A Video-based Speed Estimation Technique for Localizing the Wireless Capsule Endoscope inside Gastrointestinal Tract

Guanqun Bao, Liang Mi, Yishuang Geng, Mingda Zhou and Kaveh Pahlavan, IEEE fellow
Center for Wireless Information Network Studies
Worcester Polytechnic Institute
Worcester, MA, 01609, USA
(gbao, lmi, ygeng, kaveh)@wpi.edu

***Abstract*— Wireless Capsule Endoscopy (WCE) is progressively emerging as one of the most popular non-invasive imaging tools for gastrointestinal (GI) tract inspection. 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 intestinal disease. For the WCE, the position of the capsule is defined as the linear distance it is away from certain fixed anatomical landmarks. In order to measure the distance the capsule has traveled, a precise knowledge of how fast the capsule moves is urgently needed. In this paper, we present a novel computer vision based speed estimation technique that is able to extract the speed of the endoscopic capsule by analyzing the displacements between consecutive frames. The proposed approach is validated using a virtual testbed as well as the real endoscopic images. Results show that the proposed method is able to precisely estimate the speed of the endoscopic capsule with 93% accuracy on average, which enhances the localization accuracy of the WCE to less than 2.49 cm.**

I. INTRODUCTION

Wireless Capsule Endoscopy (WCE) is progressively emerging as one of the most popular non-invasive imaging tools for gastrointestinal (GI) tract disease diagnosis. Compared with the traditional wired colonoscopy or enteroscopy, WCE offers a patient-friendly, noninvasive and painless investigation of the entire small intestine, which other wired video endoscopic instruments can hardly reach. As a critical component of the capsule endoscopic examination, physicians need to know the precise position of the endoscopic capsule in order to identify the position of the abnormality after it is found by the video source.

A good review of existing localization techniques for the WCE is given in [1]. However, most of these methods aim to find the 3D coordinate of the endoscopic capsule, but from the physicians' point of view, knowing how much linear distance that the capsule has traveled away from certain anatomical landmarks is more useful for them to identify the position of the lesions during the open surgery [2]. An commonly used approach to measure this linear distance is to assume the capsule travels at a constant speed and the approximate position of the capsule is estimated by keeping track of the time it traveled away from landmarks, such as pylorus and ileocecal valve. But in reality, the speed of the capsule varies a lot propelled by the open and close intestinal motility ("peristalsis"). Therefore, when using this approach, the further the capsule moves away from these landmarks, the greater the error is.

To enhance the localization accuracy, a precise knowledge of how fast the capsule moves is urgently needed. In this paper, we present a novel computer vision based speed estimation technique that is able to accurately estimate the speed of the WCE by analyzing the displacements of common portion of the scene between consecutive endoscopic frames. The position of the capsule, defined as the distance away from certain landmarks, is found by integrating the speed over the elapsed time. The major contribution of this paper is that we present a quantitative way to calculate the temporary speed of the endoscopic capsule rather than assuming it travels at a constant speed, thereby, improve the localization accuracy of the WCE inside small intestine by significance. The performance of our approach is validated under a virtual testbed as well as the real images. Results show that the proposed approach is able to provide speed estimation accuracy no less than 93% and improve the localization accuracy to less than 2.49 cm in terms of linear distance.

The rest of the paper is organized as follows: in section II, we descried the detailed procedure of how to estimate the speed of the endoscopic capsule using computer vision technique. In section III, experimental results of the proposed speed estimation approach are given with analytical discussion. Finally, conclusion is drawn in section IV.

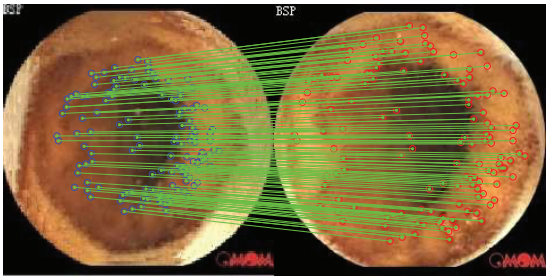
II. MATHODOLOGY

The translation of the endoscopic capsule inside the small intestine can be modeled as a tiny camera passing through a elastic cylindrical tube. Since the WCE continuously takes pictures at a rate up to 6 frames/sec, common portions of the scene may present between consecutive images [3]. These portions of the images are called "feature points" (FP). The pattern and magnitudes of the displacements of these feature points can be used as a hint to reveal the speed of the endoscopic capsule.

