AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

PLEASE NOTE: Please note that we cannot accept new source files as corrections for your article. If possible, please annotate the PDF proof we have sent you with your corrections and upload it via the Author Gateway. Alternatively, you may send us your corrections in list format. You may also upload revised graphics via the Author Gateway.

Carefully check the page proofs (and coordinate with all authors); additional changes or updates WILL NOT be accepted after the article is published online/print in its final form. Please check author names and affiliations, funding, as well as the overall article for any errors prior to sending in your author proof corrections. Your article has been peer reviewed, accepted as final, and sent in to IEEE. No text changes have been made to the main part of the article as dictated by the editorial level of service for your publication.

AQ:1 = We have updated the author name CARLES SÁNCHEZ RAMOS as per the metadata. Please check and confirm the representation of author name in all instances.

AQ:2 = Please confirm the location for Universitat Autònoma de Barcelona.

AQ:3 = Please confirm or add details for any funding or financial support for the research of this article.

AQ:4 = If you haven't done so already, please make sure you have submitted a graphical abstract for your paper. The GA should be a current figure or image from your accepted article. The GA will be displayed on your articles abstract page on IEEE Xplore. Please choose a current figure from the paper and supply a caption at your earliest convenience for the graphical abstract. Note that captions cannot exceed 1800 characters (including spaces). If you submitted a video as your graphical abstract, please make sure there is an overlay image and caption. Overlay images are usually a screenshot of your video that best represents the video. This is for readers who may not have access to video-viewing software. Please see an example in the link below: http://ieeeaccess.ieee.org/submitting-an-article/

AQ:5 = Note that Refs. [6] and [9] are identical; hence, we have deleted Ref. [9] and have renumbered the subsequent references. This will also be reflected in the citations present in the body text. Please confirm.

AQ:6 = Please provide the publisher location for Refs. [10], [11], and [19].

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 10, 2020, accepted August 7, 2020. Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2020.3018590

Intraoperative Extraction of Airways Anatomy in VideoBronchoscopy

AQ:1 **DEBORA GIL, ANTONIO ESTEBAN-LANSAQUE, AGNÉS BORRÀS, ESMITT RAMÍREZ, AND CARLES SÁNCHEZ RAMOS**

AQ:2 Computer Vision Center, Department of Computer Science, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

Corresponding authors: Debora Gil (debora@cvc.uab.cat) and Carles Sanchez Ramos (csanchez@cvc.uab.cat)

AQ:3 This work was supported in part by Spanish projects under Grant RTI2018- 095209 -B-C21 and Grant FIS-G64384969, in part by the Generalitat de Catalunya under Grant 2017-SGR-1624, in part by the CERCA-Programme, in part by the European Union Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie under Grant 712949 (TECNIOspring PLUS), and in part by the Agency for Business Competitiveness of the Government of Catalonia.. The work of Debora Gil was supported by Serra Hunter Fellow. The work of Esmitt Ramírez was supported by the Ministry of Economy, Industry and Competitiveness, Spain, under Grant BES-2016-078042.

AQ:4 **ABSTRACT** A main bottleneck in bronchoscopic biopsy sampling is to efficiently reach the lesion navigating across bronchial levels. Any guidance system should be able to localize the scope position during the intervention with minimal costs and alteration of clinical protocols. With the final goal of an affordable image-based guidance, this work presents a novel strategy to extract and codify the anatomical structure of bronchi, as well as, the scope navigation path from videobronchoscopy. Experiments using interventional data show that our method accurately identifies the bronchial structure. Meanwhile, experiments using simulated data verify that the extracted navigation path matches the 3D route.

INDEX TERMS Bronchial anatomy representation, videobronchoscopy, lung cancer, biopsy guidance.

I. INTRODUCTION

Lung cancer early-stage detection increases the survival rate over 5-years from 38% to 67% [1]. Currently, cancer diagnosis can only be achieved by analysis of tissue extracted from the lesion usually sampled using ultrathin bronchoscopic navigation. Biopsy sampling using videobronchoscopy is a two-stage procedure. First, the intervention is planned off-line using Virtual Bronchoscopy (VB) [2] to compute from computed tomography (CT) data the shortest path across bronchial levels to each nodule. Second, the bronchoscopist tries to reproduce the pre-planned route by visual identification of bronchial levels and branch orientation in the intra-operative bronchoscopy video.

Even for expert bronchoscopists it is difficult to reach distal lesions due to the lung's anatomical structure. Conventional bronchoscopic diagnostic procedures are visually guided using radiating fluoroscopy which renders a suboptimal 34% of positive results for lesions <2 cm [3]. New endoscopy techniques (like electromagnetic navigation) are expensive, require either manual intervention or special gadgets, only increase diagnostic yield to 70%, and still radiate the patient.

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello.

The 30% undiagnosed pulmonary lesions need CT follow-up or futile surgery procedures such as thoracoscopies, which induces patient anxiety, radiation exposure, invasive surgery along with associated pain, disability and rarely death. Diagnostic yield could be improved reducing radiation and costs by developing intervention support systems able to guide the bronchoscopist to the lesion.

During the past years, several technologies have been developed for on-line guiding the bronchoscopist through the planned path. Existing systems can be split into purely image-based navigation systems and systems, like electromagnetic navigation, that use specific tools that provide additional information helpful in the guidance process.

