

ManiSMC: A New Method Using Manifold Modeling and Sequential Monte Carlo Sampler for Boosting Navigated Bronchoscopy

Xiongbiao Luo¹, Takayuki Kitasaka², and Kensaku Mori^{3,1}

¹ Graduate School of Information Science, Nagoya University, Japan

² Faculty of Information Science, Aichi Institute of Technology, Japan

³ Information and Communications Headquarters, Nagoya University, Japan

Abstract. This paper presents a new bronchoscope motion tracking method that utilizes manifold modeling and sequential Monte Carlo (SMC) sampler to boost navigated bronchoscopy. Our strategy to estimate the bronchoscope motions comprises two main stages:(1) bronchoscopic scene identification and (2) SMC sampling. We extend a spatial local and global regressive mapping (LGRM) method to Spatial-LGRM to learn bronchoscopic video sequences and construct their manifolds. By these manifolds, we can classify bronchoscopic scenes to bronchial branches where a bronchoscope is located. Next, we employ a SMC sampler based on a selective image similarity measure to integrate estimates of stage (1) to refine positions and orientations of a bronchoscope. Our proposed method was validated on patient datasets. Experimental results demonstrate the effectiveness and robustness of our method for bronchoscopic navigation without an additional position sensor.

1 Introduction

During bronchoscopic interventions, physicians must know the position and orientations of a bronchoscope inside the airway trees, since they usually perform transbronchial lung biopsy (TBLB) to obtain samples of suspicious tumors for the assessment of bronchus and lung cancer. To localize and track the bronchoscope, current state of the art in navigated bronchoscopy includes two main approaches (or a combination of both): (1) image-based algorithms and (2) electromagnetic tracking (EMT). Although these methods proved good performance [1,2,3], it remains challenging to correctly localize the bronchoscope to places where it is exactly observing. Image-based schemes cannot tackle situations where problematic bronchoscopic video images (e.g., bubbles and motion blurring) happen. EMT-based methods often locate bronchoscopes incorrectly under any airway deformation, and accuracy of an EMT sensor measurements is heavily worsened by magnetic field distortion. Furthermore, no matter what approaches are used for bronchoscope motion tracking, they hardly adapt themselves to situation changes (e.g., patient coughing or dynamic errors in EMT outputs) over time during bronchoscopic interventions.

Beyond methods mentioned above, our primary idea of bronchoscope motion estimation is that bronchoscopic video sequences basically consist of three categories: (1) uninformative class, (2) inter-bronchus class, and (3) intra-bronchus class. The first class only includes problematic bronchoscopic frames (also refer to ambiguities) due to specular- or inter-reflection, bubbles, motion blurring, or collision with the bronchial walls. The informative category comprises inter- and intra-bronchus classes, which correspond to bronchoscopic frames with or without folds, and bronchoscopic images with bifurcations, respectively. A bronchoscopic image or scene can always be classified to one of three categories.

Based on these ideas, this work first learns bronchoscopic video manifolds to segment a bronchoscopic video into different clusters where each cluster represents one bronchial branch. We here construct bronchoscopic manifolds based on extending a local and global regressive mapping (LGRM) method [4] to Spatial-LGRM. By embedding an input image into these clusters, we can find an optimal bronchial branch that corresponds to its scene where a bronchoscope is observing. Hence, we can roughly obtain the pose of the camera. We then perform SMC sampling based on a selective image similarity measure to integrate the estimates of Spatial-LGRM-based learning to refine the localization of the bronchoscope.

It is worthwhile to highlight several aspects of our approach as follows. First, we propose a new framework of manifold modeling and SMC sampling to determine localizations of a bronchoscope for navigated bronchoscopy beyond image-based methods and EMT systems. Note that our manifold learning-based method for bronchoscopic scene identification provides an almost real-time means to roughly estimate the position and orientation of a bronchoscope. Next, we construct a new manifold modeling called Spatial-LGRM, which combines pose information to characterize the bronchoscope movements. Last, we introduce SMC sampling to incorporate manifold-based estimates and to tackle situations where ambiguities occur in bronchoscopic videos. Additionally, although we focused on bronchoscope motion tracking, we believe that our framework should also be appropriate to navigate other diagnostic endoscopes, e.g., colonoscope.

