Anatomy Aware-based 2.5D Bronchoscope Tracking for Image-guided Bronchoscopic Navigation

Cheng WANG [a] and Masahiro ODA [b,a] and Yuichiro HAYASHI [a] and Takayuki KITASAKA [c] and Hayato ITOH [a] and Hirotoshi HONMA [d] and Hirotsugu TAKABATAKE [e] and Masaki MORI [f] and Hiroshi NATORI [g] and Kensaku MORI [a,h,i]

[a]Graduate School of Informatics, Nagoya University, Nagoya, Japan; [b]Information and Communications, Nagoya University, Nagoya, Japan; [c]School of Information Science, Aichi Institute of Technology, Toyota, Japan; [d]Medical Examination Department, Seamen's Insurance Hokkaido Healthcare Center, Sapporo, Japan; [e]Department of Respiratory Medicine, Sapporo Minami-Sanjo Hospital, Sapporo, Japan; [f]Department of Respiratory Medicine, Sapporo-Kosei General Hospital, Sapporo, Japan; [g]Department of Internal Medicine, Keiwakai Nishioka Hospital, Sapporo, Japan; [h] Information Technology Center, Nagoya University, Nagoya, Japan; [i] Research Center for Medical Bigdata, National Institute of Informatics, Tokyo, Japan

ARTICLE HISTORY
Compiled September 14, 2022

ABSTRACT
Physicians use an endoscopic navigation system during bronchoscopy to decrease the risk of getting lost in complex tree-structure like bronchus. Most existing navigation systems based on the camera pose estimated from bronchoscope tracking and/or deep learning. However, bronchoscope tracking-based method exists tracking error, and the pre-training of the model needs massive data. This paper describes an improved bronchoscope tracking procedure by adopting image domain translation technique to improve tracking performance. Specifically, our scheme consists of three modules, an RGB-D image domain translation module, an anatomical structure classification module and a structure-aware bronchoscope tracking module. The RGB-D image domain translation module translates a real bronchoscope (RB) image to its corresponding virtual bronchoscope image and depth image. The anatomical dependency module classifies the current scene into two categories: structureless and rich structure. The bronchoscope tracking module uses a modified video-CT bronchoscope tracking approach to estimate camera pose. Experimental results showed that the proposed method achieved higher tracking accuracy than the current state-of-the-art bronchoscope tracking methods.

KEYWORDS
Depth image generation; Endoscope tracking; Bronchoscopic navigation; Anatomical structure classification

1. Introduction

Lung/bronchial cancer mortality accounts for a very high proportion of cancer deaths each year (Rebecca et al. 2020). To reduce such high death rate, physicians use bronchoscopy to diagnose lung cancer in the early stage. However, during bronchoscopy,

it is very challenging to know the location of a bronchoscope in the bronchus due to the tree-like bronchus and narrow field of camera view. Therefore, a bronchoscopic navigation system is used to provide three-dimensional (3D) navigational information to physicians to diagnose lung diseases (Ivan et al. 1998).

Many types of bronchoscopic navigation systems have been reported in the past decades since the concept of bronchoscopic navigation was proposed (Ivan et al. 1998; Mori et al. 2002; Luo et al. 2012; Shen et al. 2015; Deguchi et al. 2009; Luo et al. 2014; Merritt et al. 2013; Khanmohammadi et al. 2020; Feuerstein et al. 2010; Shen et al. 2019; Sganga et al. 2019; Wang et al. 2020; Wegner et al. 2007; Banach et al. 2021; Deguchi et al. 2012; Wang et al. 2021; Schwarz et al. 2006; Shinohara et al. 2006; Gil et al. 2020). Since navigational information mainly comes from the result of the bronchoscope tracking, we classify the existing bronchoscopic navigation systems into three categories according to the method the bronchoscope tracking based: (1) 3D-2D image registration-based; (2) sensor-based, and (3) other tracking type-based. The 3D-2D image registration-based bronchoscope tracking uses RB and virtual bronchoscope (VB) image generated from CT volumes to estimate camera pose (Mori et al. 2002; Luo et al. 2012; Shen et al. 2015; Deguchi et al. 2009; Luo et al. 2014; Merritt et al. 2013). The camera pose is estimated by maximizing the image similarity between RB and VB images (Mori et al. 2002). Deguchi et al. improved image similarity calculation by using subblocks instead of the whole images (Deguchi et al. 2009). Subblocks were selected from the area existing characteristic structures (folds, bifurcations, and so on). Luo et al. used more characteristic information of the selected subblocks (such as luminance and the contrast) to improve image similarity calculation (Luo et al. 2014). Merritt et al. sped up the processing time of the image registration procedure by re-render RB images instead of using the volume rendering technique on VB images (Merritt et al. 2013). Shen et al. used depth images to replace the color images for image similarity calculation (Shen et al. 2015, 2019). Depth images come from a *shape from shading*-based technique (Visentini-Scarzanella et al. 2012) or CycleGAN-based image domain translation (Zhu et al. 2017). Jake et al. used a residual convolutional neural network (CNN) to locate the bronchoscope in CT volume (Sganga et al. 2019). The network is used to estimate the pose of a color RB image and a rendered image independently. However, the difference between preoperative (CT volume) and intraoperative (RB) images such as organ deformation or bubble decreased the tracking accuracy.

