

Feature-based Image Patch Approximation for Lung Tissue Classification

Yang Song, *Student Member, IEEE*, Weidong Cai, *Member, IEEE*, Yun Zhou,
and David Dagan Feng, *Fellow, IEEE*

Abstract—In this paper, we propose a new classification method for five categories of lung tissues in high-resolution computed tomography (HRCT) images, with feature-based image patch approximation. We design two new feature descriptors for higher feature descriptiveness, namely the rotation-invariant Gabor-local binary patterns (RGLBP) texture descriptor and multi-coordinate histogram of oriented gradients (MCHOG) gradient descriptor. Together with intensity features, each image patch is then labeled based on its feature approximation from reference image patches. And a new patch-adaptive sparse approximation (PASA) method is designed with the following main components: minimum discrepancy criteria for sparse-based classification, patch-specific adaptation for discriminative approximation, and feature-space weighting for distance computation. The patch-wise labelings are then accumulated as probabilistic estimations for region-level classification. The proposed method is evaluated on a publicly available ILD database, showing encouraging performance improvements over the state-of-the-arts.

Index Terms—Texture, gradient, reference, adaptive

I. INTRODUCTION

The interstitial lung disease (ILD) represents a group of more than 150 disorders of the lung parenchyma [1]. Most of these cause progressive scarring of lung tissues and eventually affect breathing. Determining the specific type of disorder is important for treatment, and in conjunction with other methods, such as blood tests and pulmonary function tests, imaging scans are often used for accurate diagnosis. In particular, HRCT imaging is quickly becoming the standard practice with its high imaging quality. Different ILDs normally exhibit different combinations of tissue patterns on HRCT images [2], and differentiating the tissue patterns is critical to identify the actual type of ILD. However, interpreting the HRCT images for lung diseases is challenging even for trained radiologists [1]. Patients also have different physical conditions and medical histories, hence even those with the same type of ILD could display quite different tissue patterns. As a consequence, manual interpretation of the images could be error prone, especially when the radiologists are under heavy workload

with short time frames. It is thus suggested that an automatic system for differentiating the tissue patterns would be useful to provide initial screening or second opinions.

In this study, we focus on classification of five categories of lung tissues on HRCT images – normal, emphysema, ground glass, fibrosis and micronodules, which are highly prevalent among the main types of ILDs. Examples of these tissue patterns are shown in Fig 1. It can be seen that while in general there are perceivable differences between the different categories, the visual distinctions between different categories are sometimes subtle, and the pattern variations within the same tissue category are rather obvious. Therefore, it is quite challenging to design a robust method for automatic classification, accommodating both low inter-class distinctions and high intra-class variations.

A. Related Work

Image classification is normally performed in two stages: feature extraction for encoding the image features as feature descriptors, and labeling of image categories using supervised approaches. Being an active research field for a long time, most of the image classification techniques have been applied to a wide range of imaging problems, including the lung CT images. We will thus review mainly the recent works on lung CT images to cover the popular methodologies, and only include studies from other imaging domains if the proposed methods are not normally used for lung studies.

Visual features of lung tissues can be described numerically in a number of ways [3]. The simple gray-level distributions are often used [4]–[15], since the intensities well represent the physical properties of lung tissues. Second-order statistics such as the gray-level cooccurrence matrices (GLCM) and run length (RLE) [5], [12], [13], [16], [17] have also been widely incorporated for additional feature information. Another type of popular feature extraction techniques is based on filters, e.g. Gaussian and wavelets [6], [8]–[10], [14], [18], to highlight specific image features such as edges. Other features such as shape and spatial contexts [10], [17], [19]–[21], are also used for certain applications in lung imaging.

While these relatively traditional feature descriptors have demonstrated effectiveness, more recent descriptors in the general imaging domain have also been incorporated for lung imaging. Some examples include the scale-invariant feature transform (SIFT) [22], [23], the local binary patterns (LBP) [8], [10], and the histogram of oriented gradients (HOG) [21]. Such feature descriptors are comprehensive in extracting

This work was supported in part by ARC grant.

Y. Song is with Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, Sydney 2006, Australia (e-mail: yson1723@uni.sydney.edu.au).

W. Cai is with BMIT Research Group, School of Information Technologies, University of Sydney, Sydney 2006, Australia.

Y. Zhou is with the Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA.

D. D. Feng is with BMIT Research Group, School of Information Technologies, University of Sydney, Sydney 2006, Australia, and Med-X Research Institute, Shanghai Jiaotong University, Shanghai 200030, China.

small image details, and yet they provide multi-resolution and histogram quantization properties that are especially useful for accommodating feature variations. However, while most of these approaches deliver promising performances using the original feature descriptors, it is highly possible to achieve even better results by enhancing the descriptor designs based on medical imaging characteristics.

The above feature extraction techniques are usually defined for image patches, or often referred to as regions-of-interest (ROI) with fixed sizes. In some cases, however, the objective is to classify a larger annotated ROI area containing multiple image patches (denoted as AROI for clarity). Nevertheless, with patch-based processing, it is easier to compute features, and such features would exhibit lower intra-class variations than the AROI-level features. Furthermore, an AROI normally displays repetitive patterns with spatial correlation between patches not as informative as the local features; and hence it is quite reasonable to perform patch-level processing. An example image illustrating AROI and patch is shown in Fig. 2.

Once the feature descriptors are derived, the next stage is to perform labeling of these descriptors for image classification. The labeling is usually based on supervised approaches, and the most commonly used classifiers include k -nearest neighbor (k NN) [6], [8], [15], [16], [18], support vector machine (SVM) [7], [9], [11], [12], [14], [17], [21], [24], linear discriminant analysis (LDA) [10], [22], Bayesian classifiers [5], and artificial neural network (ANN) [4], [13]. Among these, the SVM classifier is normally highly effective, but would be error prone if the feature spaces exhibit considerable overlaps, especially with images of different categories appearing quite similar. The k NN classifier is intrinsically capable of accommodating the intra-class variations, but the feature descriptors are usually not descriptive and discriminative enough to achieve accurate classification based on simple distance measures.

A different type of classification – the sparse representation method – has been applied successfully in various imaging applications [25]–[28], and recently for medical imaging as well [29], [30]. The basic idea is that, a query image is classified based on minimum reconstruction error from a set of reference images. Such a method can be considered analogous to k NN, but the distance measure is with the optimal combination of multiple references that are selected adaptively to the query image. It is also quite effective in handling intra-class variations, with classification based on reference samples rather than learned parametric models. However, since the sparse representation is originally aimed to optimize reconstruction, it does not necessarily lead to accurate classification; and for cases with low inter-class distinctions, an additional classifier (e.g. SVM) seems still necessary [31].

B. Our Contribution

In this work, we propose a new image classification method for lung tissue patterns, based on feature-based image patch approximation. Our main methodological contributions are three-fold. First, a set of texture, intensity, and gradient (T-I-G) features are extracted for each image patch, and two new feature descriptors are proposed: 1) a new rotation-invariant Gabor-LBP (RGLBP) feature descriptor to represent

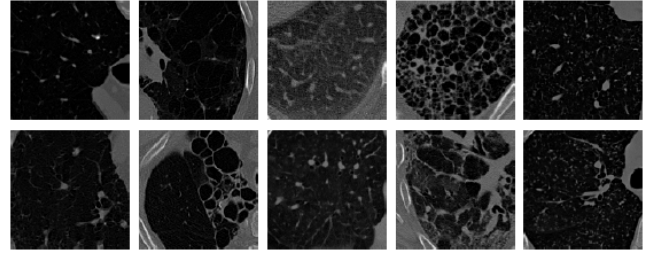


Fig. 1. Two example images of each tissue category. From left to right: normal, emphysema, ground glass, fibrosis and micronodules.

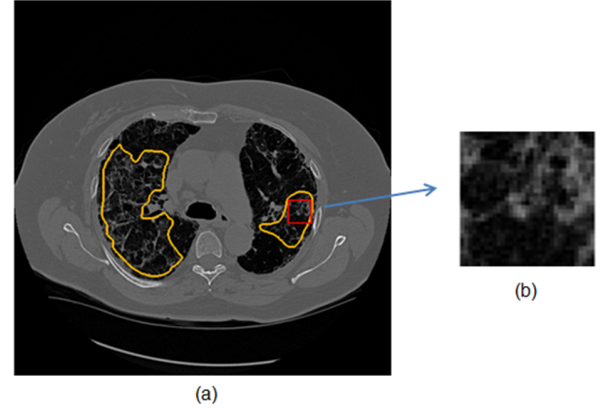


Fig. 2. A sample lung HRCT image slice shown in (a), in which the orange contours indicate two AROIs, and the red box highlights an image patch that is shown in (b).

rich texture features integrating multi-scale Gabor filters and LBP histograms; 2) a new multi-coordinate HOG (MCHOG) descriptor to extract the gradient features while accommodating rotation variance with radial-specific coordinate systems.