Electromagnetic Navigation Bronchoscopy [4] (ENB) is a medical procedure designed to localize and guide both bronchoscope and bronchoscopic tools through the bronchial tree by means of electromagnetic waves. The main disadvantage of ENB is that it increases the cost of interventions, lacks of rotational information and might not be accurate enough due to interferences between the electromagnetic waves and human tissues. Image-based navigation systems try to put into correspondence video bronchoscopy images and VB images using multimodal registration techniques [5]–[7]. Unfortunately, the synchronization of simulated navigation with the actual intra-operative video is a challenging task that often requires manual corrections during intervention.

In recent years, several alternatives to classical intensity-based registration methods have begun to be explored. These new image-based methods include Simultaneous Localization and Mapping (SLAM) [6], [8], Convolutional Neural Networks (CNNs) to compute 3D point clouds from depth maps learned from images [9] and synchronization based on visual features that can be identified in, both, CT scans an videos [10], [11].

SLAM methods estimate a 3D map of the unknown environment together with the camera position from a set of feature points matched in consecutive frames. A main challenge in endoscopic explorations is the matching step, since images lack of enough texture and salient points. In this context, ORBSLAM [6] achieves higher performance with respect to other state-of-the-art monocular SLAM approaches because it uses feature detectors adapted to the specific surgical conditions of endoscopic interventions. ORBSLAM performance in bronchoscopic interventions is not clear due to the specific scope forward-backward motion along the camera viewpoint direction and sudden rapid motions. In a recent work [8], the authors presented an improved version of VSLAM to improve point matching in case of sudden motion of the bronchoscope. A main inconvenience of SLAM approaches as navigation systems is that they still require registration of the point cloud to a segmentation of bronchi in order to localize the scope inside airways.

Despite excellent results of CNNs in the medical imaging domain [12], [13], a main inconvenience is the limited annotated data available. In the particular field of intervention guiding, gathering annotated data has the extra difficulty of intra-operative recordings probably requiring the alteration of standard protocols. The deep learning approach presented in [9] uses phantom synthetic data to train a deep SLAM that provides the depth of the bronchoscopic image. Like classic SLAM approaches, the position of the camera inside airways requires further processing of the data.

A feasible alternative which does not require huge annotated datasets in order to train complex methods is the use of anatomical landmarks [10]. In [10], lumen center lines were used to codify the route followed by the bronchoscope and indicate the path that needs to be followed to reach a target lesion.

A navigation system based on anatomical landmarks requires the codification of airways main anatomy in, both, intra-operative videos and CT-scans. This work focuses in the intra-operative identification of a route planned on CT-scans using a codification of airways main anatomy.

This work extends [11] to obtain the bronchial path followed during a bronchoscopic exploration. In [11] bronchial anatomy was encoded in single videobronchoscopy frames as a hierarchy of ellipses representing the bronchial levels observed in each frame. In the presented work we track such hierarchy across frames to dynamically extract a representation of the full anatomy navigated during the intervention. This work contributes to the identification of the scope position during intervention in two aspects:

1) **Intra-operative Extraction of the Anatomy Explored.** The anatomy of airways can be represented using a tree data structure with nodes representing bronchi branching points [10]. In bronchoscopy videos, airways anatomical structure is projected into a collection of luminal regions arranged in a hierarchy of inclusions in case frames show different bronchial levels [11]. The tracking of such hierarchy across the video defines a tree that represents the global anatomy of the patient observed during the intervention. In order that this on-line exploration tree encodes all the bifurcations traversed during the intervention, its nodes codify the traversed luminal regions as ellipses and its edges the branching levels hierarchy.
2) **Codification of the Scope Navigation Path.** Nodes also keep a flag to indicate whether they are currently tracked, in order to identify the anatomy that it is observed at each time. The intra-operative navigation path is codified as the list of the roots of the sub-tree representing such observed anatomy. Using this representation, the current position of the scope inside airways could be matched to the anatomical structure of the lung extracted from a CT [10], [14].

II. EXTRACTION OF THE ON-LINE EXPLORATION TREE

Our strategy for the extraction of the anatomy observed in a bronchoscopy (sketched in fig.1) has 3 main steps. The first step is to track the anatomy observed in a single frame across the video (fig.1 (1)). To do so, the anatomical hierarchy of ellipses extracted using [11] is matched to the current on-line tree using a measure of anatomical similarity based on ellipses overlap. Second, the exploration tree is updated depending on the type of match between the tracked tree and the image hierarchy (fig.1 (2)). In case all ellipses in the hierarchy are matched, we consider that the scope has not change the bronchial level and update the anatomical information of the nodes with the ellipse hierarchy parameters they have been matched to. In case of a partial match between one of the tracked node children and the ellipse root of the hierarchy, we consider that the scope is approaching a deeper bronchial level and add a new level to the on-line tree. Finally, the exploration path is dynamically obtained as the sequence of tracked roots of the on-line tree (fig.1 (3)). In case of adding a new tree level to the current tracked root, we have a forward motion (F) of the scope and the current tracked root is added to the exploration path. In case the children of the current tracked root are unmatched, we consider the scope moves backwards (B) and the root is removed from the exploration path.

The next sections give details about the computation of each of the steps required for the extraction of the on-line exploration tree.

