

Obstructive pulmonary function: Patient classification using 3D registration of inspiration and expiration CT images

K. Murphy, B. van Ginneken, E.M van Rikxoort, B.J. de Hoop, M. Prokop, P. Lo, M. de Bruijne, and J.P.W. Pluim

University Medical Center, Utrecht, The Netherlands.

Abstract. Chronic Obstructive Pulmonary Disorder (COPD) is a condition with 4 classes of severity, characterised by reduced airflow and diagnosed by means of pulmonary function testing. CT scanning offers far more detailed information about the underlying pathology and site(s) of the disease than is afforded by a pulmonary function test, but is not currently used in the diagnosis and classification of COPD. In this work a classification system is presented to classify 110 subjects in a COPD database based on 2 thoracic CT scans taken at full inspiration and at full expiration for each subject. Experiments with a 2-class kNN classifier (COPD/non-COPD) as well as with a 5-class kNN classifier (COPD 1-4/non-COPD) are presented. Features are derived from 3 sources 1)The inspiration scan, 2)The expiration scan, 3)Comparison of the HU values at inspiration and expiration based on a 3D, fully automatic, non-rigid registration.

Classification in a 2-class system is particularly successful achieving an area under the ROC curve of $Az=0.92$. Multi-class classification is more challenging, with 66% of cases correctly classified and a further 28% classified in a class neighbouring the true one. A discussion is presented of the difficulties of precise multi-class classification due to drawbacks of the reference standard of pulmonary function testing. It is shown that the inclusion of registration-derived features improved the performance of the classifiers in all cases.

1 Introduction

Chronic Obstructive Pulmonary Disorder (COPD) is the fourth highest cause of chronic morbidity and mortality in the United States and by 2020 it is expected to be the fifth most important illness worldwide in terms of burden of disease [10]. COPD is characterised by airflow limitation which is not fully reversible and diagnosis is confirmed by means of spirometry (pulmonary function testing) [10]. Four levels of severity are defined according to the spirometry results (COPD classes 1-4), and the status of subjects without COPD is here referred to as 'class 0'. There may be several underlying causes of the airflow obstruction including emphysema and/or narrowing of the airways. While pulmonary function tests provide useful information about the overall extent of airflow restriction they do

not infer the pathological causes of this restriction or the regions of the lungs affected.

CT scans are used to examine the physical state of the lungs in COPD patients and to investigate the causes of the disease in more detail. CT-based measurements can be expected to provide quantitative information about airflow and have been shown to do so in many studies including [5,9]. In [9], for example, analysis of CT scans is used to compare the relative contributions of emphysema and airway-narrowing. Using measurements based on CT values creates the potential to simultaneously evaluate the individual components of the disease and its location within the lungs thereby improving physician's understanding of the mechanisms of the disease [11].

Both inspiration and expiration scans have a role to play in airflow analysis since emphysema is most easily visualised in inspiration scans, while regions of air-trapping (which may be caused by airway obstruction) are more obvious at full expiration. A direct regional comparison of CT scans taken at full inspiration and expiration is expected to provide valuable additional information, but visual comparisons are typically awkward to carry out due to the difficulty of aligning small regions in both scans. For a thorough quantitative analysis a full 3D registration of the scans is required.

In this work we attempt to determine whether quantitative airflow analysis by means of CT alone is sufficient firstly to diagnose COPD and secondly to categorise the subject into the correct COPD class using a k-Nearest-Neighbour (kNN) [3] classifier. Experiments with COPD diagnosis (COPD/non-COPD) were carried out using a 2-class kNN classifier. A 5-class kNN classifier was utilised in separate experiments to attempt to classify subjects into the correct COPD class (0-4). Features are extracted from both the inspiration scan (emphysema) and the expiration scan (air-trapping). Furthermore, we perform a fully automatic 3D registration of the inspiration and expiration scans, and extract additional features from a voxel by voxel comparison of the inspiration and expiration intensities. The various feature sets are compared to determine which features are most important in achieving an accurate classification.

