Published in final edited form as:

Med Image Anal. 2021 October ; 73: 102164. doi:10.1016/j.media.2021.102164.

Visually Navigated Bronchoscopy using Three Cycle-Consistent Generative Adversarial Network for Depth Estimation

Artur Banach^{a,c,*}, Franklin King^a, Fumitaro Masaki^{a,d}, Hisashi Tsukada^b, Nobuhiko Hata^a

^aNational Center for Image-guided Therapy, Department of Radiology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States

^bDivision of Thoracic Surgery, Department of Surgery, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States

^cQUT Centre for Robotics, Queensland University of Technology, Brisbane, Australia

^dHealthcare Optics Research Laboratory, Canon U.S.A., Cambridge, MA, United States

Abstract

Background—Electromagnetically Navigated Bronchoscopy (ENB) is currently the state-of-the-art diagnostic and interventional bronchoscopy. CT-to-body divergence is a critical hurdle in ENB, causing navigation error and ultimately limiting the clinical efficacy of diagnosis and treatment. In this study, Visually Navigated Bronchoscopy (VNB) is proposed to address the aforementioned issue of CT-to-body divergence.

Materials and Methods—We extended and validated an unsupervised learning method to generate a depth map directly from bronchoscopic images using a Three Cycle-Consistent Generative Adversarial Network (3cGAN) and registering the depth map to preprocedural CTs. We tested the working hypothesis that the proposed VNB can be integrated to the navigated bronchoscopic system based on 3D Slicer, and accurately register bronchoscopic images to pre-procedural CTs to navigate transbronchial biopsies. The quantitative metrics to assess the hypothesis we set was Absolute Tracking Error (ATE) of the tracking and the Target Registration Error (TRE) of the total navigation system. We validated our method on phantoms produced from the pre-procedural CTs of five patients who underwent ENB and on two ex-vivo pig lung specimens.

Results—The ATE using 3cGAN was 6.2 ± 2.9 [mm]. The ATE of 3cGAN was statistically significantly lower than that of cGAN, particularly in the trachea and lobar bronchus ($p < 0.001$). The TRE of the proposed method had a range of 11.7 to 40.5 [mm]. The TRE computed by 3cGAN was statistically significantly smaller than those computed by cGAN in two of the five cases enrolled ($p < 0.05$).

Conclusion—VNB, using 3cGAN to generate the depth maps was technically and clinically feasible. While the accuracy of tracking by cGAN was acceptable, the TRE warrants further investigation and improvement.

*Corresponding author: artur.banach@qut.edu.au (Artur Banach).

Keywords

Bronchoscopy; Image-guided surgery; Lung cancer; CT imaging

1. Introduction

Lung cancer is the most common cause of cancer-related deaths in the United States. It is also a commonly diagnosed malignancy, second only to breast cancer in women and prostate cancer in men (Siegel et al., 2019). Early diagnosis of lung cancer is shown to improve patient outcomes (Bade and Dela Cruz, 2020), particularly in peripheral pulmonary nodules (PPNs) (Shaller and Gildea, 2020). Lately, Electromagnetically Navigated Bronchoscopy (ENB) is increasingly applied in the transbronchial biopsy of PPNs due to its excellent safety profile, with fewer pneumothoraxes, chest tubes, significant hemorrhage episodes, and respiratory failure episodes than a CT-guided biopsy strategy (Dale et al., 2012). NAVIGATE (Folch et al., 2019) is the latest and largest multicenter study conducted to investigate the efficacy of ENB; it reported that the 12-month diagnostic yield of ENB is 73% for the median lesion size of 2 cm. The reported diagnostic yield of transbronchial biopsy can be as low as 46% (Rivera et al., 2013; Wang Memoli et al., 2012; Mehta et al., 2018; Ost et al., 2016), warranting further improvement to provide accurate and timely diagnosis to patients with suspected lung cancer.

Researchers estimate that there are two technical factors causing lower diagnostic yield in ENB: dynamic deformation of the tracheobronchial tree by bronchoscope maneuvers (Leira et al., 2013) and nodule motion due to the breathing motion of the lung (Chen et al., 2015). Both issues cause CT-to-body divergence in which static preprocedural CTs no longer represent the physical shape of the airway, confusing physicians using the CT as a guidance map for the ENB (Luo and Mori, 2014; Makris et al., 2007; Mohanasundaram et al., 2013). In fact, the literature reports that the diagnostic yield is higher when the CT-to-body divergence is less than 4 mm (Makris et al., 2007), indicating to the field that CT-to-body divergence needs to be addressed before ENB can reach its full potential in diagnostic biopsy.

VNB, as opposed to ENB, has been proposed to address the aforementioned issue of CT-to-body divergence (Mirota et al., 2011). Vision-based tracking in VNB does not require an electromagnetic tracking sensor to localize the bronchoscope in CT; rather, it directly localizes the bronchoscope using the camera's view, conceptually removing the chance of CT-to-body divergence. VNB was originally proposed by Mori et al. to use intensity-based image registration between bronchoscopic images and virtual bronchoscopic images rendered from a 3-D reconstruction of the CT (Mori et al., 2006; Rai et al., 2008; Deligianni et al., 2004). These works have been extended into full navigation systems, validated through human subject studies, which have proven the clinical feasibility of VNB in bronchoscopy and have reported feature-poor images, occlusions, image artifacts, and patient's coughing as the most difficult challenges (Gibbs et al., 2009; Yu et al., 2010; Graham et al., 2010; Higgins et al., 2010; Graham et al., 2012; Merritt et al., 2013; Gibbs et al., 2014; Khare et al., 2015).

Depth estimation derived directly from bronchoscopic images has recently been proposed as an alternative to registering 3-D CT views (Shen et al., 2015). The rationale here is that depth estimation, instead of bronchoscopic images, may avoid tracking error due to varying illumination robustness, surface texture invariance, and, when applied in a trained fashion, it can also increase robustness to visual artifacts and occlusions. In fact, depth estimation is rising in fields outside of medicine (Godard et al., 2019). In general, the research on using depth estimation in VNB has been limited to evaluate the quality of depth maps (Shen et al., 2015, 2019), and has yet to be part of a comprehensive study that includes implementation to VNB to measure clinically relevant tracking accuracy.

The objective of this work, therefore, is to develop and clinically validate VNB as an alternative approach to tracking bronchoscopes without external sensors, by performing direct registration of the bronchoscopic image on the CT airway model using depth estimates. We propose a Three Cycle-Consistent Generative Adversarial Network (3cGAN), an unsupervised learning method, designed and trained in an unsupervised fashion, on virtual depth maps and real and virtual bronchoscopic images. To the best of our knowledge, this is the first time virtual bronchoscopic images have been embedded in the training process for depth estimation and navigation in bronchoscopy. We analyzed the performance of VNB, using 3cGAN by collecting Absolute Tracking Error (ATE) along the trajectories and Target Registration Error (TRE) at the targets. ATE and TRE are known to relate to the diagnostic yield (West et al., 2001; Makris et al., 2007; Karnak et al., 2013). We used clinically derived phantoms from cases involving peripheral pulmonary nodules, which are generally more challenging for physicians to target, necessitating the navigation, and two ex-vivo pig lung specimens. With the presented methodology and performance, VNB using depth estimation can proceed to a human subject study to investigate its feasibility in an in vivo clinical environment.

2. Literature Review on Visually Navigated Bronchoscopy

Vision-based tracking in bronchoscopy is a navigation method to enable VNB to register 2-D endoscopic video images to preoperative 3-D data derived from CT in real time. Attempts to use vision-based tracking in bronchoscopy started in the late 90s, led by Kensaku Mori at Nagoya University and Williams Higgins at Pennsylvania University. Mori et al. and Higgins et al. published a guiding series of bronchoscopic vision-based navigation using image intensity-based registration until the depth-based approaches started in the mid-2010s.

One of the original papers used optical flow, correlation coefficient, and mean square intensity difference to perform single frame registration between the 3-D airway model and bronchoscopic images (Mori et al., 2002). One of the interesting notes from the paper was that the success rate of image-based registration depends on the patient and their bronchoscopic paths, which vary the final success rate of the registration from 34% to 100%. Mori continued to advance his method, using metrics (Mori et al., 2006; Deguchi et al., 2009) such as the hybrid use of electromagnetic tracking sensors (Mori et al., 2005; Luó et al., 2012a; Luo et al., 2015), epipolar geometry analysis, and Kalman filtering for image registration (Luó et al., 2012b).

