

# Fast and Accurate Bronchoscope Tracking Using Image Registration and Motion Prediction

Jiro Nagao<sup>1</sup>, Kensaku Mori<sup>1</sup>, Tsutomu Enjouji<sup>1</sup>, Daisuke Deguchi<sup>1</sup>,  
Takayuki Kitasaka<sup>1</sup>, Yasuhito Suenaga<sup>1</sup>, Jun-ichi Hasegawa<sup>2</sup>,  
Jun-ichiro Toriwaki<sup>2</sup>, Hirotsugu Takabatake<sup>3</sup>, and Hiroshi Natori<sup>4</sup>

<sup>1</sup> Graduate School of Information Science, Nagoya University, Japan  
{jnagao,ddeguchi,kitasaka,mori,suenaga}@suenaga.m.is.nagoya-u.ac.jp

<sup>2</sup> School of Computer and Cognitive Sciences, Chukyo University, Japan

<sup>3</sup> Sapporo Minami-sanjyo Hospital, Sapporo, Japan

<sup>4</sup> Department of Diagnostic Ultrasound and Medical Electronics,  
Sapporo Medical University, Japan

**Abstract.** This paper describes a method for faster and more accurate bronchoscope camera tracking by image registration and camera motion prediction using the Kalman filter. The position and orientation of the bronchoscope camera at a frame of a bronchoscopic video are predicted by the Kalman filter. Because the Kalman filter gives good prediction for image registration, estimation of the position and orientation of the bronchoscope tip converges fast and accurately. In spite of the usefulness of Kalman filters, there have been no reports on tracking bronchoscope camera motion using the Kalman filter. Experiments on eight pairs of real bronchoscopic video and chest CT images showed that the proposed method could track camera motion 2.5 times as fast as our previous method. Experimental results showed that the motion prediction increased the number of frames correctly and continuously tracked by about 4.5%, and the processing time was reduced by about 60% with the search space restriction also proposed in this paper.

## 1 Introduction

A bronchoscope is a tool to observe inside the bronchi. A physician inserts a bronchoscope into a patient's airway with watching a TV monitor. During an examination or treatment with a bronchoscope, it sometimes happens that the physician gets disoriented because the lung has complex bifurcation structures. As the body of a bronchoscope can be bent, a positional sensor has to be attached to the tip of it if the position of the tip should be directly measured. However, because bronchi are very narrow, it is hard to attach a positional sensor to the tip. Even if you use one of electro-magnetic types, the measurement does not have sufficient accuracy for the navigation use. To support the physician during a bronchoscopic examination by providing information such as the position and orientation of the bronchoscope tip, we have reported a bronchoscope tracking method [1,2]. This method tracks bronchoscopic camera motion by image registration (IR) between real bronchoscopic (RB) images and virtual bronchoscopic

(VB) images generated from 3-D CT images [3,4]. This method tracks bronchoscopic camera motion by sequentially registering RB and VB images frame by frame. Although this method is robust to patient motion caused by breathing, it requires large amount of computation time. Since bronchoscope navigation should work in real time, efficient tracking algorithm is desired.

This paper proposes a fast and accurate bronchoscope tracking algorithm. Since the tracking is based on IR that finds the best camera parameter of VB that maximizes image similarity between RB and VB images, good initial guess of the camera parameter can reduce computation time. For obtaining good initial guess, we predict bronchoscopic camera motion by the Kalman filter [5]. Several research groups are working on developing bronchoscope navigation or bronchoscope tracking [6,7]. However, there exists no report on motion prediction in bronchoscope tracking. In Sect. 2, the proposed method for camera motion tracking is described. The experimental results and a brief discussion are shown in Sect. 3.

