Image–based diagnostic aid for interstitial lung disease with secondary data integration

Adrien Depeursinge[a], Henning Müller[a], Asmâa Hidki[a], Pierre–Alexandre Poletti[b], Alexandra Platon[b] and Antoine Geissbuhler[a]

[a]University & Hospitals of Geneva, Service of Medical Informatics, 24, rue Micheli–du–Crest, CH–1211 Geneva 14, Switzerland;
[b]University & Hospitals of Geneva, Service of Emergency Radiology, 24, rue Micheli–du–Crest, CH–1211 Geneva 14, Switzerland;

ABSTRACT

Interstitial lung diseases (ILDs) are a relatively heterogeneous group of around 150 illnesses with often very unspecific symptoms. The most complete imaging method for the characterisation of ILDs is the high-resolution computed tomography (HRCT) of the chest but a correct interpretation of these images is difficult even for specialists as many diseases are rare and thus little experience exists. Moreover, interpreting HRCT images requires knowledge of the context defined by clinical data of the studied case. A computerised diagnostic aid tool based on HRCT images with associated medical data to retrieve similar cases of ILDs from a dedicated database can bring quick and precious information for example for emergency radiologists. The experience from a pilot project highlighted the need for detailed database containing high-quality annotations in addition to clinical data.

The state of the art is studied to identify requirements for image–based diagnostic aid for interstitial lung disease with secondary data integration. The data acquisition steps are detailed. The selection of the most relevant clinical parameters is done in collaboration with lung specialists from current literature, along with knowledge bases of computer–based diagnostic decision support systems. In order to perform high–quality annotations of the interstitial lung tissue in the HRCT images an annotation software and its own file format is implemented for DICOM images. A multimedia database is implemented to store ILD cases with clinical data and annotated image series. Cases from the University & University Hospitals of Geneva (HUG) are retrospectively and prospectively collected to populate the database. Currently, 59 cases with certified diagnosis and their clinical parameters are stored in the database as well as 254 image series of which 26 have their regions of interest annotated.

The available data was used to test primary visual features for the classification of lung tissue patterns. These features show good discriminative properties for the separation of five classes of visual observations.

Keywords: Quantitative image analysis, database construction, content–based image retrieval, feature extraction, texture analysis, chest high–resolution CT, similar case retrieval.

1. INTRODUCTION

Interstitial Lung Diseases (ILDs) account for around 150 disorders of the lung tissue with often very unspecific symptoms. Many of the diseases are rare and establishing the differential diagnosis for ILD is considered difficult. A procedure for this is detailed in [1]. Physical examination of a patient affected by ILD is frequently abnormal but with unspecific findings. The first imaging examination used is the chest radiograph because of its low cost and weak radiation exposure. It also provides a quick overview of the whole chest. However, chest radiographs are normal in more than 10% of the patients with some forms of ILD and can provide a confident diagnosis in only 23% of the cases with lung diseases in general [2]. When first investigations on history, physical exam, routine labs and chest x–rays do not bring enough elements to fix a reliable diagnosis, a high–resolution computed

Further author information: (Send correspondence to Adrien Depeursinge)
Adrien Depeursinge: E–mail: adrien.depeursinge@sim.hcuge.ch, Telephone: +41 (0)22 372–8875

tomography (HRCT) of the chest can deliver more information. Indeed, original three–dimensional HRCT data avoid superposition of anatomic organs and provide an accurate assessment of the pattern and distribution of the lung tissue. It becomes an increasingly important method for the diagnosis of diffuse pulmonary parenchymal diseases. The diagnosis is strongly related to texture properties of the tissue.

Interpreting HRCT images of the chest represents a challenge even for trained chest radiologists and lung specialists. The three–dimensional form requires significant reading time, effort, and experience for a correct interpretation [3]. Moreover, the context is fundamental for correct interpretation: healthy tissue, for example, may have different visual aspects depending on the age or the smoking history of the patient [2,4]. This interpretation process is carried out by comparing a case with similar images in textbooks [5], which are most often organised by pathology. To do so, the radiologist must have a guess of the suspected disease present in the image and may miss the true pathology shown. Emergency radiologists have recourse to a large diversity of imaging modalities such as conventional projection radiography, computed tomography (CT), magnetic resonance imaging (MRI), functional imaging (fMRI, PET), and ultrasound applied to different organs such as the brain, colon, breast, chest, liver, kidney and the vascular and skeletal systems. They have to provide a first radiological report with ideas on the diagnosis quickly. This may result in errors by omission or confusion of diverse pathologic lung tissues [6,7].

Owing to the intrinsic complexity of the interpretation of HRCTs, a real–time image–based assistance appears useful for radiologists. A computerised diagnostic aid built on content–based image retrieval (CBIR) along with secondary data integration such as the relevant clinical parameters related to ILDs can bring quick and precious information to less experienced radiologists and non–chest experts. The information system will provide results in two steps. First the suspicious (abnormal) patterns in the new, non–interpreted HRCT are highlighted to the radiologist with a proposed tentative diagnostic. Then, the relevant patterns along with secondary data can be used as query for automatic retrieval of similar cases from an associated multimedia database of typical HRCT scans accompanied by corresponding clinical parameters of the patient [8]. The radiologist has to consider the system as a second opinion in addition to the experience gained with similar cases for providing a differential diagnosis. This differs distinctly from automated diagnosis aiming at replacing radiologists by computers [3]. By first automatically analysing the whole three–dimensional HRCT stack and showing suspicious patterns, the radiologist saves time usually devoted to browsing the entire series of images. Similar case retrieval allows not only the study of examples of the suspected disease but also cases with other diagnoses to avoid omission, confusions, and to be aware of the combination of the multiple ILD forms [9]. Another advantage of the system is to provide further examination and quantification of the evolution of the disease for the same patient by allowing access to further (anonymised) information on the case[10]. In order to store the cases to be retrieved, but also to train and evaluate the system, a dedicated database of ILD cases with HRCT containing annotations of the lung tissue patterns along with many relevant clinical parameters associated with the ILD case has to be built [6,11,12]. A well–annotated library of pulmonary HRCT cases constitutes a knowledge base offering possibilities for data mining and the perspective of sharing expertise of experienced chest radiologists and lung specialists with emergency radiologists or non–chest radiology fellows. The database can also be very useful for teaching [10,13] and research [14]. In [13], the content–based HRCT image retrieval system *ASSERT* [15] for interstitial lung disease is evaluated by counting the number of correct diagnoses established (score) with and without computerised assistance. It shows that a real–time image–based diagnostic aid improves the diagnostic quality significantly, especially when the radiologists are less experienced. Another study from the *Radiological Society of North America (RSNA)* based on receiver operating characteristics (ROC) curves brings the evidence of computer–aided diagnosis (CAD) to improve radiologists effectiveness [16].