Higgins et al. have been active in performing image-based tracking for VNB since the publication of (Rai et al., 2008). In their follow-up paper (Merritt et al., 2008), the team tested the method using a phantom, which was based on airways depicted in a human chest scan with ten synthetic targets. The accuracy measured in this paper was due to placing forceps in the biopsy site. Deligianni et al. noted that respiratory-induced misalignment significantly hampered the tracking accuracy (Deligianni et al., 2006). Therefore, they compensated for respiratory deformation by using active shape models, derived from multiple CT scans of a deformable phantom. (Soper et al., 2010) also addressed the respiration-induced error by using a hybrid of VNB and ENB on an animal model. Further advancements in this group resulted in a complete real-time bronchoscopic guidance system based on CT-video registration using gradient-based optimization (Merritt et al., 2013), automatic optimal path planning for peripheral bronchoscopy, which integrates known route constraints (Gibbs et al., 2014), and a hands-free planning and navigation system for bronchoscopy (Khare et al., 2015).

Groups outside of Mori's and Higgins' recently proposed using a depth map derived from bronchoscopic video, as opposed to the image-intensity of the source material, for registration. Among these, the depth was shown to be most robust in illumination changes and, thus, makes the 2-D/3-D registration invariant to surface texture and preserves the morphological scene information (Shen et al., 2015). While depth sensors or a camera attached to the bronchoscope is one way to obtain a depth map (Penne et al., 2009), monocular depth estimation using bronchoscopic images is more actively discussed in reference to machine learning (Godard et al., 2019). For instance, (Visentini-scarzanella et al., 2017; Mahmood et al., 2018) transformed endoscopic images into textureless (synthetic-like) images to generate depths based on supervised training of large synthetic datasets. While the approach presented in (Visentini-scarzanella et al., 2017) uses supervised learning to generate textureless images from bronchial phantom data, the model in (Mahmood et al., 2018) uses unsupervised adversarial training (Goodfellow et al., 2014) on synthetic data for the same purpose, but using pig colon endoscopic data.

As for the unsupervised approach, the state of art is mostly limited to depth estimation, not reaching the point of total navigational system (Liu et al., 2020). However, in bronchoscopic applications, Shen *et al.* (Shen et al., 2019) recently proposed the first fully unsupervised tracking, using the cGAN (Zhu et al., 2017) to generate depth maps. The study by (Shen et al., 2019) included one human subject for preliminary assessment. This model also addresses the challenge of visual artifacts in bronchoscopy by modeling and including them in the training. Our paper builds upon Shen et al. work by newly developing 3cGAN, as opposed to cGAN, to allow inclusion of virtual bronchoscopic images in training. We also intend to contribute to the field by integrating the 3cGAN based depth map creation to total navigation system based on 3D Slicer (Fedorov et al., 2012) and measure clinically relevant quantitative metrics using phantoms created from pre-procedural CT of patients who underwent ENB and ex-vivo pig lung specimens.

3. Materials and Methods

To attain the objective of our work, we tested the hypothesis that VNB can accurately register a bronchoscope to a preprocedural CT, measured by ATE and TRE. We tested our working hypothesis by developing a 3cGAN specifically designed and trained in an unsupervised fashion on bronchoscopic images, virtual bronchoscopic images, and ground truth depth maps, and by registering the estimated depth map using Iterative Closest Point (ICP) matching. The symbols used in the Fig. 1 and all the equations are explained in Tab. 1. The outline of our VNB is presented in Fig. 2.

3.1. Depth map models

We developed the 3cGAN to enable embedding of virtual images into the training pipeline and the estimation of depth maps directly from bronchoscopic images. We adopted the architectures of the adversarial networks from cGAN (Zhu et al., 2017) and used them as direct comparison models. To enable unsupervised training, we assumed lack of correspondence between the bronchoscopic images, the virtual bronchoscopic images, and the ground truth depth maps in the collected patient datasets. Those correspondences are used for the validation of depth map estimation only. We applied adversarial losses (Goodfellow et al., 2014) to all the network levels (Fig. 1). For instance, we refer to the adversarial loss of level 1 as $L_{gan_{lev1}}$.

Both models were trained in PyTorch for 50 epochs with a batch size of one, and using an Adam optimizer with a learning rate of $2 \cdot 10^{-4}$. Training took about 8 hours for cGAN and 30 hours for 3LcGAN, and less than 0.02 s for the inference of one depth map on a GTX 1080 Ti GPU for both models. Each of the training datasets contained approximately 2500 images. It took approximately 2s to estimate one camera pose using ICP in 3D Slicer on an Intel Core i7-8650U CPU.

3.1.1. cGAN—Adversarial loss of cGAN (Zhu et al., 2017) connects adversarial losses from levels 1 and 2:

$$L_{gan2} = L_{gan_{lev1}} + L_{gan_{lev2}} \quad (1)$$

The Cycle Consistency Loss of the cGAN is formulated as follows:

$$L_{cyc2} = \|A_1 - G_{B_2 \hat{A}_2}(\hat{B}_1)\|_{L2} + \|B_2 - G_{A_1 \hat{B}_1}(\hat{A}_2)\|_{L2} \quad (2)$$

The total loss function of the cGAN is formulated as follows:

$$L_2 = L_{gan2} + L_{cyc2} \quad (3)$$

3.1.2. 3cGAN—Adversarial loss of 3cGAN is a sum of adversarial losses from all six levels:

$$L_{gan6} = L_{gan2} + L_{ganlev3} + L_{ganlev4} + L_{ganlev5} + L_{ganlev6} \quad (4)$$

Cycle Consistency Loss of the proposed 3cGAN combines L_{cyc2} and cycle consistency losses, pairing level 3 with 4 and level 5 with 6:

$$L_{cyc6} = L_{cyc2} + \|C_3 - G_{B_4}\hat{C}_4(\hat{B}_3)\|_{L_2} + \|B_4 - G_{C_3}\hat{B}_3(\hat{C}_4)\|_{L_2} + \|A_5 - G_{C_6}\hat{A}_6(\hat{C}_5)\|_{L_2} + \|C_6 - G_{A_5}\hat{C}_5(\hat{A}_6)\|_{L_2} \quad (5)$$

For the 3cGAN, we also introduced a merging loss function, which combines all the levels of the network and therefore allows us to make use of all of them in the final process of depth estimation in level 1. The merging loss function is formulated as follows:

$$L_{m6} = \|B_2 - G_{A_1}\hat{B}_1(G_{C_6}\hat{A}_6(G_{B_4}\hat{C}_4(G_{C_3}\hat{B}_3(G_{A_5}\hat{C}_5(G_{B_2}\hat{A}_2(B_2))))))\|_{L_2} \quad (6)$$

The total loss function of the 3cGAN is formulated as follows:

$$L_6 = L_{gan6} + L_{cyc6} + L_{m6} \quad (7)$$

An alternative approach to the proposed 3cGAN, which would include virtual bronchoscopic images C in training, is to add only one level (half) of cGAN including a generator to estimate depth maps $\hat{B}$ from virtual bronchoscopic images C . However, such three level network (cGAN + mapping C to $\hat{B}$) would not have a reference $\hat{C}$ (as an output of the generator) for the unsupervised learning to compare with the input C . Specifically, the inclusion of C in the training process would not affect the depth estimation from real bronchoscopic images (A to $\hat{B}$ mapping). Therefore, to enable and maximize the benefit of involving virtual bronchoscopic images in training, we decided to include a cycle-GAN pair for each of the pairs of data types (AB , AC , BC) and connect them with one 'merging loss' to bring the benefit from learning on all the levels to the first level of A to $\hat{B}$ mapping.

3.2. Depth map estimation from bronchoscopic images

Both models (cGAN and 3cGAN) were trained in an unsupervised fashion on bronchoscopic images obtained from phantom and ex-vivo sequences imitating transbronchial biopsies, as well as on ground truth depth maps and virtual bronchoscopic images. The phase of depth map estimation from a bronchoscopic image is illustrated in level 1 (Fig. 1). To verify the genericity of the models in the testing phase, the network received only bronchoscopic images from cases it had not been trained on. In the ex-vivo study the 3cGAN has been trained on lung specimen #1 and tested on lung specimen #2.

