

Modeling of the Movement of the Endoscopy Capsule inside G.I. Tract based on the Captured Endoscopic Images

Guanqun Bao, Yunxing Ye, Umair Khan and Kaveh Pahlavan
Center for Wireless Information Network Studies
Worcester Polytechnic Institute
Worcester, MA, 01609, USA
(gbao, yunxingye, uikhan, kaveh)@wpi.edu

Abstract — Wireless capsule endoscopy (WCE) is a noninvasive way to provide excellent images of the lumen of the intestines as the capsule moving along in the gastrointestinal (G.I.) tract. However, the biggest drawback of this technology is its incapability of localizing of the capsule when an abnormality is found by the video source. Existing localization methods based on radio frequency (RF) and magnetic field suffer a great error due to the non-homogeneity of the human body and uncertainly of the movement of the endoscopic capsule. To complement the existing localization techniques, in this paper, we developed a series of image processing and visualization based algorithms to model the movement of the endoscopic capsule. First, a 3D map of the G.I. tract is generated to navigate the transition of the capsule. Then, by comparing the local similarity and feature points matching between the consecutive frames, the speed and rotation angels of the capsule can be roughly estimated. Finally, by mapping the pattern of the capsule's movement onto the 3D G.I. tract map, we are able to simulate the entire transition of the endoscopic capsule in the 3D space. The experimental results show that the proposed method has a good estimation of the movement of the capsule.

Keywords - capsule endoscopy, localization, movement modeling, rotation estimation

I. INTRODUCTION

Wireless capsule endoscopy (WCE) [1] has been in the clinical arena for 12 years. It provides a noninvasive imaging technology of the entire G.I. tract. Current devices, for example devices developed by Given Imaging and Olympus American, are able to provide excellent images of the lumen of the intestines as they moving along in the intestinal tract. However, none of these devices could provide accurate location information of the capsule when an abnormality is found by the video source. An accurate measurement of the capsule's position is of great benefit to the physicians in terms of reducing diagnosis time and taking immediate clinical management to obscure gastrointestinal bleeding [2]. Many efforts had been done to develop reliable localization technique inside human body. The Given Image Pillcam capsule [3] was originally developed with the potential capability of localizing the capsule on a 2-D plane at a twice-per-second rate. Eight external antennae were fixed to the anterior abdominal wall to detect the UHF-band signal that is emitted by the capsule. The position of the capsule on the 2-D plane of the abdomen is

estimated depending on the signal strength received by each antenna with an accuracy of 6 inches. However, the clinical use of this software found that the crude localization result generated by the software was not helpful and this approach was soon abandoned. Another commonly used approach for capsule localization is to assume the capsule travels at a constant speed and the approximate position of the capsule is calculated according to the time of travel away from some pre-defined land marks such like pylorus and ileo-cecal valve. Apparently, when using this approach, the further the capsule moves away from the land marks, the greater the likelihood of error is. Especially after the video capsule has entered a few centimeters of the small intestine, the localization error will increase dramatically. This is mainly due to the high complicity level of the shape of the small intestine. The distribution of small intestine is like a curled snake with its length varies from 4.6m to 9.8m [4] (the average value for human being is 7m) and the tendency of loops is highly undistinguishable. Besides, the intestinal motility is not consistent. Peristalsis may make the wireless capsule sometimes move quickly, sometimes stop or sometimes even reverse and then progress with any combination of the movement above. Furthermore, the transition of the capsule itself is not axial. It may rotate with different angles or get flipped by 180°. The unpredictable angulation of the wireless capsule creates difficulties in RSS based localization in terms of changing the antenna gain. Thus, knowing how the capsule moves inside human body will help us to analyze the radio channel and thereby enhance the accuracy of the localization. What's more, considering the abo 3-dimensional distribution of the small intestine, a 3D localization technique instead of 2D is also needed to provide sufficiently accurate spatial location information of the capsule.

