Chapter 5

Feature-Centric Lesion Detection and Retrieval in Thoracic Images

Yang Song[1], Weidong Cai[1], Stefan Eberl[2], Michael J Fulham[2,3], and David Dagan Feng[1]

[1]*Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, Australia*
[2]*Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Sydney, Australia*
[3]*Sydney Medical School, University of Sydney, Australia*
yson1723@uni.sydney.edu.au

Advances in medical digital imaging have greatly benefited patient care. Computer-aided diagnosis is increasingly being used to facilitate semi- or fully-automatic medical image analysis and image retrieval. While different tasks involve different methodologies in this domain, these tasks normally require image feature extraction as an essential component in the algorithmic framework. In this Chapter, we focus on image feature modeling in lesion detection and image retrieval for thoracic images. For both tasks, we first review the state-of-the-art and then present some of our own work in more detail.

1. Lesion Detection

Lung cancer is the most common cause of cancer-related death. Non-small cell lung cancer (NSCLC) is the most prevalent type of lung cancer, and it accounts for about 80% of all cases.[1] Staging, which assesses the degree of spread of the cancer from its original source, is critical in determining prognosis and choosing the most appropriate treatment. In the 'tumor, node, metastasis' (TNM) staging system, the size and spatial extent of the primary lung tumor and the degree of involvement of regional lymph nodes are critical factors.

Positron emission tomography – computed tomography (PET-CT) with ^{18}F-*fluoro-deoxy-glucose* (FDG) tracer is now accepted as the best imaging technique for non-invasive staging of NSCLC.[2] In PET-CT, the CT scan provides anatomical information; it has relatively low soft tissue contrast which causes difficulties in separating lesions from the surrounding tissues. On the other hand, the PET scan has high contrast and reveals increased metabolism in structures with rapidly growing cancer cells, but the localization of these foci of increase metabolism is limited by the low spatial resolution in PET. The integrated PET and CT scan thus provides

complementary functional and anatomical information. In current routine clinical workflow, the detection of abnormalities is performed manually. There may be many such abnormalities in a patient with NSCLC. To assist this time-consuming process and potentially provide a second opinion to the reading physicians, an automated system that can provide fast and robust lesion detection is desirable.

The objective is thus to design a fully automated methodology to detect primary lung tumors and disease in regional lymph nodes from PET-CT images of the thorax. Examples of lesion detection are shown in Fig. 1. There are two main challenges. First, tumor metabolism detected in PET relates to uptake of the tracer FDG; this uptake can be expressed semi-quantitatively as the standard uptake value (or SUV). The SUV normally exhibits high intra- and inter-patient variances, and can highlight non-pathological areas (e.g. in myocardium). Second, separation of the primary lung tumors from abnormal lymph nodes can be difficult, especially in complex cases where tumors invade the mediastinum or lymph nodes in the pulmonary hilar regions. In this section, we review the state-of-the-art in lesion detection, and describe our approaches[3–5] to tackle this problem.

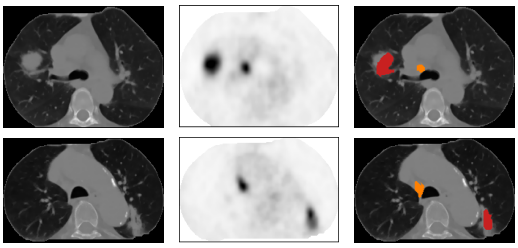

Fig. 1. Examples of primary lung tumors and involved lymph nodes.[4] Each row shows one example, using a transaxial slice view for easier visualization. The left column shows the CT slices, the middle column the PET slices, and the right column shows the lung tumors (red) and involved lymph nodes (orange).

1.1. *Review of State-of-the-art*

Currently there are no published data on the simultaneous detection of lung tumors and disease in regional lymph nodes, apart from our own studies. Existing work is mainly on lung tumor detection, without considering the involvement of lymph nodes. By first segmenting the lung fields, a threshold and fuzzy-logic based approach is then used to detect the lung tumors,[6] but the detection performance is quite sensitive to the delineation accuracy of the lung fields. Another approach attempts to handle tumors lying close to the edge of lung fields by incorporating the location, intensity, and shape information,[7] but the method could potentially result in a large number of false positives with the predefined SUV thresholds. False positives are usually detected in the mediastinum with elevated SUV. To reduce such

false positives, learning-based techniques with tumor-specific features have been proposed,[8,9] but the methods are based on empirical studies of SUV distributions and tumor sizes, and do not seem to consider abnormal lymph nodes in the thorax.

Another category of lesion detection is to detect all instances from PET images, regardless of their types. Such approaches include a texture-based classification method,[10] and a water-shed based algorithm integrated with morphological measures.[11] A common drawback with these techniques is that they operate on user-selected volume-of-interest (VOI) or potential lesions. Fully automated lesion detection has also been studied,[12] and we have previously reported several methods[13–15] based on subject-level contrast features. However, differentiation of lung tumors from abnormal lymph nodes is not investigated in these studies.

1.2. *Region-based Feature Classification*

In our work, we first proposed a region-based feature classification method[3] to detect the primary lung tumors and the abnormal lymph nodes, in a three-step approach.

1.2.1. *Region Type Identification*

A modified fuzzy c-means (FCM) approach is first used to cluster each PET-CT slice into regions of various sizes and shapes.[16] Based on anisotropic diffusion filtering (ADF),[17] Gabor and shape features, each region is then represented by a 12-dimensional feature vector F: mean and variance of ADF/Gabor filtered CT values; mean and variance of ADF/Gabor filtered PET SUVs; the size; the eccentricity; and the centroid x and y coordinates.

Next, the regions are classified into 5 types: lung field, mediastinum, involved regional lymph node (N), tumor in the lung, and border area surrounding the tumor. The last two represent the T type. The classification is based on feature F with a multi-class linear-kernel support vector machine (SVM).[18] The feature weights for each region type are also derived from the support vectors, resulting in a 5×12 matrix of feature weights w.

1.2.2. *Region Type Refinement*

The region types identified are often misclassified, particularly in terms of correct T and N classifications. Since both could be characterized by high CT values and high SUVs, the feature vector F is not sufficient to differentiate the two types. A major difference between T and N, however, could be modeled by the spatial relationships between regions. For example, T is within lung fields while N is in the mediastinum or the hilar nodal area.

We thus refine the region classifications based on spatial information. The region delineation step normally formulates several near-concentric regions at the T/N area, and several large regions corresponding to the lung fields and mediastinum surrounding the T/N area. Therefore, for each region, R_i, initially classified as T or

N, its spatial information could be defined by the regions inside and outside of R_i: $\langle R_j, R_k \rangle$. A 36-dimensional feature vector FS is hence computed by concatenating the feature vectors of R_i, R_j and R_k.

