3D Case–Based Retrieval for Interstitial Lung Diseases

Adrien Depeursinge[1], Alejandro Vargas[1], Alexandra Platon[2],
Antoine Geissbuhler[1], Pierre–Alexandre Poletti[2], and Henning Müller[1,3]

[1] Medical Informatics, Geneva University Hospitals and University of Geneva, CH,
[2] Service of Emergency Radiology, University Hospitals of Geneva, CH,
[3] Business Information Systems, University of Applied Sciences Sierre, CH.

Abstract. In this paper, a computer–aided diagnosis (CAD) system that retrieves similar cases affected with an interstitial lung disease (ILDs) to assist the radiologist in the diagnosis workup is presented and evaluated. The multimodal inter–case distance measure is based on a set of clinical parameters as well as automatically segmented 3–dimensional regions of lung tissue in high–resolution computed tomography (HRCT) of the chest. A global accuracy of 75.1% of correct matching among five classes of lung tissues as well as a mean average retrieval precision at rank 1 of 71% show that automated lung tissue categorization in HRCT data is complementary to case–based retrieval both from the user's viewpoint and also on the algorithmic side.

1 Introduction

Interstitial lung diseases (ILDs) can be characterized by the gradual alteration of the lung parenchyma leading to breathing dysfunction. They regroup more than 150 disorders of the lung parenchyma. The factors and mechanisms of the disease processes vary from one disease to another. The diagnosis of these pathologies is established based on the complete history of the patient, a physical examination, laboratory tests, pulmonary function testing as well as visual findings on chest X–ray. When the synthesis of this information arouses suspicions toward an ILD, high–resolution computer tomography (HRCT) imaging of the chest is often required to acquire a rapid and accurate visual assessment of the lung tissue. Compared to the radiograph, the tomographic image acquisition process avoids the superposition of the lung tissue with the ribs and other organs leading to three–dimensional images of the lung volumes. Nevertheless, the interpretation of HRCT is often challenging with numerous differential diagnoses and it is currently reserved to experienced radiologists [1]. A correct interpretation requires an advanced knowledge of the lung anatomy and the alterations of the interstitial tissues have wide intra and inter–disease variations. Moreover, the large number of slices contained in an HRCT image series makes the interpretation task time–consuming and subject to missing relevant lung tissue patterns.

Recent advances in medical informatics enabled access to most of the radiological exams to all clinicians through the electronic health record (EHR)

and the picture archival and communication system (PACS). This change of the medical workflow calls upon computer expert systems able to bring the right information to the right people at the right time. Interpreting HRCT of the chest is no exception to this and has been an active research domain during the last decade [2–5]. Most of the proposed systems aim at categorizing lung tissue to provide a second opinion to radiologists. This provides a quick and exhaustive scan of the large number of images and can draw the radiologist's attention on diagnostically useful parts of the images. To be useful in clinical practice such systems have to be able to detect a sufficient number of types of lung tissue [6]. Several studies obtained recognition rates of up to 90% correct matches [7] but usually while training and testing the classifier with images belonging to the same patients, which introduces a large bias compared to a real–world clinical usage with unknown incoming images [8].

In the context of medical image analysis, providing quick and precious information to the clinician is not limited to automatic recognition of abnormal tissue and/or structure. The approach of the clinician to a diagnosis if he has little experience of the domain is to compare the image under investigation with typical cases with confirmed diagnosis listed in textbooks or contained in personal collections. It allows to rule out diagnostics and, in association with clinical parameters, prevents the reader from mixing diagnoses with similar radiological findings. This process allows the clinician to partly replace a lack of experience but has two major drawbacks: searching for similar images is time–consuming and the notion of similarity may be subjective and can be ambiguous [9].

Content–based medical image retrieval (CBIR) aims at finding objectively visually similar images in large standardized image collections such as PACS [10]. For instance, the Radiology Department of the University Hospitals of Geneva (HUG) alone produced more than 80,000 images a day in 2008, representing retrospectively a potentially large repository of knowledge and experience as images are all associated with one or several diagnosis. The notion of similarity is usually established from a set of visual features describing the content of the images. Features can vary from low–level measures such as the histogram quantification of the colors to high–level semantically–related features describing the anatomical content of images. Few CBIR systems have been evaluated in clinical practice but some of them showed that they can be accepted by the clinicians as a useful tool [11, 12]. The use of a CBIR system clearly increased the number of correct diagnoses within the context of the interpretation of HRCT images associated with ILDs in [13]. A possible extension to CBIR is to carry out case–based retrieval. Most often, the clinician actually looks for similar cases as he considers the image within the context of a patient with a personal history, findings on the physical examination, laboratory tests, etc. Radiologists never interpret an image without taking into account the clinical context defined by disease–specific metadata.

