Towards On-line Quantification of Tracheal Stenosis from Videobronchoscopy

Carles Sánchez · Jorge Bernal · F. Javier Sánchez · Marta Diez · Antoni Rosell · Debora Gil

Abstract *Purpose* Lack of objective measurement of tracheal obstruction degree has a negative impact on the chosen treatment prone to lead to unnecessary repeated explorations and other scanners. Accurate computation of tracheal stenosis in videobronchoscopy would constitute a breakthrough for this non-invasive technique and a reduction in operation cost for the public health service.
Methods Stenosis calculation is based on the comparison of the region delimited by the lumen in an obstructed frame and the region delimited by the first visible ring in a healthy frame. We propose a parametric strategy for the extraction of lumen and tracheal rings regions based on models of their geometry and appearance that guide a deformable model. To ensure a systematic applicability, we present a statical framework to choose optimal parametric values and a strategy to choose the frames that minimize the impact of scope optical distortion.
Results Our method has been tested in 40 cases covering different stenosed tracheas. Experiments report a non clinically relevant 9% of discrepancy in the calculated stenotic area and a computational time allowing on-line implementation in the operating room.
Conclusion Our methodology allows reliable measurements of airway narrowing in the operating room. To fully assess its clinical impact, a prospective clinical trial should be done.

Keywords Bronchoscopy · Stenosis assessment · Parameter setting · ANOVA

Carles Sánchez, Jorge Bernal, F. Javier Sánchez and Debora Gil
Computer Vision Center, Comp. Sci. Dep., Universitat Autonoma de Barcelona, 08193, Barcelona, Spain
Tel.: +34-93-5811828
E-mail: csanchez,jbernal,javier,debora@cvc.uab.cat

Marta Diez and Antoni Rosell
Pneumology Unit, Hosp. Univ. Bellvitge, IDIBELL, CIBERES. Barcelona
E-mail: marta.diez,arosell@bellvitgehospital.cat

1 Introduction

Tracheal stenosis (or CAO) refers to an abnormal narrowing of the trachea and main bronchi that affects airflow and respiratory function. Pulmonologists have to assess CAO severity [3] in order to: 1) diagnosis of benign and malignant diseases (goiter, esophageal and lung cancer); 2) choice of the appropriate prosthesis or stent and 3) validate effectiveness of therapeutic treatments (tracheal surgery, pharmacologic antiinflamatory or antiproliferativa treatments). Inadequate CAO characterization may lead to erroneous patient treatment and additional clinical tests with a negative impact on, both, the health care system efficiency and patient emotional cost. The most extended modalities to perform such measures are analysis of Computer Tomography (CT) scanners and bronchoscopy inspections.

Multiplanar CT provides, in general, accurate static 3D anatomical information [4] of airways. However stenosis assessment results are often affected by movements and secretions prone to be confused with stenosis wall thickness [34]. High definition CT might minimize such artifact, but they are expensive and increase patient radiation. Radiation exposure makes CT poorly suited for repeat testing and follow-up, especially in younger patients [36]. Time elapsed between CT and bronchoscopy can also affect the outcome. Finally, CT is difficult in young children (who require anaesthetic) and dyspnoeic or uncooperative patients [22].

Videobronchoscopy is an endoscopic modality that allows interactive navigation inside the airways without radiating the patient and, thus, has become an important tool for assessing and treating CAO [31,8,24]. A study [2] on AABIP members about their opinion on assessing CAO concluded that most of them use visual estimation (91%). Visual inspection of stenosis presents a significant difference across observers that influences diagnosis yield independently on their experience [27]. A majority (86%) of the clinicians consulted in [2] agreed that there is an urgent need to avoid subjective visual evaluation and standardize calculations during in-vivo explorations.

A robust method for quantifying CAO in the operating room would allow immediate and accurate diagnosis and treatment [35,28]. Stenosis assessment in clinical practice is achieved by comparing the area of a reference healthy airway to the area of the obstructed segment during bronchoscopy (as illustrated in Fig.1). The area of the healthy segment is defined by the most external and complete tracheal ring (left image in Fig.1), while the obstructed one is given by the luminal area (right image in Fig.1). Current methods for measuring the size of tracheal airway can be split into contact and non-contact procedures.

Contact procedures determine the diameter of tracheal lumen by inserting rigid bronchoscopes or endotracheal tubes of increasing stepwise size [29,17]. In order to determine airway size, the endotracheal tube is inflated to fit the tracheobronchial airway . The main concerns are the need of general anesthesia, potential damage to tracheobronchial soft tissues and inaccuracies in measurements [30,41]. Existing non-contact procedures include bronchoscopic device improvement and videobronchoscopic computer analysis. In the first case, a suitable modification of the bronchoscope optics [20,15] can provide accurate measurements with micrometer precision by light triangulation. However, these technologies are still experimental and far from being commercially available. As far as computational methods for stenosis assessment are concerned, they should ensure reliable extraction of
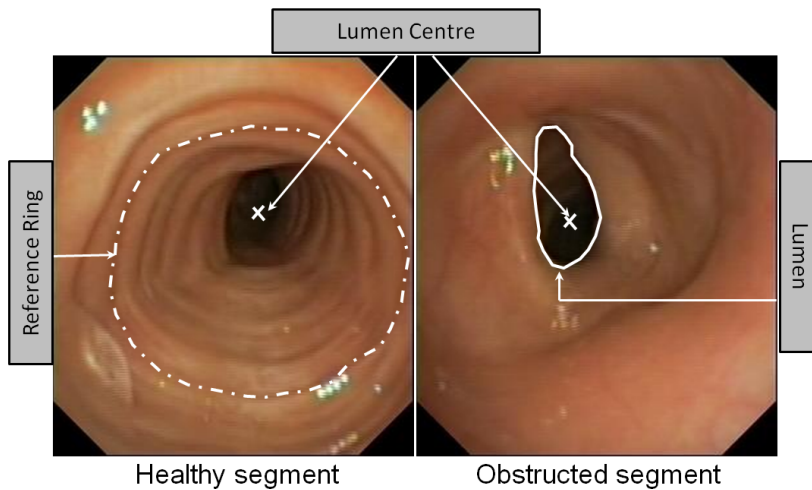
Fig. 1 Computation of stenosis degree by comparing a healthy (left image) and obstructed segments (right image)

lumen and tracheal rings during the endoscopic procedure, compensation of lens distortion and selection of frames at equal distance to target structures.

Processing of bronchoscopy videos, and in particular segmentation of tracheal structures (luminal area and rings) is a challenging task [26]. This is mainly due to the large variety of acquisition conditions (camera resolutions and digital compressions) and devices (flexible and rigid optics). Besides, processing videos acquired at the operating room adds the unpredicted presence of surgical devices (such as probe ends), as well as, illumination and camera position artifacts [26,18].

