New Image Similarity Measure for Bronchoscope Tracking Based on Image Registration

Daisuke Deguchi[1], Kensaku Mori[1], Yasuhito Suenaga[1], Jun-ichi Hasegawa[2], Jun-ichiro Toriwaki[2], Hirotsugu Takabatake[3], and Hiroshi Natori[4]

[1] Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan
{ddeguchi,mori,suenaga}@suenaga.cse.nagoyap-u.ac.jp
[2] School of Computer and Cognitive Sciences, Chukyo University, Toyota, Japan
[3] Minami-ichijyo Hospital, Sapporo, Japan
[4] School of Medicine. Sapporo Medical University, Sapporo, Japan

Abstract. This paper presents new image similarity measure for bronchoscope tracking based on image registration between real and virtual endoscopic images. A function for bronchoscope tracking is one of the fundamental functions in a bronchoscope navigation system. Since it is difficult to attach a positional sensor at the tip of the bronchoscope due to the space limitation, image registration between real endoscopic (RE) and virtual endoscopic (VE) images becomes a strong tool for bronchoscopic camera motion tracking. The summing-type image similarity measuring methods including mean squared error or mutual information could not properly estimate the position and orientation of the endoscope, since the outputs of these methods do not change significantly due to averaging. This paper proposes new image similarity measure that effectively uses characteristic structures observed in bronchoscopic views in similarity computation. This method divides the original image into a set of small subblocks and selects only the subblocks in which characteristic shapes are seen. Then, an image similarity value is calculated only inside the selected subblocks. We applied the proposed method to eight pairs of X-ray CT images and real bronchoscopic videos. The experimental showed much improvement in continuous tracking performance. Nearly 1000 consecutive frames were tracked correctly.

1 Introduction

Flexible endoscopes, such as colonoscopes or bronchoscopes, are tools for observing the insides of human bodies. A bronchoscope equips a tiny camera at the tip of a flexible tube. A medical doctor inserts the bronchoscope into a patient body with watching a TV monitor that shows video frames captured by the camera. The doctor operates the bronchoscope by referring to his anatomical knowledge. There is no guidance system that provides navigation information.

Virtual endoscopy (VE) is now widely used for visualizing the inside of a human body [1]. The user of a Virtual Endoscopy System (VES) can fly-thorough the inside of a target organ freely by using a mouse. The VES can visualize not

R.E. Ellis and T.M. Peters (Eds.): MICCAI 2003, LNCS 2878, pp. 399–406, 2003.

only the surface of the target organ's wall but also the anatomical structures existing beyond the target organ's wall by employing semi-translucent display. It is also possible to overlay anatomical names on VE images or to show quantitative measurement results. If we could fuse real endoscopy (RE) and VE, it would be possible to provide useful information, such as important organs beyond the organ's wall being currently observed or the path to the desired location for biopsy, during an bronchoscopic examination or treatment. To implement such navigation system for a bronchoscope, we should register the coordinate systems of RE and VE. Some positional sensors should be attached to the endoscope to obtain camera positions and orientations since the flexible endoscope can be bent into arbitrary forms. However, it is difficult to attach a positional sensor at the tip of the endoscope. Although wire-type positional sensors, which can be inserted into a human body through an endoscopic channel, are available, their outputs are very unstable due to magnetic interference. Image registration can become a quite useful technique for camera motion tracking of the bronchoscope. Tracking is achieved by finding rendering parameters that generate the most similar VE image to the current RE frame.