## 2 Camera Motion Tracking

### 2.1 Camera Parameters of Virtual Bronchoscope

The extrinsic camera parameters of the VB camera at the  $k$ -th frame are denoted by:

$$\mathbf{Q}^{(k)} = \begin{pmatrix} \mathbf{R}(\mathbf{r}^{(k)}) & \mathbf{t}^{(k)} \\ \mathbf{t}^{\mathbf{0}} & 1 \end{pmatrix}, \quad (1)$$

where  $\mathbf{r}^{(k)} = {}^t(r_x \ r_y \ r_z)$  is the camera orientation represented in Euler angle around the  $x$ -,  $y$ -, and  $z$ -axes in the world coordinate system and  $\mathbf{R}(\mathbf{r}^{(k)})$  is a rotation matrix constructed from  $\mathbf{r}^{(k)}$ .  $\mathbf{t}^{(k)} = {}^t(t_x \ t_y \ t_z)$  means the camera position.  $\mathbf{Q}^{(k)}$  is obtained from  $\mathbf{q}^{(k)} = {}^t({}^t\mathbf{r}^{(k)} \ t\mathbf{t}^{(k)})$ . It is this parameter vector  $\mathbf{q}$  that the method we will describe later directly seeks for. When the camera moves from the  $k$ -th frame to the  $(k+1)$ -th frame, the movement is formulated as:

$$\mathbf{Q}^{(k+1)} = \mathbf{Q}^{(k)} \Delta\mathbf{Q}^{(k)}, \quad \mathbf{Q}^{(k)} = \begin{pmatrix} \mathbf{R}^{(k)} & \mathbf{t}^{(k)} \\ \mathbf{t}^{\mathbf{0}} & 1 \end{pmatrix}, \quad \Delta\mathbf{Q}^{(k)} = \begin{pmatrix} \mathbf{R}(\Delta\mathbf{r}^{(k)}) & \Delta\mathbf{t}^{(k)} \\ \mathbf{t}^{\mathbf{0}} & 1 \end{pmatrix}, \quad (2)$$

where  $\Delta\mathbf{Q}^{(k)}$  is the movement between the  $k$ -th and  $(k+1)$ -th frames. Hence, bronchoscope camera tracking is achieved by continuously finding  $\Delta\mathbf{Q}^{(k)}$  constructed from  $\Delta\mathbf{q}^{(k)} = {}^t({}^t\Delta\mathbf{r}^{(k)} \ t\Delta\mathbf{t}^{(k)})$ .

### 2.2 Overview of Tracking Algorithm

The parameters of an RB camera are estimated by means of IR between the RB images and virtual bronchoscopic images. The input of the system is a 3-D chest CT image and a sequence of real bronchoscopic images  $\{\mathbf{B}^{(k)}\}$ , both of which

are obtained from the same patient. The CT image is used for generating VB images. The output is the sequence of tracking results  $\{\mathbf{q}^{(k)}\}$ .

The tracking method consists of two major components: (a) camera motion prediction by a Kalman filter (Motion Prediction Step) and (b) camera parameter estimation by IR (Registration Step). In order to reduce computation time, a restricted search space is configured before each Registration Step. When the camera motion at the 0-th through the  $k$ -th frames are tracked, the camera motion tracking process for the  $(k+1)$ -th frame is performed as follows:

1. **(Motion Prediction Step)** The camera parameter prediction  $\hat{\mathbf{q}}^{(k+1)}$  for the  $(k+1)$ -th frame, which is used in the succeeding Registration Step, is predicted by the Kalman filter. The Kalman filter outputs  $\hat{\mathbf{q}}^{(k+1)}$  using  $\mathbf{q}^{(0)}, \dots, \mathbf{q}^{(k)}$  acquired in the preceding Registration Steps.
2. Configure a restricted search space using  $\hat{\mathbf{q}}^{(k+1)}$ .
3. **(Registration Step)**  $\mathbf{q}_0^{(k+1)} = \hat{\mathbf{q}}^{(k+1)}$  is used as the initial search parameter for IR. Here,  $\mathbf{q}_0^{(k+1)}$  denotes the initial search parameter for IR at the  $(k+1)$ -th frame. A series of VB images  $\{\mathbf{V}(\mathbf{Q}^{(k)} \Delta \mathbf{Q}^{(k)})\}$  is generated with changing  $\Delta \mathbf{Q}^{(k)}$  that constructs  $\Delta \mathbf{Q}^{(k)}$ . The parameter generating the VB image that is the most similar to  $\mathbf{B}^{(k+1)}$  is selected as the camera parameter estimation result  $\mathbf{Q}^{(k+1)}$ .