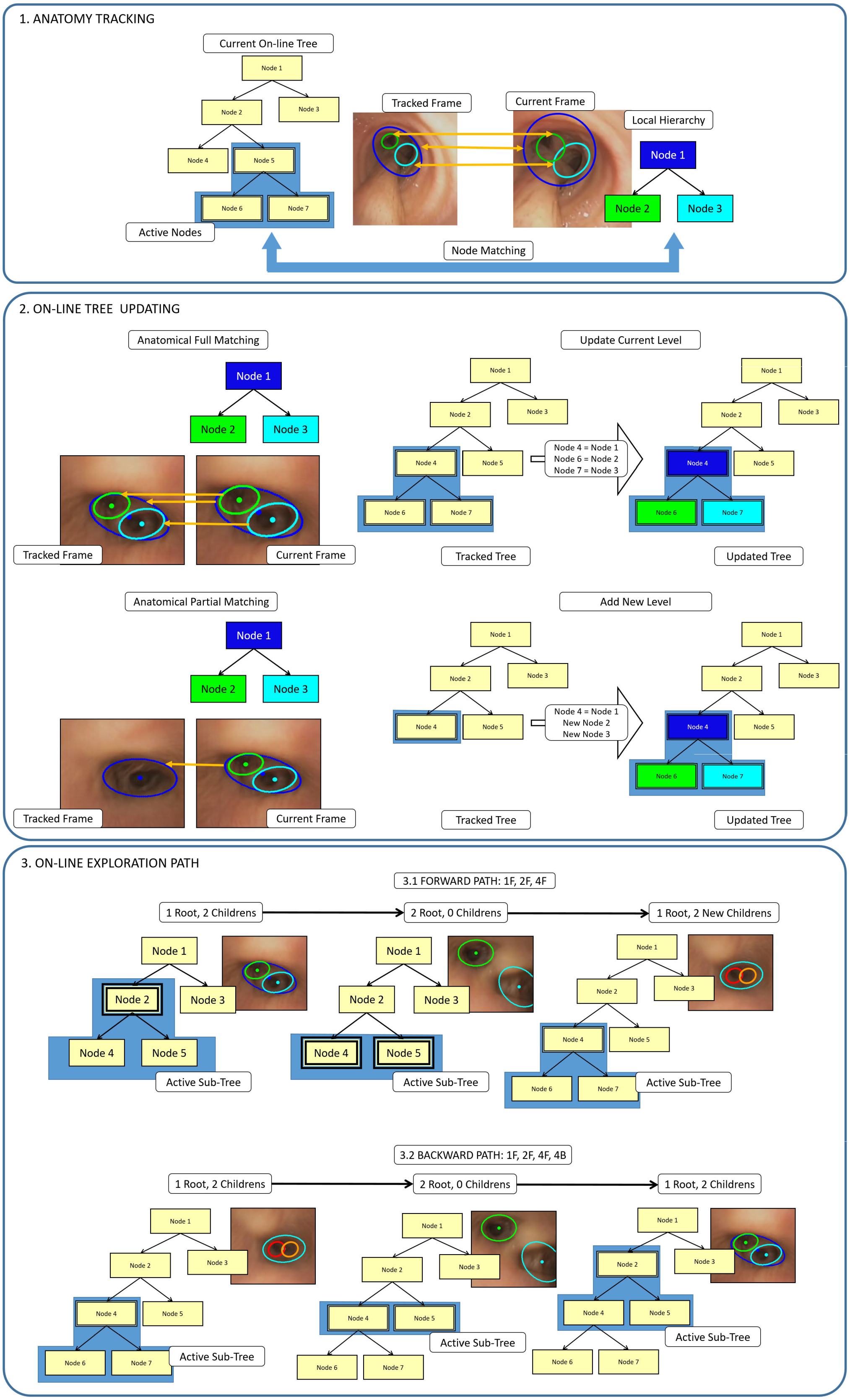


FIGURE 1. **Main steps in the extraction of the on-line exploration tree.**

A. ANATOMY TRACKING

Each node of the exploration tree stores the shape and position in the image of the luminal region it represents. Luminal shape is codified by the parameters of an ellipse representing the luminal region: center position, (x, y), mayor and minor axis, (a, b), and its orientation, θ. Such ellipses are obtained from a hierarchy of MSER regions computed using [11]. The lumen position in the image frame is codified by the image quadrant, Q, the ellipse representing the lumen is in. Image quadrants are stored in order to provide guidance instructions indicating the bronchi to follow at each bronchial level. Nodes also store temporal information. In particular, the number of frames, NFr, it has been tracked and a boolean flag (labelled *active*) indicating whether the node corresponds to a lumen that is currently being observed in the video frame (*active* $=$ *True*) or it corresponds to a bifurcation seen in previous frames (*active* $=$ *False*). A given node is activated when it has been tracked (i.e. matched to an ellipse in the hierarchy of MSER regions) for at least NFr_{Mx} frames and deactivated otherwise.

In order to track (and update) the on-line exploration tree, the active nodes sub-tree is matched to the ellipse hierarchy. To do so, we use a tree edit distance [15] that takes into account airways anatomy given by their elliptical representation.

The tree edit distance computes the optimal set of edit operations that transforms a tree T_1 into a tree T_2. In the case of trees representing bronchial anatomy, aside the identity transformation, we have two edit operations, insert and delete. Node deletion corresponds to traversing a bronchial level, while node insertion indicates that the bronchoscope is approaching a deeper bronchial bifurcation.

