Hierarchical and Binary Spatial Descriptors for Lung Nodule Image Retrieval

Gillian Ng, Yang Song, *Member, IEEE,* Weidong Cai, *Member, IEEE,* Yun Zhou, Sidong Liu, *Student Member, IEEE,* David Dagan Feng, *Fellow, IEEE*

***Abstract*— With the increasing amount of image data available for cancer staging and diagnosis, it is clear that content-based image retrieval techniques are becoming more important to assist physicians in making diagnoses and tracking disease. Domain-specific feature descriptors have been previously shown to be effective in the retrieval of lung tumors. This work proposes a method to improve the rotation invariance of the hierarchical spatial descriptor, as well as presents a new binary descriptor for the retrieval of lung nodule images. The descriptors were evaluated on the ELCAP public access database, exhibiting good performance overall.**

I. Introduction

Advances in medical imaging technology have made a large amount of data available for application areas such as staging and diagnosis. Images from previous patients can be used as a reference by physicians for the current diagnosis. However, the process of manually browsing through large image datasets is tedious and error-prone. Textual search engines are insufficient for this purpose as images contain information that is difficult to describe in text. Content-based image retrieval techniques can allow physicians to quickly retrieve images with similar contents to a query image, without relying on metadata such as keywords or text descriptions. This increases the efficiency of the diagnostic process.

A lot of work has been done on the retrieval of medical images [1, 2, 3]. It is well-known that lung cancer is one of the leading causes of cancer deaths. Thus we are interested in improving the accuracy and speed of existing lung tumor retrieval methods for use in a clinical setting. Based on a query image containing a lung nodule, we would like to retrieve images with similar types of nodules. The nodules fall into four categories — juxta-pleural, vascularized, well-circumscribed, and pleural-tail. Juxta-pleural nodules are connected significantly to the pleural surface. Vascularized nodules are located centrally in the lung, but connected to the neighboring vessels. Well-circumscribed nodules are located centrally in the lung while not being connected to vasculature. Pleural-tail nodules are connected by a thin structure to the pleural surface.

An important factor in achieving good retrieval performance is feature description, which is how the visual features extracted from images are represented. It determines whether two images are regarded as similar by the system, as well as the speed of the retrieval. Many descriptors have been developed for general content-based image retrieval as well as for specific problems like the aforementioned lung tumor retrieval.

One of the classic feature descriptors is the scale-invariant feature transform (SIFT) [4]. It is robust and invariant to scaling, orientation, and partially invariant to affine distortion and illumination changes. Although it is relatively fast, the feature vector is high-dimensional. It has been successfully applied to lung nodule classification [5, 6, 7]. Based on SIFT's work on invariance, many newer feature descriptors were developed. Speeded Up Robust Features (SURF) [8] is fast and accurate. Instead of incorporating invariance to many factors, it focuses on scale and rotation invariance, which offers a good compromise between feature complexity and robustness. It has been applied to the same lung nodule dataset used in our work, with good results [9]. Another type of feature with good performance that incorporates invariance is Local Binary Patterns (LBP) [10]. It is robust in terms of gray-scale and rotation invariance. One example of its application to the medical domain is in tumor detection [11]. Binary Robust Independent Elementary Features (BRIEF) [12] is a newer descriptor which is fast to compute and match. It was developed with a focus on speed, which is relevant to us as speed is necessary for effective use in the clinical setting. However, in contrast to the aforementioned descriptors, it is sensitive to distortions, rotation and translation. Despite this, it has provided good results to the lung image classification problem when combined with other features [13].

It has been shown in previous works that domain-specific feature descriptors are more effective [7, 14, 15]. One example of a domain-specific feature descriptor is the hierarchical contextual spatial descriptor [7]. Customized for lung tumor retrieval, it exhibits excellent performance on positron emission tomography-computed tomography (PET-CT) images, and is able to represent the spatial features accurately. Therefore, it is used as a basis for our work.

Our contributions from this work are twofold. Firstly, the hierarchical contextual spatial descriptor is improved by reducing its sensitivity to rotation. Secondly, a new feature descriptor based on the hierarchical descriptor and BRIEF, is proposed. We show that the two feature descriptors are

This work was supported in part by ARC grants.

G. Ng, Y. Song, W. Cai, and S. Liu are with the Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, Australia.

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

D. D. Feng is with BMIT Research Group, School of Information Technologies, University of Sydney, 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, China.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

discriminative enough to retrieve CT images with spatially similar lung nodules.

