# PATHOLOGY-CENTRIC MEDICAL IMAGE RETRIEVAL WITH HIERARCHICAL CONTEXTUAL SPATIAL DESCRIPTOR

*Yang Song*[1], *Weidong Cai*[1], *Yun Zhou*[2], *Lingfeng Wen*[3], *David Dagan Feng*[1]

[1]Biomedical and Multimedia Information Technology (BMIT) Research Group,
School of Information Technologies, University of Sydney, Australia
[2]The Russell H. Morgan Department of Radiology and Radiological Science,
Johns Hopkins University School of Medicine, Baltimore, USA
[3]Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Sydney, Australia

## ABSTRACT

Content-based image retrieval has been suggested as an aid to medical diagnosis. Techniques based on standard feature descriptors, however, might not represent optimally the pathological characteristics in medical images. In this paper, we propose a new approach for medical image retrieval based on pathology-centric feature extraction and representation; and patch-based local feature extraction and hierarchical contextual spatial descriptor are designed. The proposed method is evaluated on positron emission tomography – computed tomography (PET-CT) images from subjects with non-small cell lung cancer (NSCLC), showing promising performance improvements over the other benchmarked techniques.

***Index Terms***— Retrieval, tumor, local, spatial, context

## 1. INTRODUCTION

In the past three decades, medical image data have been expanded rapidly, and the vast amount of data also opens an opportunity as a case-based reasoning support [1]. A physician may choose to browse through similar images in the database as a reference set to interpret a patient's imaging scan. The ability to retrieve images with similar contents is thus important to facilitate such diagnosis supports.

It is commonly acknowledged that feature extraction and representation are the key components for achieving good retrieval performance. Depending on whether the retrieval objective is to differentiate between a wide variety of images (e.g. various imaging modalities) or between images of a certain type of disease (e.g. lung cancer), the features designed can be quite generic [2] or adaptive to the disease [3, 4, 5]. Furthermore, the similarity measure between feature vectors of different images is also essential for accurate retrievals, and learning-based methods are often used [6, 7].

In the cases of NSCLC, we would like to retrieve images showing similar pathological patterns related to lung cancer staging. In particular, tumors at different locations, such as adjacent to the mediastinum or chest wall, or well-within the lung field should be differentiated. The size and shape of a tumor are also distinctive factors. With the PET imaging, tumors can be identified relatively easily by finding areas with abnormal uptake values in the thorax. The spatial location of a tumor can then be analyzed based on the surrounding anatomical structures from the co-registered CT images. However, tumors and other structures normally exhibit large inter-subject variations; and therefore, the features designed are expected to accommodate the variations while being discriminative enough for different tumor patterns.

For this problem domain, we observe that a disease-adaptive feature extraction and representation would likely result in better retrieval performance than a generic approach. Not much works have been reported in this field. We have proposed region-localization based approaches [4, 7]. Such approaches, however, are sensitive to the localization accuracy of the tumor and surrounding structures. A further study then attempts to reduce such an accuracy requirement based on tumor centroid detection and hierarchical spatial matching [5]. The spatial features designed, however, are based on a rather rigid rectangular structure and might cause some misrepresentation of the tumor characteristics.

In this work, we propose a method for retrieving PET-CT thoracic images with spatially similar tumors. First, a set of patch-based local features based on texture and local spatial information are extracted and represented as feature words. Then, a new hierarchical contextual spatial descriptor is designed to represent the spatial information of the tumor. Briefly, originated from the pathological centroid detected, a hierarchical subdivision is constructed with circular structures and adaptive radial creation. Such a spatial descriptor is able to better represent the surrounding contexts of the tumor, with an adequate balance between its discriminative capability and geometric-transformation invariance suitable for our problem domain. Finally, a learning-based similarity measure is also incorporated to obtain good retrieval performance.

This work was supported in part by ARC grants.

## 2. METHODS

### 2.1. Local Feature Extraction

We first introduce a set of local features to represent an image $I$ using the bag-of-features (BOF) model. Since we observe that various thoracic structures can be well characterized by their textures and local spatial features, the local features are computed using Gabor filters and auto-correlation matrix [8].