A second multi-class SVM is then trained to classify the regions based on the feature vector. The final region type is determined by combining the margins computed from both multi-class SVMs. The output with the maximum combined margin is chosen as the region type.

1.2.3. *3D Object Localization*

Given the T or N regions identified from each PET-CT slice pair, we then attempt to localize the 3D T or N objects within a case. Because not all T or N areas are correctly identified from the slices due to errors from the classification and the region delineation step, we design a voting-based method for the localization. Specifically, two scores S_I^T and S_I^N are assigned to each slice I in a case:

$$S_I^{T/N} = \exp\left[-min(d(I,J), d(I,K))\right] \tag{1}$$

J and K are the slices spatially nearest to I (above and below) with detected T/N regions; and d is the normalized weighted (w) Euclidean distance between the feature vectors of the T/N regions. For each case, the mean score value and median locations of the T/N regions are computed from its set of images. T/N regions with scores higher than the mean score value are marked as valid locations.

1.3. *Multi-stage Discriminative Model*

A drawback of the region-based classification method[3] is that it requires a separate class of tumor border, to work around the issue that the surrounding areas of tumors in the lung are often misclassified as mediastinum. Such a tumor-border class complicates the training process, which is unnatural for a clinical workflow. Our later work[19] avoids this issue as we use a multi-level discriminative model and more comprehensive spatial features. However, this method requires the surrounding regions of tumors and abnormal lymph nodes to be accurately classified, involving a heuristic-based grouping operation to separate the surrounding regions from the mediastinum. In addition, the regions belonging to a tumor or an abnormal lymph node are classified individually, which could result in inconsistent labeling within a 3D volume. Therefore, we then proposed a multi-stage discriminative model,[4] as detailed in the following.

1.3.1. *Abnormality Detection*

In the first stage, after preprocessing for background removal,[20] each transaxial PET-CT slice of a 3D image set is clustered into regions using quick-shift clustering.[21] Represent each 3D image set by N_r regions from all slices $V = \{r_i : i =$

$1, ..., N_r\}$. Each region r_i is then classified into lung fields (L), mediastinum (M) or abnormality (ROI) categories based on low-level and high-level features.

Two types of low level features are computed: intensity and neighborhood. To describe the intensity features, a two-dimensional intensity vector is computed: (i) average CT density and (ii) average normalized SUV of r_i. While the average CT density is based on the raw values, for PET, we perform an extra SUV normalization $\|u_i\|$ with a sigmoid function:

$$\|u_i\| = \frac{C_1}{1 + \exp(-(u_i - \theta_V)/\theta_V)} \tag{2}$$

where u_i is the average SUV of r_i, θ_V is the adaptive reference value computed for each 3D image set V in a similar approach to our previous work,[13] and C_1 is a scaling constant controlling the range of the normalized SUV. The normalization is to rescale the SUVs across patients within a similar range, and in the process, boost the separations between the ROIs and the mediastinum. To describe the neighborhood features, the average CT density and normalized SUV of the neighboring area of region r_i in the adjacent slices (one above and one below) are also computed, to incorporate 3D information.

In some cases, the SUVs of ROIs and the mediastinum are relatively close, and some false positive ROIs could be detected in the mediastinum with the low-level features. We thus exploit the high-level features, by computing the contrast between the detected ROIs and the lung fields and mediastinum. To do this, we first classify the regions into lung fields, mediastinum and ROI – $\{R_L, R_M, R_O\}$, based on the low-level features. Let u_L, u_M and u_O be the average normalized SUVs of R_L, R_M and R_O, a four-dimensional high-level feature is then computed for each ROI region r_i: $\{u_r/u_L, u_r/u_M, u_r/u_O, u_r\}$, where u_r is the average normalized SUV of r_i.

Any misclassification at this stage would be propagated to later stages. Thus we use soft labeling to create a vector of probabilities, rather than labeling every region with a single category (L, M or ROI), to reduce the impact of possible misclassifications. The soft labeling vector, denoted as $p_i = \{p_i^L, p_i^M, p_i^O\}$ for region r_i, is obtained by combining the probability estimates of the SVM classification based on both low- and high-level features.

1.3.2. *Tumor and Lymph Node Differentiation*

In the second stage, the detected abnormalities are differentiated as tumors or abnormal lymph nodes, as illustrated in Fig. 2. A conditional random field (CRF)[22] model integrated with SVM and a comprehensive set of features are designed to achieve an accurate discrimination between the two types of abnormalities (SVM), and minimize any misclassification by exploiting 3D correlations (CRF). The use of CRF allows us to incorporate the structural information in addition to the region-based features, so that a 3D ROI volume could be classified collectively.

Based on the outputs of the first stage, the abnormalities detected from a 3D

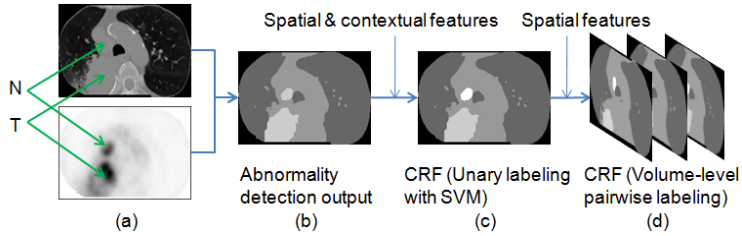

Fig. 2. Illustration of tumor and lymph node differentiation.[4] (a) A transaxial PET-CT slice. (b) The abnormality detection output with two ROIs detected. (c) Region-level labeling with the unary term based on spatial and contextual features, showing the two ROIs labeled as T (gray) and N (white). (d) Volume-level labeling with the pairwise term based on 3D spatial features.

image set V are represented as N_O regions $\{r_i : i = 1, ..., N_O\}$. Rather than individual regions created at the clustering step, r_i here represents a large ROI region created by merging regions that are labeled as ROI and spatially connected in the same slice. A set of 3D connected $\{r_i\}$ (i.e. across slices) then form a 3D ROI volume, and V could contain multiple ROI volumes, e.g. a primary lung tumor and several abnormal lymph nodes.

The objective is then to assign each r_i a binary label $a_i \in \{T, N\}$; and the probability of a labeling set $A = \{a_i : i = 1, ..., N_O\}$ is modeled as a conditional distribution in the CRF framework:

$$P(A|V) = Z^{-1} \exp(-E(A|V)) \tag{3}$$

where Z is the partition function. We define the energy $E(A|V)$ as a linear combination of a set of unary features $F_k(a_i, V)$ and a pairwise feature $G(a_i, a_{i'}, V)$:

$$E(A|V) = \sum_i \sum_k \lambda_k F_k(a_i, V) + \sum_{i,i'} \mu G(a_i, a_{i'}, V) \tag{4}$$

where λ_k is the weight of the kth feature, i and i' index the 3D connected regions (in different slices), and μ is the weight of the pairwise feature. The unary features are computed for each r_i and are the most decisive factor for labeling $a_i \in \{T, N\}$, while the pairwise features are to exploit the 3D structural information for a consistent labeling throughout an ROI volume.