Additional sensor-based bronchoscope tracking method calculates the camera pose using the output of an additional sensor (e.g.,electromagnetic (EM) sensor) (Khanmohammadi et al. 2020; Feuerstein et al. 2010; Wegner et al. 2007; Banach et al. 2021). Schwarz et al. used an EM sensor to estimate the camera pose of the bronchoscope for navigation while the diagnosis of lesion regions (Schwarz et al. 2006). Deguchi et al. used the trajectory of an electromagnetic sensor to register to the centerline of the bronchus and use the transformation matrix to estimate camera pose (Deguchi et al. 2012). Banach et al. used an improved image domain translation network (Zhu et al. 2017) to estimate depth images for 3D shape reconstruction. The 3D shape is used together with an EM sensor to register to the bronchus shape segmented from CT volume for navigation. However, sensor measurement needs additional equipment (such as magnetic field generators and sensors) and can be affected by the external environment. A bronchoscope equipped with an external sensor cannot access the terminal bronchi due to size limitations.

There are other types of navigation schemes such as RB-VB image similarity-based branch identification (Shinohara et al. 2006), using visual SLAM (Mur-Artal et al.

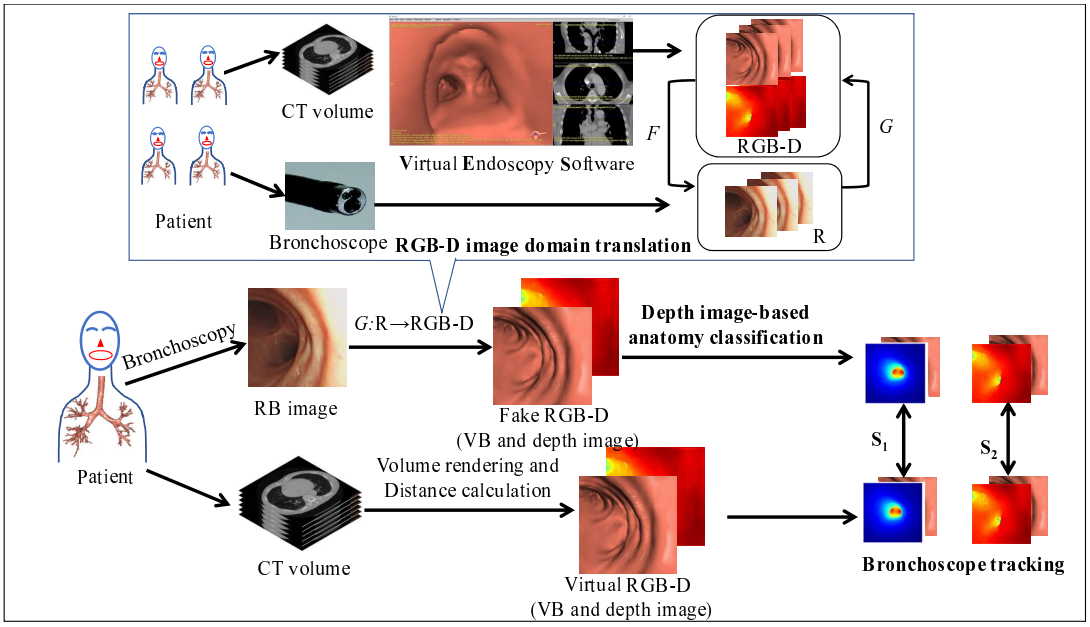

Figure 1. Structure of the proposed navigation system. There are three modules: RGB-D image domain translation, depth image-based anatomy classification and bronchoscope tracking. RGB-D image domain translation uses a pretrained model to translate a RB image into fake RGB-D images. Depth image-based anatomy classification classifies one image into two categories: lack of structure and rich structure. Bronchoscope tracking module selects appropriate similarity function to calculate the image similarity for bronchoscope.

2015) to process RB image for bronchoscope tracking (Wang et al. 2020), anatomical structure changes-based navigation schemes (Gil et al. 2020; Wang et al. 2021) have been reported. These tracking methods are still under exploration and have not been applied in clinical.