2 Bronchoscope Motion Tracking: ManiSMC

2.1 Bronchoscopic Scene Identification

(a) Preprocessing. We segment each 3D CT dataset to obtain bronchial tree structure information \mathcal{B} : $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_u, \dots, \mathbf{b}_k; u = 1, 2, \dots, k\}$, where k is the number of bronchial branches, and \mathbf{b}_u describes the centerline of one bronchial branch with its start position \mathbf{s}_u and end position \mathbf{e}_u ; branch direction \mathbf{d}_u can be computed by $\mathbf{d}_u = \mathbf{e}_u - \mathbf{s}_u$. \mathcal{B} is used to generate training data.

For each input RB image I_t at frame t , we first check whether it is an uninformative frame. In the HSB (HSB: Hue, Saturation, Brightness) color space, we compute hue and brightness deviations between current RB image and a virtual image generated by a virtual camera with an estimated pose inside the airway trees. If hue and brightness deviations are bigger than two predefined thresholds, I_t is considered as an uninformative frame. If I_t is an informative image,

we then perform morphology-based bifurcation detection to determine whether I_t is a bifurcation (intra-bronchus) image. We extract hole regions in I_t : if hole regions is more than two and the distances among these regions are constrained in a domain (here need to set minimum and maximum values for this domain), I_t is an image with observed bifurcations.

(b) Training Data Generation. By bronchial centerline information \mathcal{B} , we generate training data by changing positions and orientations of a virtual camera placed in a pre-built 3D anatomical airway model. Let \mathbf{p} denotes the camera position. Three vectors \mathbf{e}_x , \mathbf{e}_y , and \mathbf{e}_z ($\mathbf{e}_x = \mathbf{e}_z \times \mathbf{e}_y$) describe orientations of the virtual camera. We update positional parameters of the virtual camera by:

$$\mathbf{p} = \mathbf{p}_u^v + \frac{r}{3}\mathbf{e}_y, \quad \mathbf{p}_u^v = \mathbf{s}_u + \alpha_u \mathbf{d}_u, \quad (0, \mathbf{e}_y) = \mathbf{q}_\phi(0, \mathbf{e}_y^0) \mathbf{q}_\phi^{-1}, \quad \mathbf{q}_\phi = (\cos \frac{\phi}{2}, \sin \frac{\phi}{2} \mathbf{e}_z^0), \quad (1)$$

where \mathbf{p}_u^v is the v -th chosen point on the centerline of branch \mathbf{b}_u , r is the radius of \mathbf{b}_u , and $\alpha_u \in (0, 1)$ is constant coefficient. Simultaneously, we change orientational information by updating three vectors using quaternion \mathbf{q} :

$$(0, \mathbf{e}_z) = \mathbf{q}_\omega \mathbf{q}_\theta ((0, \mathbf{e}_z^0) \mathbf{q}_\theta^{-1} \mathbf{q}_\omega^{-1}), \quad (0, \mathbf{e}_y) = \mathbf{q}_\varphi \mathbf{q}_\omega \mathbf{q}_\theta ((0, \mathbf{e}_y^0) \mathbf{q}_\theta^{-1} \mathbf{q}_\omega^{-1} \mathbf{q}_\varphi^{-1}), \quad (2)$$

$$\mathbf{q}_\theta = (\cos \frac{\theta}{2}, \sin \frac{\theta}{2} \mathbf{e}_x^0), \quad \mathbf{q}_\omega = (\cos \frac{\omega}{2}, \sin \frac{\omega}{2} \mathbf{e}_z^0), \quad \mathbf{q}_\varphi = (\cos \frac{\varphi}{2}, \sin \frac{\varphi}{2} \mathbf{e}_y^0), \quad (3)$$

where $\mathbf{e}_z^0 = \mathbf{d}_u / \|\mathbf{d}_u\|$, \mathbf{e}_y^0 is a random vector that satisfies: $\mathbf{e}_y^0 \cdot \mathbf{e}_z^0 = 0$, ω and φ are rotational angles around \mathbf{e}_z^0 , θ is a rotational angle around \mathbf{e}_x^0 .

Generally, we generate two categories training data: (1) inter-bronchus and (2) intra-bronchus, by adjusting coefficient α_u : for inter-bronchus (fold) images $\alpha_u \in [0.1, 0.3]$, and intra-bronchus (bifurcation) images with $\alpha_u \in [0.7, 0.9]$.

