hidden dependencies between actions can improve the performance of human's information search and information retrieval[17]; in other words, interface appearance depends on coherent motility and the explicitly of the spatial information. In this work, to achieve the goal of the structural abnormality detection which can be referred as the spatial visualization of the GI tract, we adopt lumen detection based on mean shift method to provide robust and reliable selection of lumens.

Yichen Fan, Max Meng and Baopu Li are with the department of Electronic Engineering at the Chinese University of Hong Kong, Hong Kong SAR, China, Email: {ycfan&max&bpli}@ee.cuhk.edu.hk;

Manuscript received April 1, 2010. This work was supported by SHIAE project #8115021 of the Shun Hing Institute of Advanced Engineering of the The Chinese University of Hong Kong, awarded to Max Q.-H Meng

Together with the evaluation of coherent motility, we can provide a full and fast approach to select the informative frames for diagnosis. Here, informative frame means frame which might be convenient for review of the content of the GI tract.

We organize the remainder of this paper as below. In section II, we explain the general flow of the proposed scheme, i.e., the low level color space transformation cascade with motility assessment and adaptive lumen detection. Experiments with real data are presented in section III. Section IV gives the experimental results, and we draw some conclusions and make some discussions at the end of this paper.

II. THE PROPOSED METHOD

The general flow of the proposed method is depicted in Fig.1. It comprises four steps: color space transformation, adaptive lumen detection, motility assessment and mixture weight score.

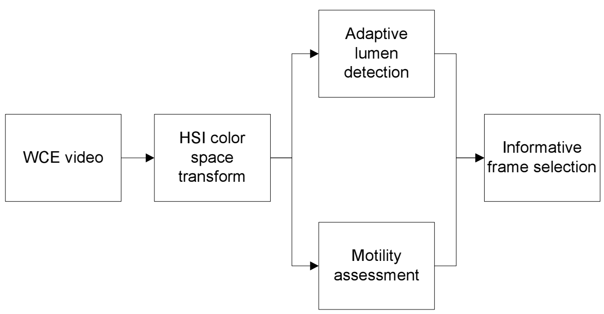

Fig.1.1st row shows the original frames, 2nd row is frames using MS

A. Adaptive Lumen Detections

As we have known, a WCE is driven by the peristalsis force of the GI wall, its movement is uncontrollable. Thus, the orientation of the camera inside it is not always focusing the central part of the lumen of the GI tract. This yields the fact that the structure abnormalities such as polyp cannot be easily detected by the physicians. The aim of this stage is to avoid the frames with insufficient lumen's areas.

The structure of the GI tract is similar to a cylindrical tube. By using the WCE lens' system, the illumination of the image plane is found to decrease away from the optical axis at least with the 4th power of the cosine of the angle of obliquity with the optical axis [13]. The illumination decays away from the optical axis can be described as the image intensity decreases. Our vision system is able to convert intensity decay in certain amount (so called lumen) to depth perception. In order to efficiently select lumen, we assume the appearances of lumen and specularity correspond to local minimums and maximums of the image intensity, respectively.

The proposed method detects the lumen and specularity by the mean shift (MS) algorithm [4], [8]. Mean Shift is a powerful and efficient non parametric iterative algorithm that has been widely adopted for clustering and segmentation. It is an efficient method to find the clusters in specified feature space. The following sections will illustrate the procedure:

1. HSI color space transformation: Since the WCE images are suffering from illumination variance, the number of bins will greatly affect the segmentation performance. HSI color space offers the most similarity to physiological perception of human visual system. It describes an image with three components: hue, saturation and intensity. Compared to the other color space models, the hue (H) and saturation (S) in HSI color space is invariant to illumination intensity and viewing orientation. We use the following equations for HSI color space representation:

$$h = \arctan \frac{\sqrt{3}(k_G - k_B)}{(k_R - k_G) + (k_R - k_B)} \quad (1)$$

(2)

$$k_c = \int_\lambda f_c(\lambda) r(\lambda) d\lambda \text{ for } c = R, G, B$$

$f_c(\lambda)$ is the channel sensor response function

$r(\lambda)$ is the surface reflectance function

Thus, we adopt the HSI color space as the color descriptor and mainly use the hue (H) value for later process.