The majority of the relevant work in lumen localization and detection is related to gastrointestinal image analysis. Under the assumptions that the largest dark blob of the images usually correspond to lumen [39], several works segment the lumen using region growing [1,33,12]. Such approaches are accurate as far as the initial seed is placed inside the luminal area and often require manual adjustments [1,33]. A common limitation is that none of the methods can address segmentation of multiple lumens, which is common in bronchoscopy videos. Most of the existing works in bronchoscopy [23,24,8,10,32] are semi-automatic procedures which are applied off-line. Off-line information is valuable to plan, but the definitive diagnosis, final clinical approach and treatment performed are done during the endoscopic procedure. Although a recent work [43] detects multiple lumen areas, it might fail in the absence of any luminal area and has a computational cost not suitable for its use in intervention time. An interesting approach to lumen modelling was introduced in [37], where the authors defined a feature space describing the geometry and appearance of the lumen. Such space was used to accurately detect the lumen center in colonoscopy and bronchoscopy images with multiple (or none) luminal areas. The works on tracheal rings segmentation in videobronchoscopy are even more sparse. As far as we know, our previous work [38] is the only one addressing tracheal ring processing. The proposed strategy used geometric and appearance features to detect tracheal rings and had a promising sensitivity within inter-observer variability.

The goal of the paper is to provide a computational system for accurate standardized measurement of CAO in the operating room. The main contributions are:

1. Our anatomical descriptors are used to define external energies in a deformable model scheme for lumen and ring segmentation. Formulation and settings are designed to ensure convergence regardless of snake internal parameters.
2. A new statistical framework based on Analysis of Variance (ANOVA) [25] for selecting the parameters that ensure a method's performance within inter-observer ranges.
3. A simple strategy for stenosis assessment in videbronchoscopy able to account for lens distortion and distance to target artifacts in a time able to run during intervention.
4. A complete annotated database and validation protocol to assess segmentations accuracy and the correlation to the stenosis assessed by experts.

The paper contents are the following. Section 2 explains our previous models of endoluminal structures. Section 3 details our strategy for the segmentation of anatomical structures, the ANOVA design for parameter setting and the selection of reference frames. Section 4 describes the databases and validation protocol and section 5 reports the results. Finally, Section 6 concludes the paper.

2 Geometric and Appearance models of the endoluminal scene

The anatomical structures to identify (see Fig. 1) in order to measure stenosis degree are the obstructed lumen and a complete tracheal ring in a healthy segment. In order to minimize the impact of videobronchoscopy artifacts, both structures are modelled using a combination of intensity and geometry.

The lumen is characterized as the dark region centre of the hub of image gradients [37]. Dark areas, DRI, are extracted by convolution of the image, $I = I(x, y)$, with a gaussian kernel, g_σ, of size σ:

$$DRI(x, y) := g_\sigma * I = \frac{1}{(2\pi)\sigma^2} e^{-\left(\frac{x^2}{2\sigma^2} + \frac{y^2}{2\sigma^2}\right)} * I(x, y) \qquad (1)$$

for $*$ the convolution operator. Image gradient centralness, DGA, is formulated using the accumulation of the image gradient lines:

$$DGA(x, y) := \sum_{\exists \lambda | (x,y) = (x_0, y_0) + \lambda \nabla I(x_0, y_0)} \nabla I(x_0, y_0) \qquad (2)$$

for $(x_0, y_0) + \lambda \nabla(x_0, y_0)$ is the of line through (x_0, y_0) oriented across the image gradient $\nabla I(x, y)$ parametrized by λ. The feature space (DRI,DGA) discriminates lumen pixels from other endoluminal structures [37] by means of a linear classifier:

$$(DRI - DRI_0)V_{DRI} + (DGA - DGA_0)V_{DGA} = 0 \qquad (3)$$

for (V_{DRI}, V_{DGA}) the slope and (DRI_0, DGA_0) the intercept.

Tracheal rings are described as image valleys with an increasing radial thickness and a concentric disposition around the lumen center [38]. In order to account for

illumination artifacts, valleys are computed using Normalized Steerable Gaussian Filter, NSGF, given by:

$$NSGF_{\sigma,\theta} := \frac{\|\partial_{\tilde{y}}^2 g_{\sigma,\theta} * I\|}{\|\partial_{\tilde{y}}^2 g_{\sigma,\theta}\| \|I\|} \tag{4}$$

for $\| \cdot \|$ the L^2 integral norm and $\partial_{\tilde{y}}^2 g_{\sigma,\theta} = (\tilde{y}^2/\sigma^4 - 1/\sigma^4) g_{\sigma,\theta}$ a second derivative of a gaussian kernel of scale σ oriented along θ [11]. In the polar domain, $NSGF$ is the maximum response for a discrete sampling of the scale in a given range, $\sigma^j = \sigma_m + j(\sigma_M - \sigma_m)/N_\sigma$, and an orientation, $\theta^i =: i\epsilon/N_\theta$, around zero:

$$NSGF := \max_{i,j} \left(NSGF_{\sigma^j,\theta^i} \right) \tag{5}$$

3 Stenosis assessment from videobronchoscopy

Guidelines proposed in [28] propose the automatic calculation of stenosis index by comparing the airway area of a healthy frame -delimited by the most external tracheal ring- and the airway area of a frame with an obstruction -delimited by the luminal area-. In this chapter we use our previously defined geometry and appearance models of the tracheal structures defined in Chapter 2 to calculate these areas. Considering this, Stenosis Index, SI, is calculated as:

$$SI = \left(\frac{AA_{Ref} - AA_{Lumen}}{AA_{Ref}} \right) \cdot 100 \tag{6}$$

for AA_{Ref}, AA_{Lumen}, the areas of the reference and abnormal segments, respectively. Being SI an area ratio, it is independent of device and image resolutions. The areas AA_{Ref} and AA_{Lumen} are given curves segmenting the reference closed ring and the obstructed lumen. Such curves are computed in the polar domain centred at the lumen as radial functions that depend on the angle $\theta \in [0, 360]$. Let $R = R(\theta)$, $L = L(\theta)$, be such functions, then AA_{Ref}, AA_{Lumen} are given by:

$$AA_{Ref} = \sum_{i=1}^{N_\theta} \frac{R(\theta_i)^2}{2} h_\theta; \quad AA_{Lumen} = \sum_{i=1}^{N_\theta} \frac{L(\theta_i)^2}{2} h_\theta \tag{7}$$

for $\theta_i = i h_\theta \in [0, 360]$ a uniform sampling of step $h_\theta = \frac{i360}{N_\theta}$. All images will be in polar coordinates with origin at the lumen center computed using [37].

