Research Article

Sparse Codebook Model of Local Structures for Retrieval of Focal Liver Lesions Using Multiphase Medical Images

Jian Wang,[1] Xian-Hua Han,[2] Yingying Xu,[3] Lanfen Lin,[3] Hongjie Hu,[4] Chongwu Jin,[4] and Yen-Wei Chen[1,3]

[1]*Graduate School of Information Science and Engineering, Ritsumeikan University, Kusatsu, Japan*
[2]*National Institute of Advanced Industrial Science and Technology, Tokyo, Japan*
[3]*College of Computer Science and Technology, Zhejiang University, Hangzhou, China*
[4]*Department of Radiology, Sir Run Run Shaw Hospital, Hangzhou, China*

Correspondence should be addressed to Yen-Wei Chen; chen@is.ritsumei.ac.jp

Characterization and individual trait analysis of the focal liver lesions (FLL) is a challenging task in medical image processing and clinical site. The character analysis of a unconfirmed FLL case would be expected to benefit greatly from the accumulated FLL cases with experts' analysis, which can be achieved by content-based medical image retrieval (CBMIR). CBMIR mainly includes discriminated feature extraction and similarity calculation procedures. Bag-of-Visual-Words (BoVW) (codebook-based model) has been proven to be effective for different classification and retrieval tasks. This study investigates an improved codebook model for the fined-grained medical image representation with the following three advantages: (1) instead of SIFT, we exploit the local patch (structure) as the local descriptor, which can retain all detailed information and is more suitable for the fine-grained medical image applications; (2) in order to more accurately approximate any local descriptor in coding procedure, the sparse coding method, instead of K-means algorithm, is employed for codebook learning and coded vector calculation; (3) we evaluate retrieval performance of focal liver lesions (FLL) using multiphase computed tomography (CT) scans, in which the proposed codebook model is separately learned for each phase. The effectiveness of the proposed method is confirmed by our experiments on FLL retrieval.

1. Introduction

Liver cancer is one of the leading causes of death worldwide. The development of medical imaging techniques, such as computed tomography (CT) and magnetic resonance imaging (MRI), gives more and more detailed information about the inner structure of human body. This detailed information prospects the possibility of early detection for some types of liver lesions, while early diagnosis and treatment is the most effective way to reduce the liver cancer death. On the other hand, multiphase contrast enhanced CT scan is generally employed as the primary technique for detection and characterization of liver lesions, due to the clinical observation that different types of liver lesions exhibit different visual characteristics at various time points after intravenous contrast injection. In multiphase contrast

enhanced CT scan procedure, the transition of visual features over time is captured from CTs scanned before and after contrast injection. The noncontrast enhanced (NC) phase image is obtained from scans before injection and three more phases' images are scanned at different times after contrast injection, including arterial (ART) phase, portal venous (PV) phase, and delayed (DL) phase, scanned at 25~40 s, 60~75 s, and 3~5 minutes after contrast injection, respectively. Due to the development of medical imaging techniques and the availability of the multiphase CT scans, medical volumes have become of higher and higher resolution and larger-scale, which prospects more accurate diagnosis while simultaneously leading to heavy burden to radiologists and medical experts for interpreting the CT images. On the other hand, with the ICT technology, it also becomes potential for medical experts to share more and more medical data
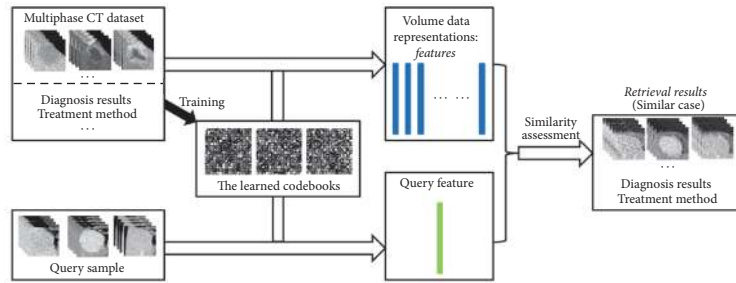
FIGURE 1: The proposed CBMIR system for FLLs retrieval.

and the corresponding diagnosis and treatment experiences. However, with the accumulated large-scale medical data with experienced clinical cases, how to search the most similar cases to a query sample (only medical image) for assistance in making decisions of diagnosis and treatment becomes an important and challenging task in both academic and clinical sites. This study concentrates on the similar case retrieval of focal liver lesions to a CT image query with multiphase CT scans.

