MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28-30, 2000, The University of Tokyo, Japan

10—2

Camera Motion Estimation Based on Real Endoscope Images and CT Images for Endoscope Navigation System

*Jun Sugiyama [†], Hidenori Shoji [†], Kensaku Mori [†], Yasuhito Suenaga [†], Jun-ichiro Toriwaki [†], and Jun-ichi Hasegawa [‡]

[†] Graduate School of Engineering, Nagoya University
[‡] School of Computer and Cognitive Science, Chukyo University

Abstract

This paper describes a method for camera motion estimation based on real endoscope images and CT images for an endoscope navigation system. We have developed an endoscope navigation system that shows the navigation information during an endoscope examination. In this system, we need to estimate the real endoscope camera motion. Our method tracks the camera motion by using an image-based registration technique. We applied the proposed method to the video images of the real bronchoscopy and 3-D X-ray CT images. Experimental results showed that the method could track the camera motion corresponding to the real endoscope correctly.

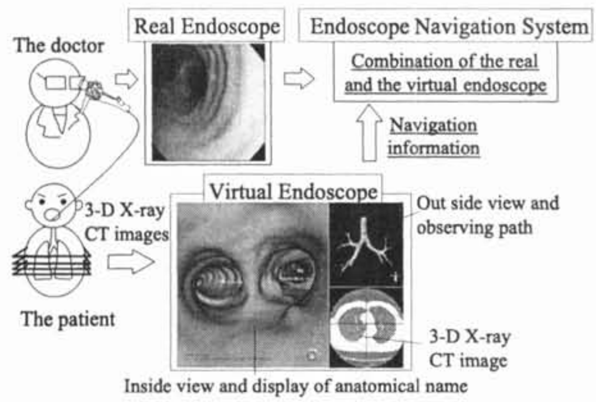

Figure 1: Endoscope navigation system

1 Introduction

An endoscope, commonly used in the medical field, is an equipment that consists of a flexible or rigid tube equipped with a small camera. A medical doctor can insert it into the inside of the human body to observe the inner wall of tube-shaped organs. However, it is hard to obtain the information of the organ such as blood vessels that are located behind the target organ wall. It is also difficult to measure the size of a tumor. Usually the doctor performs the endoscope insertion and the endoscopic biopsy only by observing the live-video images and using his anatomical knowledge. This may hurt the organ that is located under the observing organ wall during the endoscopic examination or the endoscopic biopsy.

Virtual endoscopy (VE) [1] is one of visualization techniques of 3-D medical images. The virtual endoscope system (VES) can generate endoscopic views (virtual endoscopic image) basing upon 3-D volumetric images. The VE has a lot of advantages against the real endoscope. A user can observe a target organ from arbitrary viewpoints and view directions. This system can also visualize not only the information of the target organ but also the information of the organ existing behind the target organ wall by employing a translucent rendering. Combination of the virtual endoscopy and the real endoscopy may assist the endoscopic examination if the virtual endoscope system can display the information that exactly corresponds to the current observation. The combined system, (we call it "the endoscope navigation system") , presents useful navigation information to the doctor during the examination or the operation. Figure 1 shows the concept of the endoscope navigation system.

*Address: Furo-cho, Chikusa-ku, Nagoya 464-8603 Japan. E-mail: jsugi@suenaga.cse.nagoya-u.ac.jp

2 Image based tracking

A fundamental function of the endoscope navigation system is to track or estimate the camera motion of the real endoscope in real time. However this is a difficult problem, since we can not attach a 3-D position sensor at the tip of the endoscope due to the limitation of its size, especially in the case of the flexible endoscope. Our method tracks the camera motion by using the image based registration technique for solving this dilemma. The basic idea is to find an observation parameter of VES that can generate a virtual endoscopic image that is similar to the current frame of the video endoscopic image. This process is executed for all frames of the real video endoscope. Tracking for each frame consists of two stages: (a) tracking by using texture information, and, (b) tracking by using shading information. When the real endoscope is located at a straight tube area, there is little difference of shading pattern between adjacent two frames. In such area, we can roughly estimate the camera motion from change of the organ's wall texture. In the registration process, texture information of the organ's wall is estimated from the previous real endoscopic video image and is mapped to a 3-D shape of the organ. The process finds an observation parameter of a texture mapped virtual endoscopic image that is the most similar to the current real endoscopic frame. This process assumes that the estimated observation parameter of the previous is correct. If the estimated parameter of the previous frame is incorrect, the tracking method using the texture information may cause a wrong estimation result. To avoid this problem, we also use a tracking method that uses only shading information of the bronchus.

