

Lung Nodule Classification with Multi-Level Patch-based Context Analysis

Fan Zhang, *Student Member, IEEE*, Yang Song, *Student Member, IEEE*, Weidong Cai, *Member, IEEE*, Min-Zhao Lee, Yun Zhou, Heng Huang, Shimin Shan, Michael J Fulham, and Dagan Feng, *Fellow, IEEE*

Abstract—In this paper, we propose a novel classification method for the four types of lung nodules, *i.e.*, *well-circumscribed*, *vascularized*, *juxta-pleural* and *pleural-tail*, in low dose computed tomography (LDCT) scans. The proposed method is based on contextual analysis by combining the lung nodule and surrounding anatomical structures, and has three main stages: an adaptive patch-based division is used to construct concentric multi-level partition; then, a new feature set is designed to incorporate intensity, texture and gradient information for image patch feature description and then a contextual latent semantic analysis-based classifier is designed to calculate the probabilistic estimations for the relevant images. Our proposed method was evaluated on a publicly available dataset and clearly demonstrated promising classification performance.

Index Terms—Classification, patch division, feature design, latent semantic analysis.

I. INTRODUCTION

LUNG CANCER is a major cause of cancer-related deaths in humans worldwide. Approximately 20% of cases with lung nodules represent lung cancers [1]; therefore, the identification of potentially malignant lung nodules is essential for the screening and diagnosis of lung cancer [2] [3] [4]. Lung nodules are small masses in the human lung, and are usually spherical; however, they can be distorted by surrounding anatomical structures, such as vessels and the adjacent pleura

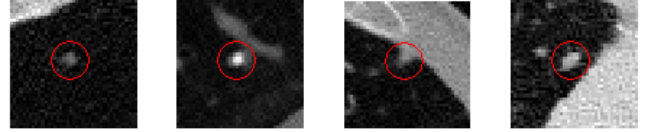


Fig. 1. Transaxial CT images with the four types of nodules, shown from left to right, well-circumscribed, vascularized, juxta-pleural and pleural-tail.

[5]. Intra-parenchymal lung nodules are more likely to be malignant than those connected to surrounding structures [6], and thus lung nodules are divided into different types according to their relative positions. At present, the classification from Diciotti et al. [7] is the most popular approach and it divides nodules into four types: well-circumscribed (W) with the nodule located centrally in the lung without any connection to vasculature; vascularized (V) with the nodule located centrally in the lung but closely connected to neighboring vessels; juxta-pleural (J) with a large portion of the nodule connected to the pleural surface; and pleural-tail (P) with the nodule near the pleural surface connected by a thin tail. Sample images are shown in Fig.1, with the nodule encircled in red.

Computed tomography (CT) is the most accurate imaging modality to obtain anatomical information about lung nodules and the surrounding structures [8] [9] [10]. In current clinical practice, however, interpretation of CT images is challenging for radiologists due to the large number of cases. This manual reading can be error-prone and the reader may miss nodules and thus a potential cancer. Computer-aided diagnosis (CAD) systems would be helpful for radiologists by offering initial screening or second opinions to classify lung nodules [11] [12]. CADs provide depiction by automatically computing quantitative measures, and are capable of analyzing the large number of small nodules identified by CT scans.

A. Related Works

While many studies have reported the detection and segmentation of lung nodules [8] [9] [13] [14] [15], there are limited data in lung nodule classification. Farag et al. reported on some of the initial studies in the classification problem [16]. We suggest, however, that improved performance could be obtained by better feature design and a more advanced classifier. In recent work [17], we designed an overlapping nodule identification procedure to help the classification, but this work mainly focused on identifying the nodules located in the intersections among different types. In prior work from our

This work was supported in part by the ARC grants. Heng Huang was partially supported by US NSF IIS-1117965, IIS-1302675, IIS-1344152.

F. Zhang, Y. Song and W. Cai are with the Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, NSW 2006, Australia.

M. Lee is with the BMIT Research Group, School of Information Technologies, University of Sydney, NSW 2006, Australia, and Royal Prince Alfred Hospital, NSW 2050, 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.

H. Huang is with Department of Computer Science and Engineering, University of Texas, Arlington, TX 76019, USA.

S. Shan is with School of Software, Dalian University of Technology, Liaoning 116620, China.

M. J. Fulham is with the Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, NSW 2050, Australia, and Sydney Medical School, University of Sydney, NSW 2006, Australia.

D. D. Feng is with BMIT Research Group, School of Information Technologies, University of Sydney, NSW 2006, Australia, and with the Center for Multimedia Signal Processing, Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong, and also with Med-X Research Institute, Shanghai Jiaotong University, Shanghai 200030, China.

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

group we suggested that contextual information surrounding the lung nodules could be incorporated to improve nodule classification [18]; however, this method required a complicated segmentation process. Contextual information refers to the complicated anatomical structures around the nodules, within which some structures are only present in certain type of nodules, and some are common across more than one type. For example, W and V nodules are similar in location and shape, which makes it difficult to distinguish them, merely based on the nodule information. V nodules are closely connected to the neighboring vessels and W nodules are isolated from other structures, so identifying connected vessels from V nodules provides an important clue to separate these from each other. Contextual patterns are similarly important for the other nodule types and patch-based approaches can be effective in tackling such a problem.

A patch-based approach, which is based on partitioning the original image into an orderless collection of smaller patches, is usually used to construct the bag-of-feature (BOF) model [19]. Compared to an overall description of the image, patch-based methods can capture local details to better represent the heterogeneous structures. Thus, such methods can be suitable for images of lung nodules that usually contain different anatomical structures. To date, patch-based approaches have been mainly designed for some medical and general imaging problems [19] [20] [21] [22] [23] other than lung nodule classification. Patch division methods normally divide the image into square patches or into circular sector patches. The partitioning is performed in a rigid manner, *i.e.*, the shape and size of each patch are pre-defined, which unavoidably groups unrelated objects together. In particular, a rigid partition on lung nodule images would mix different contextual structures in one patch and cause difficulty in context description.

Ideally, one patch should depict a single type of anatomical structure, and the formulation of patches should be adaptive to the local information. Superpixel formulation [24] [25] [26] provides an efficient tool for this aim, within each patch the pixels are closely related (*e.g.*, similar intensities). Among the various approaches, quick shift [25] has been successfully employed in medical imaging analysis [18] [27] [28]. In particular, it is more suitable for the lung nodule image analysis due to its ability to better capture the irregular contextual structures than other methods, *e.g.*, simple linear iterative clustering (SLIC) [26], which tends to generate more regular superpixels with similar size and shape. In addition, quick shift is a faster algorithm than some other methods, *e.g.*, mean shift [24], so that it could provide increased efficiency for lung nodule image classification that normally involves a large number of cases. Although quick shift shows its advantages on handling lung nodule image by incorporating the local spatial information and reducing spurious labeling due to noise, direct use of quick shift on such small images would group heterogeneous contents together. Therefore, the original quick shift process needs improvement to obtain better patches.

Then, to describe the patches numerically, a patch could be translated into a feature vector [14]. One straightforward feature is grey-level distribution represented as a histogram, which has been effectively used by a number of investigators

[29] [30] [31] to describe intensity variations. Also, the filter-based feature extraction techniques, such as Maximum Response 8 (MR8) [22] [32], are also widely applied to highlight specific image information to identify edges and shapes.

Although these conventional feature techniques have provided a great benefit, more advanced computer vision methodologies have now been incorporated into medical imaging analysis, such as scale-invariant feature transform (SIFT) [4] [16] [17] [33] [34], which is invariant to image translation, scaling, rotation and illumination changes, and robust to local geometric distortion; local binary pattern (LBP) [35] [36], which provides the texture description of objects by incorporating multi-scale and rotation-invariant property; and histogram of oriented gradients (HOG) [20] [27], which represents objects by occurrences of gradient orientation in local portions. Although such feature extraction techniques demonstrate effectiveness individually, better outcomes could still be obtained by improving the feature design.

Following the feature extraction stage, a classifier is needed to label the feature descriptors for image classification. Usually, feature-based image classification is based on supervised learning approaches [15] [17] [35] [36] [37], and the most commonly used classifiers include support vector machine (SVM), *k*-nearest neighbor (*k*-NN), *etc.* Among those, SVM has proven to be a highly effective classifier [38] [39] [40]. However, for lung nodule image classification, SVM could be error prone due to the overlapping feature spaces of the nodules [17]. Incorporating surrounding anatomical context into the feature vector would be helpful. However, since some contextual structures among inter-type nodules are similar, direct classification based on the features extracted from these structures would still be problematic. There are also topic-based models [23] [41] [42] [43] to incorporate the context information, building the relationship between the context and its label by extracting indirect knowledge. Specifically, probabilistic latent semantic analysis (pLSA) [44], which was originally used in the linguistic scenario, could extract the latent semantic topics, and further classify the image based on these hidden topics. The reason that pLSA is suitable for contextual analysis is that it tries to find the common topics shared by various contexts, and then determine the category upon these topics, instead of using the raw feature descriptors directly. Therefore, it is useful to differentiate the overlapping contextual structures shared among various types of lung nodules.