Characterization of liver lesions is gaining more and more research attention recently, where most work focused on the discriminated feature representation of the medical volume images. In [1], a CBIR system was proposed for retrieving the similar cases for three types of liver lesions by using global features and local features based on image density and texture. In [2], two attributes of margin sharpness extracted from sigmoid curves, including the intensity difference between a lesion and its surroundings, and the sharpness of the intensity transition across the lesion boundary are used for distinguishing different types of lesions. The method was evaluated in both simulations and CT scans of liver and lung lesions and competitive results were achieved. To make use of contrast enhancement features of CT scans, various methods were proposed on multiphase medical images. Quatrehomme et al. proposed to perform classification using visual features and law measures on multiphase CT images to classify five types of liver lesions. Roy et al. [3] used 4 kinds of features, including density, temporal density, texture, and temporal texture, which are derived from four-phase medical images to retrieve the most similar images in five types of liver lesions. The texture features are six coefficients computed from 3D gray-level cooccurrence matrix (GLCM). What is more, the tumors are partitioned into three subvolumes to capture spatial information and the algorithm is accelerated by performing parallel computing on the three subvolumes. Although the related work manifests the potential of retrieving the similar cases of a query FLL input, most of them only exploited low-level features such as intensity, texture, and simple shape descriptors, and thus there still is large space to improve the retrieval performance for real applications. The Bag-of-Visual-Words (BoVW) method has been successfully applied to many natural image classification and retrieval tasks [4–8]. Derived from a few representative triple-phase slices, the Bag-of-Visual-Words (BoVW) with SIFT local features, intensity, texture, and shape-based features is used

to distinguish three lesion types [9]. Yang et al. [10] has also used BoVW for retrieval of focal liver lesions in three types. Recently, Diamant et al. [11] presented an improved automated liver lesion classification method using single-phase (PV) medical images based on K-means clustering technique. All the above-mentioned methods exploited low-level feature or BoVW-based middle-level features for retrieval or classification of different types of FLLs and have proven the potential capability to FLL characterization and clinical applications. Therein, the middle-level feature, BoVW, presented the promising retrieval performance of FLLs [9–11] compared to low-level features. However the widely used BoVW-based model for medical image representation basically (1) applies the handcrafted SIFT as local features, which would lead to some tiny structure loss and is unsuitable for the fine-grained medical images; (2) uses K-means for codebook learning and coded vector calculation, which approximates any local descriptor using one learned visual word only and leads to large reconstruction error; (3) extracts middle-level feature using one-phase data only, which may be inapplicable for some types of FLL cases. Therefore, there still is large space for improving the performance for FLL characterization and representation.

In this study, we explore an improved sparse codebook-based feature representation for multiphase CT volume images and apply the extracted middle-level features for retrieving FLLs to assist diagnosis decision-making of different types of FLLs. The proposed FLL retrieval system is shown in Figure 1, where we concentrate on the discriminated feature extraction from multiphase CT volumes using the improved codebook model. In the proposed codebook model, instead of using SIFT as the local descriptors, which are calculated as the histograms of the quantized orientation and can effectively represent the distinguished part of a specific object in general images, we directly exploit the local patch as the local descriptor, which can retain all detailed information and is more suitable for the fine-grained medical image application. Furthermore, in order to more accurately approximate any local descriptor in coding procedure, the sparse coding method, instead of K-means algorithm, is employed to overcomplete codebook learning. Experiments on five types of FLL retrieval show that the proposed feature extraction method can achieve promising retrieval performance compared with state-of-the-art methods.

The rest of this paper is organized as follows. Section 2 introduces the proposed codebook model for multiphase medical image representation and the used coding strategy: sparse coding for more accurate approximation of any local descriptor. Section 3 talks about the dataset, experiments settings, and experimental results with different parameters. Finally, we conclude the paper in Section 4.