2 Materials

In this study 110 pairs of scans are used which form part of a COPD database being constructed at our research institute. The subjects are either clinical patients or were recruited as part of a CT screening trial. All subjects are either at risk of developing COPD (former/current heavy smokers) or already have COPD. The numbers of scan pairs representing each COPD level (0-4, see section 3) are as follows: 0:29, 1:19, 2:33, 3:16, 4:13.

Each subject underwent a breath-hold CT scan at full inspiration and another at full expiration. All scans were acquired in less than 12 seconds. CT scanning protocol varied slightly depending on the subject but was usually set with a beam current of 30mAs for inspiration scans (low dose) and 20mAs for expiration scans (ultra-low dose).

All scans were obtained without contrast injection on a 16 detector-row scanner (Mx8000 IDT or Brilliance 16P, Philips Medical Systems). They have a per-slice resolution of 512×512 , with the number of slices per scan varying per subject. Slice thickness is 1mm with overlapping slices and a slice-spacing of 0.7mm

3 Reference Standard

COPD is diagnosed by means of spirometry (pulmonary function testing). During pulmonary function tests subjects are instructed to exhale fully into a mouth-piece and various measurements are made, including ‘Forced Expiratory Volume in 1 Second’ (FEV1) and ‘Forced Vital Capacity’ (FVC). FEV1 measures how much air volume is released in the first second of expiration, and FVC determines the entire volume exhaled. A value known as ‘FEV1-Predicted’ is determined using look up tables based on details such as the subject’s gender, height, weight, age and race. Table 1 illustrates how COPD is diagnosed based on these measurements. FEV1 as a percentage of FVC determines the presence or absence of COPD, while FEV1 as a percentage of FEV1-Predicted is used to establish the severity of the condition.

COPD Class:	0 (No COPD)	1 (Mild)	2 (Moderate)	3 (Severe)	4 (V. Severe)
Measurement:					
FEV1 as % of FVC	≥ 70	< 70	< 70	< 70	< 70
FEV1 as % of FEV1 Predicted		≥ 80	$\geq 50, < 80$	$\geq 30, < 50$	< 30

Table 1. The criteria for determination of COPD class by pulmonary function testing

4 Methods

4.1 Pre-Processing

As an initial pre-processing step all inspiration and expiration scans were sub-sampled in order to make certain subsequent procedures more efficient. The images were reduced by block averaging to 256×256 voxels in the X-Y plane with the number of slices reduced such that the data was isotropically sampled. Linear interpolation was used to determine grey-values between voxel locations.

Segmentation masks for the lungs and lobes were computed for all scans. The lung segmentation algorithm was by Sluimer et al. [15] based on the method of Hu et al. [6]. Lobe segmentation was performed using an algorithm by van Rikxoort et al. [17]. All segmentations included in this study were visually checked.

Distance transforms from the lung boundaries were computed as follows: The lung mask was eroded with a spherical kernel of radius 5 voxels and a distance transform was calculated on the eroded segmentation. The erosion served to ensure that voxels inside the true lung boundary also obtained a distance value in the transform. Values above 10 in the distance transform image were clamped. These distance transform images were subsequently used in the registration procedure.