The idea is then to search for a labeling combination for V, so that the total energy cost is minimum, leading to a labeling set that would be optimized for the whole 3D volume. Graph cut[23] is used to derive the most probable labeling A^* that minimizes the energy function: $A^* = \text{argmin}_A \; E(A|V)$.

Unary Term:

The unary term $\sum_k \lambda_k F_k(a_i, V)$ (denoted as $\psi(a_i)$) indicates the labeling preference of individual region r_i. Specifically, given label a_i for region r_i, a higher $\psi(a_i)$ means a higher cost (i.e. lower probability) of r_i belonging to a_i. The unary

term could thus be considered as a binary classifier for r_i. We design a highly discriminative feature set describing the spatial and contextual features, and derive the unary cost $\psi(a_i)$ from the classifier output (SVM).

The main distinctive feature between T and N is the location information. In particular, it is generally true that T is in the lung fields while N is in hilar region or in mediastinum. However, quite often T might invade into the mediastinum and appear to be outside of the lung fields. For T that is near to the mediastinum, mislabeling of its surrounding areas would cause T to appear outside of the lung fields. In addition, N could be adjacent to the lung fields, appearing similar to T.

Based on these considerations, we design three types of features to extract from each region r_i that is detected as ROI: (i) Quad-radial global histogram: four radial lines are drawn at $\pm 45^\circ$ and $\pm 135^\circ$, from the geometric center of r_i, and a 12-dimensional histogram H_g is then created to compute the distribution of L, M and ROI in the four radials. (ii) Surrounding contour histogram: a closed contour is drawn outside of r_i, with a displacement of d from the boundary of r_i, and a three-dimensional histogram H_s is then created to count the percentages of L, M and ROI in the surrounding contour. (iii) Pleural distances: a four-dimensional vector D containing the distances between r_i and the lateral, medial, anterior, and posterior sides of the nearest lung field are computed.

To compute the unary cost, a binary SVM is used to classify r_i to T or N categories based on its feature vector $\{H_g, H_s, D\}$, with a probability estimate p_{a_i} for each category. The unary cost is then computed as $\psi(a_i) = 1 - p_{a_i}$, to produce two cost values for each r_i. Furthermore, we observe that the regions nearer to the boundary of the ROI volume are more prone to mislabeling. Therefore, the unary cost is refined with a Gaussian weight ω_i based on the distance between r_i and the volume center:

$$\psi(a_i) = \omega_i(1 - p_{a_i}); \quad \omega_i = \exp(-\tfrac{(z_i - z_c)^2}{2\sigma^2}) \tag{5}$$

where z_i and z_c are the z coordinates of r_i and the center of the volume, and σ is calculated as $1/2$ of the size of the volume.

Pairwise Term:

The pairwise term $\mu G(a_i, a_{i'}, V)$ (denoted as $\phi(a_i, a_{i'})$) is useful in promoting spatial consistency between spatially connected regions r_i and $r_{i'}$. Specifically, a cost is assigned as $\phi(a_i, a_{i'})$ if r_i and $r_{i'}$ are labeled differently. The regions r_i and $r_{i'}$ are considered spatially connected if they are part of the same 3D ROI volume. The pairwise term thus explores the inter-slice and volume-level information for refined labeling.

We define the pairwise cost as:

$$\phi(a_i, a_{i'}) = \delta(a_i - a_{i'}) \cdot x'; \quad x' = \tfrac{1}{1+\exp(-C(x-0.5))} \tag{6}$$

where $\delta(a_i - a_{i'})$ is 0 or 1 indicating the same or different labelings of r_i and $r_{i'}$, and x' is the cost value, which is a sigmoid normalization of the actual cost:

$x = \alpha(r_i, r_{i'}) \cdot \beta(r_i, r_{i'})$.

The factor $\alpha(r_i, r_{i'})$ measures the spatial distances between the two regions. The value of $\alpha(r_i, r_{i'})$ is in $[0, 1]$ range, and is larger if the distance between r_i and $r_{i'}$ is smaller. Such a computation is similar to the usual CRF formulation, but is based on the spatial distances, rather than intensity differences; and the pairwise cost is computed between all pairs of regions of a 3D ROI volume, rather than only for those neighboring regions. The factor $\beta(r_i, r_{i'})$ is computed as the degree of overlap in the xy plane between the two regions in different slices. This factor is introduced because in some cases, adjacent T and N volumes could actually form into one 3D volume. With the $\alpha(r_i, r_{i'})$ factor alone, the regions in T and N would be all correlated, and the lowest energy solution would tend to produce a single label for the joint volume.

1.3.3. *Tumor Region Refinement*

In the third stage, we identify high uptake in the myocardium as a false positive tumor volume and update its labeling to M. This is based on the usual assumption that images showing a high SUV in the myocardium should be considered normal, and a tumor detected in such an area should be ignored.[12]

Given a detected tumor volume T_q, if it is at the left half of the thorax, a CRF model is employed to classify T_q to either M or T category, depending on the likelihood of T_q representing a high-uptake in the myocardium or a lung tumor. Defining T_q as a series of regions $\{r_i : i = 1, ..., N_O\}$, with each r_i representing a set of connected T regions in a slice, the CRF model is designed based on the same construct as in the second stage, but with a different set of features for the unary term: (i) Pleural Distances: the signed distances between r_i and the four sides of the left lung field is computed. (ii) Shape of Lung Field: a HOG descriptor[24] is used to describe the shape of the left lung field that r_i is adjacent to. A binary SVM is then used to classify r_i to either M or T, and the unary cost is computed from the probability estimates of the classifier.

1.3.4. *Experimental Results*

The dataset used in this study comprised image scans from 85 patients diagnosed with NSCLC, acquired using a Siemens TrueV 64 PET-CT scanner at the Royal Prince Alfred Hospital, Sydney. A total of 93 lung tumors and 65 abnormal lymph nodes were annotated. During the preprocessing, the PET images were linearly interpolated to the same voxel size as the CT images, and FDG uptake normalized into SUV based on the injected dose and patient's weight.

We compared our results to the previous method,[19] which was the only work that addressed the detection of tumor and involved lymph nodes. Since the original work[19] was evaluated on a different dataset, we repeated the test on the current dataset. As shown in Table 1, our method exhibited clear improvements, which were

Table 1. Performance comparison with other methods. * indicates rerun results on current dataset. R means recall and P means precision.