To reduce the tracking error caused by the pre- and intra- operative images, we propose to use the image domain translation technique to assist the 3D-2D image registration-based bronchoscope tracking. A color RB image is translated into an image pair containing a fake VB image and a depth image. At the same time, the volume rendering technique is used to generate a candidate set of VB-depth image pairs containing VB images and depth images. The camera pose of a VB-depth image pair is chosen as the camera pose if this pair has the maximum similarity with the image pair from image domain translation. We use the depth image from image domain translation to classify if an image is lacking of structure or rich structure. According to the classified category, an appropriate image similarity function is selected to calculate the image similarity. The camera pose is considered as the one which can minimize the similarity function.

2. Method

The proposed method uses preoperative CT volume and RB images as input. After processing, the camera pose of each RB image is estimated as output. There are three modules in the proposed method: (1) RGB-D image domain translation; (2) depth image-based anatomy classification and (3) bronchoscope tracking. The system structure is shown in Fig. 1.

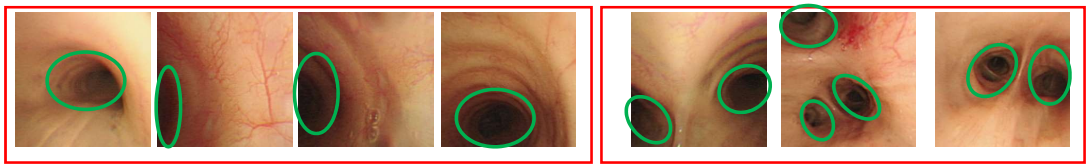

Figure 2. Example of two type of images: lack of structure and rich structure. Images are marked as lack of structure if only one BO is observed (as shown in the left four images) and images are marked as rich structure if more than one BO are observed.

2.1. RGB-D image domain translation

We use an image domain translation-based technique to excessive texture information (such as bubbles and blood) appearing on RB images. This technique involves two image domains and defines two mappings: G and F. These are used to translate the image in one domain into the opposite image domain. The training procedure makes the appearance of the translation image looks like the images in the target domain as much as possible. These two image domains in our task are the RGB image domain and the RGB-D image domain. The images in the RB image domain contains the RB images selected from RB videos. The images in the RGB-D image domain contain virtual-depth image pairs. Each pair contains a virtual image and a depth image. They are generated by using virtual endoscopy software (Mori et al. 2003). We move the virtual bronchoscope camera along the centerline of the bronchus, which is the simulation of the bronchoscopy. During this procedure, we save the generated VB image and depth image as the image pair in the RGB-D domain. To obtain the mapping $G : R \rightarrow (RGB-D)$, we use an improved image domain translation network: CycleGAN (Ito et al. 2021). We mark the generated VB image as $\hat{\mathbf{V}}$ and depth image as $\hat{\mathbf{D}}$. They are used for bronchoscope tracking in the following sections.

2.2. Anatomical analysis of bronchoscope images

The anatomical structure such as bronchial orifice and bifurcation are common observed in bronchus. These structures contribute a lot while the calculation of image similarity. Previous work concluded that the bronchoscope tracking performs poor in regions lacking of structure. Therefore, it is essential to distinguish whether an RB image has rich structure information or not. We consider an image has rich structure information if the bronchial orifice (BO) region is more than one and an image lacks structure if the number of BO region is one. Several example images are shown in Fig. 2. Considering the depth image does not contain noise such as bubbles or blood, we use the generated depth image to analyze if an image is lack of structure or not. We use the BO counting technique described in the literature (Wang et al. 2021) to classify whether an image lacks structure or not. This method projects the image pixels in two directions (vertical and horizontal) and uses the intensity profile of the depth image to count the BO number. An image is considered as lacking structure if only one BO is found (the left four images in Fig. 2) and is considered as has rich structure information if more than one BO is found. An image is marked as $C_{\hat{\mathbf{D}}} = 0$ if it lacks structure and is marked as $C_{\hat{\mathbf{D}}} = 1$ if it has rich structure information.

2.3. Structure-aware bronchoscope tracking

The task of module utilizes is to find the camera pose $\mathbf{p}$ of an RB image. This is achieved by using traditional images registration-based method (Ivan et al. 1998). The camera pose $\mathbf{p}$ is a vector $\mathbf{p} = \{\mathbf{R}, \mathbf{t}\}$, where $\mathbf{R}$ is the camera orientation and $\mathbf{t}$ is the camera position in Cartesian coordinate system. $\hat{\mathbf{p}}$ is estimated by using the following equation:

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \quad \|S(\hat{\mathbf{V}}, \phi(\mathbf{p}, \mathbb{P})) + S(\hat{\mathbf{D}}, \pi(\mathbf{p}, \mathbb{P}))\|, \tag{1}$$

where $\phi()$ is the volume rendering function that project the visible points $\mathbb{P}$ by using camera pose $\mathbf{p}$ to a 2D image plane (VB images) and $\pi()$ is the function to calculate the distance between the camera position of $\mathbf{p}$ and world points $\mathbb{P}$ (depth map). $\mathbb{P}$ are consisted by the 3D visible points located on the inner surface of bronchus. S_1 and S_2 are the similarity function.