II. METHODS

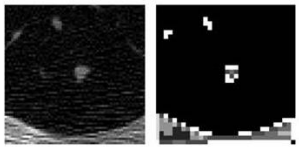
A. Feature Extraction

As in the pathology-centric method [7], the image is first represented with the bag-of-features model. Instead of using the entire nodule CT image, our method only considers a 32×32 region around the nodule, as the nodule region contains the most discriminative information for the retrieval. The nodule region is first divided into a regular grid of 2×2 patches. The mean, minimum and maximum values of each patch p_i are computed as the texture feature $f_1(i)$. The first-order autocorrelation $f_2(i)$ is used to describe the neighborhood of each patch:

$$f_2(i,a) = f_1(i)f_1(i+a)^T \quad (1)$$

where a is the displacement vector indexing the 8 immediate neighbors of i. The feature vector $f(i)$ for each patch is then the concatenation of $f_1(i)$ and $f_2(i)$.

We then represent the image using a bag-of-features approach similar to [16]. Specifically, patch feature vectors are computed on a training set of images. They are then quantized into 8 types of feature words using k-means clustering. Using more than 8 types of feature words resulted in inaccurate labeling of the patches. Each image is then represented as a 16×16 grid of feature words, on which the descriptor can be computed, as displayed in Figure 1.



(a) (b)

Figure 1. Illustration of the feature word representation. (a) CT image, (b) feature word representation.

B. Hierarchical Descriptor

To create the hierarchical contextual spatial descriptor, concentric circles are created around the lung nodule at L levels, as shown in Figure 2. As we are interested in feature description, and there is a lot of work done on nodule detection, we do not detect the location of the nodules however assume that it is provided.

Each concentric circle is divided into a number of segments, depending on the level of hierarchy. In this way, the feature descriptor incorporates pathology-centric information. Histograms of the feature words are computed on each radial, and the final descriptor is the concatenation of all the histograms. The finer details of this process are similar to the existing hierarchical descriptor [7].

Our work improves on the hierarchical descriptor by reducing its sensitivity to rotation. The radials are aligned so that the first radial always contains pixels representing parts of the anatomy close to the nodule, such as the vascular. To determine the rotation of the descriptor, only a small region (32×32) was considered, as the more discriminative parts of the anatomy are close to the nodule.

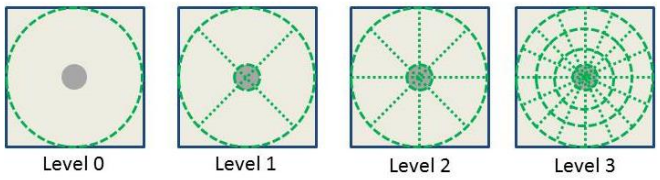


Figure 2. Illustration of the hierarchical descriptor with four levels of radial partitions. The descriptor is centered on the nodule region.

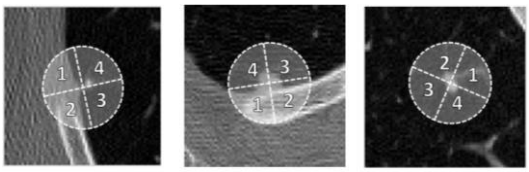


Figure 3. Illustration of the descriptor rotation computation. Three examples of CT images with the rotation of their descriptors are shown above. The radials are rotated so that radial 1 always has the largest difference in the number of ROI pixels with radial 4.

First, a Gabor filter bank with 6 radial frequencies and 4 orientations is applied. The image is then represented as a variance-weighted average from the outputs of the filter bank. The ROI pixels are then segmented using Otsu's method of binary thresholding. The region is then divided into four radials. The jth radial with the rotation θ is denoted as $R(j,\theta)$. The selected $R(j,\theta)$ becomes the first radial in the descriptor (Figure 3). The selected rotation θ is the one which minimizes the number of radials containing ROI points, and the selected radial j maximizes the difference between the number of ROI points in the current radial and the previous one. For computational efficiency, four values of θ were tested: 0, $\pi/8$, $\pi/4$, $3\pi/8$. As our implementation of the hierarchical spatial descriptor utilized four levels (L=4), the length of the feature descriptor was 8×(1+8+16+64) = 712.

C. Binary String Representation

Although the hierarchical descriptor provides good results, it is high-dimensional. Inspired by BRIEF's performance, we developed a binary descriptor which improves the retrieval speed by reducing the size of the feature vector, yet provides similar accuracy.

A total of n pairs were selected from the grid of feature words, on which tests are performed. To compare the values of the feature words of each pair $P<a, b>$ the individual tests T are defined as:

$$T(P) = \begin{cases} 1, & if\ a = b \\ 0, & otherwise \end{cases} \quad (2)$$

where a and b are the feature word values of each pair. The binary string representing each image is then the concatenation of all the tests:

$$B_n = \sum_{i=1}^{n} 2^{i-1} T(P_i) \quad (3)$$

A few different methods of pair selection were tested, including randomly selecting pairs on the grid, from each radial in the hierarchical descriptor, as well as uniformly across the grid. The selection of pairs around the nodule centroid in a circular manner (Figure 4) provided the best result. Eight concentric circles are created around the nodule, and pairs of pixels are selected at regular intervals along the circles. The pairs were selected at intervals of $2\pi/7$, determined empirically. With 7 pairs selected per circle, the length of the binary string representing each image was 56, much shorter than the hierarchical descriptor.