We have proposed a method for tracking bronchoscopic camera motion that uses epipolar geometry analysis and image registration [2]. In this method, epipolar geometry analysis is utilized for rough estimation of camera motion by solving epipolar equations. Then, precise estimation is performed by image registration. Image similarity between RE and VE images is calculated by summing gray-level differences up for all pixels of two images. However, this method could not estimate the positions and orientations of the RE camera properly, when image similarity does not change significantly due to averaging. Bricault et al. [3] reported the pioneering work in registration of RE and VE images. They aimed to construct a system for assisting transbronchial biopsy. Their method computes the camera position of a real endoscope near areas where bifurcations can be seen. The structure of the bronchial tree of the same patient was extracted from a CT image. Their method, however, has difficulty in estimating the camera position in areas where no bifurcation appears. Also, because their method uses bifurcations and the branching structure of the bronchial tree, it is not easily applicable to other organs such as the colon. Higgins et al. [4] also reported preliminary work on an endoscope navigation system. The work of both of these groups considers only the static registration of RE and VE images.

This paper presents a new method for measuring image similarity in image registration between RE and VE images. The proposed method divides an RE image into a set of subblocks and selects the subblocks that contain characteristic shapes such as folds in the computation process of image similarity. Then, an image similarity value is calculated only inside the selected subblocks. In Section 2, we show the detail of the computation process of the proposed image similarity measure. Brief description of the bronchoscope camera motion tracking process is also provided. Section 3 presents the experimental results of the proposed and previous methods. We discuss the proposed method and the obtained results in the same section.

2 Method

2.1 Overview

The entire process of bronchoscope tracking consists of four major steps: (1) select an RE frame from the RE video, (2) generate a VE image from the CT image taken before an examination, (3) find the rendering parameter that generates the most similar VE image to the selected RE image based on image similarity, and (4) proceed to the next frame.

2.2 Image Similarity Measure

The image similarity measuring process consists of the following four steps: (a) image dividing, (b) feature value computation, (c) subblock selection, and (d) image similarity computation.

Division of an Input RE Image. Let $\mathbf{B}^{(k)}$ be the k–th frame of an RE video, and the variable of $\mathbf{V}$ be a VE image. The size of an RE frame $\mathbf{B}^{(k)}$ is $W \times H$ (pixels). We generate VE images whose sizes are equal to the real ones. The RE image $\mathbf{B}^{(k)}$ is divided into $M \times N$ small subblocks. A subblock $D_{m,n}$ of the m–th row and the n–th column is defined as

$$D_{m,n} = \left\{ (p\,,\,q); (n-2)\frac{W}{M} \le p < (n+1)\frac{W}{M}, (m-2)\frac{H}{N} \le q < (m+1)\frac{H}{N} \right\}, \quad (1)$$

where m and n range $2 \le m \le M-1\,,\ \ 2 \le n \le N-1$.

Feature Value Computation. For each subblock $D_{m,n}$, we compute two types of feature values : (a) standard deviation $\sigma_{\mathbf{B}^{(k)}_{D_{m,n}}}$ and (b) local mean squared error $LoMSE(D_{m,n})$. The standard deviation, $\sigma_{\mathbf{B}^{(k)}_{D_{m,n}}}$, of the subblock, $D_{m,n}$, is given by

$$\sigma_{\mathbf{B}^{(k)}_{D_{m,n}}} = \sqrt{\frac{1}{|D_{m,n}|} \sum_{(i,j)\in D_{m,n}} \left(\mathbf{B}^{(k)}_{i,j} - \overline{\mathbf{B}^{(k)}_{D_{m,n}}} \right)^2}, \quad (2)$$

where $|D_{m,n}|$ is the number of pixels inside the $D_{m,n}$, and $\overline{\mathbf{B}^{(k)}_{D_{m,n}}}$ is the mean intensity inside the region $D_{m,n}$. The $LoMSE$ value of the subblock $D_{m,n}$ is calculated as

$$LoMSE(D_{m,n}) = \sum_{\Delta x, \Delta y} \frac{1}{|D_{m,n}|} \sum_{(i,j)\in D_{m,n}} \left(\frac{\mathbf{B}^{(k)}_{i,j} - \overline{\mathbf{B}^{(k)}_{D_{m,n}}}}{\sigma_{\mathbf{B}^{(k)}_{D_{m,n}}}} - \frac{\mathbf{B}^{(k)}_{i+\Delta x, j+\Delta y} - \overline{\mathbf{B}^{(k)}_{D'_{m,n}}}}{\sigma_{\mathbf{B}^{(k)}_{D'_{m,n}}}} \right)^2, \quad (3)$$