The optimization procedure is performed by using Powell's method (Powell 1954). This method uses the camera pose of the previous frame as the initial pose of the current frame and finds the optimized camera pose by minimizing the Eq. 1.

2.3.1. Structure aware similarity function selection

According to the structure classification result, we choose different similarity function to calculate the image similarity. As it is shown in Equation 2, if an image has a rich structure, the camera pose can be estimated by using the MoMSE function (Deguchi et al. 2009). If an image lacks structure, we construct a new similarity function containing MoMSE function and Dice function to calculate the image similarity. α and β are the weights of two functions, respectively.

$$S(\mathbf{I_1}, \mathbf{I_2}) = \begin{cases} S_1 : MoMSE(I_1, I_2) & \text{if } C_{\hat{\mathbf{D}}} = 1 \\ S_2 : \alpha Dice(I_1, I_2) + \beta MoMSE(I_1, I_2) & \text{if } C_{\hat{\mathbf{D}}} = 0 \end{cases} \tag{2}$$

To calculate the Dice score between two images, Otsu-based threshold is used to threshold the depth images (Otsu et al. 1979). We only use depth images to calculate the Dice function. The camera pose is estimated by maximizing the similarity function.

3. Experiments

3.1. Experimental setting

We used several *in vivo* cases to validate the proposed method. Each case contained the chest CT volume and RB video of the same patient. These chest CT volumes were scanned several days before bronchoscopy by a CT scanner (XVision, Toshiba Medical Systems, Tokyo) and the RB videos were taken during bronchoscopy (BF-240, Olympus, Tokyo). We used six CT volumes to generate VB images and depth images for the training of the RGB-D CycleGAN. The specifications of these cases were shown in Table 2 (from case one to case six). Table 1 showed the number of RB images, VB images and depth images used for training CycleGAN. We resized these images to 256 $\times$ 256 pixels for the training of the RGB-D CycleGAN. The mini-batch of CycleGAN

Table 1. Information of the images used in the experiment (acquisition parameters of CT volumes involved in the experiment were shown. Case one to six were used for training CycleGAN. Case seven to ten were used for validation.)

Case	Slice size (pixels)	Pixel size (mm)	Number of slices	Slice spacing (mm)	Thickness (mm)
1	512 × 512	0.6 × 0.6	183	1.0	5.0
2	512 × 512	0.6 × 0.6	76	2.0	2.0
3	512 × 512	0.6 × 0.6	195	1.0	2.0
4	512 × 512	0.4 × 0.4	63	1.0	2.0
5	512 × 512	0.3 × 0.3	63	1.0	2.0
6	512 × 512	0.4 × 0.4	667	0.3	0.5
7	512 × 512	0.6 × 0.6	183	1.0	5.0
8	512 × 512	0.6 × 0.6	72	2.0	3.0
9	512 × 512	0.6 × 0.6	76	2.0	2.0
10	512 × 512	0.5 × 0.5	209	1.3	2.5

Table 2. Information of the images used in the experiment (the number of images involved in the experiment were shown. Case one to six were used for training CycleGAN. Case seven to ten were used for validation.)

Case	Number of RB image	Number of VB image	Number of depth image
1	1945	1550	1550
2	1394	2397	2397
3	1850	991	991
4	2003	1331	1331
5	1160	1674	1674
6	1626	1444	1444
7	380	-	-
8	450	-	-
9	1121	-	-
10	1350	-	-

was set to 10 and the epoch was set to 3000. We used Adam optimizer in CycleGAN with a learning rate of 0.0002.

3.2. Experimental results

3.2.1. Validation of image domain translation

We picked several VB images and depth images generated from CycleGAN to check the image translation results. These images were shown in Fig. 3. We showed the original RB image, the generated VB image (marked as fake VB image) and the generated depth image (marked as fake depth image) of two cases in this figure. As shown in this figure, bubbles and blood disappeared in the generated VB images and depth images.