In this paper, we describe the requirements and preliminary results for building an image–based diagnostic aid for interstitial lung disease with secondary data integration, along with the construction of a library of ILD cases containing high–quality visual annotations of the diverse lung tissue patterns as well as complete clinical data for each case. The project is named *TALISMAN* ("Texture Analysis of Lung ImageS for Medical diagnostic AssistaNce").

1.1. Image–based diagnostic aid systems and databases for ILDs

Content–based image retrieval and classification in medical applications has been a very active research domain over the last 15 years and is often proposed as very important for the future of medical data management and diagnostic aid [3,8,9]. One reason for this is the ever–increasing production of medical images in today's modern hospitals. For example, the service of radiology at the University Hospitals of Geneva (HUG) produces more than 50,000 images per day in 2006. However, very few systems are used in clinical practice to fully use this amount of data. The content–based chest HRCT retrieval project *ASSERT* [15] proposes a web–based "physician–in–the–loop" diagnostic aid tool and its evaluation in [13] has demonstrated promising results. The physician delineates a suspicious region of interest (ROI) in a bidimensional (one slice) 8–bit image in order to query the image retrieval system through visual features. In such a process, the physician still has to browse the whole HRCT stack and might miss important patterns. Additionally, 8–bit grey level images instead of initial 10–14 bit output of modern scanners do not contain all appropriate information to diagnose ILDs, which are characterised by tissue patterns with a wide range of values from –1000 Hounsfield Units (H.U.) for emphysematous changes to around 300 H.U. for calcified areas. Finally, one of the main features used to characterise the ROI is surprisely the area of the delineated region, which is usually not related to pathologic patterns for lung diseases.

In [17], a CBIR system aiming at providing similar images of lung diseases is described. A coordinate system independent of lung size and shape is defined to use spatial information along with texture features based on bidimensional fast Fourier transform (FFT) to characterise and classify the query ROI. Again, the user still has to browse whole HRCT stack to select suspicious slices to query the CBIR system.

A quantitative analysis of HRCT images of diffuse lung disease is described in [18]. 315 image stacks from 105 patients containing normal and six pathologic patterns have been quantified using six physical measures. The images are divided into contiguous 32×32 blocks from which the mean and standard deviation of CT value, air density components, nodular components, line components, and multilocular components (honeycombing) are computed and used to classify each block into one of the six classes corresponding to the different patterns. The effect of the size of the blocks is also studied. This study proposes a comprehensive approach for image–based diagnostic aid. However, diagnosing ILDs and even interpreting HRCT images cannot be accurately performed without any clinical parameters of the studied case [2].

Another approach for the characterisation of interstitial lung diseases is proposed in [19]. The airway tree is segmented and quantified, functional imaging is used to provide a measure of ventilation and perfusion of the lungs. Finally, the lung tissue is segmented and quantified using grey–level histograms along with a set of texture features called Adaptive Multiple Feature Method (AMFM) [20]. No interactions between the separated modules are mentioned, though. AMFM features are evaluated in [21].

A project that is more focused on the building and management of a chest CT database is described in [10]. A visual query–by–example Image Management Environment (IME) is developed in order to assist radiologists in interpreting chest CT images. The database is populated with 2000 diverse chest CT images from which only the most relevant slice of each series is kept. Among these, 600 images (30%) are pathologic. Similar images are retrieved through features uniquely based on edge extraction, which might not be efficient to detect diffuse patterns such as ground–glass opacities, for example. With the assumption that all images have approximately the same morphology, a virtual representation of the identified objects is used for indexing images in the database.

1.2. Performance of computerised diagnostic aid systems

Although almost all systems described above provide an estimation of their retrieval or classification efficiency, it remains very problematic to compare them objectively. Indeed, diversities of the assigned tasks and the database used render the comparison of the various systems very difficult. First efforts to palliate the lack of test collections with biomedical images have been initiated by the ImageCLEF project [22]. ImageCLEF is part of the Cross–Language Evaluation Forum (CLEF), a benchmark in in multilingual information retrieval. The ImageCLEFmed 2006 collection contains images obtained with several imaging modalities, various anatomic locations, diagnoses and findings. 30 visual or/and semantic tasks were defined and tested according to different retrieval approaches and techniques. With 13 research groups participating, the mean average precision (MAP) was used to measure the accuracy of the different systems and is calculated by taking the average accuracy of the retrieved items for a given topic, and then taking the mean over all the topics.

Efforts for building up a resource for the lung imaging research community are detailed in [6]. In order to test and develop lung CADs from reliable datasets for the detection of lung nodules on CT scans, the Lung Image

Database Consortium (LIDC) is constituted of five academic institutions from across the United States. The database will include healthy and pathologic CT images, annotations and primary clinical data of the patient. A review process of the annotations is carried out within the five institutions to solve the inter–radiologist variability in interpreting the images. It is not clear so far whether the database will be available to the research community and under which conditions.