Second, each image patch is then classified based on reference dictionaries with a new patch-adaptive sparse approximation (PASA) algorithm, designed for better classification accuracy in the sparse representation: 1) the image patch labeling is enhanced with a statistical measure of the sparse coefficients to measure the minimum discrepancy; 2) a patch-specific adaptation method is designed based on pairwise feature distances to alter the feature values of the reference dictionaries for more discriminative approximation; 3) a feature-space weighting scheme is designed based on overlapping of feature distributions for feature distance computation.

Third, the labeling of the annotated ROI (AROI) is finally obtained based on probabilistic estimation from the patch-wise classification. And the proposed method is evaluated on the publicly available ILD database [2], showing promising performance improvements over the state-of-the-art results reported for the same database. Furthermore, since our proposed feature descriptors (RGLBP and MCHOG) and the approximative image classification algorithm (PASA) are designed based on few assumptions about the problem domain, these methods are thus extensible to other medical imaging problems as well.

The paper is structured as follows. Sections II, III and IV describe the proposed tissue classification method – T-I-G feature extraction, approximative patch classification and

collective region labeling. Section V describes the dataset and evaluation metrics. Section VI presents the experimental results and discussions. We conclude the paper in Section VII.

II. T-I-G FEATURE EXTRACTION

Based on our visual analysis of lung images, it is observed that texture, intensity and gradient distribution of soft tissues within an image patch are quite informative and discriminative for different categories of lung tissues. Therefore, a patch-wise T-I-G feature set, combining texture, intensity and gradient features are extracted for each image patch. And motivated by the recent works in the general imaging domain, we design a set of features based on the popular LBP and HOG descriptors, with modifications introduced for better feature descriptiveness.

Formally, denote an image patch as P comprising of $X \times Y$ pixels $P = \{p_i : i = 1, \dots, X \times Y\}$, and I_i as the intensity value of pixel p_i . The pixel p_i is also indicated by its coordinate as $p(x, y)$, and hence the corresponding $I(x, y)$. A T-I-G feature vector $f(P)$ is then derived for the image patch P with details in the following.

A. Texture Description

The LBP feature [32] describes the spatial structure of local image texture, and can be easily configured to be multi-resolution and rotation-invariant. However, the LBP feature might capture too many image details, and introduce large degree of unnecessary feature variations within the same tissue category. On the other hand, the multi-scale and multi-orientation representation of Gabor filters is often demonstrated as a highly effective texture descriptor. However, being rotation-variant, the Gabor filters become not directly suitable for our problem; but the multi-scale nature is quite useful for computing multi-resolution LBP features. Therefore, to incorporate rich texture information while attempting to minimize intra-category variations, we choose to design a new *rotation-invariant Gabor-LBP* (RGLBP) texture descriptor to incorporate the multi-scale property of Gabor filters and the rotation-invariant property of LBP features.

Let $g_{s,r}(x, y)$ represent the Gabor functions constructed from the basis function $g(x, y)$, with $s = 0, \dots, S - 1$ and $r = 0, \dots, R - 1$, where S and R denote the numbers of scales and orientations. The scale and orientation definitions follow the standard conventions [33]. A set of Gabor-filtered images $\{I^{s,r}\}$ are then computed by convolving image I with each Gabor function. To create rotation-invariant Gabor-filtered images, all R Gabor functions of a certain scale s are summed together [33]:

$$I^s(x, y) = \sum_{r=0}^{R-1} I^{s,r}(x, y) \quad (1)$$

The resultant image set thus comprises of only S Gabor-filtered images $\{I^s : s = 0, \dots, S - 1\}$, rather than the original set of $S \times R$ images, representing an S -dimensional multi-scale and rotation-invariant texture feature for each pixel $p(x, y)$.

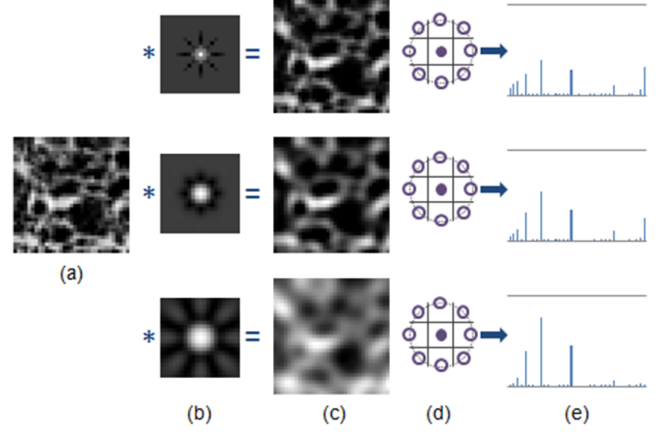


Fig. 3. Illustration of the proposed RGLBP texture descriptor. (a) An image patch. (b) Rotation-invariant Gabor filter bank with 3 scales. (c) The Gabor filtered images. (d) Structure of LBP with radius 1 and 8 neighboring pixels. (e) The derived rotation-invariant histogram for each scale, with x -axis as the feature dimension and y -axis as the feature value.

Next, for each Gabor-filtered image I^s , rotation-invariant LBP¹ feature $LBP_s(p_i)$ is computed for each pixel p_i :

$$LBP_s(p_i) = \min\{ROR(LBP'_s(p_i|8, 1), n)\} \quad (2)$$

$$LBP'_s(p_i|8, 1) = \sum_n \mathbf{1}(I_i^s > I_{neigh(i,n)}^s) 2^n$$

where I_i^s denotes the intensity of pixel p_i in image I^s , $neigh(i, n)$ indexes the neighboring pixels (equally spaced on a circle of radius 1) of the center pixel p_i , and $n = 0, \dots, 7$. The operator $ROR(\cdot)$ performs a circular bit-wise shift to produce a minimum 8-bit number $LBP_s(p_i)$ and hence the LBP feature becomes rotation-invariant. Here we simply use the 8 immediate neighbors for calculating LBP, since the multi-resolution information is encoded in the Gabor-filtered images $\{I^s\}$.

The LBP features $LBP_s(p_i)$ of all pixels in the image patch P are then accumulated as a histogram feature $RGLBP_s(P)$. As explained in [32], the total number of possible LBP values is 36, and hence the histogram $RGLBP_s(P)$ of these pixel-wise LBP features is also 36-dimensional. And the concatenation of such histogram features from all scales $\{I^s\}$ is thus the final feature vector:

$$RGLBP(P) = \{RGLBP_s(P) : s = 0, \dots, S - 1\} \quad (3)$$

which is $S \times 36$ dimensional. Here the handling for rotation invariance (i.e. computation of LBP features on rotation-invariant Gabor-filtered images I^s) is a key difference between our RGLBP descriptor and other Gabor+LBP features, which usually combine LBP and Gabor ignoring the rotation variations caused by Gabor filters (e.g. [34]). An example of the RGLBP computation is shown in Fig. 3.

B. Intensity Description

While LBP measures the differences between a pixel and its surrounding neighbors, the descriptor discards the actual

¹The LBP package is downloaded from <http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab>

intensity values in the image patch. Such intensity information, is however quite informative for lung CT images, e.g. normal tissues usually exhibit darker appearances. Therefore, an intensity histogram $IH(P)$ is also included calculated with the number of bins set to 32 empirically.

C. Gradient Description

Gradient distribution of an image is a different type of feature in complementary to the texture and intensity features. It is potentially very useful for discriminating pathological and normal lung tissues, since the former type often contains small segments that are less common in the normal lung. Among the various types of gradient-based features, the HOG feature [35] has been suggested as very effective, especially when coupled with LBP features [36], [37]. A problem with HOG features for lung images is, however, that it represents the distribution of absolute gradient orientations and hence is not invariant to rotations. While normally the rotation issue is tackled by assigning a dominant orientation based on local image statistics, such as SIFT [38], it is rather not intuitive to perceive a dominant orientation for image patches with complex textures. Therefore, inspired by the work on a SIFT-related rotation-invariant descriptor [39], we design a new *multi-coordinate HOG* (MCHOG) descriptor to accommodate the possible rotations.