The rest of the paper is organized as follows: In section II, we generated a 3D intestinal tract map by extracting the central line of the existing 3D G.I. tract models. This map provides us a clear view of how the capsule transits inside human body. Then, by comparing the local similarity and matching the feature points between the consecutive endoscopic images, information such like speed and rotation angles of the endoscopy capsule can be roughly estimated to model the capsule's movement inside the G.I. tract. In section III, both the experimental results and analysis of the proposed method are given explicitly. Conclusion and future work are addressed in section IV.

II. MOVEMENT OF THE ENDOSCOPY CAPSULE

A. 3D Map Generation

In every localization technique, map always plays a very important role in terms of refining the localization results. Existing literature [5] reported that a clear street map is able to reduce the GPS localization error from tens of meters to several meters in the urban area. In case of the localization inside human body, “map” is even more important since everything goes through the G.I tract follows the same route. Knowing a clear pattern of the intestinal tract will greatly enhance the localization accuracy. Therefore, tracing the path of intestinal tract is essential to the accurate capsule localization. Given a 3D CAD model of the G.I. tract as shown in Fig. 1, we want to trace the center of the intestinal volume so we can model the movement of the capsule inside the tract. In this sub section, some 3D image processing techniques are applied to accomplish this goal. For the large intestine (as shown in the middle of Fig. 1), since it already has a very clear pattern which looks like a big hook, we applied 3D skeletonization technique [6] to extract the path of it. As for the small intestine, since the shape of the small intestine is much more complicated (the trend of the small intestine can be hardly recognized by human eyes), we developed an element sliding technique to trace the path. The basic idea behind this technique is to define an element shape (ES) with its radius automatically adjustable to the radius of the small intestine. This ES is propelled forward by a factor associated proportional to the average distance between the vertices within certain range and the physical center of the ES. As the ES goes along the small intestine, the position of its physical center is recorded and this will give us a clear path of the small intestine. The preliminary result of the path extracted from the 3D model is shown on the right of Fig. 1.

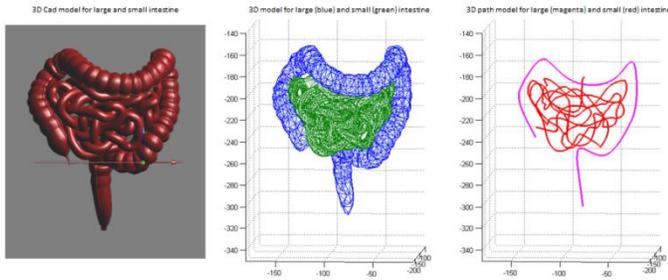


Fig. 1. 3D path generation from a 3D G.I. tract model

B. Speed Estimation of the Endoscopy Capsule

After the endoscopy capsule was swallowed by the patient, it travels through the G.I. tract propelled by peristalsis [4]. From the dataset we observe that whenever the intestinal lumen contracts, the difference between consecutive frames is high, which to some level reflects a higher speed of the capsule’s movement. Based on this observation, we developed an automatic peristalsis detection method based on the color histogram similarity between the two consecutive frames [7]. The reason why we choose color histogram as the similarity metric is that it’s more robust in this particular application and insensitive to the texture noise caused by the bubbles and little pieces of food. However, the overall color histogram is not a

good descriptor of the color feature since frames with different contents may have the similar histogram distribution. One example is given in Fig.2, although the contents of the two frames are totally different, the overall histograms of the two still look similar. Thus, instead of using the overall color histogram, we developed a local color histogram comparison algorithm. The algorithm divides the captured image into 16 non-overlapping blocks and calculates the local similarity for each pair of the corresponding blocks. Besides, the original frames captured by the endoscopy capsule are encoded by RGB color space, which is difficult to describe the nature of a color by the amounts of each channel. Thus, we convert the histograms into HSV color space, which is more similar to human’s visual perception [8, 9]. Then, the similarity between two frames can be calculated by:

$$Sim(Img_1, Img_2) = \frac{1}{M * N} \sum_{i=1}^M \sum_{j=1}^N \left(1 - \frac{|H_{i1}(j) - H_{i2}(j)|}{Max(H_{i1}(j), H_{i2}(j))} \right) \quad (1)$$

where H_1 and H_2 represent the color histograms of the corresponding blocks. M is the number of the non-overlapping blocks which equals to 16. N is the sample number of the histogram which equals to 768. Since the WCE uses a pinhole camera, when calculate the similarity, only the area covered by the red circle (as shown in Fig. 3) is under consideration. A sample partitioned endoscopic frame and its HSV histograms is shown in Fig.3.

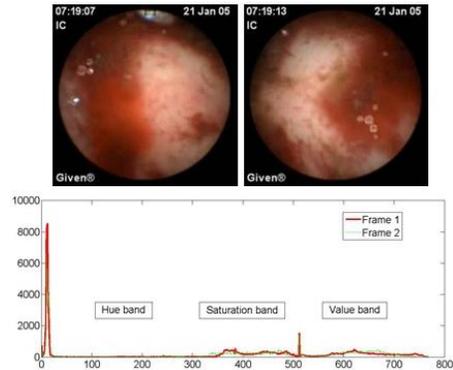


Fig.2. Overall histogram comparison

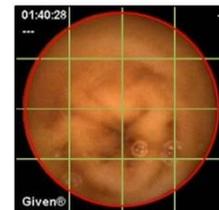


Fig.3. Local similarity comparison based on partitioned endoscopic image

C. Rotation Estimation of the Endoscopy Capsule

If the similarity between two frames is high, it can be inferred that the endoscopy capsule stayed still in the same position or just rotated slightly during measured time slot. In the case when the same pattern appears in both frames, it’s

possible to calculate the rotation angel between the two frames by feature matching. First, a set of feature points needs to be identified in the reference (first) frame. The simplest statistical measurement is to calculate the variance σ^2 of grey levels in a square neighborhood ($P \times P$, $P=16$) centroid on a pixel. By sliding the square window back and forth on the frame, points with maximum local grey level variance (mostly on the edges) are selected to be the initial feature points.

$$\sigma^2 = \frac{1}{P^2} \sum_{a=-P/2}^{P/2} \sum_{b=-P/2}^{P/2} (f(i+a, j+b) - \mu)^2 \quad (2)$$

$$\mu = \frac{1}{P^2} \sum_{a=-P/2}^{P/2} \sum_{b=-P/2}^{P/2} f(i+a, j+b) \quad (3)$$

where σ^2 is the variance of the local grey level and μ is the average grey level within the sliding window.

Then, to find the corresponding feature points in the second frame, a cross-correlation matching technique [10] is applied. In signal processing, cross-correlation is a classical method of estimating the degree to which two series of signals are correlated. In 2D pattern recognition, cross-correlation can be used for identifying the target pattern in the image. Consider the image below in black and the mask shown in red. The mask is centered at every pixel in the image and the cross correlation is calculated, this forms a 2D array of correlation coefficients. The un-normalized correlation coefficient at position (i, j) on the image is given by:

$$r(i, j) = \sum_{a=-Q/2}^{Q/2} \sum_{b=-Q/2}^{Q/2} \left(\text{mask} \left(a + \frac{Q}{2}, b + \frac{Q}{2} \right) - \overline{\text{mask}} \right) (f(i+a, j+b) - \overline{\text{image}}) \quad (4)$$

where Q is the size of the mask, $f(i, j)$ represents the intensity value at (i, j) , $\overline{\text{mask}}$ is the mean value of the mask pixels and $\overline{\text{image}}$ is the mean value of the image pixels covered by the mask. The mask itself is a cropped image which needs to have the same appearance as the pattern to be found. If the mask and the pattern being sought are similar, the cross correlation between the two will be high. The peak $r(i, j)$ is the position of the best match in the searching image.

