Stable Anatomical Structure Tracking for video-bronchoscopy Navigation

A.Esteban-Lansaque[1], C.Sánchez[1], A.Borras[1], M.Diez-Ferrer[2], A.Rosell[2] and D.Gil[1]

[1]Comp. Vision Center, Comp. Science Dep. UAB.
[2]Pneumology Unit, Hosp. Univ. Bellvitge, IDIBELL, CIBERES.

Abstract. Bronchoscopy allows to examine the patient airways for detection of lesions and sampling of tissues without surgery. A main drawback in lung cancer diagnosis is the difficulty to check whether the exploration is following the correct path to the nodule that has to be biopsied. The most extended guidance uses fluoroscopy which implies repeated radiation of clinical staff and patients. Alternatives such as virtual bronchoscopy or electromagnetic navigation are very expensive and not completely robust to blood, mocus or deformations as to be extensively used. We propose a method that extracts and tracks stable lumen regions at different levels of the bronchial tree. The tracked regions are stored in a tree that encodes the anatomical structure of the scene which can be useful to retrieve the path to the lesion that the clinician should follow to do the biopsy. We present a multi-expert validation of our anatomical landmark extraction in 3 intra-operative ultrathin explorations.

Keywords: Lung cancer diagnosis, video-bronchoscopy, airway lumen detection, region tracking.

1 Introduction

Lung cancer is one of the most diagnosed cancers among men and women. Actually, lung cancer accounts for 13% of the total cases with a 5-year global survival rate in patients in the early stages of the disease of 38% to 67% and in later stages of 1% to 8% [1]. This manifests the importance of detecting and treating lung cancer at early stages, which is a challenge in many countries [2]. Computed tomography (CT) screening programs may significantly reduce the risk of lung cancer death. Diagnostic of solitary peripheral lesions can be diagnosed via bronchoscopy biopsy avoiding complications of other interventions such as transthoracic needle aspiration [3]. However, bronchoscopy navigation is a difficult task in case of solitary peripheral small lesions since according to the Am. Coll. Chest Phys., diagnostic sensitivity of lesions is 78%, but drops to 34% for lesions < 2 cm [4]. Actually, to reach a potential lesion bronchoscopists plan the shortest and closest path to the lesion exploring a pre-operative CT scan and, at intervention time, try to reproduce such a path by visual identification of bronchial levels and branch orientation in the bronchoscopy video.

Even for expert bronchoscopists it is difficult to reach a lesion due to the lung's anatomical structure. Images are commonly symmetrical so given a rotated broncoscope the direction to follow is not clearly defined. To assess the navigated path, bronchoscopists use a technique called fluoroscopy to obtain real-time X-ray images of the interior of the lungs. Aside from errors arising from visual interpretation, fluoroscopy implies repeated radiation for, both, clinical staff and patients [5]. In very recent years several technologies (like CT Virtual Bronchoscopy VB or electromagnetic navigation) have been proposed to reduce radiation at intervention time. Virtual Bronchoscopy VB (VB LungPoint or NAVI) is a computer simulation of the video bronchoscope image from CT scans to assess the optimal path to a lesion that, at intervention time, guides the clinician across the planned path using CT-video matching methods. Electromagnetic navigation (inReachTM, SpinDrive) uses additional gadgets which act as a GPS system that tracks the tip of the bronchoscope along the intervention. Although promising, these alternative technologies are not as useful as physicians would like. VB LungPoint and NAVI require manual intra-operative adjustments of the guidance system [6, 7], while electromagnetic navigation specific gadgets increase the costs of interventions limiting it use to resourceful entities.

Despite having increased interest in recent years, image processing has not been fully explored in bronchoscopy guiding system. Most of the methods are based on multi-modal registration of CT 3D data to video 2D frames. In [8], shape from shading (SFS) methods are used to extract depth information from images acquired by the bronchoscope to match them to the 3D information given by the CT. One of the disadvantages of such methods is that SFS is very time consuming so it cannot be implemented in real time systems. Other methods try to directly match virtual views of the CT to the current frame of the bronchoscope (2D-3D registration) to find the bronchoscope location [9]. Finally, there are hybrid methods [10] that use a first approximation using epipolar geometry that it is corrected by 2D-3D registration. These 2D-3D registration methods are also very time consuming and can lead to a mismatch in case images are obscured by blood or mucus and bronchi are deformed by patient's coughing.

