BRANCH:Bifurcation Recognition for Airway Navigation based on struCtural cHaracteristics

Mali Shen[1](✉), Stamatia Giannarou[1], Pallav L. Shah[2], and Guang-Zhong Yang[1]

[1] Hamlyn Centre for Robotic Surgery, Imperial College London, London, UK
mali.shen09@imperial.ac.uk
[2] National Heart and Lung Institute, Imperial College London, London, UK

Abstract. Bronchoscopic navigation is challenging, especially at the level of peripheral airways due to the complicated bronchial structures and the large respiratory motion. The aim of this paper is to propose a localisation approach tailored for navigation in the distal airway branches. Salient regions are detected on the depth maps of video images and CT virtual projections to extract anatomically meaningful areas that represent airway bifurcations. An airway descriptor based on shape context is introduced which encodes both the structural characteristics of the bifurcations and their spatial distribution. The bronchoscopic camera is localised in the airways by minimising the cost of matching the region features in video images to the pre-computed CT depth maps considering both the shape and temporal information. The method has been validated on phantom and in vivo data and the results verify its robustness to tissue deformation and good performance in distal airways.

1 Introduction

Lung cancer remains a challenging disease with high mortality despite of the increasing knowledge of its aetiology. Data from the US National Lung Screening Trial suggests that early identification of lung cancer can lead to 20% reduction in mortality [1]. Trans-thoracic procedures such as CT guided biopsy have reasonable accuracy for targeting nodules greater than 20 mm but with high complication rates and surgical risks [2]. As an alternative, bronchoscopy provides a less invasive way for sampling pulmonary nodules but navigation in distal airways is particularly challenging due to the size and complexity of the bronchial tree anatomy.

To assist navigation during bronchoscopic procedures, Electromagnetic (EM) tracking and image registration approaches have been extensively investigated to localise the bronchoscopic camera in the airways [3]. The accuracy of EM tracking is limited by field distortions, inaccurate sensor calibration and most importantly airway deformation due to respiration and patient's motion. Image registration approaches essentially create a virtual camera using the patient specific pre-operative CT airway model and estimate its pose by minimising the difference between the video image and the virtual camera view [3]. The accuracy

M. Descoteaux et al. (Eds.): MICCAI 2017, Part II, LNCS 10434, pp. 182–189, 2017.
DOI: 10.1007/978-3-319-66185-8_21

of the image-based tracking approaches relies on the selection of the similarity measure between the video and virtual images [4,5]. Geometry-based similarity measures such as pq-space based registration [6] or depth-based registration [7] have also been proposed and shown to be more robust than intensity-based methods. Moreover, salient feature tracking has been used to estimate the motion of the bronchoscope. Luo et al. [8] proposed a tracking system combining Kalman filter, SIFT feature tracking and image registration. Wang et al. [9] proposed an endoscopic tracking approach based on Adaptive Scale Kernel Consensus (ASKC) estimator and feature tracking. The accuracy of these feature-based methods depends on the amount of correctly detected feature points on the bronchoscopic video. Due to the paucity of surface structure, illumination artefacts and tissue deformation in distal airways, the conventional image registration and feature-based approaches have limited clinical feasibility.

Thus far, the above navigation techniques have been mostly validated near the proximal airways. In this paper, we focus on the tracking towards segmental airways with increasing number of bifurcations, smaller bronchial size and larger respiratory displacement. A new approach is proposed for bronchoscope localisation during navigation in distal airways based on the matching of bronchoscopic data with virtual camera views from CT data. The Maximally Stable Extremal Region (MSER) detector [10] is applied in a novel fashion on depth maps instead of images of the airways to extract salient regions which are further filtered to identify bifurcations. A robust airway descriptor is proposed to encode both the structural characteristics of the airway bifurcations and their global spatial relationships. The proposed descriptor is based on shape context [11] and is tolerant to certain degree of airway deformation. Camera location is estimated by computing the optimal match between the airway features detected in the video images and those detected in the CT virtual views. Particle swarm optimisation was applied to minimise the matching cost for continuous tracking. The proposed localisation framework has been validated on phantom and in vivo data and the results verify the advantage of the method in recovering the location of the bronchoscope in distal airways.

2 Method

The proposed BRANCH approach consists of three parts: the detection of anatomically meaningful regions that represent airway bifurcations; the description of the shape characteristics and spatial relationship of airway regions and the localisation of the bronchoscopic camera using airway feature matching between CT and video data.