3.3. Localization and navigation by registration of depth map to preoperative 3-D models

The cGAN-estimated depth map ($\hat{B}_1$) was re-projected into a point cloud using bronchoscopic camera intrinsics and registered to the preoperative CT using ICP matching. The estimated i -th camera pose $\hat{T}_i$ is calculated as follows:

$$\hat{T}_i = \begin{cases} ICP(proj(\hat{D}_i, T_0)), & i = 1 \\ ICP(proj(\hat{D}_i, \hat{T}_{i-1})), & i > 1 \end{cases} \quad (8)$$

where $proj(\hat{D}_i, \hat{T}_{i-1})$ represents the re-projection of the i -th estimated depth map $\hat{D}_i$ from the previous camera pose $\hat{T}_{i-1}$. We use the initial ground truth camera pose (T_0) for the initialization. The ICP registration process using ICP matching finishes tracking and gives us the location and orientation of the bronchoscopic camera in the CT coordinate system.

4. Validation studies

4.1. Patient selection criteria

We retrospectively enrolled five patients from the ENB cases performed by our co-author HT in the month of January 2020, under IRB-approved #2020P001835. Attempts were made to enroll at least two successful cases of lesion in the right upper lobe (RUL), one successful case of left upper lobe (LUL), and two unsuccessful cases of either RUL or LUL. We selected only upper lobe nodules, as they are the most difficult to reach; thus, bronchoscopic navigation is most useful for them. A case was labeled as successful if the surgeon was able to reach the nodule in ENB. If there were multiple candidates from the January 2020 cases, attempts were made to choose cases from throughout the month to maximize randomization. The selected cases, with CT parameters, nodule location, diameter, reachability, and distance from the last navigation frame to the nodule, are presented in Table 2.

4.2. Bronchoscopic operations using patient-specific phantoms

We used (Nardelli et al., 2015)'s method to segment the lung airways from CT scans in 3-D Slicer (Fedorov et al., 2012). The segmented airway and the artificial 3-D airway model were used to produce 3-D clinical phantoms. They were produced first by 3-D printing the segmentations in ABS plastic. The 3-D printed segmentation models were mounted inside a container that was then filled using a tin-cured silicone rubber compound (Mold Max, Smooth-On, Inc., Macungie, PA). After curing was completed, the 3-D printed plastic models were cut out of the silicone, leaving a void that simulated an airway. Bronchoscopic operations were performed using a Scope 4 bronchoscope (Ambu Inc, Columbia, MD), virtual bronchoscopic images, and ground truth depth maps that were generated in Unity (Unity Technologies, San Francisco, CA). The validation design replicates real surgical scenarios, including the equipment commonly used in operating theaters.

Two bronchoscopic operations per case were recorded. The first was the training sequence, where the bronchoscope was inserted into every possible airway in order to collect video for the cGAN training. In the second, the navigation sequence imitated a typical transbronchial biopsy from carina toward the nodule identified by the clinician. All navigation sequences

were recorded using an electromagnetic tracker (Aurora, NDI, Waterloo, Canada). In this paper, the position and orientation tracked by the electromagnetic sensor are treated as ground truth. The coordinate frame of the ENB tracking device was registered to the CT coordinate frame using fiducial registration. Eight random fiducial points around the trachea and carina were marked on the phantom using an electromagnetically tracked pointer and were registered to their corresponding points in the CT image. The achieved average fiducial registration error was 1.46 mm. If the target could not be reached, the navigation sequence was stopped at the point where the bronchoscope was immersed. Due to scope limitations, we were not able to reach the target directly in any of the cases except case #4, but we reached the upper lobe in each of the cases except case #2. The distance from the last camera frame to the center of the nodule for each phantom navigation sequence is presented in Tab. 2. Only one of the navigation sequences (case #4) reached close proximity to the target, which directly influenced reported TRE. Moreover, the limiting scope size also caused movement of the electromagnetic sensor attached to the bronchoscopic camera in some parts of the navigation sequences. This led to spatial divergence between the bronchoscopic image and ground truth depth maps, which can be observed in the second and last column in Fig. 5. This phenomenon could have influenced the depth similarity metrics, but it did not influence the localization performance.

4.3. Bronchoscopic operations using ex-vivo specimens

The lung segmentation procedure and the registration process of the ENB tracking device to the CT coordinate frame were identical to those performed in the phantom study. The achieved average fiducial registration error was 0.72 mm. Bronchoscopic operations were performed using a snake robot developed in our lab (Dupourqué et al., 2019; Kato et al., 2020) with the OVM6946 bronchoscopic camera (OmniVision, CA, USA). The training sequence was recorded by inserting the bronchoscope into Right Upper Lobe, Right Middle Lobe, Left Upper Lobe and Left Middle Lobe of pig lung specimen #1. The navigation sequence imitated a typical transbronchial biopsy from the carina towards the imaginary target in the LML in ex-vivo specimen #2. The navigation sequence was recorded using ENB tracking (Aurora, NDI, Waterloo, Canada) to obtain the ground truth poses with respect to the CT frame. The experimental setup of the ex-vivo data collection has been presented in Fig. 3.

4.4. Evaluation Metrics

Two sets of metrics were collected in the validation study. The first set of the data collection relates to the depth map estimated by two methods: cGAN and 3cGAN. We collected Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Normalized Cross-Correlation (NCC) between the estimated and ground truth depth maps for all image frames as depth similarity metrics.

The second set of the data collection relates to localization performance of the proposed approach. The metric collected was the ATE of the cGAN and 3cGAN from the trachea to the lobar bronchus. The number of successful localizations was also counted by defining "success" as an ATE under 5.0 mm. The threshold of 5 mm is in line with the

bronchoscopic standard of accuracy required for navigation (Khan et al., 2016). Here, ATE is defined as the distance between the ground truth and the estimated camera location.

In addition, we calculated TRE for all the frames at the lobar bronchus for each case. Specifically, TRE was computed as follows:

$$TRE = d \cdot \sin(RE) + ATE \quad (9)$$

where d stands for the distance from the position of the bronchoscope to the center of the nodule and ATE and RE represent the position and orientation error, respectively.

4.5. Statistical analysis

First, the average and standard deviation of the depth similarity metrics computed by cGAN and 3cGAN were tabulated at three locations of the airway: the trachea, the main bronchus, and the lobar bronchus, and with all locations combined. The Anderson-Darling Normality Test was performed to test if the depth similarity metrics of cGAN and 3cGAN were distributed normally. If they were normally distributed, a paired t-test was performed to compare the two models at each location. Otherwise, a Wilcoxon signed-rank test was performed to compare the two models at each location.

Second, average and standard deviation of ATE and its box plots representation were reported. For all ATE data collected, an Anderson-Darling Normality Test was performed to test if ATE computed by cGAN and 3cGAN was distributed normally. If the ATE was normally distributed, a paired t-test was performed to compare the ATE of each method. Otherwise, a Wilcoxon signed-rank test was performed to compare the ATEs of the two models as paired data.

Third, ATE computed by cGAN and 3cGAN was decomposed into three locations of the airway: the trachea, the main bronchus, and the lobar bronchus. Then, contingency tables were created, showing the number of frames yielding ATE with two thresholds. The threshold of 5 mm is in line with the standard (Khan et al., 2016). The threshold of 10 mm was selected to provide more information about the frames registered above the 5 mm error. A chi-square test was performed at each location to assess if the difference between the number of successful frames computed by cGAN and 3cGAN was statistically significant.

Lastly, box plots of TRE computed by cGAN and 3cGAN at the lobar bronchus were presented for each case. For all TRE data collected, an Anderson-Darling Normality Test was performed to test if TRE computed by cGAN and 3cGAN was distributed normally. If TRE was normally distributed, a paired t-test was performed to compare the TRE of each method for each case. Otherwise, a Wilcoxon signed-rank test was performed to compare the TRE of the two models as paired data.

All statistical analyses were performed using Python version 3.7. In all tests, p values of 0.05 or less were considered to be statistically significant.

5. Results

Both cGAN and 3cGAN were successfully completed to create depth maps with depth similarity metrics of 454 frames in five cases, and to compute ATE of all 454 frames. TRE was computed at the lobar bronchus in all cases except case #2 because we did not reach the lobar bronchus in that case.