3.2.2. Validation of tracking accuracy

We implemented several previous methods for comparison of the tracking. These methods were: (1) CT-video registration-based tracking (Deguchi et al. 2009) (marked as

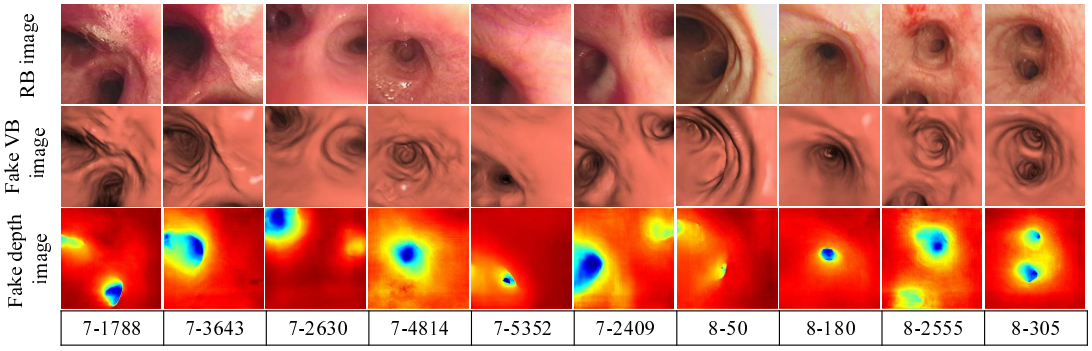

Figure 3. Sample images generated from CycleGAN. The first row is the RB images, the second row and the third rows are the generated VB images and depth images, respectively. The format of the text below images is case-frame number (e.g. 7-1788 means the 1788^{th} frame in case 7). No bubbles and blood were observed in generated VB images and depth images.

Table 3. The average MSE value between RB image and the VB images generated using different methods in four cases. The smaller the MSE value, the better the method. The proposed method showed lowest MSE value.

Case	RB-VB	VB-VB	D-D	VD-VD (Proposed method)
7	4428.8	3961.0	5713.5	2532.2
8	992.8	1385.7	1355.0	650.8
9	10144.2	13040.2	14038.2	7993.3
10	1480.2	1058.5	1106.0	965.2

RB-VB); (2) virtual image registration-based tracking (marked as VB-VB), (3) depth image registration-based tracking (Shen et al. 2019) (marked as D-D) and the proposed method (marked as VD-VD). For quantitative evaluation, we calculated the image similarity (mean squared error (MSE)) between RB images and the virtual images generated from different tracking methods (we did not create the ground truth of the camera pose because we think the camera pose may not so accurate). The average MSE value in four cases were shown in Table 3. The MSE value of the case 7 was shown in Fig. 4. The average MSE value of the proposed method (VD-VD) in case 7 was 2532.2, while the method RB-VB is 4428.8, the method VB-VB is 3961.0 and the method RB-VB is 5713.5.

We also used the tracked camera pose to generate VB images for comparison. The generated VB images were shown in Fig. 5. The generated VB images of the proposed method showed higher similarity than other methods.

4. Discussion

The generated VB image and depth image were shown in Fig. 3 and the comparison of different tracking methods were shown in Fig. 5. By comparing these images, we summarized the following advantage of the proposed method.

4.1. Better tracking performance

The proposed method showed better tracking performance than these previous methods, which we think it benefits from the decreased image difference and the using of

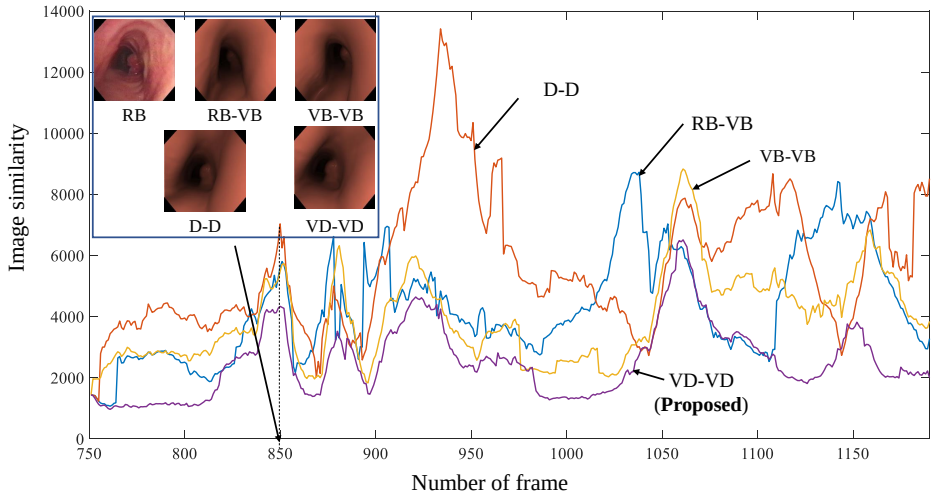

Figure 4. The MSE value between RB image and VB image generated using the tracking result of different methods of case 7. The proposed method showed a lower image similarity than other methods. We showed the visualized images of the 850-th frame. The VB image of the proposed method showed best similarity.

