Measurement of Motion Detection of Wireless Capsule Endoscope Inside Large Intestine

Mingda Zhou, Guanqun Bao, and Kaveh Pahlavan
Center for Wireless Information Network Studies
Worcester Polytechnic Institute
Worcester, MA, 01609, USA
(mzhou2, gbao, kaveh)@wpi.edu

***Abstract*— Wireless Capsule Endoscope (WCE) provides a noninvasive way to inspect the entire Gastrointestinal (GI) tract, including large intestine, where intestinal diseases most likely occur. As a critical component of capsule endoscopic examination, physicians need to know the precise position of the endoscopic capsule in order to identify the position of detected intestinal diseases. Knowing how the capsule moves inside the large intestine would greatly complement the existing wireless localization systems by providing the motion information. Since the most recently released WCE can take up to 6 frames per second, it's possible to estimate the movement of the capsule by processing the successive image sequence. In this paper, a computer vision based approach without utilizing any external device is proposed to estimate the motion of WCE inside the large intestine. The proposed approach estimate the displacement and rotation of the capsule by calculating entropy and mutual information between frames using Fibonacci method. The obtained results of this approach show its stability and better performance over other existing approaches of motion measurements. Meanwhile, findings of this paper lay a foundation for motion pattern of WCEs inside the large intestine, which will benefit other medical applications.**

I. INTRODUCTION

Since first invented by Given Imaging in 2000, wireless capsule endoscopy (WCE) has become one of the most popular inspection techniques inside the gastrointestinal(GI) tract, including large intestine [1]. WCE is a swallowable device at the size of a pill and equipped with one or two miniature cameras, going through the GI tract within typically 5-8 hours. During the journey, it takes dozens of thousands of color images with a frame rate varies from 2 to 8 frames per second which enables those frames perform as a real video [2]. As a critical component of capsule endoscopic examination, physicians need to know the precise position of the endoscopic capsule in order to identify the position of detected intestinal diseases.

During the past few years, many attempts have been made to develop accurate localization techniques for the WCE [3], [4]. However, none of the existing localization methods is able to provide accurate position information of the endoscopic capsule due to non-homogeneous body tissues and un-uniformly distributed organs [5]. To complement the existing wireless localization infrastructures, researchers are investigating using compute vision techniques to track the motion of the video capsule [6]. Two of the most popular computer vision based motion tracking methods are scale-invariant feature transform (SIFT) [7], [8] and speeded up robust features (SURF) [9]. Both of them are usually formed in four steps: set up scale spaces, extract local features, generate descriptors utilizing surrounding pixels and map corresponding feature points.

Local features perform robustly when analyzing images taken inside the small intestine, at least in the upper small intestine [7]. However, things are different when WCE goes into large intestine where local features are not as clear as those in the small intestine and the peristalsis slows down which can be observed clearly from the video taken inside large intestine. Meanwhile, with higher frame rate [2], successive images perform more globally, resembling real video rather than individual images.

In this paper, we propose a novel approach estimating the orientation and displacement of the track of WCE in large intestine, only based on the information extracted from consecutive frames taken by WCE. This approach proceeds in four steps: (1) Apply low-pass filter on input images to smooth them, eliminating the noises and preparing for the next steps. (2) Calculate mutual information (MI) between input images, record the maximum MI and corresponding parameters such as orientation and scale. (3) Estimate the rotation and relative displacement of WCE according to calculated orientation and scale. (4) Performance evaluation. The main contribution of this paper is that we introduce a more global solution to analyze the relative displacement and rotation of WCE with better performance than that of feature based applications proposed in [9]–[11]. Also, our approach is the measurement of WCE in large intestine while most of the related works are designed for WCE in small intestine. Moreover, this approach is much easier to be applied due to its higher linearity and stability.

The rest of this paper is organized as follows: Section II includes details about the image analysis algorithm applied to WCE images. In section III, we talk about the experimental results and analytical comparison with other algorithms to validate the performance of our approach. Finally, in section IV, conclusion is drawn.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

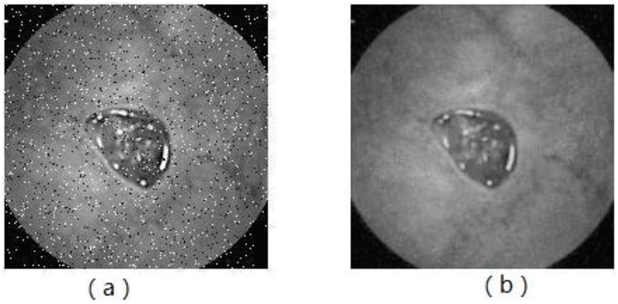

Fig. 1. (a) Frame with salt and pepper noise (b) Frame applied Butterworth filter.