5.1. Depth Similarity Metrics

The average and standard deviation of MAE, RMSE, and NCC between the depth maps of ground truth and the depth maps estimated by cGAN and 3cGAN were 11.6 ± 4.7 [mm] and 7.8 ± 2.5 [mm], 15.0 ± 5.5 [mm] and 10.6 ± 3.0 [mm], and 0.77 ± 0.15 and 0.90 ± 0.07 , respectively (Tab. 3). The Anderson-Darling normality test showed that the distributions of MAE, RMSE, and NCC were not distributed normally. Therefore, we performed the Wilcoxon signed-rank test to compare the depth similarity metrics computed by the two models and found that 3cGAN produces statistically significantly smaller values in MAE ($p < 0.001$) and RMSE ($p < 0.001$) at the main bronchus and lobar bronchus, and statistically significantly higher values in NCC ($p < 0.001$) at each location. Fig. 5 illustrates the estimated depths for both cGAN and 3cGAN models on images along the trajectory from case # 4. The example outputs of the generators at each level of the network has been presented in Fig. 4

5.2. Absolute Tracking Error

Absolute Tracking Error (ATE) was defined as the absolute distance between the estimated and camera positions measured by the electromagnetic tracking. We recognized this camera position as the ground truth in our study. The average and standard deviation of the ATE from cGAN and 3cGAN was 7.5 ± 3.8 [mm] and 6.2 ± 2.9 [mm], respectively. ATEs in cGAN and 3cGAN were not represented by normal distribution. Therefore, a Wilcoxon signed-rank test was applied to compare ATE values in cGAN and 3cGAN, which found that the ATE of 3cGAN was statistically significantly lower than that of cGAN (Fig. 6). The plots of example trajectories have been presented in Fig. 8.

The contingency tables (Tab. 4) show the number of frames with an ATE of 5 mm or lower, or 10mm or lower, at three locations in the airway: the trachea, the main bronchus, and the lobar bronchus. A chi-square test was performed on each contingency table, presenting the statistically significant difference between cGAN and 3cGAN at the trachea and the lobar bronchus for ATEs below 5 mm and at the trachea and main bronchus for ATEs below 10 mm.

Fig. 7 supports our statistical finding that 3cGAN performs better, particularly in the trachea and lobar bronchus. We have selected three representative time frames in the navigational trajectory and illustrated the pose estimation error of these three frames within each of the bronchi. The first, the trachea illustration, shows us that 3cGAN localizes the bronchoscopic camera with a significantly smaller pose error than cGAN. In the trachea, the diameter of the visible airway is the largest of all; however, the visible airway depth is not. The visible airway depth is the largest in the main bronchus, which leads to the lower and upper lobes.

The pose estimation errors in the main bronchus are presented in the second illustration, indicating that cGAN and 3cGAN perform equally in this part of the lung. The last, the lobar bronchus illustration, shows that 3cGAN estimates the pose better than cGAN in terms of both position and orientation. The lobar bronchus is characterized by narrower airways and lower visible airway depths.

5.3. Target Registration Error in Visually Navigated Bronchoscopy

The average and standard deviation of TRE in the four cases (case #1, #3, #4, and #5) of 3cGAN in the lobar bronchus was 40.5 ± 3.2 [mm], 32.0 ± 3.0 [mm], 11.7 ± 3.2 [mm], and 12.9 ± 9.5 [mm]. Box plots of the TRE of the four cases for cGAN and 3cGAN are presented in Fig. 9. The Anderson-Darling normality test showed that the distribution of the TRE was not distributed normally. The Wilcoxon signed-rank test showed that the TRE computed by 3cGAN for cases #3 and #4 was statistically significantly smaller than those computed by cGAN.

5.4. Ex-vivo Results

The average and standard deviation of the ATE from 3cGAN was 6.98 ± 4.66 mm with TRE = 4.1 mm. The number of frames with ATE of 5 mm or lower was 82 (48%), and with ATE of 10 mm or lower was 134 (79%).

5.5. Total Navigation System

VNB in 3D Slicer at three bifurcation points from case # 4 are presented in Fig. 10. The rows of this figure represent entries to the 2nd, 3rd and 4th airway generations with bronchoscopic camera views, virtual bronchoscopic views from the pose estimated by our VNB, and virtual bronchoscopic views from the pose determined by ENB. All of the virtual views have been generated by our 'FlyThrough' navigation software in 3-D Slicer and contain center line representation of the airway leading toward the nodule. In the last frame, both virtual views show augmented nodule visualization for the surgeon to better understand the optimal position from which to deploy the biopsy needle. Note that at this particular frame, when close proximity to the nodule has been reached, the VNB view is better aligned in the bronchoscopic view and not ENB view. This proves the potential of VNB to eventually outperform ENB.

Fig. 11 presents 'FlyThrough' mode in the VNB software within 3D Slicer from case #1. The software presents a bronchoscopic camera view, a virtual bronchoscopic view produced using the proposed 3cGAN, and a ground truth virtual bronchoscopic view produced using ENB. The software user can see the center line of the airway leading to the nodule and can use the 3-D view of the airway model with the camera's pose displayed to perceive the bigger picture of the bronchoscope's location at every moment of the procedure. This figure presents what the clinician would see when trying the proposed VNB in the operating theater.

6. Discussion

Our results indicate that the presented VNB, using depth estimation and registration in the pre-procedural CT, has potential as a feasible alternative to external tracking in bronchoscopy. We have also found that embedding virtual bronchoscopic images in training and deploying the proposed loss functions allowed for significantly better depth estimation. The proposed 3cGAN provided a statistically significant localization improvement in the trachea and lobar bronchi in comparison to cGAN. We inferred that 3cGAN is best suited for smaller visible depth ranges, such as those in the trachea and lobar bronchi. The visible depth in the main bronchus has the widest range, as it is the bronchi leading to the lower lobe. We hypothesized that increasing the number of training frames recorded in the area of the main bronchus could improve the 3cGAN performance in that part of the lung.

The presented method is in accordance with the latest paper that presented a hybrid method of image-based tracking for bronchoscopy (Luó et al., 2012b). They combined SIFT, epipolar geometry, Kalman filtering, and image registration and validated their method in two phantom videos. Their tracking accuracy was 2.95 ± 1.29 mm for case #1 and 2.74 ± 0.86 mm for case #2. Compared to our accuracy of 6.2 ± 2.9 mm, their performance was significantly better. However, the authors did not provide information on what parts of the lung their images were recorded or if they were aiming at some clinical target. That information is crucial for a comprehensive comparison, as in our study we aimed for clinical targets in the upper lobes. Moreover, the method proposed in (Luó et al., 2012b) is based on traditional image processing methods like gradient feature search and epipolar geometry. The depth map based methods are proven to be more robust to illumination artifacts and low surface textures in bronchoscopy than methods based on gradient features (Shen et al., 2015). Recent developments in deep learning have allowed us to improve depth estimation from monocular endoscopic cameras like bronchoscopes and move heavy computational image processing to a learning, pre-experimental phase, significantly improving the tracking results. In this paper, we aim to further expand and improve deep learning based bronchoscopic depth estimation and prove its feasibility in a clinical setting, which is another reason why we do not provide a direct comparison with (Luó et al., 2012b).

Our paper also adds value to the learning-based methods used for VNB. Recently, two studies introduced supervised learning-based VNB, validated on phantom videos (Sganga et al., 2019; Zhao et al., 2020). The mean tracking accuracy reported in (Sganga et al., 2019) was 10 mm (if the airway was correctly recognized). However, there was no accuracy reported for particular bronchi, and their method got lost in the early upper lobe. The median tracking accuracy reported in (Zhao et al., 2020) was 1.64 mm; however, there was no insight on the airway generation in which the accuracy was investigated. These works are far from clinical adoption in terms of training datasets and technical metrics used for validation. Moreover, their supervised nature requires manually labeled data, which constrains them significantly from further clinical advancement. Note that the approach presented in (Zhao et al., 2020) shares the same unsupervised Cyclic-Gan core as (Shen et al., 2019) and our approach. However, they use it to estimate virtual bronchoscopic images to further train another network for pose estimation in a supervised way. To avoid constraining our

method to supervised training, we decided to proceed with geometry-based ICP camera pose estimation. Also, due to the unsupervised nature of the proposed depth estimation method we have not recorded ground truth poses for our training sequences, which does not allow us to train a pose estimator presented in (Zhao et al., 2020).

The literature reports only one VNB method using unsupervised learning-based depth estimation (Shen et al., 2019). Even though their work was validated on in vivo bronchoscopic videos, their validation included only technical metrics of tracking performance and focused on dealing with the image artifacts. The cGAN method we compare against in the paper, is the core of the depth estimation used in (Shen et al., 2019), without the artifact robustness. Our phantom datasets do not contain artifacts typical of bronchoscopy and the ex-vivo datasets contain a negligible number of them, which means that these datasets are not applicable to a full version of the (Shen et al., 2019) approach. As we did not compare our system to this method directly in a randomized control trial, we can only make a historical comparison between the approaches. Their reported mean ATE was 3.18 ± 2.34 mm, and the percentage of successfully tracked (< 5 mm) frames was 85% (Shen et al., 2019). Our mean ATE was 6.2 ± 2.9 mm, and the percentage of successfully registered frames was 40%. Note that our video sequences represent clinical target-aimed trajectories and (Shen et al., 2019) did not provide any clinical details about their videos.