2. Mean shift (MS) discontinuity preserving filtering: MS formulates the points in color space to a probability density function. Thus, dense regions in color space may indicate local maxima. For a given kernel which is uniform and with a bandwidth parameter h, kernel density estimator for a given set of d-dimensional points is defined by:

$$f(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \quad (3)$$

To find its maximum, its gradient ascent is equal to zero, we will have:

$$m(x) = \frac{\sum_{i=1}^{n} K'\left(\frac{x - x_i}{h}\right) x_i}{\sum_{i=1}^{n} K'\left(\frac{x - x_i}{h}\right)} - x \quad (4)$$

Where $m(x)$ is the MS. We will then move the kernel by $m(x)$ and repeat this procedure until convergence reached. The information of convergence points is recorded as z_i.

3. Determine clusters: To segment the image into clusters, all feature points associated with the same convergence point are sorted into the same cluster. Clusters with less than β pixels will be eliminated. The clusters have been labeled with not only with its area A, but also the average intensity I. We are only interested in the lumen and the specularity for each image; we empirically choose the number of clusters.

4. Clusters evaluation: The cluster with the smallest or the largest mean be chosen to represent the lumen or specularity, respectively. We use the following formula to calculate the lumen detection score:

$$S_i^L = I^2 \times \frac{A_L}{A} \tag{5}$$

Where I is the normalized average intensity, A_L and is the lumen area.

Fig.2. shows the frames segmented by the MS algorithm. The lumen and specularity can be easily adopted for calculation of lumen detection score.

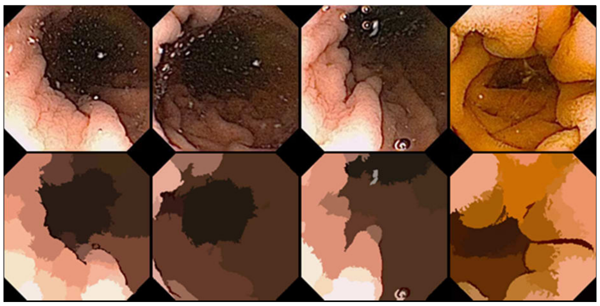

Fig.2.1st row shows the original frames, 2nd row is frames using MS

B. *Motility Assessment*

As we have mentioned, interface appearance depends not only on explicitly of spatial information, but also coherent motility. To evaluate the coherent motility, we calculate the motion vector between consecutive frames and among consecutive sequences to avoid the irregular motion of the WCE. We use the block matching algorithm (BMA) [15] to extract and evaluate the motion vectors. BMA estimate the amount of motion on a block by block basis based on certain criterion [14][15][16].

Instead of giving an exhaustive search, it is reasonable to make use of the fact that the WCE motion in frame basis is usually coherent, thus we use ARPS [14] algorithm to simplify the prediction of the motion vector with deduction of the size of search pattern. The choice of the region of support and algorithm for computing the predicted motion vector will affect the computation efficiency and accuracy of the result. [14]

There are lots of cost functions to evaluate the motion vectors, such as mean squared error (MSE), or peak signal-to-noise ratio (PSNR) and mean absolute difference (MAD). For the simplicity and effectiveness, we use MAD as the cost function to find the motion vector difference between consecutive frames:

$$MAD = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|C_{ij} - P_{ij}\right| \tag{5}$$

C. *Weighted score function*

Combined the lumen detection, motility assessment, we can finally find the frames with positive score:

$$P_i = w_1 S_i^L + w_2 S_i^M \tag{6}$$

w_j is the weight of stages,

$S_i^C = \frac{1}{1+e^{-S_i^C}}$, $S_i^M = \frac{1}{1+e^{S_i^M}}$ are the normalized score of Lumen detection and motility assessment, respectively.

III. Experiment

In order to evaluate the performance of the proposed method, we apply the proposed method to five real patient videos taken by the Olympus WCE. The video consist 500 frames (part of video is shown in Fig.3) and of approximately 250 seconds in duration. The resolution of the WCE images is 288*288. The informative frames are manually selected as ground truth.