The above process is iterated for each of the subsequent RB frames. We assume that the camera parameters  $\mathbf{q}_0^{(0)}$  at the starting frame  $\mathbf{B}^{(0)}$  are known.

### 2.3 Camera Motion Prediction by Kalman Filter

**Kalman Filtering.** The Kalman filter is formulated as follows:

$$\mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}_k + \mathbf{G}\mathbf{w}_k \quad (\text{state equation}) , \quad (3)$$

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{n}_k \quad (\text{observation equation}) , \quad (4)$$

where  $\mathbf{x}_k$  is a system state vector at time step  $k$  and  $\mathbf{y}_k$  is an observation vector. Although  $\mathbf{F}$ ,  $\mathbf{G}$ ,  $\mathbf{H}$ , the state transition, driving and observation matrices, respectively, can change in time, they are assumed to be invariant in our model. The covariance matrices of Gaussian noises  $\mathbf{n}_k$  and  $\mathbf{w}_k$ ,  $\Sigma_{\mathbf{n}_k}$  and  $\Sigma_{\mathbf{w}_k}$ , respectively, are assumed to be known.

*The Kalman Filter Algorithm.* Hereafter,  $\hat{\mathbf{x}}_{t|t'}$  and  $\hat{\Sigma}_{t|t'}$ ;  $((t, t') = (k, k-1), (k, k), (k+1, k) \text{ or } (0, -1))$  denote the estimated values of  $\mathbf{x}_t$  and covariance matrix of  $\mathbf{x}_t$ , respectively, given the observation signals  $\mathbf{y}_0, \dots, \mathbf{y}_{t'}$ . When  $\mathbf{y}_k$  is received,  $\hat{\mathbf{x}}_{k+1|k}$  is predicted by the two steps described below. The iteration starts with  $k=0$ , and the starting initial values are  $\hat{\mathbf{x}}_{0|-1}$  and  $\hat{\Sigma}_{0|-1}$ .

[STEP 1] Estimate  $\hat{\mathbf{x}}_{k|k}$  and  $\hat{\Sigma}_{k|k}$  from  $\mathbf{y}_k$ ,  $\hat{\mathbf{x}}_{k|k-1}$  and  $\hat{\Sigma}_{k|k-1}$ .

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \hat{\Sigma}_{k|k-1} {}^t\mathbf{H} \left( \mathbf{H}\hat{\Sigma}_{k|k-1} {}^t\mathbf{H} + \Sigma_{\mathbf{n}_k} \right)^{-1} (\mathbf{y}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1}) , \quad (5)$$

$$\hat{\Sigma}_{k|k} = \hat{\Sigma}_{k|k-1} - \hat{\Sigma}_{k|k-1} {}^t\mathbf{H} \left( \mathbf{H}\hat{\Sigma}_{k|k-1} {}^t\mathbf{H} + \Sigma_{\mathbf{n}_k} \right)^{-1} \mathbf{H}\hat{\Sigma}_{k|k-1} . \quad (6)$$

[STEP 2] Predict  $\hat{\mathbf{x}}_{k+1|k}$  and  $\hat{\Sigma}_{k+1|k}$  using  $\hat{\mathbf{x}}_{k|k}$  and  $\hat{\Sigma}_{k|k}$ .