where $D'_{m,n} = \{(i+\Delta x, j+\Delta y)\ ;\ (i,j) \in D_{m,n}\}$, $\overline{\mathbf{B}^{(k)}_{D_{m,n}}}$ and $\overline{\mathbf{B}^{(k)}_{D'_{m,n}}}$ are the mean intensities inside the subblocks $D_{m,n}$ and $D'_{m,n}$, $|D_{m,n}|$ is the number of pixels inside the $D_{m,n}$. $(\Delta x\,,\,\Delta y)$ can take the combination of

$$(\Delta x\,,\,\Delta y) = \left\{ \begin{array}{ccccc} (-\frac{W}{2H}\,,\,-\frac{W}{2H}), & (0\,,\,-\frac{W}{2H}), & (\frac{W}{2H}\,,\,-\frac{W}{2H}), & (-\frac{W}{2H}\,,\,0), & \\ (\frac{W}{2H}\,,\,0), & (-\frac{W}{2H}\,,\,\frac{W}{2H}), & (0\,,\,\frac{W}{2H}), & (\frac{W}{2H}\,,\,\frac{W}{2H}) & \end{array} \right\}. \tag{4}$$

Subblock Selection. A subblock $D_{m,n}$ that satisfies either of the following two conditions is selected. The selected subblcoks are appended to the list $A^{(k)}$ which holds the selected subblocks for $\mathbf{B}^{(k)}$,

$$\left(\sigma_{\mathbf{B}^{(k)}_{D_{m,n}}} \geq T_{SD_1}\right) \wedge \left(\neg\Big(LoMSE(D_{m,n}) \geq T_{LoMSE_2}\Big)\right), \tag{5}$$

$$\left(LoMSE(D_{m,n}) \leq T_{LoMSE_1}\right) \wedge \left(\neg\Big(\sigma_{\mathbf{B}^{(k)}_{D_{m,n}}} \leq T_{SD_2}\Big)\right), \tag{6}$$

where the symbol $\neg$ means NOT, T_{SD_1}, T_{LoMSE_1}, and T_{LoMSE_2} are threshold values.

Image Similarity Computation. As image similarity between $\mathbf{B}^{(k)}$ and $\mathbf{V}$, we compute an image similarity value called modified mean squared error ($MoMSE$) by

$$MoMSE\Big(\mathbf{B}^{(k)}\,,\,\mathbf{V}\Big) = \frac{1}{|A^{(k)}|}\sum_{D\in A^{(k)}}\frac{1}{|D|}\sum_{(i,j)\in D}\Big((\mathbf{B}^{(k)}_{i,j} - \overline{\mathbf{B}^{(k)}_{D}}) - (\mathbf{V}_{i,j} - \overline{\mathbf{V}_D})\Big)^2, \tag{7}$$

where $|A^{(k)}|$ is the number of subblocks stored in $A^{(k)}$, and $\overline{\mathbf{B}^{(k)}_{D}}$ and $\overline{\mathbf{V}_D}$ are the mean intensities inside the subblock D of $\mathbf{B}^{(k)}$ and $\mathbf{V}$.