In order to define a criterion for the selection of the optimal transformation, each edit operation is assigned a cost which is used to compute the total cost as the sum of the costs of all edit operations of the transformation. Then, the edit distance is the transformation of minimal cost. In our case, the identity has zero cost and the cost of the other edit operations (insert and delete) is set to 1. Nodes in T_1 are deleted and inserted depending on whether they are matched to a node in T_2 or not. A node in T_1 is deleted if it can not be matched to any node in T_2. All unmatched nodes in T_2 are, then, inserted into T_1. Node matching is given in terms of a similarity measure. If we note by $(n_j^i)_{i=1}^{N_j}$ the nodes of tree T_j, $j = 1, 2$, and $sim(\cdot, \cdot)$ the similarity measure, then $n_1^{i_1}$ is matched to $n_2^{i_2}$ if:

$$n_2^{i_2} = \max_i \left\{ sim(n_1^{i_1}, n_2^i),\ \text{for } sim(n_1^{i_1}, n_2^i) \geq Th \right\} \tag{1}$$

being Th a tolerance parameter on a minimum similarity between nodes. In case $sim(n_1^{i_1}, n_2^i) < Th$, $\forall i = 1, \ldots N_2$, then $n_1^{i_1}$ is deleted.

In our case, the similarity between nodes is given by the overlap between the ellipses they represent. This overlap is computed as the volumetric overlap error (VOE) between the masks of the nodes' ellipses. If n_j^{ij} represents an ellipse with parameters $(x_j^{ij}, y_j^{ij}, a_j^{ij}, b_j^{ij}, \theta_j^{ij})$, $j = 1, 2$, then its elliptical region, namely $E_{nj}^{ij}(x, y)$, fulfills the following inequality:

$$E_{nj}^{ij}(x, y) = \frac{(x\cos(\theta_j^{ij}) - y\sin(\theta_j^{ij})) - x_j^{ij}}{a_j^{ij}} + \frac{(y\cos(\theta_j^{ij}) + x\sin(\theta_j^{ij})) - y_j^{ij}}{b_j^{ij}} \leq 1 \tag{2}$$

It follows that the VOE defining the similarity between $n_1^{i_1}$ and $n_2^{i_2}$ is given by:

$$\begin{aligned} sim(n_1^{i_1}, n_2^{i_2}) &:= E_{n_1}^{i_1}(x, y) \cap E_{n_2}^{i_2}(x, y) \\ &= \frac{2 \cdot |(E_{n_1}^{i_1}(x, y) < 1) \cap (E_{n_2}^{i_2}(x, y) < 1)|}{|E_{n_1}^{i_1}(x, y) < 1| + |E_{n_2}^{i_2}(x, y) < 1|} \end{aligned} \tag{3}$$

where $|\cdot|$ indicates the number of pixels of the mask approximating the ellipse area.

B. EXPLORATION TREE UPDATING

In order to update the exploration tree, we compute the edit distance transformation between the active sub-tree, T_1, defined by the active nodes and all their children and the ellipse hierarchy encoded as a tree, T_2. The active sub-tree is updated depending on the edit operation (identity, delete or insert) that it should be applied to transform it to the hierarchy tree.

For those nodes having a match (identity edit operation) to one of the ellipses in the hierarchy, we update all their values. The anatomical parameters (ellipse and quadrant) are set to the values of the ellipse they have been matched to. The number of tracked frames, NFr, is increased by one and the flag *active* is updated if $NFr \geq NFr_{Mx}$. We observe that a full match of all nodes in T_1 indicates that we are still in the same bronchial level.

For those nodes in the active subtree that should be deleted, the anatomical information remains unchanged, NFr, is decreased by one and the flag active is updated if $NFr < NFr_{Mx}$. A deactivation of the deleted nodes indicates the possibility of a change in the bronchial level. If the deactivated node is the root, the scope is approaching a deeper bronchial level, while in case of deactivating the children, the scope is moving backwards to a previous bronchial level. We observe that in the first case, the active sub-tree could have two roots, one for each children of the deactivated root.

Finally, the unmatched nodes of the hierarchy are inserted into the active sub-tree as new nodes with anatomical values equal to the values of the hierarchy nodes' ellipses, $NFr = 1$ and *active* $=$ *False*. We observe that the inserted nodes will be children of an active root and,, thus, indicate the scope is moving forward towards the next bronchial level of the active root. The final exploration tree is given only by those nodes that have been active at least once.

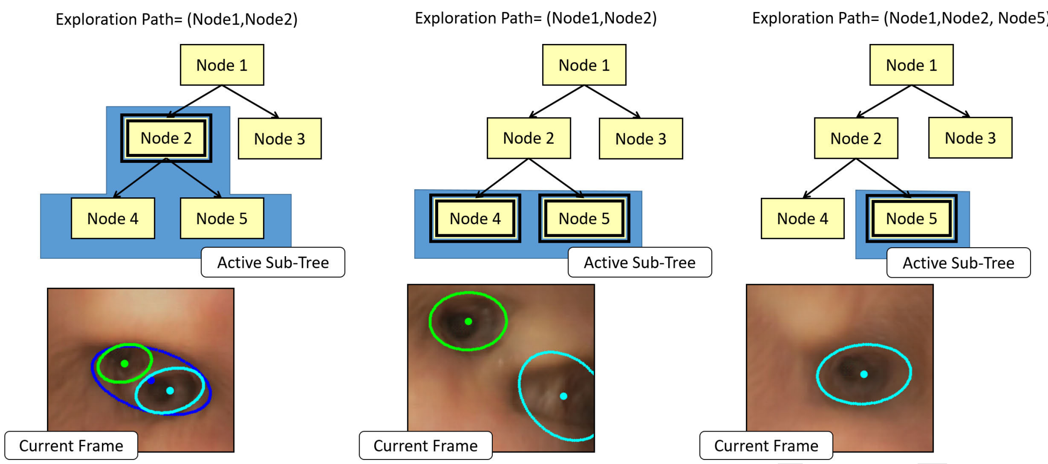


FIGURE 2. **Definition of the exploration path from the active roots in forward motion.**