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{F}\hat{\mathbf{x}}_{k|k} \text{ ,} \tag{7}$$

$$\hat{\Sigma}_{k+1|k} = \mathbf{F}\hat{\Sigma}_{k|k}{}^t\mathbf{F} + \mathbf{G}\Sigma_{\mathbf{w}_k}{}^t\mathbf{G} \text{ .} \tag{8}$$

**Camera Motion Prediction by Kalman Filter.** We assume that the camera motion is in a constant state of acceleration. The motion is written as:

$$\mathbf{v}^{(k)} = \frac{d\mathbf{q}^{(k)}}{dt} = \frac{\mathbf{q}^{(k)} - \mathbf{q}^{(k-1)}}{\Delta t} \text{ ,} \tag{9}$$

$$\mathbf{a}^{(k)} = \frac{d\mathbf{v}^{(k)}}{dt} = \frac{\mathbf{v}^{(k)} - \mathbf{v}^{(k-1)}}{\Delta t} = \frac{\mathbf{q}^{(k)} - 2\mathbf{q}^{(k-1)} + \mathbf{q}^{(k-2)}}{\Delta t} \text{ ,} \tag{10}$$

where  $\mathbf{v}^{(k)}$  and  $\mathbf{a}^{(k)}$  represent the velocity and acceleration, respectively, and  $\Delta t$  the interval between time steps. Using (9) and (10),  $\mathbf{q}^{(k+1)}$  and  $\mathbf{v}^{(k+1)}$  can be rewritten in terms of  $\mathbf{a}^{(k)}$  and  $\mathbf{v}^{(k)}$  as:

$$\mathbf{q}^{(k+1)} = \mathbf{q}^{(k)} + \mathbf{v}^{(k)}\Delta t + \frac{1}{2}\mathbf{a}^{(k)}(\Delta t)^2 \text{ ,} \tag{11}$$

$$\mathbf{v}^{(k+1)} = \mathbf{v}^{(k)} + \mathbf{a}^{(k)}\Delta t \text{ .} \tag{12}$$

Letting  $\mathbf{x}_k = {}^t(\mathbf{q}^{(k)} \ \mathbf{v}^{(k)} \ \mathbf{a}^{(k)})$  and constructing  $\mathbf{F}$  from (11) and (12), we obtain:

$$\mathbf{F} = \begin{pmatrix} \mathbf{Id}_6 & (\Delta t)\mathbf{Id}_6 & \frac{1}{2}(\Delta t)^2\mathbf{Id}_6 \\ \mathbf{0}_6 & \mathbf{Id}_6 & (\Delta t)\mathbf{Id}_6 \\ \mathbf{0}_6 & \mathbf{0}_6 & \mathbf{Id}_6 \end{pmatrix} \text{ ,} \quad \mathbf{w}_k = \begin{pmatrix} \mathbf{w}_q \\ \mathbf{w}_v \\ \mathbf{w}_a \end{pmatrix} \text{ ,} \tag{13}$$

$$\mathbf{G} = (\mathbf{Id}_6 \ \mathbf{Id}_6 \ \mathbf{Id}_6) \text{ ,} \quad \mathbf{H} = (\mathbf{Id}_6 \ \mathbf{0}_6 \ \mathbf{0}_6) \text{ ,}$$

where  $\mathbf{Id}_6$  and  $\mathbf{0}_6$  are  $6 \times 6$  identity and zero matrices, respectively. The predicted camera parameter is given by  $\hat{\mathbf{q}}^{(k+1)} = \mathbf{H}\hat{\mathbf{x}}_{k+1}$  and provided to the Registration Step as the initial search parameter  $\mathbf{q}_0^{(k+1)}$ .