3 Method

3.1 Observation parameters of VES

Figure 2 shows the endoscope observation parameters which consist of the camera position, the camera posture, and the lighting parameter. Observation parameter $\mathbf{Q} = (\mathbf{p}, \mathbf{w})$ means the camera position $\mathbf{p} = (x, y, z)$ and the camera posture $\mathbf{w} = (w_x, w_y, w_z)$ represented as Euler angle.

3.2 Camera motion estimation

The following describes the camera motion estimation method. Figure 3 shows the flow chart of this method. The inputs of our method are real endoscopic video images and X-ray CT images of the corresponding patient. We represent a frame of the endoscopic video as $\mathbf{R}_k$ (k means the frame number) and a virtual endoscopic image as $\mathbf{V}$. In the tracking process, VES generates several endoscopic images for finding an observation parameter. An observation parameter of the first real endoscopic frame is adjusted manually.

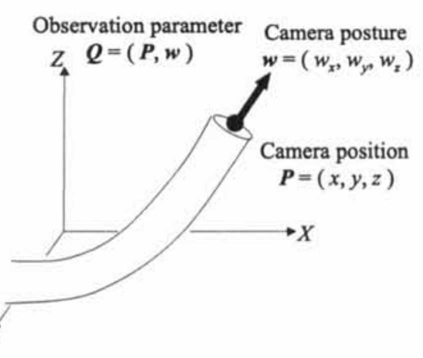

Figure 2: Observation parameter of endoscope

3.2.1 Estimation using texture information

When the method performs an estimation for a real endoscopic image $\mathbf{R}_k$, we use a real endoscopic image $\mathbf{R}_{k-1}$ as a texture image for virtual endoscopic images in the registration process of $\mathbf{R}_k$ and map this texture image to a virtual endoscopic image. The 3-D shape of the organ is represented by a set of triangular patches in VES. Texture mapped virtual endoscopic images are generated by assigning a texture image captured from a triangular region of $\mathbf{R}_{k-1}$ to each triangular patch. The triangular region is found by projecting vertices of the triangular patch onto the projection plane at the estimated observation parameter of $\mathbf{R}_{k-1}$. This technique assumes that the estimation was performed correctly at the previous frame. Then we execute the next process, that is, to find $\mathbf{Q}$ which gives

$$\max_{\mathbf{Q}} \mathbf{E}_1(\mathbf{R}_k, \mathbf{V}_{tex}(\mathbf{Q})). \quad (1)$$

This process means that we search the parameter space to find the observation parameter maximizing $\mathbf{E}_1$. We generate several texture-mapped virtual endoscopic images with changing observation parameters, and measure the similarity $\mathbf{E}_1$ between $\mathbf{R}_k$ and a virtual endoscopic images $\mathbf{V}_{tex}(\mathbf{Q})$. This process regards $\mathbf{Q}_{tex}$ that takes the highest similarity value

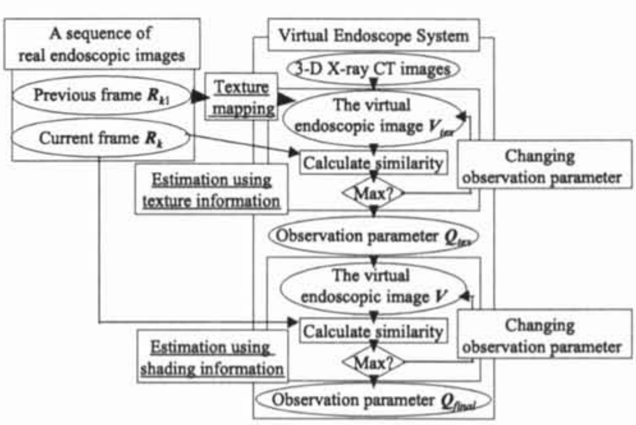

Figure 3: Processing flow of camera motion estimation

as the estimated parameter of the real endoscope. We employ the correlation for measuring similarity $\mathbf{E}_1$ and the Powell's method[2] in the maximization operation.