Figure 4. Illustration of pair selection around the nodule centroid. The points are selected at regular intervals.

D. Similarity Measurement

Our retrieval process consists of two steps. First, a query image is classified with support vector machines (SVM). With the image type from the SVM, the top 10 most similar images in the predicted class are then retrieved using a similarity metric. For the hierarchical descriptor, the histogram intersection is used as a similarity metric. For the binary string representation, the Hamming distance is used. The SVM helps to identify the image type, thus improving the retrieval accuracy over using just the histogram intersection or the Hamming distance. This method is also advantageous as the number of comparisons between the query image and the images in the database is reduced, due to the query image being compared with only those in the predicted class.

III. Experiments and Results

A. Datasets

Our work was evaluated on the Early Lung Cancer Action Program (ELCAP) Public Lung Image Database [17]. The dataset contains 50 sets of low-dose whole-lung CT scans obtained in a single breath hold with a 1.25mm slice thickness. Each image in the database is annotated with the nodule type and location. A total of 379 lung nodules are annotated. The nodules are of four types — 57 juxta-pleural nodules, 61 vascularized, 114 well-circumscribed, and 147 pleural-tail. During preprocessing, a window of 32×32 pixels is cropped as the region-of-interest (ROI) with the annotated nodule at the centroid. Image retrieval is then performed on these ROIs.

B. Comparisons and Results

Following the leave-one-subject-out scheme, only images from different subjects would be retrieved. Three query examples using the modified hierarchical descriptor as well as the binary string representation are shown in Figure 5. It can be seen that the results retrieved by the hierarchical descriptor are rotation invariant. The chest wall is present in all three of the retrieved images, although the rotation is different. This may or may not be desirable depending on whether it is necessary to differentiate nodules based on their location in the lungs. Despite the difference in descriptor used, there is some overlap between the top 3 images, and the chest wall is present in all the images.

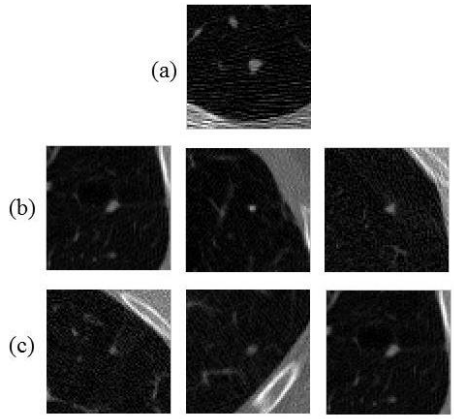


Figure 5. Illustration of top 3 retrieval results. (a) query image, (b) retrieval results with the modified hierarchical descriptor, (c) retrieval results with the binary descriptor.

The retrieval performance is quantitatively compared between the proposed hierarchical and binary descriptors with SVM for classification (H-SVM and B-SVM), as well as (1) the bag-of-features model (Section II.A) with SVM, (2) the state-of-the-art in lung nodule retrieval using SIFT with linear discriminant analysis (LDA) [5], and (3) SIFT with SVM. A training set composed of 50% of each nodule type is used for training the SVM and LDA. For SVM, the polynomial kernel is found to provide the best results.

The F-scores for each of the nodule types are given in Figure 6, and the overall precision, recall, and F-score for each of the descriptors are listed in Table 1. It can be seen that the hierarchical descriptor provides the best results. For most of the nodule types, the binary string descriptor performs slightly worse than the hierarchical descriptor, which is to be expected, as invariance is not incorporated into the descriptor. In situations where a small drop in accuracy is acceptable, the binary string descriptor provides a balance between speed and performance. Overall, the binary descriptor provides similar performance to SIFT. However, the binary descriptor is faster to compute than SIFT. The results also show that SVM is a more suitable classifier than LDA in this problem domain.

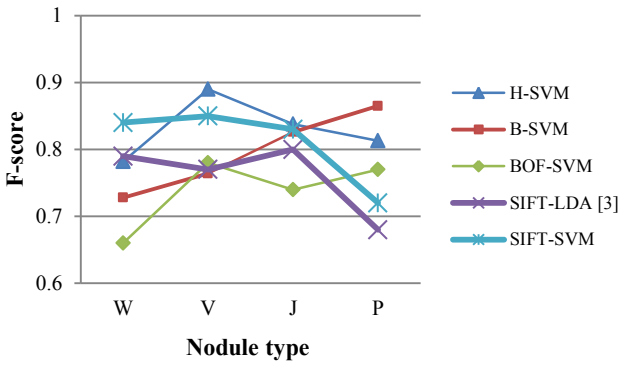


Figure 6. F-scores for each of the nodule types, comparing the proposed hierarchical (H-SVM) and binary (B-SVM) descriptors combined with SVM, and BOF-SVM, SIFT-LDA [5] and SIFT-SVM. The nodule types are indicated by W: well-circumscribed, V: vascularized, J: juxta-pleural, and P: pleural-tail.