(c) Learning Bronchoscopic Video Manifolds. After generating training data, we calculate low dimensional embedding spaces or eigenspaces and mapping functions or eigenmaps using LGRM that was proved to provide better performance more than other manifold learning methods in [4].

Suppose n training images $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n\}$, where $\mathbf{x}_i \in \mathbb{R}^D$, $D = w \times h$, $w \times h$ is the image size. Eigenspace $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_i, \dots, \mathbf{y}_n\}$ are low dimensional embedding manifolds of \mathcal{X} , where $\mathbf{y}_i \in \mathbb{R}^d$, ($d \ll D$). Manifold learning aims to map \mathcal{X} to \mathcal{Y} by finding eigenmap $\mathcal{M} \in \mathbb{R}^{D \times d}$: $\mathbf{x}_i \rightarrow \mathbf{y}_i, \mathbb{R}^D \rightarrow \mathbb{R}^d$. Note that \mathbf{X} , \mathbf{Y} , and \mathbf{M} represent matrices of \mathcal{X} , \mathcal{Y} , and \mathcal{M} .

In LGRM, since each image \mathbf{x}_i is first transformed to a Hilbert space \mathcal{H} and assumes that there exists a linear projection between \mathcal{H} and \mathbb{R}^d , for any \mathbf{x}_i , its low dimensional embedding \mathbf{y}_i satisfies:

$$\mathbf{y}_i = \psi(\mathbf{M})^T \psi(\mathbf{x}_i) + \mathbf{N}, \quad (4)$$

where $\psi(\mathbf{M})^T$ maps \mathbf{M} from \mathcal{H} to \mathbb{R}^d and \mathbf{N} is a residual term.

Specially, finding \mathbf{Y} and \mathbf{M} can be formulated the following optimization problem in terms of LGRM [4]:

$$\min_{\mathbf{Y}^T \mathbf{Y} = \mathbf{I}} \text{Tr} \left[\mathbf{Y}^T (\mathbf{L}_l + \mu \mathbf{L}_g) \mathbf{Y} \right], \quad (5)$$

where Tr is the trace operator, \mathbf{I} is an identity matrix, $(\mathbf{L}_l + \mu\mathbf{L}_g)$ is the Laplacian matrix that is the key component in manifold learning, and μ is a constant. \mathbf{L}_l that preserves the local manifold structure is calculated by Laplacian Eigenmaps [5]. \mathbf{L}_g denotes a kernelized global regression regularization, which is the main different part from other manifold learning algorithms; it can be computed by $\mathbf{L}_g = \xi\mathbf{H}(\mathbf{H}\mathbf{K}\mathbf{H} + \xi\mathbf{I})^{-1}\mathbf{H}$, where ξ is a regularization parameter, \mathbf{H} is the global centering matrix, and \mathbf{K} is the kernel matrix with its component $K_{ij} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|/\sigma^2)$ between two training images \mathbf{x}_i and \mathbf{x}_j .

Performing eigen decomposition on $(\mathbf{L}_l + \mu\mathbf{L}_g)$, we can obtain eigenspace \mathbf{Y} . Seeking partial derivatives of Eq. 5, $\psi(\mathbf{M})$ and \mathbf{N} can be determined by [4]:

$$\psi(\mathbf{M}) = \psi(\mathbf{X})\mathbf{H}(\mathbf{H}\psi(\mathbf{X})^T\mathbf{X}\mathbf{H} + \xi\mathbf{I})^{-1}\mathbf{Y}, \quad \mathbf{N} = \frac{1}{m}\mathbf{Y}^T\mathbf{1}_m - \frac{1}{m}\mathbf{M}^T\psi(\mathbf{X})\mathbf{1}_m, \quad (6)$$

when constant m reckons on \mathbf{L}_l and $\mathbf{1}_m \in \mathbb{R}^m$ is a vector of one.

