Lung Texture Classification Using Bag of Visual Words

Marina Asherov, Idit Diamant, Hayit Greenspan
Department of Biomedical Engineering, Tel-Aviv University, Tel-Aviv, Israel

ABSTRACT

Interstitial lung diseases (ILD) refer to a group of more than 150 parenchymal lung disorders. High-Resolution Computed Tomography (HRCT) is the most essential imaging modality of ILD diagnosis. Nonetheless, classification of various lung tissue patterns caused by ILD is still regarded as a challenging task. The current study focuses on the classification of five most common categories of lung tissues of ILD in HRCT images: normal, emphysema, ground glass, fibrosis and micronodules. The objective of the research is to classify an expert-given annotated region of interest (AROI) using a bag of visual words (BoVW) framework. The images are divided into small patches and a collection of representative patches are defined as visual words. This procedure, termed dictionary construction, is performed for each individual lung texture category. The assumption is that different lung textures are represented by a different visual word distribution. The classification is performed using an SVM classifier with histogram intersection kernel. In the experiments, we use a dataset of 1018 AROIs from 95 patients. Classification using a leave-one-patient-out cross validation (LOPO CV) is used. Current classification accuracy obtained is close to 80%.

Keywords: Visual words, Image classification**,** Interstitial Lung Diseases, High-Resolution Computed Tomography.

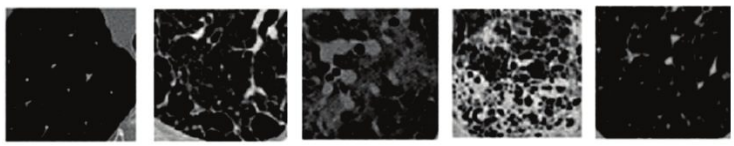
1. INTRODUCTION

Interstitial lung diseases (ILD) represent a heterogeneous group of more than 150 disorders of the lung parenchyma[1]. These disorders are characterized by alternation of the lung parenchyma resulting in breathing dysfunction. Correct clinical diagnosis of the specific type of ILD is essential to determine the appropriate treatment and prognosis. In order to obtain confident diagnosis information needs to be integrated from a variety of sources, such as clinical history, physical examination, measurements of lung functions, appropriate blood test, thoracic imaging and sometimes invasive procedures. Currently the most crucial imaging modality in ILD diagnosis is the High-Resolution Computed Tomography (HRCT)[2,3]. This technique enables to identify different lung tissue patterns. The appearance and quantification of these textures are essential to characterize the specific type of ILD. Since the disorders are rare, ILD diagnosis using HRCT images is a challenging task even for trained radiologists. Moreover, ILD have low inter-class distinction and high intra-class variation which make this task more challenging. Therefore, there is a need to develop an automatic system for categorization of lung tissue patterns to support more accurate and fast diagnosis.

Image classification is usually performed in two main parts: feature extraction and supervised classification using machine learning techniques. The first step is to extract salient features that are encoded as feature descriptors to represent the image. A variety of features to describe different lung tissue patterns associated with ILDs have been suggested: gray-level histogram[4,7] , gray-level co-occurrence matrices (GLCM)[5,6], run length (RLE)[5,6], filter bank (Gaussian, Gabor, wavelets)[8,9,10], shape and spatial content[12] , scale-invariant feature transform (SIFT)[13] , local binary patterns (LBP)[7,11] , and histogram of oriented gradients (HOG)[7]. The next step is to perform labeling of these descriptors for image classification using supervised techniques, such as k-nearest neighbor (kNN)[4,10] , support vector machine (SVM)[4,8,9] , Bayesian classifiers[4], artificial neural network (ANN)[4], decision trees[4], linear discriminant analysis (LDA)[13], and sparse representation methods[7]. A variety of algorithms have been developed to classify lung tissue types in CT images. For example, Depeursinge et al.[4] compared five common classifiers - naive Bayes, k-NN, J48 decision trees, multilayer perceptron, and SVM using gray-level histograms, air components and quincunx wavelet frame coefficients with B-spline wavelets as features. Song et al.[7] presented two novel feature descriptors combined with gray-level histogram - rotation-invariant Gabor-LBP (which integrates multi-scale Gabor filters and LBP histograms) and a multi-coordinate HOG descriptor (which is based on extraction of gradient features while accommodating rotation variance with radial-specific coordinate systems).