3.1 Reference Ring Segmentation

The reference ring is segmented in the three steps sketched in Fig. 2. First, a set of candidate points (yellow and red lines in image 2)) is selected from $NSGF$ response. Second, the largest unconnected segments (red lines) are completed using geometric information [14] (grey line in image 3)). Finally, a snake is used to model a closed curve (blue lines in images 3) and 4) of Fig. 2) for AA_{Ref} computation.

The $NSGF$ operator (eq. 5) gives 3 values for each pixel: the maximum value and the scale, σ_{NSGF}, and orientation, θ_{NSGF}, of the filter that achieved such
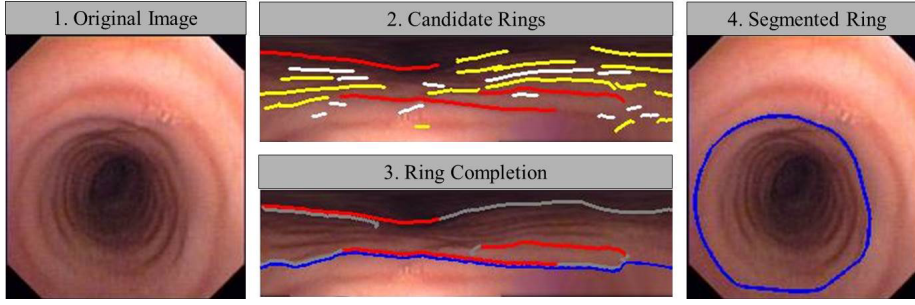
Fig. 2 Ring segmentation: 1) Original image; 2) Detected rings (white), candidate ones (yellow) and reference ring (red) in polar coordinates; 3) Reference ring ACC closing (grey) and snake completion (blue), and 4) Reference ring in cartesian domain.

maximum. Local maxima are computed using Non-Maxima Supression, NMS, oriented along θ_{NSGF} as:

$$NMS_{NSGF}(x,y) = \begin{cases} \sigma_{NSGF} & \text{if } NSGF > \max(NSGF_{V+}, \ NSGF_{V-}) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

for $V = (\cos(\theta_{NSGF}), \sin(\theta_{NSGF}))$, and $NSGF_{V+} = NSGF(x + V_x, y + V_y)$, $NSGF_{V-} = NSGF(x - V_x, y - V_y)$ computed using linear interpolation.

Points on the candidate reference ring are selected by filtering NMS_{NSGF} in a range of scales, $[\sigma^j, \sigma^{j+k}]$, set using the ANOVA described in Section 3.3. The mask of candidate points, noted by χ_{NSGF}, is given by:

$$\chi_{NSGF}(x,y) = \begin{cases} 1 & \text{if } NMS_{NSGF}(x,y) \in [\sigma^j, \sigma^{j+k}] \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Candidate points (see image 2) in Fig.2) are a fragmented set that should be completed to compute AA_{Ref}. Completion is achieved two-stages to consider as much anatomical information as possible.

First, we apply an Anisotropic Contour Closing, ACC [14], to complete χ_{NSGF} connected components using gradient information. The ACC operator is given by a restricted heat operator with unitary Dirichlet conditions on $u_0 = \chi_{NSGF}(x,y)$:

$$u_t = \text{div}(Q\Lambda Q^t \nabla u), \quad u_{|\{\chi_{NSGF}(x,y)>0\}} = 1 \tag{10}$$

for div the divergence operator and the diffusion tensor $Q\Lambda Q^t$ defined as:

$$Q := \left(\xi^\perp, \xi\right) \text{ and } \Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{11}$$

where $\xi, \xi^\perp$ are the eigenvectors (ordered in decreasing eigenvalues) of the structure tensor, $ST(\rho_1, \rho_2)$, computed over the image in polar coordinates:

$$ST(\rho_1, \rho_2) := \begin{pmatrix} g_{\rho_1} * (\partial_x g_{\rho_2} * I)^2 & g_{\rho_1} * (\partial_x g_{\rho_2} * \partial_y g_{\rho_2} * I) \\ g_{\rho_1} * (\partial_x g_{\rho_2} * I \partial_y g_{\rho_2} * I) & g_{\rho_1} * (\partial_y g_{\rho_2} * I)^2 \end{pmatrix}$$

for g_{ρ_1}, g_{ρ_2} gaussian kernels of variance ρ_1 and ρ_2, respectively. By the properties of heat equations in manifolds, the solution to (10) is a binary map of a closed model of the uncomplete initial contour that admits an efficient implementation

[19] able to run in intervention time. Image 3) in fig.2 shows the ACC completion (gray lines) of the yellow fragmented curves shown in image 2).

Contours completed with ACC are connected curves that match the C-shape of tracheal rings and, thus, are open at the lower part limiting with the esophagus. In order to get a closed curve, we use a deformable model [21] guided by the distance map to the reference ring. The reference ring is the most complete one covering the 180 degrees of the C-shape arc. In case of having more than one, the most external ring having the maximum radial coordinate is chosen. If R_{Ext} notes the snake external energy, $\partial_\theta R$, $\partial_{\theta\theta} R$ the first and second derivatives of the radial function $R(\theta)$ modelling the ring, then $R(\theta)$ is the steady state of:

$$R_t = \partial_r(R_{Ext}) + \alpha\partial_\theta(R_{t-1}) + \beta\partial_{\theta\theta}(R_{t-1}) \text{ with } R_0 \equiv 1 \tag{12}$$

for α, β the weights for the stiffness and elasticity snake terms. The snake modelling ACC closing is shown in blue lines on images 3) and 4) in fig.2.

3.2 Obstructed Lumen Segmentation

The curve, $L = L(\theta)$, that segments the lumen is also computed using a snake:

$$L_t = \partial_r(L_{Ext}) + \alpha\partial_\theta(L_{t-1}) + \beta\partial_{\theta\theta}(L_{t-1}) \tag{13}$$

In this case, the external energy L_{Ext} is given in terms of the linear classifier (3) as follows. The linear classifier defines a 1-parametric family of likelihood maps depending on its intercept, l:

$$LK_l := |DRI \cdot V_{DRI} + DGA \cdot V_{DGA} - (DRI_0 \cdot V_{DRI} + DGA_0 \cdot V_{DGA})| =$$
$$= |DRI \cdot V_{DRI} + DGA \cdot V_{DGA} - l| \tag{14}$$

for $|\cdot|$ the absolute value. The left part of Fig. 3 shows (DRI, DGA) feature space with two classifiers (dashed lines) at different intercept values, l_k and l_{k+n}.