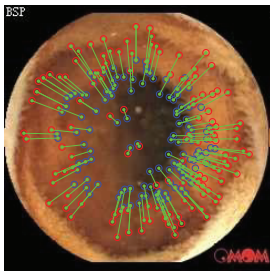
A. Feature Points Matching

To make an accurate estimation of the capsule's speed, it's very important that the FPs extracted from the reference (first) frame can be accurately located in the follow-

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



(a) Corresponding feature points between two consecutive frames



(b) Formation of motion vectors

Fig. 1. Feature matching between two consecutive images using A-SIFT

ing frames. The Affine Scale-invariant Feature Transform (ASIFT) defined by the affine camera model in Eq. 1, is a perfect matching tool for the WCE images due to its immune property to viewpoint changes, blur, noise and spatial deformations.

$$A = H_\lambda R_1(\Psi) T_t R_2(\Phi) = \lambda \begin{bmatrix} cos\Psi & -sin\Psi \\ sin\Psi & cos\Psi \end{bmatrix} \begin{bmatrix} t & 0 \\ t & 1 \end{bmatrix} \begin{bmatrix} cos\Phi & -sin\Phi \\ sin\Phi & cos\Phi \end{bmatrix} \quad (1)$$

where R represents rotation and T represents tilt. Ψ is rotation angle of camera around optical axis. Φ is longitude angle between optical axis and a fixed vertical plane. λ is zoom parameter. Detailed procedure of FPs matching using ASIFT can be found in [4]. An example of feature points matching is given in Fig. 1 (a), in which blue "o" represents the coordinates of detected FPs in the reference frame, red "o" represent the coordinates of matched FPs on the second frame. If we connect the corresponded FP pairs on the same frame (as shown in Fig. 1 (b)), a bunch of motion vectors will be generated representing the displacements of FPs between frames.

B. Quantitative Calculation of Speed

We use Fig. 2 to illustrate the procedure of calculating the speed of WCE. In Fig. 2, point P is a FP detected at a distance D from the initial position of the camera C with its angular depth equal to θ_1. After the camera has moved forward by a distance d to a new position C', the angular depth of P changes to θ_2. The changes in angular depth can be used to calculate the displacement of the capsule.

$$\theta_1 = tan^{-1}\frac{R}{D} \quad \Longrightarrow \quad D = \frac{R}{tan\theta_1} \quad (2)$$

$$\theta_2 = tan^{-1}\frac{R}{D-d} \quad (3)$$

replacing D in Eq. 3 with Eq. 2, we get:

$$d = \frac{R}{tan\theta_2}\left(1 - \frac{tan\theta_2}{tan\theta_1}\right) \quad (4)$$

where R is the radius of the small intestine. Since the small intestine is a nonrigid environment, R changes as the small intestine contracts. In our previous work [5], we developed an algorithm that can automatically determine the radius R by

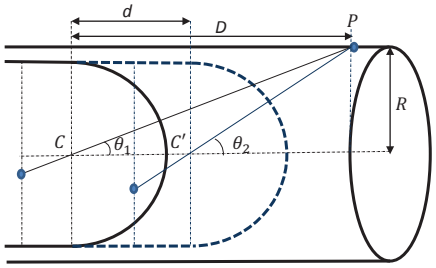


Fig. 2. Geographic model for speed estimation

measuring the size of black hole in the image. If we assume the time interval between the frames is half a second (which can be up to 6 frames / sec for the most recent released capsules), the speed of the capsule can be calculated by:

$$v = \frac{\frac{1}{N}\sum_{i=0}^{N} d_i}{0.5} = \frac{2}{N}\sum_{i=0}^{N}\frac{R}{tan\theta_{2i}}\left(1 - \frac{tan\theta_{2i}}{tan\theta_{1i}}\right) \quad (5)$$

where N equals to the total number of all detected FPs.

From Eq. 4 it can be seen that information on depth of FP is factored into the final expression of distance moved by the capsule. In this way, the actual displacement d of the camera is independent of the location of the FP chosen in the image.

C. Image Unrolling

To facilitate the derivation of angler depth, we map the coordinate (x,y) of any point on the original cylindrical image plane to the unrolled image plane (x',y') by:

$$x' = \frac{L\phi}{2\pi} \qquad y' = r \quad (6)$$

where ϕ is the angle between point P and horizontal axis.

$$\phi = tan^{-1}\left(\frac{y-y_0}{x-x_0}\right) \quad (7)$$

r is the radius of the circular ring associated with point P that can be calculated by:

$$r = \sqrt{(x-x_0)^2 + (y-y_0)^2}. \quad (8)$$

L and H are length and height of the unrolled image plane respectively. Fig. 3 shows the result of image unrolling.