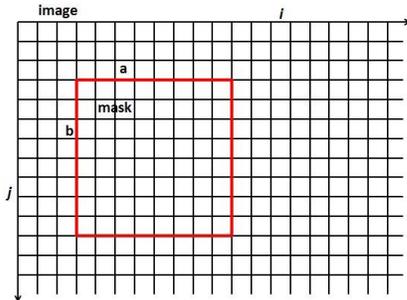


Fig.4. Cross correlation sliding window

Knowing the positions of the corresponding points in two consecutive frames, to find how much the capsule rotated, we need to calculate the rotation matrix between the two frames. Rotation matrix, generally indicated by \mathbf{R} , is a 3×3 matrix shown as below:

$$\mathbf{R} = \begin{bmatrix} \cos\theta\cos\phi & \sin\psi\sin\theta\cos\phi - \cos\psi\sin\phi & \cos\psi\sin\theta\cos\phi + \sin\psi\sin\phi \\ \cos\theta\sin\phi & \sin\psi\sin\theta\sin\phi + \cos\psi\cos\phi & \cos\psi\sin\theta\sin\phi - \sin\psi\cos\phi \\ -\sin\theta & \sin\psi\cos\theta & \cos\psi\cos\theta \end{bmatrix} \quad (5)$$

where ϕ , θ , and ψ represent the rotation angel around x-axis, y-axis and z-axis respectively.

There are many ways to recover the rotation matrix. In this paper, we used Singular Value Decomposition (SVD) [11] method due to its easy implementation. The basic principle of SVD is to decompose a matrix (defined as \mathbf{H} , in this particular application \mathbf{H} is a 3×3 square matrix) into 3 separate matrices:

$$[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{SVD}(\mathbf{H}) \quad (6)$$

$$\mathbf{H} = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (7)$$

where the columns of \mathbf{U} are the left singular vectors, \mathbf{S} has singular values and is diagonal and \mathbf{V}^T has rows that are the right singular vectors. The SVD represents an expansion of the original data in a coordinate system where the covariance matrix is diagonal.

The next step involves accumulating a matrix, called \mathbf{H} . One thing needs to be pointed out during this step is the re-center of both dataset so that both centroids can be placed at the origin, like shown below:

$$\mathbf{H} = \sum_{i \in N} (P_A(i) - \bar{A})(P_B(i) - \bar{B})^T \quad (8)$$

$$\bar{A} = \frac{1}{N} \sum_{i=1}^N P_A(i) \quad (9)$$

$$\bar{B} = \frac{1}{N} \sum_{i=1}^N P_B(i) \quad (10)$$

where P_A and P_B are corresponding point sets in first frame and second frame respectively in $[x, y, z]^T$ style. This step removes the translation component, leaving only the rotation component to deal with. After \mathbf{H} is factorized by SVD, the rotation matrix can be calculated by multiplication of \mathbf{V} and \mathbf{U}^T :

$$\mathbf{R} = \mathbf{V} \mathbf{U}^T \quad (11)$$

where \mathbf{R} can be expressed in the following form:

$$\mathbf{R} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \quad (12)$$

From the matrix \mathbf{R} , the rotation angle ψ around the optical axis, as illustrated in Fig. 5, can be calculated by:

$$\psi = \arctan\left(\frac{R_{32}}{R_{33}}\right) \quad (13)$$

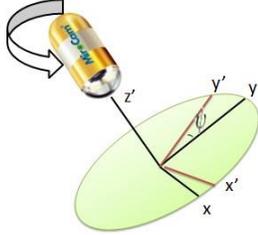


Fig.5. Rotation of the capsule around main optical axis

III. EXPERIMENTAL RESULTS

According to the clinical data provided by Umass Memorial Hospital, the average transit time of the capsule from the duodenum to the cecum was 240 ± 40 min. Our experimental results show that average 2.4 peristalsis were detected per min from the endoscopic video and each peristalsis takes up around 6 ± 2 seconds. In terms of time, around 20% of the time the similarity between consecutive images drops below 65%, which means the capsule proceeding very fast propelled by the peristalsis. Around 30% of the time the similarity between consecutive images stays beyond 75%, which means the capsule either stays still or rotates very slowly. Fig.6 shows a sample video clip of 50 seconds. During this transit time, two peristalsis were detected.