Medical Imaging 2014: Computer-Aided Diagnosis, edited by Stephen Aylward, Lubomir M. Hadjiiski,
Proc. of SPIE Vol. 9035, 90352K · © 2014 SPIE · CCC code: 1605-7422/14/$18 · doi: 10.1117/12.2044162

A comparison across works is difficult due to the fact that each work uses a different dataset of different sizes and categories. Additionally, some of the works focus on classifying a given annotated region of interest (AROI) while others classify smaller constant size regions (e.g. ROIs of 32X32 pixels) which were extracted from the AROIs. In the current work we focus on using the bag-of-visual-words (BoVW) model for the classification task. The objective of our study is to evaluate the implementation of BoVW algorithm for classification of five most widespread categories of lung tissues: normal, emphysema, ground glass, fibrosis and micronodules (see Figure 1 for examples). Classification is applied on annotated region of interest (AROI) manually drawn by a radiologist.



(a) (b) (c) (d) (e)

Figure 1. Examples of lung textures: (a) normal, (b) emphysema, (c) ground glass, (d) fibrosis, (e) micronodules.

2. METHODS

2.1 Dataset

We use a database of Interstitial Lung Diseases (ILD) cases[14] with HRCT images series with a slice thickness of 1mm, inter–slice distance of 10mm and in-plane resolution of 512x512 pixels. This database is comprised of 2-D hand-drawn regions annotated by two radiologists with 15 and 20 years of experience, that are used as ground truth. Each AROI is labeled as one texture category (normal, emphysema, ground glass, fibrosis and micronodules). In our work, HRCT image series of 91 patients with 1018 AROIs are used (this set contains all images in the database with AROIs above 1000 pixels). The distribution of the lung texture classes are detailed in Table 1.

Table 1. Data distribution of each tissue category

	Normal (N)	***Emphysema (E)***	***Ground Glass (G)***	***Fibrosis (F)***	***Micronodules (M)***	***Total***
Patients	14	9	27	27	14	91
AROIs	168	100	247	267	236	1018

2.2 BoVW algorithm

Image representation using bag-of-visual-words (BoVW) model is adapted to the visual analysis field from the bag-of-words (BoW) representation of text documents[15]. Using the BOW model, a document is represented as a word distribution using a frequency occurrence word histogram. When representing an image using BOVW, we treat it as a document and therefore represent it as a distribution of visual elements. Thus we first find these visual elements, discretize their space and create a visual word dictionary. Then, using this dictionary, we can represent an image as a visual word histogram, i.e., the frequency of occurrence of visual words that appear in the image.

We implemented the BoVW model as follows, generating feature descriptor for each AROI, while dividing the experiment to training and testing phases (illustrated in Figure 2):

a. Patches extraction:
   Square image patches are extracted from each AROI (as shown in Figure 3). The patches are extracted from every third pixel in the image, which was found out to be the optimized pixel step size in terms of classification accuracy and computational complexity. Each image patch is represented by a feature vector of the pixels gray-level intensity.

A Principal Component Analysis (PCA) is applied on feature vectors to reduce the computational complexity of the algorithm.

b. Creating a texture dictionary:
This step is performed only at the training phase. For each tissue category a texture dictionary is built. The vector-represented patches are converted into visual words and a representative texture dictionary is generated using K-means clustering algorithm. The output of K-means algorithm is the K most significant feature vectors (or visual words) representing each category. Then the dictionaries of all categories are concentrated to a single global dictionary. Since our dataset is unbalanced, we chose to build a different dictionary for each category with the same amount of words to make sure that each category will be represented well with no bias to the largest category.

c. Quantization step:
Each image patch is categorized to the closest visual word in the dictionary. The quantization is based on Euclidian distance.

d. Creating texture word histogram:
In order to represent the whole annotated ROI, a normalized texture words histogram is generated, using the global texture dictionary, for every training and testing AROI (as shown in Figure 4). The histogram represents the distribution of visual words in an AROI.