TABLE 1. RESULTS ON OVERALL RETRIEVAL PRECISION, RECALL AND F-SCORE, COMPARING THE VARIOUS DESCRIPTORS.

	Precision	**Recall**	**F-score**
H-SVM	0.83	0.83	0.83
B-SVM	0.79	0.80	0.80
BOF-SVM	0.74	0.74	0.74
SIFT-LDA [3]	0.75	0.76	0.76
SIFT-SVM	0.80	0.80	0.80

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed an improvement to the hierarchical descriptor, to improve its invariance to rotation. We also presented a new feature descriptor based on BRIEF. A summary of the workflow is as follows. Feature words are computed on each image, and the features represented by the hierarchical or binary string descriptor. Retrieval is then done by first detecting the nodule image type using SVM, and then using the Hamming distance as a similarity metric. The top 10 images can then be retrieved.

The two feature descriptors provided similar results, although the binary string representation performed slightly worse. However, the shorter binary string representation provides a good compromise in favor of speed. Additional work can be done to incorporate rotation invariance into the binary string representation, similar to LBP. The proposed method can also be applied to other medical imaging domains, with domain-specific modifications.

In future work, we will investigate extending the feature descriptor design to 3D, to fully utilize the 3D information provided in CT images. We will also perform more comprehensive evaluation by comparing with the other more recent feature descriptors, such as SURF.

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. Medical Informatics, vol 73, pp.1-23, 2004.

[2] W. Cai, J. Kim, and D. Feng, "Content-based medical image retrieval," Biomedical Information Technology, pp.211–227, 2008.

[3] C. Akgül, D. Rubin, S. Napel, C. Beaulieu, H. Greenspan, and B. Acar, "Content-Based Image Retrieval in Radiology: Current Status and Future Directions", Journal of Digital Imaging, vol. 24, no. 2, pp.208-222, 2011.

[4] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, vol. 60, no. 2, pp.91-110, 2004.

[5] A. Farag, S. Elhabian, J. Graham, A. Farag, and R. Falk, "Toward Precise Pulmonary Nodule Descriptors for Nodule Type Classification", in Medical Image Computing and Computer-Assisted Intervention - MICCAI 2010, Pt III, LNCS, vol. 6363, Springer-Verlag Berlin, Berlin, 2010.

[6] Y. Song, W. Cai, Y. Wang, and D.D. Feng, "Location classification of lung nodules with optimized graph construction", in ISBI, pp.1439-1442, 2012.

[7] Y. Song, W. Cai, Y. Zhou, L. Wen, and D.D. Feng, "Pathology-centric medical image retrieval with hierarchical contextual spatial descriptor", in ISBI, pp.198-201, 2013.

[8] H. Bay, T. Tuytelaars, and L. Gool, "SURF: Speeded Up Robust Features", in ECCV, LNCS, vol. 3951, pp. 404-417, 2006.

[9] A. Farag, A. Ali, J. Graham, S. Elhabian, A. Farag, and R. Falk, "Feature-based lung nodule classification", in Advances in Visual Computing, LNCS, vol. 6455, Springer, 2010.

[10] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, no. 7, pp.971-987, 2002.

[11] L. He, L.R. Long, S. Antani, and G.R. Thoma, "Histology image analysis for carcinoma detection and grading", Computer Methods and Programs in Biomedicine, vol. 107, no. 3, pp.538-556, 2012.

[12] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary Robust Independent Elementary Features", in ECCV, LNCS, vol. 6314, pp. 778-792, 2010.

[13] Y. Song, W. Cai, S. Huh, M. Chen, T. Kanade, Y. Zhou, and D. Feng, "Discriminative Data Transform for Image Feature Extraction and Classification", MICCAI 2013, LNCS, vol. 8150, pp.452-459, 2013.

[14] 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 CVPRW, pp.1-7, 2012.

[15] 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", The Journal of Nuclear Medicine, vol. 52(suppl), 211, 2011.

[16] Y. Song, W. Cai, S. Eberl, M. F. Fulham, and D. Feng, "A Content-based Image Retrieval Framework for Multi-Modality Lung Images", in IEEE Symp. Computer-Based Medical System, pp. 285-290, 2010.

[17] ELCAP Public Lung Image Database. Available from: http://www.via.cornell.edu/lungdb.html.