Method	Test size	# T	# N	T-R	T-P	N-R	N-P
Proposed method	85 cases	93	65	97.9	88.4	86.2	88.9
Song et.al.[19]	50 cases	53	36	84.4	83.8	77.8	76.9
Song et.al.[19]*	85 cases	93	65	89.3	79.1	72.3	82.5

mainly attributed to three factors: (i) fewer abnormal lymph nodes mislabeled as tumors, especially for those nodes lying close to the lung fields, with the spatial and contextual features (unary term); (ii) fewer tumors mislabeled as abnormal lymph nodes, especially those previously caused by inconsistent labeling of regions in one tumor volume, with the spatially-smoothed 3D volume labeling (pairwise term); and (iii) fewer high-uptake areas in the myocardium mislabeled as abnormalities.

1.4. *Data Adaptive Structure Estimation*

While the multi-stage discriminative model is fairly effective for lesion detection, it relies mainly on complicated and domain-specific feature design. These features are designed based on domain knowledge and empirical studies, and so their effectiveness may be restricted to the limited scenarios available in the datasets, and might be difficult to generalize to a larger variation of cases.

Therefore, we proposed a different approach to the detection problem – after detecting all abnormalities, if we can identify the actual lung fields (tumors inclusive), then we can differentiate lung tumors and abnormal lymph nodes based on the degree of overlap between the detected abnormality and the lung fields.[5] The main problem is how to estimate the original lung fields if the subject had been healthy. Limited studies exist in this area, and are mostly based on statistical shape model,[25,26] with time-consuming registration[25] or complex landmark detections.[26]

Since precise lung segmentation is not required for lesion detection, but a fair estimation of the overlap would suffice, we design a simpler atlas-based approach. Our design is similar to the approach[27] that obtains brain segmentation masks from multiple weak segmenters. Different from local-level computation,[28] the regression-based combination[27] minimizes the weighted difference for the whole image. However, its direct derivation of segmentation from other labeling outputs might impose a stringent requirement on the weight learning, which would be difficult to apply to the thoracic images due to the large variations of anatomical structures caused by lesions. This motivates us to opt for an indirect approach, with intermediate multi-atlas modeling of the feature space and a further classification for final labeling.

Our approach[5] consists of the following steps. All abnormalities are detected via region-based classification. Then, the actual lung structure is estimated, with regression and graph-based techniques. Thirdly, the detected abnormalities are

classified as tumors or abnormal lymph nodes based on contextual features. Fig. 3 illustrates the overall method.

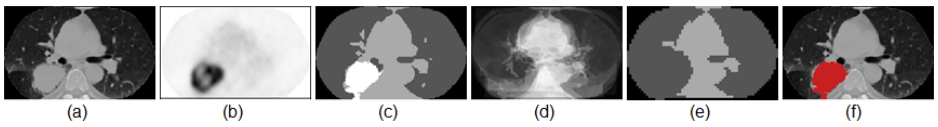

Fig. 3. Method illustration.[5] (a) An axial CT slice (after preprocessing). (b) The co-registered PET slice, where the dark region indicates a lung tumor with increased FDG uptake and a high SUV. (c) Output of the initial abnormality detection, showing the lung fields, mediastinum and abnormality with increasing grayscale values. (d) The appearance model generated with regression, approximating the CT intensities if without the lung tumor. (e) Output of the graph-based structure labeling for lung fields and medaistinum. (f) The detection output after tumor/lymph node classification, with tumor highlighted in red.

1.4.1. *Initial Abnormality Detection*

The thoracic PET-CT images are first preprocessed to remove the background and soft tissues outside of the lung and mediastinum with morphological operations. All images are then aligned and rescaled to the same size. Next, the abnormalities are detected by classifying the mean-shift clustered regions into lung fields (L), mediastinum (M) or abnormalities (O)[19] (Fig. 3c). The high-uptake in the myocardium is masked-out based on its size, spatial location within the thorax and the shape of the left lung field.

1.4.2. *Adaptive Structure Estimation*

To differentiate between lung tumors and abnormal lymph nodes, a general rule is that lung tumors should be inside the lung fields, while lymph nodes are outside. However, as shown in Fig. 3c, due to the lung tumor, only a portion of the right lung field is correctly identified, and the tumor then appears outside the lung fields. Therefore, we need to estimate the actual lung fields before the tumor growth (Fig. 3e). Given a 3D thoracic PET-CT volume I, our objective is to label each voxel i to the lung field or mediastinum type. To do this, the thoracic appearance is first modeled from a set of reference images, then the voxels are classified as L/M.

Regression-based Appearance Model:

Although patient-specific conditions, such as body weights, introduce variational factors, there is great similarity between images for the normal structures. It is thus a fair assumption that one image can be approximated by a weighted combination of multiple images. Therefore, we model the CT appearance of the original thoracic structures (Fig. 3d) based on other reference images.

We first introduce a basic formulation for the appearance model. Let $y \in \mathbb{R}^{n\times 1}$

be the n-dimensional feature vector (i.e. voxel-wise CT intensities) of I, and $D \in \mathbb{R}^{n\times K}$ be the matrix of K feature vectors from K reference images I_k ($n \gg K$). The difference between y and the weighted combination of D should then be minimized: $\min_x \parallel y - Dx \parallel_2^2$, where $x \in \mathbb{R}^{K\times 1}$ is the weight vector; and Dx is the original appearance of I approximated.

With the derived x, each reference image I_k is assigned one weight x_k, and hence all voxels in I_k contribute equally to the approximated appearance. However, due to the non-rigid structure of the thorax and presence of the abnormalities, it is normal that only a portion of I_k is similar to I and the rest should take lower weights. Therefore, we incorporate a voxel-wise similarity-based weight vector for each I_k. For voxel i_k of image I_k, the weight $w_{i,k}$ is computed as:

$$w_{i,k} = \frac{1}{\alpha_i}\exp(-\frac{1}{\beta_i} \parallel i - i_k \parallel_2),\ \beta_i = \sum_{k=1}^{K} \parallel i - i_k \parallel_2 \tag{7}$$

where α_i is to normalize $\sum_k w_{i,k} = 1$. With the weight matrix $W = \{w_{i,k}\} \in \mathbb{R}^{n\times K}$, the regression formulation thus becomes: $\min_x \parallel y - (W \circ D)x \parallel_2^2$.