B. Outline Of The Proposed Method

In light of the above, this paper presents a novel image classification method for the four common types of lung nodules. We suggest that the major contributions of our work are as follows: i) a patch-based image representation with multi-level concentric partition, ii) a feature set design for image patch description, and iii) a contextual latent semantic analysis-based classifier to calculate the probabilistic estimations for each lung nodule image.

More specifically, a concentric level partition of the image is designed in an adaptive manner with: (1) an improved

superpixel clustering method based on quick shift is designed to generate the patch division; (2) multi-level partition of the derived patches is used to construct level-nodule (*i.e.*, patches containing the nodules), and level-context (*i.e.*, patches containing the contextual structures). A concentric level partition is thus constructed to tackle the rigid partitioning problem.

Second, a feature set of three components is extracted for each patch of the image that are as follows: (1) a SIFT descriptor, depicting the overall intensity, texture, and gradient information; (2) a MR8+LBP descriptor, representing a richer texture feature incorporating MR8 filters before calculating LBP histograms; (3) a multi-orientation HOG descriptor, describing the gradients and accommodating rotation variance in a multi-coordinate system.

Third, the category of the lung nodule image is finally determined with a probabilistic estimation based on the combination of the nodule structure and surrounding anatomical context: (1) SVM is used to compute the classification probability based on level-nodule; (2) pLSA with contextual voting is employed to calculate the classification probability based on level-context. The designed classifier can obtain better classification accuracy, with SVM capturing the differences from various nodules, and pLSA further revising the decision by analyzing the context.

The structure of this paper is as follows. Sections II, III and IV describe the three stages of the proposed method concentric level construction, feature extraction, and context analysis classification. Section V introduces the experiment dataset and the outline of the experiment design, and Section VI presents the experimental results and discussions. Finally, the conclusion and future work are given in Section VII.

II. CONCENTRIC LEVEL PARTITION

Our method is built upon a patch-based image representation. The current approaches are usually based on patches with fixed shape and size, such as dividing the image into the square patches [22] [23] or into circular sectors based on radial partitions [20] [21] with a predefined number of pixels in these areas. However, such rigid partition methods would unavoidably group unrelated pixels together, as illustrated in Fig.2(a).

Ideally, pixels in the same patch should share similar information, such as intensities, as shown in Fig.2(b). Therefore, we designed an adaptive patch partitioning method formulating superpixels using an improved quick shift clustering method. Then, a concentric level partition model is constructed based on the distances from patches to the centroid of the lung nodule. The shape and size of our patches are derived adaptively according to the local intensity variation, instead of being predefined by rigid partitioning.

A. Superpixel Formulation

Superpixel formulation is the process of dividing an image into multiple segments, which can incorporate local spatial information and reduce spurious labeling due to noise [18]. This can perfectly fit our requirement of patch partitioning. In particular, quick shift is a mode seeking algorithm that

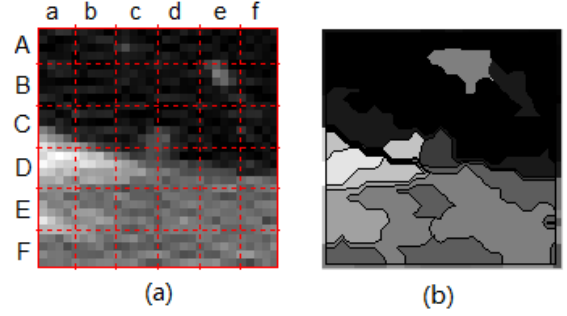


Fig. 2. Comparison between a rigid and adaptive partition on a juxta-pleural nodule. (a) Rigid square partition, which might group irrelevant structures together. Taking lattice (D, e) as an illustration, half of it is pleural surface and the remaining part belongs to parenchyma. Also, the lung nodule is seen over four different lattices, *i.e.*, (C, c), (C, d), (D, c) and (D, d). Such a partition would obscure the subject of the patch, *e.g.*, (C, e) is clearly parenchyma and (E, e) is the pleural surface, but it is difficult to determine what (D, e) expresses. (b) Superpixel formulation by clustering the intensity-related pixels together.

can be used to partition an image into a set of superpixels forming a tree of links to the nearest neighbor which increases an estimate of the density. However, due to the small size of lung nodules, a poor partition is often obtained when directly applying the quick shift method. To tackle this problem, we employ quick shift in an iterative way with image amplification and downsampling.

At the first stage, the image is amplified with nearest neighbor interpolation [45]. A problem with direct use of quick shift on the original nodule image is that the image is so small that only a few pixels could present a particular anatomical structure, leading to the possibility of incorporating trans-regional pixels, as illustrated in Fig.3(a). The bottom-left area in the sample image, corresponding to the white patch in (a), should be further divided because of the high contrast (white and grey). To obtain such an effect, our idea is to amplify the original image based on the local intensity information. Quad-amplification generated the best performance by amplifying the image twice with twofold amplification each time through the experiments.

Then, the quick shift method¹ is applied to the amplified image in an iterative way. Two parameters are introduced in quick shift [46]: *kernelsize*, the size of the kernel used to estimate the density, and *maxdist*, the maximum distance between points in the feature space that may be linked if the density is increased. Fixing them at particular values, *i.e.*, the best performing parameter settings that obtain the highest classification rate with the standard quick shift, generates too many patches, as shown in Fig.3(b). Dividing the image into too many patches would not only separate the integrated anatomical structure but also reduce the efficiency of the method, specifically during the feature extraction stage which extracts the feature set individually for each patch. Therefore, an iterative scheme is designed to handle this problem by increasing the initial values step by step to combine the clusters obtained in the previous iteration. Both *kernelsize* and *maxdist* were initialized at 2, increased by 0.3 and 1

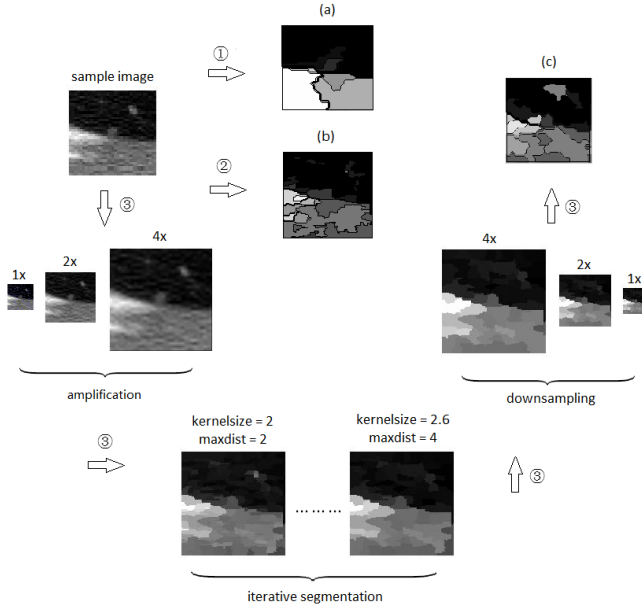


Fig. 3. Procedure for the superpixel formulation scheme with amplification, iterative formulation, and downsampling stages with illustrations reflecting the three approaches: (1) direct use of quick shift on the original image, (2) applying quick shift based on amplification and downsampling without iterative formulation, (3) our proposed scheme. Using under-formulation too few patches are generated to build the multi-level partition (shown in (a)); with the over-formulation, (shown in (b)), there are so many patches produced that individual patch contains insufficient information; with our method, (shown in (c)), there is a more suitable superpixel formulation result for the next step of concentric level partitioning.

respectively in each of the three iterations in our experiments.

Lastly, the downsampling stage is employed to restore the superpixel image to the original size. The clusters are thus the desired image patches. The illustration of the proposed superpixel formulation approach is shown in Fig.3.

B. Concentric Level Partition Construction

Next, we divide the patches in one image into multiple concentric levels, based on the distances between the patches and the centroid of the nodule patch. The nodule patch is the patch that contains the nodule centroid, which is given in the dataset.