Specifically, a total of 24 Gabor filters with six radial frequencies and four orientations are used. An image is then represented as a variance-weighted average from all outputs of the filter bank, to produce a feature image $\overline{G}$:

$$\overline{G} = (\sum_{k=1}^{24} \sigma_k G_k)/(\sum_{k=1}^{24} \sigma_k) \quad (1)$$

where $\sigma_k$ is the standard deviation of the $k$th filtered image $G_k$. The feature image $\overline{G}$ is then divided into a regular grid of image patches ($4 \times 4$ pixels), and for each patch $p_i$, we compute its mean, minimum and maximum PET and CT values as the texture feature $f_1(i)$.

The local spatial feature $f_2(i)$ of an image patch $p_i$ describes the co-occurrence statistics of the neighborhood of $p_i$. The first-order auto-correlation is used here:

$$f_2(i, a) = f_1(i) f_1(i + a)^T \quad (2)$$

where $a$ is the displacement vector indexing the 8 immediate neighbors. And $f_2(i)$ is thus the concatenation of all $f_2(i, a)$.

The combined local feature vector of an image patch $p_i$ is thus $f(i) = \{f_1(i)|f_2(i)\}$, with $(6 + 8 \times 6 \times 6) = 294$ dimensions. Each $f(i)$ is then converted into a feature word representing one of $K = 16$ feature categories using $k$-means clustering. A number larger than 16 would sometimes produce empty clusters based on our experiments.

Denote an image $I$ as a grid of $N$ patches $P_I = \{p_i : i = 1, ..., N\}$, and $w_i$ (of dimension $K$) as the feature word of $p_i$ with only one nonzero element in $w_i$ indicating the assigned feature word. The image $I$ is thus now represented as a grid of feature words $W_I = \{w_i : i = 1, ..., N\}$ (example shown in Fig. 1).

### 2.2. Hierarchical Contextual Spatial Descriptor

Next, based on the local feature words generated, a hierarchical contextual spatial descriptor is designed to encode the spatial information surrounding the pathological object. The descriptor is also partially size-, translation- and rotation-invariant, optimized for the thoracic imaging domain.

#### *2.2.1. Pathological Centroid Detection*

To build the spatial descriptor, we first detect the centroid of the pathology, which is the geometric center of a tumor. A straightforward idea is that if we can detect a region that roughly defines the tumor, we can then compute the geometric center of the region as the tumor centroid.

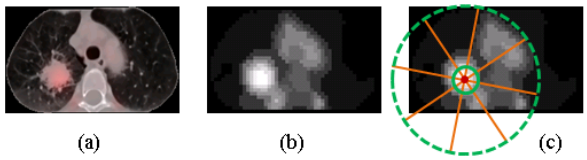


**Fig. 1**. Illustration of the proposed feature extraction and representation. (a) a fused PET-CT slice, showing a primary lung tumor. (b) the feature-word grid, representing each image patch with its feature word value. (c) the hierarchical partition of the spatial contexts, showing only the level-2 structure for simplicity, with the red dot indicating the tumor centroid, green circles depicting the concentric circles, and orange lines dividing the radials.

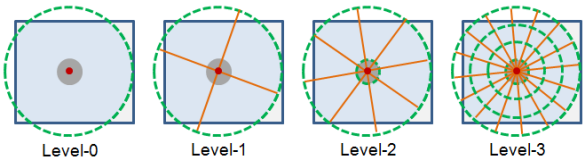


**Fig. 2**. Illustration of the proposed hierarchical contextual spatial descriptor, showing four levels of radial partitions.

To do this, we first extract the maximally stable extremal regions (MSER) [9] from the image grid of feature words $W_I$. The MSER algorithm is particularly effective for detecting the tumor area, since tumors exhibit much higher uptake values from the rest of the thorax, and rather than defining a global optimal threshold, MSER can determine the local threshold. 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 in Fig. 1).