II. METHODOLOGY

Our approach basically consists of 4 steps: Pre-processing, mutual information calculation, parameter searching and performance evaluation.

(1) Low-pass 2-D Butterworth filter are implemented to eliminate the noises and guarantee the quality of the third step.

(2) After the pre-processing, we calculate the mutual information of pairs of images. The more similar two images have, the more mutual information value is calculated.

(3) Via searching the maximum mutual information and corresponding parameters such as scales and angles, rotation and relative displacement can be discovered so that we could tell the movement of WCE.

(4) In the process of searching and performance evaluation, bicubic interpolation is frequently engaged, which can delicately help reconstruct float images.

A. Pre-processing

When the WCE passes through the large intestine, it takes images and transmits them to sensor array located outside the human body, but noises may be generated and shown in images.

Then after extraction of images from the signals sent by WCE, we transform those images into frequency domain and then apply a low-pass 2-D Butterworth filter on each frame to filter the noise (e.g. pepper or Gaussian). This step effectively guarantees the results of analysis to be robust enough. Define W as pass band and n to be order of the 2-D filter. Then the 2-D Butterworth filter can be described as

$$G^2(w) = |H(w)|^2 = \frac{1}{1+(\frac{w}{w_c})^2 n} \tag{1}$$

Following Eq.1, 2-D IFFT is applied to transform frames back to spatial domain. Fig. 1(a) shows one of the frames with pepper noise and (b) is filtered by Butterworth filter.

Consequently, the filtered frame is obviously more smooth and is better for statistical analysis.

B. Mutual Information Calculation

The similarity of two images can be directly expressed as a function of their mutual information, which is proposed in Shannons information theory [12]. Formally, the mutual information of two images X and Y can be defined as:

$$I(X,Y) = \sum_{x\in X, y\in Y} p(x,y) log(\frac{p(x,y)}{p(x)p(y)}) \tag{2}$$

where p(x) and p(y) are marginal probability distribution functions of gray scales in images X and Y respectively. Eq.2 can be equivalently expressed as

$$I(X,Y) = H(X) + H(Y) - H(X \cup Y) \tag{3}$$

$X \cup Y$ represents the gray scale mapping matrix of image X and Y [13]. H(X) in Eq. 3 is the Shannon entropy of image X which can be expressed as

$$H(X) = \sum_{x\in X} p(x) log(p(x)) \tag{4}$$

As can be observed from Eq.2 to Eq.4, the mutual information provides a global view of similarity via statistical analysis, which is supposed to be more effective on global scale.

C. Fibonacci Searching Technique

Since the mutual information only reflects the similarity of two frames, a searching technique is required to find the maximum mutual information and the corresponding deformation. Fibonacci searching technique is a robust iterative method for searching extreme value to achieve this goal, performing better and is less time consuming than binary searching technique. Fig. 2 is given as the flow chart of Fibonacci searching technique, where the ratio 0.618 is called golden ratio.

To insure the outcomes of Fibonacci searching technique, the threshold should be small enough (e.g. 0.005). In addition, Butterworth filters smoothing effect prevents the searching procedure from being stuck in a short interval with violent jitters.

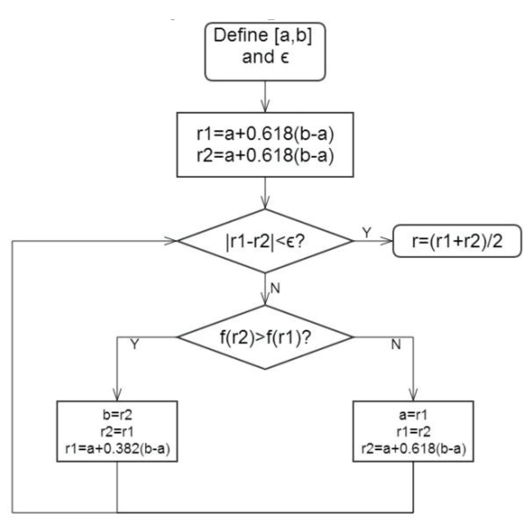

Fig. 2. Flow chart of Fibonacci searching technique

TABLE I

ESTIMATION OF ROTATION

Rotation angle	Calculated orientation error (degree)	Calculated orientation error in [9] (degree)	Calculated orientation error in [10] (degree)	Calculated orientation error in [11] (degree)
5	0.0352	0.022	0.63	0.0673
10	0.0320	0.022	0.83	0.179
15	0.0453	0.026	0.90	0.403
20	0.0913	0.028	1.02	1.775
25	0.0227	0.689	0.86	0.829
30	0.0410	1.558	0.96	Large error
35	0.0844	4.719	0.92	-
40	0.0557	20.084	0.93	-
45	0.0821	30.771	0.78	-