e. Classification:
Each test AROI is classified into one of five tissue categories using a nonlinear multi-class support vector machine (SVM) classifier with histogram intersection kernel:

$$K(x, y) = \exp\left[-\sum_i \min(x_i, y_i)\right] \tag{1}$$

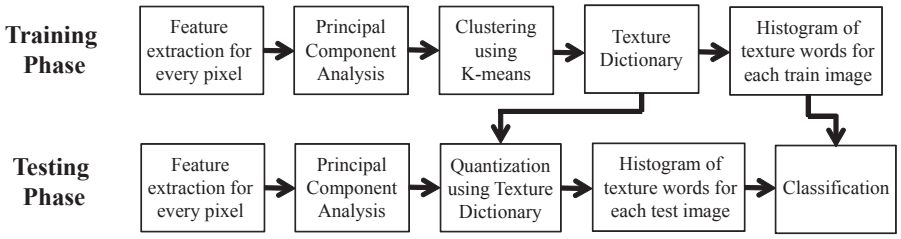
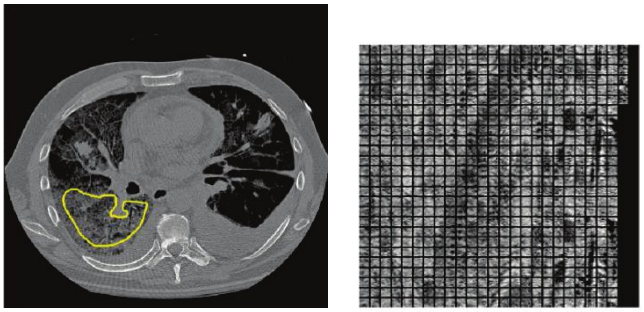


Figure 2. Classification using bag of visual words - training and testing phases



(a) (b)

Figure 3. (a) A sample of a HRCT image slice with AROI (yellow). (b) Extracted patches that are completely enclosed in the AROI.

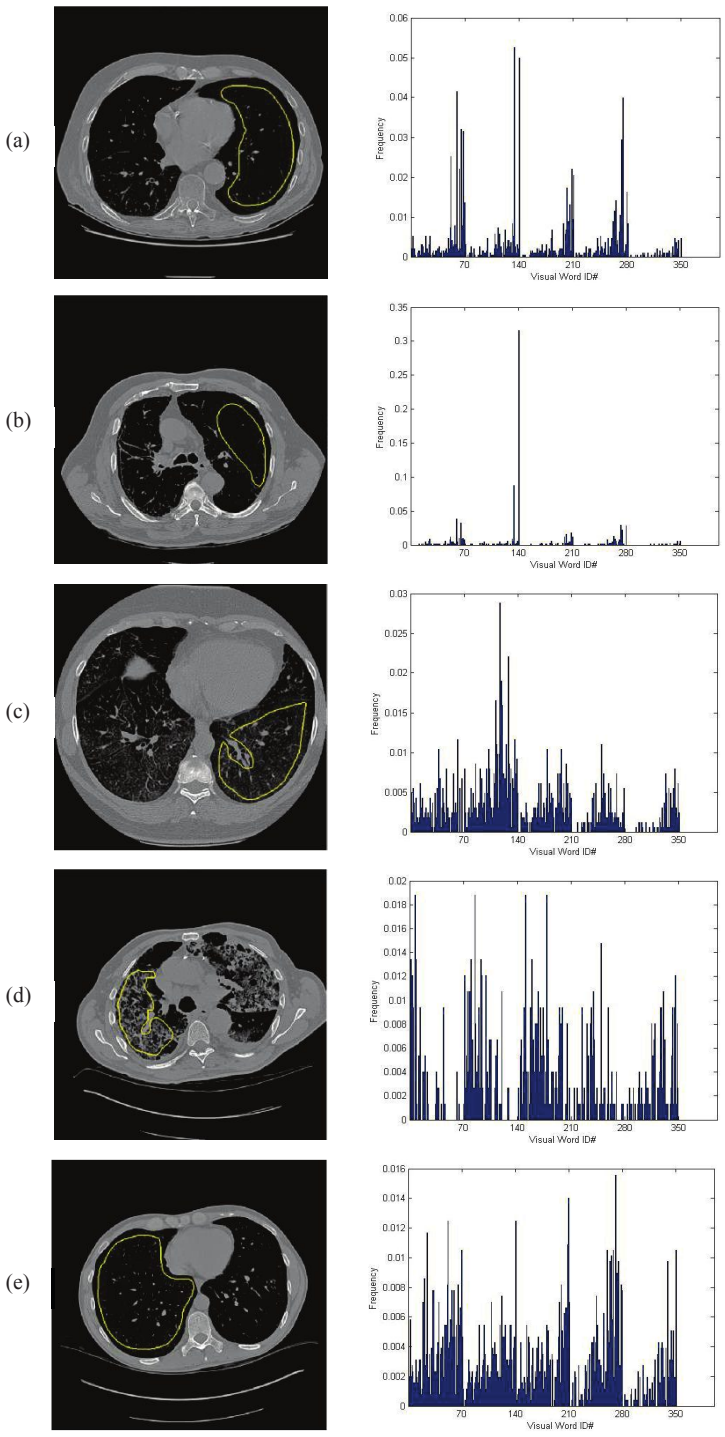


Figure 4. Examples of HRCT image slices with AROI (left) and the appropriate normalized texture words histogram (right) for each texture category: (a) normal, (b) emphysema, (c) ground glass, (d) fibrosis, (e) micronodules. The main assumption is that different textures are characterized by a different distribution of visual words in the AROI.