Furthermore, while the above formulation is sufficient to obtain a closely matching appearance model, the L/M labeling information is not utilized. Since the final objective is to achieve accurate structure labeling, it is natural to integrate the supervised information to enhance the discriminative power:

$$\begin{aligned} &\min_x \parallel y - (W \circ D)x \parallel_2^2 + \parallel h - (W \circ A)x \parallel_2^2 \\ &\quad = \min_x \parallel \begin{pmatrix} y \\ h \end{pmatrix} - \begin{pmatrix} W \circ D \\ W \circ A \end{pmatrix} x \parallel_2^2 \qquad = \min_x \parallel f - \Omega x \parallel_2^2 \end{aligned} \tag{8}$$

where $h \in \{1, 2, 1.5\}^{n\times 1}$ is the label vector of I from the initial detection outputs (1=L, 2=M, and 1.5=O), and $A \in \{1, 2\}^{n\times K}$ for the reference images from the ground truth. The value 1.5 is chosen to have equal distance between O/L and between O/M, to assign no preference for matching such areas with L or M . Both h and A are normalized to the same range as y and D, and the approximated appearance model is then $(W \circ D)x$ and the labeling $(W \circ A)x$.

Finally, to avoid overfitting, we choose to not have all reference images contributing to the appearance approximation, with a sparse regularization:

$$\min_x \parallel f - \Omega x \parallel_2^2,\ s.t. \parallel x \parallel_0 \leq C \tag{9}$$

where C is the constant number of reference images we limit to. The OMP algorithm[29] is then used to solve x.

To compute the feature vector y, we divide I into multiple sections, each with three slices, and y is then derived for each section. To construct D, since the reference images also contain lung tumors or abnormal lymph nodes, rather than simply concatenating all voxels, the annotated tumor voxels are replaced with the average intensity of the lung fields labeled at the initial detection step.

Graph-based Structure Labeling:

Next, based on the appearance model (Fig. 3d), we classify the lung fields and mediastinum (Fig. 3e). A straightforward idea is to use the approximated labeling $(W \circ A)x$ as the classification output. However, such labelings can be erroneous especially for the boundary areas of tumors. Therefore, we design a further graph-based classification step for the structure labeling.

We first define a notation for the appearance model: $G = \{g_i\} = (W \circ D)x$, where g_i is the approximated intensity for voxel i. The problem is thus to derive a label set $V = \{v_i \in \{L, M\}\}$, to classify each voxel to category L or M.

We observe that the mislabeled parts usually appear lighter in G but still darker than the real mediastinum. This thus motivates us to encode contrast information for the labeling. To do this, from G, we first calculate the mean values (m) and the graylevel histograms (d) of the lung field and mediastinum (labeled during the initial abnormality detection). A 5-dimensional feature vector q_i is then computed for each voxel i: (i) g_i; (ii) g_i/m_L; (iii) g_i/m_M; (iv) $Pr[g_i \leq d_L \leq 256]$; and (v) $Pr[1 \leq d_M \leq g_i]$.

In addition to q_i, which incorporates the global-level information m and d, contrast information can also be described in a pairwise fashion. Specifically, for two voxels i and j, if g_i and g_j are similar and they are spatially close, they would likely take the same label. Hence we define the difference $s_{i,j}$ between i and j based on their intensity $\mid g_i - g_j \mid$ and spatial $\parallel i - j \parallel_2$ distances:

$$s_{i,j} = \log(\parallel i - j \parallel_2 +1) \times \log(\mid g_i - g_j \mid +1) \tag{10}$$

A lower $s_{i,j}$ would imply a higher probability of $v_i = v_j$.

We then design a CRF construct to integrate both q_i and $s_{i,j}$ to label G, with the following energy function:

$$E(V|G) = \sum_i \phi(v_i) + \sum_{i,j} \psi(v_i, v_j) \tag{11}$$

Here $\phi(v_i)$ represents the cost of i taking the label v_i, computed as $1 - p(v_i|q_i)$; and $p(.)$ is the probability estimate from a binary liner-kernel SVM classifier based on q_i. The pairwise term $\psi(v_i, v_j)$ penalizes the labeling difference between i and j.

Note that our pairwise term connects longer distance voxels to encourage consistent labelings for similar voxels, not limited to neighboring voxels as the traditional CRF construct. However, to ensure a sparse graph, a voxel i should be linked to a small number of other voxels only. Therefore, we introduce a constant threshold tr, so that $s_{i,j} = 0$, if $\mid g_i - g_j \mid > tr$. The labeling set V is then derived by minimizing $E(V|G)$ using graph cut.

1.4.3. *Feature Extraction and Classification*

Based on the estimated thoracic structure V (Fig. 3e), we then classify the detected abnormalities (O) into tumors (T) or abnormal lymph nodes (N) (Fig. 3f). A simple

Table 2. The detection recall and precision.

	Tumor	Node	Tumor[19]	Node[19]
Recall (%)	90.7	88.6	84.4	77.8
Precision (%)	89.1	88.6	83.8	76.9

4-dimensional feature vector is designed: (i) size of O; (ii) size of overlap between O and lung field labeled in V; (iii) size of overlap between O and mediastinum labeled in V; and (iv) size of overlap between O and the convex hull of lung field detected during initial abnormality detection. Features (ii)–(iv) are also normalized by the size of O. A binary linear-kernel SVM is then trained to classify O to T or N. To enhance the error tolerance, the classification is performed on a section basis as well, and the final T/N label is produced based on a weighted averaging of the probability estimates from each section. The weights are computed as $\exp(-d/\eta)$, where d is the distance between the section and center of O, and η is the maximum distance possible for O.

1.4.4. *Experimental Results*

The experiment was performed on 50 sets of 3D thoracic PET-CT images from patients with NSCLC. A total of 54 lung tumors and 35 abnormal lymph nodes were annotated as the ground truth. For each data set, the contour of lung field was also roughly delineated – we allowed some error margins in the delineation since we did not expect precise lung segmentation. Five images were selected as the training set for both structure labeling and classification between tumors and lymph nodes. The data sets were then randomly divided into five sets; and within each set, each image was used as the testing image, with the other nine as the reference images.

The usefulness of each components in the structure estimation was first analyzed. With the proposed graph-based structure labeling, we evaluated the appearance model with different constructs of the regression method. With the fixed regression-based appearance model, we then evaluated the structure labeling with different graphical constructs. The results clearly demonstrated the advantage of our proposed appearance model and the structure labeling. The overall detection recall and precision are shown in Table 2. The results exhibited marked improvement over our previous work,[19] especially for the abnormal lymph nodes; and it suggested the effectiveness of our approach for differentiating the two abnormalities, by mainly analyzing the degree of overlap between the detected abnormality and the estimated lung structures.