4.2 Registration

Registration was carried out on the down-sampled images in order to reduce memory consumption. The registration was implemented using `elastix` version 4.0 (<http://elastix.isi.uu.nl>) which is a registration toolkit based on the National Library of Medicine Insight Segmentation and Registration Toolkit (ITK). The expiration scan was transformed to match the inspiration scan and the calculated transform was subsequently applied to the original full resolution expiration data. The registration procedure consisted of an initial affine registration step followed by a non-rigid registration to handle the deformations of the lung tissue. Both steps involved a multi-resolution strategy with 5 resolution levels for the affine procedure and 7 for the non-rigid. The cost function used was a combination of the mutual information (MI) [16] of the down-sampled scans and the sum of squared differences (SSD) of the lung boundary distance images described in section 4.1. The lung boundary distance images were included as an additional guide to the registration to ensure that the lung boundaries were well aligned. The MI cost was weighted more heavily (0.75) than the SSD cost (0.25). Optimisation was by means of a stochastic gradient descent optimizer [7]. The non-rigid registration deformations were modelled by a B-Spline grid [12]. The grid-size varied per resolution-level with the finest grid at the last level having a spacing of 10mm in each dimension. The lung segmentation described in section 4.1 was used as a mask for the inspiration scan to ensure that during registration the cost function was calculated on samples from the volume of the segmented lungs only. The registration settings were chosen bearing in mind that there may be very large deformations between the inspiration and expiration scans. Registration using these settings took approximately 45 minutes per scan pair.

4.3 Vessel Segmentation

In order to exclude vessels from subsequent feature calculations (in which only lung parenchyma is of interest) a vessel segmentation is performed. Although the registration was reasonably accurate for most vessel structures, on some occasions vessel alignment was imperfect. For this reason vessel segmentation was carried out on the inspiration and transformed expiration images separately. Vessel segmentation was also carried out on the original expiration image.

The vessels were segmented using the algorithm of Lo et. al. [8]. Since the majority of the scans used a low-dose or an ultra-low dose protocol, the best results were obtained by using the full resolution scan data and applying a noise

filter [13] prior to the segmentation. The vessel segmentation was dilated with a spherical kernel of radius 1 voxel to ensure that undersegmentation was not an issue. In addition a conservative thresholding was applied to the original image to segment the brightest voxels. The final segmentation comprised the union of the dilated vessel segmentation and the thresholded bright voxels. The threshold t was set at $t = CT_{min} + ((CT_{max} - CT_{min}) \times 0.5)$ where CT_{min} and CT_{max} represent the minimum and maximum intensity values within the lung volume. In practise, to avoid CT_{min} and CT_{max} being influenced by noise or minor segmentation errors, CT_{min} is given the value of the 1st centile and CT_{max} the value of the 99th centile. Examples of vessel segmentations in inspiration and expiration are shown in figure 1. Note that there are occasionally voxels on the pleural surface which are included in the vessel segmentation due to the thresholding step and inaccuracies in the lung segmentation. These are not problematic in our application as our vessel segmentation serves only to exclude these voxels during feature calculation.

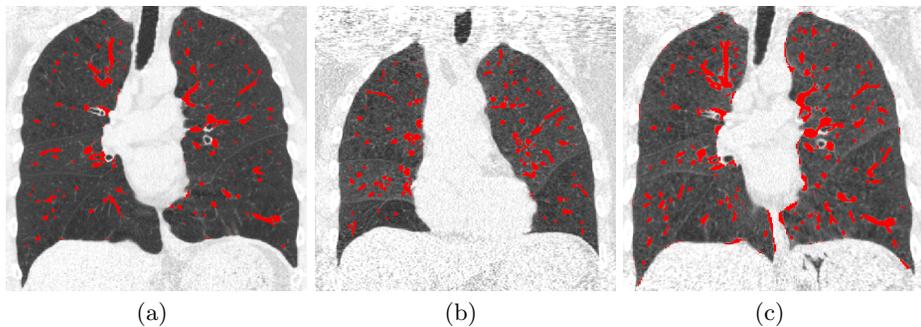


Fig. 1. Vessel Segmentations in (a) inspiration, (b) expiration and (c) transformed expiration

4.4 Feature Calculation

With the aim of classifying the subjects based on their COPD severity a number of features were calculated for each image pair. The features can be divided into 3 subsets as follows: Feature set A - Features of the inspiration image only, Feature set B - Features of the expiration image only, Feature set C - Features based on a voxel-by-voxel comparison of the inspiration image with the registered (transformed) expiration image. The features consist of percentages and averages which are calculated over both lungs, over each lung individually and over each lobe individually. Per-lung and per-lobe features are included since they may prove more useful than global features in the classification of COPD. The full list of 48 calculated features is given in table 2.

