Capsule Endoscope Localization Based on Computer Vision Technique

Li Liu, Chao Hu, Wentao Cai

Shenzhen Institute of Advanced Integration Technology (SIAIT),Chinese Academy of Sciences/Chinese University of Hong Kong
Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
Key Lab for Biomedical Informatics and Health Engineering, CAS

li.liu1@sub.siat.ac.cn

{ chao.hu, wt.cai}@siat.ac.cn,

Max Q.-H. Meng

Department of Electronic Engineering
The Chinese University of Hong Kong
Shenzhen Institute of Advanced Integration Technology, CAS/CUHK
Shatin, N.T. Hong Kong

max@ee.cuhk.edu.hk

Abstract - **To build a new type of wireless capsule endoscope with interactive gastrointestinal tract examination, a localization and orientation system is needed for tracking 3D location and 3D orientation of the capsule movement. The magnetic localization and orientation method produces only 5 DOF, but misses the information of rotation angle along capsule's main axis. In this paper, we presented a complementary orientation approach for the capsule endoscope, and the 3D rotation can be determined by applying computer vision technique on the captured endoscopic images. The experimental results show that the complementary orientation method has good accuracy and high feasibility.**

Index Terms - Capsule Endoscope, Localization and Orientation, Feature Tracking, Rotation Decomposition

I. INTRODUCTION

One great breakthrough in the endoscopic techniques over last 20 years is the advent of "wireless capsule endoscope" [1] named as M2A, announced in May, 2000, by Israel scientist G. Iddan et al. When swallowed by the patient, it travels through the gastrointestinal (GI) tract propelled by peristalsis. More than 50,000 images are taken over its 7-hour journey. Comparing with the traditional GI examination endoscopes, the wireless capsule endoscope has its advantages, which makes it possible for doctor to inspect the patient's entire GI tract without causing much pain [2]. However, the current type of capsule endoscope still has its limitation. It can only send out the endoscopic images, but can not provide additional pathological information, such as the position and direction of the diseases, the size of the pathological tissues, and etc. Thereby it makes it difficult for the doctor to realize more accurate diagnose and medical operation (or other medical measures). Therefore, it is desirable to develop a new type of capsule endoscope with some motion abilities, such as the view-angle adjustment, the illumination control, the locomotion control, and etc. To reach this objective, one key technique is to accurately track the position and movement of target capsule endoscope [4, 5].

To address this problem, a system for tracking the capsule has been built by using magnetic sensor array (as shown in

Fig.1) [3, 4]. In the system, a small magnet is enclosed in the capsule endoscope. As the magnetic field intensity originating from the magnet is related to the distance, the magnetic sensors are disposed outside the human body, to detect the magnetic signal from the magnet. With the sampled signals of these sensors, the capsule's location and orientation can be calculated through an appropriate algorithm based on the mathematic model of the magnet's magnetic field.
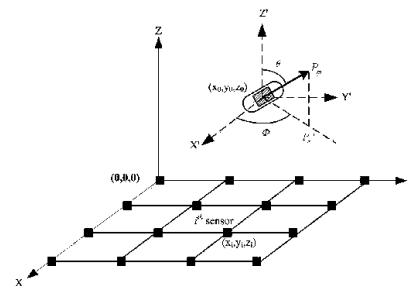

Fig.1 Localization and Orientation System of Wireless Capsule Endoscope

However, this magnetic localization system provides only five degrees of freedom (three for translation and two for rotation) [5], and one rotation message is missing. This rotation information is helpful for locating the pathologic tissue in the GI tract, the 3D recovery of the GI tract, and the feedback control of the external actuation on the capsule. Here, we propose a supplemental localization method by computer vision technique. This technique does not require any additional device, but only the images presented by the capsule endoscope.

The rest of this paper is organized as follows: In section II, we propose the Supplemental Localization and Orientation Approach based on computer vision technique, including the correspondent feature point selection and their tracking, the Epipolar Geometry Analysis by using 8-point algorithm, and the rotation decomposition by quaternion. In section III, we present the experimental results and give some analysis, followed by the conclusion given in section IV.