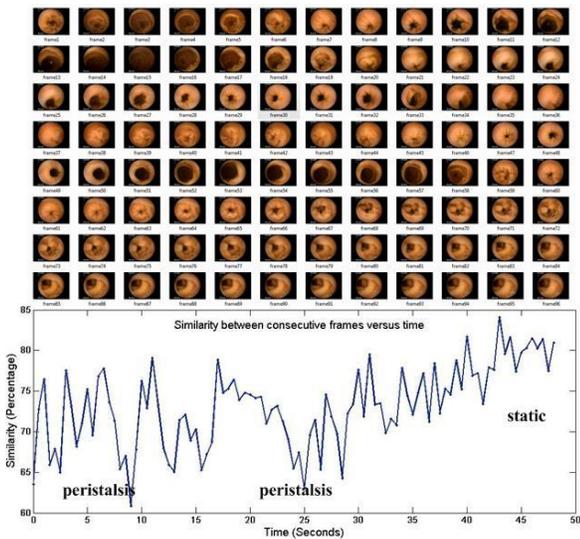


Fig.6. Speed estimation according to the similarity between consecutive images

Fig. 7 shows two consecutive endoscope images and the corresponding feature points found by the algorithm introduced in section II. Overall 52 feature points were detected in the example shown in Fig.6 where circles indicate the original positions of the feature points and lines pointed to the rotated positions in the next frame. The calculated results for the rotation matrix and rotation angle are shown in table 1.

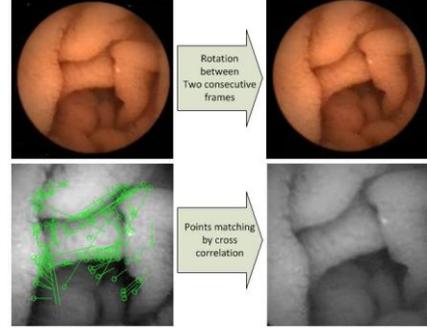


Fig.7. Rotation Estimation by feature matching between two consecutive images

Table 1. Calculated parameters for the example shown in Fig. 7

Feature Points Detected	Rotation Matrix	Translation Matrix	Rotated Angle
52	$\begin{bmatrix} 0.9635 & 0.2677 & 0 \\ -0.2677 & 0.9635 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$T = \begin{bmatrix} -14.8879 \\ 19.3272 \\ 0 \end{bmatrix}$	15.5354

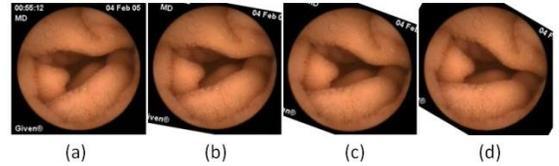


Fig.8. (a) original frame (b) rotated by 10° (c) rotated by 20° (d) rotated by 30°

Table 2. Calculated parameters for the example shown in Fig. 8

Real Rotation Angel	Feature Points Detected	Rotation Matrix	Calculated Rotation Angle	Error
10°	79	$\begin{bmatrix} 0.9852 & -0.1713 & 0 \\ 0.1713 & 0.9852 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	9.8697°	0.1303
20°	18	$\begin{bmatrix} 0.9244 & -0.3814 & 0 \\ 0.3814 & 0.9244 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	22.4324°	2.4324
30°	10	$\begin{bmatrix} -0.9065 & -0.4222 & 0 \\ 0.4222 & 0.9065 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	180.0913°	collapse

To verify the accuracy of the results, we artificially generated a set of rotated frames with different angles from the

same endoscopic image. The rotated angles are 10° , 20° and 30° respectively. From Table 2 it can be seen that when the rotation angle is small ($<30^\circ$), the calculated results are very accurate with an average error below 1.2° . However, as the rotation angle goes up, the accuracy drops down, finally the algorithm collapses at the rotation angle $>30^\circ$.