Feature calculation was carried out on the down-sampled versions of the images to conserve memory usage. Prior to calculating the features the inspiration, expiration and transformed expiration scan were noise-filtered [13]. Blood mass correction (BMC) [4] was applied to the transformed expiration image before the calculation of feature set C where the inspiration and transformed expiration images are compared directly. The theory of BMC is based on the assumption that the total mass of the lungs is not changed by the actions of inhaling and exhaling air. Calculating the masses according to the CT intensity values however, it can be demonstrated that a minor change in mass occurs. This change in mass is due to alterations in blood-flow in the lungs during the breathing process [2]. The BMC calculates the difference in mass for each lung between inspiration and expiration and thereby derives a correction factor for voxels in that lung. The grey-value of each voxel in the lung is adjusted in the transformed expiration image such that the corrected mass of the lung is equal to that of the same lung in inspiration.

Comparison of subjects based on the relationship between their inspiration and expiration scans can cause difficulties. For example, subject A may show a smaller change in intensity values between inspiration and expiration than subject B, which might suggest that subject A has a comparatively reduced airflow. However such a difference may also be due to the fact that subject A did not inhale or exhale as completely or deeply as subject B for the purposes of the CT scans. For this reason, all features in feature set C were normalised by multiplication with the value C calculated by $C = (Vol_{INSP} - Vol_{EXP})/Vol_{INSP}$ where Vol_{INSP} is the total lung volume at inspiration and Vol_{EXP} the total lung volume at expiration. Vessel exclusion during calculation of feature set C was achieved by excluding any voxel which was segmented as vessel in either the inspiration image or the transformed expiration image.

ID	Description	Feature set
1	Emphysema(-950). Percentage of voxels below -950HU in insp.	A
2	Emphysema(-910). Percentage of voxels below -910HU in insp.	A
3	Air-trapping(-850). Percentage of voxels below -850HU in exp.	B
4	Ventilation. Average of ventilation values v where $v = (1000(CT_{INSP} - CT_{EXP}))/CT_{EXP}(CT_{INSP} + 1000)$ [14]	C
5	Subtraction. Average of subtraction values s where $s = CT_{EXP} - CT_{INSP}$	C
6	Ratio. Average of ratio values r where $r = CT_{EXP}/CT_{INSP}$	C

Table 2. Calculated features. Each of the 6 listed features is calculated over 8 regions (The total lung volume (1), each lung independently (2) and each lobe independently (5)), making 48 features in total.

4.5 Classification

All classification experiments used k-Nearest-Neighbour (kNN) classifiers [3]. Briefly, each subject in the training set is plotted in feature-space based on its calculated feature values. New subjects are plotted similarly and classified based on the classes of the k nearest training subjects in feature-space.

Two types of classification were performed. Firstly 2-class classification was carried out, where subjects were classified as either having COPD (Class 1) or not (Class 0). Secondly, multi-class classification was attempted, giving each subject one of five COPD classes (0-4) as per table 1. In each case experiments were carried out using (1) a combination of all feature sets (feature sets A+B+C), (2) Only features which do not require registration (feature sets A+B) and (3) Only features which are derived from registration (feature set C). For all experiments values of k between 4 and 15 were tested and the k giving the best results was selected in each case. A leave-one-out training and testing procedure was applied in all cases.

5 Results

5.1 2-class Classification

The best results for 2-class classifications with various datasets are shown in figure 2. An ROC curve is shown in each case, with the area under the curve (A_z) and the value of k used listed in the legend. The best results were obtained using all possible features (feature sets A+B+C) where an A_z value of 0.92 was achieved. Using only registration-related features (feature set C) gave a reduced A_z of 0.89, while using only features of the individual inspiration and expiration images (feature set C) gave the worst result with $A_z=0.88$.