Anatomical landmarks identified in, both, CT scans and videobronchoscopy frames might be a fast alternative to match the off-line planed path to interventional navigation. Landmark extraction in intra-operative videos is challenging due to the large variety of image artifacts and the unpredicted presence of surgical devices. Recent works [11] have developed efficient video processing methods to extract airways lumen centres that have been used in a matching system [12]. The system codified CT airways using a binary tree and used multiplicity of centres tracked in videos to retrieve the navigation path. In spite of promising results, the method was far from clinical deployment. A main criticism is a too simple matching criteria only based on lumen multiplicity which omitted the airway scene structure and the false positive rate in tracking.

We propose a method that extracts not just lumen centres but also stable lumen regions. Lumen regions are a better strategy for bronchoscopic navigation because they provide more information such as the area of the lumen (proximal,
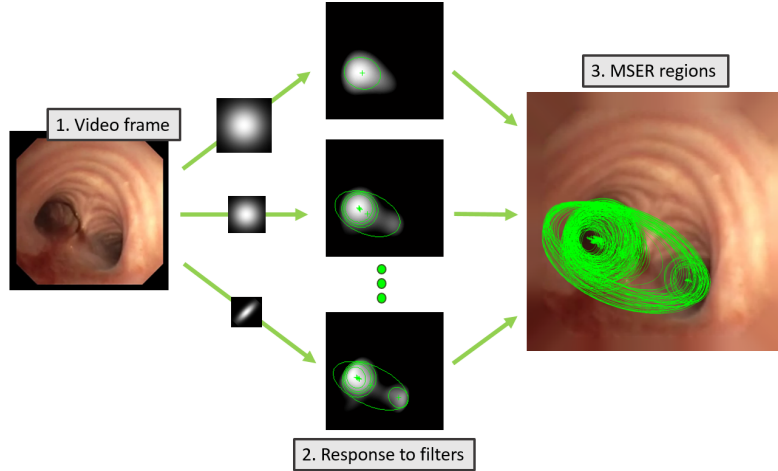
Fig. 1: MSER regions from likelihood maps.

distal) and altogether the hierarchy of the regions. In fact, we represent these regions in a tree that encodes the anatomical structure of bronchoscopic images. Besides, assuming slow motion, we track all the regions using a modified Kalman filter with no velocity and no acceleration that tracks the hierarchy of luminal regions. The capability of intra-operative luminal region tracking is assessed by a multi-expert validation in 3 intra-operative ultrathin explorations.

2 Stable Bronchial Anatomy Tracking

To retrieve bronchial anatomy from videos, lumen regions are extracted using maximally stable extremal regions (MSER) over a likelihood map of lumen center location. These regions are encoded with a hierarchical tree structure that filters regions inconsistent with bronchi anatomy in video frames. Finally, anatomically consistent regions are endowed with temporal continuity across the sequence using a modified Kalman filter.

2.1 Bronchial Anatomy Encoding in Single Frames

The first step to encode the anatomical structure of bronchoscopic images is to find lumen regions candidates. Extraction of lumen regions is based on likelihood maps [11] which indicate the probability of a point to be a lumen centre. In [11], such maps are computed using a single isotropic Gaussian kernel to characterize dark circular areas which under the assumption of central navigation are more probable to be lumen. The use of one single Gaussian kernel limits the extraction of lumen regions to circular regions of the same size which is not fully sensible in interventional videos. To model non-circular lateral bronchi and small distal levels, we compute several likelihood maps using a bank of
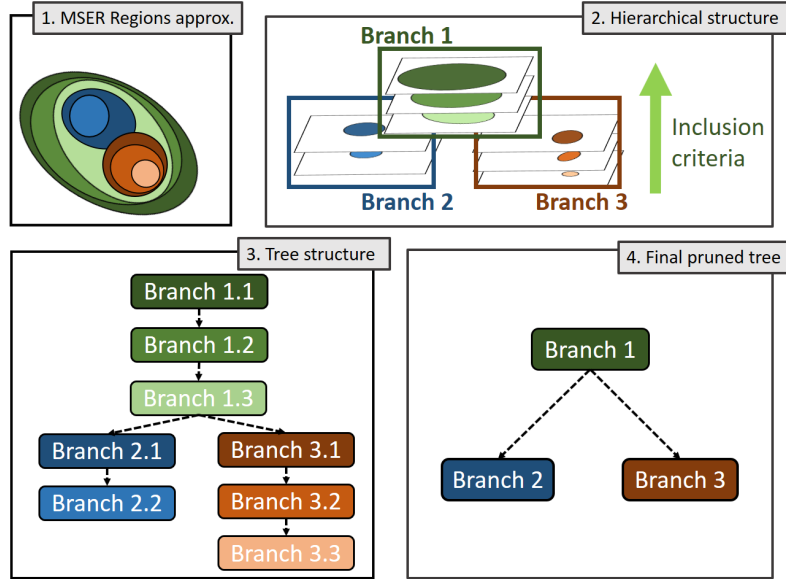
Fig. 2: Tree structure from MSER regions.