After obtaining eigenspace \mathbf{Y} and eigenmap \mathbf{M} , for any input RB image I_t , its embedding \mathbf{y}_t can be easily obtained by Eqs. 4 and 6:

$$\mathbf{y}_t = (\psi(\mathbf{X})\mathbf{H}(\mathbf{H}\psi(\mathbf{X})^T\mathbf{X}\mathbf{H} + \xi\mathbf{I})^{-1}\mathbf{Y})^T\psi(\mathbf{x}_t) + \frac{1}{m}\mathbf{Y}^T\mathbf{1}_m - \frac{1}{m}\mathbf{M}^T\psi(\mathbf{X})\mathbf{1}_m. \quad (7)$$

However, LGMR only use intensity information of training images. This means \mathbf{L}_l and \mathbf{L}_g (or \mathbf{K}) only preserve intensity information of training images. From our experiences, intensity of inter-bronchus (fold) images are quite similar, although they may generate from totally different observation positions and orientations. This results in similar embedding representations of inter-bronchus images in \mathbf{Y} ; it may collapse clusters to wrongly identify bronchoscopic scenes and incorrectly estimate bronchoscope localizations.

To overcome such a drawback of LGRM, we extend it to Spatial-LGRM that integrates spatial information included camera position \mathbf{p} and orientation matrix $\mathbf{r}(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$, i.e., we add \mathbf{p} and $\mathbf{r}(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ to \mathbf{y}_i w.r.t \mathbf{x}_i :

$$(\mathbf{y}_i) \rightarrow \mathbf{y}_i^s = (\mathbf{y}_i, \mathbf{p}^i, \mathbf{e}_x^i, \mathbf{e}_y^i, \mathbf{e}_z^i), \quad \mathbf{Y} \rightarrow \mathbf{Y}^s, \quad \mathbb{R}^d \rightarrow \mathbb{R}^{(d+12)}. \quad (8)$$

Finally, we obtain eigenspace \mathbf{Y}^s and eigenmap \mathbf{M} for RB scene clustering.

(d) Bronchoscopic Scene Clustering. After preprocessing input RB image I_t , we cluster I_t to recognize current bronchoscopic scene by: (a) embedding I_t to \mathbf{Y}^s using \mathbf{M} in terms of Eq. 7 and obtain \mathbf{y}_t^s , (b) calculating Euclidean distance C_{t-1}^i between estimated position \mathbf{p}_{t-1} of I_{t-1} and $\mathbf{p}^i \in \mathbf{y}_i^s \in \mathbf{Y}^s$: $C_{t-1}^i = \sqrt{\|\mathbf{p}_{t-1} - \mathbf{p}^i\|^2}$, and choosing J_p nearest neighbors for \mathbf{y}_t^s by C_{t-1}^i : $\{\mathbf{y}_i^s\}_{i=1}^{J_p}$, (c) computing orientation deviation O_{t-1}^h between estimated rotation matrix \mathbf{r}_{t-1} of I_{t-1} and $\mathbf{r}^h(\mathbf{e}_x^h, \mathbf{e}_y^h, \mathbf{e}_z^h) \in \mathbf{y}_h^s \in \{\mathbf{y}_i^s\}_{i=1}^{J_p}$: $O_{t-1}^h = \arccos((\text{Tr}(\mathbf{r}^h\mathbf{r}_{t-1}^T) - 1)/2)$, and selecting J_o nearest neighbors by O_{t-1}^h from $\{\mathbf{y}_i^s\}_{i=1}^{J_p}$: $\{\mathbf{y}_h^s\}_{h=1}^{J_o}$, and (d) calculating distance E_t^h between \mathbf{y}_t^s and $\mathbf{y}_h^s \in \{\mathbf{y}_i^s\}_{i=1}^{J_o}$: $E_t^h = \sqrt{\|\mathbf{y}_t^s - \mathbf{y}_h^s\|^2}$. Finally, the output of the cluster is the pose parameters $(\mathbf{p}_*^s, \mathbf{r}_*^s)$ that correspond to the optimal embedding \mathbf{y}_*^s ($\mathbf{y}_*^s \in \{\mathbf{y}_h^s\}_{h=1}^{J_o}$) that is the closest to \mathbf{y}_t^s .

2.2 SMC-Based Bronchoscope Motion Estimation

This section fuses pose parameters ($\mathbf{p}_*^s, \mathbf{r}_*^s$) estimated from the step of bronchoscopic scene identification to determine current RB pose using a SMC sampler. Since this stage is quite similar to our previous work [6], we here briefly review the processing of SMC-based bronchoscope motion estimation.