IV. CONCLUSION

In this paper, we presented a novel image processing based approach to analyze the movement of the endoscopy capsule inside human body. The major contribution of this work includes introducing the concept of “3D map” into the localization inside human body and modeling the movement of endoscopy capsule. The proposed technique is very easy to implement, low cost, and with high accuracy. No extra device is needed for this technique other than the video camera itself. The experimental results show that the proposed speed and rotation estimation methods have a good performance especially when the capsule moves slowly. In the future, we will focus on refining this algorithm according to the clinical data and combining this technique with the existing RF localization approaches to provide a hybrid solution to the localization inside human body.

ACKNOWLEDGEMENT

The authors would like to thank Dr. David Cave for his precious suggestions and the colleagues at the CWINS laboratory for their directly or indirectly help in preparation of the results presented in this paper.

REFERENCE

[1] D. Faigel and D. Cave, “Capsule Endoscopy”, November 29, 2007

[2] L. Liu, C. Hu, W. Cai and M.Q.-H. Meng. “Capsule endoscope localization based on computer vision technique,” Engineering in Medicine and Biology Society, Annual International Conference of the IEEE. pp. 3711-3714, Nov. 2009

[3] F. De Lorio, C. Malagelada, F. Azpiroz, M. Maluenda, C. Violanti, L. Igual, J. Vitrià, J.-r. Malagelada. “Intestinal motor activity, Endoluminal Motion and Transit”, Neurogastroenterology & Motilit, Vol. 21, Issue 12, pp. 1264-1271, Dec. 2009

[4] K. Pahlavan, G. Bao, Y. Ye, et al. Rf localization for wireless video capsule endoscopy. International Journal of Wireless Information Networks, 2012, 19(4): 326-340.

[5] K. Pahlavan and P. Krishnamurthy, “Principles of Wireless Networks”, Prentice Hall, 2002

[6] S. Andrei, L. Thomas, S. Ariel and K. Leif. “On-the-fly Curve-skeleton Computation for 3D Shapes”, Computer Graphics Forum, Vol. 26, Issue 3, pp. 323-328, Sep. 2007

[7] C. Poh, Z. Zhang, Z. Liang, L. Li and J. Liu. “Feature selection and classification for Wireless Capsule Endoscopic frames,” International Conference on Biomedical and Pharmaceutical Engineering, 2009. ICBPE '09, pp. 1-6, Dec. 2009

[8] B. Li, M.Q.-H. Meng, “2010 IEEE International Conference on Capsule Endoscopy Images Classification by Color Texture and Support Vector Machine,” Automation and Logistics (ICAL), pp. 126-131, Aug. 2010

[9] C. Poh, T. Htwe, L. Li, W. Shen, J. Liu, J. Lim, K. Chan and P. Tan. “Multi-level Local Feature Classification for Bleeding Detection in Wireless Capsule Endoscopy Images”, Cybernetics and Intelligent Systems (CIS), 2010 IEEE Conference on, pp. 76-81, June 2010

[10] J. N. Sarvaiya, S. Patnaik and S. Bombaywala, “Image Registration by Template Matching Using Normalized Cross-Correlation”, International Conference on Advances in Computing, Control, & Telecommunication Technologies, 2009. ACT '09, pp. 819-822, 28-29 Dec. 2009

[11] M. Jun, K. Parhi and E. Deprettere, “A Unified Algebraic Transformation Approach for Parallel Recursive and Adaptive Filtering and SVD Algorithms”, Signal Processing, IEEE Transactions on, Vol.49, Issue.2, pp. 424-437, Feb. 2001