5.2 Multi-Class Classification

Multi-class classification results are illustrated in figure 3. The bar-chart illustrates for all items in each class (and overall) what percentage of the items were classified correctly and incorrectly. Incorrect classifications are further divided into 1-Class errors (meaning that the chosen COPD class was a neighbour of the true COPD class) and larger errors (mainly 2-class errors, see figure 4 for more detail). The leftmost bar in each group represents classification using the full feature set (A+B+C), the central bar using feature sets A+B and the rightmost bar using feature set C alone. Overall the best performance is achieved with feature set A+B+C, while feature set C alone achieves almost the same result. Excluding the registration-based features (using feature set A+B) gives a distinct reduction in the performance.

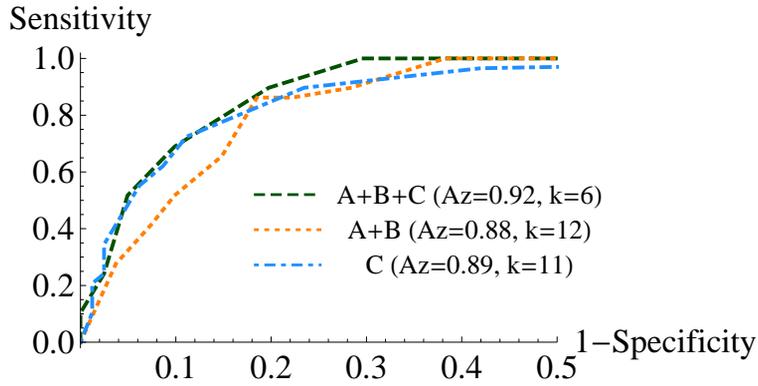


Fig. 2. ROC curves illustrating the performance of the 2-class leave-one-out classification systems with various feature sets.

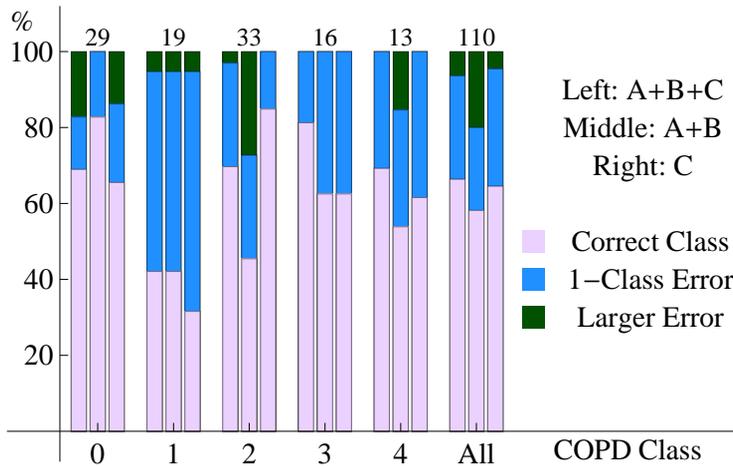


Fig. 3. Bar charts illustrating the performance of the multi-class leave-one-out classification systems with various feature sets. The number of subjects in each group is indicated by the figures above the bars.

6 Discussion

It has been shown that using automatic analysis of inspiration and expiration CT scans, subjects can be classified in a 2-class (COPD/non-COPD) system with

a high degree of accuracy ($A_z=0.92$). Furthermore, multi-class classification is fully successful in 66% of cases, and chooses a COPD class neighbouring the correct one in a further 28%.

Precise multi-class classification of this data is difficult, partly due to the nature of the reference standard. Although pulmonary function tests are used as a gold standard to classify COPD cases they have a number of associated drawbacks and issues with reproducibility. Firstly, the test results may vary depending on how well the subject understands and adheres to the given instructions, the condition of the subject on the day of testing, the instrument used and various other factors [1]. Secondly, there is always variation in the population around the FEV1-predicted values which are based on subject height, age, weight, gender etc. [1]. Finally, since the measures used to classify COPD are on a continuous scale, even the most accurate measurements may show a subject to be very close to the border of 2 COPD classes.