Except for the promising depth estimation and thus localization accuracy, the unsupervised training makes our approach practical for clinical translation. We also predict that including virtual images would provide a good base from which to develop a depth estimation method that would be capable of identifying and addressing the image artifacts in the depth estimation process. At this stage, we think it would require pairing virtual bronchoscopic images with real bronchoscopic images for the network to address the artifacts in the depth estimation process, which would result in compromising unsupervised training. Unsupervised training is a big advantage in clinical applications due to the less complicated data collection process. However, if including partially supervised training would add a significant advantage to address image artifacts, it could be a reasonable compromise to make. Another alternative would be to deploy an approach like (Shen et al., 2019) and manually annotate the artifacts as an additional input for training. On the contrary, the advantage of learning artifact prediction implicitly is that thanks to the presence of the virtual bronchoscopic images the network could learn how real bronchoscopic images look without the artifact and that way see ‘through’ the artifact. This would provide more genericity than manual labeling in (Shen et al., 2019).

The performance of the proposed 3cGAN (Fig. 5) is validated on the images from the navigation (testing) dataset, which the network has not seen before. The discriminator learns to distinguish the estimated depth maps and ground truth depth maps in the training process. Since the navigation images come from a different sequence and from a different phantom/lung than the training images, the relationship between the estimated depth and visual conditions is also different. That difference explains why ground truth and estimated depth maps have unequal distributions (Fig. 5). Consequently, our validation study showed that the generator still produces depth maps sufficient for visual navigation providing satisfactory tracking accuracy.

Examples of estimations on all the network's levels are presented in Fig. 4. Compared to the depth estimation in the first network's level, other mappings seem to provide less accurate estimations. We predict this is due to the fact that the proposed merging loss function (see eq. 6) was designed to accumulate the benefits from cyclic training of all three level pairs, for the best quality of the depth estimation in the first network's level. We predict that adding equivalent merging loss function with the focus on other levels could bring estimation improvement to these levels, and potentially improve the depth mapping as well.

The presented 3cGAN is only able to estimate depth maps up-to-scale, which results in the reprojected point cloud not being in the metric space. The proposed navigation method is expected to be generalizable due to the fact that the scale of the point cloud is optimized during ICP matching. This expectation is supported by the results obtained in this study and by the fact that ICP has been deployed in multiple clinical settings as a reliable 3D registration method in bronchoscopy (Klein et al., 2007; Hofstad et al., 2014; Visentini-scarzanella et al., 2017), laparoscopic nephrectomy (Vagvolgyi et al., 2008; Su et al., 2009), and sinus surgery (Burschka et al., 2005).

In our work we aimed to address the CT-to-body divergence hindering ENB from reaching its highest potential. The technical metrics that directly influenced the diagnostic yield of ENB biopsies were the reachability of the target and the TRE that described the difference between the expected and the actual location of the catheter at the completion of navigation (West et al., 2001; Gex et al., 2014; Karnak et al., 2013). Despite the report that TRE of the ENB was 6.1 mm (Becker et al., 2005), 6.6 mm (Gildea et al., 2006), 4.0 mm (Makris et al., 2007), and 4.7 mm (Eberhardt et al., 2007), Karnak et al. report that mean TRE is lower in diagnostic cases (4.33 ± 0.71 mm) and not non-diagnostic cases (5.16 ± 0.05 mm, $P = 0.008$) (Karnak et al., 2013). The future of ENB depends on its expansion against its limitations and defending against competing technologies, such as radial ultrasound, fluoroscopy, confocal microendoscopy (Reynisson, 2014), and vision-based systems (Shen et al., 2019). The TRE achieved by the proposed VNB (Fig. 9) did not outperform the ENB accuracy, as mentioned above, nor was it within clinical standards (nodule diameter ≥ 10 mm (Jeong et al., 2007; Merritt et al., 2008)) for any of the presented cases. If the bronchoscopic video acquisition was not constrained by the scope size, and the last frame reached close proximity to the target, the TRE could have significantly better results. Moreover, high TREs were the result of the imperfection of the presented ATE and orientation error, which, as we observed during this research, often increases with the depth of airway penetration. We inferred that the orientation error increasing in the deeper parts of the lung represents the rising number of visible airways and the decreasing distance between them. That, merged with even a slight depth estimation error, can cause ICP registration to end up in a local minimum and return a misleading result. Improving that aspect would significantly decrease the TRE, as it directly depends on the estimation of the orientation.

The required orientation accuracy depends on the distance between the nodule and the camera, and ATE. The distance between the nodule and the camera is determined by the actual location of the nodule within the lung anatomy. Let's assume that the nodule has a spherical shape, the position of the camera is as close to the center of the nodule as the airway allows, and that the nodule is located $d = 10\text{mm}$ away from the camera position.

If we assume $ATE = 0\text{ mm}$, the needle placement will look like as it is in Fig. 12 (a), and the maximum allowed orientation error would be defined as $\alpha = \tan^{-1}\left(\frac{r}{d}\right)$. If we take $r = 5\text{ mm}$ (nodule diameter $\approx 10\text{ mm}$ (Jeong et al., 2007; Merritt et al., 2008)) and $d = 10\text{ mm}$ from the assumption above, then the maximum allowed orientation error is $\alpha = 27^\circ$. The requirement for the orientation error becomes more complex when $ATE \neq 0$. α then depends on the direction of the ATE with respect to the center of the nodule. The worst-case scenario is when $ATE = 5\text{ mm}$. This situation is illustrated in Fig. 12 (b) and $\alpha = 0\text{ mm}$ because every deviation away from the nodule would result in missing the biopsy target. In summary, under the aforementioned assumptions, the maximum allowed orientation error is $\alpha \in (0, 27)^\circ$ depending on the direction and amplitude of ATE.

The achieved mean ATE is larger than the required 5 mm (Khan et al., 2016). However, the percentage (48%) of frames with ATE lower than 5 mm is promising and encourages to further explore the presented methodological path. It is also yet to be investigated if this method can provide the required level of accuracy in higher lung generations.

This study is limited by the number of clinical cases and experiments not reaching close proximity of the nodules (except case #4). We also admit that the ICPs should be further investigated, and potentially modified, to improve the registration performance.

In summary, we have shown that depth estimation-based VNB is a clinically feasible alternative to ENB. We have presented a new unsupervised depth estimation method for bronchoscopy that embeds virtual bronchoscopic images in training, and we performed a clinical validation of the navigation accuracy in five retrospective biopsy cases and two ex-vivo animal specimens.

Acknowledgements

Artur Banach is supported by QUT Centre for Robotics at Queensland University of Technology in Brisbane, Australia. FM is an employee of Canon U.S.A., Inc. NH has a financial interest in Harmonus, a company developing Image Guided Therapy products. NH's interests were reviewed and are managed by Brigham and Women's Hospital and Partners Healthcare in accordance with their conflict of interest policies.

References

- Bade BC, Dela Cruz CS, 2020. Lung Cancer 2020: Epidemiology, Etiology, and Prevention. Clin. Chest Med 41, 1–24. URL: 10.1016/j.ccm.2019.10.001, doi:10.1016/j.ccm.2019.10.001. [PubMed: 32008623]
- Becker HD, Herth F, Ernst A, Schwarz Y, 2005. Bronchoscopic Biopsy of Peripheral Lung Lesions Under Electromagnetic Guidance A Pilot Study. J. Bronchol URL: <https://journals.lww.com/bronchology>.
- Burschka D, Li M, Ishii M, Taylor RH, Hager GD, 2005. Scale-invariant registration of monocular endoscopic images to ct-scans for sinus surgery. Medical Image Analysis 9, 413–426. [PubMed: 16009593]
- Chen A, Pastis N, Furukawa B, Silvestri GA, 2015. The effect of respiratory motion on pulmonary nodule location during electromagnetic navigation bronchoscopy. Chest 147, 1275–1281. URL: 10.1378/chest.14-1425, doi:10.1378/chest.14-1425. [PubMed: 25357229]
- Dale CR, Madtes DK, Fan VS, Gorden JA, Veenstra DL, 2012. Navigational bronchoscopy with biopsy versus computed tomography-guided biopsy for the diagnosis of a solitary pulmonary

nodule: A cost-consequences analysis. *J. Bronchol. Interv. Pulmonol*19, 294–303. doi:10.1097/LBR.0b013e318272157d.