Specifically, first, we denote the center pixel of image patch P as p_0 . With p_0 as the origin, P is divided into $V = 8$ radial sections, with the first section covering $[\pi/8, 3\pi/8)$, and the rest covering the subsequent angle ranges. The image patch P can be thus represented by:

$$\begin{aligned} P &= \{RS_v(\alpha_{v,1}, \alpha_{v,2}) : v = 1, \dots, V\} \\ \alpha_{v,1} &= 2v\pi/V - \pi/8 \\ \alpha_{v,2} &= 2v\pi/V + \pi/8 \end{aligned} \quad (4)$$

And for each radial section RS_v , a coordinate system is defined:

$$y_v = \langle 1, \angle \frac{\alpha_{v,1} + \alpha_{v,2}}{2} \rangle \quad (5)$$

$$x_v = \perp y_v \quad (6)$$

A visual illustration of the radial-specific coordinate systems is shown in Fig. 4b.

Then, for each RS_v , the vectors for computing gradients are rotated according to the section-specific coordinate system, and the gradient orientation $GO_v(p_i)$ and magnitude $GM_v(p_i)$ are computed for each pixel $p_i \in RS_v$ as:

$$\begin{aligned} GO_v(p_i) &= \arctan\{h_{y_v}(p_i)/h_{x_v}(p_i)\} \\ GM_v(p_i) &= \sqrt{h_{y_v}(p_i)^2 + h_{x_v}(p_i)^2} \end{aligned} \quad (7)$$

where $h_{x_v}(p_i)$ and $h_{y_v}(p_i)$ are the gradient values at p_i in x_v and y_v directions. Note that the gradient orientation $GO_v(p_i)$ is now aligned to the radial-specific coordinates.

Next, the image patch P is divided into K overlapping cells $P = \{CS_k : k = 0, \dots, K-1\}$ (Fig. 4d). And a histogram $MCHOG(CS_k)$ of gradient orientations weighted by gradient magnitudes is computed for each cell CS_k . Then patch-level histogram feature $MCHOG(P)$ is constructed

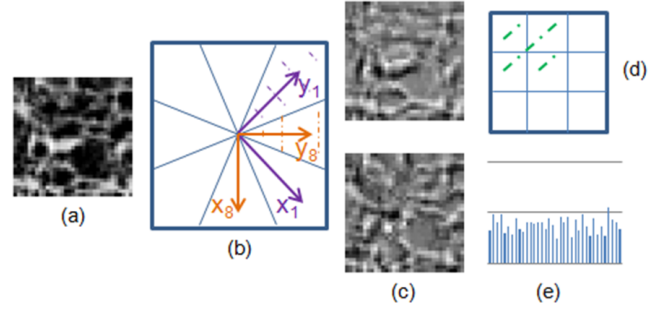


Fig. 4. Illustration of the proposed MCHOG gradient descriptor. (a) An image patch. (b) The radial-specific coordinate systems, with the light blue lines indicating the division of the radial sections, and purple and orange arrows representing the xy coordinates of radial section P_1 and P_8 . (c) The gradients computed on y (upper) and x (lower) directions, based on the radial-specific coordinate systems. (d) Subdivision of 4 overlapping cells, with the green shaded area representing the first cell. (e) The derived gradient histogram, with x -axis as the feature dimension and y -axis as the feature value. Note that the square boxes of (b) and (d) are actually of the same size as the image patch, but drawn with different scales for easier viewing.

by concatenating the cell-level histograms, and the order of concatenation is derived by choosing the cell with maximum mean intensity as the starting position k_0 :

$$\begin{aligned} MCHOG(P) &= \{MCHOG(CS_{k'})\} \\ k' &= k_0, \dots, K-1, 0, \dots, k_0-1 \\ k_0 &= \operatorname{argmax}_k \sum_i I_i : \forall p_i \in CS_k \end{aligned} \quad (8)$$

Here if $k_0 = 0$, then $k' = 0, \dots, K-1$. By default, the number of histogram bins in HOG is 9, hence the gradient descriptor $MCHOG(P)$ is of $K \times 9$ dimensions.

Here the radial-specific coordinate systems and hence the gradient orientations are rotation-invariant between radials, but not within a certain radial. Although the descriptor can be made completely rotation-invariant by creating pixel-specific coordinate systems as suggested in [39], our feature design is motivated by attempts to avoid losing too much discriminative power of features and establish a better balance with the feature descriptiveness. Such a way of computing the gradient orientations is also the fundamental difference between our MCHOG feature and the circular-cell HOG feature [35], [40], even though the radial structures seem similar.

D. Feature Vector

In summary, a patch-wise T-I-G feature vector $f(P)$ of $H = (S \times 36 + 32 + K \times 9)$ dimensions is extracted for each image patch P :

$$f(P) = \{RGLBP(P), IH(P), MCHOG(P)\} \quad (9)$$

Considering the small size of P , it is intuitive to select small number of scales (S) and cells (K) for the feature computation. Furthermore, based on our initial study on a rather small set of images, it is observed that in order not to smooth out the small tissue patterns, 3 spatial scales ($S = 3$) are quite adequate; and in order not to introduce too much spatial information causing extra feature variations, 4 square cells ($K = 4$) are used. Performance comparison with other parameter settings for the entire dataset are later presented in the results. The

three parts of the feature vector – $RGLBP(P)$, $IH(P)$, and $MCHOG(P)$ – are also normalized to have a common mean value, so they would carry similar weights in the overall feature description. The mean of a certain feature, e.g. $RGLBP(P)$, is computed by dividing the sum of its feature elements with its feature dimension. Then by linear rescaling of $IH(P)$ and $MCHOG(P)$, their means are aligned to the same mean as $RGLBP(P)$.

III. APPROXIMATIVE PATCH CLASSIFICATION

With the patch-wise T-I-G feature vector $f(P)$, the next step is to classify each image patch P into one of the five tissue categories. Considering that lung images normally exhibit quite different patterns even within the same tissue category, we expect that even with the comprehensive feature design, large intra-class variations would still exist. Therefore, we would like to use a classification scheme that is especially effective in handling such issues. And we thus design a data-adaptive and non-parametric approach, namely the *patch-adaptive sparse approximation* (PASA) method, to classify an image patch P based on the closeness of approximation by other image patches from each tissue category.

Denoting the five tissue categories – normal, emphysema, ground glass, fibrosis and micronodule – as T_N , T_E , T_G , T_F and T_M , the objective here is to assign each image patch P a category label $L(P) \in \{T_N, T_E, T_G, T_F, T_M\}$. Details of the method are described in the following.

A. Minimum Discrepancy Criterion

Assume that there are Q image patches belonging to a certain tissue category $l \in \{T_N, T_E, T_G, T_F, T_M\}$ in the training set. Denote this set of reference image patches as $\{P_q : q = 1, \dots, Q\}$. An overcomplete feature dictionary matrix D_l for the Q reference image patches is then constructed by concatenating their features as column vectors:

$$D_l = \{f(P_q) : q = 1, \dots, Q\} \in \mathbb{R}^{H \times Q} \quad (10)$$

where H is the feature dimension and $H < Q$, and each column vector $f(P_q)$ is referred to as an atom. Here five dictionaries are created with one for each tissue category, and the number of atoms Q in each dictionary could be different depending on the size of training data. More details of dictionary construction are described in a later subsection.

To classify an image patch P , a sparse-regularized linear model is formulated to compute an approximated vector $f'_l(P)$ of its feature $f(P)$ from the feature dictionary D_l :

$$\omega_l = \arg\min_{\omega_l} \|f(P) - D_l \omega_l\|_2^2 \text{ s.t. } \|\omega_l\|_0 \leq C \quad (11)$$

$$f'_l(P) = D_l \omega_l \quad (12)$$

where $\omega_l \in \mathbb{R}^{Q \times 1}$ is a sparse coefficient vector with C nonzero elements, and C is a constant. The optimization problem Eq. (11) can be solved efficiently using the *orthogonal matching pursuit* (OMP) [41] algorithm².

²The OMP package is downloaded from <http://www.cs.technion.ac.il/~ronrubin/software.html>

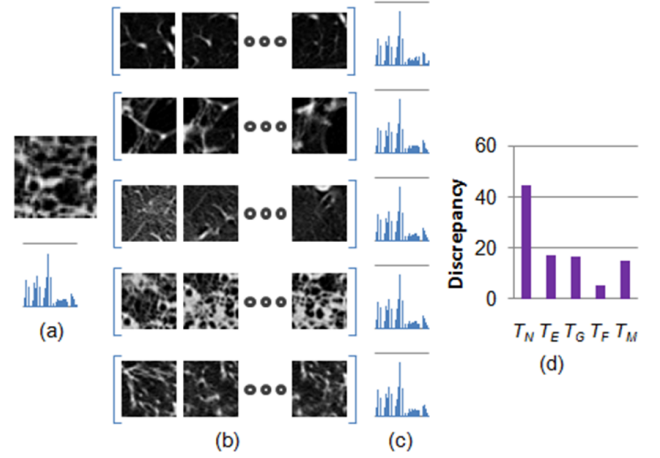


Fig. 5. Illustration of the proposed approximative patch classification. (a) An image patch P of category T_F , and its feature descriptor $f(P)$ with x -axis as the feature dimension and y -axis as the feature value. (b) The feature dictionary D_l constructed for each tissue category, with rows from top to bottom showing tissue categories T_N , T_E , T_G , T_F and T_M . (c) The approximated features $f'_l(P)$ from the corresponding dictionaries D_l . (d) Measure of discrepancy between the approximated features and the original feature vector, and the image patch P is thus labeled as T_F .