For one image I comprising of O patches $PA = \{pa_o : o = 1, \dots, O\}$, we define L as the total number of concentric levels, with levels $LV \in \{lv(l) : l = 0, 1, \dots, L\}$ in which $lv(l)$ contains the patches whose distances are l to the nodule patch. Here, the distance refers to the smallest number of patches passed from patch pa_o to the nodule patch. For $lv(0)$, it comprises only one patch which contains the centroid of the lung nodule; for $lv(l) (l > 0)$, it comprises the immediate outside neighboring patches of $lv(l-1)$. To facilitate contextual analysis, we divided the various levels into two categories: level-nodule, which is $lv(0)$ composed by the lung nodule patch, and level-context, which is $lv(l) (l > 0)$ composed by the context patches.

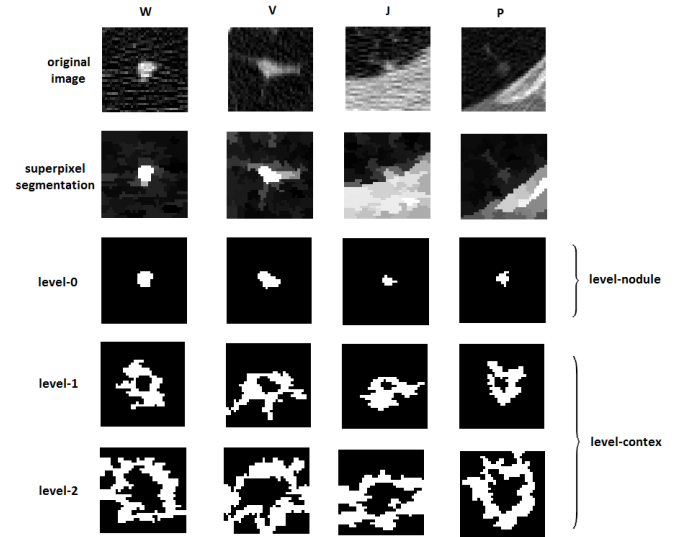


Fig. 4. Sample concentric level partition for each type shown in columns. The first 2 rows give the original images and superpixel formulation, as labelled, and the next 3 rows show the partition outcomes constituted by level-nodule, containing lung nodule patches, and level-context, containing context patches.

While level-nodule tends to represent the lung nodule for each image, level-context tries to indicate different surrounding anatomical structures. Taking the sample images in Fig.4 as examples, $lv(1)$ of type W, V, J, and P represents the (1) parenchyma (2) vessel, (3) pleura and parenchyma, and (4) pleural tail and parenchyma patches, respectively. For $lv(2)$, while type W and J contain the same structures with $lv(1)$, type V and J contain parenchyma patches and pleura and parenchyma patches.

Whether level-nodule and level-context can capture the nodule and surrounding context is crucial for describing the lung nodule image, in which level-nodule contributes more to the category decision (as discussed in Section VI.A). On the one hand, the level-nodule could exactly include the whole nodule if it appears isolated from other structures, such as type W and P, because the nodules usually have high contrast with the surrounding anatomical structures. On the other hand, the level-nodule might cover other undesirable structures if the nodule is very similar to the surrounding tissues, such as type V and J. In these circumstances, the over-segmentation property of quick shift-based approach can better describe the nodule patch by extracting the central region of the nodule which is normally used as the most significant characteristic to differentiate various nodules [16] [17] [18]. However, this would also introduce the problem that part of nodule will be incorporated into contextual patches. Furthermore, the reverse problem that level-nodule contains some surrounding tissues would emerge as well so that level-nodule and level-context might not precisely depict the corresponding structures. Context analysis classification (in particular the level type identification and contextual voting, introduced in Section IV) is designed to tackle these problems by discriminating the different combination of contextual structures.

¹The quick shift package is from VLfeat project, downloaded from: <http://www.vlfeat.org/index.html>

III. FEATURE EXTRACTION

The effectiveness of image feature description depends on: distinction and invariance [16], which means that the descriptor needs to capture the distinctive characteristics and be robust to adapt to the various imaging conditions. Based on our visual analysis the lung nodules, we suggest that intensity, texture and gradient can characterize the various nodules and the diverse contextual structures. We thus designed the feature set of the combination of SIFT for overall description, MR8+LBP for texture, and multi-orientation HOG for gradient. For convenience, we refer to this feature set as the FS3 feature.

Formally, denote an image as I comprising of O patches $PA = \{pa_o | o = 1, \dots, O\}$. The FS3 feature $fs3(pa_o)$ is extracted from each patch pa_o , as:

$$fs3(pa_o) = \{SIFT(pa_o), MR8+LBP(pa_o), MHOG(pa_o)\} \quad (1)$$

where $SIFT(pa_o)$, $MR8+LBP(pa_o)$ and $MHOG(pa_o)$ are the three component descriptors. The three sections of FS3 feature are adjusted into the same scale by linear rescaling so that they have similar effects on the feature description.

A. SIFT Descriptor For Overall Description

The SIFT process generates a 128-length vector for each key point. Since SIFT is invariant to image translation, scaling, rotation and illumination changes, and robust to local geometric distortion, it provides valuable lung nodule data [4] [16] [17]. SIFT is robust and is able to carry out semantic classification due to its ability to capture the texture and gradient information [47] [48]. Besides, it identifies the key points by computing extremum pixels in the image local area to incorporate the intensity information. Thus, SIFT² descriptor was adopted as the first component of FS3 to give an overall description from intensity, texture and gradient perspectives.

In our case, we extract only one 128-length vector near the centroid of each patch. Specially, since the shapes of the extracted patches are not uniform, we selected the smallest rectangle sub-window to cover all pixels for each patch and then ran SIFT on this window. The final SIFT descriptor $SIFT(pa_o)$ of patch pa_o is calculated by selecting one key point near the centroid of the rectangle widow.

B. MR8+LBP Descriptor For Texture Description

The combination of MR8 filters and LBP feature is designed to provide richer texture description of patches by incorporating multi-scale and rotation-invariant properties. LBP is a powerful feature for texture based image classification [35] [49]. Although LBP can be easily configured to describe the local texture structure with multi-resolution and rotation-invariance, it captures too many trivial image variations. Therefore, we incorporate the MR filter set before computing LBP histogram. The MR set contains both isotropic and anisotropic filters at multiple orientations and multiple scales and records the angle

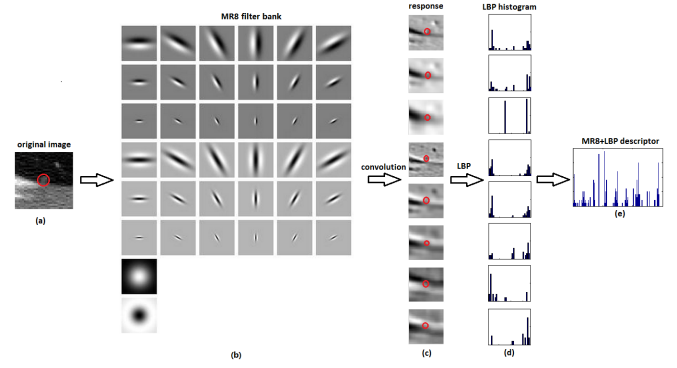


Fig. 5. Illustration of the generation of the MR8+LBP descriptor: (a) the original image with the target patch pa_o with a red circle, (b) 38 filters with 6 orientations at 3 scales for 2 oriented filters, and 2 isotropic filters ($6 \times 3 \times 2 + 2$), (c) 8 image filter responses with 3 scales for 2 filters and 2 isotropic filters ($3 \times 2 + 2$), (d) histograms of the select patch pa_o in the 8 responses by computing the LBP descriptor, (e) the final MR8+LBP descriptor produced by concatenating all histograms.

of maximum response, which makes it possible to discriminate textures that appear to be very similar.

Specifically, MR8 bank [50] is used in our method, which consists of 38 filters but produces only 8 filter responses by recording only the maximum filter response across all orientations for one scale. This yields rotation invariance. The final filter response contains 2 anisotropic filters for each of 3 different scales and 2 isotropic filters ($2 \times 3 + 2$). MR8 filters are directly applied to the original image, as shown in Fig.5. For image I , we get 8 filter responses represented by $I_{MR8}(f)$ where $f \in [1, 8]$.

Next, as for patch pa_o in image I , LBP descriptor $LBP(pa'_o)$ is computed for the corresponding patch pa'_o in each filter response of I_{MR8} . As explained by Song et al. [51], the total number of possible values is 36, and hence the histogram of each response patch contains 36 dimensions, as shown in Fig.5(d). With 8 filter responses, patch pa_o in image I thus gives 8×36 -dimension histograms. All histograms are concatenated to obtain the final MR8+LBP descriptor $MR8+LBP(pa_o)$ for patch pa_o , as shown in Fig.5(e). This generates a 288-length vector.