1.3. Experiences from a pilot project

A pilot study of our approach for lung CT analysis and retrieval is described in [23]. Selected slices showing either healthy or one of the seven abnormal lung tissue patterns are classified through visual features with a Support Vector Machine (SVM) classifier. The image is divided into several blocks of size 16×16 from which the following visual features are extracted:

- average grey level and standard deviation of the grey levels,
- grey level histogram using 16 bins,
- features derived from co–occurrence matrices (four directions and two distances),
- responses of Gabor filters in four directions and at three scales,
- Tamura texture features.

This pilot project highlighted the need for a detailed database containing exact annotations in addition to clinical data [12] in order to perform multimodal classification and retrieval of three–dimensional HRCT images. Currently, few CADs are using visual features along with clinical parameters to perform multimodal retrieval.

2. MATERIAL & METHODS

The methods for building the image–based diagnostic aid tool are consituted by several connected subtasks. First, requirements were defined in collaboration with a lung specialist and the radiologists. The state of the art on image–based diagnostic aid was studied to identify challenges and solutions proposed. Clinical parameters associated with interstitial lung diseases were selected by a medical doctor of the project, a lung specialist, and an emergency radiologists. Based on this, a first version of a relational database for data acquisition was implemented in MySQL to start the collection of ILD cases through HTML forms and PHP scripts. Two HTML forms were created to fill the database. One is intended for clinical data and the other for image and ROI data. Each form executes MySQL queries through PHP scripts. A PHP reader for ROI files was implemented to automatically fill the database. This structure evolved based on the available clinical parameters in the electronic patient record and internal discussions. In parallel, the annotation process of images in the database were carried out with a visual interface implemented in Java. The interface was adapted in order to fit the needs of the radiologists for the various annotation tasks. Once a sufficient amount of annotated regions was reached, visual features were implemented in order to characterise the lung tissues patterns for the classification.

3. RESULTS

3.1. Data acquisition

3.1.1. Selection of the clinical parameters

Selection of clinical parameters was performed in two parts. A first step was to specify the scope of the diagnostic aid tool, determined by the selected subset of interstitial lung diseases included in the study. In collaboration with the pneumology and emergency radiology of the University Hospitals of Geneva (HUG), 15 different diseases that are described as the most frequent causes of lung parenchymal disorders [4] were selected. Based on each pathology, the most discriminative clinical parameters for the establishment of the differential diagnostic were kept. This selection process was carried out from the literature [1,5], along with knowledge bases of computer–based diagnostic decision support systems [24]. Discussions and remarks from lung specialists, radiologists and the medical informatics research group (SIM*) at the HUG allowed an iterative review of the selected parameters as well as standardised units and data formats to be used. Finally, the parameters that were not available from the electronic patient record were removed. 99 clinical parameters were selected to characterise the subgroup of ILDs. Terminology used in the HTML form corresponds to Medical Subject Headings (MeSH) and Systematised Nomenclature of Human Medicine (SNOMED).

3.1.2. Annotation process of regions of interest

The purpose of HRCT annotation is to establish the *ground truth* for lung tissue classification. In order to supervise the training of the system and to find effective visual features for the characterisation of the HRCT patterns, experienced radiologists delineate typical pathologic regions in HRCT scans.

The pilot study [23] highlighted the need for high–quality annotations. Indeed, since the annotations are intended for computerised pattern recognition, the regions of interest (ROIs) have to delineate pathologic patterns very precisely. Moreover, the perspective of further development of three–dimensional texture features requires the possibility to visualise and delineate ROIs in the entire HRCT volume. In addition, the possibility to set the window level used for displaying the 16–bit image on a computer screen was required. These specifications were integrated into an existing graphical software originally developed for delineating hepatic tumors in CT scans (see Figure 1). The radiologist opens an entire DICOM series and then draws precise ROIs in any layer of the CT volume in the axial view. Sagittal and coronal views are only available for visualisation purposes. Depending on the spacing between slices used, anisotropy in the vertical direction prevents from delineating ROIs in sagittal and coronal views. A file format was developed to save or load ROIs of the tool.

The CT images to be annotated were selected from cases with certified diagnoses. Depending on the disease textbooks [5] were consulted to identify the expected patterns. Patterns used to describe lung parenchymal disorders are not standardised among radiology communities. A detailed description of encountered patterns is given in [2]. In [15], patterns are classified into 4 perceptual categories. However, these categories may overlap in practice. Terminology used was mainly derived from [5] and is described in table 1. Examples of five patterns are shown in Figure 2. Localisation of the parenchymal disorders is relevant for several diseases. Thus, localisations of the ROIs are stored along with pattern labels in the ROI files. Table 2 lists the localisations used. All relevant documents available in the electronic patient record such as reports from radiology, pneumology, histology, biopsy and the final letter were anonymised and consulted for a better interpretation of the images. Only series with a layer thickness $< 3mm$ were annotated. Series with several co–morbidities, or with blur caused by breathing or movements of the patient or containing artifacts were not selected for annotations. Some images taken with contrast agent were annotated, but were left aside for further analysis. When possible, healthy tissue was delineated in the studied series to provide a wide range of the aspects of normal lung parenchyma.

Currently, 254 image series are selected and 26 annotated.

1	healthy
2	emphysema
3	bronchiectasis
4	cysts
5	pneumothorax
6	ground glass
7	consolidation
8	reticulation
9	fibrosis
10	increased attenuation
11	macronodules
12	micronodules

Table 1. Terminology for lung tissue patterns.

1	basal
2	apical
3	perihilar
4	sub pleural
5	diffuse
6	non relevant

Table 2. Localisation of the ROIs.