2. The Sparse Codebook Model of Local Structures for Medical Image Representation

Typically, the codebook model, BoVW representation, involves four major steps: (1) key-point detection for interested image regions; (2) local descriptor extraction for representing the interested image regions around key points; (3) coding via quantizing the local descriptors in terms of a predefined dictionary (also called codebook), which generally learned from a large amount of prepared local descriptors with K-means algorithm; (4) pooling operation by accumulating the coded feature vectors into a fixed-length representation feature of an image. The local descriptors in the conventional BoVW model generally employed SIFT, which is calculated as the histogram of the quantized orientation and can effectively represent the distinguished part of a specific object in general images. However, this study concentrates on extraction of middle-level features for the fine-grained medical FLL data, and the rough quantization of orientation in SIFT would lead to a lot of detailed information loss, which may not orient the effective representation of the interested image regions for a medical FLL image. Furthermore, the conventional BoVW model used K-means algorithm to learn the codebook (set of visual words) and approximated any local descriptor with only one visual word, which would lead to large reconstruction error. Therefore, this study proposes a sparse codebook model oriented for fine-grained FLL image representation, where the local patch of the interested image region is directly used as the local descriptor for retaining all detailed texture and the sparse coding, instead of K-means algorithm, is employed for codebook learning and coding any local descriptors. Furthermore, we also combine the codebook-based feature representations of multiphase CT volumes. The proposed sparse codebook model with local structures is shown in Figure 2.

2.1. Local Descriptors for Representing the Interested Image Regions. Recent research [12, 13] in the computer vision field showed that it is possible to discriminate between textures using pixel neighborhoods such as a small patch, an $l \times l$ pixel region. Awate et al. [14] explored nonparametric neighborhood statistics and manifested promising performance for texture segmentation. Pietikäinen et al. [13, 15] showed that despite the global structure of the textures, very good discrimination can be achieved by exploiting the distributions of such pixel neighborhoods. Therefore, the exploitation of these microstructures for representing images in the distributions of local descriptors has gained much attention and has led to state-of-the-art performances [13, 15,
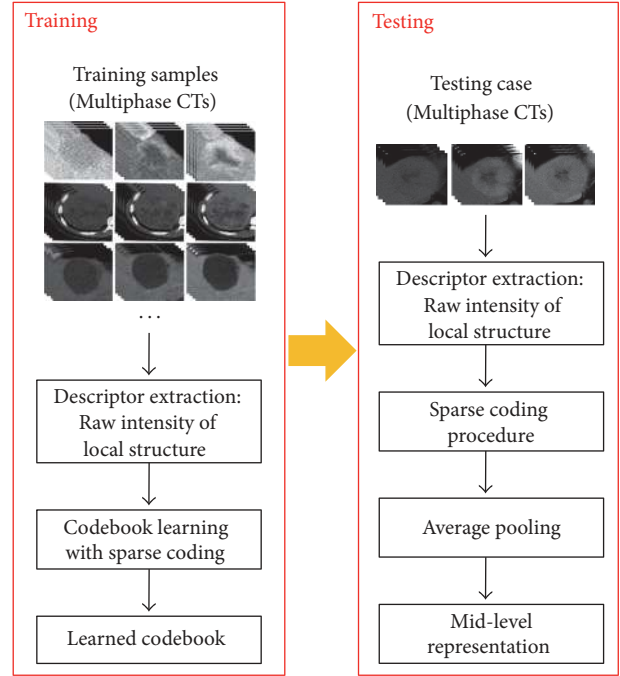

FIGURE 2: Proposed sparse codebook model with local structures.

16] for different classification and segmentation problems in computer vision. As we know different types of FLLs mainly manifest different intensity variance, which means that there are different textures for different types of FLLs. Thus, this study explores the local patches (texture structures) as local descriptors in the codebook model.