2.3 Bronchoscope Camera Motion Tracking Based on Image Registration

We input an RE video and a 3-D chest X-ray CT image to the tracking process. Bronchoscope tracking is achieved by sequentially obtaining the extrinsic camera parameters, $\mathbf{Q}^{(k)} = (\mathbf{P}^{(k)}, \mathbf{w}^{(k)})$ (k is the frame number), for the total amount of input RE video frames. Here, $\mathbf{P} = \left(p_x^{(k)}\,,\,p_y^{(k)}\,,\,p_z^{(k)}\right)$ is the bronchoscopic camera position and $\mathbf{w} = \left(w_x^{(k)}\,,\,w_y^{(k)}\,,\,w_z^{(k)}\right)$ is the orientation. These parameters are represented in the coordinate system of the input CT image so that the obtained parameters can generated the most similar VE images to the real ones. The

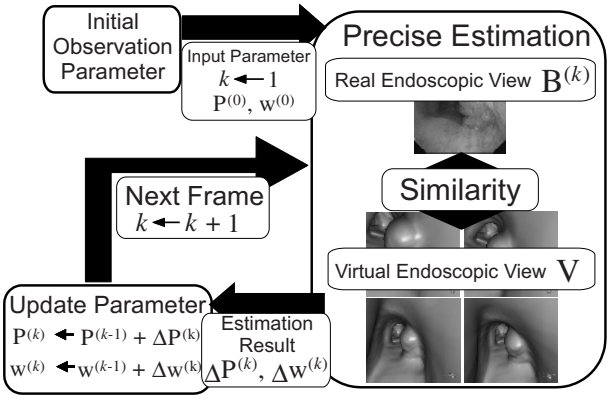

Fig. 1. Process flow.

tracking is performed by sequentially finding the parameter by using the finding result of the previous frame. The best parameter for each frame is calculated as the parameter that minimizes image similarity $MoMSE\left(\mathbf{B}^{(k)}, \mathbf{V}\right)$ between the RE frame $\mathbf{B}^{(k)}$ and the VE image $\mathbf{V}$ rendered by it. The tracking process for each frame is formulated as

$$\mathbf{Q}^{(k)} = \arg \min_{\mathbf{Q}^{(k)}} MoMSE\Big(\mathbf{B}^{(k)}, \mathbf{V}(\mathbf{Q}^{(k)})\Big). \tag{8}$$

The Powell method is employed here for executing the above minimization process. Volume rendering method is utilized here for generating VE images. Fast software-based volume rendering module presented in Ref. [5] is used for obtaining VE images that have less artifacts. To improve the speed of the registration process, the ray casting process of volume rendering is performed only inside the selected subblocks. Figure 1 shows the entire flow of the tracking process.

3 Experimental Results and Discussion

The proposed method was implemented on a conventional PC platform (CPU: AMD dual Athlon MP 1900+ processors, 2GByte main memory). We applied the proposed method to eight pairs of X-ray CT images and real bronchoscopic video images for evaluating the efficiency of the proposed method. Bronchoscopic videos were recorded onto digital videotapes in operation rooms during examinations and transferred to the host computer. We divided a video frame into $M \times N = 30 \times 30$ subblocks. Acquisition parameters of CT images are : 512×512 pixels, $72 \sim 209$ slices, $2 \sim 5$ mm collimation, and $1 \sim 2$ mm reconstruction pitch. We presently performed the tracking process as an off-line job. Evaluation of the proposed method is performed in three ways. The first method (Method I) uses the image registration which employs mean squared error for image similarity measure. The second one (Method II) performs tracking

Table 1. Results of endoscopic camera motion tracking. Method I uses image registration by employing the image similarity computed by mean squared error. Method II uses epipolar geometry analysis and image registration based on mean squared error. Method III employs the proposed image similarity.