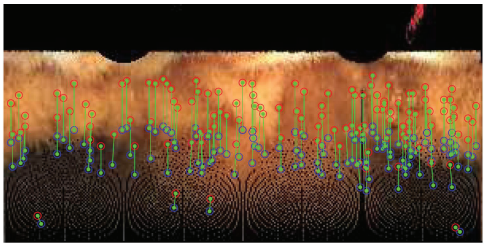


Fig. 3. Image unrolling

In this unrolled image plane, x' axis represents the radian angle ϕ whose value ranges from 0 (when $x' = 0$) to 2π (when $x' = L$). y' axis represents angular depth which reflect the distance away from the camera. $y' = 0$ represents a 0 angular depth and $y' = H$ gives the maximum field of view η of the camera. Under this new coordinate system, the angular depth of any point P can be calculated directly through its y' value by:

$$\theta \cong \left(\frac{y'}{H}\right)\eta \tag{9}$$

III. EXPERIMENTAL RESULTS AND ANYLYSIS

Validation of the proposed speed estimation algorithm is extremely difficult. One of the major reasons is that we have limited control of the capsule after it is swallowed by the patient, so we could not verify the performance of the algorithms [2], [6]. Besides, experiments on human subjects are strictly restricted by law. Thus, the only way to quantitatively test our speed estimation algorithm is to create a virtual emulation testbed. In our previous work [7], we established a virtual testbed for validation purpose (as shown in Fig. 4). The virtual test bed shared the same size, texture and topology with the real small intestine and we move a camera view point inside the testbed at known time-variant speed and takes pictures at 2 frames / sec. Similar emulation set up can be found in [6], [8], [9]. In this way, by applying the proposed speed estimation algorithm to the emulated image sequence, the quantitative performance of the proposed approach can be validated.

A summary of the experimental results is given in Table 1. The average speed of the capsule inside small intestine is 0.48 mm/s. It can be seen from Table 1 that more FPs are detected when the capsule moves slowly. As long as the speed of the capsule stay below 4 mm/s, the proposed speed estimation algorithm is able to provide estimation accuracy no less than 93%. Thus, it fullfills the accuracy requirement for the WCE application. As the speed increases, since the overlap portion between consecutive frames becomes smaller, less FPs are detected, which affects the accuracy of the proposed speed estimation algorithm. When the capsule's speed reach to 10 mm/s, almost no common portion present between the frames, thus, no FPs are detected the algorithms fails.

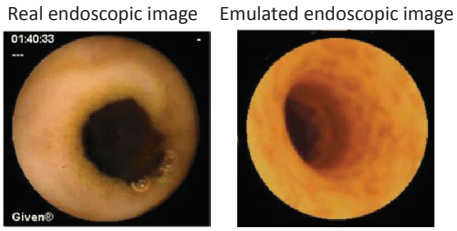


Fig. 4. Emulation testbed for speed estimation

TABLE I
SPEED ESTIMATION RESULTS UNDER TESTBED

Average speed	Number of detected FPs	Average estimation errors
1 (mm/s)	79	0.022 (mm/s)
:	:	:
4 (mm/s)	32	0.298 (mm/s)
10 (mm/s)	NA	NA

Then we compare the localization results using the proposed speed estimation method with the previously mentioned constant average speed method. 5000 artificial frames were generated by moving the view point inside the virtual testbed with speeds various from 0 to 4 mm/s. The position of the capsule is determined by calculating the distance it has traveled away from the starting point. Fig. 5 shows the comparative localization results of using both methods. We found that the average error of using the proposed speed estimation is able to provide accuracy less than 2.71 cm on average compared with up to 27 cm error that constant speed algorithm provides.

We also tested our algorithm with the real clinical data. Fig. 6 shows a sequence of 60 endoscopic frames. After applying our speed estimation algorithm, a plot of the corresponding speed is given underneath. Since we do not have precise control of the capsule, we are not able to perform quantitative evaluation, but we can examine the images manually by visual inspection. As shown in Fig. 6, the whole image sequence can be divided into 4 sections marked by A, B, C, and D. It can be seen that during section A and C, the scene almost stays still which indicates

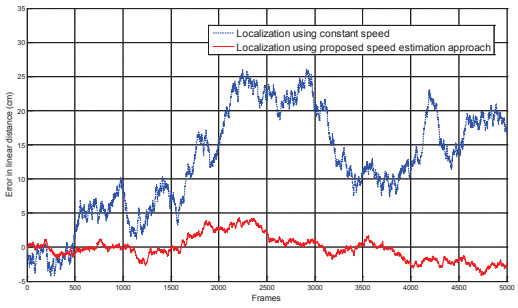


Fig. 5. Comparative localization results between the proposed method and constant speed method