3D distance. We showed several visualized virtual bronchoscope images of case 7 by using the camera pose from these tracking methods in Fig. 5. We can found that the proposed method (method 4) showed a better than other methods (such as 0800-th, 1000-th and 1100 and 1150-th frame). These example images mainly contains three parts: bronchial orifice, lesion region and tracheal wall (1100-th frame: bronchial orifice in blue, lesion region in yellow and the others in red is tracheal wall). The tracheal wall of four method do not have many difference. However, the proposed method (method VD-VD) showed a better appearance in lesion region (the lesion region in method VD-VD is nearer to camera than its in others methods).

4.2. Decreased image difference between pre- and intraoperative images

The proposed method showed better tracking results in images with poor image conditions. We think it is because the proposed method uses a VB-depth image instead of using an RB image for image similarity calculation. The most difference between preoperative and intraoperative images disappeared using the image domain translation technique.

Both of the four methods are based on the image registration-based bronchoscope tracking. However, the method RB-VB mainly consider the image similarity and does not take the complex image conditions between pre- and intra operative into consideration. The method VB-VB used the transformed virtual images for image registration, which decreased the influence from image difference. The method D-D performed the registration on depth domain, which decreased the influence from image difference more. However, since the depth image-based registration does not contain color information, the tracking result become poor. The proposed method performs the registration on virtual image and use the depth image to judge the anatomical structure, the tracking result becomes better (such as 0850-th frame in case 7, the VB image from the proposed method showed a higher similarity with RB image). The method VB-VB performed the image registration on virtual images without depth informa-

tion, therefore, it can be considered as ablation study of the proposed method when the anatomical analysis is not used.

4.3. Shortcomings of the proposed method

One disadvantage of the proposed method is that bronchoscope tracking is time-consuming, which is the same as other image registration-based bronchoscope tracking methods. We measured the average processing time of 1000 frames using a laptop computer. The average processing time of CycleGAN (RGB-D image domain conversion) is about 0.005 s on GPU, 0.005 s for anatomical analysis of bronchoscope images, and 0.56 s per frame for structure-aware bronchoscope tracking on CPU, which may not be possible to run in real time. Therefore, future work of the proposed method includes reducing the processing time. This can be achieved by implementing the bronchoscope tracking part on the GPU.

We used the image similarity instead of using the camera pose for quantitative evaluation because it is difficult to create the ground truth of the camera pose. In the future, we need to consider a way to make the ground truth for evaluation such as using the EM sensor to capture the camera pose.

5. Conclusions

We use the image domain translation network to decrease the influence from different image types and use the anatomical structure to assist image similarity calculation. The translated virtual image and depth image has high similarity with the images from CT image in appearance. The proposed method shows high tracking accuracy than the previous method. In the future, more effects are needed to decrease the difference between pre- and intraoperative images, such as the implementation of a deformation simulation procedure and the generation of more navigational information such as branching name.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Parts of this research were supported by the MEXT/JSPS KAKENHI Grant Numbers 26108006, 17H00867, 17K20099, the JST CREST Grant Number JPMJCR20D5, Japan, and the JSPS Bilateral International Collaboration Grants.

References

Rebecca LS, Kimberly DM, Ahmedin J. 2020. Cancer statistics, 2020. CA: Cancer J. Clin.. **70**(1):7–30.

Ivan, B., Gilbert, F., Philippe, C.: Registration of real and CT derived virtual bronchoscopic images to assist transbronchial biopsy. IEEE Trans. Med. Imaging, **17**(5):703–714 (1998)

Mori, K., Deguchi, D., Sugiyama, J., Suenaga, Y., Toriwaki, J.I., Jr, C.R.M., Takabatake, H., Natori, H.: Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. Med. Image Anal., **6**(3):321–336 (2002)

Luo, X.B., Feuerstein, M., Kitasaka, T., Mori, K.: Robust bronchoscope motion tracking using sequential monte-carlo methods in navigated bronchoscopy: dynamic phantom and patient validation. Int. J. comput. Assist. Radiol. Surg., **7**(3):371–387 (2012)

Shen, M., Giannarou, S., Yang, G.Z.: Robust camera localisation with depth reconstruction for bronchoscopic navigation. Int. J. Comput. Assist. Radiol. Surg., **10**(6):801–813 (2015).