In this paper we show that automated lung classification in HRCT data is complementary to case–based retrieval, both from the user's viewpoint and also on the algorithmic side. In a first step, healthy and four abnormal tissue

types associated with 7 of the most common ILDs are automatically detected in HRCT. These latter are *emphysema*, *ground glass*, *fibrosis*, *micronodules* and *healthy*. Then, based on the volumes of the segmented tissues and a set of selected clinical parameters, similar cases are retrieved from a multimedia database of ILD cases built at the HUG within the context of the Talisman project [1] .

2 Methods

This section describes the various steps of our computer–aided diagnosis (CAD) system for ILDs consisting of semi–automatic segmentation of the lung volumes, classification of the lung tissue based on texture properties, and multimodal retrieval of similar cases.

2.1 Semi–Automatic Segmentation of the Lung Volumes

Segmentation of the lung volumes is a required preliminary step to lung tissue categorization. The result of this step is a binary mask M_{lung} that indicates the regions to be analyzed by the texture analysis routines. Since the geometries and shapes of the lungs are subject to large variations among the cases, semi–automatic segmentation based on region growing and mathematical morphology is carried out. The region growing routine contained in YaDiV [2] is used. Starting from a seed point $s(x, y, z)$ defined by the user, each 26–connected neighbor is added to the region M_{lung} if the summed value of its own neighbors differs of less than a given variance v defined by the user. At this stage, M_{lung} describes the global lung regions well but contains many holes where the region growing algorithm was stopped by denser regions such as vessels or consolidations of the lung parenchyma. To fill these holes, a closing operation is applied to M_{lung} using a spherical structuring element. Two parameters require attention from the user: the radius r of the structuring element in millimeters and N_{op} which defines the number of closing operations.

2.2 Automated Lung Tissue Categorization

Our approach for categorizing lung tissue patterns associated with ILDs in HRCT relies on texture analysis. Most of the patterns depict nonfigurative and cellularly organized areas of pixels. To describe texture properties, features based on grey–level histogram in Hounsfield Units (HU) as well as statistical measures from a tailored wavelet transform are extracted. A support vector machine (SVM) classifier is used to draw boundaries among the distinct classes of lung tissue represented in the feature space.

---

[1] Talisman: Texture Analysis of Lung ImageS for Medical diagnostic AssistaNce, `http://www.sim.hcuge.ch/medgift/01_Talisman_EN.htm`, as of 8 November 2009

[2] YaDiV: Yet Another DIcom Viewer, `http://www.welfenlab.de/en/research/fields_of_research/yadiv/`, as of 8 November 2009

Grey–Level Histogram CT scanners deliver DICOM images with pixel values in HU that can be univoquely mapped to the density of the observed tissue. Thus, essential information is contained directly in the grey–levels. To encode this information, 22 histogram bins bin_j of grey–levels in $[-1050; 600[$ corresponding to the interval of the lung HU values (including pathological) are used as texture features. An additional feature pix_{air} counts the number of air pixels which have value below -1000 HU.

Wavelet–Based Features To be complementary to the grey–level histogram, attributes describing the spatial distribution of the pixels are required. Multi-scale analysis using wavelet transforms proved to be adequate for texture analysis [14] but requires to control the three essential affine parameters: translation, scale–progression and directionality. For lung tissue analysis, we assume that lung tissues patterns in transversal slices are translation and rotation–invariant. Moreover, a fine and initializable scale–progression is necessary to distinguish between vessels and micronodules. To assess translation invariance, a wavelet frame transform is used. Isotropic polyharmonic B–spline wavelets along with the nonseparable quincunx subsampling scheme yield a near isotropic wavelet transform with fine and tunable scale–progression [15]. The classical separable wavelet transform tends to favor the vertical and horizontal directions, and produces a so–called "diagonal" wavelet component, which does not have a straight-forward directional interpretation. The quincunx scale–progression is finer than the widely used dyadic one with a subsampling factor of $\sqrt{2}$ instead of 2. In addition to be near isotropic by implementing a multiscale smoothed version of the Laplacian Δ, isotropic polyharmonic B–spline wavelets can be scaled easily using the order γ that iterates Δ:

$$\psi_\gamma(\mathbf{D}^{-1}\mathbf{x}) = \Delta^{\frac{\gamma}{2}}\{\phi\}(\mathbf{x}), \tag{1}$$