Then, by deriving the feature approximation for each tissue category, the labeling $L(P)$ of image patch P is the one producing the minimum discrepancy between the original feature $f(P)$ and the approximated ones $\{f'_l(P)\}$:

$$L(P) = \arg\min_l \|f(P) - f'_l(P)\|_2 \sigma(\omega_l) \quad (13)$$

Here we incorporate the standard deviation $\sigma(\omega_l)$ of the nonzero coefficients in ω_l to improve the labeling accuracy, and $\sigma(\omega_l)$ is computed by first gathering all nonzero elements in ω_l then calculating the standard deviation on this set. A smaller $\sigma(\omega_l)$ is preferred since it implies that the approximation is more of a consensus between the dictionary atoms; while a large $\sigma(\omega_l)$ could hint a disagreement. In this way, the classification is data-adaptive, i.e. the labeling $L(P)$ is determined based on the feature similarity between P and the other image patches. The approach is also non-parametric, meaning there is not a parametric model generated from the training set. Fig. 5 shows an overview of the classification process of an image patch.

B. Patch-specific Adaptation

A problem with the sparse model Eq. (11) is that it is possible to obtain a good approximation of an image patch P from a linear combination of several quite different image patches, leading to mislabeled tissue category. To obtain accurate labeling, assuming an image patch P is of category T_N , we would expect that its feature approximation $f'_l(P)$ based on dictionaries $\{D_l : l \neq T_N\}$ would produce larger discrepancies with $f(P)$ than $D_l : l = T_N$ and hence the labeling of P would be T_N . To achieve such an effect, our idea is to alter the feature values of the reference dictionaries based on the degree of similarity between P and each reference image patch P_q , so that the dictionaries representing the incorrect

tissue categories would tend to generate more distant feature approximations. This patch-specific adaptation is thus helpful to achieve discriminative feature approximation, and is also the main difference between our model and the more standard sparse representation schemes.

To do this, first, a feature distance $d(P, P_q)$ (a scalar number) between image patch P and each reference patch P_q is computed. And to reduce the risk of consistent misrepresentation of feature similarities, histogram-intersection distance is computed instead of the Euclidean distance used for discrepancy-based labeling:

$$d(P, P_q) = \lambda \cdot \{|f(P) - f(P_q)| ./ (f(P) + f(P_q))\} \quad (14)$$

where λ is a vector of feature weights (derivation details in the next subsection), and $||$ and $./$ denote element-wise absolute and division operations. And based on the feature distances between P and all atoms in the dictionary D_l , the distance value $d(P, P_q)$ is normalized to $[0, 1]$.

Then, a scaling coefficient $\phi(P, P_q)$ (a scalar number) is derived to rescale the feature vector $f(P_q)$ into $f^\phi(P_q)$, and the design of $\phi(P, P_q)$ is based on two factors. First, the degree of similarity between P and P_q – if they are quite similar, $\phi(P, P_q)$ should be close to 1 so that the scaled feature $f^\phi(P_q)$ would exhibit minimum variation from the original feature vector. Take Fig. 6c as an example, with an image patch P of the T_F category, the scaling coefficients computed for the reference patches from non- T_F categories are more distant from 1. Second, the sign of distance between $f(P)$ and $f(P_q)$ – if $f(P)$ is larger, $\phi(P, P_q)$ should be less than 1 so that $f^\phi(P_q)$ would become even smaller to increase the feature distance between P and P_q (as shown in Fig. 6b). The scaling coefficient $\phi(P, P_q)$ is thus computed as:

$$\phi(P, P_q) = \begin{cases} \exp(-d(P, P_q)), & \text{if } f(P) \geq f(P_q) \\ 2 - \exp(-d(P, P_q)), & \text{otherwise} \end{cases} \quad (15)$$

Here $\phi(P, P_q) \in (0, 2)$, and $f(P) \geq f(P_q)$ is true if at least half of the feature dimensions of $f(P)$ take values not smaller than the corresponding dimensions in $f(P_q)$. Then the feature vector $f(P_q)$ is rescaled into vector $f^\phi(P_q)$ by:

$$f^\phi(P_q) = \phi(P, P_q) f(P_q) \quad (16)$$

Note that $\phi(P, P_q)$ is computed for every pair of image patches P and P_q , and hence the feature scaling for a certain reference patch P_q is adaptive to both image patches P and P_q .

Finally, the scaled feature vectors $f^\phi(P_q)$ are assembled as the transformed reference dictionary as:

$$D_l^\phi = \{f^\phi(P_q) : q = 1, \dots, Q\} \in \mathbb{R}^{H \times Q} \quad (17)$$

and integrated into sparse linear model as the following:

$$\omega_l = \underset{\omega_l}{\operatorname{argmin}} \|f(P) - D_l^\phi \omega_l\|_2^2 \quad \text{s.t. } \|\omega_l\|_0 \leq C \quad (18)$$

The same approximation procedure Eq. (13) is then used to obtain the labeling of image patch P . And Fig. 6d gives an example that with the patch-specific adaptation, the image patch P is now correctly labeled as T_F .

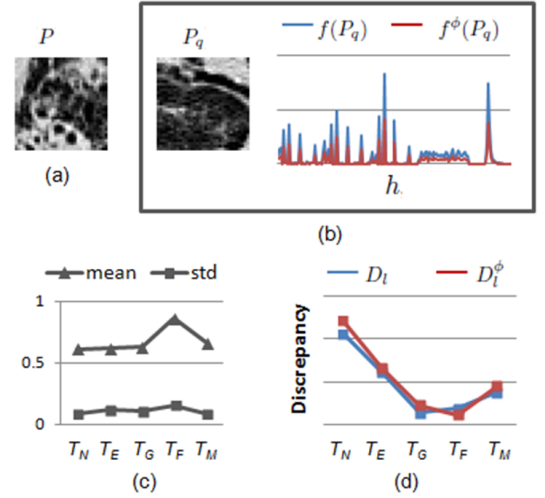


Fig. 6. Illustration of the proposed patch-specific adaptation. (a) An image patch P of category T_F . (b) A reference image patch P_q of category T_G , with its original (blue) and scaled (red) descriptors $f(P_q)$ and $f^\phi(P_q)$. (c) The mean (triangle) and standard deviation (square) of all scaling coefficients computed for P relative to P_q in each tissue category. (d) The discrepancy measures for P based on original (blue) and scaled (red) dictionaries D_l and D_l^ϕ , showing labeling of T_F with the latter case.

C. Feature-space Weighting

Different feature dimensions should contribute differently to the distance computation between feature vectors, and hence naturally feature weights are desired. To derive the feature weight vector $\lambda = \langle \lambda_1, \lambda_2, \dots, \lambda_H \rangle$ with one weight value λ_h corresponding to one feature dimension, a density estimation-based approach is designed based on the reference set.

First, we denote a feature element in the T-I-G feature vector $f(P_q)$ as $fv_{q,h}$ and $f(P_q) = \{fv_{q,h} : h = 1, \dots, H\}$. By regrouping the atoms in the feature dictionary based on the feature dimensions, the dictionary D_l is then denoted as a concatenation of row vectors $\{fr_h\}$:

$$D_l = \{fr_h : h = 1, \dots, H\} \in \mathbb{R}^{H \times Q} \quad (19)$$

$$fr_h = \{fv_{1,h}, fv_{2,h}, \dots, fv_{Q,h}\} \in \mathbb{R}^{1 \times Q}$$

Then for a certain feature dimension h , a probability distribution $PD_h(D_l)$ of the Q reference patches is created using the *kernel density estimation* (KDE):

$$PD_h(D_l, \mu_h) = Q^{-1} \sum_q \mathcal{K}(\mu_h - fv_{q,h}) \quad (20)$$

where μ_h takes a value in the feature range of fr_h :

$$\mu_h = [\min_q fv_{q,h}, \max_q fv_{q,h}] \quad (21)$$

And $\mathcal{K}(\cdot)$ is the Gaussian kernel with bandwidth computation following normal distribution of all data samples.