anisotropic Gaussian filters with different orientations and scales. Gaussian filters have been normalized by their L2 norm to obtain more uniform responses comparable across different scales and degrees of anisotropy. Figure 1 shows the likelihood maps (Fig.1.2) computed by convolving the left-hand side frame with the bank of Gaussian filters shown in side small images. To suppress outlying small local maxima, likelihood maximal regions are computed using maximally stable extremal regions (MSER) [13]. Finally, all MSER regions are put together (Fig.1.3) in order to be post-processed in next stages. We note that the collection of MSER regions are a set of elliptical regions following a hierarchy inclusions that should correspond to airways projected from different bronchial levels.

To extract projected airways anatomical structure from MSER regions, we encode their hierarchy using a n-tree using the strategy sketched in Fig.2. To better illustrate the tree creation we show a synthetic image (Fig.2.1) that simplifies the image in Fig.1.3 and a scheme of MSER hierarchy in Fig.2.2. Since each MSER region should be represented as a node of the tree, we iteratively construct the tree by keeping a list of root and children regions. First, MSER regions are sorted regarding their area in ascending order and the first region of the sorted list is added to the root node list and marked as current root. Then, we iteratively consider the next region in the sorted list, add it to the root list and update the children list according to whether the region contains any of the current roots. All roots contained in it are added in the tree structure as child of the node we are examining and are removed from the root list. The tree generated from the hierarchical structure of Fig.2.2 is shown in Fig.2.3. Ideally, we would like that each of the bronchial branches that represents a lumen region

(a) Frame i   (b) Frame i + 1

(c) Cost matrix
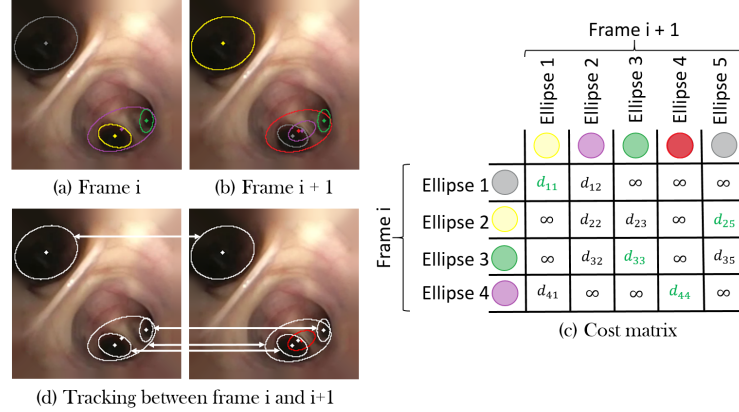
(d) Tracking between frame i and i+1

Fig. 3: Region tracking between two consecutive frames and cost matrix.

would correspond to a tree node. This is not the case due to the multiple MSER regions coming from different likelihood maps that lie on a bronchial lumen. Hence, the algorithm also prunes the tree keeping just the highest node of each branch to produce the final tree (Fig.2.4) encoding bronchial anatomy in images.

2.2   Bronchial Anatomy Tracking Across Sequence Frames

To endow MSER regions with temporal continuity, they are tracked using a modified Kalman filter. For each lumen region at a given frame, a Kalman filter [14] predicts its location in the next frame according to a motion model (constant velocity, constant acceleration) prone to fail in intra-operative videos because lumen movement does not fulfil any of those models. To reduce the impact of sudden variations in motion model, we have implemented a constant position tracker that uses a state vector with zero velocity and zero acceleration. In addition, instead of only using proximity of region centres to match them, we consider their overlap. In this way, our modified tracker matches nearby lumen regions only in case they maintain shape and area so that there is no mismatch in case lumens at different bronchial levels appear. To do so we compute a modified cost matrix with the euclidean distance between the centres of lumen regions at a time i and lumen regions at a time $i+1$. The trick is, when similarity ratio between those regions is small the distance is set to ∞ so that there is no matching between those regions. Finally, the Hungarian algorithm [15] is applied to the cost matrix for optimal matching.