II. SUPPLEMENTAL LOCALIZATION AND ORIENTATION APPROACH

Because the capsule endoscope changes position and direction very slow, some identical areas exist in the

successive two endoscopic images, so we can find the correspondent point pairs in these two images. Using the image correspondences, we can determine the motion (rotation and translation) parameters of capsule endoscope with an appropriate algorithm. This approach can be a complementary method for improving the magnetic localization and orientation method. Typically, the proposed approach can be divided into several steps:

1) Measuring capsule's magnetic signals and compute its 3D localization and 2D orientation parameters by applying the magnetic localization and orientation algorithm.
2) Adjusting the capsule's camera parameters on the captured images, e.g., the nonlinear distortion correction. This step is based on the procedure of the camera calibration.
3) Establishing feature correspondence for all pairs of the consecutive endoscopic images.
4) Estimating motion and rotation parameters of capsule endoscope based on the correspondent point pairs.
5) Recurring the step 3) and 4) from the calculated motion and rotation parameters to improve the accuracy.

This computer vision method results in a relative orientation with a rotation matrix from one camera's direction changing to another one in true three dimensions.

A. Correspondent Feature Points Tracking

Some feature abstraction algorithms as SIFT is not well suited for endoscope video tracking since endoscopic images are smooth (as shown in Fig.2) and the corresponding points are difficult to be determined.

Fig.2 Endoscopic Image

Optical Flow is a method for extracting the apparent motion in a sequence of images. There are different implementations of optical flow. An iterative implementation of the Lucas-Kanade optical flow computation (KLT) provides sufficient local tracking accuracy. Correspondent feature tracking of capsule endoscope image is the new application of conventional KLT algorithm.

Suppose A and B be two consecutive endoscopic images. $A\left(\vec{x}\right) = A(x,y)$ and $B\left(\vec{x}\right) = B(x,y)$ are then the grayscale value of the two images in location $\vec{x} = [x \quad y]^T$, where x and y are the two pixel coordinates of a generic image point $\vec{x}$. For clarity purposes, denote displacement vector $\bar{V} = [v_x, v_y]^T$, as well as the image position vector $\vec{p} = [p_x, p_y]^T$. The goal of correspondent feature points tracking is to find the displacement vector $\bar{v} = [v_x, v_y]^T$ that minimizes the matching function:

$$\varepsilon(\bar{v}) = \varepsilon(v_x, v_y) = \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} \left(A(x,y) - B(x+v_x, y+v_y)\right)^2. \quad (1)$$

At the optimum, the first derivative of ε with respect to $\bar{V}$ is zero:

$$\left.\frac{\partial \varepsilon(\bar{v})}{\partial \bar{v}}\right|_{\bar{v}=\bar{v}_{opt}} = \begin{bmatrix} 0 & 0 \end{bmatrix} \quad (2)$$

After expansion of the derivative, we obtain:

$$\frac{\partial\varepsilon(\bar{v})}{\partial \bar{v}} = -2 \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} \left(A(x,y) - B(x+v_x, y+v_y)\right) \bullet \left[\frac{\partial B}{\partial x}, \frac{\partial B}{\partial y}\right]. \quad (3)$$

Let us now substitute $B(x+v_x, y+v_y)$ by its first order Taylor expansion about the point $\bar{v} = [0,0]^T$:

$$\frac{\partial\varepsilon(\bar{v})}{\partial \bar{v}} \approx -2 \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} \left(A(x,y) - B(x,y) - \left[\frac{\partial B}{\partial x}, \frac{\partial B}{\partial y}\right]\right) \bullet \left[\frac{\partial B}{\partial x}, \frac{\partial B}{\partial y}\right]. \quad (4)$$

Observe that the quantity $A(x,y) - B(x,y)$ can be interpreted as the temporal image derivative:

$$\forall(x,y) \in [p_x - w_x, p_x + w_x] \times [p_y - w_y, p_y + w_y],$$

$$\delta I(x,y) \cong A(x,y) - B(x,y). \quad (5)$$

The matrix $\left[\frac{\partial B}{\partial x}, \frac{\partial B}{\partial y}\right]$ is merely the image gradient vector.