3. EXPERIMENTS AND RESULTS

3.1 Algorithm performance evaluation

We evaluated the algorithm performance using recall, precision, F-score and accuracy measures according to the following equations:

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{2}$$

$$\mathrm{Pr}ecision = \frac{TP}{TP + FP} \tag{3}$$

$$F - score = \frac{2TP}{2TP + FN + FP} \tag{4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where TP, FN, and FP are the values of true positive, false negative and false positive classifications of texture types.

In order to estimate the generalization performance of the AROI classification, leave-one-patient-out cross validation (LOPO CV) was used. We experimented with various parameter values to optimize performance: Size of dictionary per category (hereon termed sub-dictionary) was tested in the range of 5 to 100 words, Several visual word sizes (number of principal components) were tested (5, 10, 15, 20), as well as several patch sizes (7x7, 9x9, 11x11). The classification accuracy results are shown in Figure 5(a)-(c). It can be seen that when using varying sub-dictionary sizes from 30-80 words the highest accuracy rate is achieved for word size 10 (especially for 7x7 and 11x11 patches). Figure 5(d) demonstrates the patch size value influence on the performance when using word size 10. It can be seen that for each dictionary size the best performance was achieved for a patch size of 7X7 pixels.

We obtained 79% classification accuracy using a 7x7 patch size, visual word size of 10 and sub-dictionary size of 70 visual words per category (total dictionary size is 350 words). Tables 2 and 3 present the AROI classification results. The global dictionary that is created for this optimal set of parameter values is presented in Figure 6.

Table 2. Results of the AROI classification

	N	**E**	**G**	**F**	**M**
Recall [%]	88.7	65.0	75.3	88.0	71.2
Precision [%]	76.0	94.2	76.5	77.0	82.0
F-score [%]	81.9	76.9	75.9	82.2	76.2
Accuracy [%]	79.0				

Table 3. Confusion matrix of AROI classification

		Prediction [%]				
		N	**E**	**G**	**F**	**M**
Ground Truth	**N**	89	1	5	1	5
	E	12	65	4	12	7
	G	8	0	75	13	4
	F	0	1	7	88	4
	M	7	0	11	11	71