2. Thoracic Image Retrieval

In the past three decades, the volume of medical image data has rapidly expanded. Patient images are normally stored in the picture archiving and communication systems (PACS); and the vast amount of these data also opens an opportunity for case-based reasoning or evidence-based medicine support.[30] When a physician makes a diagnosis based on a patient scan, the physician may choose to browse through similar images in the database as a reference set to help reach the correct diagnosis. The ability to find images with similar content is thus important to facilitate diagnosis. In this section, we review the state-of-the-art in medical image retrieval, and describe our approaches[3,31] for the thoracic PET-CT.

2.1. *Review of State-of-the-art*

A typical content-based image retrieval (CBIR) system comprises three main components: feature extraction for image representation, similarity measure between feature descriptors, and image indexing for retrieval.[32] In the medical domain, since different disease patterns and imaging modalities (e.g. X-ray, CT, magnetic resonance imaging, PET-CT) are best characterized by different types of features, the majority of image retrieval studies focus on feature extractions,[33–39] designed for specific anatomical structures. The various types of features explored include intensity values,[35] textures describing the tissue appearances,[34,36,37,39,40] and shapes of anatomical structures.[38] Contextual features describing organ-related spatial information have also been recently incorporated.[36,39,41] However, most of these contextual features designed still largely resemble the standard feature descriptors, not optimized for the particular medical imaging problem. As a result, the retrieved images would still be mainly similar in their visual appearances, but not in terms of anatomical or pathological information.

In thoracic PET-CT, images representing similar pathological characteristics can be used as references to achieve accurate staging of NSCLC. The retrieved images are thus expected to exhibit a similar spatial extent for the lung tumor and disease spread in regional lymph nodes. Fig. 4 shows tumors in various locations. Studies in thoracic PET-CT image retrieval include design of overlapping[42] and graph-based features.[43] Both approaches rely on segmentation of thoracic structures and lesions, which however, might not error-prone with simple thresholding-based techniques. Our initial work[44] does not require precise segmentation, but it is designed for 2D image retrieval only. We have developed various learning-based similarity measure techniques,[19,45,46] which can be integrated with any feature representation.

2.2. *Pathological Feature Description*

In our work, we first proposed an image retrieval method,[3] which finds images exhibiting similar pathological features that are extracted based on the lesion de-

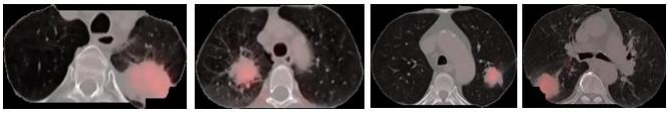

Fig. 4. Four examples of fused thoracic PET-CT images (transaxial slices). The images depict tumors at different locations in the lung fields, and in two images showing the tumors adjacent to the chest wall and the mediastinum.

tection results. As described in Section 1.2, tumors (T) and abnormal lymph nodes (N) are detected as 3D objects. The feature of a 3D T or N object F_{obj} is computed as a weighted combination of the feature vectors F_r of the comprising image slices. F_r is computed from the bounding box of the T/N region in the image slice. The weight of an image slice is adaptively assigned as its mean CT intensity and PET SUV, so that slices with more salient features would carry higher weights. A case is then represented by two feature vectors F_T and F_N, characterizing its tumor and lymph node appearances. The distance between the query case X and the reference case Y is then defined as the weighted intersection differences, with weights defined based on the volumes of the T or N objects.

2.3. *Spatial Feature Encoding*

The simple pathological feature description is usually insufficient to represent the large variations in spatial contexts of different cases, and would thus impact the retrieval performance. The most widely used spatial feature descriptor is the circular or square histogram.[36,47,48] The drawback of these feature descriptors is that they might not accommodate spatial variations well, due to the fixed grid structures. The hierarchical subdivision scheme, such as spatial pyramid matching (SPM),[49,50] are able to balance between the subdividing and disordering to a certain extent with a multi-scale design. We have designed spatial descriptors[51,52] based on the concept of SPM. However, such approaches might be still too rigid to handle the large inter-subject variations because of the even subdivision.

To better describe the pathological features, we then proposed a hierarchical contextual spatial descriptor,[31] with an adequate balance between its discriminative capability and geometric-transformation invariance suitable for the thoracic imaging domain. Fig. 5 gives an overview of the proposed method.

2.3.1. *Pathological Centroid Detection*

To build the spatial descriptor, we first detect the centroid of the pathology, which is the geometric center of a tumor. To do this, we first extract the maximally stable extremal regions (MSER)[53] from the image grid of feature words W_I. Next, the region of pathology is selected from the MSER outputs, by choosing the inner-most region that normally represents the center area of the tumor. The geometric center

of this detected region is thus the tumor centroid, and denoted by the patch index as p_o (example as illustrated in Fig. 5).

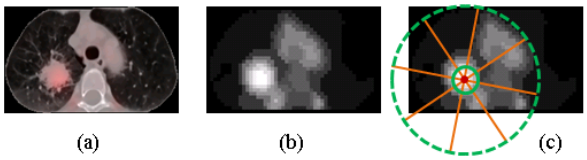

Fig. 5. Illustration of the proposed feature extraction and representation.[31] (a) A fused PET-CT slice, showing a primary lung tumor. (b) The feature-word grid, representing each image patch with its feature word value. (c) The hierarchical partition of the spatial contexts, showing only the level-2 structure for simplicity, with the red dot indicating the tumor centroid p_o, green circles depicting the concentric circles, and orange lines dividing the radials.

2.3.2. *Context-based Partitioning Model*

Based on the grid of feature words W_I and the pathological centroid p_o, we then formulate the spatial features using a context-based partitioning model. We incorporate the hierarchical partitioning concept similar to SPM; but rather than dividing a rectangle cell evenly into 4 sub-cells as in SPM, we design a hierarchical circular structure, as illustrated in Fig. 5. The circular model is more suitable here mainly because: (1) usually tumors are close to blob-like shapes; and (2) anatomical structures surrounding the tumors can be better fitted into radials rather than rectangles.

We define L as the total levels of hierarchy, with individual levels $l \in \{0, 1, ..., L-1\}$. At each level l, 4^l radials are created with the partitions of 2^{l-1} concentric circles while each circle is divided into 2^{l+1} radials. The overall spatial descriptor H_I of image I is concatenated from the feature vectors of individual levels: $H_I = \{H_I^l : l = 0, ..., L-1\}$, with a dimension of $K \sum_{l=0}^{L-1} 4^l$.

At level-0, there is actually no partitioning, and a circle O_I centered at p_o is created with radius r_I the largest distance between p_o and the image border. All feature words in O_I are accumulated into a weighted histogram $H_I^{l=0}$, with each feature word w_i Gaussian weighted according to its Chebyshev distance from p_o. At level-1, the circle O_I is evenly divided into $J = 4$ radials $O_I = \{R_{I,j} : j = 1, ..., J\}$ from the centroid p_o. The feature vector of each radial is then computed in the same way as level-0. At a higher level ($l \geq 2$), 4^l radials are then created from O_I in a similar way.