3.2.2 Estimation using shading information

The initial observation parameter of the virtual endoscopy in this estimation is the result of the estimation using texture information. This registration process generates virtual endoscopic images with only a shading technique and without a texture mapping technique. We consider that the light intensity is inversely proportional to the square of the distance between the light source and a point on the object surface in VES. The attenuation factor is manually adapted so as to the lighting condition in real endoscopic images. This process performs basically in the same way as the estimation using texture information except for the type of virtual endoscopic images (non-Texture mapped images are used here) and the function for measuring the similarity. We execute the operation,

$$\max_{\mathbf{Q}} \mathbf{E}'(\mathbf{R}_k, \mathbf{V}(\mathbf{Q})), \tag{2}$$

where $\mathbf{V}$ means a non texture mapped virtual endoscopic image at the observation parameter $\mathbf{Q}$, $\mathbf{R}_k$ the real endoscopic image at the frame k, and $\mathbf{E}'$ the function for measuring similarity. $\mathbf{E}'$ is defined here as

$$\mathbf{E}'(\mathbf{R}_k, \mathbf{V}(\mathbf{Q})) = \frac{\mathbf{E}_1(\mathbf{R}_k, \mathbf{V}(\mathbf{Q}))}{\mathbf{E}_2(\mathbf{R}_k, \mathbf{V}(\mathbf{Q}))}, \tag{3}$$

where $\mathbf{E}_2$ is the mean square error between two images. This estimation also regards the observation parameter $\mathbf{Q}_{final}$ that takes the highest similarity value $\mathbf{E}'$ as the final estimated observation parameter for $\mathbf{R}_k$.

4 Result

We evaluated the proposed method by applying it to real video bronchoscopic images and 3-D chest X-ray CT images. Three cases were tested. Table 1 shows the specification of CT images. Real endoscopic video images were recorded on the digital video tape during examinations. The video was captured frame by frame and inputted into the computer. The average processing time was about 30 seconds per frame. In the real examination, it is hard to know the real endoscope camera motion, so we evaluated the tracking results by comparing real and virtual endoscopic images frame by frame with our observations. Figure 4 shows the results of the camera motion tracking. From this figure we can confirm that our method could track the camera motion about correctly. We also applied our previous method based on optical flow patterns[3] and the method based on only shading information to the same real video bronchoscopic images and 3-D chest X-ray CT images. Table 2 shows the number of frames that were estimated correctly by those methods. The method proposed here was successful in tracking camera motion for about 550 frames.

Table 1: Specification of CT images

3-D X-ray CT image	Case A	Case B	Case C
No. of pixel	512x512	512x512	512x512
Thickness of slice(mm)	5	3	2
Reconstruction pitch(mm)	1	2	1
Slices	183	72	76

Table 2: Tracking results

Case	No. of frame	No. of successful frames		
		Shading	Optical-flow	This method
A	544	232	285	544
	110	12	110	53
	451	51	349	206
B	300	93	190	139
C	191	89	128	96
	192	50	146	102
	149	119	149	113

5 Discussion

The camera motion tracking by our method was successful for the real endoscopic frames including apparent shapes like tumors. By using texture information, the proposed method could track camera motion correctly in the frames where there are few shading features. However, it may fail at frames where bubbles exist in front of the endoscope camera or when the endoscope camera is too close to the organ wall. Once the tracking of camera motion fails, tracking of the following frames is also likely to fail.

Though the optical flow based method gives the best result for tracking the camera motion for the area of the tube-like parts having few shading features, it estimates only forward or backward motion of the camera. The proposed method using texture information estimates all the camera motions successfully. Regarding the tracking accuracy itself, our former optical-flow based method is better than the proposed method. The texture based method can be improved further by considering the frequency of mapping texture and the similarity measure in the estimation using texture.