Let's make a slight change of notation:

$$\nabla I = [I_x, I_y]^T \cong \left[\frac{\partial B}{\partial x}, \frac{\partial B}{\partial y}\right]^T. \quad (6)$$

Observe that the image derivatives I_x and I_y may be computed directly from the first image $A(x,y)$ in the $(2w_x + 1) \times (2w_y + 1)$ neighborhood of the point p independently from the second image $B(x,y)$ (the importance of this observation will become apparent later on when describing the iterative version of the flow computation). If a central difference operator is used for derivative, the two derivative images have the following expression:

$$\forall(x,y) \in [p_x - w_x, p_x + w_x] \times [p_y - w_x, p_y + w_y],$$

$$I_x(x,y) = \frac{\partial A(x,y)}{\partial x} = \frac{A(x+1,y) - A(x-1,y)}{2}, \quad (7)$$

$$I_y(x,y) = \frac{\partial A(x,y)}{\partial y} = \frac{A(x,y+1) - A(x,y-1)}{2}. \quad (8)$$

In practice, the Sharr operator is used for computing image derivatives (very similar to the central difference operator). Following this new notation, equation (4) may be written:

$$\frac{1}{2}\left[\frac{\partial \varepsilon(\overline{v})}{\partial \overline{v}}\right]^T \approx \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y}\left(\left[\begin{array}{cc} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{array}\right]\overline{v} - \left[\begin{array}{c} \delta I \bullet I_x \\ \delta I \bullet I_y \end{array}\right]\right). \quad (9)$$

Denote

$$G \cong \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y}\left[\begin{array}{cc} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{array}\right]$$

$$\overline{b} \cong \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y}\left[\begin{array}{c} \delta I \bullet I_x \\ \delta I \bullet I_y \end{array}\right]. \quad (10)$$

Then, equation (9) may be written:

$$\frac{1}{2}\left[\frac{\partial \varepsilon(\overline{v})}{\partial \overline{v}}\right]^T \approx G\overline{v} - \overline{b} \quad (11)$$

From (2), the optimum optical flow vector can be written as

$$\overline{v}_{opt} = G^{-1}\overline{b}. \quad (12)$$

This expression is valid only if the matrix G is invertible. That is meaning that the image $A(x,y)$ contains gradient information in both x and y directions in the neighborhood of the point p.

B. Correspondent Feature Points Selection

Now we need to select the feature point p on A, and this step is called feature selection. Indeed, the central step of tracking is the computation of the optical flow vector $\overline{v}$. At this step, the G matrix is required to be invertible. In other words, the minimum eigenvalue of G must be large enough (larger than a threshold). It characterizes the pixels that are "easy to track". The process of selection goes as follows:

1) Compute the G matrix and its minimum eigenvalue λ_m at every pixel in the image A.

2) Call λ_{max} the maximum value of λ_m over the whole image.

3) Retain the image pixels that have a λ_m value larger than a percentage of λ_{max}. This percentage can be 10% or 5%.

4) From those pixels, retain the pixels with the local max (a pixel is kept if its λ_m value is larger than that of any other pixel in its 3×3 neighborhood).

5) Keep the subset of those pixels so that the minimum distance between any pair of pixels is larger than a given threshold distance (e.g. 10 or 5 pixels).

After that process, the remaining pixels are typically "good to track". They are the selected feature points that are fed to the tracker.

C. Epipolar Geometry Analysis

Because current capsule endoscope has a single camera, the localization and orientation must be realized by the images taken at different camera position or orientation. Epipolar Geometry Analysis can be applied to calculate the localization/orientation parameters by two-projection motion method [6], named as 8-point algorithm [7], which is realized by finding 8 or more correspondent points in the two successive images.
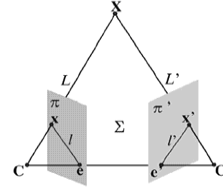

Fig.3 Epipolar Geometry

Let $P = (x,y,z)^T$ be an arbitrary object point before the objective motion and $P' = (x',y',z')^T$ the same object point after the motion. We have $(\overline{x} = x/z, \overline{y} = y/z)$ and $(\overline{x}' = x'/z', \overline{y}' = y'/z')$, the projective coordinates of P and P' onto the image plane; the capsule endoscope's imaging center coordinate be (X_m, Y_m) and its view angle be θ=140º/2=70º, such as Given M2A capsule endoscope here. To apply the 8-point algorithm, we need to change above coordinate represented by pixel number to the coordinate (u,v) relative to focal length (f=1).

$$u = \frac{x - X_m}{Y_m} \bullet \tan 70° \quad \text{and} \quad v = \frac{y - Y_m}{Y_m} \bullet \tan 70°$$

The rigid body motion equation relates P to P' is

$$P' = R \bullet P + T \quad (13)$$

Where R is the 3×3 rotation matrix; T is the translation vector $T = [t_x \quad t_y \quad t_z]^T$.