Next, by computing a probability distribution for each tissue category, a set of five probability distributions are created for a feature dimension h : $\{PD_h(D_l)\}$, as illustrated in Fig. 7a. And a degree of feature overlapping FO_h (a scalar number) between the five tissue categories is computed as:

$$FO_h = \frac{\sum_{\mu_h} [\sum_l PD_h(D_l, \mu_h) \cdot \wedge_l PD_h(D_l, \mu_h)]}{\sum_{\mu_h} \sum_l PD_h(D_l, \mu_h)} \quad (22)$$

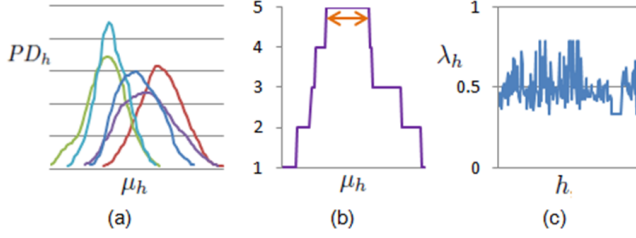


Fig. 7. Illustration of the proposed feature-space weighting. (a) The KDE curves of a certain feature dimension h for five tissue categories with different colors. (b) The number of nonzero probabilities (y -axis) in (a) for every feature value μ_h , and the orange arrow indicates the feature range with five nonzero probabilities accumulating to the overlapping degree FO_h . (c) The derived feature weight vector λ of all feature dimensions.

Here $\wedge_l PD_h(D_l, \mu_h) = 1$ if μ_h is common to all tissue categories $\forall l : PD_h(D_l, \mu_h) > 0$ (as illustrated in Fig. 7b) and otherwise it is 0. The value in $[\cdot]$ represents the overlaps among the tissue categories at μ_h . If there is a large amount of μ_h exhibiting such a common presence, the degree of overlapping FO_h would be high and the discriminative power of feature dimension h would thus be lower.

Finally, based on the overlapping measure FO_h , the feature weight λ_h is derived following a sigmoid curve, so that feature dimensions with smaller FO_h would take larger weights, and the weights are shifted to center at 0.5 based on the mean overlaps of all feature dimensions:

$$\lambda_h = \{1 + \exp[-2(1 - FO_h - |FO|)]\}^{-1} \quad (23)$$

$$|FO| = H^{-1} \sum_h (1 - FO_h)$$

The derived feature weight is visualized in Fig. 7c.

It is also worth mentioning that other means of generating the feature weights are also initially experimented, but found to be much less desirable. The attempted ways are mainly based on discriminative learning to obtain feature weights using 1) SVM and 2) triplet-learning [42]. In this case, the SVM training does not perform well in a linear model; and it is hard to obtain good feature weights using triplet-learning due to huge number of triplet data.

D. Dictionary Construction

To construct the reference dictionary, the whole database is first divided sequentially into three sets of roughly equal numbers of 3-D images. And assuming one set contains Z 3-D images, for each 3-D image, $Z - 1$ images are then used as its reference set. Therefore, for each image patch P , $1/3$ of the datasets, excluding the 3-D image set that contains P , are used as reference images. The patch-wise feature vectors $\{f(P_q)\}$ from all of the reference images are concatenated together according to their ground truth labeling to construct five feature dictionaries $\{D_l : l = T_N, T_E, T_G, T_F, T_M\}$. Note that the dictionary construction is only a process of vector concatenation without real computations, hence it is very low-cost. Here only $1/3$ of the datasets are used as references in order to reduce the dictionary size and speed up the optimization process. Since the number of patches is much bigger than the feature dimension, this dictionary size is

large enough to satisfy the overcomplete constraint for sparse representation. In addition, the reference images are also used to generate the feature weights.

IV. COLLECTIVE REGION LABELING

Since the feature computation and approximation are designed at image-patch level, an AROI containing multiple image patches would sometimes exhibit a mixture of labelings of tissue categories. Therefore, we further perform a region-level classification to achieve a unanimous label for each AROI, based on collective probabilistic estimation of its image patches. Formally, an AROI comprising A image patches is denoted as $AROI = \{P_a : a = 1, \dots, A\}$. With the patch-wise labeling $L(P_a)$ obtained from the approximative approach, we would then like to obtain a single label for the AROI $L(AROI) \in \{T_N, T_E, T_G, T_F, T_M\}$.

First, rather than using a discrete labeling $L(P_a)$, five probability values $\{\forall l : PR(P_a, l)\}$ are computed for each image patch P_a , representing the probabilities of P_a belonging to each tissue category. The probability value $PR(P_a, l)$ is derived based on the discrepancy between its feature vector and the approximation:

$$PR(P_a, l) = \exp\left(-\frac{2\|f(P_a) - f'_l(P_a)\|_2 \sigma(\omega_l)}{\sum_{l'} \|f(P_a) - f'_{l'}(P_a)\|_2 \sigma(\omega_{l'})}\right) \quad (24)$$

where $l' = \{T_N, T_E, T_G, T_F, T_M\}$. Then, the final labeling $L(AROI)$ is thus the category with the highest total probability from all image patches:

$$L(AROI) = \operatorname{argmax}_l \sum_a PR(P_a, l) \quad (25)$$

V. DATASET AND EVALUATION METRICS

The publicly available database of ILD cases [2] is used in this study. The database contains 113 sets of high-resolution CT (HRCT) images with 512×512 pixels per slice. The database also indicates 2062 2-D AROIs that are manually drawn by two radiologists with 15 and 20 years of experience. For each AROI, a tissue pattern annotation is provided, with altogether 17 different tissue patterns. Among these, 5 commonly seen tissue patterns – normal, emphysema, ground glass, fibrosis and micronodule – are studied by the researchers who create the database. We thus also focus on differentiating between these five tissue patterns in our study, involving 1458 AROIs from 95 image sets. A summary of the dataset is listed in Table I. Note that one image set might contain multiple types of tissue patterns, and hence the sum of images of the five categories is larger than the actual number of images.

In this study, our design of 2-D methodology is motivated by the following considerations: 1) the inter-slice spacing is quite large (i.e. 10 mm), 2) the ground truth is provided as 2-D AROIs, and 3) existing works in similar domains are normally based on 2-D slices with 2-D image features [8], [14]. In addition, similar to the settings in [14], an image slice is divided into half-overlapping image patches of 31×31 pixels. And image patches with centroids residing outside of annotated AROIs are excluded from the evaluation. Table I shows the resulting number of patches in each tissue category.

TABLE I
SUMMARY OF THE DATASET USED.

Tissue category	# Images	# AROIs	# Patches
Normal (T_N)	15	157	6934
Emphysema (T_E)	9	108	1474
Ground glass (T_G)	35	416	2974
Fibrosis (T_F)	35	479	4456
Micronodule (T_M)	18	298	7893

To evaluate our method, we first analyze the effects of each component design over the basis techniques, such as the standard LBP and HOG descriptors and sparse-based classifiers. Various settings of the major parameters are also experimented. Furthermore, we compare our proposed method with more standard approaches for both feature extraction and classification, and with the state-of-the-art results reported for the same dataset as well. The performance of tissue classification is measured by recall, precision and F-score:

$$\text{Recall} = TP / (TP + FN) \quad (26)$$

$$\text{Precision} = TP / (TP + FP) \quad (27)$$

$$\text{F-score} = 2TP / (2TP + FN + FP) \quad (28)$$

where TP, FN, and FP are the numbers of true positive, false negative and false positive classifications of tissue categories. As described previously, each image patch is classified based on a reference dictionary constructed with 1/3 of the database excluding the 3-D image set that contains this image patch. Therefore, the TP, FN and FP measures could be computed by accumulating over all image patches in the dataset, in a similar manner to the leave-one-subject-out evaluation technique. The same metrics are also used for AROI-level classifications.

VI. RESULTS AND DISCUSSIONS

A. Tissue Pattern Classification

We first report the patch-wise classification results using our proposed T-I-G feature descriptor and approximative patch classification. As shown in Table II, all of the five tissue categories achieve over 80% true positive rates. Recall we have explained earlier that the main difficulties in robust classification of tissue patterns are 1) low inter-class feature distinctions and 2) high intra-class variations. Regarding the first challenge, some tissue categories often exhibit quite similar features, such as T_E and T_N , T_G and T_F , and T_M and T_N . In particular, both T_E and T_N appearing quite dark in the overall lung field, the parenchyma details differentiating these two could be sometimes difficult to perceive even visually. The small increase of nodules in some T_M patches compared to the normal lung also makes them difficult to identify from T_N patches. It is also found that if the T_G patches show relatively high densities, they could look similar to T_F . Consequently, misclassifications are expected to occur between these three pairs of tissue categories.