2.1 Detection of Airway Bifurcations on Depth Maps

Since geometric characteristics have been proven to be more robust to illumination artefacts and surface texture for bronchoscopic navigation than image appearance features [6,7], in our work depth maps are generated and used to extract features that represent airway structure. To generate depth maps from pre-operative data, a patient specific airway model is segmented from 3D chest CT scans. Fast marching is used to compute the centreline of the bronchial model. A virtual camera with the same intrinsic parameters as the bronchoscope is simulated and moved along the centreline from the trachea to each bronchiole. A CT reference depth map z_{CT} is generated at each point on the centreline with the camera direction being tangential to the centreline. The depth maps from the bronchoscopic video data z_V are recovered using a Shape From Shading (SFS) method tailored for the endoscopic environment [12].

The aim of our approach is to detect bifurcations and represent each part of the airway along the centreline based on the number of bifurcations, their shape, size and spatial association. For this purpose, the MSER detector is applied to extract a set of salient regions R_i from each depth map. A SVM classifier is used to distinguish between detected regions that correspond to airway bifurcations and noise detections such as wall regions which should be eliminated. Shape features including solidity, extent, eccentricity, as well as the minimal, median and maximal depth values of each region were used to train the classifier. The regions that have been classified as airway bifurcations are then organised in a tree structure. Region R_i is a child of Region R_j if $R_i \cap R_j = R_i$.

To further remove multiple detections of the same airway bifurcation regions which give redundant information, of all the child regions, only the largest region representing a unique airway bifurcation remains and will be considered in the airway description. Region R_i is a duplicated detection of region R_j if $R_i \cap R_j = R_i$ and $(R_j - R_i) \cap R_k = \{\}$ where $k \neq i, j$. The regions that have survived the above filtering stages correspond to airway branches and their contour represents the border of each bifurcation. The detection and filtering process of MSER regions on the depth maps of a video frame and its corresponding CT virtual image are illustrated in Fig. 1.

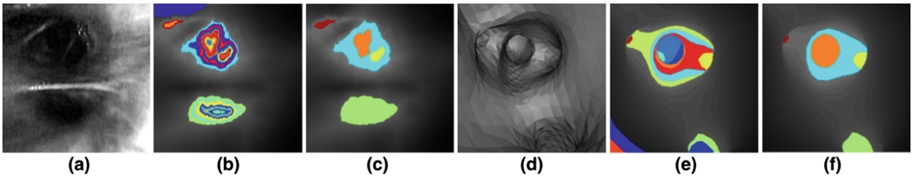

Fig. 1. The detection and filtering of MSER regions on the depth maps of (a)–(c) a video frame and (d)–(f) its corresponding CT virtual image.

2.2 Airway Bifurcation Representation

A novel airway descriptor based on shape context [11] is proposed in this work to characterise both the shape of each airway region and the geometrical association between them. The use of geometrical association provides more robust airway representation to shape variations between the regions detected on the CT and bronchoscopic video data due to lumen deformation under respiratory effect. In order to do that, boundary points ∂R are extracted from the airway regions R. For a point p_i on the boundary shape, its shape context is defined as a coarse histogram h_i of the relative coordinates of the remaining $n-1$ boundary points.

$$h_i\left(k\right) = \#\left\{q \neq p_i : \left(q - p_i\right) \in bin\left(k\right)\right\}, p_i, q \in \partial R \tag{1}$$

To incorporate scale information in our representation, the shape context histograms of all the boundary points are estimated with the same radius r_{ref} which is equal to the mean distance d_{mean} between all the boundary points of the detected bifurcations on the depth map of the CT virtual image. Regarding the orientation of the shape context histograms, for the video data, the reference orientation axis for the angular bins is the horizontal axis of the image. For the CT data descriptor, different orientations are considered in order to find the orientation θ that gives the best matching to the video data during camera localisation. The cost of matching point c_i on the CT boundary shape to point v_j on the video boundary shape based on their shape context is estimated using the χ^2 test statistic. The Hungarian method [13] is applied to minimise the total pairwise cost of matching those two sets of points to achieve the optimal permutation π.

$$H\left(\pi\right) = \sum_i C\left(c_i, v_{\pi(i)}\right) \tag{2}$$

For camera localisation as it will be explained in the next section, our aim is to estimate the pairwise cost of matching the airway regions detected in the video and those in CT data. The cost of matching an airway region R_{CT} in the CT virtual view to a region R_V in the video depth map is computed from the optimal permutation as:

$$C^{SC}(R_{CT}, R_V) = 1 - \frac{1}{2}\left(\frac{m}{n_{CT}} + \frac{m}{n_V}\right) \tag{3}$$

where n_{CT} is the number of boundary points of region R_{CT}, n_V is the number of boundary points of region R_V, and m is the number of matched pairs of boundary points between R_{CT} and R_V.