### 2.4 Camera Parameter Estimation by Image Registration

**Image Registration.** The parameters of the bronchoscope camera is estimated by using  $\mathbf{q}_0^{(k+1)} = \hat{\mathbf{q}}^{(k+1)}$  as the initial search parameter. IR for finding camera parameter at the  $(k + 1)$ -th frame is formulated as the maximization problem:

$$\Delta\mathbf{Q}^{(k)*} = \arg \max_{\Delta\mathbf{Q}^{(k)}} S\left(\mathbf{B}^{(k+1)}, \mathbf{V}\left(\mathbf{Q}^{(k)}\Delta\mathbf{Q}^{(k)}\right)\right) \text{ ,} \tag{14}$$

where  $S(\mathbf{B}, \mathbf{V})$  is the image similarity between  $\mathbf{B}$  and  $\mathbf{V}$ . The camera parameter at the  $(k + 1)$ -th frame is obtained by

$$\mathbf{Q}^{(k+1)} = \mathbf{Q}^{(k)}\Delta\mathbf{Q}^{(k)*} \text{ .} \tag{15}$$

This maximization problem is numerically solved by Powell method [8]. The search starts from  $\Delta\mathbf{Q}_0^{(k)}$  obtained from the predicted camera parameter  $\hat{\mathbf{q}}^{(k+1)}$  and  $\mathbf{Q}^{(k)}$ .

**Search Space Restriction.** Also, we restrict the search space by using the prediction results. The restricted search space is calculated by:

$$\left( \left| \mathbf{t}^{(k+1)} - \hat{\mathbf{t}}^{(k+1)} \right| \leq \left| \mathbf{t}^{(k)} - \hat{\mathbf{t}}^{(k+1)} \right| \right) \wedge \left( \left| \mathbf{r}^{(k+1)} - \hat{\mathbf{r}}^{(k+1)} \right| \leq \left| \mathbf{r}^{(k)} - \hat{\mathbf{r}}^{(k+1)} \right| \right), \quad (16)$$

where  $\mathbf{t}^{(k)}$  and  $\mathbf{r}^{(k)}$  are the position and orientation of the bronchoscopic camera defined by (1).

## 3 Experimental Results and Discussion

### 3.1 Experiments

We applied the proposed method to eight pairs of RB video sequences and 3-D CT images. The RB video and the CT image in each pair were obtained from the same patient. The acquisition parameters of the CT images are:  $512 \times 512$  pixels of slice size, 72-209 slices, 2.0-5.0 mm of slice thickness, 1.0-2.0 mm of reconstruction pitch. A conventional PC hosting two Intel XEON 3.06GHz processors and 2GBytes of main memory was used for this experiment. Table 1 and Fig.1 show the tracking results of: (1) Our previous method [2] without motion prediction or search space restriction (abbreviated to “Previous” in the table), (2) the proposed method with motion prediction, without the search space restriction (“MP”), and (3) the proposed method with motion prediction and search space restriction (“MP+SSR”). The tracking results were evaluated in terms of the number of frames correctly and continuously tracked by visual inspection. Continuity of the tracking was judged to end when the VB images generated from the estimated camera parameters looked greatly different from the corresponding RB images for several frames, or when the tracking no longer followed the motion of the RB camera. Figure 1 shows examples of RB images and the corresponding VB images generated from the camera parameters calculated by each method.

### 3.2 Discussion

The number of frames tracked by the “MP” method increased in the six of the fifteen paths used for the experiment, and the total number of tracked frames increased by about 4.5%, reducing the processing time by about 20%. Furthermore, with the search space restriction (i.e. “MP+SSR” method), the computation time was reduced by about 60%, maintaining the larger number of tracked frames compared to those of the “Previous” method. We assumed that the real bronchoscope camera was in a state of constant acceleration. In reality, however, the motion of the camera is not in such a state. Therefore, it is expected that there might be some reduction in the number of frames tracked by the “MP” method; this was the case in the two of the fifteen paths. In this sense, it can be said that there is a tradeoff between speed and accuracy and robustness.