C. EXPLORATION PATH EXTRACTION

The sequence of active roots provides information about the current position of the scope, as well as, the path navigated so far. The exploration path is codified as a list of the nodes that have been single roots in the active sub-tree. The quadrants of these nodes codify the navigation path. We add an extra attribute to the nodes in the list in order to indicate the direction of the scope motion at the moment of traversing the bifurcation: *F* for forward motion, *B* for backwards motion. The position of the scope is given by the last node inserted in the exploration path list. In particular, the level of the last node inside the whole exploration tree defines the current bronchial level.

The exploration path list is computed as follows. The path is initialized with the root (which represents the tracheal entry point) of the full exploration tree with forward motion. Nodes are added to the list as new levels are traversed using the following criterion. In forward motion, each time the scope approaches a deeper level, two new children are added to the current active root which is the last node inserted in the path list. At the final approach phase, such root is deactivated and its children become active roots. The moment the scope traverses to the next bronchial level, only the root representing the lumen the scope has entered into remains active. In backwards motion, we would have the inverse sequence of activations/desactivations. In any case, traversing a bronchial level can be detected as an increment in the number of active roots followed by a decrement. Every time this condition is satisfied, the current active root is added to the path list. Forward and backward motion is determined depending on whether the added active root is a children or a parent of the last inserted node. In case of being a children, we have a forward motion to a deeper level, while in case of being a parent, the scope is moving backwards to a level already visited.

Figure 2 graphically sketches the process of adding a new node to the exploration path list in case of forward motion. The nodes of the active sub-tree are indicated with a blue frame with the root in double line. We also show the ellipses the the active sub-tree nodes represent in a bronchoscopy image for better interpretation of the anatomical changes that take place during level traversal. In fig.2 (a), the scope is placed at a point of Node2 bronchi that allows the visualization of the next level luminal areas, represented as the light ellipses included in Node2 dark ellipse. As the scope moves forward (fig.2(b)), Node2 lumen does not show any more in images and, thus, it is deactivated. At this moment, the active sub-tree has two roots, Node4 and Node5. Finally, in fig.2(c), the scope has entered into Node5 bronchial level, so that this node becomes the only root of the active sub-tree and is added to the path list.

III. EXPERIMENTS

Our experiments have been designed to evaluate our method in two aspects:

1) **Assessment of the Exploration Tree.** The extraction of the on-line exploration tree has been tested on 8 interventional videos acquired at Hospital de Bellvitge (Barcelona, Spain) using an Olympus Exera III HD Ultrathin videobronchoscope. Videos were acquired during 4 biopsy sampling procedures. For each procedure, 2 different videos were recorded, one navigating a lower lobe and the other one navigating an upper lobe.
In order to validate the extraction of the on-line exploration tree, a clinical expert visually inspected each of the interventional videos to create a Ground Truth (GT) tree with all the bronchi seen across the video. The extracted on-line tree was compared to this GT tree using a standard tree edit distance in order to compute false positives, FP, false negatives, FN, precision and recall of the on-line exploration tree. False positives correspond to structures wrongly identified

as bronchial lumens, while false negatives are missed luminal regions.

2) **Assessment of the Exploration Path.** To assess the exploration path, we have computed the path list in VB to compared it to the list of bifurcations traversed by the VB camera [16]. Virtual bronchoscopies were generated using an own developed software from CT scans [16] and augmented with intra-operative appearance using the method presented in [17], [18]. For each virtual bronchoscopy of a patient, four virtual explorations were generated, covering the four main lobes: left and right upper lobes, noted LUL, RUL, and left and right lower lobes, noted LLL, RLL. Exploration paths reached between the sixth and twelfth bronchial level. Simulations were performed using central navigation without rotation around the scope.
We have compared the quality of our exploration path to the paths computed using [14]. Following [19], the metrics used for the comparison of both methods are True Positives Nodes (TPN) and True Path Representations (TPR). For a given exploration, a node is considered to be a TPN if its label coincides with the GT node label. The number of consecutive TPN achieved from the 1st node divided by the path node length defines TPR. We have used a T-test to detect significant differences across methods and Confidence Intervals (CIs) at significance $\alpha = 0.05$ to report average precision and recall ranges.
Using the same metrics and database we also did a comparison with a SLAM state of the art method [6], which has been proven to perform well in scope tracking in videobronchoscopy [8] In order to obtain the exploration path from ORBSLAM camera tracking, the path described by ORBSLAM camera 3D position was registered to the path described by the virtual camera which follows the center line of airways. A node representing a traversal of a bronchial level was considered correctly identified and considered TPN if the range of positions the virtual camera traverses a new bronchial level intersects the range of positions the ORBSLAM camera traverses the same bronchial level.
Also, and with the aim of increasing the performance of ORBSLAM, we computed virtual exploration paths transformed to enhance edges and corners of triangles. ORBSLAM needs to identify key points between frames to track the camera. Since our virtual images have been computed to have an appearance as close as possible to intra-operative videos, there might not be enough key points to a good performance of ORBSLAM even using the method described in [6]. Thus, we did not apply [17] and used flat illumination in virtual simulations. The right image in figure 3 shows an example of a frame transformed to enhance corners and edges in comparison to the realistic virtual frame used to validate our method shown in the right image of the figure.