2.3 Camera Localisation

Camera localisation is achieved by finding the virtual camera view with the highest similarity to the examined video frame. For computational efficiency, the CT airway feature descriptors are pre-computed on the depth maps of the virtual camera views densely sampled along the airway skeleton from the trachea to the peripheral airways. Moreover, the camera is localised only at the video frames where the scene context on the video data changes significantly. To detect any context change of the video data, the detected airway regions are tracked along consecutive video frames using the Kalman filter based on a constant velocity model. The state of the Kalman filter is defined as $[x, y, u, v]$ where x and y

are the 2D location coordinates of the centroid of each region and u and v are the velocity of the centroid along the x and y axis, respectively. The Hungarian algorithm is applied to find the optimal match between the regions detected on consecutive video frames taking into account the distance between their centroid location and their size of area. The average matching cost is thresholded to identify the frames where a significant scene context change occurs in order to update the camera location with respect to the CT airway model.

To localise the bronchoscopic camera, both the shape context and temporal correspondence information are considered. The cost of region matching based on the shape context information is estimated as in Eq. 3. The temporal correspondence information is established by tracking the bifurcation regions on the CT and video data separately, using the Kalman filter described above. This is to deal with fast camera motion and partial occlusion of any airway regions in the video due to image artefacts. If region R_{CT} on the CT and region R_V on the video data have been previously matched and also successfully tracked on each data modality, a matching cost of 0 is assigned. Otherwise, the cost of matching two new regions is set to 1.

$$C^T(R_{CTi}, R_{Vj}) = \begin{cases} 0 & \text{if } \pi_R(i) = j \\ 1 & \text{otherwise} \end{cases} \tag{4}$$

The total pairwise cost of matching individual regions between a CT frame and a video image is defined as $C(R_{CTi}, R_{Vj}) = C^{SC}(R_{CTi}, R_{Vj}) + C^T(R_{CTi}, R_{Vj})$. Both cost matrices have been normalised within the range of $[0, 1]$.

In our work, the camera state is defined as $s = [d, \theta, l]$ where d is the distance of the camera location from the trachea point along the centreline, θ is the rotation around the centreline with respect to the initial orientation of the virtual camera along the centreline, and l is the centreline branch where the camera is located. For a given state s, there will be a unique feature descriptor which represents the pre-operative CT model of the airway. The estimation of the camera state is solved by minimising the total cost of matching the regions in the pre-computed CT depth maps to the video frames in Eq. 5.

$$\varphi\left(z_{CT}, z_V\right) = \min_{d,\theta,l} \left\{ \sum_i C\left(R_{CT_i}\left(d, \theta, l\right), R_{V_{\pi_R(i)}}\right) \right\} \tag{5}$$

$\pi_R(i)$ is the index of the matched region on the video image that corresponds to region i on the CT data. Particle swarm optimisation was applied to find the optimal camera state because the cost function of feature matching is not differentiable. The camera state of the previous frame is used to initialise the camera state for the next frame. The variation range of l is defined based on the current d and its variation range.

3 Results

The proposed tracking approach based on airway bifurcation recognition was validated on data from a silicon human airway phantom and a bronchoscopic examination. The bronchoscopic video data were performed with an Olympus BF-260 bronchoscope with an outer diameter of 5.5 mm and a field of view of 120°. Airway models were segmented from HRCT scans with a slice thickness of 1 mm acquired with a Siemens Somatom Definition Edge CT scanner. The BRANCH framework was implemented in MATLAB and runs at 3.7 s per video frame on a PC with i7-4770 CPU at 3.40 GHz without code optimisation.