D. Bicubic Interpolation

Since the float frame needs to be iteratively deformed to discover the maximum mutual information with reference frame, the new gray scales of pixels in float frame could be calculated by bicubic interpolation with higher accuracy. Define f(i+u,j+v) as the corresponding gray scale in reference frame of a pixel from float frame where i+u and j+v are the corresponding coordinate. Then it can be written as:

$$f(i+u, j+v) = \vec{A} * \vec{B} * \vec{C} \tag{5}$$

$$\vec{A} = [S(u+1) \quad S(u) \quad S(u-1) \quad S(u-2)] \tag{6}$$

$$\vec{B} = \begin{bmatrix} f(i-1,j-1) & f(i-1,j) & f(i-1,j+1) & f(i-1,j+2) \\ f(i,j-1) & f(i,j) & f(i,j+1) & f(i,j+2) \\ f(i+1,j-1) & f(i+1,j) & f(i+1,j+1) & f(i+1,j+2) \\ f(i+2,j-1) & f(i+2,j) & f(i+2,j+1) & f(i+2,j+2) \end{bmatrix} \tag{7}$$

$$\vec{C} = [S(v+1) \quad S(v) \quad S(v-1) \quad S(v-2)]^T \tag{8}$$

where S(x) is the primary function representing weights of pixels which is given by:

$$S(x) = \begin{cases} |x|^3 - 2|x|^2 + 1 & |x| < 1 \\ -|x|^3 + 5|x|^2 - 8|x| + 4 & 1 \leq |x| < 2 \\ 0 & 2 \leq |x| \end{cases} \tag{9}$$

Thus, the deformed frame can be reconstructed pixel by pixel smoothly.

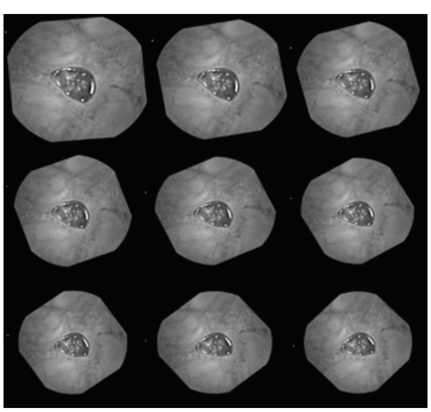

Fig. 3. An example of tested frames, 9 rotation angles from 5 to 45 in different resolutions

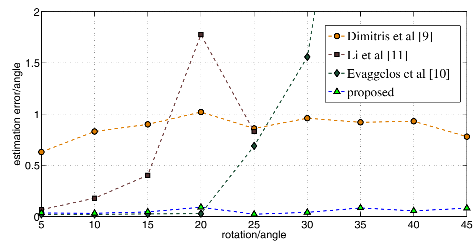

Fig. 4. Compared experimental results of rotation estimation with [9, 10, 11]

III. RESULTS AND ANALYSIS

Our approach is evaluated based on 119 consecutive WCE video frames, each with 531*531 pixels resolution donated by Given Imaging.

Because of the shortage of ground truth data set to compare with, the only way is using rotation transformations and scale simulations which are also adopted by [9].

A. Orientation

Nine rotation angles from 5 to 45 with a step of 5are tested in this paper (shown in Fig. 3) and obtained results are shown in table 1. This table indicates a stable performance and relatively low error throughout all the tested angles while those outcomes from [9]–[11] acquire larger errors following the increases of rotation angle.

To be more directly, we plot the statistical analysis in Fig. 4 from which severe error ratios can be observed in other approaches when the rotation angle is equal or higher than 30.

B. Relative Displacement

In this paper, we test 12 different scale values varying from 0.2 to 3.0, part of which is shown in Fig. 5. To transform the results from scale values to relative displacement in order to measure the motion tracking of WCE, the displacement estimation method according to projective transformation in [9] is introduced.