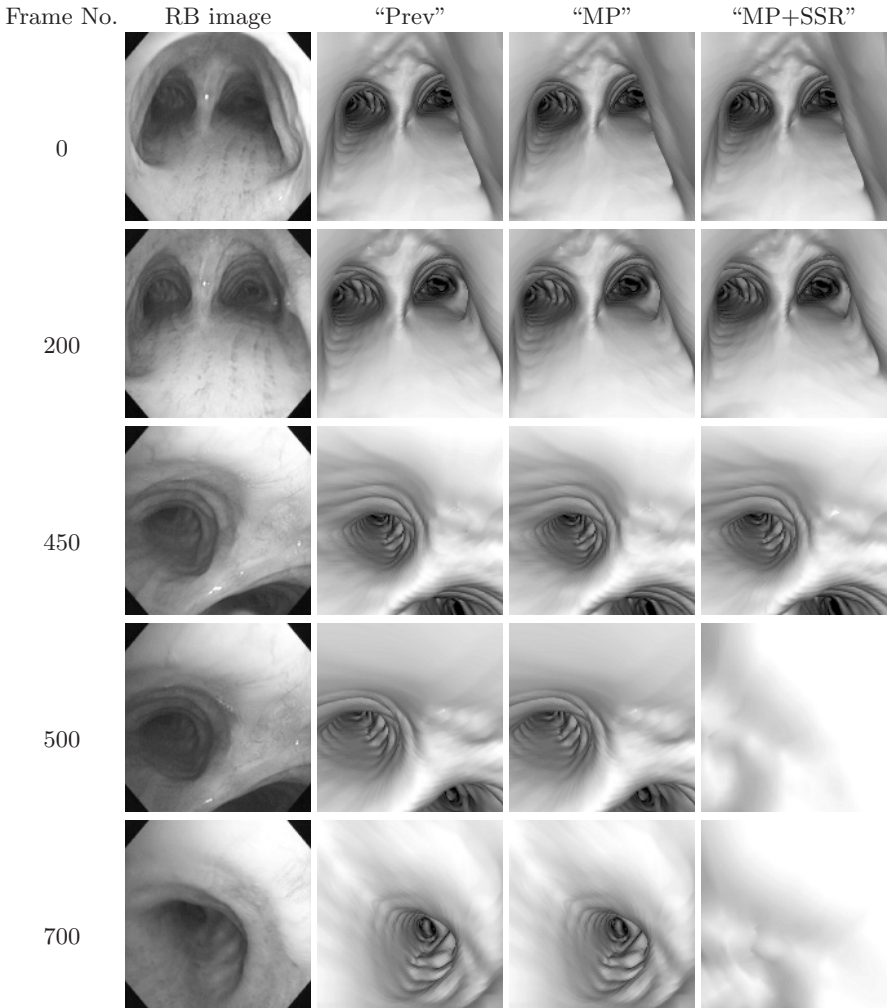
**Table 1.** The tracking results. The results are shown in [frames]. Figures in the parentheses show the average processing time per frame [sec]

Case	Path	Frames	Number of tracked frames (avg. time [sec])		
			Previous	MP	MP+SSR
1	A	500	500 (3.42)	500 (3.21)	500 (2.25)
	B	257	116 (6.82)	106 (2.63)	106 (1.97)
	C	200	180 (7.08)	180 (5.98)	180 (1.85)
2	A	430	407 (3.68)	407 (4.04)	407 (2.66)
3	A	300	255 (6.90)	261 (5.32)	261 (2.54)
	B	1100	973 (4.74)	973 (4.89)	973 (2.97)
	C	1000	873 (9.05)	873 (4.62)	873 (2.97)
4	A	205	205 (2.21)	205 (2.27)	205 (1.57)
5	A	300	69 (6.93)	70 (2.69)	70 (1.93)
6	A	400	300 (7.94)	353 (7.69)	353 (2.05)
	B	800	715 (7.64)	715 (6.72)	478 (2.29)
	C	524	200 (4.50)	351 (4.97)	351 (1.43)
7	A	225	140 (6.94)	161 (5.62)	161 (1.92)
	B	300	140 (2.96)	141 (3.14)	141 (2.09)
8	A	300	282 (4.59)	263 (1.92)	263 (1.54)
Total (avg. time)			5355 (5.94)	5559 (4.77)	5322 (2.38)

It is considered that the improvement of the tracking performance was owing to the Kalman filter. Because it gives good initial guess for the search by IR, the Powell method can reach the solution more quickly, avoiding undesirable local minima. In Path C of Case 6, the number of tracked frames significantly increased. Since the “Previous” method used the estimation result of the previous frame as the initial search parameters, the search by the Powell method fell into a local minimum. On the other hand, since the Kalman filter gives better initial parameters closer to the correct minimum, the Powell method could find the correct solution.