C. Multi-orientation HOG Descriptor For Gradient

Gradient distribution provides helpful supplementary information to texture for discriminating various anatomical structures in nodule images. Among various gradient-based methods, HOG is being widely used and can also improve performance considerably when coupled with LBP [27]. However, unlike SIFT and MR8+LBP descriptors, the raw HOG descriptor cannot handle rotation-invariant problems. Therefore, we designed a multi-orientation HOG descriptor inspired by our previous work [51] to provide further an advanced gradient description in addition to that from SIFT. The designed descriptor is adaptive to the locations of patches relative to the centroid of the nodule, rather than having the same initial orientation for all patches [51].

Assuming that the center of patch pa_o is c_{pa_o} , we built 8 coordinate systems that share the same origin c_{pa_o} but have

²The SIFT package is from VLFeat project, downloaded from: <http://www.vlfeat.org/index.html>

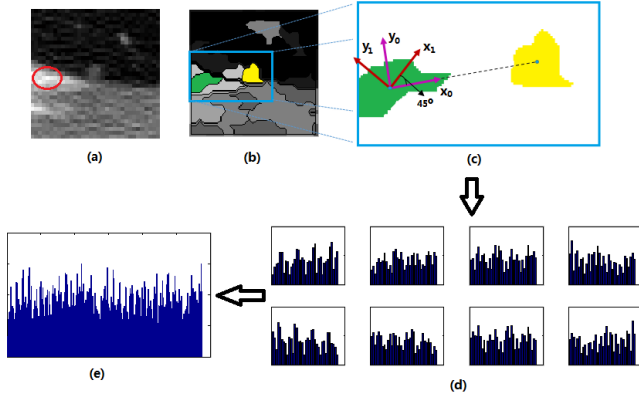


Fig. 6. Depiction of the generation of multi-orientation HOG descriptor: (a) the original image with target patch in red circle; (b) the location relationship between the target patch (green area) and the lung nodule patch (yellow area); (c) the design of the multi-coordinate systems with blue dashed line as the initial orientation of the first coordinate system and 2 of the 8 coordinate systems are shown with purple and red arrows; (d) the derived gradient histogram for each coordinate system by computing the HOG descriptor and, (e) the final multi-orientation HOG descriptor by concatenating all histograms.

different initial orientations (0 degree). A visual illustration of the multi-coordinate systems is shown in Fig.6(c). Two of them are shown with (x_0, y_0) and (x_1, y_1) . Contra-rotating the first coordinate system (*i.e.*, (x_0, y_0)) by 45 degree generates the next one (*i.e.*, (x_1, y_1)). Instead of predefining the initial orientation of the first coordinate system, we set it as the direction (blue dash line) from the centroid of the patch (green area) to the centroid of lung nodule (yellow area).

Next, for each coordinate system, patch pa_o is divided into 9 cells, within which gradient orientations of the pixels in 9 undirected histograms are counted to encode the gradient distribution. Instead of adopting the histogram statistics directly, we apply the *UOCTTI* variant from Felsenszwalb et al. [52] which computes the *L1* norm of each of the *L2* normalized undirected histograms to get a four-dimensional texture-energy feature for each coordinate system, as shown in Fig.6(d). As a result, for patch pa_o , we obtain a 288-length (8 systems \times 9 cells \times 4-D features) multi-orientation HOG descriptor $MHOG(pa_o)$, as shown in Fig.6(e).

IV. CONTEXT ANALYSIS CLASSIFICATION

With the concentric level partition and feature set, the next stage is to label each image with one of the four nodule categories. Considering that the morphology of lung nodules forms a continuum [17], which means the structures of lung nodules among different categories are similar, even with the comprehensive feature design, it remains difficult to classify the images precisely. So to aid classification we incorporated the contextual information.

The proposed method involves SVM analysis for lung nodule patches, and pLSA analysis for context patches. In a supervised manner, besides the explicit label information (with SVM), we also extracted the implicit latent semantic information hidden in the relationship between the images and their categories (with pLSA). In this way, the training data are used twofold, which acquires much more information.

The first step is lung nodule probability estimation using SVM. This step works on level-nodule that focuses on lung nodule description. The proposed feature sets are extracted for all patches in level-0, and the SVM classification procedure is performed with a probability estimate [17]. For each lung nodule image I , we thus compute its probability of each of the four types $TP = \{tp_t | t \in \{w, v, j, p\}\}$ based on level-nodule, called the level-nodule probability, as:

$$p_{level-nodule}(tp_t|I) = P_{svm}(tp_t|I) \quad (2)$$

where $P_{svm}(tp_t|I)$ is the probability estimates from SVM. Specifically, a four-type SVM was trained with polynomial kernel by *C-SVC* (with the default parameters, *i.e.*, $\gamma = 1/\text{number of features}$, $\text{coef0} = 0$, and $\text{degree} = 3$) from [47] in our experiments.

The second step is the context probability estimates using the proposed topic-based classification method. Topic model was originally used for natural language processing by extracting the semantic topics between the documents and words. The underlying idea is that each document can be considered as mixture of topics, where the topic is a probability distribution of words. pLSA [44] is one of the techniques to extract the latent semantic topics hidden between documents and words, which means that it infers the visual topics, as illustrated in Fig.7. As mentioned in Section II.B, each level lv in level-context tends to represent certain anatomical structures, which can be used to determine the type of lung nodule. pLSA is thus used to identify the potential type of level lv , called level type $LTP = \{ltp_t : t = \{w, v, j, p\}\}$, *i.e.*, to which type level lv belongs, by calculating the probability of level lv given certain level type ltp_t , called level type probability $P(ltp_t|lv)$. Specifically, in level type identification process (introduced Section IV.A), treating the levels as documents and the patches as words (shown in Fig.7(b)(e)), the representation of level in terms of latent semantic topics can be derived by pLSA, and the probabilities of the types upon these topics can be obtained in the training stage. Therefore, level type probability is calculated based on these two components.

The combination of all obtained level type probabilities of context levels $lv(l) (l > 0)$ can be used to describe the level-context. However, the level types might overlap because similar anatomical structures can be shared among the nodule types. Therefore, rather than a simple concatenation, the contextual voting (introduced in Section IV.B) is designed to obtain the combined level context. The probability $P_{level-context}$ of the lung nodule image given the four types is then derived based on the voting result.

The final step is to calculate probability of image I given the type tp_t through level-nodule and level-context probabilities with a weighted parameter $\lambda \in (0, 1)$, as:

$$P(tp_t|I) = \lambda * p_{level-nodule}(tp_t|I) + (1 - \lambda) * p_{level-context}(tp_t|I) \quad (3)$$

The nodule image I is classified into the type tp_t that has the highest probability.

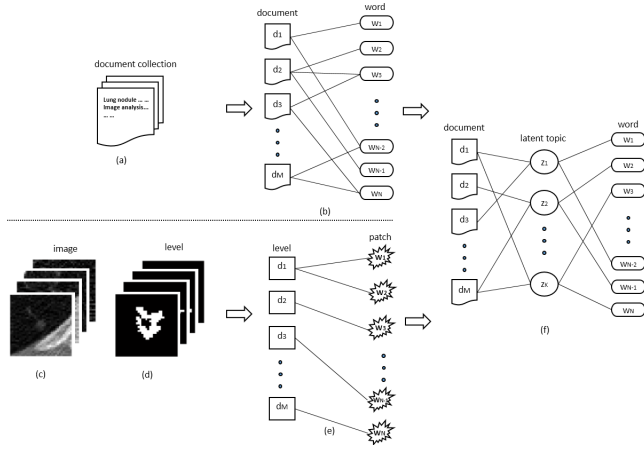


Fig. 7. Outline of applying pLSA for level type discovery: (a) document collection; (b) the bipartite relationship between documents and words constructed by co-occurrence of words on documents; (c) the image collection; (d) the level collection by extracting the same level from images; (e) the bipartite relationship between levels and patches constructed by co-occurrence patches in levels and, (f) latent semantic topics discovery by using pLSA.

A. Level Type Identification With pLSA

As for the level type identification stage, the first step is dictionary construction. Here, we apply the conventional k -means clustering strategy to all patches in the same level across the whole dataset to construct the dictionary.