As can be seen in table 2, estimated displacement errors stay in the same magnitude when our approach is applied while the errors in [9] are unacceptably large when actual scale values are relatively small and decrease alone with the

TABLE II

ESTIMATION OF RELATIVE DISPLACEMENT

Actual scale	Calculated displacement error	Calculated displacement error in [9]	Calculated displacement error in [10]
0.2	0.0194	Large error	0.13
0.3	0.0313	13.3085	-
0.4	0.0079	0.0193	0.12
0.6	0.0031	0.0005	0.02
0.8	0.0026	0.0004	0.17
1.2	0.0011	0.0005	-
1.4	0.0013	0.0004	-
1.6	0.0016	0.0007	-
1.8	0.0097	0.0007	-
2.0	0.0009	0.0005	0.19
2.5	0.0026	0.0010	-
3.0	0.0005	0.0016	0.37

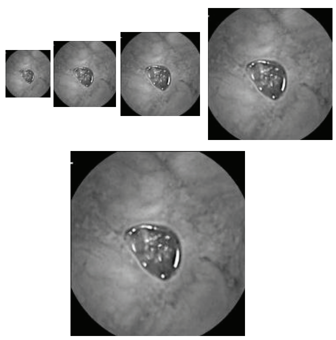

Fig. 5. An example of frames in difference scales: 0.2, 0.3, 0.4, 0.6 and 0.8

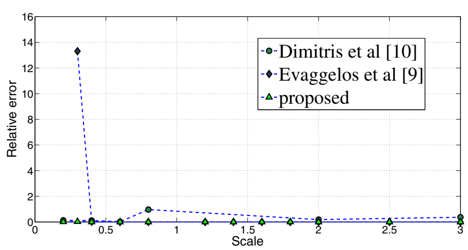

Fig. 6. Compared experimental results of displacement estimation with [9, 10]

augment of actual scale. On the other hand, the calculated errors in [10] are about 10 times larger through the procedure of estimation. Fig.6 shows the overall performance quality of 3 approaches.

IV. CONCLUSIONS

In this paper, we proposed a novel approach to measure the motion tracking of WCE inside the large intestine. We applied maximum mutual information theory as theoretical support and Fibonacci technique as searching technique so that the measurements focus more on global information of frames. The major contribution of our research is that we utilize the global statistical information of frames rather than local features to measure the motion of WCE and achieve a higher accuracy and stability.

REFERENCES

[1] A. Karargyris and N. Bourbakis, "Wireless capsule endoscopy and endoscopic imaging: A survey on various methodologies presented," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 29, no. 1, pp. 72–83, 2010.

[2] S. N. Adler and Y. C. Metzger, "Pillcam colon capsule endoscopy: recent advances and new insights," *Therapeutic advances in gastroenterology*, vol. 4, no. 4, pp. 265–268, 2011.

[3] S. Li, Y. Geng, J. He, and K. Pahlavan, "Analysis of three-dimensional maximum likelihood algorithm for capsule endoscopy localization," in *Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on*. IEEE, 2012, pp. 721–725.

[4] K. Pahlavan, G. Bao, Y. Ye, S. Makarov, U. Khan, P. Swar, D. Cave, A. Karellas, P. Krishnamurthy, and K. Sayrafian, "Rf localization for wireless video capsule endoscopy," *International Journal of Wireless Information Networks*, vol. 19, no. 4, pp. 326–340, 2012.

[5] G. Ciuti, A. Menciassi, and P. Dario, "Capsule endoscopy: from current achievements to open challenges," *Biomedical Engineering, IEEE Reviews in*, vol. 4, pp. 59–72, 2011.

[6] G. Bao, L. Mi, and K. Pahlavan, "A video aided rf localization technique for the wireless capsule endoscope (wce) inside small intestine," in *Proceedings of the 8th International Conference on Body Area Networks*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2013, pp. 55–61.

[7] ——, "Emulation on motion tracking of endoscopic capsule inside small intestine," in *14th International Conference on Bioinformatics and Computational Biology, Las Vegas*, 2013.

[8] G. Bao, Y. Ye, U. Khan, X. Zheng, and K. Pahlavan, "Modeling of the movement of the endoscopy capsule inside gi tract based on the captured endoscopic images," in *Proceedings of the IEEE International Conference on Modeling, Simulation and Visualization Methods, MSV*, vol. 12, 2012.

[9] E. Spyrou and D. K. Iakovidis, "Video-based measurements for wireless capsule endoscope tracking," *Measurement Science and Technology*, vol. 25, no. 1, p. 015002, 2014.

[10] D. K. Iakovidis, E. Spyrou, D. Diamantis, and I. Tsiompanidis, "Capsule endoscope localization based on visual features," in *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on*. IEEE, 2013, pp. 1–4.

[11] L. Liu, C. Hu, W. Cai, and M.-H. Meng, "Capsule endoscope localization based on computer vision technique," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 3711–3714.

[12] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

[13] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *Neural Networks, IEEE Transactions on*, vol. 5, no. 4, pp. 537–550, 1994.