When creating the radials, rather than partitioning at fixed angles, our approach is to create a partition that minimizes the co-occurrences of multiple structure types within one radial, to reduce fragmented segments due to dividing in homogeneous regions. Denote the jth radial as $R_I(j, \theta)$ with θ representing the direction of the

basis radial. The total variance $v_I(J,\theta)$ of such a partition is:

$$v_I(J,\theta) = \sum_{j=1}^{J}\sum_{i=1}^{N}(w_i - m_j)^2,\ s.t.\ p_i \in R_I(j,\theta) \tag{12}$$

where m_j is the mean of the feature words of radial $R_I(j,\theta)$. The best partition structure is thus the one resulting in the smallest variance: $\theta = argmin_\theta v_I(J,\theta),\ \forall\theta = \{0, \frac{\pi}{2J}, \frac{\pi}{J}, \frac{3\pi}{2J}\}$. Here we choose to test four possible θ only, for convenient implementation and better efficiency.

2.3.3. *Similarity Measure*

To measure the degree of similarity between two images I (the query image) and J (the reference image), we compute the difference between feature descriptors H_I and H_J as a weighted histogram-intersection distance.[19] Images with smaller distances with the query image I are then retrieved as the searching results.

2.3.4. *Experimental Results*

For evaluation, a PET-CT database containing 50 sets of image scans from subjects diagnosed with NSCLC was used. We selected three key slices depicting the primary lung tumor from each patient scan, forming a database of 150 PET-CT slices. The ground truth indicating the similar or dissimilar relationships between each pair of slices were annotated. Two images were considered similar, if the tumors were in similar locations in the thorax (e.g. anterior or posterior), and showed similar spatial relationships relative to the chest wall and the mediastinum.

The retrieval performance was quantitatively compared with other methods, including two previous approaches proposed for tumor retrieval on PET-CT images,[19,52] and several techniques based on more standard algorithms (bag-of-words, SPM, and SIFT[54] features). Our proposed approach demonstrated clear performance improvements.

3. Summary

We have presented a brief review of our recent work on lesion detection and image retrieval for thoracic PET-CT imaging. For lesion detection, we described three different methods: region-based feature classification, multi-stage discriminative model, and data-adaptive structure estimation. While the second method is highly effective with its structural labeling and more discriminative features, the third approach requires only a simple feature set that is less empirical. For image retrieval, we described two different methods: pathological feature description, and spatial feature encoding. The second method presents a more discriminative hierarchical contextual spatial descriptor and improved the retrieval performance significantly. It is also worth noting that while some of the approaches, such as the region-based

classification and multi-stage discriminative model, were customized towards the specific problem domain, the data adaptive structure estimation and spatial feature encoding are more general and can be extended to other imaging applications.

References

1. S. B. Edge, D. R. Byrd, C. C. Compton, A. G. Fritz, F. L. Greene, and A. Trotti (Eds.), *AJCC cancer staging handbook, 7th ed.* (Springer, 2010).
2. W. D. Wever, S. Stroobants, J. Coolen, and J. A. Verschakelen, Integrated PET/CT in the staging of nonsmall cell lung cancer: technical aspects and clinical integration, *Eur. Respir. J.* **33**, 201–212, (2009).
3. Y. Song, W. Cai, S. Eberl, M. J. Fulham, and D. Feng, Thoracic image case retrieval with spatial and contextual information, *in Proc. ISBI.* pp. 1885–1888, (2011).
4. Y. Song, W. Cai, J. Kim, and D. Feng, A multi-stage discriminative model for tumor and lymph node detection in thoracic images, *IEEE Trans. Med. Imag.* **31**(5), 1061–1075, (2012).
5. Y. Song, W. Cai, Y. Zhou, and D. Feng, Thoracic abnormality detection with data adaptive structure estimation, *in MICCAI, LNCS.* **7510**, 74–81, (2012).
6. I. Jafar, H. Ying, A. F. Shields, and O. Muzik, Computerized detection of lung tumors in PET/CT images, *in Proc. EMBC.* pp. 2320–2323, (2006).
7. Y. Cui, B. Zhao, T. J. Akhurst, J. Yan, and L. H. Schwartz, CT-guided automated detection of lung tumors on PET images, *in SPIE Med. Imaging.* **6915**, 69152N, (2008).
8. C. Ballangan, X. Wang, S. Eberl, M. Fulham, and D. Feng, Automated lung tumor segmentation for whole body PET volume based on novel downhill region growing, *in SPIE Med. Imaging.* **7623**, 76233O, (2010).
9. J. Gubbi, A. Kanakatte, T. Kron, D. Binns, B. Srinivasan, N. Mani, and M. Palaniswami, Automatic tumour volume delineation in respiratory-gated PET images, *J. Med. Imag. Radia. Oncol.* **55**, 65–76, (2011).
10. G. V. Saradhi, G. Gopalakrishnan, A. S. Roy, R. Mullick, R. Manjeshwar, K. Thielemans, and U. Patil, A framework for automated tumor detection in thoracic FDG PET images using texture-based features, *in Proc. ISBI.* pp. 97–100, (2009).
11. S. Renisch, R. Opfer, and R. Wiemker, Towards automatic determination of total tumor burden from PET images, *in SPIE Med. Imaging.* **7624**, 76241T, (2010).
12. H. Gutte, D. Jakobsson, F. Olofsson, M. Ohlsson, S. Valind, A. Loft, L. Edenbrandt, and A. Kjaer, Automated interpretation of PET/CT images in patients with lung cancer, *Nucl. Med. Commun.* **28**(2), 79–84, (2007).
13. Y. Song, W. Cai, S. Eberl, M. Fulham, and D. Feng, Automatic detection of lung tumor and abnormal regional lymph nodes in PET-CT images, *J. Nucl. Med. 52.* **52** (Supplement 1), 211, (2011).
14. Y. Song, W. Cai, and D. Feng, Global context inference for adaptive abnormality detection in PET-CT images, *in Proc. ISBI.* pp. 482–485, (2012).
15. Y. Song, W. Cai, H. Huang, Y. Wang, and 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).
16. Y. Song, W. Cai, S. Eberl, M. J. Fulham, and D. Feng, Region and learning based retrieval for multi-modality medical images, *in Proc. Biomed.* pp. 723–063, (2011).
17. P. Perona and J. Malik, Scale-space and edge detection using anisotropic diffusion, *IEEE Trans. Pattern Anal. Mach. Intell.* **12**, 629–639, (1990).