Let $\mathbf{Q}_t(\mathbf{p}_t, \mathbf{r}_t)$ with translation \mathbf{p}_t and rotation matrix \mathbf{r}_t denotes transformation matrix from bronchoscope coordinates to CT coordinates at frame t .

We generate set of random samples $\mathcal{S}_t^g = \{(\mathbf{Q}_t^g, w_t^g) : t = 1, 2, \dots, N; g = 1, 2, \dots, M\}$ (N and M are the number of frames and samples, and w_t^g is a sample weight) to approach the posterior distribution of bronchoscope motion \mathbf{Q}_t . These samples are deterministically drifted and stochastically diffused by $\mathbf{Q}_t^g = \mathbf{A}\mathbf{Q}_{t-1}^g + \mathbf{B}n_t^g$, where matrix \mathbf{A} is calculated from ($\mathbf{p}_*^s, \mathbf{r}_*^s$) and $\mathbf{B}n_t^k$ is a noise term. Next, using \mathbf{Q}_t^g to generate virtual image I_V , we compute weight w_t^g by a selective image similarity measure [1]: $w_t^g = MoMSE(\mathbf{I}_t, I_V(\mathbf{Q}_t^g))$. Finally, in our case, the output parameters $\hat{\mathbf{Q}}_t$ with position and rotation of the SMC sampler for determining the pose of RB frame \mathbf{I}_t can be determined in terms of w_t^g : $\hat{\mathbf{Q}}_t = \max_{\hat{w}_t^g} \{(\mathbf{Q}_t^g, w_t^g)\}$, i.e., sample $\hat{\mathbf{Q}}_t$ with maximal weight \hat{w}_t^g corresponds to the maximal similarity between the current bronchoscopic image and the virtual frame generated by placing a virtual camera with the estimated pose including translation vector and rotation matrix inside the 3D airway tree anatomical model that was constructed by volume rendering techniques.

3 Experiments

For validation of our proposed method, we applied it to five cases of patient datasets that include bronchoscopic video frames and their corresponding 3-D chest CT images. The acquisition parameters of CT images are 512×512 pixels, 72-361 slices, and 1.0-2.0 mm slice thickness.

In Section 2.1(a), after pre-processing CT data, we obtain bronchial branch structure information \mathcal{B} and a 3D anatomical airway model. We generate training images with 30×30 pixels in gray-scale space by adjusting the following parameters in terms of Eqs. 1-3 in Section 2.1(b): for inter-bronchus images, α_u is set to 0.15, 0.20, and 0.25; for intra-bronchus images, α_u is set to 0.80, 0.85, and 0.90; $\phi = 30^\circ$, $\theta = 15^\circ$, $\omega = 0^\circ, 15^\circ, 30^\circ, \dots, 345^\circ$, and $\varphi = 0^\circ, 30^\circ, 60^\circ, \dots, 330^\circ$. For each branch $\mathbf{b}_u \in \mathcal{B}$, we generates 7488 inter-bronchus and 7488 intra-bronchus frames. During learning bronchoscopic video in Section 2.1, we constructed ten-dimensional embedding manifolds ($d = 10$ and $D = 30 \times 30$). Hence, the dimensions of matrices \mathbf{X} , \mathbf{Y} , and \mathbf{M} are 7488×900 , 7488×10 , and 900×10 . We set $\mu = 10^{-4}$, $\xi = 10^{-5}$, and $\sigma = 100$ in Spatial-LGRM according to [4].

Currently, for each patient case, we use six bronchial branches to create bronchoscopic video manifolds. They are (1) TR: trachea, (2) LM: left main bronchus, (3) RM: right main bronchus, (4) LU: left upper lobe bronchus, (5) RU: right upper lobe bronchus, and (6) RT: right trunchus intermedius. Hence, we totally obtain 12 clusters (inter- and intra-bronchus classes) for one patient and each cluster includes one eigenspace and one eigenmap.

After uninformative and bifurcation detections of input RB image I_t in Section 2.1(a), we convert I_t to a gray-scale image (362×370 or 256×263 pixels) and interpolate it to 30×30 pixels. During scene clustering in Section 2.1(d), the parameters of nearest neighbors are set to: $J_p = 200$ and $J_o = 100$. Additionally, to evaluate the successful rate of uninformative, we generate ground truth by manually inspecting real bronchoscopic images from three observers, one bronchoscopist and two scientists. Moreover, the tracking results of our proposed method are also manually and visually examined by the same three experts.