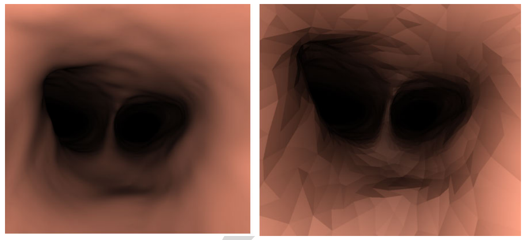


FIGURE 3. Preprocessing for ORBSLAM input. Virtual image, left, and the processed for ORBSLAM, right.

The computation of the exploration tree requires setting the values of the number of frames, NFr_{Mx}, for node activation and deactivation and the threshold, Th, determining node matching in (1). The number of frames NFr_{Mx} was set to $NFr_{Mx} = 5$, while the threshold was set to $Th = 0.2$ in order to manage sudden abrupt motions.

Several parameters were tuned in order to find the optimal performance of the ORBSLAM. The algorithm is based on the extraction of FAST keypoints and ORB descriptors. The keypoints are extracted following a pyramid scheme using different scales. The scaling factor between pyramid levels is 1.2 and the number of levels used is 8. For each level, the image is divided in a grid. At each cell, FAST keypoints are extracted imposing a minimum response. Firstly an initial threshold is imposed. The value of the initial threshold is $iniThFAST = 9$. If no corners are detected a lower value is imposed. The minimum value imposed is $minThFAST = 2$.

TABLE 1. Exploration tree assessment. Quality numbers for each patient and video.

	FP	FN	Num. Nodes	Precision	Recall
Patient 1 - Video 1	2	0	27	92.59%	100.00%
Patient 1 - Video 2	6	2	21	71.43%	90.48%
Patient 2 - Video 1	0	0	17	100.00%	100.00%
Patient 2 - Video 2	0	2	11	100.00%	91.82%
Patient 3 - Video 1	0	2	11	100.00%	81.82%
Patient 3 - Video 2	0	2	19	100.00%	89.47%
Patient 4 - Video 1	8	2	15	46.67%	86.67%
Patient 4 - Video 2	4	0	11	63.64%	100.00%
Total	20	10	132	82.61%	92.35%

A. ASSESSMENT OF THE EXPLORATION TREE

Table 1 reports, for each patient and path, false positives, FP, false negatives, FN, the total number of nodes of the GT exploration tree, precision and recall. The last row reports total numbers for the 8 paths. Our on-line exploration tree has an overall recall of 92% and a precision of 82%, but its performance substantially varies across patients and videos. Meanwhile Patient 2 and 3 reach 100% in precision with 90.7% of average recall, in the remaining cases the average precision drops to 68.7% with only a slightly increase in average recall (94.3%). In particular, Patient 4 has the lowest precision scores, with only 46.67% in the first video. Most FPs are due to shines in bubbles and sudden abrupt changes in scope motion, which mainly appear at most distal levels.

Meanwhile FNs are mostly attributed to substantial deviation in central navigation, which is a main requirement for the assumption of inclusion of lumens from different bronchial levels.

Table 2 reports the average, μ, standard deviation, σ, and CIs for the precision and recall. The confidence interval for the recall indicates that our algorithm detects most of the bronchi in all explorations ($\sigma = 0.0759$). In terms of precision, its CI indicates that the precision varies depending on the patient due to the reasons explained in Table 1.

TABLE 2. Precision and recall statistics for the on-line exploration tree extraction.

	μ	σ	CI
Precision	0.8261	0.2072	$[66.80\%, 98.43\%]$
Recall	0.9235	0.0759	$[84.71\%, 100.0\%]$

TABLE 3. Comparison of TPN and TPR obtained for [14] and the proposed on-line path.

	On-line path	Paths from [15]	p-val
TPN	$[78.06\%, 94.43\%]$	$[71.07\%, 80.95\%]$	0.0235
TPR	$[47.31\%, 85.28\%]$	$[54.25\%, 74.82\%]$	0.8086

B. ASSESSMENT OF THE EXPLORATION PATH

Table 3 reports CIs for average TPN and TPR percentages for both methods and p-values of the T-test for the difference. The detection rate of the correct bifurcations along the whole path (TPN) is significantly different with a CI for the difference equal to [1.44, 19.03] between the method proposed in this work and [14]. This improvement is due to the fact that the proposed algorithm not only detects luminal regions but it also encodes the hierarchy relationships between them. Hierarchy relationships allow to increase the robustness when luminal regions are not detected in some frames. Concerning the percentage of correct paths reached from the trachea (TPR), there are not significant differences with a CI for the difference equal to $[-16.39, 20.92]$. Although, our algorithm outperforms the approach from [14] in terms of TPN, there are not significant differences due to the TPR metric itself. Given a path with all the bifurcations but the first one correct, the TPR value is 0 since there is none correct value starting from the trachea. In both cases binary trees are used to encode the path followed during the intervention. Even so, lungs does not contain only binary bifurcations, they also contains ternary bifurcations which are not well defined in binary trees. Such ternary bifurcations appear in the proximal levels leading to low TPR values. Although there is a problem of representation of such ternary bifurcations, proximal levels are not important for a guiding system as doctors does not get lost.