The values LK_l can be interpreted as the distance (in feature space) to the set of pixels that define the border ($LK_l = 0$) between non-lumen and lumen points. We note that only for $l \in [min(V_{DRI}, V_{DGA}), max(V_{DRI}, V_{DGA})] = [l_m, l_M]$ likelihood maps that a well defined curve of minimal points. Outside this interval, all pixels are classified either lumen or non-lumen and, thus, LK_l has not any minimum. The set of LK_l mimimal points progressively approximate the lumen border as the interval $[l_m, l_M]$ is swept. The optimal intercept, $\widetilde{l}_j$, is computed using the ANOVA described in Section 3.3.

To minimize the impact of $LK_{\widetilde{l}}$ local minima and dependency on initial curve, we consider a family of snakes, $(L^j)_{j=1}^{j=N_{\widetilde{l}}}$, that solve (13) for a sampling of size $N_{\widetilde{l}}$ of the interval $[l_m, \widetilde{l}]$ given by $l^j = l_m + j/N_{\widetilde{l}}(\widetilde{l} - l_m)$. Each snake L^j is formulated as the steady state of:

$$L_t^j = \partial_r(LK_{l^j}) + \alpha\partial_\theta L_{t-1}^j + \beta\partial_{\theta\theta} L_{t-1}^j \text{ with } L_0^j = L^{j-1} \tag{15}$$

for L^{j-1} the solution for $L_{Ext} = LK_{l^{j-1}}$ and $L^0 \equiv 1$. This progressive scheme avoids premature convergence and minimizes the impact of initialization and internal parameters α, β. The snake for $j = N_{\widetilde{l}}$ is our final lumen segmentation. Middle images in fig.3 show two different LK_{l^k} energies in polar coordinates for consecutive values l^{k_1} and l^{k_2}. The snakes solving (15) for l^{k_1} and l^{k_2} are plotted on the polar images shown under each LK^{l_k} map and on the right cartesian image.
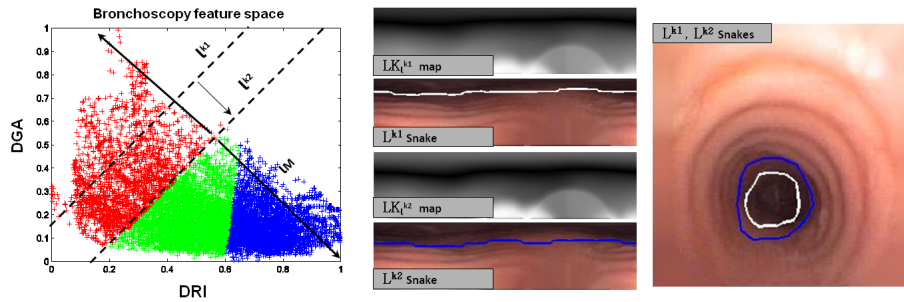
Fig. 3 Lumen segmentation. From left to right: feature space, snakes (l^{k1}, l^{k2}) segmenting 2 consecutive energy maps ($LK_{l^{k1}}$, $LK_{l^{k2}}$) in polars and over the original frame.

3.3 Parameter Setting

In medical diagnosis a new method is accepted as a diagnosis tool [42] provided that it can substitute (compares to) the gold standard or ground truth, GT. In the context of anatomical structures segmentation, GT is provided by experts manual annotations, while comparisons are given by a metric quantifying differences between manual and automatic regions [16]. Given that different observers produce different GTs, a method is optimal if its metric ranges are comparable to a gold standard GT considering differences across manual annotations. In image processing, this is achieved by considering to different experts GTs to define a gold standard and a comparable method is said to be within inter-observer ranges [1] . We propose to use Analysis of variance (ANOVA) to compare multiple methods to inter-observer ranges. ANOVA [6] explores the variability of a quantitative variable across several categorical explanatory variables (called factors) and provides a statistical way to decide wether differences in factors means are significant enough with a given confidence level α. The design to set optimal parameters is as follows.

A 2-way ANOVA for the quality metric and factors given by the several methods (row factor) and observers (column factor) is used to check if methods performance varies across experts. The desired result of this test would be a significance in the methods factor, possibly a significance across observers and, most important, no significant interaction. A non-significant difference across observers would indicate that there is no evidence of an inter-observer variability for the chosen set of experts and, thus, that any of them could be used as GT. In case of significant interaction, a 1-way ANOVA with the combined method-observer factor should be used to detect the sources of bias. Otherwise, we can compare the methods output to observer ranges by using the following 1-way ANOVA.

In case of no interaction, a 1-way ANOVA with groups given by the different methods and the observers evaluated against each other provides the final comparison. For the methods, the ANOVA variable is the metric averaged over all observers, while for the observers control group is the average of the metrics obtained by evaluating each observer against the remaining ones. Given that in this case, the observer group acts like a control group, the correction used for the multicomparison test is Dunnet [9]. The optimal methods configuration are the ones such that Dunnet test does not find significant differences.

---

[1] http://tinyurl.com/kmqmfys

3.4 Frame Selection and Lens Distortion

One of the main challenges that a given stenosis assessment method has to face is to assure that measurements taken from obstructed frames are comparable. Measurements are comparable if they use the same measurement unit, which in our case is the projected area in each pixel of the image. If we consider perspective projections sharing the same orientation with the camera, differences in projected areas depend on the distance between the camera and the target structure. Thus, measures in different frames are comparable provided that they are acquired at the same distance from target structures. We propose to use the following protocol to select frames. First the reference ring frame is selected by navigating the scope through a healthy segment until the reference ring is out of the camera field of view. From this point, the scope is pulled back a distance measured using the scope marks (at 1 cm each) until such ring is completely visible. The use of marks [7] is the most accepted way of measuring distances (tumors to carina, length of a stenosed segment or fistulae size) in clinical procedures. Regarding the obstructed frame, the scope is placed at the same distance as before by pulling it back using the same protocol.

We have also to consider that endoscope optics produce barrel distortion as they are wide angle. This distortion depends on the specific endoscope we are working with: for instance, rigid endoscopes produce almost no distortion whereas flexible ones present noticeable distortion. Barrel distortion makes area projected of pixels near projection center being smaller that area projected in pixels far from projection center. This implies that reference ring area may be underestimated with respect to the obstructed region one. Although our measurements are comparable to the ones provided by clinicians some optic distortion correction methods [18] could be applied to obtain precise anatomically results.

4 Databases and validation protocol

We have tested our methodology for the segmentation of anatomical structures and computation of the stenosis index using the following experimental settings:

4.1 Accuracy of Tracheal Structures Segmentation

The goal is to compare our method to annotations made by physicians. First, we will apply the ANOVA design of Section 3.3 to a training set in order to set parameters that reach a performance within inter-observer ranges. Second, the optimal parameter setting will be assessed on a test set of frames.