- Deguchi D, Mori K, Feuerstein M, Kitasaka T, Maurer CR Jr., Suenaga Y, Takabatake H, Mori M, Natori H, 2009. Selective image similarity measure for bronchoscope tracking based on image registration. *Medical Image Analysis*13, 621–633. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1361841509000450>, doi:10.1016/j.media.2009.06.001. [PubMed: 19592291]
- Deligianni F, Chung A, Yang G.z., 2004. Patient-specific bronchoscope simulation with pq -space-based 2D/3D registration. *Comput. Aided Surg*9, 215–226. doi:10.3109/10929080500144927. [PubMed: 16192063]
- Deligianni F, Chung AJ, Yang GZ, 2006. Non-rigid 2D-3D Registration with Catheter Tip EM Tracking for Patient Specific Bronchoscope Simulation, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer, pp. 281–288. URL: 10.1007/11866565_{_}35, doi:10.1007/11866565_35.
- Dupourqué L, Masaki F, Colson YL, Kato T, Hata N, 2019. Transbronchial biopsy catheter enhanced by a multisection continuum robot with follow-the-leader motion. *International journal of computer assisted radiology and surgery*14, 2021–2029. [PubMed: 31289997]
- Eberhardt R, Anantham D, Ernst A, Feller-Kopman D, Herth F, 2007. Multimodality bronchoscopic diagnosis of peripheral lung lesions: A randomized controlled trial. *Am. J. Respir. Crit. Care Med*176, 36–41. doi:10.1164/rccm.200612-1866OC. [PubMed: 17379850]
- Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller JV, Pieper S, Kikinis R, 2012. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*30, 1323–1341. doi:10.1016/j.mri.2012.05.001. quantitative Imaging in Cancer. [PubMed: 22770690]
- Folch EE, Pritchett MA, Nead MA, Bowling MR, Murgu SD, Krinsky WS, Murillo BA, LeMense GP, Minnich DJ, Bansal S, et al., 2019. Electromagnetic navigation bronchoscopy for peripheral pulmonary lesions: one-year results of the prospective, multicenter navigate study. *Journal of Thoracic Oncology*14, 445–458. [PubMed: 30476574]
- Gex G, Pralong JA, Combescure C, Seijo L, Rochat T, Soccal PM, 2014. Diagnostic yield and safety of electromagnetic navigation bronchoscopy for lung nodules: A systematic review and meta-analysis. *Respiration*87, 165–176. doi:10.1159/000355710. [PubMed: 24401166]
- Gibbs JD, Graham MW, Bascom R, Cornish DC, Khare R, Higgins WE, 2014. Optimal procedure planning and guidance system for peripheral bronchoscopy. *IEEE Trans. Biomed. Eng*61, 638–657. doi:10.1109/TBME.2013.2285627. [PubMed: 24235246]
- Gibbs JD, Graham MW, Higgins WE, 2009. 3D MDCT-based system for planning peripheral bronchoscopic procedures. *Comput Biol Med*200939, 266–279. doi:10.1038/jid.2014.371.
- Gildea TR, Mazzone PJ, Karnak D, Meziane M, Mehta AC, 2006. Electromagnetic navigation diagnostic bronchoscopy: A prospective study. *Am. J. Respir. Crit. Care Med*174, 982–989. doi:10.1164/rccm.200603-344OC. [PubMed: 16873767]
- Godard C, Mac Aodha O, Firman M, Brostow GJ, 2019. Digging into self-supervised monocular depth estimation, in: *Proceedings of the IEEE international conference on computer vision*, pp. 3828–3838.
- Goodfellow IJ, Pouget-abadie J, Mirza M, Xu B, Warde-farley D, 2014. Generative Adversarial Nets. *arXiv*.
- Graham MW, Gibbs JD, Cornish DC, Higgins WE, 2010. Robust 3-D airway tree segmentation for image-guided peripheral bronchoscopy. *IEEE Trans. Med. Imaging*29, 982–997. doi:10.1109/TMI.2009.2035813. [PubMed: 20335095]
- Graham MW, Gibbs JD, Higgins WE, 2012. Computer-based route-definition system for peripheral bronchoscopy. *J. Digit. Imaging*25, 307–317. doi:10.1007/s10278-011-9433-7. [PubMed: 22083553]
- Higgins WE, Bascom R, Gibbs JD, Graham MW, Cornish DC, Khan M, Khare R (Eds.), 2010. *Image-Guided Bronchoscopic Sampling Of Peripheral Lesions: A Human Study*. volume 06, ATS. doi:10.1164/ajrccm-conference.2010.181.1_meetingabstracts.a5171.

- Hofstad EF, Sorger H, Leira HO, Amundsen T, Langø T, 2014. Automatic registration of ct images to patient during the initial phase of bronchoscopy: a clinical pilot study. *Medical physics*41, 041903. [PubMed: 24694134]
- Jeong YJ, Yi CA, Lee KS, 2007. Solitary pulmonary nodules: detection, characterization, and guidance for further diagnostic workup and treatment. *American Journal of Roentgenology*188, 57–68. [PubMed: 17179346]
- Karnak D, Çiledağ A, Ceyhan K, Atasoy Ç, Akyar S, Kayacan O, 2013. Rapid on-site evaluation and low registration error enhance the success of electromagnetic navigation bronchoscopy. *Ann. Thorac. Med*8, 28–32. doi:10.4103/1817-1737.105716. [PubMed: 23440066]
- Kato T, King F, Takagi K, Hata N, 2020. Robotized catheter with enhanced distal targeting for peripheral pulmonary biopsy. *IEEE/ASME Transactions on Mechatronics*, 1–1 doi:10.1109/TMECH.2020.3040314.
- Khan KA, Nardelli P, Jaeger A, O'Shea C, Cantillon-Murphy P, Kennedy MP, 2016. Navigational Bronchoscopy for Early Lung Cancer: A Road to Therapy. *Advances in therapy*33, 580–96. doi:10.1007/s12325-016-0319-4. [PubMed: 27084723]
- Khare R, Bascom R, Higgins WE, 2015. Hands-Free System for Bronchoscopy Planning and Guidance. *IEEE Trans. Biomed. Eng*62, 2794–2811. doi:10.1109/TBME.2015.2401514. [PubMed: 25675452]
- Klein T, Traub J, Hautmann H, Ahmadian A, Navab N, 2007. Fiducial-free registration procedure for navigated bronchoscopy, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 475–482.
- Leira HO, Langø T, Hofstad EF, Amundsen T, 2013. Bronchoscope - induced displacement of lung targets. *J. Pulmonol. Interv. Bronchol.*
- Liu X, Sinha A, Ishii M, Hager GD, Reiter A, Taylor RH, Unberath M, 2020. Dense depth estimation in monocular endoscopy with self-supervised learning methods. *IEEE Transactions on Medical Imaging*39, 1438–1447. doi:10.1109/TMI.2019.2950936. [PubMed: 31689184]
- Luó X, Feuerstein M, Kitasaka T, Mori K, 2012a. Robust bronchoscope motion tracking using sequential Monte Carlo methods in navigated bronchoscopy: Dynamic phantom and patient validation. *Int. J. Comput. Assist. Radiol. Surg*7, 371–387. doi:10.1007/s11548-011-0645-6. [PubMed: 21785944]
- Luo X, Mori K, 2014. A discriminative structural similarity measure and its application to video-volume registration for endoscope three-dimensional motion tracking. *IEEE Trans. Med. Imaging*33, 1248–1261. doi:10.1109/TMI.2014.2307052. [PubMed: 24893255]
- Luo X, Wan Y, He X, Mori K, 2015. Observation-driven adaptive differential evolution and its application to accurate and smooth bronchoscope three-dimensional motion tracking. *Med. Image Anal*24, 282–296. doi:10.1016/j.media.2015.01.002. [PubMed: 25660001]
- Luó X.o., Feuerstein M, Deguchi D, Kitasaka T, Takabatake H, Mori K, 2012b. Development and comparison of new hybrid motion tracking for bronchoscopic navigation. *Med. Image Anal*16, 577–596. doi:10.1016/j.media.2010.11.001. [PubMed: 21334250]
- Mahmood F, Chen R, Durr NJ, 2018. Unsupervised Reverse Domain Adaptation for Synthetic Medical Images via Adversarial Training. *IEEE Trans. Med. Imaging*37, 2572–2581. doi:10.1109/TMI.2018.2842767. [PubMed: 29993538]
- Makris D, Scherpereel A, Leroy S, Bouchindhomme B, Faivre JB, Remy J, Ramon P, Marquette CH, 2007. Electromagnetic navigation diagnostic bronchoscopy for small peripheral lung lesions. *Eur. Respir. J*29, 1187–1192. doi:10.1183/09031936.00165306. [PubMed: 17360724]
- Mehta AC, Hood KL, Schwarz Y, Solomon SB, Lincoln A, 2018. The Evolutional History of Electromagnetic Navigation Bronchoscopy State of the Art. *CHEST*154, 935–947. URL: 10.1016/j.chest.2018.04.029, doi:10.1016/j.chest.2018.04.029. [PubMed: 29723514]
- Merritt SA, Gibbs JD, Yu KC, Patel V, Rai L, Cornish DC, Bascom R, Higgins WE, 2008. Image-Guided Bronchoscopy for Peripheral Lung Lesions. *Chest*134, 1017–1026. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0012369208603645>, doi:10.1378/chest.08-0603. [PubMed: 18583513]