The CT airway descriptors are computed on the depth maps sampled with a distance interval of 0.01 mm along the centreline of each CT airway model. The generated depth maps were normalised before computing the airway regions. 147 and 187 video frames with labelled bifurcation and noise regions were used to train the SVM classifier for phantom data and in vivo data, respectively. The number of radial and angular bins for the shape context estimation is set to 12 and 5, respectively. A log scale was used for binning the angular distances in the range of $[1/8, 2] \times d_{mean}$.

Ground truth data was manually generated for the in vivo experiments. The examined in vivo video sequences correspond to the longest sequences where continuous ground truth data could be manually generated. EM data was used as ground truth for the static phantom data. The performance of BRANCH has been compared to the state-of-the-art depth-based registration approach (Depth-Reg) proposed in [7]. The camera location estimated by Depth-Reg was projected to the closest centreline point for comparison with the ground truth data labelled on the centreline. Two sets of phantom data and two sets of in vivo data including 1330 phantom video frames and 374 in vivo video frames in total covering airway generation from 0(trachea) to 4 were used in the validation. The distance errors of the estimated trajectories on a set of phantom data and a set of in vivo data near distal airways are shown in Fig. 2. Quantitative analysis of the tracking accuracy of the two methods at different airway generations is provided in Fig. 3.

As shown in Fig. 3, the proposed BRANCH approach outperformed Depth-Reg approach with significantly higher accuracy in distal airway locations for both phantom and in vivo validation. Depth-Reg performed well for proximal airway locations only for the static phantom validation. The presence of tracheal cartilages on the phantom enhanced the accuracy of Depth-Reg approach near the trachea. However, BRANCH method outperformed Depth-Reg at higher airway generations for phantom data where more bifurcations could be observed.

The in vivo data used in our experiment are particularly challenging as they were collected from a subject with Excessive Dynamic Airway Collapse (EDAC) which causes airway obstructions during exhalation. In addition, the large deformation of the distal airways due to respiratory motion causes the shape and size of airway bifurcations appearing in the bronchoscopic video data varies significantly from the reconstructed CT model. Also, sudden camera movement from one airway branch to another are highly likely at the distal airways due to higher branch distribution. Despite of these challenges, BRANCH provides superior accuracy for all airway generations for the in vivo data (Fig. 3). The defined airway description based on shape context allows certain degree of variation

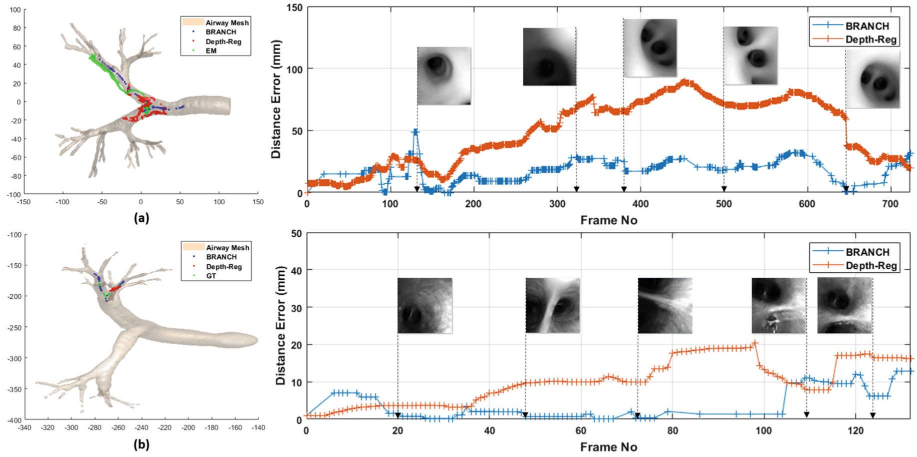

Fig. 2. Bronchoscope localisation accuracy for Depth-Reg [7] and BRANCH. (a) Phantom data using EM data as ground truth, (b) In vivo data using manually generated data as ground truth. Left: 3D trajectories of the camera movement shown in the CT airway mesh. Right: distance errors of the estimated camera locations.