Given an image I and the patch region with size $l \times l$ centered at the ith pixel, we directly use intensities, I_i^j, $j = 1, 2, \ldots, l \times l$, of all pixels in the patch regions as a local descriptor $\mathbf{y}_i = [I_i^1, I_i^2, \ldots, I_i^{l \times l}]$. The representation of a pixel directly uses the neighboring pixels' intensities, which not only considers the intensity but also retains the variation in intensity (texture) without any detained structure loss and thus would adapt to the fine-grained medial FLL retrieval application. In our study, we take the $l \times l$ local patches of all pixels in a medical image as the local descriptor set for coding.

2.2. Codebook Learning Algorithm: From K-Means to Sparse Coding. Given a set of prepared local descriptors, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]$, the codebook model firstly learns a small set of visual words (prototype features) for coding any local descriptor. A common strategy for codebook learning in the BoVW model usually applies the K-means algorithm. A family of signals $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N$ can be represented by the nearest neighbor in a codebook $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K]$, $N \gg K$, in which a codeword $\mathbf{d}_i$ is a column vector. The codebook $\mathbf{D}$ is learned in the K-means algorithm by solving the least-square problem as follows:

$$\arg\min_{\mathbf{D}, \mathbf{x}_i} \sum_{i=1}^N \left\| \mathbf{y}_i - \mathbf{D}\mathbf{x}_i \right\|^2,$$

$$\text{s.t.} \quad \left\| \mathbf{x}_i \right\|_0 = 1,$$

$$\|\mathbf{x}_i\|_1 = 1,$$

$$\mathbf{x}_i \geq 0, \quad \forall i, \tag{1}$$

where $\|\cdot\|_0$ and $\|\cdot\|_1$ refer to L_0-norm and L_1-norm, separately. $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$ are the coded vectors for $\mathbf{Y}$. Single nonzero entry in each coded vector $\mathbf{x}_i$ is ensured by the constraint $\|\mathbf{x}_i\|_0 = 1$, $\forall i$, and the coding weight is always 1, formulated as the second constraint term: $\|\mathbf{x}_i\|_1 = 1$, $\forall i$. The K-means is widely used in codebook training because of its simplicity. However, it is too restricted to approximate signals properly by allowing only one codeword from the codebook.

As a generalization of the K-means algorithm, the sparse coding technique employs a linear combination of codewords for the representation of each signal, which means more than one nonzero entries in coding, and the weights can be calculated to be arbitrary values but not limited to 1. The intuitive way for the sparse coding problem can be formulated to optimize the following objective function:

$$\arg\min_{\mathbf{D}, \mathbf{X}} \quad \|\mathbf{Y} - \mathbf{DX}\|_2^2,$$

$$\text{s.t.} \quad \|\mathbf{X}\|_0 \leq \alpha K, \tag{2}$$

where $\mathbf{X}$ is the sparse approximation of $\mathbf{Y}$ on codebook $\mathbf{D}$. α is a sparsity measure, which is a ratio between the number of nonzero entries in $\mathbf{x}_i$ and the total number of codewords in $\mathbf{D}$. $\alpha * K$ controls the maximum number of codewords that can be used for approximation of the input signal $\mathbf{y}_i$.

Similar to realization of K-means algorithm, there are two stages: sparse coding stage and codebook update stages, in the codebook learning of sparse coding, as shown in Algorithm 1. In consideration of simplicity and efficiency, we employ the Orthogonal Matching Pursuit (OMP) algorithm for coefficient calculation and K-SVD method for codebook updating in the two stages, respectively. The detailed implementation of these two methods is described in the following section. Furthermore, we also implement the alternating direction method of multipliers (ADMM) for solving our sparse coding problem, which is an algorithm to solve convex optimization problems by breaking them into smaller pieces for easier handling. Please refer to the existing work [17] for details.

2.3. The Implementation Algorithms: OMP and K-SVD in Sparse Coding

2.3.1. OMP for Coefficient Calculation. With the fixed codebook, OMP algorithm is a simple and efficient way to solve the sparse approximation problem, which is NP-hard because of the overcomplete codebook $\mathbf{D}$. OMP is an iterative greedy algorithm, in which at each step a column codebook vector is selected, which is most correlated with the current residuals, and then the selected column codebook vector is add to the set of selected visual word set being used for calculating the coded coefficients, which can be effectively solved with the least-square problem due to the smaller number of the selected visual words than the dimension of the input signal $\mathbf{x}_i$. The algorithm updates the residuals by projecting the input

signal onto the linear subspace spanned by the visual word set that has already been selected and the algorithm then iterates until a predefined stop criterion is satisfied.