Next, latent semantic topics for certain level are extracted by pLSA³. Assuming there are M images and the dictionary size is N , for each level- l we could obtain the dataset of M levels $LV = \{lv_m : m = 1, \dots, M\}$ represented by subsets of N patches $PA = \{pa_n : n = 1, \dots, N\}$. The dataset can be summarized by a co-occurrence matrix X of size $M \times N$, where $X(lv_m, pa_n)$ denotes the weighted occurrences of patch pa_n appeared in level lv_m .

With the co-occurrence matrix X , the occurrence of patches in a set of levels can be interpreted by a reduced size of hidden variables, i.e., latent semantic topics, $Z = \{z_k : k = 1, \dots, K\}$ through pLSA. Formally, the occurrence of patch pa_n in level lv_m can be represented as:

$$P(pa_n, lv_m) \cong \sum_k P(pa_n|z_k) * P(z_k|lv_m) * P(lv_m) \quad (4)$$

We use the Expectation-Maximization (EM) [44] algorithm to maximize the likelihood Lh so that we can learn the parameters Z ,

$$Lh = \prod_m \prod_n P(pa_n, lv_m)^{X(pa_n, lv_m)} \quad (5)$$

Then, the relationship between latent topics z_k and level types ltp_t can be estimated:

$$P(ltp_t|z_k) \cong \sum_m P(ltp_t|lv_m) * P(lv_m|z_k) \quad (6)$$

where $P(ltp_t|lv_m)$ is obtained through training dataset and $P(lv_m|z_k)$ is obtained through solving Eq.(5).

The above $P(ltp_t|z_k)$ can be regarded as a weighting factor of latent topic z_k for determining the level type ltp_t . On the

whole, the proportion of the factor should be increased if the number of times z_k appears in one type, but is offset by the frequency of z_k in the whole dataset. For instance, the parenchyma patches appear commonly on all level-contexts, which are usually inconsequential for topic prediction, especially for type J and P. Hence, we need to find a way to control the case that some latent topics are generally more common than others.

To manage this problem, we make an adjustment of $P(ltp_t|z_k)$ based on the term frequency-inverse document frequency (TF-IDF) [53] algorithm. For latent topic z_k , we firstly compute its TF value in determining level type ltp_t with the following equation,

$$TF(ltp_t, z_k) = \frac{P(ltp_t|z_k)}{\sum_t P(ltp_t|z_k)} \quad (7)$$

Next, IDF value is calculated to measure whether the latent topic is common or rare across the whole dataset,

$$IDF(z_k) = \log\left(\frac{\sum_k P(ltp_t|z_k)}{1 + P(ltp_t|z_k)}\right) \quad (8)$$

At last, the conditional probability of type ltp_t given the latent topic z_k is obtained by calculating the TF-IDF value,

$$\begin{aligned} P(ltp_t|z_k) &= TF-IDF(ltp_t, z_k) \\ &= TF(ltp_t, z_k) * IDF(z_k) \end{aligned} \quad (9)$$

Above all, for a unlabeled level lv' at level- l , we calculate its conditional probability of level types, i.e., $P(ltp_t|lv')$, which is also called level type probability $P_{level-type}(ltp_t|lv')$:

$$P(ltp_t|lv') \cong \sum_k P(ltp_t|z_k) * P(z_k|lv') \quad (10)$$

B. Contextual Voting

As for the contextual voting stage for L -level-context, there are 4^L voters voting for 4 candidates, where voters are level type sequences S and candidates are lung nodule types TP . Here, level type sequence is the combination of level types from the inside to outside level, which has 4^L possible combinations. For instance, if a 2-level partition is obtained for image I with $lv(1)$ labeled as type W and $lv(2)$ labeled as type J, the level-context of image I is encoded by type sequence W-J.

The level-context probability is determined by two parts: credit and voting score. Credit value is defined for one level type sequence, and describes its reliability on determining the lung nodule type. Formally, for any sequence s_q in sequence set $S = \{s_q : q = 1, \dots, Q, Q = 4^L\}$, its credit on nodule type tp_t can be formulated as:

$$Cre(tp_t, s_q) = co-occur(tp_t, s_q) / occur(s_q) \quad (11)$$

where, $co-occur(tp_t, s_q)$ is the number of occurrences of sequence s_q in images of type ltp_t , and $occur(s_q)$ is the total number of occurrence of sequence s_q . Credits are computed from training dataset.

Then, voting score is defined as the likelihood of the image representing a certain nodule type based on the level type probability. Formally, for any sequence $s_q = \{ltp_t(l) : l =$

³The pLSA package is downloaded from <http://www.robots.ox.ac.uk/~vgg/software/>

$1, \dots, L\}$, its voting score towards type tp_t for image I can be formulated as:

$$Sco_{tp_t}(s_q) = \prod_l P(ltp_t(l)|lv(l)) \quad (12)$$

Finally, a weighted scheme for computing level-context probability is formulated as:

$$P_{level-context}(tp_t|I) = \max_q \left(\cos\left(\frac{Rank(Sco_{tp_t}(s_q))}{Q} * \frac{\pi}{2}\right) * \exp(Sco_{tp_t}(s_q) * Cre(tp_t, s_q)) \right) \quad (13)$$

where $Rank(Sco_{tp_t}(s_q))$ represents the rank index of $Sco_{tp_t}(s_q)$ among all sequences, *i.e.*, $Sco_{tp_t}(s_q) = i$ if $Sco_{tp_t}(s_q)$ is the i th largest value. The weighted cosine value is used to increase the proportion of high possibility level type sequences.

V. DATASET AND EXPERIMENT DESIGN

We used the publicly available Early Lung Cancer Action Program (ELCAP) [54] database for experiments. The ELCAP database contains 50 sets of low-dose CT lung scans with 379 unduplicated lung nodules annotated at the centroid. An expert reader of the images was invited to classify the nodules into well-circumscribed (W-15%), vascularized (V-16%), juxta-pleural (J-30%) and pleural-tail (P-39%). As observed in this problem domain, the manual classification done by experts is objective according to the visual inspections of the four nodule types. Based on experience, the readers can immediately recognize the anatomical structures that are very similar. The available image viewing software can be used to adjust the images, *e.g.*, the size, view and contrast, windowing, *etc.*, so that the noise and undesirable structures can be eliminated, leading to a more accurate interpretation. Therefore, we suggest that the inter-observer variability would be low. The nodules in this dataset were relatively small, with an axial size across the centroid of 4×4 pixels on average. Thus to restrict the scope to lung nodule classification with contextual information only, a sub-window of 33×33 pixels was cropped from each image slice with the annotated centroid of the nodule appearing in the center. This is similar to the handling in the related works by other researchers [4] [17] [18].

A set of 80 randomly selected images, with 20 for each type, was used for training. The training set was used for three tasks: (1) level-nodule predication through SVM with probability estimates, (2) level type probability calculation based on discovered latent topics, and (3) credit computation for each level type sequence. During the experiments, the selected training images were also included in the testing stage for the proposed method as well as all control methods. In this way, the global dictionary for contextual analysis classification could be obtained through the whole dataset, and more images were incorporated for testing.

The experiments included the following aspects: (1) parameter setting; (2) analyzing the effects of level partition; (3) evaluating the performance of the designed FS3 descriptor; (4) assessing the context analysis classifier. The performance

TABLE I
APPROXIMATE ANATOMICAL CONTEXT OF LEVEL PATCH FOR EACH TYPE

Type	W	V	J	P
lv-0	no	no	no	no
lv-1	pa	pa/ ve	pa/ pl	pa/ ve/ plt
lv-2	pa/ pl	pa/ ve	pa/ pl	pa/ pl
lv-3	pa/ pl/ ve	pa/ ve	pa/ pl	pa/ pl
lv-4	pa/ pl/ ve/ ots	pa/ ve/ pl/ ots	pa/ pl/ ots	pa/ pl/ ots

no = nodule; pa = parenchyma; ve = vessel; pl = pleura; plt = pleural tail; ots = others

of nodule image classification was done with an overall classification rate, as:

$$\text{Classification Rate} = N_{right}/N_{all} \quad (14)$$

where N_{right} and N_{all} are the numbers of correctly labeled images and all images. Also, to illustrate the performance on images from different types, recall and precision were computed:

$$\text{Recall} = TP/(TP + FN) \quad (15)$$

$$\text{Precision} = TP/(TP + FP) \quad (16)$$

where TP , FN and FP are the numbers of true positive, false negative and false positive classifications of nodule images.