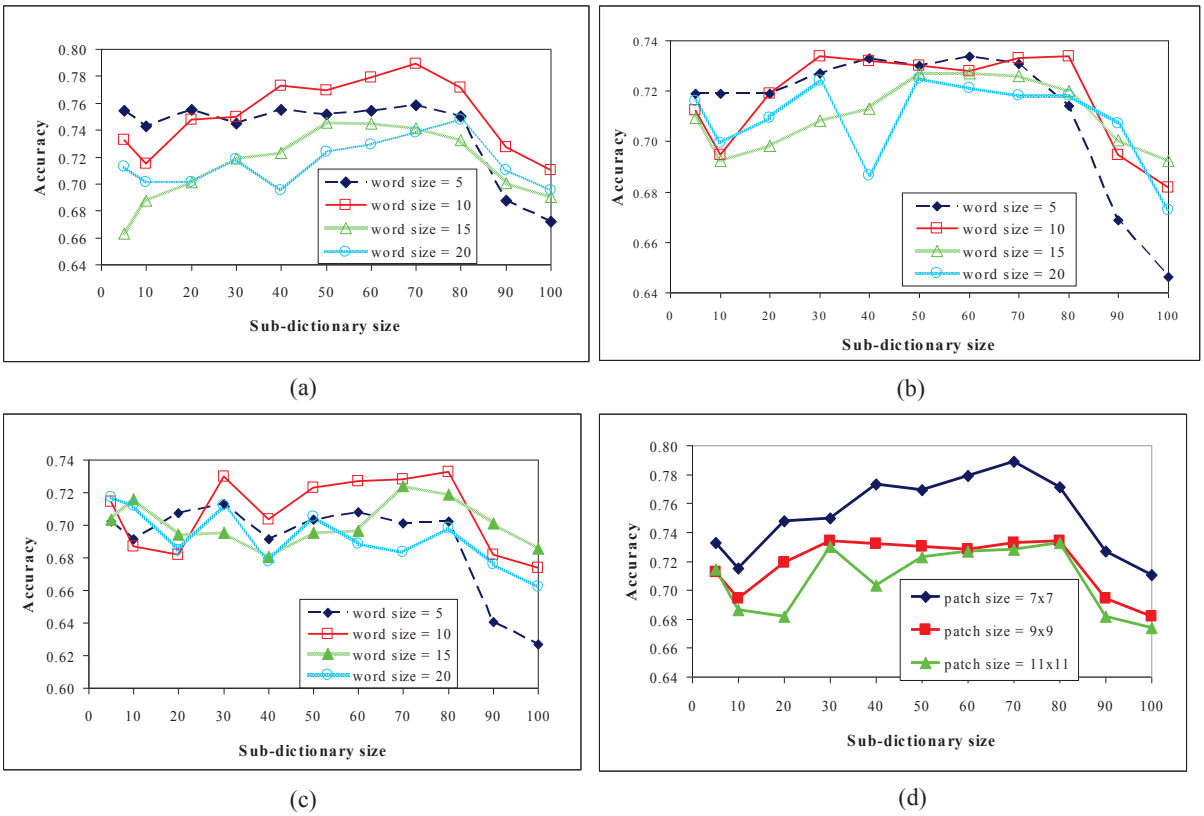


Figure 5. (a)-(c) Classification performance analysis evaluation using 7x7, 9x9 and 11x11 patch sizes respectively, for various word and sub-dictionary sizes. (d) Classification performance analysis evaluation using word size 10 for various patch and sub-dictionary sizes.

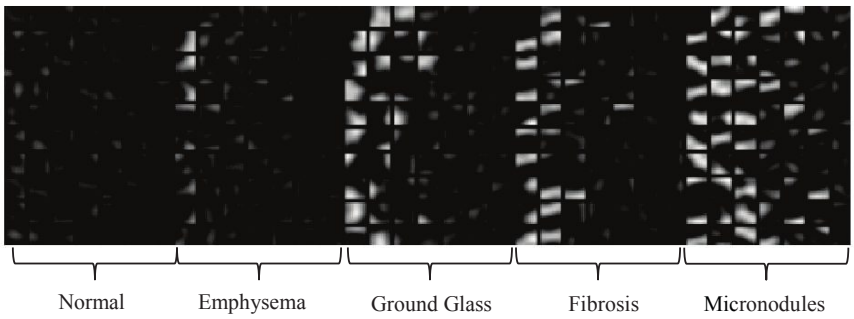


Figure 6. Texture dictionary is built by concatenating sub-dictionaries for every category.

3.2 Comparison to previous work

We compared our classification performance to a recent work in the literature that deals with ILD tissue classification on the AROI level [7]. Results are shown in Table 6.

Table 6. Comparison of ILD tissue classification performance to state-of-the-art for input of AROIs

Reference	Data amount	[%]	N	E	G	F	M
Song et al.[7]	1458 AROIs from 95 image sets (dataset divided to 3 parts)	Recall	92.4	77.1	67.7	73.4	76.1
		Precision	40.4	82.7	79.4	93.3	92.3
		F-score	56.2	79.8	73.1	82.2	83.4
Our Results	1018 AROIs from 91 image sets	Recall	88.7	65.0	75.3	88.0	71.2
		Precision	76.0	94.2	76.5	77.0	82.0
		F-score	81.9	76.9	75.9	82.2	76.2

4. DISCUSSION AND CONCLUSIONS

In our study we evaluated the performance of BoVW approach for classification of five most common lung tissue types in HRCT images of healthy tissue and four patterns associated with ILDs: emphysema, ground glass, fibrosis and micronodules. The AROI is divided into small image patches that are used to build a dictionary of representative patches, defined as visual words. In order to represent the entire AROI, a visual word histogram is created. The classification of histogram of an AROI is performed using a SVM classifier with histogram intersection kernel. The system classification performance was evaluated using various algorithm parameters values. The best results of 79% accuracy rate were achieved using 7x7 patch size, visual word size of 10 and total dictionary size of 350 words.