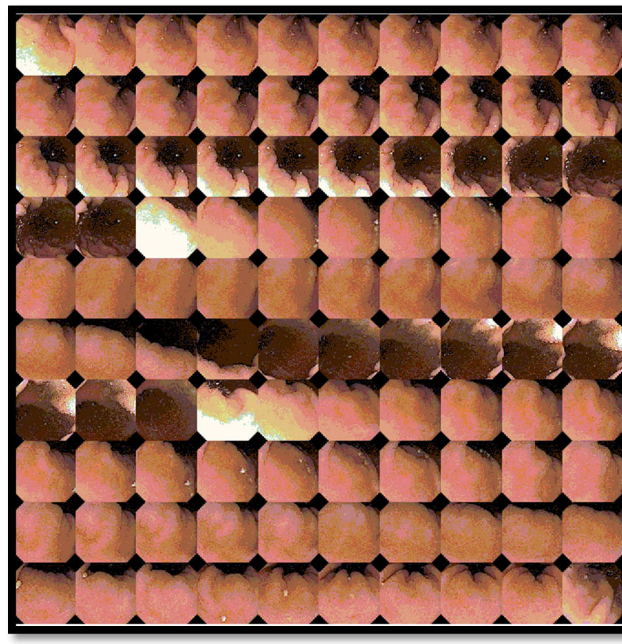

Fig.3.WCE image sequences

In the experiments for Section II.A, the spatial bandwidth is set to 7. The lumen detection score is normalized by a factor of $1/1000$. In order to eliminate the distortion given by overexpose, we set an area constrain by $S_i^L = 0$ if $A_S / A > 0.5$. In the experiments for Section II.B, we empirically choose the value for search parameter of size 5 and the Marco Block of size 12. The score of lumen detection and motility assessment is shown in Fig .4. We normalize the scores to make them share the same monotonic property and be in the same range. For score of LD, it is proportional to the area of the lumen and the normalized intensity. As expected, higher score leads to higher possibility of lumens. For score normalization of MA, we modify the logistic function to reverse monotonicity. We can tell from the result: as the MA score increase, the more coherent motility we have. In fig.5, with specularity area constrain given above, overexposed frame were filtered. In the experiments for Section II.C, The weights w_i can affect the final selection. If $w_1 > w_2$, the result is more suitable for lumen detection. Otherwise the motility assessment takes leading role. Fig.5. demonstrates the comparison for different weight. The informative frames is selected with the positive score that is larger than 0.6. We choose this value as threshold empirically. The performance is evaluated in terms of sensitivity, specificity and precision. The result is demonstrated in Table I.

The performance is evaluated in terms of sensitivity, specificity, precision and FAR. The result is demonstrated in Chart I. According to the detection statistical evaluation, the proposed method yielded an overall sensitivity of 76.41%, specificity of 87.5% and precision of 84.51%. The false alarm ratio (FAR) is 57.89% which is used to indicate the ability of system to avoid false positives. Considering the challenges mentioned before, the experiment evaluations demonstrated that the proposed method can achieve promising results.

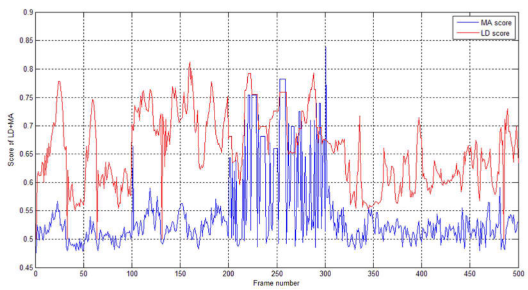

Fig.4.LD score and MA score calculated within 500 frames

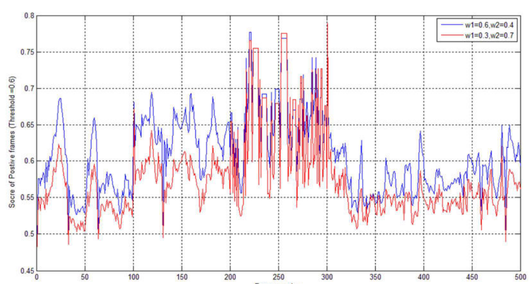

Fig.5. Frame selection based on different weight functions

TABLE I
OVERALL PERFORMANCE OF THE PROPOSED METHOD

Sensitivity	Specificity	Precision	FAR
TP/(TP+FN)	TN/(FP+TN)	TP/(TP+FN)	FP/(FP/TP+FN)
76.41%	87.5%	84.51%	57.89%