Deguchi, D., Mori, K., Feuerstein, M., Kitasaka, T., Jr, C.R.M., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Selective image similarity measure for bronchoscope tracking based on image registration. Med. Image Anal., **13**(4):621–633 (2009)

Powell, M. J. D. : An efficient method for finding the minimum of a function of several variables without calculating derivatives. Computer Journal. **7**(2): 155–162 (1964)

Luo X.B., Mori, K.: A discriminative structural similarity measure and its application to video-volume registration for endoscope three-dimensional motion tracking. IEEE Trans. Med. Imaging, **33**(6):1248–1261 (2014)

Merritt, S.A., Khare, R., Bascom, R., Higgins, W.E.: Interactive CT-video registration for the continuous guidance of bronchoscopy. IEEE Trans Med Imaging, **32**(8):1376–1396 (2013)

Khanmohammadi, A., Aghaie, A., Vahedi, E., Qazvini, A., Ghanei, M., Afkhami, A., Hajian, A., Bagheri, H.: Electrochemical biosensors for the detection of lung cancer biomarkers: A review. Talanta, **206**:120251(2020)

Feuerstein, M., Sugiura, T., Deguchi, D., Reichl, T., Kitasaka, T., Mori, K.: Marker-free registration for electromagnetic navigation bronchoscopy under respiratory motion. In: International Workshop on Medical Imaging and Virtual Reality, pp. 237–246. (2010)

Shen, M., Gu, Y., Liu, N., Yang, G.Z.: Context-aware depth and pose estimation for bronchoscopic navigation. IEEE Robot. Autom. Lett., **4**(2):732–739 (2019)

Jake, S., David, E., Chauncey, G., David C.: Offsetnet: Deep learning for localization in the lung using rendered images. In: 2019 The International Conference on Robotics and Automation (ICRA), pp. 5046–5052 (2019)

Wang, C., Oda, M., Hayashi, Y., Villard, B., Kitasaka, T., Takabatake, H., Mori, M., Honma, H., Natori, H., Mori, K.: A visual SLAM-based bronchoscope tracking scheme for bronchoscopic navigation. Int. J. Comput. Assist. Radiol. Surg. **15**(10):1619–1630 (2020)

Wegner, I., Biederer, J., Tetzlaff, R., Wolf, I., Meinzer, H.P.: Evaluation and extension of a navigation system for bronchoscopy inside human lungs. In: Proc. SPIE, **65091H**:522–533 (2007)

Banach, A., King, F., Masaki, F., Tsukada, H., Hata, N.: Visually navigated bronchoscopy using three cycle-Consistent generative adversarial network for depth estimation. Med. Image Anal., **73**, 102164 (2021)

Deguchi, D., Feuerstein, M., Kitasaka, T., Suenaga, Y., Ide, I., Murase, H., Mori, K.: Real-time marker-free patient registration for electromagnetic navigated bronchoscopy: a phantom study. Int. J. Comput. Assist. Radiol. Surg., **7**(3), 359–369 (2012)

Wang, C., Hayashi, Y., Oda, M., Kitasaka, T., Takabatake, H., Mori, M., Honma, H., Natori, H., Mori, K.: Depth-based branching level estimation for bronchoscopic navigation. Int. J. Comput. Assist. Radiol. Surg. **16**(10) 1795–1804 (2021)

Schwarz, Y., Greif, J., Becker, H.D., Ernst, A., Mehta, A.: Real-time electromagnetic navigation bronchoscopy to peripheral lung lesions using overlaid CT images: the first human study. Chest, **129**(4):988–994 (2006)

Shinohara, R., Mori, K., Deguchi, D., Kitasaka, T., Suenaga, Y., Takabatake, H., Mori, M., Natori H.: Branch identification method for CT-guided bronchoscopy based on eigenspace image matching between real and virtual bronchoscopic images. In: Proc. SPIE, **6143**, pp: 614314 (2006)

Gil, D., Esteban-lansaque, A., Borras, A., Ramırez, E., Sanchez, C.: Intraoperative extraction of airways anatomy in videobronchoscopy. IEEE Access, **8**, pp. 159696–159704 (2020)

Visentini-Scarzanella, M., Stoyanov, D., Yang G.Z.: Metric depth recovery from monocular images using shape-from-shading and specularities. In: 19th IEEE International Conference on Image Processing (ICIP), pp. 25–28 (2012)

Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In:2017 International Conference on Computer Vision (ICCV), pp. 2242–2251 (2017)

Mur-Artal, R., Montiel, J.M.M, Tardos, J.D.: ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Trans Robot, **31**(5):1147–1163 (2015)

Mori, K., Suenaga, Y., Toriwaki, J.: Fast software-based volume rendering using multimedia instructions on PC platforms and its application to virtual endoscopy. In: Proc. SPIE, **5031**:111–122 (2003)