Our tracking of luminal regions is illustrated in Fig.3. Fig.3a and Fig.3b show two frames at time i and $i + 1$ respectively with the luminal regions plotted as ellipses of different colors. Distances across ellipses at time i and $i + 1$ are given in the cost matrix shown in Fig.3c. In those images we can see that there are two regions which might be mismatched because of its proximity (ellipse 4 at frame i and ellipse 2 at frame $i + 1$) but their distance is set to ∞ because of its non-

	Proximal	Distal	Total
Prec	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]
Rec	[0.84, 0.99]	[0.91, 0.98]	[0.90, 0.97]

Table 1: Average precision and recall confidence intervals for region tracking.

similarity (region overlap). Since our tracker takes into account the position and the region overlap, it can clearly define the right match. This region matching allows to track lumens of different bronchi's levels and maintain the anatomical structure in the image. Finally, in Fig.3c we can see the regions which have been correctly matched (white) and those which have not (red).

3   Results

We have compared under intervention conditions the quality of the proposed tracking according to Section 2.2. Method has been applied to 8 sequences extracted from 3 ultrathin bronchoscopy videos performed for the study of peripheral pulmonary nodules at Hospital de Bellvitge. Videos were acquired using an Olympus Exera III HD Ultrathin videobronchoscope. We have split the 8 sequences into proximal (up to 6th division) and distal (above 6th) sets to compare also the impact of the distal level. The maximum bronchial level achieved in our ultrathin explorations was within 10th and 12th, which is in the range of the maximum expected level reachable by ultrathin navigation. Sequences contain bronchoscope collision with the bronchial wall, bubbles due to the anaesthesia and patient coughing.

For each sequence, we sampled 10 consecutive frames every 50 frames. Such frames were annotated by 2 clinical experts to set false detections and missed centres. To statistically compare our tracker, ground truth was produced by intersecting the experts' annotations. Ground truth sets were used to compute precision (Prec) and recall (Rec) for each set of consecutive frames. These scores are taken for all such sets in distal and proximal fragments for statistical analysis. We have used a Wilcoxon test data to assess significant differences and confidence intervals, CI, to report average expected ranges. Table 1 reports CIs for each set of consecutive frames score at proximal, distal and all together (both proximal and distal) levels. According to these results, it is worth noticing that the proposed method always has a 100% of precision and a recall over 84%. We can see that Recall at proximal levels is a bit smaller than recall at distal levels. This is due to more frames with collisions at proximal levels that distort the likelihood model (see Discussion Section). Even so, proximal and distal levels present non-significant differences between them ($p - val > 0.8$ for a Wilcoxon test).

Figure 4 shows regions tracked in consecutive frames selected at distal and proximal levels. It is worth noticing the capability of our strategy to capture most distal and lateral bronchi without introducing false positives.
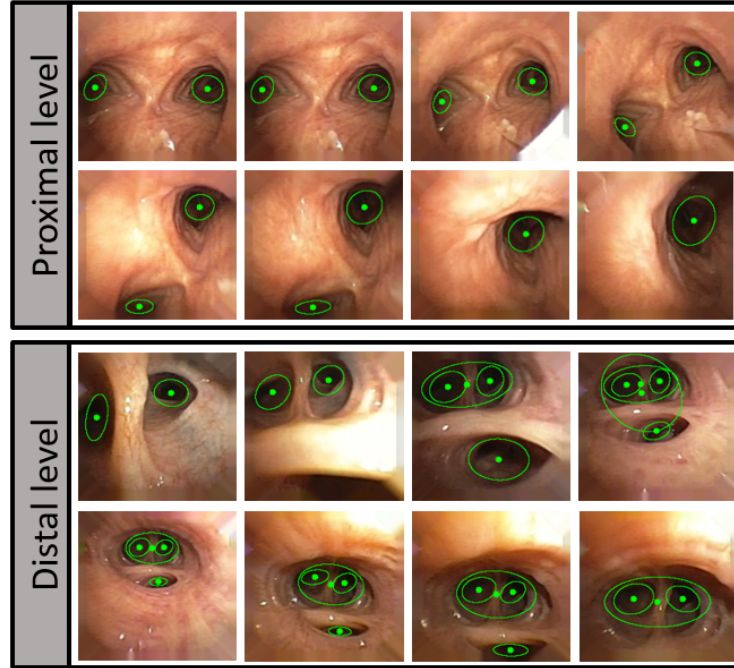
Fig. 4: Frames of tracked regions at proximal and distal levels.