In an attempt to understand the relationship between pulmonary function test scores and our system mis-classifications, all 37 cases that were misclassified in the multi-class experiment (using all features) were analysed. Figure 4 illustrates the results of this analysis with FEV1/FVC scores shown in the upper part of the figure, and FEV1/FEV1-Predicted scores in the lower part. It can be seen that the pulmonary function test scores of a number of the mis-classified cases place them very close to the boundary between two COPD classes. Cases 34, 35 and 36 (on the X-axis) for example, have FEV1/FEV1-Predicted scores placing them at the boundary between COPD classes 3 and 4. The reference standard narrowly places them in class 4, while our classifier placed them in class 3. Case number 30 has an FEV1/FEV1-Predicted score which places him in class 2, however his FEV1/FVC score is very close to the 70% boundary line, above which he would have been classified into class 0. Our classifier places this subject in class 0 indicating that his CT scans do not reveal classic signs of COPD. Further investigation would be required to determine whether the pulmonary function test scores or the CT scans of this patient truly reflect his condition. Ultimately there may be a case for using quantitative analysis of CT data in combination with results of pulmonary function tests when diagnosing COPD subjects.

All experiments illustrated that both 2-class and multi-class classification are improved by the inclusion of feature set C, based on the automatic 3D registration of the inspiration and expiration images. Direct comparison of the percentage of air present in a voxel at inspiration and at expiration is impossible without an accurate registration, and provides a qualitative measure of the pulmonary function at that precise location.

The relationship between CT-derived features and COPD classification which has been demonstrated is an extremely important one. Further research opportunities now exist to determine whether CT should be routinely used in the diagnosis and severity classification of COPD. In addition, more precise relationships between CT features and COPD can be investigated to determine, for

example, whether impairment in specific lobes affects the spirometry measurements more seriously than in others.

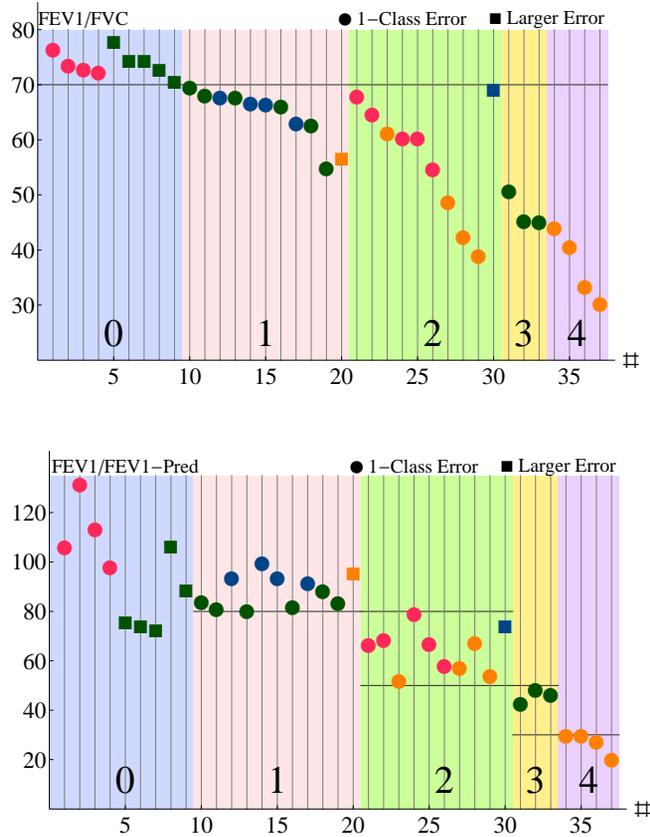


Fig. 4. Analysis of pulmonary function scores for 37 mis-classified subjects - Above: FEV1/FVC scores, Below: FEV1/FEV1-Predicted scores. Subjects are ordered by their reference standard COPD class from left to right (see background colours). The colour of each point indicates the class into which the subject was incorrectly classified. (Colours are as per the background, thus: 0=blue, 1=pink, 2=green, 3=orange, 4=purple). Horizontal lines depict the boundaries of the COPD class as defined in table 1.