		Case 1				Case 2	Case 3		
		A	B	C	D	A	A	B	C
Video Frame Size ($W \times H$) (pixels)		362 × 370				362 × 370	362 × 370		
Number of Frames		544	500	320	200	430	973	873	50
Number of Successive Frames	I	370	99	56	100	93	240	140	12
	Method II	544	199	254	183	198	240	680	50
	III	395	500	116	180	407	973	873	50

Case 4	Case 5			Case 6		Case 7		Case 8
A	A	B	C	A	B	A	B	A
362 × 370	362 × 370			362 × 370		256 × 253		256 × 253
200	400	800	400	200	200	200	205	500
66	150	60	50	190	140	141	205	142
85	231	60	149	192	301	146	205	142
69	300	658	200	140	140	149	205	282

by using both epipolar geometry analysis and image registration based on mean squared error. In this method, the tracking process roughly estimates the camera movement by solving epipolar equations based on the corresponding point pairs on the two consecutive RE images. Then, image registration is performed as precise estimation. The third one (Method III) uses the proposed image similarity measure for image registration. Tracking performance was evaluated by counting the number of successive frames that were tracked correctly by our visual inspection. Processing time of the proposed method for one frame was four seconds in average. Table 1 presents the tracking results. In this table, 'Case 1A', 'Case 1B', 'Case 1C', and 'Case 1D', for example, represent different video clips of a same patient. Examples of the tracking results are also shown in Fig. 2. In this figure, the left columns are frames selected from a sequence of RE frames. The columns C–I, C–II, and C–III present the subblocks selected by the conditions $\sigma_{\mathbf{B}_D^{(k)}} \geq T_{SD_1}$, $LoMSE(D) \leq T_{LoMSE_1}$, and $LoMSE(D) \geq T_{LoMSE_2}$.

The proposed image similarity measure showed great success in tracking as seen in the results of Case 1A, Case 2A, Case 3A, and Case 3B, Method III, which employs the proposed image similarity measure. The previous method could not estimate the camera motion appropriately because it could not catch the change of the characteristic shapes of RE frames such as folds. In contrast, the proposed method can estimate camera motion by computing the similarity only in the subblocks which characteristic shapes are observed.

The Methods I and II took about fifteen seconds to process one frame. In contrary to these results, the Method III only needs four seconds to process one frame. This is because VE image generation and image similarity computation are performed only inside the selected subblocks. This elimination much