Table 4 reports CIs for average TPN and TPR percentages for both methods and p-values of the T-test for the difference. The detection rate of TPN and TPR is significantly different with a CI for the difference between our method and ORBSLAM equal to $[30.1183, 57.4840]$ and $[0.4922, 47.20]$ respectively, between the method proposed in this work and [6]. Further, we note that that ORSLAM intervals for TPN and TPR are the same. Both issues are due to the fact that ORBSLAM underestimates depth during navigation and, thus, the moment it does not reach a given level, it is unable to catch up and deeper levels are also missed. Depth underestimation could be attributed to tracking key points which are not good descriptors of bronchial anatomy. In fact, ORBSLAM is only able to detect the first 1-4 bifurcations on average. From the point of view of intervention guidance, these levels are the least relevant ones since bronchoscopists have little difficulties identifying proximal airways. Another issue influencing ORBSLAM performance is that, like other motion estimation methods, it requires a minimum amount of motion in consecutive frames to properly estimate camera position. Since motion is more pronounced at main airways, initial bifurcations are better reconstructed and reached by ORBSLAM. At deeper levels, scope motion is more subtle and, thus, ORBSLAM systematically underestimates camera position.

TABLE 4. Comparison of TPN and TPR obtained for ORBSLAM and the proposed on-line path.

	On-line path	Paths from ORBSLAM [9]	p-val
TPN	$[78.06\%, 94.43\%]$	$[34.02\%, 50.86\%]$	$3.3728e-06$
TPR	$[47.31\%, 85.28\%]$	$[34.02\%, 50.86\%]$	0.0459

IV. CONCLUSION

With the final goal of an image-based navigation system for bronchoscopy guidance, we have presented a method for the extraction and codification of the bronchial anatomy and exploration path from videobronchoscopy. Our method bases on a graph encoding the hierarchy of bronchial levels traversed during the exploration. This simple representation of the geometry of airways allows the localization of the scope and the reconstruction of the path navigated during the intervention. This is an advantage over SLAM methods that require further registration of the reconstructed 3D point cloud to an anatomical reconstruction of the patient's airways in order to locate the scope.

Experiments conducted in interventional videos show that our method is highly accurate retrieving bronchial anatomy (recall $= 90\%$). However, there is a substantial variability in precision across patients due to bubbles, shines and abrupt changes in bronchial level due to patient cough. Bubbles introduce false lumen detections in the extraction of MSER regions, which could introduce false nodes and levels in the exploration tree. This could be solved by either using alternative descriptors of the lumen (like deep features) or pre-processing images to remove shines. Abrupt motion due to patient cough is prone to change the level of the scope at most distal airways which, like all trackers, is usually missed. This artifact is common to all trackers and it is a limitation of the technique. The impact of abrupt motion could be minimized if image guidance was complemented with inertial sensors deployed in the working channel in a hybrid system.

Comparison to existing guidance methods based on image analysis, shows that the proposed on-line tree outperforms the existing methods based on landmarks [14] and also SLAM approaches [6].

The results of this work are so promising, that encourage testing the system in clinical premises. This is a current work in cooperation with Hospital Germans Trias i Pujol from Barcelona, Spain.

ACKNOWLEDGMENT

The Titan X Pascal used for this research was donated by the NVIDIA Corporation. The D. Gil would like to dedicate this work to her mother, E. R. Enfedaque.

REFERENCES

[1] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA: A Cancer J. Clinicians*, vol. 61, no. 2, pp. 69–90, Mar. 2011.

[2] P. J. Reynisson, H. O. Leira, T. N. Hernes, E. F. Hofstad, M. Scali, H. Sorger, T. Amundsen, F. Lindseth, and T. Langø, "Navigated bronchoscopy: A technical review," *J. Bronchol. Intervent. Pulmonol.*, vol. 21, no. 3, pp. 242–264, 2014.

[3] E. F. Donnelly, "Technical parameters and interpretive issues in screening computed tomography scans for lung cancer," *J. Thoracic Imag.*, vol. 27, no. 4, pp. 224–229, Jul. 2012.

[4] T. R. Gildea, P. J. Mazzone, D. Karnak, M. Meziane, and A. C. Mehta, "Electromagnetic navigation diagnostic bronchoscopy: A prospective study," *J. Respiratory Crit. Care Med.*, vol. 174, no. 9, pp. 982–989, 2006.

[5] X. Luó, M. Feuerstein, D. Deguchi, T. Kitasaka, H. Takabatake, and K. Mori, "Development and comparison of new hybrid motion tracking for bronchoscopic navigation," *Med. Image Anal.*, vol. 16, no. 3, pp. 577–596, Apr. 2012.

[6] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015. AQ:5

[7] R. Eberhardt, N. Kahn, D. Gompelmann, M. Schumann, C. P. Heussel, and F. J. Herth, "LungPoint—A new approach to peripheral lesions," *J. Thoracic Oncol.*, vol. 5, no. 10, pp. 1559–1563, 2010.

[8] C. Wanga, M. Odab, and Y. Hayashi, "Visual SLAM for bronchoscope tracking and bronchus reconstruction in bronchoscopic navigation," *Proc. SPIE*, vol. 10951, Mar. 2019, Art. no. 109510A.

[9] M. Visentini-Scarzanella, T. Sugiura, T. Kaneko, and S. Koto, "Deep monocular 3D reconstruction for assisted navigation in bronchoscopy," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 7, pp. 1089–1099, Jul. 2017.