#### *2.2.2. 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 the spatial pyramid matching (SPM) [10]; but rather than dividing a rectangle cell evenly into 4 sub-cells as in SPM, we design a new hierarchical circular structure, as illustrated in Fig. 2. 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 individ-

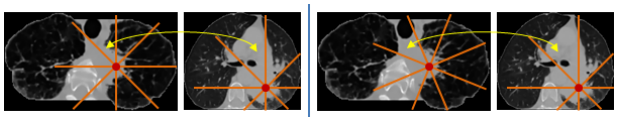

**Fig. 3**. Illustration of the radial creation with level-2 radials. The left and right pairs show the standard and the proposed subdivision schemes, respectively. The arrows indicate the corresponding radial sections, and a better matching between the two images can be seen from the right pair.

ual 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 $K \sum_{l=0}^{L-1} 4^l$. The details of the descriptor construction are as following.

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:

$$H_I^{l=0} = \sum_{i=1}^{N} \lambda_i w_i \tag{3}$$

where each feature word $w_i$ is Gaussian weighted ($\lambda_i$) according to its Chebyshev distance $d_i$ from $p_o$:

$$\lambda_i = \exp\{-(2d_i^2)/(r_I^2)\} \tag{4}$$

With the weights $\lambda_i$, image patches nearer to the centroid would contribute more to the feature histogram.

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 and the level-1 feature vector $H_I^{l=1}$ is thus the concatenation of all four feature vectors. With four radials, each would be more close to the medial, anterior, lateral or posterior of the thorax, and hence the feature concatenation also follows this order to obtain an anatomical-meaning representation of the surrounding structure.

When creating the radials, a basis radial line is usually first chosen at a certain direction, such as $0^\circ$, and the rest of radial lines can be drawn at even intervals. Such standard approaches, however, introduce a random partition, potentially causing mismatches between two images with slight rotations (example in Fig. 3). Our approach is to reduce fragmented segments due to dividing in homogeneous regions; and the idea is to create a partition that minimizes the co-occurrences of multiple structure types within one radial by minimizing the total variance. Denote the $j$th radial as $R_I(j, \theta)$ with $\theta$ representing the direction of the basis radial. The total variance $v_I(J, \theta)$ of such a partition is:

$$v_I(J, \theta) = \sum_{j=1}^{J} \sum_{i=1}^{N} (w_i - m_j)^2, \ s.t. \ p_i \in R_I(j, \theta) \tag{5}$$

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:

$$\theta = argmin_\theta v_I(J, \theta), \ \forall \theta = \{0, \frac{\pi}{2J}, \frac{\pi}{J}, \frac{3\pi}{2J}\} \tag{6}$$

Here we choose to test four possible $\theta$ only, for convenient implementation and better efficiency.

At a higher level ($l \geq 2$), we create $4^l$ radials from $O_I$ by the following two steps. First, $O_I$ is divided into $2^{l-1}$ concentric circles, with the inner most circle of radius $r_{I,1}$ and the other circles dividing the remaining area of $O_I$ into even-interval contours. The value $r_{I,1}$ is determined as the radius of the MSER region. Next, the second inner circle is divided into $2^{l+1}$ radials in the same adaptive way as the four-radial division at level-1, and radial lines are then extended into the inner most and outer circles to create all $4^l$ radials.

### 2.3. 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 [7]. And images with smaller distances with the query image $I$ are then retrieved as the searching results.

## 3. RESULTS

For evaluation, a PET-CT database containing 50 sets of image scans from subjects diagnosed with NSCLC is 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, based on if both tumors appearing at similar locations in the thorax (e.g. anterior or posterior), and with similar spatial relationships relative to the chest wall and mediastinum. Following the leave-one-subject-out scheme, for each query image, only images from different subjects would be retrieved.

Three query examples using the proposed approach are shown in Fig. 4. Take the first example for brief explanations: the query image contains a primary lung tumor of considerably large size at the posterior lung field and adjacent to the mediastinum. The first image retrieved exhibits the most similarity, and the second image retrieved is ranked higher than the third one due to its similarity in overall thoracic appearances with the query image.