Regarding the second challenge, within each tissue category, there is normally large degree of feature variations. One common type of variations is the background intensity.

TABLE II
CONFUSION MATRIX OF PATCH-WISE TISSUE CLASSIFICATION.

Ground Truth	Prediction (%)				
	T_N	T_E	T_G	T_F	T_M
T_N	87.6	3.9	1.9	0.4	6.1
T_E	12.4	80.6	0.4	4.9	1.7
T_G	3.5	0.2	82.7	10.5	3.1
T_F	3.3	4.0	9.3	81.2	2.2
T_M	12.3	0.5	4.3	1.8	81.1

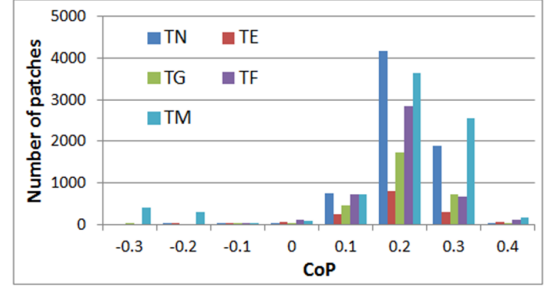


Fig. 8. Histogram with x -axis as the confidence of prediction of an image patch, and y -axis as the number of patches.

For example, the background intensity in T_N and T_M often becomes higher than usual, and can become more difficult to differentiate from the other tissue categories. The more complex types, such as T_E and T_F , contain more irregular local structures and thus exhibit large variations across images of the same category. Especially if considering the patch-level, even the patches from the same image would display varying visual patterns, with patches dividing at different places of the image. Such difficulties would thus also affect the classification performances.

While the patch-wise classification performs reasonably well as shown in Table II, we would like to further analyze if the classification is accurate with large margins. In other words, we want to make sure that an accurate labeling is not by chance, but with good confidence. For this evaluation, we define a confidence of prediction (CoP) measure for an image patch P – the difference between the probability estimates of the expected tissue category $\hat{l}(P)$ and the category l with the highest labeling probability and $l \neq \hat{l}(P)$. Using the annotations defined in Section IV, then $CoP(P) = PR(P, \hat{l}(P)) - \max_{l, l \neq \hat{l}(P)} PR(P, l)$. With this measure, $CoP(P) < 0$ means P is mislabeled, and a large positive $CoP(P)$ implies a large confidence level of labeling and is thus much desirable. As shown in Fig. 8, most image patches exhibit CoP values of at least 0.2. We would not expect a much higher value for CoP since the labeling probability $PR(\cdot)$ is usually a bit larger than 0 (but smaller than 1) in a sparse representation construct. And considering there are five tissue categories to differentiate, such a confidence measure is thus quite good. Based on this result together with Table II, we could infer that the proposed classification approach is quite robust in the presence of the low inter-class distinctions and high intra-class variations.

Table III shows the recall, precision and F-score measures

TABLE III
RESULTS OF PATCH-WISE TISSUE CLASSIFICATION.

	T_N	T_E	T_G	T_F	T_M
Recall (%)	87.6	80.6	82.7	81.2	81.1
Precision (%)	80.7	70.7	74.2	87.1	90.8
F-score (%)	84.0	75.3	78.2	84.1	85.7

TABLE IV
CONFUSION MATRIX OF AROI-LEVEL TISSUE CLASSIFICATION.

Ground Truth	Prediction (%)				
	T_N	T_E	T_G	T_F	T_M
T_N	92.4	0.6	1.9	0	5.1
T_E	21.0	77.1	0	0.9	1.0
T_G	24.9	0.3	67.7	5.7	1.4
T_F	15.1	2.8	7.8	73.4	0.9
T_M	13.4	1.1	8.8	0.7	76.1

of the patch-wise classification. The lower precisions of T_E and T_G categories are partly due to the smaller numbers of image patches in these two tissue categories, compared to the other three categories. The high precision of T_M also indicates that image patches of non- T_M categories are less likely to be misclassified as T_M . After the collective region labeling, the classification metrics calculated at AROI-level are summarized in Table IV and V. Here due to the much smaller number of AROIs than patches, the recall rates shown in both tables are lower than the patch-wise classification results. However, if examining the patch-level results after the collective labeling, the recall rates actually show on average 6% improvement compared to before the collective labeling. An unwanted side effect now is that more T_G and T_F patches are now misclassified as T_N , and hence more AROIs are misclassified as T_N as well. As a result, lower precision and F-score are observed for the T_N category.

TABLE V
RESULTS OF AROI-LEVEL TISSUE CLASSIFICATION.

	T_N	T_E	T_G	T_F	T_M
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

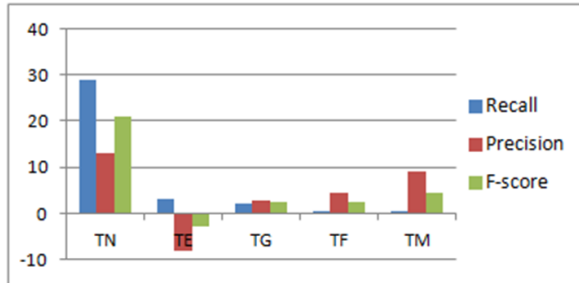


Fig. 9. Performance differences (y -axis) between our patch-wise classification results and the state-of-the-art results reported in [14] for a smaller subset of the same database used in this work.

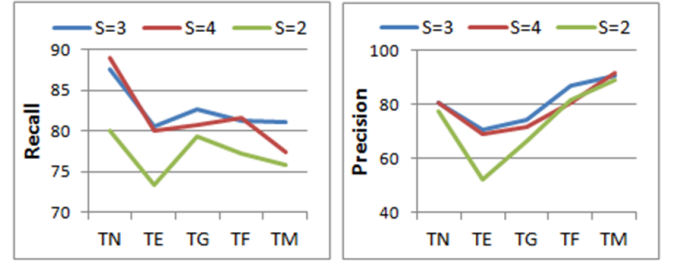


Fig. 10. The patch-wise classification recall (left) and precision (right) with various settings of scale parameter S for feature descriptor RGLBP.

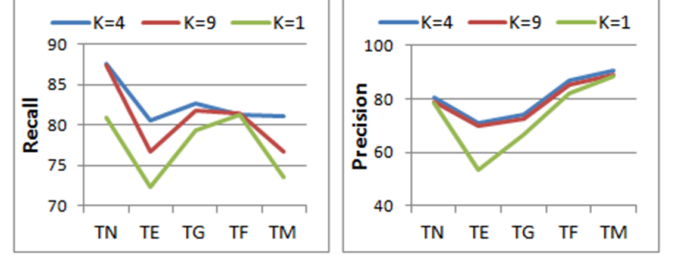


Fig. 11. The patch-wise classification recall (left) and precision (right) with various numbers of cells K for feature descriptor MCHOG.

We then assess our classification performance with the state-of-the-art for ILD tissue classification [14]. The results reported in [14] are taken directly for performance comparison here. As shown in Fig. 9, our results exhibit better performances especially for the T_N category. The higher recall of T_N tissues also contributes to the higher precisions of T_F and T_M categories. For the T_E category, however, our precision measure is lower, and hence causing lower F-score as well. This is mainly due to misclassifications of T_N and T_F tissues as the T_E category. Note that since the results reported in [14] are based on patches, here we conduct the comparison using our patch-wise classification results as well. We would also like to mention that the results in [14] are based on a smaller subset (85 image sets) of the same ILD database used in this work (we use 95 image sets); therefore, it is possible that the comparison could be somewhat biased.

Our method is currently implemented in Matlab, running on a standard PC with a 2.66-GHz dual core CPU. For one image patch, the feature set computation takes about 0.06 s, and the classification needs about 0.12 s. For one AROI, the collective region labeling takes on average 0.02 s.