References

1. M.R. Becklake. Concepts of normality applied to the measurement of lung function. *Am J Med*, 80(6):1158–1164, 1986.
2. R. Brower, R.A. Wise, C. Hassapoyannes, B. Bromberger-Barnea, and S. Permutt. Effect of lung inflation on lung blood volume and pulmonary venous flow. *J Appl Physiol*, 58(3):954–963, 1985.
3. T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Trans. Information Theory*, 13(1):21–27, 1967.
4. T. Guerrero, K. Sanders, E. Castillo, Y. Zhang, L. Bidaut, T. Pan, and R. Komaki. Dynamic ventilation imaging from four-dimensional computed tomography. *Phys Med Biol*, 51(4):777–791, 2006.
5. A. Heremans, J. A. Verschakelen, L. van Fraeyenhoven, and M. Demedts. Measurement of lung density by means of quantitative CT scanning. A study of correlations with pulmonary function tests. *Chest*, 102(3):805–811, 1992.
6. S. Hu, E.A. Hoffman, and J.M. Reinhardt. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans Med Imaging*, 20:490–498, 2001.
7. S. Klein, M. Staring, and J.P.W. Pluim. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *IEEE Trans Image Process*, 16:2879–2890, 2007.
8. P. Lo, J. Sporning, H. Ashraf, J.J. Holst Pedersen, and M. de Bruijne. Vessel-guided airway segmentation based on voxel classification. In *The First International Workshop on Pulmonary Image Analysis - MICCAI*, pages 113–122, 2008.
9. Y. Nakano, S. Muro, H. Sakai, T. Hirai, K. Chin, M. Tsukino, K. Nishimura, H. Itoh, . D. Paré, J.C. Hogg, and M. Mishima. Computed tomographic measurements of airway dimensions and emphysema in smokers. *Am J Respir Crit Care Med*, 162(3):1102–1108, 2000.
10. K. F. Rabe, S. Hurd, A. Anzueto, P.J. Barnes, S.A. Buist, P. Calverley, Y. Fukuchi, C. Jenkins, R. Rodriguez-Roisin, C. van Weel, and J. Zielinski. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: Gold executive summary. *Am J Respir Crit Care Med*, 176(6):532–555, 2007.
11. J. Reilly. Using computed tomographic scanning to advance understanding of chronic obstructive pulmonary disease. *Proc Am Thorac Soc*, 3(5):450–455, 2006.
12. D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Trans Med Imaging*, 18(8):712–721, 1999.
13. A.M.R. Schilham, B. van Ginneken, H. Gietema, and M. Prokop. Local noise weighted filtering for emphysema scoring of low-dose CT images. *IEEE Trans Med Imaging*, 25:451–463, 2006.
14. B.A. Simon. Non-invasive imaging of regional lung function using X-ray computed tomography. *J Clin Monit Comput*, 16(5-6):433–442, 2000.
15. I.C. Sluimer, M. Prokop, and B. van Ginneken. Towards automated segmentation of the pathological lung in CT. *IEEE Trans Med Imaging*, 24(8):1025–1038, 2005.
16. P. Thévenaz and M. Unser. Optimization of mutual information for multiresolution image registration. *IEEE Trans. Image Proc.*, 9:2083 – 2099, 2000.
17. E.M. van Rikxoort, M. Prokop, B. de Hoop, M.A. Viergever, J.P.W. Pluim, and B. van Ginneken. Automatic segmentation of the pulmonary lobes from fissures, airways, and lung borders: evaluation of robustness against missing data. In *Medical Image Computing and Computer-Assisted Intervention (in press)*, 2009.