Denote $E = T \times R$, motion equation can be expressed:

$$(u' \quad v' \quad 1)E(u \quad v \quad 1)^T = 0 \quad (14)$$

E is called the "Essential Matrix", which is composed of rigid body's translation and rotation information through R and T. We can determine rotation matrix R and unit vector T_0 (collinear to T) by finding 8 or more correspondences in the two successive images[7] (discussed in section A and B).
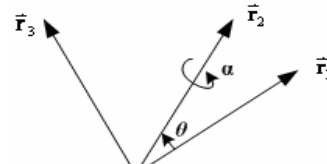
D. Rotation Decomposition Using Quaternion

Now, we need to decompose the rotation gotten by above method into these 2 sub-rotations: one for the rotation of the main axis of the endoscope, and the other one is rotation around the main axis (this rotation can not be determined by the magnetic localization method). For simplifying the problem, we use the quaternion [8] method that represent above two rotations with two sets of quaternion parameters. Using an appropriate algorithm, we can solve these quaternion parameters from the rotation matrix, and compute the two angles corresponding to the two sub-rotations.

If the rotation matrix R is $R = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}$, the quaternion is $\vec{q} = [q_0 \quad q_x \quad q_y \quad q_z]^T$ with

$$q_0 = \sqrt{\frac{1+m_{11}+m_{22}+m_{33}}{4}}; \quad q_x = \frac{m_{32}-m_{23}}{4q_0}; \quad q_y = \frac{m_{13}-m_{31}}{4q_0};$$

$q_z = \frac{m_{21}-m_{12}}{4q_0}$. We can determine the resultant main axis orientation vector $\vec{r}_2 = (m_2 \quad n_2 \quad p_2)^T$ by using initial orientation vector $\vec{r}_1 = (m_1 \quad n_1 \quad p_1)^T$ and the known rotation matrix R (or the quaternion $\vec{q} = \lfloor q_0 \quad q_x \quad q_y \quad q_z \rfloor^t$), and finally obtain the rotation angle α :

Fig.4 Quaternion Representation for Rotation

$$\alpha = 2\tan^{-1}\left(\frac{m_1 q_x + n_1 q_y + p_1 q_z}{q_0}\right) \qquad (15)$$

III. EXPERIMENTAL RESULTS

Fig. 5 shows the endoscope images and the correspondent feature points when given rotation angle around the main axis of the capsule is 5º, 10º, 15º, 20º, 25 º and 30º. Table I shows the calculated results for these angles. We observed that the calculated rotation angle errors are 0.0673°, 0.179°, 0.403°, 1.775°, 0.829°, and an acceptable value.

From the results, we conclude that the results are acceptable when rotation angle is not so large. In other word, successive two endoscopic images must have small variation relative to the capsule's position and orientation; otherwise there will result in a large error. Endoscopic images tested in the experiment are transmitted wireless with 2 fr/s, and the correspondent feature points in consecutive images only have very small variation, so the rotation angle calculation has relative high accuracy.
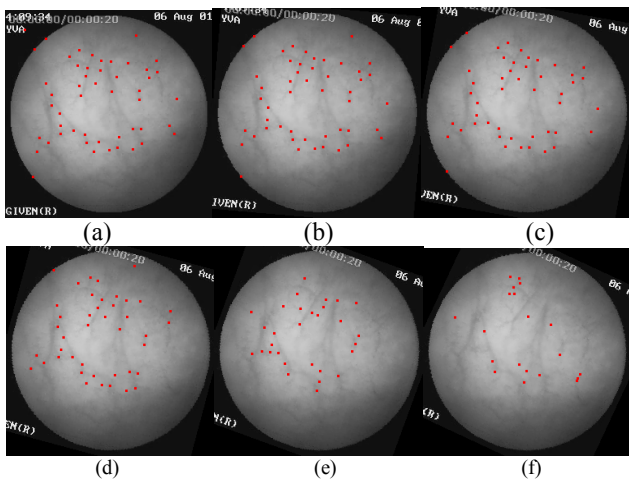

(a)　　　　　(b)　　　　　(c)

(d)　　　　　(e)　　　　　(f)