[10] C. Sánchez, M. Diez-Ferrer, J. Bernal, F. J. Sánchez, A. Rosell, and D. Gil, "Navigation path retrieval from videobronchoscopy using bronchial branches," in *Proc. Workshop Clin. Image-Based Procedures*. Springer, 2015, pp. 62–70. AQ:6

[11] A. Esteban-Lansaque, C. Sánchez, A. Borrás, M. Diez-Ferrer, A. Rosell, and D. Gil, "Stable anatomical structure tracking for video-bronchoscopy navigation," in *Proc. Workshop Clin. Image-Based Procedures*. Springer, 2016, pp. 18–26.

[12] J.-P. Charbonnier, E. M. V. Rikxoort, A. A. A. Setio, C. M. Schaefer-Prokop, B. V. Ginneken, and F. Ciompi, "Improving airway segmentation in computed tomography using leak detection with convolutional networks," *Med. Image Anal.*, vol. 36, pp. 52–60, Feb. 2017.

[13] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[14] C. Sánchez, A. Esteban-Lansaque, A. Borrás, M. Diez-Ferrer, A. Rosell, and D. Gil, "Towards a videobronchoscopy localization system from airway centre tracking," in *Proc. VISIGRAPP (VISAPP)*, 2017, pp. 352–359.

[15] M. Pawlik and N. Augsten, "Tree edit distance: Robust and memory-efficient," *Inf. Syst.*, vol. 56, pp. 157–173, Mar. 2016.

[16] E. Ramírez, C. Sánchez, A. Borràs, M. Diez-Ferrer, A. Rosell, and D. Gil, "BronchoX: Bronchoscopy exploration software for biopsy intervention planning," *Healthcare Technol. Lett.*, vol. 5, no. 5, pp. 177–182, Oct. 2018.

[17] D. Gil, A. Esteban-Lansaque, A. Borras, and C. Sanchez, "Enhancing virtual bronchoscopy with intra-operative data using a multi-objective GAN," *IJCAR*, vol. 14, no. 1, p. S7, 2019.

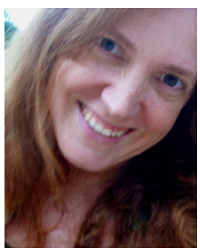
[18] A. Esteban-Lansaque, C. Sanchez, A. Borras, and D. Gil, "Augmentation of virtual endoscopic images with intra-operative data using content-nets," *BioRxiv*, 2019, Art. no. 681825.

[19] E. Ramírez, C. Sánchez, A. Borrás, M. Diez-Ferrer, A. Rosell, and D. Gil, "Image-based bronchial anatomy codification for biopsy guiding in video bronchoscopy," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer, 2018, pp. 214–222.

DEBORA GIL received the Ph.D. degree in mathematics from the Computer Science Department, UAB.

She is currently a Professor with the Computer Science Department, UAB, and in charge of the Interactive and Augmented Modeling (IAM, iam.cvc.uab.es) research group in the Computer Vision Center (CVC). She has a wide multidisciplinary experience and is an expert in mathematical and statistical modeling of heterogeneous data in clinical (diagnosis and intervention) decision support systems. She has coauthored more than 60 articles in journals (41 indexed in JCR) and over 100 international conferences. She has been IP of several national (seven competitive and five with private companies) and international (WP Leader in E-pilots, H2020-CS2-CFP08-2018-01 and a Coordinator of H2020 Topiomics-ATTRACT).

ANTONIO ESTEBAN-LANSAQUE received the Ph.D. degree in computer science from the Computer Science Department, UAB. He is currently an Temporal Associate Professor with the Computer Science Department, UAB. His main contributions have been the development of intelligent bronchoscopic support systems.



AGNÉS BORRÀS received the degree in computer science engineering and the Ph.D. degree in computer vision from the Universitat Autonoma de Barcelona (UAB), the master's degree in computer game design from UAB, and the Ph.D. degree in motion graphics from the Open University of Catalonia (UOC). Her research interests comprise the image analysis, virtual environments, augmented reality, and 3D representation. She has published 23 scientific papers, six in indexed journals (four of them Q1). She has worked as a Collaborate Professor at the Image and Multimedia Technology Center (CITM) and as a Consultant at UOC. She is a member of the R + D Department of the Computer Vision Center (CVC) and she has a strong experience in technological transfer. She has participated in 12 projects and she has an European Inventor's Patent.

ESMITT RAMÍREZ received the M.Sc. degree in computer science from the Central University of Venezuela (UCV). He is a member of the Interactive and Augmented Modeling (IAM, iam.cvc.uab.es) research group in the Computer Vision Center, UAB. He is also a member of the Computer Graphics Center (CCG), UCV. He has coauthored more than 20 papers in journals and international conferences. He currently holds a fellowship granted by the Ministry of Economy, Industry and Competitiveness, Spain. His main researches are focused on medical imaging and 3D rendering algorithms.

CARLES SÁNCHEZ RAMOS received the Ph.D. degree in computer science from the Computer Science Department, UAB.

He is currently a Professor with the Computer Science Department, UAB, and a member of the Interactive and Augmented Modeling (IAM, iam.cvc.uab.es) research group in the Computer Vision Center (CVC). He has been working in research projects as well technological transfer for more than eight years. He has long experience in medical imaging, anatomy modeling, and image processing applied in the development of intelligent endoscopic support systems. He has coauthored more than ten papers in journals and over 20 international conferences. He currently has a TecnioSpring Plus fellow and has been IP of two national projects and participated in 12.

• • •