We have used two different databases, one for tracheal ring and another one for lumen region. For each structure 80 images were selected from 20 explorations made at Bellvitge Hospital. The lumen database covers different kinds of obstructions and SI degrees [27] classified according to airflow resistance decay [3] as 15 Mild ($SI < 50\%$), 15 Moderate ($SI \in (50\%, 70\%)$), 30 Severe ($SI > 70\%$) and 20 healthy frames. Tracheal rings include representative appearance artifacts. For each data set, a random sampling of 25 frames was used as training set to adjust the optimal scale, $[\sigma_j, \sigma_{j+k}]$ and intercept, $\widetilde{l}$, using the ANOVA of Section 3.3. The

optimal parameter setting is tested in the remaining 55 images, using the same Dunnet test used for training and then a paired t-test to compare the selected optimal parameters to observer's range. All tests have been done at significant level $\alpha = 0.05$.

The random variable for statistical tests is a measure of segmentation accuracy against off-line manual annotations made by two different experts. Manual curves were compared to our segmentations by means of the average distances computed across the curve angular parameter. Given two curves in polars, $\gamma_A(\theta) = (\theta, r_A(\theta))$, $\gamma_B(\theta) = (\theta, r_B(\theta))$, the average distance between them is:

$$AvD_B^A = ((\frac{1}{N_\theta} \sum_i abs(r_A(\theta_i) - r_B(\theta_i)))/s_x) \cdot 100 \qquad (16)$$

for $\theta_i = 360i/N_\theta$ a uniform sampling of the angle of step $h_\theta = 360/N_\theta$ and s_x the height dimension of the original image to account for differences in device resolutions. For a given curve annotated by observer O_k, we will note by $AvD_\rho^{O_k}$ the distance between the observer curve and the output of our method using a parameter setting ρ. Such ρ represents the scale interval $[\sigma_j, \sigma_{j+k}]$ for ring segmentation and the intercept $\widetilde{l}_k$ for lumen segmentation.

The variable and group factors for the ANOVA parameter setting is defined as follows. For the 2-ANOVA, groups are the observers, O_k k=1,2, for the first factor and the different parameters settings, ρ, for the second one. For each group, (O_k, ρ), the individuals are $AvD_\rho^{O_k}$ computed for the 25 training frames. For the 1-ANOVA test, groups are defined by the different parameters and the average observer. In this case the ANOVA variable is given by averaging distances: $AvD_O = (AvD_{O_2}^{O_1} + AvD_{O_1}^{O_2})/2$ for the observer group and $AVD_\rho = (AvD_\rho^{O_1} + AvD_\rho^{O_2})/2$ fot the method ones. The samplings defining ANOVA groups for the method are given by uniform samplings of size N_σ for ring scales and N_l for lumen intercept:

$$(([\sigma_j, \sigma_{j+k}])_{j=1}^{N_\sigma - k})_{k=1}^{N_\sigma}; \; \sigma_j = \sigma_m + j(\sigma_M - \sigma_m)/N_\sigma \qquad (17)$$

$$\widetilde{l}_k = l_m k/N_l \cdot (l_M - l_m) \qquad (18)$$

4.2 Accuracy of the Stenosis Index

The accuracy of SI has been quantified by comparing (6) to the percentage obtained by the manual annotations of obstructed lumen and reference ring made by the same two experts. Such off-line morphological analysis is considered the gold standard [27] to be automatically computed on-line. Each expert stenosis index will be noted by SI_{O_1} and SI_{O_2} respectively.

The data set used for this experiment are 20 new cases of patients with several degrees of stenosis [2] different from the ones used in the first experiment and visually classified as 3 mild, 10 moderate and 7 severe. Images showing a normal and an abnormal airway were extracted following the protocol described in Section 3.4. The comparison between our SI and SI_{O_k} was done using a 1-way ANOVA with groups given by SI, SI_{O_1} and SI_{O_2}.

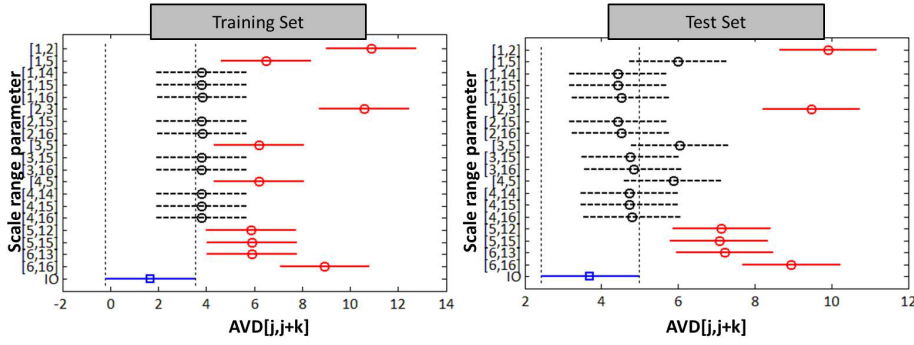[2] Bellvitge Hospital Barcelona. Sampling size for clinical trial.

Fig. 4 Ring segmentation multicomparison plots for train and test sets. X-axis is the average distance, AvD_ρ, for $\rho = [\sigma_j, \sigma_{j+k}]$ represented in Y-axis by $[j, j+k]$ and IO is AVD_O.

5 Experiments

5.1 Tracheal Structures Segmentation Accuracy

For tracheal rings, the 2-ANOVA test detects no interaction between observers and scale ranges ($p - val \simeq 1$), no significant differences between observers ($p - val = 0.9917$) but a significant difference across scale values ($p - val < 10^{-16}$). These results indicate that the performance of our method does not depend on the observer so that best performers can be selected by comparing to an average observer using the 1-ANOVA test described in Section 4.