2.3.2. K-SVD for Codebook Updating. K-SVD was proposed for generating a dictionary of spare representation, via singular value decomposition (SVD). It is a generalization of the K-means clustering method. K-SVD works by iteratively alternating between coefficient calculation of the input data based on the current dictionary and updating the atoms in the dictionary to better fit the data. As introduced above the coefficient calculation is implemented via OMP algorithm; this part mainly describes the atom updating procedure, where each of the codewords will be updated once a time, assuming that all the other codewords and the coded vector $\mathbf{x}_i$ are fixed. The update of the kth atom is done by rewriting the penalty term as

$$\|\mathbf{Y} - \mathbf{DX}\|^2 = \left\| \left(\mathbf{Y} - \sum_{j \neq k}^{K} \mathbf{d}_j \mathbf{x}_j^T \right) - \mathbf{d}_k \mathbf{x}_k^T \right\|^2$$

$$= \left\| \mathbf{E}_k - \mathbf{d}_k \mathbf{x}_k^T \right\|^2, \tag{3}$$

where K is the total number of codewords and the row vector $\mathbf{x}_k^T$ contains the coefficients of all input signals on codeword $\mathbf{d}_k$. $\mathbf{E}_k$ represents the reconstruction residuals using all codewords excepting $\mathbf{d}_k$. After adding $\mathbf{d}_k$ for approximating the input signals, we expect that the reconstruction error in (3) is minimized with the fixed $\mathbf{E}_k$, which is formulated as the following formula:

$$\arg\min_{\mathbf{d}_k, \mathbf{x}_k^T} \quad \left\| \mathbf{E}_k - \mathbf{d}_k \mathbf{x}_k^T \right\|^2 \tag{4}$$

which can be easily solved by applying SVD. However, $\mathbf{x}_k^T$ can be filled and the sparsity will be destroyed if simply SVD is applied on $\mathbf{E}_k$. To address this problem, a constraint function $\boldsymbol{\omega}_k$ is defined to remove the zero entries of $\mathbf{x}_k^T$ and the corresponding elements in $\mathbf{E}_k$, achieving $\mathbf{x}_k^R$ and $\mathbf{E}_k^R$, respectively. Applying SVD on $\mathbf{E}_k^R$, $\mathbf{E}_k^R = \mathbf{U}\Delta\mathbf{V}^T$, the codeword $\mathbf{d}_k$ is updated by the first eigenvector of $\mathbf{U}$, and the coefficient vector $\mathbf{x}_k^T$ is updated by zero padding of the multiplication of the first column of $\mathbf{V}$ with $\Delta(1, 1)$.

The above process, SVD on the reconstruction residuals matrix, is applied K times to update the codewords each at a time. The details of learning codebook by OMP and K-SVD are shown in Algorithm 1.

2.4. Middle-Level Feature Extraction for Medical Images. With the prepared training local descriptors $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]$, the codebook $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K]$ can be learned using the conventional K-means or sparse coding strategy. In order to code the local descriptors $\mathbf{y}_m \in \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M\}$ from any test image and extract the middle-level feature, we fix the