18. C. Cortes and V. Vapnik, Support-vector networks, *Machine Learning.* **20**(3), 273–297, (1995).
19. 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.* **6893**, 191–198, (2011).
20. Y. Song, W. Cai, S. Eberl, M. Fulham, and D. Feng, A content-based image retrieval framework for multi-modality lung images, *in Proc. CBMS.* pp. 285–290, (2010).
21. A. Vedaldi and S. Soatto, Quick shift and kernel methods for mode seeking, *in ECCV, LNCS.* **5305**, 705–718, (2008).
22. J. Lafferty, A. McCallum, and F. Pereira, Conditional random fields: probabilistic models for segmenting and labeling sequence data, *in Proc. ICML.* pp. 282–289, (2001).
23. V. Kolmogorov and R. Zabih, What energy functions can be minimized via graph cuts?, *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(2), 147–159, (2004).
24. N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, *in Proc. CVPR.* pp. 886–893, (2005).
25. I. Sluimer, M. Prokop, and B. van Ginneken, Toward automated segmentation of the pathological lung in CT, *IEEE Trans. Med. Imag.* **24**(8), 1025–1038, (2005).
26. M. Sofka, J. Wetzl, N. Birkbeck, J. Zhang, T. Kohlberger, J. Kaftan, J. Declerck, and S. K. Zhou, Multi-stage learning for robust lung segmentation in challenging CT volumes, *in MICCAI, LNCS.* pp. 667–674, (2011).
27. T. Chen, B. C. Vemuri, A. Rangarajan, and S. J. Eisenschenk, Mixture of segmenters with discriminative spatial regularization and sparse weight selection, *in MICCAI, LNCS.* pp. 595–602, (2011).
28. F. Rousseau, P. A. Habas, and C. Studholme, Human brain labeling using image similarities, *in Proc. CVPR.* pp. 1081–1088, (2011).
29. J. Tropp, Greed is good: Algorithmic results for sparse approximation, *IEEE Trans. Inform. Theory.* **50**, 2231–2242, (2004).
30. H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler, A review of content-based image retrieval systems in medical applications - clinical benefits and future directions, *Int. J. Med. Inform.* **73**, 1–23, (2004).
31. Y. Song, W. Cai, Y. Zhou, L. Wen, and D. D. Feng, Hierarchical spatial matching for medical image retrieval, *in Proc. ISBI.* pp. 1–4, (2013).
32. W. Cai, J. Kim, and D. Feng, Content-based medical image retrieval, *Biomedical Information Technology, Chapter 4.* pp. 83–113, Edited by D. Feng, Elsevier, (2008).
33. W. Cai, D. Feng, and R. Fulton, Content-based retrieval of dynamic PET functional images, *IEEE Trans. Inf. Technol. Biomed.* **4**(2), 152–158, (2000).
34. W. Cai, S. Liu, L. Wen, S. Eberl, M. J. Fulham, and D. Feng, 3D neurological image retrieval with localized pathology-centric CMRGIc patterns, *in Proc. ICIP.* pp. 3201–3204, (2010).
35. B. Fischer, A. Brosig, P. Welter, C. Grouls, R. W. Gunther, and T. M. Deserno, Content-based image retrieval applied to bone age assessment, *in Proc. SPIE.* p. 762412, (2010).
36. D. Unay, A. Ekin, and R. S. Jasinschi, Local structure-based region-of-interest retrieval in brain MR images, *IEEE Trans. Inf. Technol. Biomed.* **14**(4), 897–903, (2010).
37. L. Sorensen, M. Loog, P. Lo, H. Ashraf, A. Dirksen, R. P. W. Duin, and M. D. Bruijne, Image dissimilarity-based quantification of lung disease from CT, *in MICCAI, LNCS.* **6361**, 37–44, (2010).
38. X. Qian, H. D. Tagare, R. K. Fulbright, R. Long, and S. Antani, Optimal embedding for shape indexing in medical image databases, *Med. Imag. Anal.* **14**, 243–254, (2010).

39. U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, X-ray image categorization and retrieval on the organ and pathology level, using patch-based visual words, *IEEE Trans. Med. Imag.* **30**(3), 733–746, (2011).
40. 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. Imag. Graph. (2011)*. (2011).
41. S. Yu, C. Chiang, and C. Hsieh, A three-object model for the similarity searches of chest CT images, *Comput. Med. Imag. Graph.* **29**, 617–630, (2005).
42. J. Kim, L. Constantinescu, W. Cai, and D. Feng, Content-based dual-modality biomedical data retrieval using co-aligned functional and anatomical features, *in MICCAI Workshop, LNCS*. pp. 45–52, (2007).
43. A. Kumar, J. Kim, L. Wen, and D. Feng, A graph-based approach to the retrieval of volumetric PET-CT lung images, *in Proc. EMBC*. pp. 5408–5411, (2012).
44. Y. Song, W. Cai, S. Eberl, M. Fulham, and D. Feng, Structure-adaptive feature extraction and representation for multi-modality lung images retrieval, *in Proc. DICTA*. pp. 152–157, (2010).
45. Y. Song, W. Cai, and D. Feng, Disease-specific context modeling and retrieval with fast structure localization, *in MMBIA*. pp. 89–94, (2012).
46. W. Cai, Y. Song, and D. Feng, Regression and classification based distance metric learning for medical image retrieval, *in Proc. ISBI*. pp. 1775–1778, (2012).
47. S. Belongie, J. Malik, and J. Puzicha, Shape matching and object recognition using shape contexts, *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(24), 509–522, (2002).
48. M. M. Rahman, S. K. Antani, and G. R. Thoma, A medical image retrieval framework in correlation enhanced visual concept feature space, *in Proc. CBMS*. pp. 1–4, (2009).
49. S. Lazebnik, C. Schmid, and J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, *in Proc. CVPR*. pp. 2169–2178, (2006).
50. J. Feulner, S. K. Zhou, E. Angelopoulou, S. Seifert, A. Cavallaro, J. Hornegger, and D. Comaniciu, Comparing axial CT slices in quantized N-dimensional SURF descriptor space to estimate the visible body region, *Comput. Med. Imag. Graph.* **35**(3), 227–236, (2011).
51. Y. Song, W. Cai, S. Eberl, M. Fulham, and D. Feng, Thoracic image matching with appearance and spatial distribution, *in Proc. EMBC*. pp. 4469–4472, (2011).
52. Y. Song, W. Cai, and D. D. Feng, Hierarchical spatial matching for medical image retrieval, *in Proc. ACM MM Workshop*. pp. 1–6, (2011).
53. J. Matas, O. Chum, M. Urban, and T. Pajdla, Robust wide baseline stereo from maximally stable extremal regions, *in Proc. BMVC*. pp. 384–396, (2002).
54. D. G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.* **60**(2), 91–110, (2004).