B. Evaluation of Features

To evaluate the feature design, we first measure the patch-wise classification recall and precision based on different parameter settings for computing the feature descriptors RGLBP and MCHOG, for the entire dataset. As shown in Fig. 10, the classification performances using scale parameters $S = 3$ and $S = 4$ are comparable, with obvious advantages over the fewer number of scales ($S = 2$). This indicates the benefit of having a multi-resolution approach for feature computation. Fig. 11 shows the effects of using different numbers of cells for computing the gradient features. The results imply that

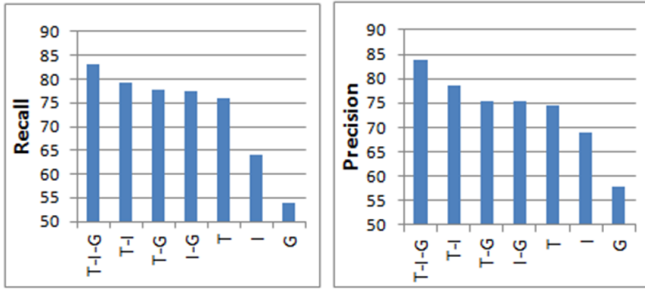


Fig. 12. The average patch-wise classification recall (left) and precision (right) with various feature combinations: proposed feature set T-I-G, texture and intensity (T-I), texture and gradient (T-G), intensity and gradient (I-G), texture (T), intensity (I), and gradient (G).

having multiple cells ($K = 4$ or $K = 9$) makes the feature more descriptive and discriminative than using a single cell only. We would also not suggest using even larger S or K values, since the image patch is relatively small.

The contribution from each type of feature – texture (RGLBP), intensity and gradient (MCHOG) – is then analyzed by comparing the classification results using different combinations of features. Fig. 12 shows that integrating all three types of features delivers much better performance. If classifying using a single type of feature, the proposed texture feature outperforms the gradient feature significantly, suggesting that lung images are better characterized by textures rather than gradients. And the difference between T-I-G and T-I indicates the advantage of introducing the proposed gradient feature, in addition to the texture and intensity features.

By comparing with the classification results using standard LBP and HOG descriptors, the advantages of the proposed RGLBP and MCHOG feature descriptors are shown in Fig. 13. To make fair comparisons with equal feature dimensions (i.e. $S = 3$ and $K = 4$ for RGLBP and MCHOG features), the LBP descriptor is computed rotation-invariantly using 8 neighbors (same as RGLBP) and 3 resolutions (radius of 1, 2 and 3); and HOG descriptor is computed using 9 orientation bins and 4 cells (same as MCHOG). And all compared feature sets are coupled with the proposed PASA classifier. If using only the standard LBP and HOG, the classification tends to have trouble identifying the T_E and T_G patches and also often mislabel other tissue patterns as these two categories. The advantage of having RGLBP or MCHOG rather than the respective standard descriptor can be also seen by comparing the performance of using the T-I-G feature set and the others. The overall results suggest that with RGLBP and MCHOG, the feature descriptions are more comprehensive and thus very helpful for enhancing the classification performance.

C. Evaluation of Classifiers

To evaluate our design of the approximative classifier, we compare the patch-wise classification performances on the entire dataset using the following approaches: 1) the proposed PASA method; 2) sparse representation with standard Euclidean reconstruction error; 3) the proposed minimum discrepancy criterion with the measure of standard deviation; and

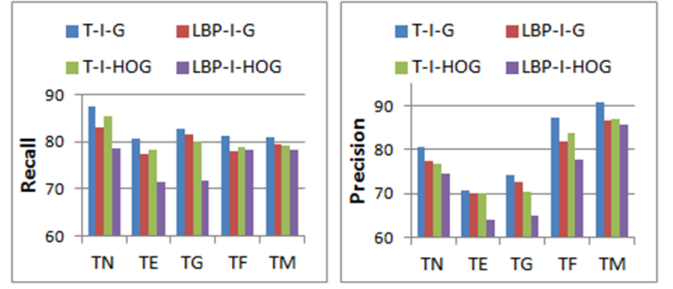


Fig. 13. The patch-wise classification recall (left) and precision (right) with: our proposed T-I-G feature set (blue), replacing RGLBP with standard LBP (red), replacing MCHOG with standard HOG (green), and using standard LBP and standard HOG with the intensity feature (purple).

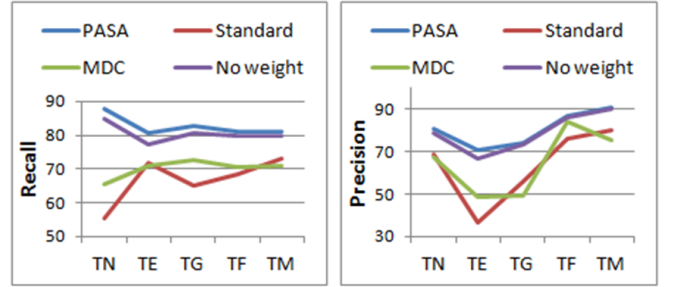


Fig. 14. The patch-wise classification recall (left) and precision (right) with: 1) the proposed PASA method (blue), 2) standard sparse-based classifier (red), 3) the proposed minimum discrepancy criterion (green), and 4) PASA without feature-space weighting (purple).

4) the proposed minimum discrepancy criterion plus the patch-specific adaptation but without the feature-space weighting. In all these experiments, the same T-I-G feature descriptor and reference dictionaries are used, and the sparsity constraint factor C (Eq. (11)) is set to 10 to achieve a good balance between classification accuracy and computational efficiency. As shown in Fig. 14, the improvements of 3) and 4) over the basic sparse-based classification 2) are obvious. The benefit of including the patch-specific adaptation is especially prominent, comparing 4) with 3). With the patch-specific adaptation, about 14% more of the patches are correctly labeled, but resulting in mislabeling of about 2% of patches. So overall, such a component helps to improve the labeling on about 12% of the patches. Relatively small performance enhancements are introduced by the feature-space weighting though, seen from the differences between 1) and 4).

We then use a couple of totally different and popular classifiers, i.e. SVM and k NN, to compare with the proposed sparse-based classification approach. The feature set and reference images used for all classifiers are the same as our proposed method, i.e. T-I-G feature vector with default parameter settings $S = 3$ and $K = 4$ and leave-one-subject-out in 1/3 of the dataset. The best-performing SVM kernel is found to be polynomial, while the linear kernel fails to converge during training and the classification using Gaussian radial basis function (RBF) kernel is very slow. The order of polynomial used is 3 with regularization parameter $C = 1$, which are the default settings for polynomial kernel in LIBSVM [43]. The Euclidean

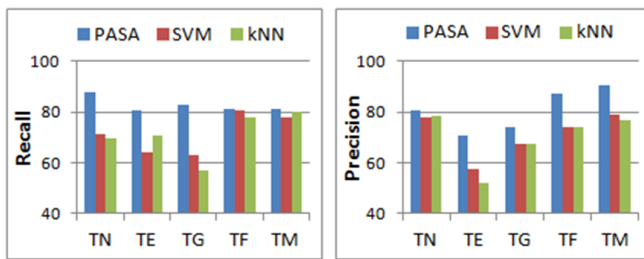


Fig. 15. The patch-wise classification recall (left) and precision (right) with the proposed PASA method, SVM, and k NN.

distance function is used in the k NN classifier with k set to 3. The setting $k = 3$ is found to deliver the best classification performance, while numbers smaller or larger than 3 (1 to 10) exhibit a consistent trend of decrease in classification recall. As shown in Fig. 15, our proposed PASA approach outperforms both the SVM and k NN classifiers considerably, especially for the recalls of the first three tissue categories and in turn most of the precisions as well. This suggests that while PASA does not involve discriminative learning like SVM, it can be actually more discriminative with its non-parametric formulation. Without requiring a parametric modeling of the feature space separation, such an approach is more data-adaptive and is thus able to accommodate the intra-class feature variations well, particularly in this problem domain. And the advantage over k NN implies that with low inter-class feature distinctions, the sparse representation method is more effective with optimal combination of multiple references than using direct distance-based classification.

In terms of computational complexity, compared to SVM, both PASA and k NN have lower training costs, only requiring construction of reference dictionaries. During testing, the two main components of PASA are the pairwise distance computation for patch-specific adaptation and dictionary-based feature approximation for classification. It is less efficient than SVM and k NN, since SVM does not require the pairwise distance computation and works with much fewer support vectors than the dictionary size while k NN involves only pairwise distance computation. However, since both main components in PASA are roughly linear to the dictionary size in complexity, the additional load is fairly low. Therefore, we would suggest that it is worthwhile to give a little computational efficiency for the advantage of better classification.

VII. CONCLUSION AND FUTURE WORK

An automatic classification method for lung HRCT images is presented in this paper. Five categories of lung tissues – normal, emphysema, ground glass, fibrosis and micronodules – that are important for ILD disease diagnosis, are the main objects to be differentiated. To tackle the challenges in low inter-class distinctions and high intra-class variations, we have designed a feature-based image patch approximation method. First, an image patch is represented as a feature vector, based on our proposed RGLBP texture and MCHOG gradient descriptors. Then, the image patch is classified into one of the five tissue categories, using our proposed PASA classifier

based on reference image patches. Finally, a single labeling is assigned for each AROI based on collective probabilistic estimation. Using a publicly available ILD HRCT image database, we have conducted extensive experiments to evaluate the overall method design and the proposed feature descriptors and sparse-based classification, and demonstrated promising performance improvements. We also suggest that the proposed method, in its whole or some components, can be easily extensible to other medical imaging domains. In our future work, we will further investigate more robust techniques of parameter selection for the feature set other than the current default settings, and more adaptive ways of reference dictionary construction other than the current concatenation using 1/3 of the database.