VI. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Parameter Setting

In this section, we first discuss parameter L , the number of total levels of partition. The criterion for the selection of L comes from whether the extracted level-context is descriptive enough to represent the contextual information. Based on our visual inspection, the main anatomical structures of patches at each level are shown in Table I. Note that this is just an approximation because the thorax is a complex structure that includes some other structures located near the nodules. However, it can still give rich information about the selection of L . The final value L is settled at 2 based on the above observation. If L is 1, the level-context would contain too few patches to tell the difference between various types, especially for type V and P because some type V nodules are attached to the vessels with the structures that are very similar to the pleural tails for type P. If L is larger than 2, the higher levels would have similar contexts to the lower ones, which decreases the discriminative ability of the lower levels and reduces the efficiency of our method since the higher levels would include more patches.

The dictionary for constructing the bipartite relationship between levels and patches is calculated by clustering the patches in the same level together. In the experiments, the dictionaries were constructed from the whole dataset (*i.e.*, the testing dataset) so that both training and testing images could share the same vocabulary. With $L = 2$, two global dictionaries were extracted with k -means algorithm, with the parameters $N_{level-1}$ and $N_{level-2}$ indicating the sizes of the two dictionaries, from a total of 2534 and 5937 patches for the two levels in the whole dataset. Fig.8 shows the distribution

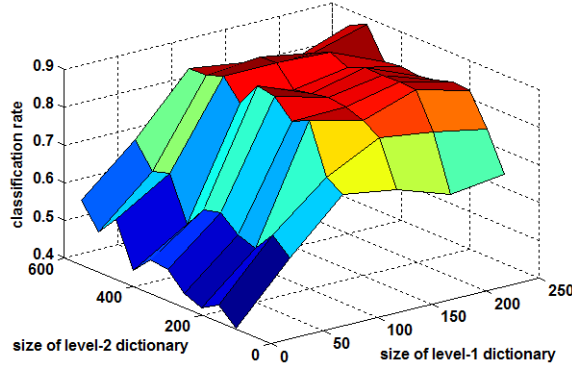


Fig. 8. Distribution of classification rates given sizes of the two dictionaries: the rate tends to be stable after $N_{level-1}$ is 100 and $N_{level-2}$ is 200, shown in the red area.

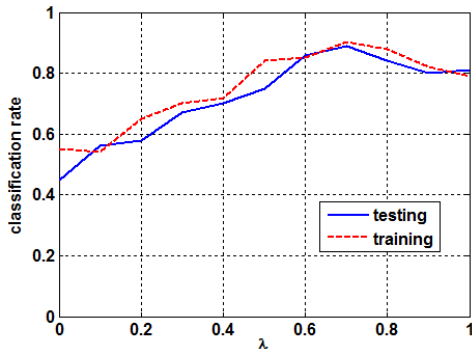


Fig. 9. The distributions of classification rates given parameter λ on the training and testing datasets.

of classification rates on different sizes of level-1 and level-2 dictionaries. The improvement becomes insignificant after $N_{level-1}$ is 100 and $N_{level-2}$ is 200, hence we selected these two values as sizes of the two dictionaries.

As for the parameter K , the number of latent topics, we found a similar classification rate distribution across dictionary size N , just as illustrated in [23], *i.e.*, the classification rate curve tends to be stable as the number of topics increases. We selected $K_{level-1}$ and $K_{level-2}$ with the same values of $N_{level-1}$ and $N_{level-2}$, because the k -means method tries to cluster the related patches into the same group, within which the patches together represent the same topic. By equating N and K , we could map the dictionary and the latent semantic topics together to improve the descriptive ability of latent semantic topics [55].

Finally, the effect of weight parameter λ on the classification results is shown in Fig.9. Although the classification rates on the training dataset tend to be a little higher than that on the testing dataset when λ is smaller (from 0 to 0.5), they are similar when λ is relatively larger (from 0.6 to 1). The similar classification rate distributions for the training and testing datasets suggest the balanced performance of the proposed method on both the labeled and unlabeled images. In addition, it can be observed that combining both level-nodule probability and level-context probability resulted in better performance than using the individual contribution separately.

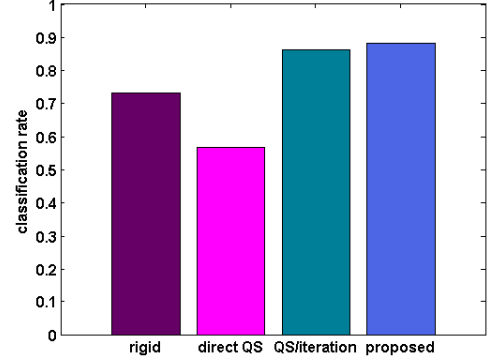


Fig. 10. Comparisons of classification rates from the results of various level partition methods: (1) rigid partition, (2) direct use of quick shift on original image, (3) quick shift based on amplification and downsampling without iterative process and, (4) our method.

Specifically, the best results were achieved when $\lambda = 0.7$, indicating level-nodule contributes more to the final decision. Therefore, λ was fixed at 0.7.

B. Evaluation Of Concentric Level Partition

To illustrate the effect of the proposed concentric level partition, we present the quantitative comparisons among the following methods: the rigid partition, quick shift on the original image, quick-shift based on amplification and downsampling without iterative formulation, and our method. Except for the second approach that extracted too few patches to construct the multi-level partition structure, the other two were handled in the following manner to construct a 2-level partition structure as for the proposed method: (1) For the first one, 3×3 pixels were regarded as one patch. The center patch was used as level-nodule, and the remaining patches were equally divided into level-1 and level-2 to build level-context; (2) For the third one, a 4-level partition structure was formed. The first two context levels and the last two context levels were then combined to derive a 2-level partition. For the second method, we built a 1-level partition, which means the level type probability from level-1 was directly used as level-context probability. As for the two quick shift-based control approaches (the second and third ones), the best classification rates were found across all *kernelsize* and *maxdist* (*kernelsize*: from 1 to 5, *maxdist*: from 1 to 10) at *kernelsize* = 1 and *maxdist* = 1, *kernelsize* = 2.6 and *maxdist* = 5 respectively, which implies the best parameter settings. Quad-amplifications were selected to amplify the image since the smaller amplification was still not enough to present certain structures and larger amplification reduced the efficiency by introducing too many patches. The comparison results are shown in Fig.10.

The rigid partition generated a lower rate than the last two, which proves that concentric level partition based on adaptive patch extraction can obtain better description of the image contextual information, and further improve the classification result. The second approach obtained the lowest classification rate at about 55%, suggesting that the amplification and iterative process can improve classification performance. Also

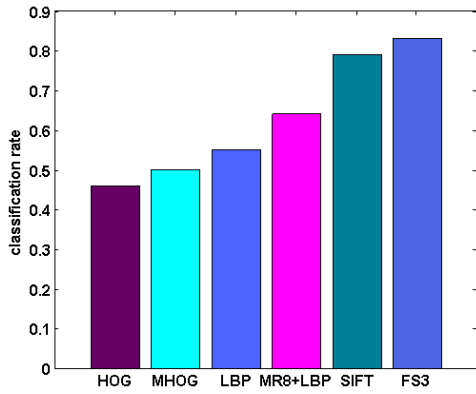


Fig. 11. Comparison of classification rates from results of level-nodule prediction based on different feature descriptors: (1) original HOG; (2) multi-orientation HOG; (3) original LBP; (4) MR8+LBP; (5) SIFT and, (6) FS3.

while the last two had quite similar results, our iterative scheme outperformed the other ones, with a shorter running time due to a smaller number of extracted patches (the third method generated more than 5 times as many patches as those from the proposed method). Except for the advantage of the improved quick shift process, the comparison with the second approach that uses level type probability to directly predict level-context probability also shows the necessity of the contextual voting, as we discussed in Section IV that it is undesirable to estimate the type of lung nodule image directly based on level types because similar contexts could be shared among different types of lung nodules.

C. Evaluation Of Feature Extraction

The designed FS3 feature describes the image content from different aspects, *i.e.*, intensity, texture and gradient, with the three component descriptors. To measure if the combination of them provides an advanced description of the image, we compared it with those three components individually. Also, the original LBP and HOG were tested to show the advantages of the designed MR8+LBP and multi-orientation HOG descriptors. Here, instead of using the proposed classifier, we simply classified the images according to level-nodule probability, which was computed from the nodule patch with SVM. In this way, we hoped to eliminate the influence from the context analysis, and merely focus on feature design comparison. The comparison results are shown in Fig.11.

Raw HOG generated the worst result among all approaches, since it works on uniformly spaced cells instead of focusing on particular points, and thus it is deficient for acquiring the most distinctive characterization among the various patches. The designed multi-orientation HOG descriptor outperformed the raw HOG descriptor, indicating the benefit of handling the rotation-invariant problem for feature description. However, multi-orientation HOG achieved a result with only close to 50% of images accurately classified, suggesting that simply accommodating the rotation problem is still not sufficient to capture the spherical shape of lung nodules.