4 Results and Discussion

Table 1 summarizes the processed results of our methods. Detection rates of uninformative and intra-bronchus images are about 76.3% and 89.7%. The successful or correct scene recognition by only using bronchoscopic scene identification (BSI) described in Section 2.1 is about 3839 frames (59.6 %), which was improved to 4522 frames (70.2 %) using ManiSMC (Section 2). Fig. 1 visually compares the processed results of methods of BSI and ManiSMC. Generally, our experimental results demonstrate the effectiveness of ManiSMC that shows a good performance to understand bronchoscopic videos.

However, our method still fails to correctly estimate the bronchoscope localizations. Several reasons must be clarified as follows. First, sometimes uninformative images are wrongly detected, which results in incorrect embedding in manifolds; e.g., an image with bubbles can never find a correct correspondence in eigenspaces. In the future, we will improve uninformative frame detection by the work of Atasoy et al. [7] or the methods presented in [8,9]. Next, intra-bronchus images are wrongly classified. If an inter-bronchus image is detected to be an intra-bronchus one, it will never obtain a correct embedding in manifold clustering. Fig. 2 (a) (*top*) shows a successful detection of uninformative and bifurcation images. In some cases, it is difficult to detect whether an image is bifurcation, e.g., in Fig. 2 (a) (*bottom*), the RB frame is collided with a bronchial wall. Since detections of uninformative and bifurcation images are important to BSI and ManiSMC, we must improve current detection methods to

Table 1. Quantified processed results of bronchoscope motion tracking by visual inspection that manually checks if a RB image is similar to a virtual one

Patient cases	Number of frames	Moving path	Detection rates		Successful frames	
			Uninformative	Bifurcation	BSI	ManiSMC
1	1436	LM→LU	76.2%	89.7%	59.0%	77.0%
2	1685	TR→LM	74.6%	89.2%	51.4%	63.4%
3	1167	TR→RM→LM	70.8%	88.5%	45.8%	53.7%
4	1053	RM→RU	79.3%	90.3%	71.5%	77.5%
5	1101	TR→RM→RT	80.8%	90.8%	76.1%	82.2%
Total	6442		76.3%	89.7%	59.6%	70.2%

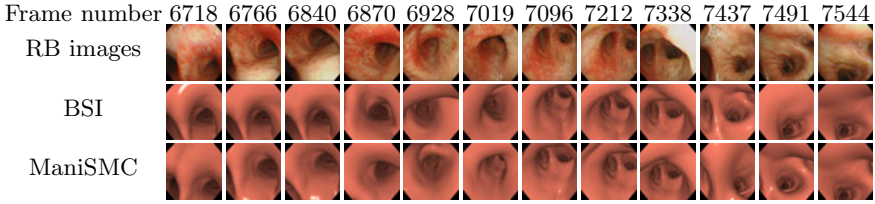


Fig. 1. Visual inspection of processed results of Case 4 by our bronchoscope motion tracking methods. Top row shows selected frame numbers and second row shows their corresponding patient RB images. Other rows display virtual images generated from processed results using methods of BSI and ManiSMC.

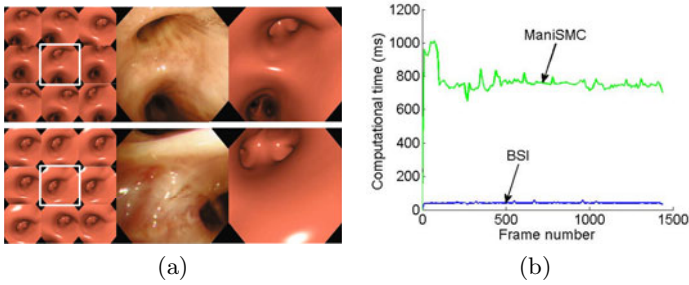


Fig. 2. (a) Examples of bronchoscopic scene identification: successfully recognized (*top*) and unsuccessfully identified (*bottom*) images. They show 9-nearest neighbors (*left*) of BSI, the frame with “□” used to SMC sampling, and RB images (*median*) with corresponding virtual images (*right*) generated by estimates of ManiSMC. (b) Computational times of Case 1. Average processing times of methods of BSI and ManiSMC are about 42 and 750 milliseconds per frame. Note that BSI can process one frame almost in real time (about 25 fps).