, where ϕ is an appropriate smoothing (low–pass) function and $\mathbf{D} = [1\ 1; 1\ -1]$ is the quincunx subsampling matrix. Statistical measures of the wavelet coefficients are extracted as texture features. Two variances $\sigma^i_{1,2}$ and fixed means $\mu^i_{1,2} = \mu^i$ of a mixture of two Gaussians are estimated using the expectation–maximization (EM) algorithm for each subband $i \in [1; 8]$. The low–pass filtered images are left aside. $\gamma = 3$ obtained the best accuracy of the lung tissue patterns in [16].

Blockwise Classification In order to automatically categorize every pixel of M_{lung}, each 2D slice is divided into overlapping blocks. Preliminary results using block sizes of $\{8\times8; 16\times16; 24\times24; 32\times32; 40\times40; 48\times48; 56\times56; 64\times64\}$ showed that optimal blocks of size 32×32 is the best trade–off between classification performance and localization. For each block, 22 histogram bins bin_j of GLH in [-1050; 600] and the number of air pixels pix_{air} are concatenated into one hybrid feature vector $\mathbf{v}$ along with GMM parameters of 8 iterations of the quincunx wavelet transform using β_γ of order $\gamma = 3$:

$$\boldsymbol{v} = (bin_1 \ldots bin_{22} \quad pix_{air} \quad \mu^1\ \sigma^1_1\ \sigma^1_2 \ldots \mu^8\ \sigma^8_1\ \sigma^8_2) \tag{2}$$

An SVM classifier learns from the space spanned by $\boldsymbol{v}$ to find the decision boundaries among five classes of lung tissue. The optimal cost of the errors C and the width of the Gaussian kernel σ_{kernel} are found using a grid search with $C \in [0; 100]$ and $\sigma_{kernel} \in [10^{-2}; 10^2]$.

2.3 Multimodal Case–Based Retrieval

In order to retrieve similar cases from a database to assist the clinician in diagnosis of ILDs, a distance measure based on the volumes of segmented tissue groups as well as on clinical parameters is used. Case–based retrieval is enabled by the automated categorization of the entire HRCT image series. The three–dimensional map of the lung tissue obtained with the blockwise classification of the lung regions yields a semantically–related basis for the comparison of the cases. The percentages of the respective volumes v_i of the five classes of lung tissue are used to assess the visual similarity between HRCT image series from two patients. The respective volumes of lung tissue are semantically related to the ILDs as each histological diagnosis is associated to a given combination of HRCT findings. This allows to reduce the semantic gap between the user's intentions and the visual features, which is often a bottleneck in CBIR [17]. The Euclidean distance is computed from the percentages of the five volumes of tissue as follows:

$$d_{vol} = \sqrt{v_h^2 + v_e^2 + v_g^2 + v_f^2 + v_m^2} \tag{3}$$

with v_h corresponding to *healthy* tissue, v_e to *emphysema*, v_g to *ground glass*, v_f to *fibrosis* to v_m for *micronodules*.

44 clinical parameters with two levels of importance are used to assess the "meta–similarity" between the cases. The levels of importance are defined by a physician according to the relevance for establishing the diagnosis of eight common ILDs. 3 clinical parameters of first importance include age, gender and smoking history. Another 41 parameters of second importance included physical findings, medical history, and laboratory results. The parameters associated with biopsy outcomes were not included as the goal of the CAD is to provide quick information to the radiologists before any biopsy.

The multimodal distance measure d_M between two cases is computed as a linear combination of three modalities:

$$d_M = a_1 d_{vol} + a_2 d_{param1} + a_3 d_{param2}, \tag{4}$$

with a_j being the weights of each modality. d_{vol} is the Euclidean distance in terms of percentages of the volumes of segmented tissue according to (3) and $d_{param1,2}$ the euclidean distance in terms of clinical parameters of importance 1 and 2, respectively. d_{vol} and $d_{param1,2}$ are normalized before being combined in (4).