- Merritt SA, Khare R, Bascom R, Higgins WE, 2013. Interactive CT-video registration for the continuous guidance of bronchoscopy. *IEEE Trans. Med. Imaging*32, 1376–1396. doi:10.1109/TMI.2013.2252361. [PubMed: 23508260]
- Mirota DJ, Ishii M, Hager GD, 2011. Vision-Based Navigation in Image-Guided Interventions. *Annu. Rev. Biomed. Eng*13, 297–319. doi:10.1146/annurev-bioeng-071910-124757. [PubMed: 21568713]
- Mohanasundaram U, Ho LA, Kuschner WG, Chitkara RK, Canfield J, Canfield LM, Krishna G, 2013. The Diagnostic Yield of Navigational Bronchoscopy Performed with Propofol Deep Sedation. *ISRN Endosc.* 2013, 1–5. doi:10.5402/2013/824693.
- Mori K, Deguchi D, Akiyama K, Kitasaka T, Maurer CR, Suenaga Y, Takabatake H, Mori M, Natori H, 2005. Hybrid Bronchoscope Tracking Using a Magnetic Tracking Sensor and Image Registration, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer, pp. 543–550. URL: 10.1007/11566489{_}67, doi:10.1007/11566489_67.
- Mori K, Deguchi D, Kitasaka T, Suenaga Y, Takabatake H, Mori M, Natori H, 2006. Image-based bronchoscope tracking using motion predication and multiple searches. *International Journal of Computer Assisted Radiology and Surgery*.
- Mori K, Deguchi D, Sugiyama J, Suenaga Y, Toriwaki J, Maurer CR, Takabatake H, Natori H, 2002. Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. *Med. Image Anal*6, 321–336. doi:10.1016/S1361-8415(02)00089-0. [PubMed: 12270236]
- Nardelli P, Khan KA, Corvò A, Moore N, Murphy MJ, Twomey M, O'Connor OJ, Kennedy MP, Estépar RSJ, Maher MM, Cantillon-Murphy P, 2015. Optimizing parameters of an open-source airway segmentation algorithm using different CT images. *Biomed. Eng. Online*14, 1–24. doi:10.1186/s12938-015-0060-2. [PubMed: 25564100]
- Ost DE, Ernst A, Lei X, Kovitz KL, Benzaquen S, Diaz-Mendoza J, Greenhill S, Toth J, Feller-Kopman D, Puchalski J, et al., 2016. Diagnostic yield and complications of bronchoscopy for peripheral lung lesions. results of the aquire registry. *American journal of respiratory and critical care medicine*193, 68–77. [PubMed: 26367186]
- Penne J, Höller K, Stürmer M, Schrauder T, Schneider A, Engelbrecht R, Feuñner H, Schmauss B, Hornegger J, 2009. Time-of-Flight 3-D Endoscopy, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer, pp. 467–474. URL: 10.1007/978-3-642-04268-3{_}58, doi:10.1007/978-3-642-04268-3_58.
- Rai L, Helferty JP, Higgins WE, 2008. Combined video tracking and image-video registration for continuous bronchoscopic guidance. *International Journal of Computer Assisted Radiology and Surgery*3, 315–329. URL: 10.1007/s11548-008-0241-6, doi:10.1007/s11548-008-0241-6.
- Reynisson P, 2014. Navigated Bronchoscopy, A Technical Review. *J. Bronchol. Intervent. Pulmonol*21, 242–264.
- Rivera MP, Mehta AC, Wahidi MM, 2013. Establishing the diagnosis of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest*143, e142S–e165S. URL: 10.1378/chest.12-2353, doi:10.1378/chest.12-2353. [PubMed: 23649436]
- Sganga J, Eng D, Graetzel C, Camarillo DB, 2019. Autonomous Driving in the Lung using Deep Learning for Localization. *RAL*, 1–10.
- Shaller BD, Gildea TR, 2020. What is the value of electromagnetic navigation in lung cancer and to what extent does it require improvement? *Expert Rev. Respir. Med*14, 655–669. URL: 10.1080/17476348.2020.1748012, doi:10.1080/17476348.2020.1748012. [PubMed: 32216487]
- Shen M, Giannarou S, Yang GZ, 2015. Robust camera localisation with depth reconstruction for bronchoscopic navigation. *Int. J. Comput. Assist. Radiol. Surg*10, 801–813. doi:10.1007/s11548-015-1197-y. [PubMed: 25903774]
- Shen M, Gu Y, Liu N, Yang GZ, 2019. Context-Aware Depth and Pose Estimation for Bronchoscopic Navigation. *IEEE Robot. Autom. Lett*4, 732–739. doi:10.1109/LRA.2019.2893419.

- Siegel RL, Miller KD, Jemal A, 2019. Cancer statistics, 2019. CA. Cancer J. Clin69, 7–34. doi:10.3322/caac.21551. [PubMed: 30620402]
- Soper TD, Haynor DR, Glenny RW, Seibel EJ, 2010. In vivo validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways. IEEE Trans. Biomed. Eng57, 736–745. doi:10.1109/TBME.2009.2034733. [PubMed: 19846362]
- Su LM, Vagvolgyi BP, Agarwal R, Reiley CE, Taylor RH, Hager GD, 2009. Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3d-ct to stereoscopic video registration. Urology73, 896–900. [PubMed: 19193404]
- Vagvolgyi B, Su LM, Taylor R, Hager GD, 2008. Video to ct registration for image overlay on solid organs. Proc. Augmented Reality in Medical Imaging and Augmented Reality in Computer-Aided Surgery (AMIARCS), 78–86.
- Visentini-scarzanella M, Koto S, Sugiura T, Kaneko T, 2017. Deep monocular 3D reconstruction for assisted navigation in bronchoscopy. Int. J. Comput. Assist. Radiol. Surg7, 1089–1099. doi:10.1007/s11548-017-1609-2.
- Wang Memoli JS, Nietert PJ, Silvestri GA, 2012. Meta-analysis of Guided Bronchoscopy for the Evaluation of the Pulmonary Nodule. CHEST142, 385–393. URL: <http://journal.publications.chestnet.org/>, doi:10.1378/chest.11-1764. [PubMed: 21980059]
- West JB, Fitzpatrick JM, Toms SA, Maurer CR, Maciunas RJ, 2001. Fiducial point placement and the accuracy of point-based, rigid body registration. Neurosurgery48, 810–817. doi:10.1227/00006123-200104000-00023. [PubMed: 11322441]
- Yu KC, Gibbs JD, Graham MW, Higgins WE, 2010. Image-based reporting for bronchoscopy. J. Digit. Imaging23, 39–50. doi:10.1007/s10278-008-9170-8. [PubMed: 19050956]
- Zhao C, Shen M, Sun L, Yang GZ, 2020. Generative Localization with Uncertainty Estimation through Video-CT Data for Bronchoscopic Biopsy. IEEE Robot. Autom. Lett5, 258–265. doi:10.1109/LRA.2019.2955941.
- Zhu JY, Park T, Isola P, Efros AA, 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. Proc. IEEE Int. Conf. Comput. Vis 2017-Octob, 2242–2251. doi:10.1109/ICCV.2017.244, arXiv:1703.10593.

Fig. 1.

Block diagram of all levels of 3cGAN. The generator (G) and discriminator (D) are the main components of a Generative Adversarial Network. The symbols used are explained in Tab. 1. In the training process, G tries to generate an image that is so similar to the ground truth that D is not able to distinguish between them. Each D has access to the ground truth images of its input for the *real/fake?* comparison. The superscripts under G and D represent the input with the level number and the output of G with the level number, respectively. For example, $G_{A_1 \hat{B}_1}$ represents a generator that estimates depth $\hat{B}_1$ from bronchoscopic image

A_1 . $D_{\hat{B}_1}$ is a discriminator trying to decide if the estimated depth $\hat{B}_1$ is estimated (fake) or ground truth (real). Models in the other levels can be interpreted analogically. Note that the training is unsupervised, so there is no correspondence needed between the inputs at any of the network's levels. The black dashed lines represent the process of calculating the cycle-consistency loss functions (eq. (5)) for all three cycle-consistency level pairs. The grey lines illustrate the network backbone related to the merging loss function (eq. (6)). The grey circles highlight the beginning and end of the flow in eq. (6). The generator $G_{A_1\hat{B}_1}$ is used in the testing phase, which has been highlighted in the figure.