IV. CONCLUSIONS

In this paper, we have introduced a novel WCE informative frame selection scheme based on weighted lumen detection and motility assessment function. The informative frames are chosen according to their depth perception and motility assessment. Only those frames with enough depth information together with smooth motion will be considered as informative frames in our problem since our final goal of this scheme is to facilitate the diagnosis of the physician. We employed the MS to describe the lumen in WCE image sequences and ARPS to depict the motion characteristics of WCE images. Such a combination integrates the advantages of these two schemes and is suitable for our problem. Initial experiments validated a promising performance of informative frame detection in WCE video. The detected informative frames in a WCE video can also build a good basis for our previous work that is concentrated on 3D reconstruction of GI tract using WCE images.

In the future work, we will further testify the robustness of the proposed method by using more videos with longer duration and try different scoring method to improve the accuracy of the approach. We will also trying to tackling the selection based on machine learning approach.

REFERENCES

[1] Jamie S. Barkin, "Wireless capsule endoscopy" in *gastrointestinal endoscopy clinics of North America*, Jan 2004, vol.14.

[2] Chao Hu, Max Meng "A linear algorithm for tracing magnet position and orientation by using three-axis magnetic sensors," in *IEEE trans.* On *Magnetic*, vol.43, Dec. 2007

[3] G. Iddan, G. Meron, "*Wireless capsule endoscopy*" in *Nature*, May 2000. Vol. 405, pp.417.

[4] Cheng, Yizong, "*Mean Shift, Mode Seeking, and Clustering*" in *IEEE Transactions on Pattern Analysis and Machine Intelligence* (IEEE) 17 (8): 790–799, 1995.

[5] S. J. Phee, W.S. Ng, I. M. Chen, F. Seow-Choen, and B.L. Davies. "*Automation of colonoscopy"* in *IEEE Engineering in Medicine and Biology*, pp. 81-88, 1998.

[6] L. Sucar and D. Gillies. "*Knowledge-based assistant for colonoscopy*" in *IEA.AIE*, volume 2, pages 665-672, 1990.

[7] K. Asari. "*A fast and accurate segmentation technique for the extraction of gastrointestinal lumen from endoscopic images"* in *Medical Engineering and Physics*, 22:89-96, 2000.

[8] Z. Xenophon, A. Argyros and Dimitris P. Tsakiris "*Lumen detection for capsule endoscopy*" in *IEEE International Conference of Intelligent Robots and Systems*, pp. 3921-3926, 2008.

[9] K. Asari. "*A fast and accurate segmentation technique for the extraction of gastrointestinal lumen from endoscopic images*" in *Medical Engineering and Physics*, pp. 89-96, 2000.

[10] C. Kwoh and D. Gillies. "*Using Fourier information for the detection of the lumen in endoscope images.*" In *IEEE TENCON*, pp 981-985, 1995.

[11] S. Panagiota, V. Fernando, V. Jordi and R. Petia. "*Identification of intestinal motility events of capsule endoscopy video analysis*" in *Proc. ICPR,* 2006, vol.4, pp. 719-722

[12] V. Fernando, et.al "*Intestinal motility assessment with video capsule endoscopy: automatic annotation of phasic intestinal contractions*" in *IEEE Transactions on Medical Imaging*, vol. 29, NO.2, Feb. 2010.

[13] Kingslake, R. "*Optical System Design"*, Academic Press, New York, 1983.

[14] Nie. Yao, "*Adaptive rood pattern search for fast block-matching motion estimation*" in *IEEE Transactions on Image Processing,* Vol.11, NO.12, Dec. 2002.

[15] L.M.Po and W.C. Ma, "*A novel four-step search algorithm for fast block motion estimation,*", in *IEEE Trans. Circuits Syst. Video Technol.,* vol. 6, pp.313-317, June 1996.

[16] L.K. Liu and E. Feig, "*A block-based gradient descent search algorithm for block motion estimation in video coding"*, in *IEEE Trans. Circuits Syst. Video Technol.,* vol. 6, pp.419-423, Aug. 1996.

[17] Downing, R.E., Moore, J.L., and Brown, S.W. "*The effects and interaction of spatial visualization and domain expertise on information seeking*" in *Computers in Human Behavior* p195 – 209, 2005