further enhance performances of BSI and ManiSMC. Third, similar images such as collision of bronchial walls and convolution of bronchial bifurcations in training data usually confuses clusters to determine accurate embeddings, although correct embeddings are included in the nearest neighbors of input RB images. Forth, we generated training data by updating the virtual camera observation poses in terms of bronchial centerline, i.e., most training images converges the bronchial centerline; however, a bronchoscope is usually not moving along the centerline, which causes actual bronchoscope poses that are difficult to correspond to manifolds. We need to improve the diversity of training data by adding more different virtual camera poses. Moreover, loss of centerline information due to airway segmentation algorithms also contributes to failures of scene identification. Finally, training data were generated from static CT slices that were acquired without airway deformation but bronchoscopic videos include patient breathing or coughing, this also causes unsuccessful bronchoscope tracking. Additionally, computational times of our methods are shown in Fig. 2 (b). Interestingly, BSI can almost process one frame in real time, about 25 frames per second

(fps). ManiSMC needs 0.75 seconds per frame since it requires to compute each sample weight during SMC-based motion estimation that is time-consuming.

5 Conclusion

This work proposed a new method that introduces LGRM-based manifold learning and SMC sampling for bronchoscope motion estimation. We constructed a Spatial-LGRM modeling with camera pose information to learn bronchoscopic video manifolds and use them to identify bronchoscopic video scenes where a bronchoscope is located and observing. Such a method can almost process video frames in real time (about 25 frames per second). By integrating a SMC sampler, our method can tackle situations where ambiguities occur in bronchoscopic videos. We may conclude that our proposed method provides a perspective means to boost bronchoscopic navigation without an additional position sensor.

Acknowledgments. This work was supported in part by Grant-in-Aid for Science Research funded by MEXT and JSPS, and “Chi-no-kyoten” by Aichi Prefecture. Authors would like to thank Dr. H. Tahabatake for providing datasets.

References

1. Deguchi, D., Mori, K., Feuerstein, M., Kitasaka, T., Maurer Jr., C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Selective image similarity measure for bronchoscope tracking based on image registration. *MedIA* 13(4), 621–633 (2009)
2. Soper, T.D., Haynor, D.R., Glenn, R.W., Seibel, E.J.: In vivo validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways. *IEEE TBME* 57(3), 736–745 (2010)
3. Luo, X., Feuerstein, M., Sugiura, T., Kitasaka, T., Mori, K., Imaizumi, K., Hasegawa, Y.: Towards hybrid bronchoscope tracking under respiratory motion: evaluation on a dynamic motion phantom. In: Wong, K.H., Miga, M.I. (eds.) *SPIE Medical Imaging 2010*, California, USA, vol. 7625, p. 76251B (2010)
4. Yang, Y., Nie, F., Xiang, S., Zhuang, Y., Wang, W.: Local and global regressive mapping for manifold learning with out-of-sample extrapolation. In: Fox, M., Poole, D. (eds.) *Proceedings of AAAI 2010*, Atlanta, USA, pp. 649–654 (2010)
5. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* 15(5), 1371–1396 (2003)
6. Luo, X., Reichl, T., Feuerstein, M., Kitasaka, T., Mori, K.: Modified hybrid bronchoscope tracking based on sequential monte carlo sampler: Dynamic phantom validation. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) *ACCV 2010*, Part III. LNCS, vol. 6494, pp. 409–421. Springer, Heidelberg (2011)
7. Atasoy, S., Mateus, D., Lallemand, J., Meining, A., Yang, G.Z., Navab, N.: Endoscopic video manifolds. In: Jiang, T., et al. (eds.) *MICCAI 2010*. LNCS, vol. 6362, pp. 437–445. Springer, Heidelberg (2010)
8. Oh, J., Hwang, S., Lee, J., Tavanapong, W., Wong, J., de Groen, P.C.: Informative frame classification for endoscopy video. *MedIA* 11(2), 110–127 (2007)
9. Bashar, M., Kitasaka, T., Suenaga, Y., Mekada, Y., Mori, K.: Automatic detection of informative frames from wireless capsule endoscopy images. *MedIA* 14(3), 449–470 (2010)