Fig. 2.

Illustration of the proposed system's pipeline. The network estimates a depth map from a bronchoscopic image. That depth map is re-projected into a point cloud using bronchoscopic camera intrinsics. The camera location is determined using ICP registration of a point cloud of the preoperative CT 3-D lung model.

Fig. 3.
The ex-vivo experimental setup. The figure shows the ex-vivo pig lung #2 used for collecting bronchoscopic images to validate the navigation performance of the proposed Visually Navigated Bronchoscopy.

Fig. 4.

Outputs of the generators at each level of the network in 3cGAN approach. The first and second rows represent the inputs and outputs of the generators respectively. The same notation as in Fig. 1 and all equations has been used for clarity. The depth maps are scaled in mm.

Fig. 5.

The effect of cGAN and 3cGAN on depth estimation in Visually Navigated Bronchoscopy. We distinguished 5 example bronchoscopic images, with case #4 trajectory representing different camera poses. Rows represent respectively: bronchoscopic images, ground truth depth maps (GT), corresponding depth maps estimated by cGAN, corresponding depth map estimated by 3cGAN, and the illustration of the estimated camera pose overlaid onto the airway representation of case #4. Depth maps are scaled in mm. This figure supports the claim of VNB's clinical feasibility in navigating toward peripheral lesions and addresses the advantage of 3cGAN over cGAN in depth estimation.

Fig. 6.

ATE of cGAN and 3cGAN approaches. The box extends from the 25% to 75% quartile values of the data with a line at the median (6.7 [mm] and 5.9 [mm] for cGAN and 3cGAN). A Wilcoxon signed-rank test showed statistical difference between ATEs computed from the models ($p < 0.05$).

Fig. 7.

The effect of the depth estimation method (cGAN vs 3cGAN) on Absolute Tracking Error (ATE) of the trachea, main bronchus, and lobar bronchus in Visually Navigated Bronchoscopy. The plot in the left upper corner represents ATE for both depth estimation methods. We distinguished three points to illustrate examples of statistically significant differences in the performance between cGAN and 3cGAN over the whole trajectory, as well as in the trachea and the lobar bronchus. The presented trajectory comes from case #5.

Fig. 8.

Ground truth (red) and estimated with 3cGAN (green) trajectories of the navigation sequences for all datasets. The tracking error in phantom #1 increased in the second part of the trajectory but at the end of it the registration stayed within the acceptable 5 mm. For phantom #2 the pose estimation error was higher than expected at the carina and main bronchus but converged when getting closer to the target. The tracking error for the ex-vivo sequence was high in the first part of the trajectory but stayed close to the ground truth from the main bronchus all the way to the target.

Fig. 9.

The TREs of the cGAN and the 3cGAN approaches at the lobar bronchus. The box extends from the 25% to 75% quartile values of the data with a line at the median. A Wilcoxon signed-rank test showed statistical differences in case #3 and case #4 ($p < 0.05$).

Fig. 10.

The outcome of our 'FlyThrough' navigation software illustrating three bifurcation points in case #4. Rows 1, 2, and 3 represent entry to the 2nd, 3rd and 4th airway generations, respectively. The first column represents the bronchoscopic camera view, the second column represents the virtual bronchoscopic view from the pose estimated by our VNB, and the third column represents the virtual bronchoscopic view from the pose determined by electromagnetic tracking. Both virtual views contain the center line of the airway leading to the nodule. Note that the visual navigation view (column 2) in row 3 resembles the

bronchoscope view more than the electromagnetic navigation view (column 3). Virtual images in row 3 contain augmented nodule visualization (green) for biopsy assistance.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 11.

Our 'FlyThrough' navigation software illustrating case #1 in 3-D Slicer. The presented software provides: Bronchoscope = bronchoscopic camera image, Visual Navigation = virtual bronchoscopic view from the pose estimated by our VNB, and Electromagnetic Navigation = virtual bronchoscopic view from the pose determined by electromagnetic tracking. At the top of the navigation window, the software provides a three-dimensional view of the lung phantom with the poses of both the Visual Navigation view (red camera model) and the Electromagnetic Navigation view (blue camera model), as well as the green nodule representation. Our navigation software presented the utility of the proposed VNB method in a scenario resembling the surgical environment.

Fig. 12.

Illustration of the two scenarios of the nodule targeting in terms of the orientation error. (a) represents a scenario when Absolute Tracking Error $ATE = 0$, where r is the radius of the nodule, d represents the distance between the camera position and the center of the nodule, and α is the maximum allowed orientation error. (b) illustrated a situation when the $ATE = 5\text{mm}$, and there is no space for the orientation error, otherwise the biopsy target will be missed. It is assumed that the nodule has a spherical shape.

Table 1.Imagery in training^{*}

Item	Source	Symbol
Bronchoscopic image	Clinical bronchoscope	A_i
Ground truth depth map	CT	B_i
Virtual bronchoscopic image	CT	C_i
Estimated bronchoscopic image	$G_{B_2}\hat{A}_2, G_{C_6}\hat{A}_6$	$\hat{A}_2, \hat{A}_6$
Estimated depth map	$G_{A_1}\hat{B}_1, G_{C_3}\hat{B}_3$	$\hat{B}_1, \hat{B}_3$
Estimated virtual bronchoscopic image	$G_{B_4}\hat{C}_4, G_{A_5}\hat{C}_5$	$\hat{C}_4, \hat{C}_5$

^{*} The superscript number represents the level of the network. The symbols presented are used in the Fig. 1 and equations (1)-(7) to describe the training process.

Table 2.

Characteristics of patient-specific video sequences and target nodules^{*}

Case #	CT dimensions	CT spacing	Lobe	Generation	Diameter [mm]	d [mm]
1	512x512x403	0.678x0.678x0.8	RUL	5	19	31
2	512x512x409	0.754x0.754x0.8	RUL	5	22	65
3	512x512x430	0.684x0.684x0.8	RUL	5	31	38
4	512x512x465	0.875x0.875x0.8	LUL	4	13	16
5	512x512x426	0.748x0.748x0.8	RUL	6	8	78

^{*}RUL = Right Upper Lobe, LUL = Left Upper Lobe, diameter = diameter of the nodule, d = distance from the last frame to the nodule.

Table 3.

Depth similarity metrics computed with cGAN and 3cGAN*

Location	MAE [mm]			RMSE [mm]			NCC		
	Model		p value	Model		p value	Model		p value
	cGAN	3cGAN		cGAN	3cGAN		cGAN	3cGAN	
Trachea	11.7±5.7	10.7±2.0	0.80	15.4±7.2	14.4±2.2	0.74	0.73±0.12	0.87±0.07	<0.001
Main Bronchus	11.9±4.6	7.2±2.1	<0.001	15.3±5.2	9.9±2.5	<0.001	0.77±0.15	0.9±0.05	<0.001
Lobar Bronchus	10.6±3.5	6.7±1.7	<0.001	13.4±3.9	9.0±1.7	<0.001	0.81±0.15	0.92±0.08	<0.001
All location	11.5±4.6	7.8±2.5	<0.001	14.9±5.5	10.6±3.0	<0.001	0.77±0.15	0.90±0.07	<0.001

* Values were given as the mean±standard deviation. Models were compared on each metric at each location using Wilcoxon signed-rank test. MAE, RMSE, and NCC stand for Mean Absolute Error, Root Mean Square Error and Normalized Cross Correlation.

Table 4.

The number of frames with thresholds at three locations

Location	Model	ATE		ATE	
		5 [mm]	p value	10 [mm]	p value
Trachea	cGAN	11 (12%)	<0.001	70 (74%)	0.0095
	3cGAN	46 (48%)		84 (88%)	
Main bronchus	cGAN	93 (35%)	0.78	191 (73%)	<0.001
	3cGAN	90 (34%)		250 (95%)	
Lobar bronchus	cGAN	33 (33%)	0.04	94 (93%)	0.11
	3cGAN	47 (47%)		87 (86%)	

*
P values were produced using Chi-square test comparing the models at each location.