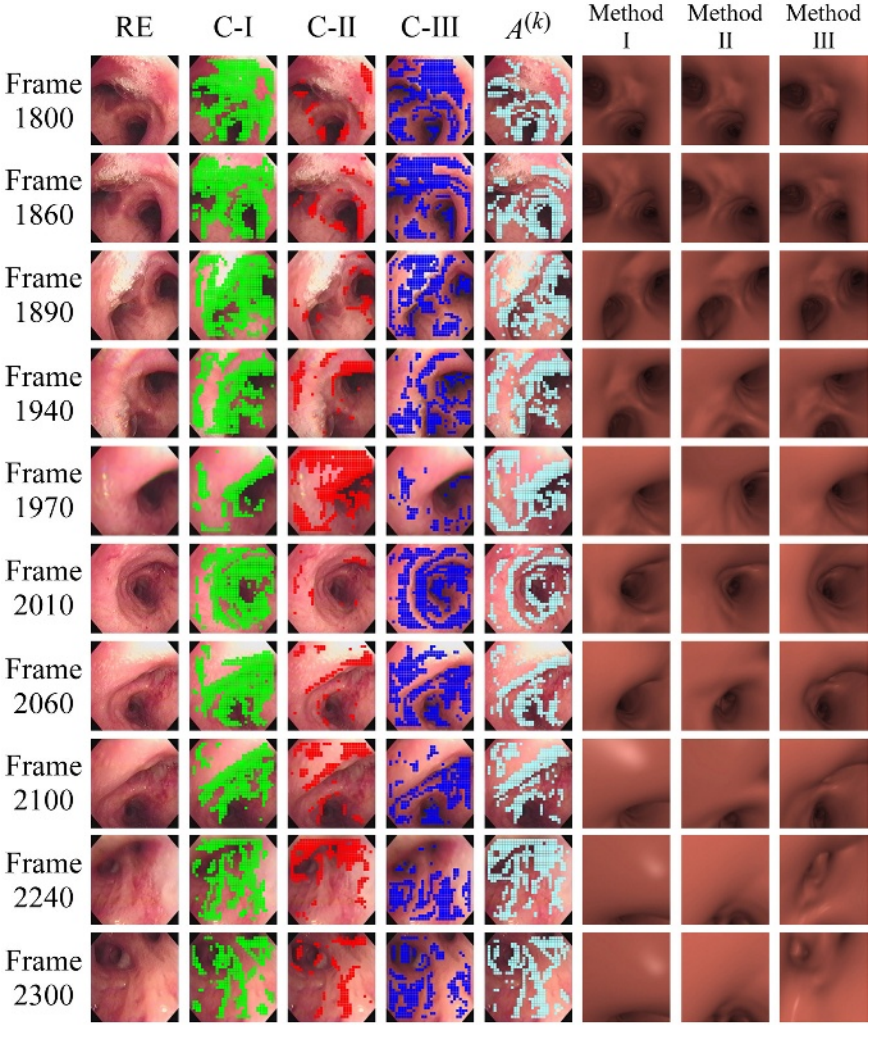

Fig. 2. Tracking results of the bronchoscope camera motion of Case 1B. The left column shows a sequence of the selected RE frames. Frame numbers are also presented. Columns of C–I, C–II, and C–III show the conditions of $\sigma_{\mathbf{B}_D^{(k)}} \geq T_{SD_1}$, $LoMSE(D) \leq T_{LoMSE_1}$, $LoMSE(D) \geq T_{LoMSE_2}$ for selecting subblocks. The right columns show VE images generated by using the estimated observation parameters. The results of Method I, II, and III are displayed on the right side. Method I uses the image registration by employing the image similarity computed by mean squared error. Method II performs tracking by using both epipolar geometry analysis and image registration based on mean squared error. Method III uses the proposed image similarity measure for image registration.

contributed to the reduction of processing time. The proposed image similarity computation scheme showed a very sharp minimum at the registered point in comparison with the previous method. This also improved the computation time.

4 Conclusion

This paper presented a new image similarity measure for bronchoscope tracking based on image registration between real and virtual endoscopic images. The proposed image similarity effectively used characteristic structures observed in RE images in the similarity computation. We applied the proposed method to eight pairs of X-ray CT images and real bronchoscopic videos. The experimental showed significant improvement in continuous tracking performance. Future work includes: (a) evaluation by the large number of cases, (b) development of quantitative evaluation method, (c) development of a stable method for selecting characteristic regions, and (d) reduction of computation time.

Acknowledgments. The authors thank to our colleagues for their useful suggestions and discussions. D. Deguchi and K. Mori thank to Dr. Calvin R. Maurer, Jr. for his many useful comments and suggestions.Parts of this research were supported by the 21st century COE program, the Grant-In-Aid for Scientific Research from Japan Society for Promotion of Science, and the Grant-In-Aid for Cancer Research from the Ministry of Health and Welfare of Japanese Government.

References

1. P. Rogalla, J. Terwisscha van Scheltinga, B. Hamm, eds., "Virtual endoscopy and related 3D techniques", Springer, Berlin, 2001
2. K.Mori, D.Deguchi, J.Sugiyama, et al., "Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images", Medical Image Analysis, 6, pp. 321–336, 2002
3. I. Bricault G. Ferretti, P. Cinquin, "Registration Real and CT-Derived Virtual Bronchoscopic Images to Assist Transbronchial Biopsy", IEEE Trans. on Medical Imaging, 17, 5, pp. 703–714, 1998
4. J.P.Helferty, W.E.Higgins, "Technique for Registering 3D Virtual CT Images to Endoscopic Video", Proceedings of ICIP (International Conference on Image Processing), pp. 893–896, 2001
5. K.Mori, Y.Suenaga and J. Toriwaki, "Fast volume rendering based on software optimization using multimedia instructions on PC platform", Proceedings of Computer Assisted Radiology and Surgery (CARS)2002, pp. 467–472, 2002