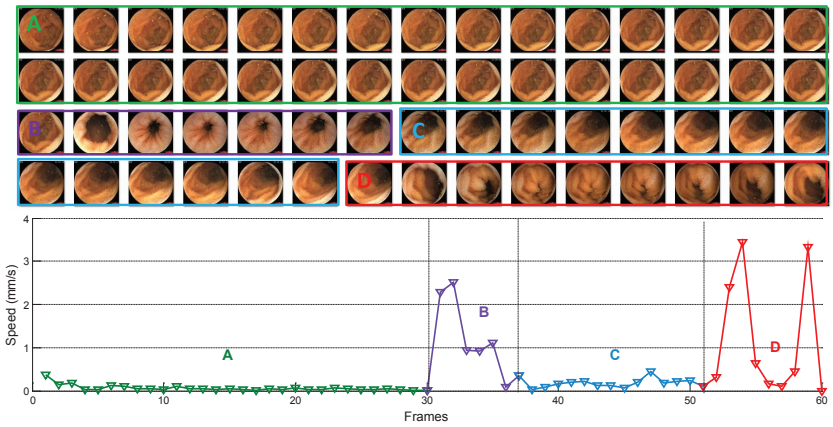
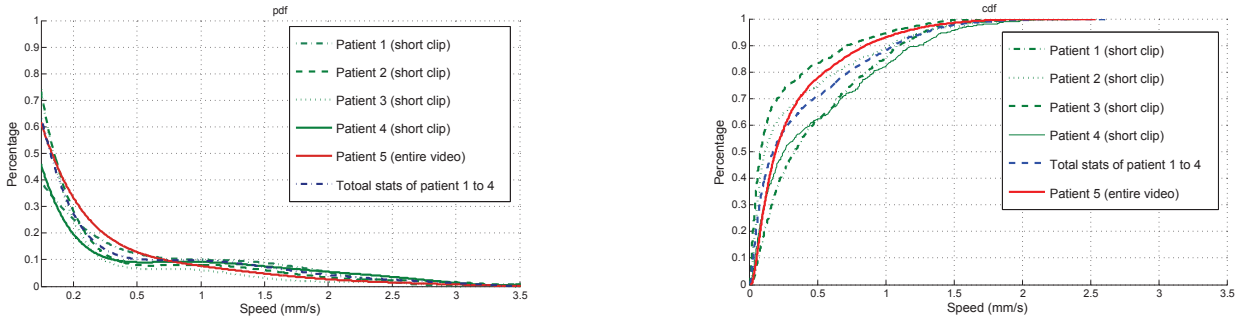


Fig. 6. Speed estimation results of a sequence of real endoscopic images



(a) PDF of the speed estimation from different individuals

(b) CDF of the speed estimation from different individuals

Fig. 7. Statistical results of speed estimation result using clinical data

a slow motion of the capsule, while in section B and D, the capsule moves faster. The trend of the corresponding speed estimation plot matches this observation. To further validate our algorithm, we compared the statistical results of clinical data (4 short video clips and 1 long video clip) from 5 different patients. The results are shown in Fig. 7. It can be seen although the video clips are from different individuals, after applying our speed estimation algorithm, the estimated speed shares very similar distribution in term of probability density function (PDF) and cumulative distribution function (CDF).

IV. CONCLUSIONS

In this paper, we present a novel computer vision based speed estimation technique to facilitate the localization of WCE inside small intestine. The proposed method extracts the speed of the endoscopic capsule by processing the video source that comes with the endoscopic examination, therefore, no extra device is needed. Experimental results show that our method is able to precisely estimate the speed of endoscopic capsule and enhance the localization accuracy of WCE by significance.

REFERENCES

[1] T. D. Than, G. Alici, H. Zhou, and W. Li, "A review of localization systems for robotic endoscopic capsules," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 9, pp. 2387–2399, 2012.

[2] 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.

[3] C. S. Bell, K. L. Obstein, and P. Valdastri, "Image partitioning and illumination in image-based pose detection for teleoperated flexible endoscopes," *Artificial intelligence in medicine*, vol. 59, no. 3, pp. 185–196, 2013.

[4] J.-M. Morel and G. Yu, "Asift: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, 2009.

[5] L. Mi, G. Bao, and K. Pahlavan, "Geometric estimation of intestinal contraction for motion tracking of video capsule endoscope," 2013.

[6] L. France, J. Lenoir, A. Angelidis, P. Meseure, M.-P. Cani, F. Faure, and C. Chaillou, "A layered model of a virtual human intestine for surgery simulation," *Medical image analysis*, vol. 9, no. 2, pp. 123–132, 2005.

[7] L. Mi, G. Bao, and K. Pahlavan, "Design and validation of a virtual environment for experimentation inside the 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. 35–40.

[8] S. Seshamani, W. Lau, and G. Hager, "Real-time endoscopic mosaicking," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*. Springer, 2006, pp. 355–363.

[9] P. M. Szczypiński, R. D. Sriram, P. V. Sriram, D. N. Reddy *et al.*, "A model of deformable rings for interpretation of wireless capsule endoscopic videos," *Medical Image Analysis*, vol. 13, no. 2, p. 312, 2009.