MR8+LBP descriptor achieved about a 5% improvement on classification rate compared with raw LBP, suggesting the

benefit of MR8 filters in discriminating the similar textures through its multi-orientation and multi-scale properties and its ability to eliminate some unnecessary imaging details. The better performances of LBP-based approaches over the HOG-based approaches indicate that lung nodule structure can be better characterized by textures other than gradients; however, its lower classification rate compared with SIFT suggests integrating the intensities, textures and gradient can provide a better feature descriptor.

The FS3 descriptor obtained the best results, an demonstrated that the feature description with intensity, texture and gradient description of the image content extracted by the designed feature set is more comprehensive for capturing the characteristics of the lung nodule structure and thus is very helpful for enhancing the classification performance.

D. Evaluation Of Classifier

The proposed classifier categorizes the lung image based on the concentric level partition and the extracted feature. The level-nodule probability and level-context probability are combined to classify the images. To evaluate our design of the classifier, comparisons were conducted across our proposed method (PM), the SVM on level-nodule (S1) and the SVM with context analysis (S2) on the overall classification rate and also to measure the recall and precision for each individual type. The second approach is the calculation of level-nodule probability. The third approach contains level-nodule probability and level-context probability; however, the level-context only contains a single level which is composed of all context patches, and a SVM model is trained, based on the distribution of topics calculated with pLSA to compute the level-context probability. The dictionary size N , topic size K , and weighed parameter λ for this approach were settled as 350, 350, and 0.6, which obtained the highest classification rate. Fig.12 illustrates the comparison results.

The classification rates of the two SVM-based approaches were quite similar at about 82% and 83%. The relatively small enhancement with S2 shows the benefit of context analysis, and also indicates the limit of SVM with its direct classification manner for lung nodule context identification. About 7% more of the images were correctly classified using the proposed classifier (89%) than only adopting level-nodule probability (S1, 82%), suggesting that integrating the level-context probability with level-probability is helpful for lung nodule image classification. Also about 6% improvement was obtained over S2, indicating that the designed context analysis scheme based on multi-level partition and contextual voting is more suitable for lung nodule context analysis.

Furthermore, apart from the improvement on overall performance, our method achieved more balanced distributions of recall and precision across the four types (with standard deviations of recall of 0.096, 0.086 and 0.023, of precision of 0.106, 0.104 and 0.052 for S1, S2 and PM respectively). The two SVM-based approaches had similar distribution patterns and standard deviations on recall and precision, but the designed classifier resulted in smaller standard deviations by balancing the distribution over all the four types. This more balanced

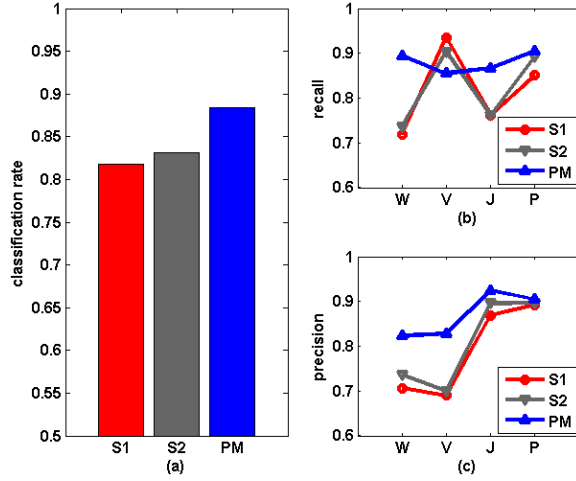


Fig. 12. Comparison between SVM and the proposed classifier: (a) classification rate, (b) recall and, (c) precision.

performance was due to the proposed context analysis, which helped to restore the misclassified nodules based on the level-nodule probability to their actual categories. Table II shows the confusion matrix. It can be observed that the misclassification rates of type W and J to type V are much higher than the others in the two control methods (as shown with bold in Table II). However, the proposed method made the misclassified images evenly distributed by classifying more of these two types nodules correctly.

This enhancement strongly depends on level-type identification and contextual voting. Taking the type W and V as examples, although images from both of them contain a similar level type sequence V-J, the credit of V-J on image type W is 0.36 that is twice as the credit on image type V, showing more images with sequence V-J belong to type W. Therefore, with the help of the voting process, more images in type W were correctly classified, as well as type J. Meanwhile, images that are not clearly distinguished between these three categories would also be more likely to be classified into type W and J, thus more of the actual type V were misclassified than the two control methods, leading to a relatively lower recall of type V. Ultimately, although the misclassification rate of type V to other types was increased in the proposed method, those of the other types were decreased to a larger extent, leading to the more balanced distributions of recall and precision and a better overall classification rate.

VII. CONCLUSIONS

We present a supervised classification method for lung nodule LDCT images in this paper. The four main categories of lung nodules well-circumscribed, vascularized, juxta-pleural and pleural-tail were the objects to be differentiated. We designed a novel method to overcome the problem of the lung nodule overlapping adjacent structures. Our method had 3 components: concentric level partition, feature extraction, and context analysis classification. A concentric level partition was constructed by an improved quick shift superpixel formulation. Then, a FS3 feature set including SIFT, MR8+LBP and

TABLE II
CONFUSION MATRIX ON PERCENTAGE OF THE IMAGES CLASSIFIED INTO OTHER TYPES

	W	V	J	P
	S1/S2/PM	S1/S2/PM	S1/S2/PM	S1/S2/PM
W	71.9/73.7/89.5	14.0/10.5/3.5	7.0/7.0/1.8	7.0/8.8/5.3
V	3.2/3.2/4.8	93.5/90.3/85.5	1.6/3.2/3.2	1.6/3.2/6.5
J	7.1/7.1/2.7	9.7/9.7/4.4	76.1/76.1/86.7	8.8/7.1/6.2
P	4.8/3.4/3.4	4.8/4.8/2.7	5.4/2.7/3.4	85.0/89.1/90.5

multi-orientation HOG was generated to describe the image patch from various perspectives. Finally, a supervised classifier was designed through combining level-nodule probability and level-context probability. The results from the experiments on the ELCAP dataset showed promising performance of our method. We also suggest that the proposed method can be generally applicable to other medical or general imaging domains. For instance, the improved quick shift formulation process could be applied as the preprocessing stage for patch-based imaging analysis; the extracted feature set could be employed as a feature descriptor for other kinds of images; and the latent semantic analysis with the voting process could be used for analyzing hierarchical image patches.

REFERENCES

- [1] J. J. Erasmus, J. E. Connolly, H. P. McAdams, and V. L. Roggli, "Solitary pulmonary nodules: Part I. morphologic evaluation for differentiation of benign and malignant lesions," *Radiographics*, vol. 20, no. 1, pp. 43–58, 2000.
- [2] D. Wu, L. Lu, J. Bi, Y. Shinagawa, K. Boyer, A. Krishnan, and M. Salganicoff, "Stratified learning of local anatomical context for lung nodules in CT images," in *Proc. CVPR*, 2010, pp. 2791–2798.
- [3] R. A. Ochs, J. G. Goldin, F. Abtin, H. J. Kim, K. Brown, P. Batra, D. Roback, M. F. McNitt-Gray, and M. S. Brown, "Automated classification of lung bronchovascular anatomy in CT using adaboost," *Medical Image Analysis*, vol. 11, no. 3, pp. 315–324, 2007.
- [4] A. Farag, S. Elhabian, J. Graham, A. Farag, and R. Falk, "Toward precise pulmonary nodule descriptors for nodule type classification," in *MICCAI LNCS*, vol. 13, no. 3, 2010, pp. 626–633.
- [5] A. A. Farag, "A variational approach for small-size lung nodule segmentation," in *Proc. ISBI*, 2013, pp. 81–84.
- [6] D. Xu, H. J. van der Zaag-Loonen, M. Oudkerk, Y. Wang, R. Vliegenhart, E. T. Scholten, J. Verschakelen, M. Prokop, H. J. de Koning, and R. J. van Klaveren, "Smooth or attached solid indeterminate nodules detected at baseline CT screening in the NELSON study: Cancer risk during 1 year of follow-up," *Radiology*, vol. 250, no. 1, pp. 264–272, 2009.
- [7] S. Diciotti, G. Picozzi, M. Falchini, M. Mascalchi, N. Villari, and G. Valli, "3-D segmentation algorithm of small lung nodules in spiral CT images," *IEEE Trans. Information Technology in Biomedicine*, vol. 12, no. 1, pp. 7–19, 2008.
- [8] B. Zhao, "Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm," *Journal of Applied Clinical Medical Physics*, vol. 4, no. 3, pp. 248–260, 2003.
- [9] Y. Lee, T. Hara, H. Fujita, S. Itoh, and T. Ishigaki, "Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique," *IEEE Trans. Medical Imaging*, vol. 20, no. 7, pp. 595–604, 2001.
- [10] A. A. Farag, S. Y. Elhabian, S. A. Elshazly, and A. A. Farag, "Quantification of nodule detection in chest CT: A clinical investigation based on the ELCAP study," in *The Second International Workshop on Pulmonary Image Processing*, 2010, pp. 149–160.
- [11] S. G. Armato III, M. L. Giger, and H. MacMahon, "Automated detection of lung nodules in CT scans: preliminary results," *Medical physics*, vol. 28, no. 8, pp. 1552–1561, 2001.