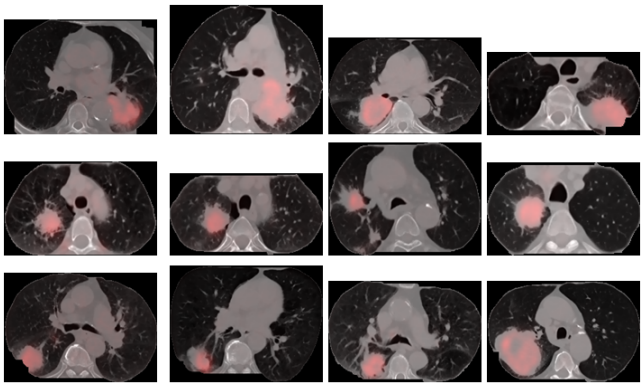

**Fig. 4**. From top to bottom, each row shows one retrieval example with the tumor: adjacent to the mediastinum, within the lung field but near to the mediastinum, and attached to the chest wall. The left most column presents the query image, and the top three similar images are then shown.

The retrieval performance is also quantitatively compared with other methods, including two previous approaches proposed for tumor retrieval on PET-CT images – the hierarchical spatial matching (HSM) [5] with similar design of local features and similarity measure as the proposed method but with different spatial feature representation, and a region localization approach [7]. Several techniques based on more standard algorithms and combinations with the proposed components are also compared: (1) BOF-LF – BOF model with the proposed local features; (2) SPM-LF – SPM based on BOF-LF; (3) SPM-Cent – SPM-LF with a subdivision center at the tumor centroid detected; (4) BOF-SIFT – BOF model with SIFT features; and (5) SPM-SIFT – SPM based on BOF-SIFT. As shown in Fig. 5, the proposed approach exhibits clear performance improvements. The advantage of BOF-LF over BOF-SIFT suggests the effectiveness of the local features designed. The further improvement by SPM-LF indicates the suitability of integrating a hierarchical feature representation. And the performance difference between our proposed method and SPM-LF demonstrates the benefit of the hierarchical contextual spatial descriptor.

## 4. CONCLUSIONS

In this paper, we presented a new pathology-centric feature extraction and representation method for medical image retrieval. The local features describing textures and local spatial information are first extracted to obtain feature words. A hierarchical contextual spatial descriptor is then designed to effectively represent the spatial features pertaining to the lung tumors. A learning-based similarity measure is also incorporated. The proposed method can also be extended to other medical imaging domains, with slight modifications for the specific diseases, such as the pathology centroid detection and the radius computation of the concentric circles.

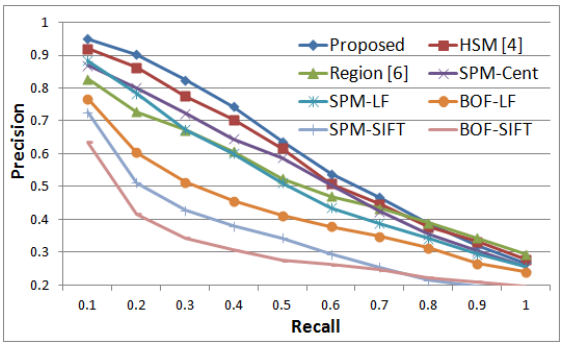


**Fig. 5**. Retrieval precisions, compared with other methods.

## 5. REFERENCES


[1] 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.*, vol. 73, pp. 1–23, 2004.

[2] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using path-based visual words," *IEEE Trans. Med. Imag.*, pp. 733–746, 2011.

[3] 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.

[4] 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.

[5] Y. Song, W. Cai, and D. D. Feng, "Hierarchical spatial matching for medical image retrieval," *in Proc. ACM MM Workshop*, pp. 1–6, 2011.

[6] M. M. Rahman, S. K. Antani, and G. R. Thoma, "A biomedical image retrieval framework based on classification- driven image filtering and similarity fusion," *in Proc. ISBI*, pp. 1905–1908, 2011.

[7] Y. Song, W. Cai, S. Eberl, M.J. Fulham, and D. Feng, "Discriminative pathological context detection in thoracic images based on multi-level inference," *in MICCAI, LNCS*, pp. 185–192, 2011.

[8] T. Harada, H. Nakayama, and Y. Kuniyoshi, "Improving local descriptors by embedding global and local spatial information," *in ECCV, LNCS*, pp. 736–749, 2010.

[9] 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.

[10] 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.