*http://www.sim.hcuge.ch/

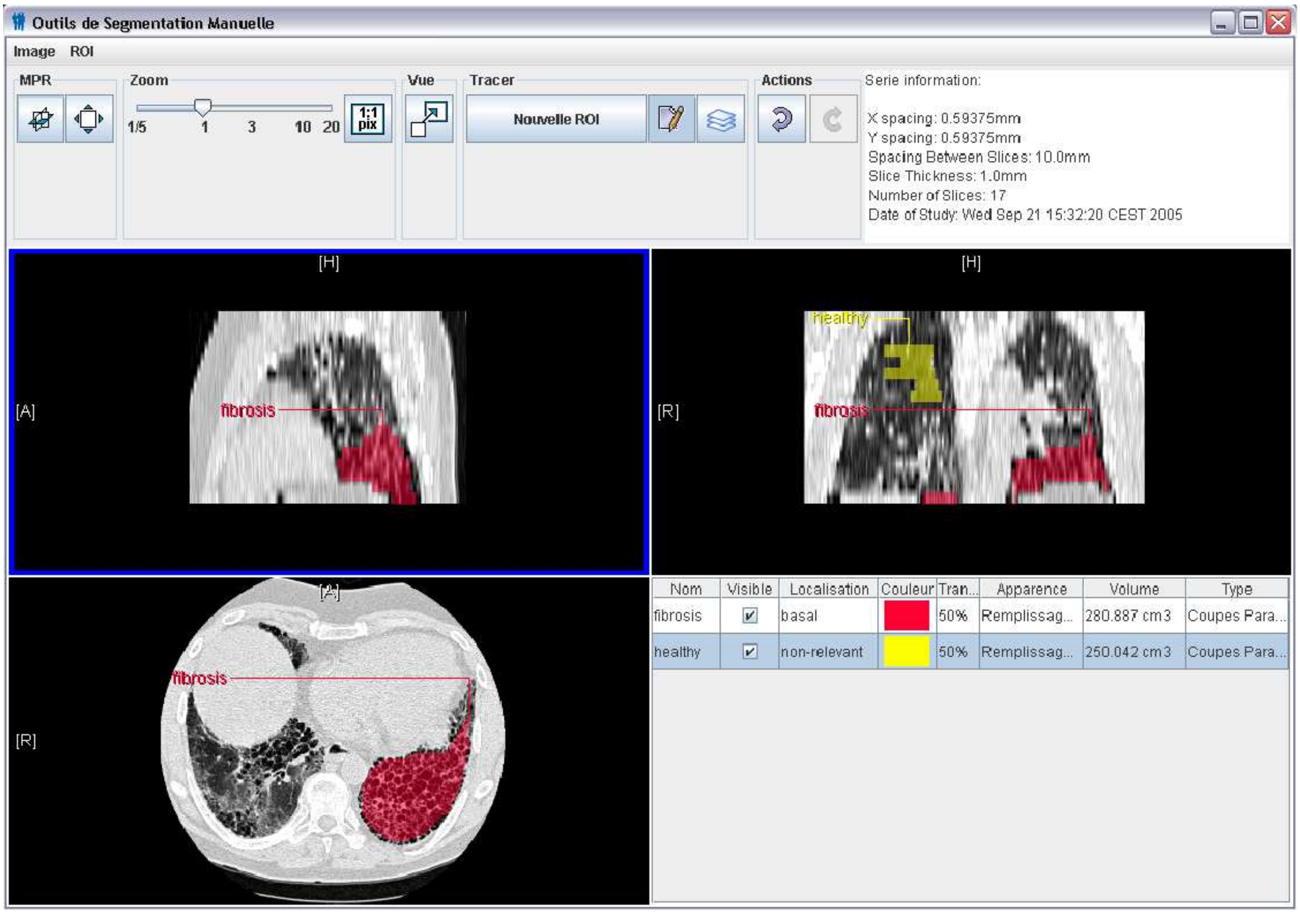

Figure 1. A screen shot of the graphical tool for the annotation of image regions.

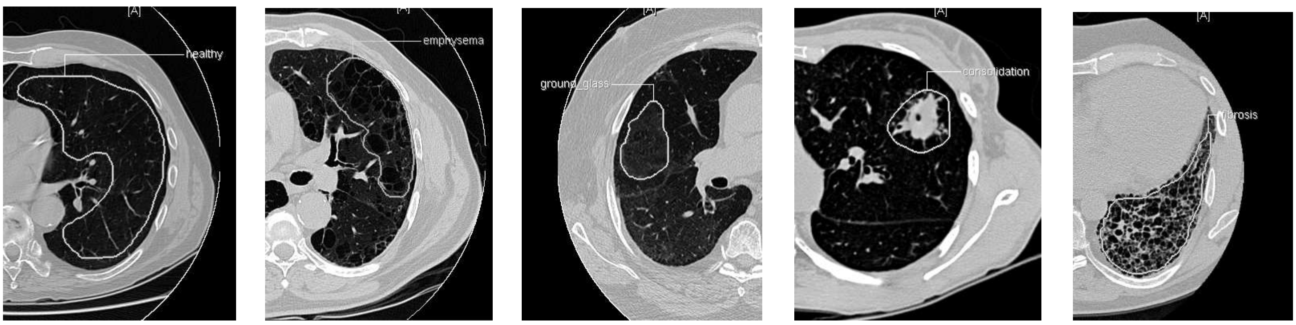

Figure 2. Examples of healthy tissue and four pathologic patterns.

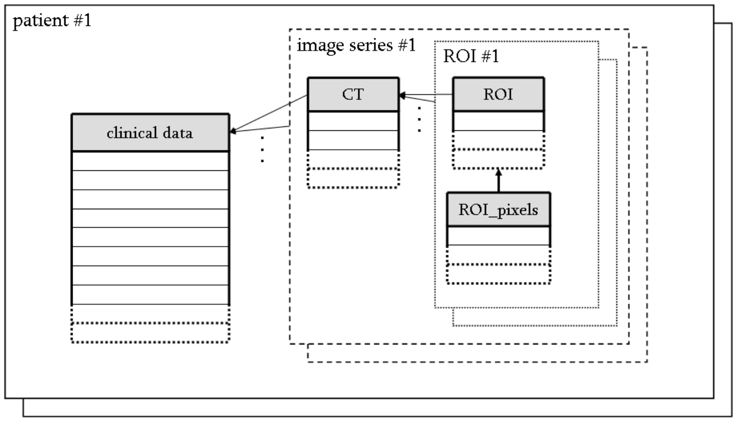

Figure 3. The architecture of the multimedia database for pulmonary CT case with clinical data.