- [12] L. Fan, C. L. Novak, J. Qian, G. Kohl, and D. Naidich, "Automatic detection of lung nodules from multislice low-dose CT images," in *Proc. SPIE, Medical Imaging 2001*, 2001, pp. 1828–1835.
- [13] A. A. Farag, J. Graham, A. y A. Farag, S. Elshazly, and R. Falk, "Parametric and non-parametric nodule models: Design and evaluation," *Third International Workshop on Pulmonary Image Processing*, pp. 151–162, 2010.
- [14] I. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken, "Computer analysis of computed tomography scans of the lung: a survey," *IEEE Trans. Medical Imaging*, vol. 25, no. 4, pp. 385–405, 2006.
- [15] S. L. A. Lee, A. Z. Kouzani, and E. J. Hu, "Automated detection of lung nodules in computed tomography images: a review," *Machine vision and applications*, vol. 23, no. 1, pp. 151–163, 2012.
- [16] A. Farag, A. Ali, J. Graham, S. Elshazly, and R. Falk, "Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose CT scans of the chest," in *Proc. ISBI*, 2011, pp. 169–172.
- [17] F. Zhang, W. Cai, Y. Song, M.-Z. Lee, S. Shan, and D. Feng, "Overlapping node discovery for improving classification of lung nodules," in *Proc. EMBC*, 2013, pp. 5461–5464.
- [18] Y. Song, W. Cai, Y. Wang, and D. Feng, "Location classification of lung nodules with optimized graph construction," in *Proc. ISBI*, 2012, pp. 1439–1442.
- [19] S. O'Hara and B. A. Draper, "Introduction to the bag of features paradigm for image classification and retrieval," *Computer Vision and Pattern Recognition*, 2011.
- [20] D. Unay and A. Ekin, "Dementia diagnosis using similar and dissimilar retrieval items," in *Proc. ISBI*, 2011, pp. 1889–1892.
- [21] Y. Song, W. Cai, Y. Zhou, L. Wen, and D. Feng, "Pathology-centric medical image retrieval with hierarchical contextual spatial descriptor," in *Proc. ISBI*, 2013, pp. 202–205.
- [22] J. Galaro, A. Judkins, D. Ellison, J. Baccon, and A. Madabhushi, "An integrated texton and bag of words classifier for identifying anaplastic medulloblastomas," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 3443–3446.
- [23] A. Cruz-Roa, F. Gonzlez, J. Galaro, A. Judkins, D. Ellison, J. Baccon, A. Madabhushi, and E. Romero, "A visual latent semantic approach for automatic analysis and interpretation of anaplastic medulloblastoma virtual slides," in *MICCAI LNCS*, vol. 7510, 2012, pp. 157–164.
- [24] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, 1995.
- [25] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," in *Proc. ECCV*, 2008, pp. 705–718.
- [26] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk, "SLIC superpixels," *Ecole Polytechnique Federal de Lausanne (EPFL), Tech. Rep.*, vol. 149300, 2010.
- [27] Y. Song, W. Cai, J. Kim, and D. Feng, "A multistage discriminative model for tumor and lymph node detection in thoracic images," *IEEE Trans. Medical Imaging*, vol. 31, no. 5, pp. 1061–1075, 2012.
- [28] Y. Song, W. Cai, and D. Feng, "Microscopic image segmentation with two-level enhancement of feature discriminability," in *Proc. DICTA*, 2012, pp. 1–6.
- [29] 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*, vol. 30, pp. 2440–2454, 2003.
- [30] 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. Medical Imaging*, vol. 25, no. 4, pp. 464–475, 2006.
- [31] W. Cai, J. Kim, and D. Feng, "Content-based medical image retrieval," *Biomedical Information Technology*, pp. 211–227, 2008.
- [32] P. Khurd, C. Bahlmann, P. Maday, A. Kamen, S. Gibbs-Strauss, E. M. Genega, and J. V. Frangioni, "Computer-aided gleason grading of prostate cancer histopathological images using texton forests," in *Proc. ISBI*, 2010, pp. 636–639.
- [33] F. Zhang, Y. Song, W. Cai, Z. Yun, S. Shan, and D. Feng, "Context curves for classification of lung nodule images," in *Proc. DICTA*, 2013, pp. 185–191.
- [34] F. Zhang, Y. Song, W. Cai, Y. Zhou, M. Fulham, S. Eberl, S. Shan, and D. Feng, "A ranking-based lung nodule image classification method using unlabeled image knowledge," in *Proc. ISBI*, 2014.
- [35] L. Srensen, S. B. Shaker, and M. De Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *IEEE Trans. Medical Imaging*, vol. 29, no. 2, pp. 559–569, 2010.
- [36] C. Jacobs, C. I. Snchez, 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*, 2011, pp. 207–214.
- [37] S. Liu, L. Zhang, W. Cai, Y. Song, L. Wen, and D. Feng, "A supervised multiview spectral embedding method for neuroimaging classification," in *Proc. ICIP*, 2013, pp. 602–605.
- [38] 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," *Academic radiology*, vol. 18, no. 3, pp. 306–314, 2011.
- [39] 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*, 2011, pp. 183–190.
- [40] A. Depeursinge, D. Van 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. Information Technology in Biomedicine*, vol. 16, no. 4, pp. 665–675, 2012.
- [41] U. Castellani, A. Perina, V. Murino, M. Bellani, G. Rambaldelli, M. Tansella, and P. Brambilla, "Brain morphometry by probabilistic latent semantic analysis," in *MICCAI LNCS*, vol. 13, no. 2, 2010, pp. 177–184.
- [42] S. Liu, W. Cai, L. Wen, and D. Feng, "Semantic-word-based image retrieval for neurodegenerative disorders," *The Journal of Nuclear Medicine*, vol. 53, p. 2309, 2012.
- [43] S. Liu, W. Cai, Y. Song, S. Pujol, R. Kikinis, and D. Feng, "A bag of semantic words model for medical content-based retrieval," in *MICCAI Workshop on Medical Content-Based Retrieval for Clinical Decision Support*, 2013.
- [44] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Machine learning*, vol. 42, no. 1-2, pp. 177–196, 2001.
- [45] J. W. Hwang and H. S. Lee, "Adaptive image interpolation based on local gradient features," *IEEE Signal Processing Letters*, vol. 11, no. 3, pp. 359–362, 2004.
- [46] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," in *Proc. ACM MM*, 2010, pp. 1469–1472.
- [47] Y. Xu, S. Huang, H. Ji, and C. Fermler, "Scale-space texture description on SIFT-like textons," *Computer Vision and Image Understanding*, vol. 116, no. 9, pp. 999–1013, 2012.
- [48] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. ICCV*, vol. 2, 1999, pp. 1150–1157.
- [49] Y. Song, W. Cai, and D. Feng, "Hierarchical spatial matching for medical image retrieval," in *Proc. ACM Multimedia 2011 Workshop on MMAR*, 2011, pp. 1–6.
- [50] M. Varma and A. Zisserman, "Classifying images of materials: Achieving viewpoint and illumination independence," in *Proc. ECCV*, 2002, pp. 255–271.
- [51] Y. Song, W. Cai, Y. Zhou, and D. Feng, "Feature-based image patch approximation for lung tissue classification," *IEEE Trans. Medical Imaging*, vol. 32, no. 4, pp. 797–808, 2013.
- [52] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.
- [53] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of documentation*, vol. 28, no. 1, pp. 11–21, 1972.
- [54] "ELCAP public lung image database." [Online]. Available: <http://www.via.cornell.edu/databases/lungdb.html>
- [55] F. Zhang, S. Shan, and Y. He, "Tag semantic analysis by utilizing latent semantic analyzing method of pLSA," *Journal of Information and Computational Science*, vol. 9, no. 18, pp. 5765–5776, 2012.