Fig.5 Ccorrespondent Feature Point Extraction

IV. CONCLUSION

In this paper, we presented a complementary orientation approach for the capsule endoscope. We apply Lucas-Kanade

TABLE I   TEST RESULTS OF ROTATION ANGLE ALONG CENTRAL AXIS

Given Rotation angle	Rotation Matrix			Calculated Rotation angle	Error
5°	$\begin{bmatrix} 0.9963 & -0.0860 & -0.0048 \\ 0.0860 & 0.9963 & -0.0042 \\ 0.0052 & -0.0038 & 1.0000 \end{bmatrix}$			4.9327°	0.067°
10°	$\begin{bmatrix} 0.9852 & -0.1707 & -0.0122 \\ 0.1704 & 0.9850 & -0.0256 \\ 0.0164 & 0.0232 & 0.9996 \end{bmatrix}$			9.8207°	0.179°
15°	$\begin{bmatrix} 0.9665 & -0.2515 & -0.0505 \\ 0.2522 & 0.9677 & -0.0068 \\ 0.0472 & -0.0193 & 0.9987 \end{bmatrix}$			14.5972°	0.403°
20°	$\begin{bmatrix} 0.9249 & -0.3701 & -0.0871 \\ 0.3704 & 0.9288 & -0.0130 \\ 0.0857 & -0.0203 & 0.9961 \end{bmatrix}$			21.7749°	1.775°
25°	$\begin{bmatrix} 0.8885 & -0.3789 & 0.2590 \\ 0.4221 & 0.8961 & -0.1372 \\ -0.1802 & 0.2312 & 0.9561 \end{bmatrix}$			24.1724°	0.829°
30°	$\begin{bmatrix} 0.9654 & -0.0602 & -0.2537 \\ -0.1184 & 0.9681 & -0.2206 \\ 0.2323 & -0.2430 & 0.9418 \end{bmatrix}$			**-0.52784°**	**Big Error**

optical flow computation to track the featured points on the captured endoscopic images, then use 8-point algorithm to obtain the localization/orientation parameters, and finally decompose the rotation by the quaternion method. The experimental results show that the complementary orientation method has good accuracy and high feasibility.

In the near future we will further improve it by finding a reasonable calibration technique to determine the camera intrinsic parameters; correcting the camera's distortion; and improving the matching algorithm.

REFERENCES

[1] F. Gong, P. Swain, and T. Mills, "Wireless endoscopy", Gastrointestinal Endoscopy. Vol.51, No.6, pp725~729, June, 2000
[2] M. Q.-H. Meng, T. Mei, J. Pu, C. Hu, X. Wang,Y. Chan, "Wireless Robotic Capsule Endoscopy: State-of-the-Art and Challenges", Proceedings of the 5th World Congress on Intelligent Control and Automation, June 15-19, 2004, pp.5561-5565.
[3] W.Andra, H. Danan, W. Kirnbe, H.-H. Kramer, P. Saupe, R. Schmieg, and M. E Bellemann, "A novel method for real-time magnetic marker monitoring in the gastrointestinal tract", Physics in Medicine and Biology, 45(2000), 3081-3093.
[4] C. Hu, M. Q.-H. Meng, M. Mandal, "The Calibration of 3-Axis Magnetic Sensor Array System for Tracking Wireless Capsule Endoscope", Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, October 9-15, 2006,pp.162-167
[5] C. Hu, M. Q.-H. Meng. M. Mandal, "Efficient Linear Algorithmfor Magnetic Localization and Orientation in Capsule Endoscopy", Proceedings of the 2005 IEEE, Engineering in medicine and Biology 27th Annual Conference Shanghai, China, September 1-4, 2005. 162-167
[6] H. C. Longuet-Higgins, "A Computer Algorithm for Reconstructing a Scene From Two Projections," Nature, vol. 293, pp. 133–135, Sept 1981.
[7] R. I. Hartley, "In Defense of the Eight-Point Algorithm", IEEE Trans. On Pattern Analysis and Machine Intelligence. Vol.19, No.6, pp580~593, June 1997.
[8] C. Hu, M. Q.- H. Meng, and M. Mandal, "Robot Rotation Decomposition Using Quaternions". Proceedings of the 2006 IEEE, International Conference on Mechatronics and Automation, Luoyang, China, June 25 - 28, 2006, pp.1158~1163.