3.1.3. Multimedia database architecture

First requirements for building a library of annotated pulmonary CT cases with clinical data integration are described in [12]. The MySQL database is constituted of four tables (see Figure 3). The main table contains all clinical data of the case. No confidential data is stored in the database, except the patient number, which requires authorised access to be used in the electronic patient record by an MD. For images, data, and settings, a second table named *CT* is linked to the main table to store image parameters such as *slice thickness* and the *date of the study.* One patient can have several image series. Linked to the *CT* table, the *ROI* table contains all information about ROIs such as *slice number*, *label* or *localisation.* Finally, the table *ROI_pixels* is linked to *ROI* and contains the (x, y) coordinates of the boundary of the ROIs. The database contains 166 fields among the four tables.

3.1.4. Data entry

An agreement from the ethical board of the hospitals was obtained to allow retrieving ILD cases from the patient record of the HUG. Cases with an ICD–10 code corresponding to one of the 15 pathologies described in [4] were retrospectively collected from the year 2003. Only cases with one or several HRCT exams were kept. Although ICD–10 diagnosis codes have been double–checked by professional coders since 2003, the diagnosis of several cases were completely correct. Our medical doctor along with the radiologists reviewed every diagnosis in the electronic patient record and conserved only typical cases of ILD. Cases from paediatrics were left aside. Remaining cases were subsequently stored in the database through the HTML forms. As often as possible pull–down menus were used for textual data. Units for laboratory tests and other numerical data were chosen depending on the formats used in the electronic patient record. The clinical parameters stored were those corresponding to the most acute manifestation of the disease. Currently, 59 cases with certified diagnoses and their clinical parameters are captured in the database.

3.2. Visual features

Since HRCT images are produced by measuring the attenuation of an x–ray beam passing through biological tissue, various visual aspects of lung tissue are mainly characterised by texture properties. Vocabulary used by radiologists to interpret patterns in HRCT images is mainly related to texture analysis and object recognition. Texture analysis is even more relevant for the characterisation of ILDs since many of them are diffuse. For example, emphysematous change is defined as the destruction of the lung parenchyma, which results in decreased attenuation in the HRCT image.

Texture analysis in digital image processing has been a very active research domain over more than thirty years [25]. This includes statistical approaches of autocorrelation function, spectral analysis, wavelet transforms [26], textural edginess, mathematical morphology, gray–level co–occurrences, run length, and autoregressive models. However, a first analysis of image patterns can be carried out by studying the grey–level distribution of image

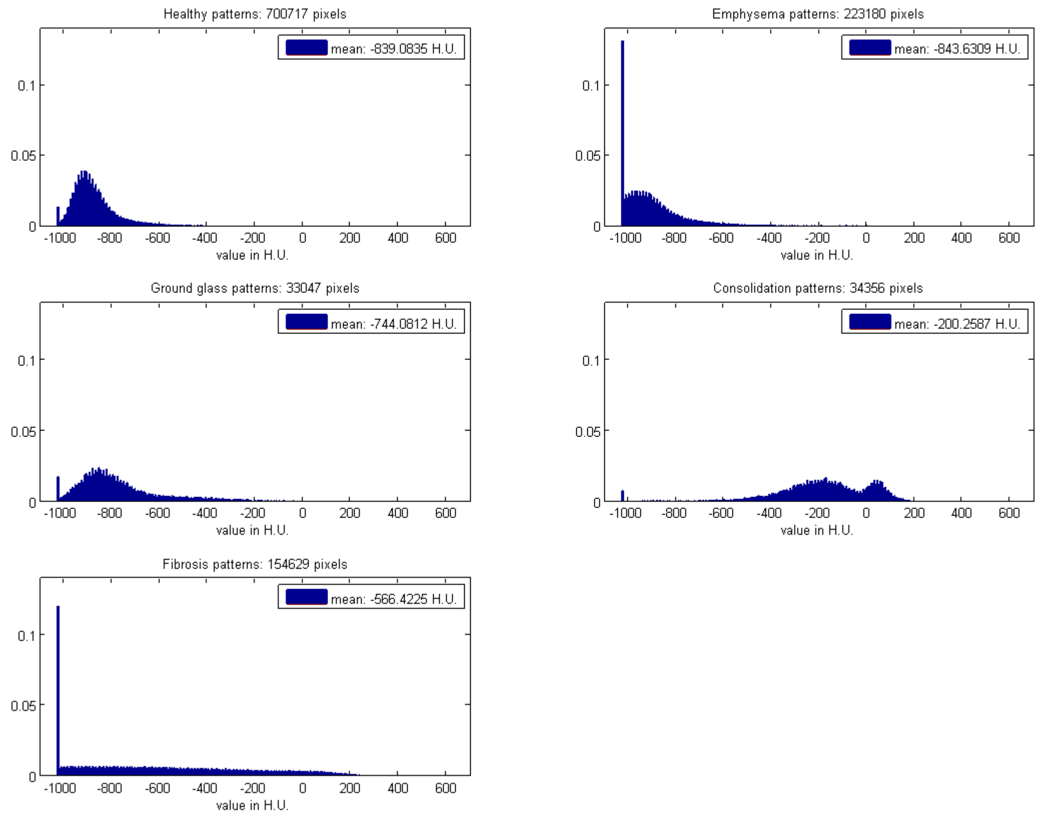

Figure 4. Histograms of patterns *healthy*, *emphysema*, *ground glass*, *consolidation* and *fibrosis* observed over all cases.

pixels. This can be done by building grey–level histograms from annotated ROIs. Full resolution (12–bit grey values) DICOM files are containing values in Hounsfield Units (H.U.) in the interval $[-2000; 2000]$. Theses values are corresponding univoquely to densities of the anatomic organs and thus allow the identification of lung tissue components. Spatial organisation of image pixels is neglected in this approach. A Fourier transform was applied to study spatial periodicity of tissue patterns.