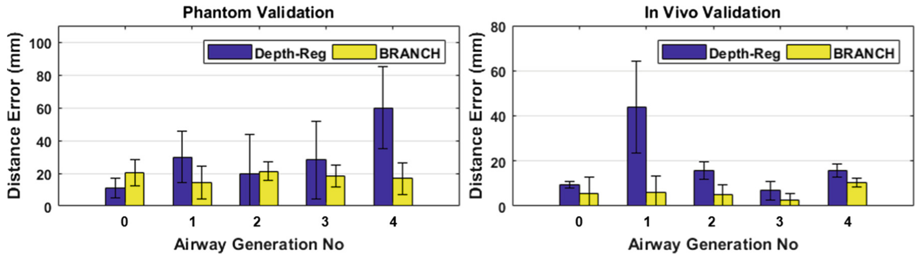

Fig. 3. Bar plot of camera localisation accuracy at different airway generations.

in the shape of bifurcations which can be modelled by affine transformation. Moreover, the spatial association of the bifurcations incorporated in the airway representation is not significantly affected by the tissue deformation.

Tracking can be temporarily affected if the airway features detected on the video images include noise or no bifurcations have been detected. Due to the poor lighting, lower lobe bifurcations are not clearly visible on the video images when the scope is inside the left or right main bronchi (long straight airway segments) such as at Frame 130 and 322 in Fig. 2(a)). However, when bifurcations appear again as the scope moves to the distal airways, correct tracking resumes (Frame 378, 500 and 645 in Fig. 2(a)). Water bubbles can cause false positive detections of bifurcations (Frame 109 in Fig. 2(b)). Video airway features are matched to the wrong CT airway features with false temporal location information. Correct tracking resumes when the scope moved back to the previously successfully tracked airway location (Frame 124 in Fig. 2(b)).

4 Conclusion

In this paper, the BRANCH framework has been proposed for robust bronchoscope localisation in distal airways. Airway bifurcations have been detected and a novel descriptor has been introduced based on the shape characteristics of bifurcations and their spatial relationship. The performance of the proposed method has been validated on phantom and in vivo data with significant tissue deformation, fast camera motion and image artefacts. The results verify the improved robustness of the BRANCH method in dealing with tissue deformation and distal airway tracking compared to the Depth-Reg method. The presented performance evaluation analysis shows the potential clinical value of the technique.

Acknowledgements. Dr. Giannarou is grateful for the support from the Royal Society (UF140290).

References

1. Aberle, D.R., Adams, A.M., Berg, C.D., et al.: Reduced lung-cancer mortality with low-dose computed tomographic screening. N. Engl. J. Med. **365**(5), 395–409 (2011)
2. Wiener, R.S., Schwartz, L.M., Woloshin, S., Welch, H.G.: Population-based risk for complications after transthoracic needle lung biopsy of a pulmonary nodule: an analysis of discharge records. Ann. Intern. Med. **155**(3), 137–144 (2011)
3. Reynisson, P.J., Leira, H.O., Hernes, T.N., et al.: Navigated bronchoscopy: a technical review. J. Bronchol. Interv. Pulmonol. **21**(3), 242–264 (2014)
4. Deguchi, D., Mori, K., Feuerstein, M., et al.: Selective image similarity measure for bronchoscope tracking based on image registration. MedIA **13**(4), 621–633 (2009)
5. Luo, X., 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)
6. Deligianni, F., Chung, A., Yang, G.Z.: pq-Space based 2D/3D registration for endoscope tracking. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2878, pp. 311–318. Springer, Heidelberg (2003). doi:10.1007/978-3-540-39899-8_39
7. Shen, M., Giannarou, S., Yang, G.Z.: Robust camera localisation with depth reconstruction for bronchoscopic navigation. IJCARS **10**(6), 801–813 (2015)
8. Luo, X., Feuerstein, M., Deguchi, D., Kitasaka, T., Takabatake, H., Mori, K.: Development and comparison of new hybrid motion tracking for bronchoscopic navigation. MedIA **16**(3), 577–596 (2012)
9. Wang, H., Mirota, D., Ishii, M., Hager, G.D.: Robust motion estimation and structure recovery from endoscopic image sequences with an adaptive scale kernel consensus estimator. In: IEEE Conference on CVPR, pp. 1–7 (2008)
10. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image Vis. Comput. **22**(10), 761–767 (2004)
11. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Trans. PAMI **24**(4), 509–522 (2002)
12. Visentini-Scarzanella, M., Stoyanov, D., Yang, G.Z.: Metric depth recovery from monocular images using shape-from-shading and specularities. In: IEEE International Conference on ICIP, pp. 25–28 (2012)
13. Papadimitriou, C.H., Steiglitz, K.: Combinatorial Optimization: Algorithms and Complexity. Courier Corporation, Mineola (1982)