3 Results

In this section, the techniques described in Section 2 are applied to a multimedia dataset consisting of 76 cases with at least one annotated HRCT image series
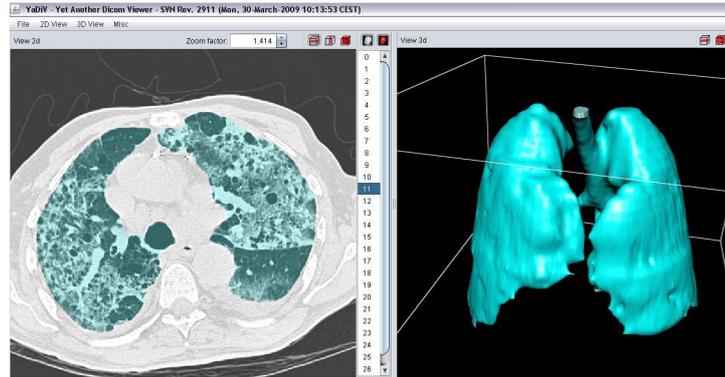
Fig. 1. An example of the segmentation of the lung volumes using a modified version of YaDiV.

with slice thickness of $1mm$ and no contrast agent. Annotation of regions was performed by two experienced radiologists at the HUG. The diagnoses of each of the cases was confirmed either by pathology (biopsy, bronchoalveolar washing) or by a laboratory/specific test confirming the diagnosis. For 69 of the 76 cases, 99 clinical parameters were collected from the EHR, describing the patient's clinical state at the time of the stay when the HRCT image series was acquired. 46 of these parameters were used for the retrieval of similar cases in (4).

To obtain recognition rates of the lung tissue patterns that are representative for actual clinical situations, a leave–one–patient–out cross–validation [8] is used to avoid training and testing the SVM classifier with images belonging to the same patient. All images from the selected case are left aside for testing whereas the remaining images are used to train the SVMs. For each case, lung volumes are segmented using YaDiV, where a tab was added for the closing operation. An example of the segmentation is depicted in Figure 1.

Then, the whole lung volume is segmented using a distance between the centers of the blocks equal to 4 pixels, leading to an overlap of 87.5% (see Figure 2). Table 1 shows the confusion matrix of the segmented tissue sorts. The associated performance measures are listed in Table 2. Note that some patients may contain several sort of lung tissue. To assess case–based retrieval performance, mean precisions at rank 1, 5, 10 and at rank equal to the number of instances of the diagnosis N_r are computed using a leave–one–patient–out cross–validation with 69 cases (see Table 3). A grid search for optimal weights of the modalities in (4) is carried out with $a_j \in [0 : 50[$.

4 Discussion

The results obtained with the various components of the proposed CAD are discussed in this section. Our experience with the segmentation of 69 lung volumes
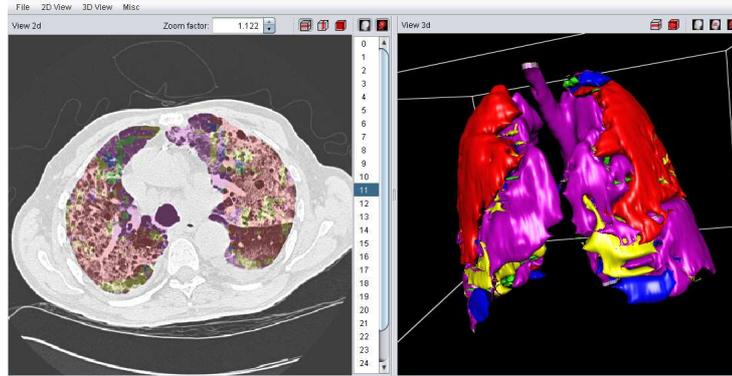
Fig. 2. Automated segmentation of the lung tissue patterns of a patient affected with pulmonary fibrosis. The 3D segmented regions are displayed to the clinician using YaDiV. Green: *healthy* (0.1 liters), blue: *emphysema* (0.39 l), yellow: *ground glass* (0.53 l), red: *fibrosis* (1.91 l), pink: *micronodules* (1.77 l).

Table 1. Confusion matrix of the blockwise classification of lung tissue patterns using a leave–one–patient–out cross–validation in %. Global arithmetic and geometric means of 75.1% and 74.7% are obtained respectively. N_{vox} denotes the number of manually segmented voxels used for evaluation and N_{cases} the number of patients.

	healthy	emphysema	ground glass	fibrosis	micronodules	N_{vox}	N_{cases}
healthy	**78.1**	2.8	0.7	0.2	18.1	63'914	7
emphysema	0.9	**70.1**	0	4.7	24.2	61'578	5
ground glass	4.6	1.6	**76**	14.7	3.1	644'814	21
fibrosis	2.3	1.9	17	**73.5**	5.3	860'474	28
micronodules	13.7	1.8	2.2	6.7	**75.7**	1'436'055	10

Table 2. Performance measures of the blockwise classification of the lung tissue patterns using QWF and GLH features.

	recall	precision	F–measure	accuracy
healthy	78.4	78.2	78.3	91.3
emphysema	89.6	70.2	78.7	92.4
ground glass	79.2	76	77.6	91.2
fibrosis	73.6	73.5	73.6	89.4
micronodules	59.9	75.6	66.8	85

Table 3. Mean precisions based on the diagnosis of the retrieved cases. The values of the weight a_i that allowed best global precisions show the respective importances of the modalities. Abbreviations: BOOP: bronchiolitis obliterans organizing pneumonia, PCP: pneumocystis pneumonia.