3.2.1. Global grey–level histograms and air component

In order to study the composition of the lung tissue, grey–level histograms of the global ROI were built. Histograms of five patterns observed over all cases are shown in Figure 4. The mean, variance, skewness and kurtosis were extracted from the distribution as visual features. The air component of patterns (in %) was determined by computing the number of pixels with values inferior to -1000 H.U. divided by total number of pixels in the ROI.

Grey–level histograms in Figure 4 show high variability of distributions among the five patterns. Healthy patterns are constituted of soft tissues with H.U. values in the range $[-950; -700]$ and a weak air component. Emphysema is characterised by very low density of the tissue until destruction of the lung parenchyma, which results in a very high air component: more than 13% of the pixels are constituted by air. Distribution of the ground glass pixels is close to those of healthy pixels. However, the progressive increase of the opacity which characterises ground glass patterns is translated into an upper mean and lower kurtosis compared to a healthy pixel distribution. Consolidation pixels are corresponding to a very high mean (around -200) as well as a weak air component. At last, the distribution of fibrosis pixels shows a uniform dispersion of the tissue density (low kurtosis), with a strong air component. As a result, features such as mean and kurtosis of the distribution along with air component may have strong discriminative potential for the classification of the patterns.

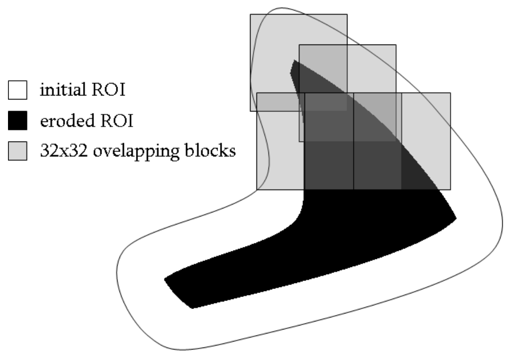

Figure 5. Division of ROIs into overlapping blocks for local analysis.

3.2.2. Local FFT and smoothness

ROIs were divided into overlapping 32×32 blocks in order to carry out local analysis. Overlapping was used to avoid omission of relevant patterns between two blocks. The size of the blocks (32×32) was chosen as a trade–off between being as local as possible and allowing detection of low frequencies of the patterns. Overlapping blocks are organised as follows: a binary image containing the initial ROI eroded with a 32×32 structuring element is horizontally visited. Each time the eroded ROI is reached, a block is extracted and then subtracted from the binary image. This procedure is repeated until the eroded ROI is completely deleted (see Figure 5).

The Fourier decomposition of each block was computed with an FFT to carry out a spectral analysis of the patterns. In order to be independent of the directionality, Fourier coefficients are averaged over same radius on the coefficient map. The mean Fourier coefficients are computed over all extracted blocks. The spectrum of five patterns are represented in Figure 6. Distribution of the coefficients look alike among patterns. Coefficients of fibrosis patterns have higher amplitudes compared to other patterns, which reflects the composition of fibrosis tissue characterised by quick and periodic transitions between high densities and air. Smoothness, definite by the amplitude of the zero–frequency coefficient was extracted as a visual feature.

3.2.3. Classification: preliminary results

First representations of distributions of the visual features are shown in Figures 7 and 8. Figure 7 shows the air component and mean values for each ROI from five patterns. These two visual features allow good separation of emphysema, fibrosis and consolidation patterns. Healthy and ground glass patterns are better separated using kurtosis and mean values of the distributions as shown in Figure 8. No classifier is implemented yet. The pilot study [23] showed promising results using a Support Vector Machine (SVM) classifier.

4. DISCUSSION

A comprehensive approach including the context of medical images is required to carry out qualitative analysis of HRCT images. Only very rarely radiologists can interpret images without taking into account the medical context. Integrating this context contained in the clinical data requires an environment with medical skills and needs access to the electronic patient record.

Since pathologic lung tissue patterns are constituted by varying changes of the lung parenchyma, a single kind of visual feature can not be effective to describe all patterns. For example, emphysema and fibrosis patterns are diffuse alterations of the lung parenchyma, which are strongly related to texture properties of the tissue. On the other hand, nodules and bronchiectasis are characterised by shapes. Thereby, a set of features based on texture and shape is mandatory for a correct assessment of various lung tissue patterns. Limits of such an image–based approach appears when lung tissue shows multiple overlayed patterns which might result in unknown visual aspects.