A variety of algorithms have been developed to classify lung tissue types in CT images. However, it is difficult to have a fair comparison between the algorithms since each work has been applied on different datasets of different sizes and categories. Additionally, some of the works classified an annotated region of interest (AROI) and some smaller constant size regions such as ROIs of 32X32 pixels which were extracted from the AROIs. We chose to compare our results to a state-of-the-art method[7] that performed lung texture classification in the AROI-level as was done in our work using the same database and categories (although we used 1018 AROIs from 91 patients while Song et al.[7] used 1458 AROIs from 95 patients). We have demonstrated overall similar classification results. Our method performed well for ground glass and fibrosis categories in terms of recall values and for normal tissue in terms of precision and F-score values. It is important to note that Song et al.[7] used texture, intensity and gradient features, while in the current work we use gray-level intensity only.

In future work, we will augment the feature representation to include additional features, such as a Gabor filter bank, to explore if this can augment the results. In addition, we will evaluate the impact of classifying a fixed size (small) ROIs instead of an AROI with varying sizes. We will also investigate the influence of voting of classified ROIs in order to determine each AROI category using its interior ROIs.

REFERENCES

[1] Baughman, R.P, du Bois, R. M. (Eds.), [Diffuse Lung Disease: A Practical Approach], 2nd ed., Springer (2012).
[2] Ryu, J. H.., Olson, E. J., Midthun, D. E. and Swensen, S. J., "Diagnostic Approach to the Patient with Diffuse Lung Disease" Mayo Clin Proc., 77(11), 1221-1227 (2002).
[3] King, T.E, "Clinical advances in the diagnosis and therapy of the interstitial lung diseases", Am J Respir Crit Care Med, 72(3), 268-79 (2005).
[4] Depeursinge, A., Iavindrasana, J., Hidki, A., Cohen, G., Geissbuhler, A., Platon, A., Poletti P. A. and Müller, H., "Comparative Performance Analysis of State-of-the-Art Classification Algorithms Applied to Lung Tissue Categorization", J. Digital Imaging, 23(1), 18-30 (2010).
[5] Xu, Y., Sonka, M., McLennan, G., Guo, J. and Hoffman, E. A., "MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies," IEEE Trans Med Imaging, 25(4), 464–475 (2006).
[6] Yao, J., Dwyer, A., Summers, R. M. and Mollura, D. J., "Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification," Acad. Radiol., 18 (3), 306–314 (2011).
[7] Song, Y., Cai, W., Zhou, Y. and Feng, D. D., "Feature-based image patch approximation for lung tissue classification". IEEE Trans Med Imaging. 32(4), 797-808 (2013).
[8] Depeursinge, A., de Ville, D. V., Platon, A., Geissbuhler, A., Poletti, P. A., and Muller, H., "Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames," IEEE Trans Inf Technol Biomed, 16(4), 665-675 (2012).
[9] Depeursinge, A., Foncubierta–Rodriguez, A., Van de Ville, D., and Muller, H., "Multiscale Lung Texture Signature Learning Using The Riesz Transform," MICCAI 2012, 7512, 517-524 (2012).
[10] Foncubierta-Rodríguez , A., Depeursinge, A. and Müller, H., "Using Multiscale Visual Words for Lung Texture Classification and Retrieval", Medical Content-Based Retrieval for Clinical Decision Support, 7075, 69-79 (2012)
[11] Sørensen, L., Shaker, S.B. and de Bruijne, M., "Quantitative analysis of pulmonary emphysema using local binary patterns", IEEE Trans Med Imaging, 29(2), 559-569 (2010).
[12] C. Jacobs, C. I. Sanchez, S. C. Saur, T. Twellmann, P. A. de Jong and B. van Ginneken, "Computer-aided detection of ground glass nodules in thoracic CT images using shape, intensity and context features," in MICCAI LNCS, 6893, 207–214 (2011).
[13] Farag, A., Elhabian, S., Graham, J., Farag, A. and Falk, R., "Toward precise pulmonary nodule descriptors for nodule type classification," in MICCAI LNCS, 6363, 626–633, (2010).
[14] Depeursinge, A., Vargas, A., Platon, A., Geissbuhler, A., Poletti, P. A., and Muller, H., "Building a reference multimedia database for interstitial lung diseases," Comp Med Imaging and Graph., 36 (3), 227–238 (2012).
[15] Manning, C.D., Raghavan, P., and Schutze, H., Introduction to information Retrieval, 1st ed., Cambridge University Press (2008).