Itoh, H., Oda, M., Mori, Y., Misawa, M., Kudo, S.E., Imai, K., Ito, S., Hotta, K., Takabatake, H., Mori, M., Natori, H., Mori, K.: Unsupervised colonoscopic depth estimation by domain translations with a lambertian-reflection keeping auxiliary task. Int. J. Comput. Assist. Radiol. Surg., **16**(6): 989–1001 (2021)

Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern.: Syst., **9**(1): 62–66 (1979)

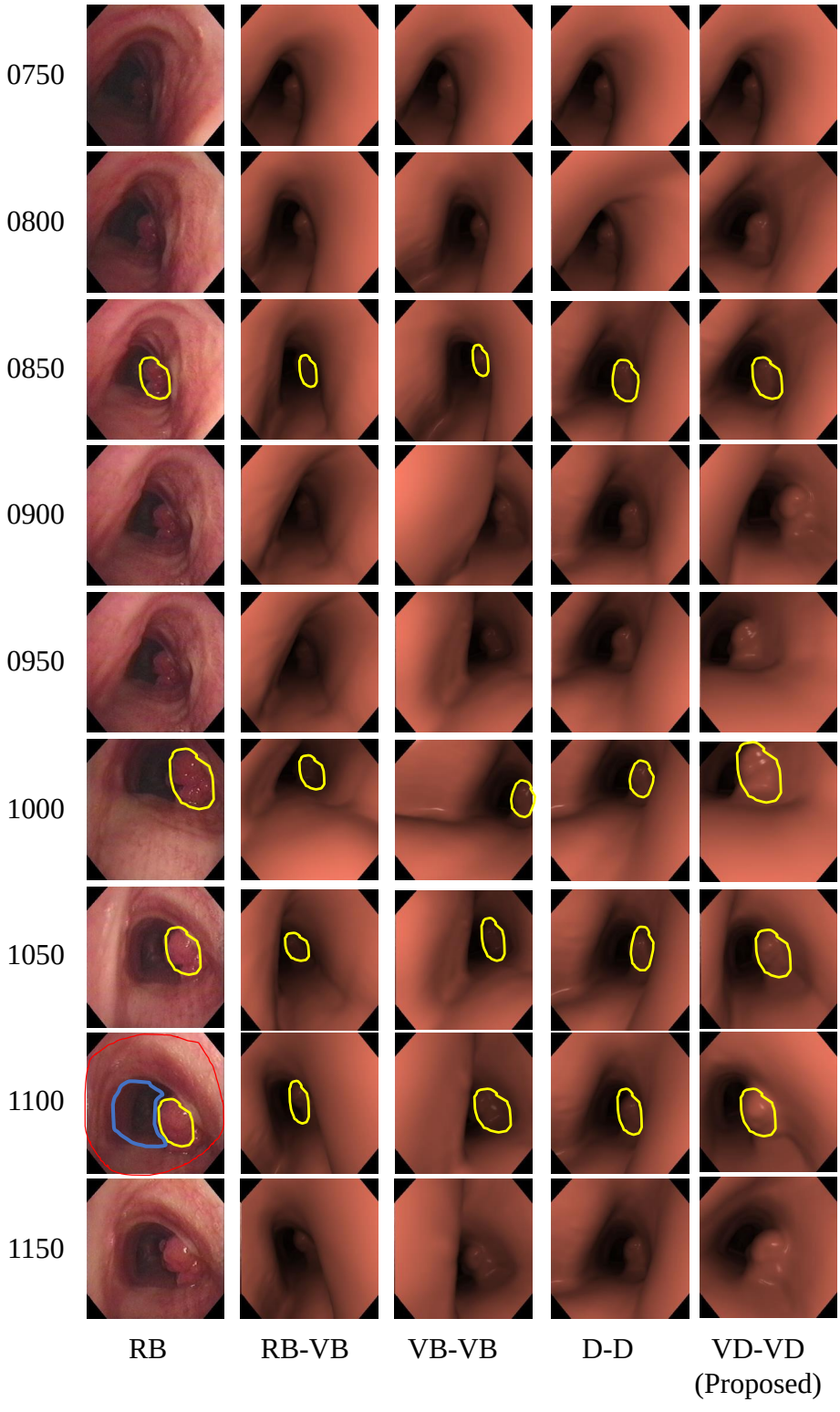

Figure 5. Visualized virtual bronchoscope images by using the tracking result from different tracking methods. We showed the original RB images, the generated VB images using different tracking results. We drew the bronchial orifice (regions in blue), the lesion region (regions in yellow) and the tracheal wall (regions between red and blue) on 1100-th frame. The regions in lesion showed that the proposed method performs better than other methods. The proposed method showed two aspects better than other methods: (1) several regions have higher appearance in RB image and the VB image from the proposed method (such as the lesion in 1100-th frame). (2) the camera orientation is similar between RB image and the propose method (such as the 1050-th frame).