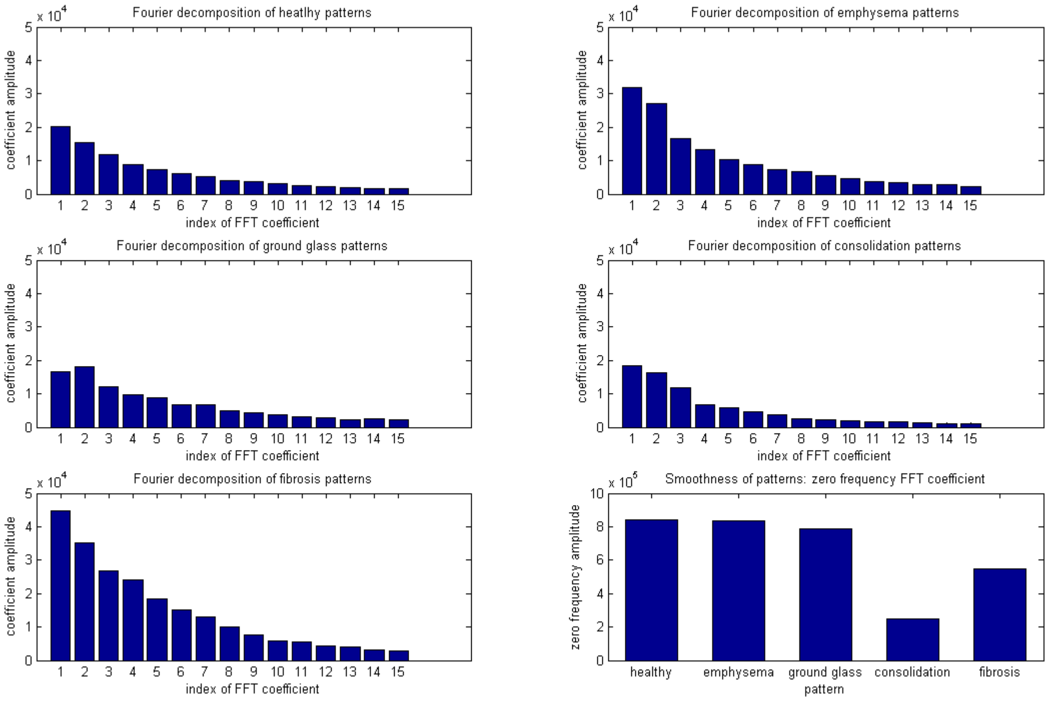

Figure 6. FFT transforms of patterns *healthy*, *emphysema*, *ground glass*, *consolidation* and *fibrosis*. The mean Fourier coefficients are computed over all extracted blocks from all cases. Smoothness of patterns definite by the amplitude of the zero–frequency coefficient are represented in the bottom right figure.

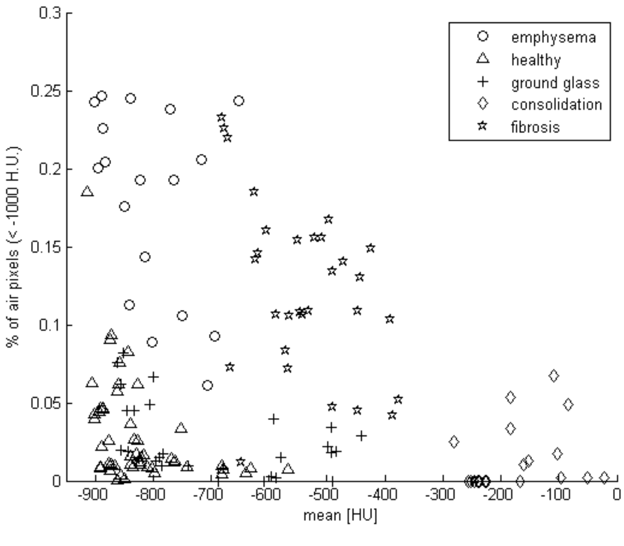

Figure 7. 2D plot of air components and mean values for each ROI from five patterns.

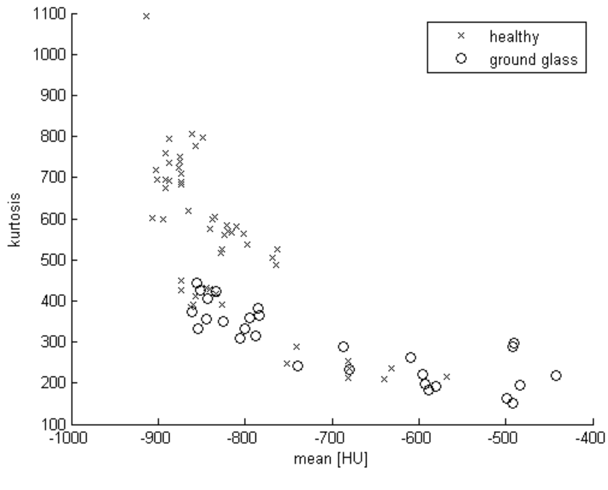

Figure 8. 2D plot of kurtosis and mean of distributions of ROIs from *healthy* and *ground glass* patterns.

5. CONCLUSION & FUTURE WORK

This paper describes an overview of requirements for building an image–based diagnostic aid with secondary data integration. The use of clinical parameters along with HRCT images is in accordance with the usual procedure for diagnosing interstitial lung diseases. The built multimedia database provides ground truth to train and validate the diagnostic aid system. First collected data was used to test simple visual features for further classification of lung tissue patterns. These features showed good discriminative properties for the separation of five patterns.

There still remains an important need for a specialised visual feature set to describe pathologic tissue patterns of the lung. Moreover, a classification framework has to be developed to allow automatic characterisation of unknown HRCT images. A user-friendly graphical interface is required for an optimized use of the system in clinical practice. Our plan is to develop 3D texture features to allow for a better characterisation of the tissue and use SVMs to perform the classification. Another goal is to integrate the clinical parameters acquired for the database to allow for the retrieval of similar cases as a first help for the emergency radiologists. The framework will integrate a graphical interface to show suspicious patterns to radiologists and to retrieve similar cases from the multimedia database.

6. ACKNOWLEDGEMENTS

This work was partially supported by the Swiss National Science Foundation (FNS) with grant 205321–109.