The number of tracked frames did not increase in some paths. In these paths, the IR might not have been performed properly due to distortion of the bronchus or bubbles appeared in the bronchi. To avoid such tracking failure because of improper IR, skipping IR and merely using the motion prediction result is one solution. Rapid changes of entropy of the RB images might be used to detect appearance of large bubbles.

In Path B of Case 6 with “MP+SSR,” the number of tracked frames was fewer than that of the “Previous” method. In this path, the estimation result by IR was affected by the bubbles. Then the Kalman filter used the improper result for motion prediction, and consequently improper restricted search space was configured. On the other hand, the tracking result of the path without search space restriction (“MP” method) was acceptable. Therefore this problem can be solved by detecting bubbles as we mentioned before. In Path B of Case 1, the motion prediction could not give proper prediction when the camera moved



**Fig. 1.** Examples of tracking results. VB images are rendered using the estimated camera parameters. The columns labeled “Prev,” “MP,” and “MP+SSR” show the results obtained by the previous method [2], the method using the Kalman filter’s output as initial guesses, and the method using restricted search space, respectively.

quickly. Because our Kalman filter assumes the camera to be in a constant state of acceleration, it can not predict the camera parameter properly when the motion violates the assumption too badly. In such cases, we might need another method to estimate the amount of camera motion.

## 4 Conclusions

We proposed a method for faster and more accurate bronchoscope camera motion tracking by camera motion prediction. To predict camera motion, we used a Kalman filter, assuming that the camera motion is in a constant state of acceleration. Also, the search space for IR was restricted using the prediction results. The motion prediction increased the number of frames correctly and continuously tracked by about 4.5%, and with the search space restriction the processing time was reduced by about 60%. Our future work includes: (1) accuracy improvement of the camera motion prediction by the Kalman filter, (2) reduction of computation time, (3) experiments on many more cases, and (4) quantitative evaluation of the tracking accuracy.

**Acknowledgments.** This study was partly supported by the Grants-in-Aid for Scientific Research and the 21st Century COE Program from Japan Society for the Promotion of Science, Grants-in-Aid for Cancer Research from the Ministry of Health, Labor and Welfare of Japan.

## References

1. K. Mori, Y. Suenaga, J. Toriwaki et al., "Tracking of camera motion of real endoscope by using the Virtual Endoscope System," In H. U. Lemke, M. W. Vannier, K. Inamura et al. (Editors), CARS2000, International Congress Series 1214, pp.85-90
2. D. Deguchi, K. Mori, Y. Suenaga et al., "New Image Similarity Measure for Bronchoscope Tracking Based on Image Registration," MICCAI 2003, LNCS vol.2878, pp.399-406
3. D. J. Vining, R. Y. Shitritin, E. F. Haponik et al., "Virtual Bronchoscopy," Radiology, 193 (P), Supplement to Radiology (RSNA Scientific Program), p.261, 1994
4. P. Rogalla, J. Terwisscha van Scheltinga, and B. Hamm, eds., "Virtual endoscopy and related 3D techniques," Springer, Berlin, 2001
5. D. A. Forsyth, and J. Ponce, "Computer Vision A Modern Approach," Pearson Education, 2003
6. 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
7. J. P. Helferty, W. E. Higgins, "Technique for Registering 3D Virtual CT Images to Endoscopic Video," Proceedings of ICIP, pp.893-896, 2001
8. W. H. Press, S. A. Teukolsky, W. T. Vetterling et al., "Numerical Recipes in C, The Art of Scientific Computing Second Edition," CAMBRIDGE UNIVERSITY PRESS, pp.321-336, 1999