	$P@1$	$P@5$	$P@10$	$P@N_r$	N_r
Fibrosis	79.2	58.3	51.7	42.7	24
BOOP	60	20	18	20	5
Miliary tuberculosis	71.4	48.6	34.3	42.9	7
PCP	25	20	10	25	4
Hypersensitivity pneumonitis	54.5	40	39.1	38	11
Acute interstitial pneumonia	66.7	33.3	25.5	27.2	9
Sarcoidosis	100	66.6	52.2	56.8	9
average/total	59.4	39.7	34.2	32.4	69
weights $a_{1,2,3}$ in (4)	8;1;39	6;9;38	8;5;48	10;4;41	

shows that the 3D region growing associated with closing allows an almost fully–automatic segmentation. However, the trachea is included as lung tissue in most cases, which may bias the volume percentages of the five patterns in (4). Manual corrections are required when the closing operation cannot fill large regions of consolidated tissue.

The automatic segmentation of the lung tissue is a crucial step for the success of the CAD. The accuracies obtained in Table 2 show that the SVM classifier can learn efficiently from the hybrid feature space. However, the recurrent confusion between *healthy* and *micronodules* patterns suggests that the decision boundaries are not trivial in some cases (see also Figure 2). Table 1 also shows recurrent confusions between *fibrosis* and *ground glass*. This may be partially explained by the fact that *fibrosis* patterns are most often accompanied with small regions of *ground glass* because of the re–distribution of the perfusion to the functional tissue remaining. This has the effect to overload the healthy tissue which thus has the visual appearance of *ground glass* because of increased attenuation. However, during the annotations sessions, the label *fibrosis* was assigned to the whole ROI leading to classification errors when the system correctly detects the small *ground glass* regions. Using the clinical context of the images such as the age of the patient showed to allow clarifications between visually similar patterns in [18]. For instance, *micronodules* in a 20–year–old subject are very visually similar to *healthy* tissue surrounded by vessels of a 80–year–old man. In case of unbalanced classes, SVMs classifiers with asymmetric margins can be used to favor minority classes. At the border of the lungs, missclassifications occur due to the response of the wavelets to the sharp change of intensity. A solution to this is to use the symmetry of the lung tissue using the tangent to the lung border as axis. To remove noise in the blockwise classification, a 3D averaging of the lung regions may avoid small isolated regions.

The retrieval precisions presented in Table 3 are currently fairly low to be used in clinical routine but show the feasibility of indexing ILD cases using the

volumes of automatically segmented lung tissue as well as clinical parameters. It is important to note that the link between visual similarity of two HRCT scans and their associated diagnoses is not straightforward. The values of the weights $a_{1,2,3}$ that allowed best performances reflect the importance of the each modality. High values obtained for a_3 shows the unexpectedly high importance of the clinical parameters of secondary priority. High variations of the precision can be explained by the fact that the number of cases is still fairly small, particularly for BOOP and PCP. The link between visual similarity of two HRCT scans and their associated diagnosis is not straightforward. Further improvements are required to highlight the importance of the visual similarity: the low–level feature vector v can be used directly as in (4), under the condition to overcome the difficulty in setting up a standardized localization system for the lung anatomy.

5 Conclusion and Future Work

Image-based diagnosis aid tools for ILDs are available for evaluation to clinicians at the Emergency Radiology Service of the HUG. A web–based graphical interface is available to submit visual and textual queries. The recognition rate is obtained with an experimental setup that is similar to actual clinical situations. By analyzing every slice of the image series, it minimizes the risk of missing of important lesions. Future work is required to reduce false detections of *micronodules* as well as to improve the precision of case–based retrieval.