REFERENCES

1. T. King, "Approach to the adult with interstitial lung diseases," *UpToDate* **December**, 2005.
2. P. Stark, "High resolution computed tomography of the lungs," *UpToDate* **February**, 2005.
3. K. Doi, "Current status and future potential of computer–aided diagnosis in medical imaging," *British Journal of Radiology* **78**, pp. 3–19, 2005.
4. A. Hidki, H. Müller, A. Depeursinge, P.-A. Poletti, and A. Geissbuhler, "Putting the image into perspective: The need for domain knowledge when performing image–based diagnostic aid," in *Swiss conference on medical informatics (SSIM 2006)*, (Basel, Switzerland), April 2006.
5. W. R. Webb, N. L. Müller, and D. P. Naidich, eds., *High-Resolution CT of the Lung*, Lippincott Williams & Wilkins, Philadelphia, PA, USA, 2001.
6. S. G. Armato, G. McLennan, M. F. McNitt-Gray, and et. al., "Lung image database consortium: Developing a resource for the medical imaging research community," *Radiology* **232**, pp. 739–748, 2004.
7. R. Kakinuma, H. Ohmatsu, M. Kaneko, K. Eguchi, T. Naruke, K. Nagai, Y. Nishiwaki, A. Suzuki, and N. Moriyama, "Detection failures in spiral ct screening for lung cancer: Analysis of ct findings," *Radiology* **212**, pp. 61–66, July 1999.
8. U. Sinha, A. Bui, R. Taira, J. Dionisio, C. Morioka, D. Johnson, and H. Kangarloo, "A review of medical imaging informatics," *Annals of the New York Acedemy of Sciences* **980**, pp. 168–197, December 2002.
9. H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content–based image retrieval systems in medicine – clinical benefits and future directions," *International Journal of Medical Informatics* **73**, pp. 1–23, 2004.
10. G. Sasso, H. R. Marsiglia, F. Pigatto, A. Basilicata, M. Gargiulo, A. F. Abate, M. Nappi, J. Pulley, and F. S. Sasso, "A visual query–by–example image database fro chest ct images: Potential role as a decision and educational support tool for radiologists," *Journal of Digital Imaging* **18**, pp. 78–84, March 2005.
11. H. J. Lowe, B. G. Buchanan, G. F. Cooper, and J. K. Vreis, "Building a medical multimedia database system to integrate clinical information: An application of high–performance computing and communications technology," *Bulletin of the Medical Library Association* **83**(1), pp. 57–64, 1995.
12. A. Depeursinge, H. Müller, A. Hidki, P.-A. Poletti, T. Rochat, and A. Geissbuhler, "Building a library of annotated pulmonary CT cases for diagnostic aid," in *Swiss conference on medical informatics (SSIM 2006)*, (Basel, Switzerland), April 2006.
13. A. M. Aisen, L. S. Broderick, H. Winer-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, C.-R. Shyu, and A. Marchiori, "Automated storage and retrieval of thin–section CT images to assist diagnosis: System description and preliminary assessment," *Radiology* **228**, pp. 265–270, 2003.

14. A. Horsch, M. Prinz, S. Schneider, O. Sipilä, K. Spinnler, H.-P. Vallée, V. de Leeuw, R. Vogl, T. Wittenberg, and G. Zahlmann, "Establishing an international reference image database for research and development in medical image processing," in *Bildverarbeitung in der Medizin*, pp. 409–412, (Erlangen, Germany), March 2003.
15. C.-R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick, "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**, pp. 111–132, July/August 1999.
16. H. Abe, H. MacMahon, R. Engelmann, Q. Li, J. Shiraishi, S. Katsuragawa, M. Aoyama, T. Ishida, K. Ashizawa, C. E. Metz, and K. Doi, "Computer–aided diagnosis in chest radiography: Results of large–scale observer tests at the 1996–2001 RSNA scientific assemblies," *RadioGraphics* **23**(1), pp. 255–265, 2003.
17. C.-T. Liu, P.-L. Tai, A. Y.-J. Chen, C.-H. Peng, T. Lee, and J.-S. Wang, "A content–based CT lung retrieval system for assisting differential diagnosis image collection," in *Proceedings of the second International Conference on Multimedia and Exposition (ICME'2001)*, pp. 241–244, IEEE Computer Society, IEEE Computer Society, (Tokyo, Japan), August 2001.
18. Y. Uchiyama, S. Katsuragawa, H. Abe, J. Shiraishi, F. Li, Q. Li, C.-T. Zhang, K. Suzuki, and K. Doi, "Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography," *Medical Physics* **30**, pp. 2440–2454, September 2003.
19. E. A. Hoffman, J. M. Reinhardt, M. Sonka, B. A. Simon, J. Guo, O. Saba, D. Chon, S. Samrah, H. Shikata, J. Tschirren, K. Palagyi, K. C. Beck, and G. McLennan, "Characterization of the interstitial lung disease via density-based and texture-based analysis of computed tomography images of lung structure and function," *Academic Radiology* **10**, pp. 1104–1118, October 2003.
20. R. Uppaluri, E. A. Hoffman, M. Sonka, G. W. Hunninghake, and G. McLennan, "Interstitial lung disease: A quantitative study using the adaptive multiple feature method," *American Journal of Respiratory and Critical Care Medicine* **159**, pp. 519–525, February 1999.
21. R. Uppaluri, E. A. Hoffman, M. Sonka, P. G. Hartley, G. W. Hunninghake, and G. McLennan, "Computer recognition of regional lung disease patterns," *American Journal of Respiratory and Critical Care Medicine* **160**, pp. 648–654, August 1999.
22. W. Hersh, J. Jensen, H. Müller, P. Gorman, and P. Ruch, "Trans–atlantic collaboration for evaluating image retrieval systems in the ImageCLEF biomedical image retrieval task," in *eChallenges 2005*, (Ljubljana, Slovenia), October 2005.
23. H. Müller, S. Marquis, G. Cohen, and A. Geissbuhler, "Lung CT analysis and retrieval as a diagnostic aid," *Studies in Health Technology and Informatics* **116**, pp. 453–458, 2005.
24. C. P. Friedman, A. S. Elstein, F. M. Wolf, G. C. Murphy, T. M. Franz, P. S. Heckerling, P. L. Fine, T. M. Miller, and V. Abraham, "Enhancement of clinician's diagnostic reasoning by computer-based consultation," *Journal of the American Medical Association* **282**, pp. 1851–1856, November 1999.
25. R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE* **67**, May 1979.
26. M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Transactions on Image Processing* **4**(11), pp. 1549–1560, 1995.