4 Discussion and Conclusions

We have introduced a method that extracts and tracks stable lumen regions at different levels of the bronchial tree. The tracked regions encode the anatomical structure of the scene which can be useful to retrieve the path to the lesion that the clinician should follow to do the biopsy. Results in ultrathin bronchoscopy videos indicate high equal performance of our lumen region tracker based on MSER at proximal and distal levels. Particularly, there are not any false detections (Prec=1) and the rate of missed lumen regions is under 16% (Rec > 0.84). Although, non-significant according to a Wilcoxon test, we can appreciate a slight deviation between proximal and distal recall. The reason for such bias is that our model does not satisfy the illumination conditions in carina when collisions happen. This could be solved by making the likelihood maps less restrictive at proximal levels, but does not invalidate our system for bronchoscopic navigation. Clinicians need guiding systems for distal levels in which we obtain a recall grater than 90%, at proximal levels, they navigate without any tool just by visually assessing the path.

We conclude that results are promising enough (see the full exploration at `https://www.youtube.com/watch?v=CWEHX2KP8YI`) to encourage the use of anatomical landmarks in a biopsy guidance system. In Fig. 4 we can see 8 sample images from two videos at distal and proximal level. Images are ordered according

to its occurrence in time from left to right and from up to down. As we can see, at proximal levels the anatomical structure of bronchi is easy but at distal levels it becomes more complex. This complex anatomical structure could be used to put in correspondences the anatomical structure extracted from the CT and the anatomical structure extracted from frames recorded by the bronchoscope.

ACKNOWLEDGMENTS

References

1. A. Jemal, F. Bray, M. M. Center, J. Ferlay, and et al., "Global cancer statistics," *CA: a cancer journal for clinicians*, vol. 61(2), pp. 69–90, 2011.
2. P. J. Reynisson, H. O. Leira, T. N. Hernes, and et al., "Navigated bronchoscopy: a technical review," *J Bronchology & Interv Pulmo*, vol. 21(3), pp. 242–264, 2014.
3. A. Manhire, M. Charig, C. Clelland, F. Gleeson, R. Miller, and et al., "Guidelines for radiologically guided lung biopsy," *Thorax*, vol. 58(11), pp. 920–936, 2003.
4. E. F. Donnelly, "Technical parameters and interpretive issues in screening computed tomography scans for lung cancer," *J Thor Imag*, vol. 27(4), pp. 224–229, 2012.
5. R. W. Shepherd, "Bronchoscopic pursuit of the peripheral pulmonary lesion: navigational bronchoscopy, radial endobronchial ultrasound, and ultrathin bronchoscopy," *Current opinion in pulmonary medicine*, vol. 22 (3), pp. 257–264, 2016.
6. R. Eberhardt, N. Kahn, D. Gompelmann, and et al., "Lungpointâa new approach to peripheral lesions," *J Thor Onco*, vol. 5 (10), pp. 1559–1563, 2010.
7. F. Asano, Y. Matsuno, and et al., "A virtual bronchoscopic navigation system for pulmonary peripheral lesions," *Chest*, vol. 130 (2), pp. 559–566, 2006.
8. M. Shen, S. Giannarou, and G.-Z. Yang, "Robust camera localisation with depth reconstruction for bronchoscopic navigation," *IJCARS*, vol. 10 (6), pp. 801–813, 2015.
9. L. Rai, J. P. Helferty, and W. E. Higgins, "Combined video tracking and image-video registration for continuous bronchoscopic guidance," *IJCARS*, vol. 3 (3-4), pp. 315–329, 2008.
10. X. Luó, M. Feuerstein, D. Deguchi, and et al., "Development and comparison of new hybrid motion tracking for bronchoscopic navigation," *MedIma*, vol. 16 (3), pp. 577–596, 2012.
11. C. Sánchez, J. Bernal, D. Gil, and F. J. Sánchez, "On-line lumen centre detection in gastrointestinal and respiratory endoscopy," in *CLIP*, pp. 31–38, Springer, 2013.
12. C. Sánchez, M. Diez-Ferrer, J. Bernal, and et al., "Navigation path retrieval from videobronchoscopy using bronchial branches," in *CLIP*, pp. 62–70, Springer, 2015.
13. J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Im Vis Comp*, vol. 22 (10), pp. 761–767, 2004.
14. S. S. Haykin, S. S. Haykin, and S. S. Haykin, *Kalman filtering and neural networks*. Wiley Online Library, 2001.
15. H. W. Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.