The 1-ANOVA detects a significant ($p - value < 10^{-16}$) difference in accuracy across scales that is further explored using a Dunnet multicomparison test. The left plot in fig. 4 shows the multicomparison for the training and the right one for the test. Both plots show intervals for mean differences. Each level mean is represented as an horizontal line centered at the mean group and vertically distributed according to the group number. Those scale ranges that are not significantly different from IO (in blue and square) are plotted in dashed black, while scales with a significant different performance are depicted in red and circle. As expected, trends in significant differences are the same for both plots that select
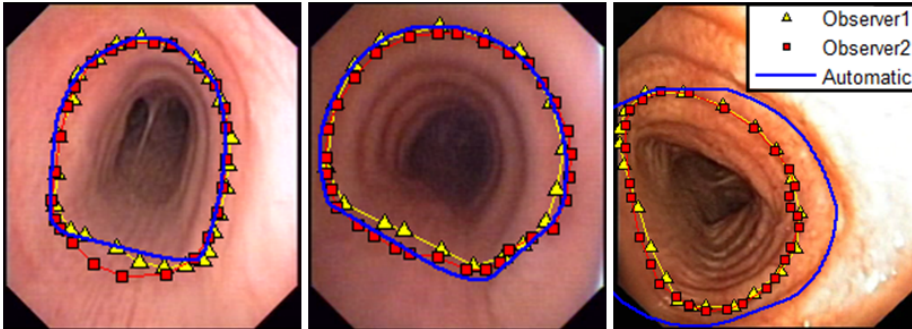

Fig. 5 Ring segmented with the optimal scale range $[\sigma_2, \sigma_{2+15}]$ with blue lines for methods' output and yellow triangles, red squares lines for the 2 manual annotations.
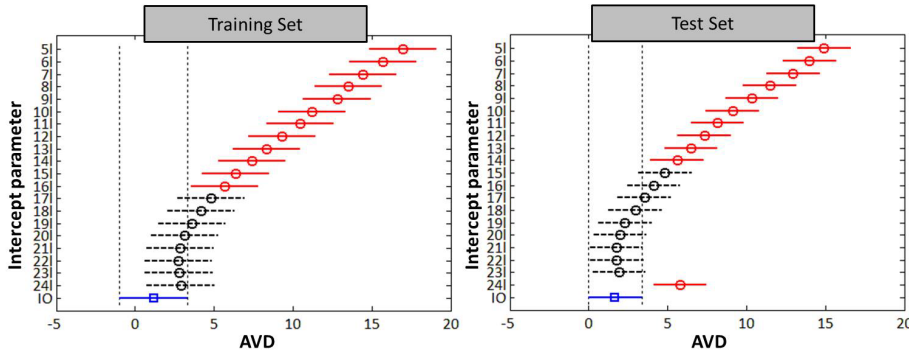
Fig. 6 Lumen segmentation multicomparison plot for train and test sets. X-axis is the average distance, AvD_ρ, for $\rho = l_k$ represented in Y-axis by kl and IO is AVD_O.

the scale $[\sigma_j, \sigma_{j+k}] = [2, 15]$ as the most coincident with AvD_O. The t-test comparing AvD_O to $AvD_{[2,15]}$ is not significant ($p-value = 0.0883$) though it has a small negative bias (see Section 6 for discussion) in confidence intervals for the difference $AvD_O - AvD_{[2,15]}$ ($CI = [-1.58, 0.11]$). Figure 5 shows some examples of rings segmented with $[\sigma_j, \sigma_{j+k}] = [2, 15]$.

Concerning lumen segmentation, 2-ANOVA detects no interaction between observers and intercept ($p-val = 1$), no significant differences between observers ($p-val = 0.9653$) and significant differences ($p-val < 10^{-16}$) across intercept values. Like tracheal ring segmentation, this implies that the performance of our method does not depend on the observer and, thus, a 1-ANOVA can be applied. The 1-ANOVA test is significant ($p-value < 10^{-16}$) and best performers are selected using the Dunnet multicomparison test shown in fig. 6. Plots compare intervals for mean performances as in Fig.4. As before, train and test plots have the same trend and select $lk = l21$ as the best intercept. The t-test comparing AvD_O with AvD_{l21} detects no significant differences ($p-value = 0.18$) with a confidence interval for the difference $AvD_O - AvD_{l21}$ equal to $CI = [-0.28, 0.05]$. Fig.7 shows segmentation results using the same color coding as in Fig. 5.
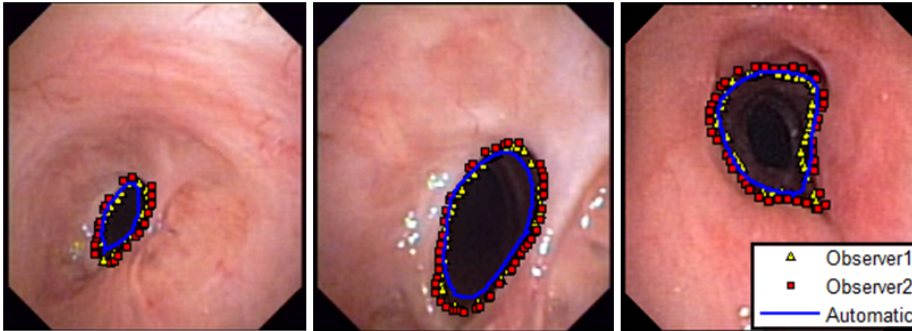

Fig. 7 Lumen segmented with the optimal intercept $\widetilde{l}_{21}$ with blue lines for methods' output and yellow triangles, red squares lines for the 2 manual annotations.
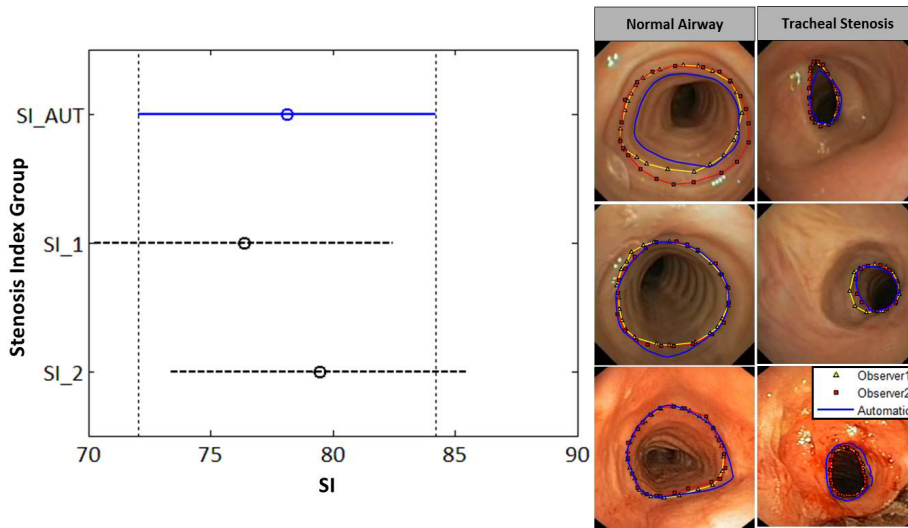
Fig. 8 SI assessment: multicomparison plots (left) and qualitative results (right) with blue lines for methods' output and yellow triangles, red squares lines for manual annotations.

5.2 Accuracy of the Stenosis Index

According to 1-ANOVA there are not significant differences across observers and our automatic SI index ($p - value > 0.83$). This is also confirmed in the multi-comparison plot shown in fig. 8 that indicates a clear overlapping across the 3 indexes computed by experts and our method. Also a t-test for paired data comparing SI_{AUT} with the average expert score SI_O could not find any significant differences ($p - value = 0.96$). Besides, the confidence interval for the difference $SI_O - SI_{AUT}$ is $CI = [-10.3, 9.8]$, which is clinically acceptable [27]. Right images in fig. 8 show representative examples of our method output compared to with manual annotations with the same color coding as in fig.7.