REFERENCES

- [1] W. R. Webb, N. L. Muller, and D. P. Naidich, *High-resolution CT of the lung*. Lippincott Williams Wilkins, 2008.
- [2] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P. A. Poletti, and H. Muller, “Building a reference multimedia database for interstitial lung diseases,” *Comput. Med. Imaging Graph.*, vol. 36, no. 3, pp. 227–238, 2012.
- [3] I. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken, “Computer analysis of computed tomography scans of the lung: a survey,” *IEEE Trans. Med. Imag.*, vol. 25, no. 4, pp. 385–405, 2006.
- [4] 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,” *Med. Phys.*, vol. 30, no. 9, pp. 2440–2454, 2003.
- [5] Y. Xu, M. Sonka, G. McLennan, J. Guo, and E. A. Hoffman, “MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies,” *IEEE Trans. Med. Imag.*, vol. 25, no. 4, pp. 464–475, 2006.
- [6] I. C. Sluimer, M. Prokop, I. Hartmann, and B. van Ginneken, “Automated classification of hyperlucency, fibrosis, ground glass, solid, and focal lesions in high-resolution CT of the lung,” *Med. Phys.*, vol. 33, no. 7, pp. 2610–2620, 2006.
- [7] M. J. Gangeh, L. Sorensen, S. B. Shaker, M. S. Kamel, M. de Bruijne, and M. Loog, “A texture-based approach for the classification of lung parenchyma in CT images,” in *MICCAI LNCS*, vol. 6363, pp. 595–602, 2010.
- [8] L. Sorensen, S. B. Shaker, and M. de Bruijne, “Quantitative analysis of pulmonary emphysema using local binary patterns,” *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 559–569, 2010.
- [9] A. Depeursinge, A. Foncubierta-Rodriguez, D. V. de Ville, and H. Muller, “Lung texture classification using locally-oriented riesz components,” in *MICCAI LNCS*, vol. 6893, pp. 231–238, 2011.
- [10] 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*, vol. 6893, pp. 207–214, 2011.
- [11] R. Xu, Y. Hirano, R. Tachibana, and S. Kido, “Classification of diffuse lung disease patterns on high-resolution computed tomography by a bag of words approach,” in *MICCAI LNCS*, vol. 6893, pp. 183–190, 2011.
- [12] J. Yao, A. Dwyer, R. M. Summers, and D. J. Mollura, “Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification,” *Acad. Radiol.*, vol. 18, no. 3, pp. 306–314, 2011.
- [13] S. C. Park, J. Tan, X. Wang, D. Lederman, J. K. Leader, S. H. Kim, and B. Zheng, “Computer-aided detection of early interstitial lung diseases using low-dose CT images,” *Phys. Med. Biol.*, vol. 56, pp. 1139–1153, 2011.
- [14] A. Depeursinge, D. V. de Ville, A. Platon, A. Geissbuhler, P. A. Poletti, and H. Muller, “Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 4, pp. 665–675, 2012.
- [15] C. S. Mendoza, G. R. Washko, J. C. Ross, A. A. Diaz, D. A. Lynch, J. D. Crapo, E. K. Silverman, B. Acha, C. Serrano, and R. S. J. Estepar, “Emphysema quantification in a multi-scanner HRCT cohort using local intensity distributions,” in *Proc. ISBI*, pp. 474–477, 2012.

- [16] P. D. Korfiatis, A. N. Karahaliou, A. D. Kazantzi, C. Kalogeropoulou, and L. I. Costaridou, "Texture-based identification and characterization of interstitial pneumonia patterns in lung multidetector CT," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 3, pp. 675–680, 2010.
- [17] U. Bagci, J. Yao, A. Wu, J. Caban, T. N. Palmore, A. F. Suffredini, O. Aras, and D. J. Mollura, "Automatic detection and quantification of tree-in-bud (TIB) opacities from CT scans," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 6, pp. 1620–1632, 2012.
- [18] L. Sorensen, M. Nielsen, P. Lo, H. Ashraf, J. H. Pedersen, and M. de Bruijne, "Texture-based analysis of COPD: a data-driven approach," *IEEE Trans. Med. Imag.*, vol. 31, no. 1, pp. 70–78, 2012.
- [19] Y. Song, W. Cai, S. Eberl, M. Fulham, and D. Feng, "Thoracic image case retrieval with spatial and contextual information," in *Proc. ISBI*, pp. 1885–1888, 2011.
- [20] A. El-Baz, A. Soliman, P. McClure, G. Gimel'farb, M. A. El-Ghar, and R. Falk, "Early assessment of malignant lung nodules based on the spatial analysis of detected lung nodules," in *Proc. ISBI*, pp. 1463–1466, 2012.
- [21] Y. Song, W. Cai, J. Kim, and D. D. Feng, "A multistage discriminative model for tumor and lymph node detection in thoracic images," *IEEE Trans. Med. Imag.*, vol. 31, no. 5, pp. 1061–1075, 2012.
- [22] A. Farag, S. Elhabian, J. Graham, A. Farag, and R. Falk, "Toward precise pulmonary nodule descriptors for nodule type classification," in *MICCAI LNCS*, vol. 6363, pp. 626–633, 2010.
- [23] Y. Song, W. Cai, Y. Wang, and D. D. Feng, "Location classification of lung nodules with optimized graph construction," in *Proc. ISBI*, pp. 1439–1442, 2012.
- [24] Y. Song, W. Cai, S. Eberl, M. Fulham, and D. Feng, "Discriminative pathological context detection in thoracic images based on multi-level inference," in *MICCAI LNCS*, vol. 6893, pp. 185–192, 2011.
- [25] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.
- [26] X. Yuan and S. Yan, "Visual classification with multi-task joint sparse representation," in *Proc. CVPR*, pp. 3493–3500, 2010.
- [27] Z. Jiang, Z. Lin, and L. S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent k-svd," in *Proc. CVPR*, pp. 1697–1704, 2011.
- [28] D. Xu, Y. Huang, Z. Zeng, and X. Xu, "Human gait recognition using patch distribution feature and locality-constrained group sparse representation," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 316–326, 2012.
- [29] M. Liu, L. Lu, X. Ye, S. Yu, and M. Salganicoff, "Sparse classification for computer aided diagnosis using learned dictionaries," in *MICCAI LNCS*, vol. 6893, pp. 41–48, 2011.
- [30] Y. Song, W. Cai, Y. Zhou, and D. Feng, "Thoracic abnormality detection with data adaptive structure estimation," in *MICCAI LNCS*, vol. 7510, pp. 74–81, 2012.
- [31] Y. Song, W. Cai, H. Huang, Y. Wang, and D. D. Feng, "Object localization in medical images based on graphical model with contrast and interest-region terms," in *Proc. CVPR Workshop*, pp. 1–7, 2012.
- [32] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.
- [33] J. Han and K. K. Ma, "Rotation-invariant and scale-invariant gabor features for texture image retrieval," *Image and Vis. Comput.*, vol. 25, pp. 1474–1481, 2007.
- [34] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition," in *Proc. ICCV*, pp. 786–791, 2005.
- [35] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. CVPR*, pp. 886–893, 2005.
- [36] X. Wang, T. X. Han, and S. Yan, "A HOG-LBP human detector with partial occlusion handling," in *Proc. ICCV*, pp. 32–39, 2009.
- [37] Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, L. Cao, and T. Huang, "Large-scale image classification fast feature extraction and SVM training," in *Proc. CVPR*, pp. 1689–1696, 2011.
- [38] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [39] B. Fan, F. Wu, and Z. Hu, "Aggregating gradient distribution into intensity orders: a novel local image descriptor," in *Proc. CVPR*, pp. 2377–2384, 2011.
- [40] H. Skibbe and M. Reiser, "Circular fourier-hog features for rotation invariant object detection in biomedical images," in *Proc. ISBI*, pp. 450–453, 2012.
- [41] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [42] A. Frome, Y. Singer, F. Sha, and J. Malik, "Learning globally-consistent local distance functions for shape-based image retrieval and classification," in *Proc. ICCV*, pp. 1–8, 2007.
- [43] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 1–27, 2011.