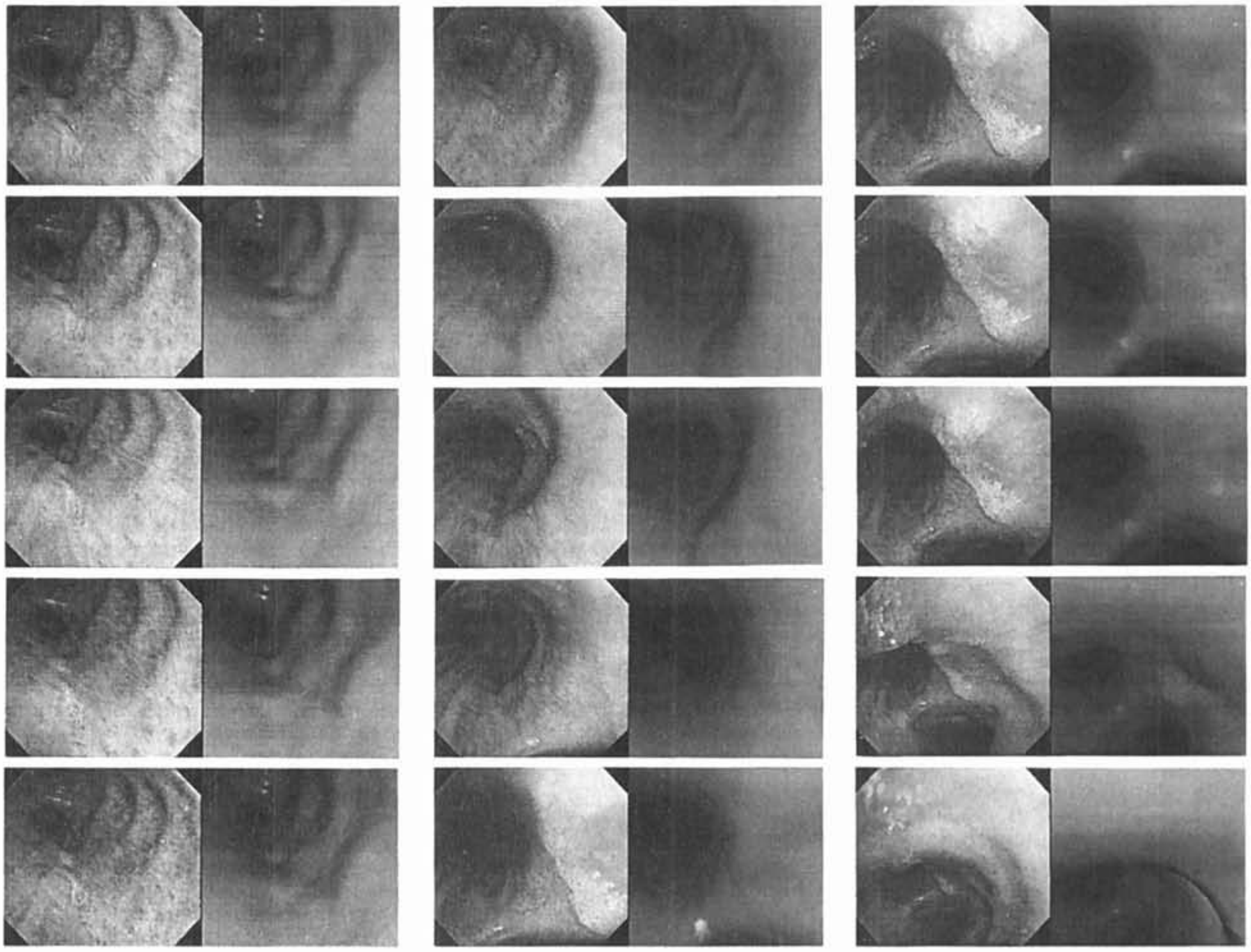

Figure 4: Tracking results

Left figures of each column are real endoscopic images. Right ones are corresponding virtual endoscopic images that were rendered by using result of the tracking.

6 Conclusion

This paper proposed a new method for the camera motion estimation of a real endoscope for the endoscope navigation system, which uses the texture information. Experimental results showed that the method could track the camera motion successfully by using the image based registration technique. Future work includes to consider the frequency of texture mapping and the similarity measure in the estimation using texture. We need to improve the tracking accuracy of scenes where bubbles appear in the endoscopic frame or when the endoscopic camera is very close to the organ wall. It is also desired to evaluate the method by large number of cases.

Acknowledgments. The authors would like to thank Dr.K.Katada of Fujita Health University, Dr.H.Takabatake of Hokkaido Keiai-kai Minami Ichi-jyo Hospital, Dr.H.Natori of Sapporo Medical University for providing CT images, and clinical validation of our system. We would also like to thank members of Suenaga laboratory and Toriwaki laboratory for useful suggestions and discussions. Parts of this research were supported by the Grant-In-Aid for Scientific Research from the Ministry of Education and the Grant-In-Aid for Scientific Research from Japan Society for Promotion of Science.

References

[1] K.Mori, A.Urano, J.Hasegawa, et al., "Virtualized Endoscope System -an Application of Virtual Reality Technology to Diagnostic Aid", IEICE Transactions on Information and Systems, E79-D, pp.809-819, 1996.

[2] William H.Press, Saul A. Teukolsky, William T. Vetterling, et al., "Numerical recipes in C: the art of scientific computing -2nd ed.", Cambridge University Press, Melbourne, 1996.

[3] K.Mori, Y.Suenaga, J.Toriwaki, et al., "A method for tracking camera motion of real endoscope by using virtual endoscopy system", Proceedings of SPIE Medical Imaging Vol. 3978, pp.122-133, 2000.