Finally, the average processing time for structure segmentation using a sub-optimal Matlab code on a 12-core Intel i7 processor with 16GB of RAM is around 5 seconds per frame. Taking into account that stenosed and healthy images are not consecutive frames and require a previous navigation, this computational cost allows an implementation in intervention time.

6 Discussion and Conclusions

This paper introduces, up to our knowledge, the first automatic tool for systematic quantification of CAO from analysis of videobronchoscopy. CAO assessment is based on the comparison of the areas delimited by curves segmenting the most external ring and the obstructed lumen, called stenosis index (SI). We have assessed the performance of each tracheal structure characterization in databases which contain a balanced presence of all possible appearances that we may find in procedures, including extremely abnormal examples. Experimental results show that our method is comparable to clinicians in, both, tracheal structure characterization and CAO assessment. Aside our method has a low computational cost that
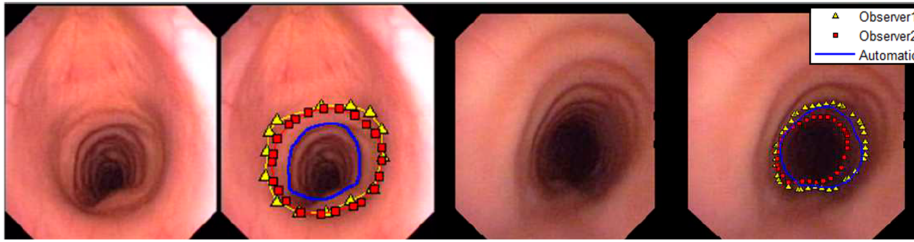
Fig. 9 Discrepancies between experts' annotation and our detections.

could have a real-time implementation to be used in the operating room. There are several positive aspects that make our methodology well-suited for clinical deployment.

A main key feature of our method is that it does not need a previous CT scan. Being exclusively based on bronchoscopy processing, it is especially appropriate for stenosis analysis in children, as it allows for accurate calculation without patient radiation [17]. Being fully automated, it allows the dynamic assessment of stenosis, a condition in which airflow is decreased or even stopped as intrathoracic pressure increases [5]. In these cases patients do not present a fix obstruction degree but a variable one which can change as the patient breathes or coughs. Current procedures based on CT scan information can only provide a static image at maximum inspiration and at maximum expiration. Our methodology allows the computation of obstruction dynamic ranges, which can be key to certify tracheomalacia.

Still, there are some issues that could be improved. First, the computation of SI relies on the identification of a complete ring in images of a healthy segment or in the accurate computation of obstructed airway area. This can be difficult to obtain in severely ill patients (Fig. 9) with pathologies such as diffuse tracheal inflammation (post radiotherapy), infiltration (amyloidosis) or cartilage ring dismorfia (Tracheobronchopathia osteochrondoplastica). In such cases, our method could underestimate the degree of obstruction, yielding an optimistic diagnose. On one hand, these cases are not common in tracheal stenosis diagnosis and manual identification is also difficult. On the other hand, they can be solved by computing the reference area using the lumen segmented from a non-obstructed segment.

Second, stenosis visual assessment has two main error sources: mental area computation and differences in distance to target structures. In this paper, we have focussed on the development of a method that suppresses area errors and compares to the gold standard off-line manual annotations, which cannot account for distance errors. In order to obtain more accurate and measurements and relax the acquisition protocol, the whole video should be analyzed [27] to perform 3D measurements and reconstruction from videobronchoscopy frames. This could be achieved by using structure from motion techniques (SLAM [40]), which allow the computation of, both, camera parameters and 3D point position from a set of correspondent pixels across consecutive frames. These techniques have proven to be robust provided that point correspondence is accurate enough and have few outliers. Usual correspondences based on local image features might fail in case of structures having a uniform appearance and geometry, like tracheal rings and lumen. An alternative is using a normalized sampling on previously segmented structures ensuring point correspondence [13]. In this context, our strategy constitutes a solid first step towards 3D stenosis measurements from videobronchoscopy.

Finally, a prospective multicenter control randomized clinical trial should be carried out in order to assess the clinical impact of the proposed methodology. This clinical trial could also be used to measure the effect of some of the potential sources of error such as differences in distances to target for healthy and stenosed frame or distortion. Moreover, if the clinical trial of the method is performed by clinicians with different level of expertise we could assess whether our methodology can be used to guide novel clinicians as a training tool.

Conflict of interest

The authors declare that they have no conflict of interest. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

1. Asari, K.V.: A fast and accurate segmentation technique for the extraction of gastrointestinallumenfrom endoscopic images. Medical engineering & physics **22(2)**, Elsevier (2000)
2. Begnaud, A., Connett, J.E., Harwood, E.M., Jantz, M.A., Mehta, H.J.: Measuring central airway obstruction: What do bronchoscopists do? Annals of the American Thoracic Society (ja) (2014)
3. Brouns, M., Jayaraju, S.T., Lacor, C., De Mey, J., Noppen, M., Vincken, W., Verbanck, S.: Tracheal stenosis: a flow dynamics study. JAP **102**(3), 1178–1184 (2007)
4. Brown, Robert H and Herold, Christian J and Hirshman, Carol A: In vivo measurements of airway reactivity using high resolution computertomography. Am. Re. Respir Dis **144**, 208–12 (1991)
5. Carden, Kelly A and Boiselle, Philip M and Waltz, David A and Ernst, Armin: Tracheomalacia and tracheobronchomalacia in children and adults an in-depth review. CHEST Journal **127**(3), 984–1005 (2005)
6. Cohen, J.: Statistical powe analysis for the behavioral sciences. Lawrence Erlbaum Associates (1988)
7. Colt, H., Murgu, S.: Bronchoscopy and central airway disorders. Elsevier (2012)
8. Dörffel, WV. and Fietze, I: A new bronchoscopic method to measure airway size. ERJ **14(4)**, 783–788 (1999)
9. Dunnett, C.W.: A multiple comparison procedure for comparing several treatments with acontrol. Journal of American Statistical Association **50**, 1096–1121 (1980)
10. Forkert, Lutz and Watanabe, Hiroshi and Sutherland, Kenneth and Vincent, Sandra and Fisher, John T: Quantitative videobronchoscopy: a new technique to assess airway caliber. AJRCCM **154(6)**, 1794 (1996)
11. Freeman, W., Adelson: The design and use of steerable filters. PAMI **13(9)**, 891–906 (1991)
12. Gallo, G., Torrisi, A.: Lumen detection in endoscopic images: a boosting classification approach. IJAIS **5(1-2)**, 127–134 (2012)
13. Garcia-Barnes, J., Gil, D., Badiella, L., Hernandez-Sabate, A., Carreras, F., Pujades, S., Martí, E.: A normalized framework for the design of feature spaces assessing the left ventricular function. TMI **29**(3), 733–745 (2010)
14. Gil, D., Radeva, P.: Extending anisotropic operators to recover smooth shapes. Comp. Vis. Imag. Unders. **99**, 110–125 (2005)
15. Hayashi, A and Takanashi, S: New method for quantitative assessment of airway calibre using a stereovision fibreoptic bronchoscope. BJA **108(3)**, 512–516 (2012)
16. Heimann, T. and van Ginneken, B: Comparison and evaluation of methods for liver segmentation from ct datasets. TMI **28(8)**, 1251–1265 (2009)
17. Hein, E., Rutter, M.: New perspectives in pediatric airway reconstruction. International anesthesiology clinics **44(1)**, 51 (2006)
18. Helferty, James P and Zhang, Chao and McLennan, Geoffrey and Higgins, William E: Videoendoscopic distortion correction and its application to virtual guidance of endoscopy. TMI **20**(7), 605–617 (2001)