6 Acknowledgments

References

1. Aziz, Z.A., Wells, A.U., Hansell, D.M., Bain, G.A., Copley, S.J., Desai, S.R., Ellis, S.M., Gleeson, F.V., Grubnic, S., Nicholson, A.G., Padley, S.P., Pointon, K.S., Reynolds, J.H., Robertson, R.J., Rubens, M.B.: HRCT diagnosis of diffuse parenchymal lung disease: inter–observer variation. Thorax **59**(6) (June 2004) 506–511
2. Sluimer, I.C., Schilham, A., Prokop, M., van Ginneken, B.: Computer analysis of computed tomography scans of the lung: a survey. Medical Imaging, IEEE Transactions on **25**(4) (April 2006) 385–405
3. Shyu, C.R., Brodley, C.E., Kak, A.C., Kosaka, A., Aisen, A.M., Broderick, L.S.: ASSERT: A physician–in–the–loop content–based retrieval system for HRCT image databases. Computer Vision and Image Understanding (special issue on content–based access for image and video libraries) **75**(1/2) (July/August 1999) 111–132

4. Uppaluri, R., Hoffman, E.A., Sonka, M., Hunninghake, G.W., McLennan, G.: Interstitial lung disease: A quantitative study using the adaptive multiple feature method. American Journal of Respiratory and Critical Care Medicine **159**(2) (February 1999) 519–525

5. Fetita, C., Chang-Chien, K.C., Brillet, P.Y., Prêteux, F., Grenier, P.: Diffuse parenchymal lung diseases: 3D automated detection in MDCT. In LNCS, S., ed.: Proceedings of MICCAI 2007. Volume 4791. (2007) 825–833

6. Webb, W.R., Müller, N.L., Naidich, D.P., eds.: High–Resolution CT of the Lung. Lippincott Williams & Wilkins, Philadelphia, PA, USA (2001)

7. Kim, N., Seo, J.B., Lee, Y., Lee, J.G., Kim, S.S., Kang, S.H.: Development of an automatic classification system for differentiation of obstructive lung disease using HRCT. Journal of Digital Imaging **22**(2) (April 2009) 136–148

8. Dundar, M., Fung, G., Bogoni, L., Macari, M., Megibow, A., Rao, B.: A methodology for training and validating a cad system and potential pitfalls. International Congress Series **1268** (2004) 1010–1014 CARS 2004 – Computer Assisted Radiology and Surgery. Proceedings of the 18th International Congress and Exhibition.

9. Müller, H., Clough, P., Hersh, W., Geissbuhler, A.: Variations of relevance assessments for medical image retrieval. In: Adaptive Multimedia Retrieval (AMR). Volume 4398 of Springer Lecture Notes in Computer Science (LNCS)., Geneva, Switzerland (2007) 233–247

10. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content–based image retrieval systems in medicine – clinical benefits and future directions. International Journal of Medical Informatics **73**(1) (2004) 1–23

11. Müller, H., Rosset, A., Garcia, A., Vallée, J.P., Geissbuhler, A.: Benefits from content–based visual data access in radiology. RadioGraphics **25**(3) (2005) 849–858

12. Caritá, E.C., Seraphim, E., Honda, M.O., Mazzoncini de Azevedo-Marques, P.: Implementation and evaluation of a medical image management system with content–based retrieval support. Radiologia Brasileira **41**(5) (October 2008) 331–336

13. Aisen, A.M., Broderick, L.S., Winer-Muram, H., Brodley, C.E., Kak, A.C., Pavlopoulou, C., Dy, J., Shyu, C.R., Marchiori, A.: Automated storage and retrieval of thin–section CT images to assist diagnosis: System description and preliminary assessment. Radiology **228**(1) (July 2003) 265–270

14. Unser, M.: Texture classification and segmentation using wavelet frames. IEEE Transactions on Image Processing **4**(11) (November 1995) 1549–1560

15. Van De Ville, D., Blu, T., Unser, M.: Isotropic polyharmonic B–Splines: Scaling functions and wavelets. IEEE Transactions on Image Processing **14**(11) (November 2005) 1798–1813

16. Depeursinge, A., Van De Ville, D., Unser, M., Müller, H.: Lung tissue analysis using isotropic polyharmonic B–spline wavelets. In: MICCAI 2008 Workshop on Pulmonary Image Analysis, New York, USA, Lulu (September 2008) 125–134

17. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content–based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(12) (2000) 1349–1380

18. Depeursinge, A., Racoceanu, D., Iavindrasana, J., Cohen, G., Platon, A., Poletti, P.A., Müller, H.: Fusing visual and clinical information for lung tissue classification in HRCT data. Journal of Artificial Intelligence in Medicine (2009–to appear)