19. Hernàndez-Sabaté, A.: Exploring Arterial Dynamics and Structures in IntraVascular Ul-traSound Sequences. PhD thesis, Universitat Autònoma de Barcelona (2009)
20. Jowett, Nathan and Weersink, Robert A and Zhang, Kai and Campisi, Paolo and Forte, Vito: Airway luminal diameter and shape measurement by means of an intraluminal fiberoptic probe:a bench model. Archives Otolaryngology- Head and Neck Surgery **134(6)**, 637 (2008)
21. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. IJCV **1**(4), 321–331 (1988)
22. Lee, Karen S and Sun, Maryellen RM and Ernst, Armin and Feller-Kopman, David and Majid, Adnan and Boiselle, Phillip M: Comparison of dynamic expiratory ct with bron-choscopy for diagnosing airwaymalacia: a pilot evaluation. Chest **131**, 758–64 (2007)
23. Masters, I.B. and Eastburn, M.M: A new method for objective identification and measure-ment of airway lumeninpaediatricflexible videobronchoscopy. Thorax **60(8)**, 652 (2005)
24. McFawn, P.K. and Forkert, L. and Fisher, J.T.: A new method to perform quantitative measurement of bronchoscopic images. ERJ **18(5)**, 817–826 (2001)
25. Miller, Jr. and Rupert, G.: Beyond ANOVA: basics of applied statistics. CRC Press (1997)
26. Mori, Kensaku and Deguchi, Daisuke and Sugiyama, Jun and Suenaga, Yasuhito and Toriwaki, Jun-ichiro and Maurer, CR and Takabatake, Hirotsugu and Natori, Hiroshi: Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. Medical Image Analysis **6**(3), 321–336 (2002)
27. Murgu, S. and Colt, H.: Subjective assessment using still bronchoscopic images misclas-sifies airwaynarrowingin laryngotracheal stenosis. Interactive cardiovascular and thoracic surgery **16(5)**, 655–660 (2013)
28. Murgu, S. and Colt, H. G.: Morphometric bronchoscopy in adults with central airway obstruction: caseillustrationsand review of the literature. The Laryngoscope **119(7)**, 1318–1324 (2009)
29. Myer 3rd, C., O'connor, D., Cotton, R.: Proposed grading system for subglottic stenosis based on endotracheal tubesizes. The Annals of otology, rhinology, and laryngology **103(4 Pt 1)**, 319 (1994)
30. Norwood, S. and Vallina, V.L. and Short, K.: Incidence of tracheal stenosis and other late complications after percutaneoustracheostomy. Annals of surgery **232(2)**, 233 (2000)
31. Nouraei, SAR and McPartlin, DW and Nouraei, SM and Patel, A and Ferguson, C and Howard, DJ and Sandhu, GS: Objective sizing of upper airway stenosis: a quantitative endoscopic approach. The Laryngoscope **116**, 12–17 (2006)
32. Odry, B.L., Kiraly, A.P., Slabaugh, G.G., Novak, C.L., Naidich, D.P., Lerallut, J.F.: Active contour approach for accurate quantitative airway analysis. In: Medical Imaging, pp. 691,613–691,613. International Society for Optics and Photonics (2008)
33. Phee, SJ and Ng, WS and Chen, IM and Seow-Choen, F and Davies, BL: Automation of colonoscopy. ii. visual control aspects. EMBM **17(3)**, 81–88 (1998)
34. Polverosi, R and Vigo, M and Baron, S and Rossi, G: Evaluation of tracheobronchial lesions with spiral ct: comparison betweenvirtual endoscopy and broncoscopy. Radiol Med **102**, 313–9 (2001)
35. Rozycki, H.J., Van Houten, M.L., Elliott, G.R.: Quantitative assessment of intrathoracic airway collapse in infants andchildrenwithtracheobronchomalacia. Pediatric pulmonology **21(4)**, 241–245 (1996)
36. S., N.: Computed tomography and radiation exposure. N engl J Med **33**, 850–1 (2008)
37. Sánchez, C., Bernal, J., Gil, D., Sánchez, F.J.: On-line lumen centre detection in gastroin-testinal and respiratory endoscopy. In: MICCAI, *LNCS*, vol. 8361, pp. 31–38 (2014)
38. Sánchez, C., Gil, D., Rosell, A., Andaluz, A., Sánchez, F.J.: Segmentation of tracheal rings in videobronchoscopy combining geometry andappearance. In: VISAPP, vol. 1, pp. 153–161 (2013)
39. Sucar, L. E. and Gillies, D. F: Knowledge-based assistant for colonscopy. In: IEAIES, vol. 2, pp. 665–672 (1990)
40. Thrun, S., Leonard, J.J.: Simultaneous localization and mapping. Springer handbook of robotics pp. 871–889 (2008)
41. Vergnon, J.M., Costes, F., Bayon, M.C., Emonot, A.: Efficacy of tracheal and bronchial stent placement on respiratory functionaltests. Chest **107(3)**, 741–746 (1995)
42. Wieand, S., Gail, M.H., James, B.R., James, K.L.: A family of nonparametric statistics for comparing diagnostic markers withpairedor unpaired data. Biometrika **76(3)**, 585–592 (1989)
43. Zabulis, X. and Argyros, A. A. and Tsakiris, D. P.: Lumen detection for capsule endoscopy. In: IROS. IEEE/RSJ InternationalConference on, pp. 3921–3926. IEEE (2008)