(1) **Task**: Find the optimal codebook that represent the data
samples $\{\mathbf{y}_i\}_{i=1}^N$ as sparse compositions. Objective function is:
$$\min_{\mathbf{D},\mathbf{X}}\{\|\mathbf{Y}-\mathbf{DX}\|_F^2\}, \quad \text{s.t.}, \|\mathbf{x}_i\|_0 \le \alpha K, \ \forall i$$
(2) **Init**: Set dictionary matrix $\mathbf{D}^{(0)}$ with l^0 normalized columns
(3) **while** (stopping rule not met) **do**
(4)   *Sparse Coding Stage*: Compute $\mathbf{x}_i$ to approximate example
    $\mathbf{y}_i$ using OMP, by solving:
$$i = 1, 2, \ldots, N \quad \min_{\mathbf{x}_i}\{\|\mathbf{y}_i-\mathbf{Dx}_i\|_2^2\}, \quad \text{s.t.}, \|\mathbf{x}_i\|_0 \le \alpha K$$
(5)   *Codebook Update Stage*: update each dictionary atom $\mathbf{d}_k$,
    $k = 1, 2, \ldots, K$
(6)   - -Define the group of indices corresponding to atom $\mathbf{d}_k$,
$$\boldsymbol{\omega}_k = \{i \mid 1 \le i \le N, \ \mathbf{x}_k^T(i) \ne 0\}$$
(7)   - -Compute the overall representation error:
$$\mathbf{E}_k = \mathbf{Y} - \sum_{j \ne k}\mathbf{d}_j\mathbf{x}_k^T$$
(8)   - -Obtain $\mathbf{E}_k^R$ by restricting $\mathbf{E}_k$ using $\boldsymbol{\omega}_k$
(9)   - -Apply SVD: $\mathbf{E}_k^R = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T$. Atom $\mathbf{d}_k$ is updated by the first
      column of $\mathbf{U}$ and coefficient vector $\mathbf{x}_k^T$ is updated by the
      first column of $\mathbf{V}$ multiplied by $\boldsymbol{\Delta}(1, 1)$
(10)   - -Padding $\mathbf{d}_k$ and $\mathbf{x}_k^T$ to their original size by 0 according
      to $\boldsymbol{\omega}_k$
(11) **end while**

ALGORITHM 1: Codebook learning via OMP and K-SVD.

codebook $\mathbf{D}$ and obtain the coded vector $\mathbf{x}_m$ by minimizing the following cost function:

$$\underset{\mathbf{x}_m}{\arg\min} \quad \|\mathbf{y}_m - \mathbf{Dx}_m\|_2^2, \tag{5}$$
$$\text{s.t.} \quad \|\mathbf{x}_m\|_0 \le \alpha K.$$

This optimization problem is also implemented by the OMP method, which is called coding strategy and then M coded vectors with K-dimension for each image can be obtained.

Finally, the M coded vectors are aggregated to form a fixed-length feature for image representation by average operator, which is also called pooling procedure. In our application, we apply multiphase CT volumes for FLL retrieval, and thus we extract the codebook-based feature for each phase CT image and directly concatenate all features for multiphase CT representation.

3. Experiments and Results

3.1. Dataset. The proposed codebook model is used to learn features for retrieval of focal liver lesions (FLLs) from the dataset which we have constructed with the help of radiologists. The dataset consists of 5 types of liver lesions, and there is 137 medical cases in total in the dataset, including 38 lesions in cyst class, 28 cases for both hepatocellular carcinoma (HCC) and hemangioma (HEM), 22 cases in focal nodular hyperplasia (FNH), and 21 liver lesions for metastasis (METS). Due to the fact that most of the medical cases supplied by radiologists contain only three phases (NC, ART, and PV), while only small amount of them have the DL phase,
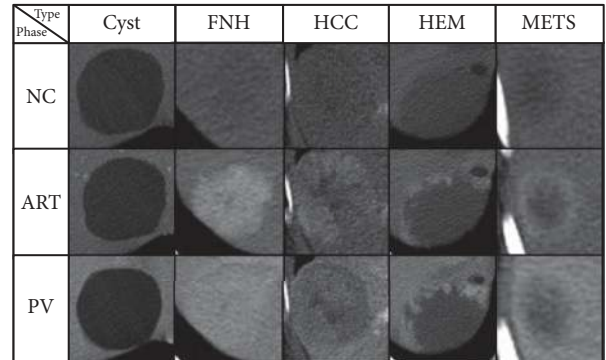

FIGURE 3: Examples of each lesion type on 3 phases. Rows are images that belong to same contrast phase, while columns are images from the same lesion: cyst, FNH, HCC, HEM, and METS.

we choose to use only 3 phases for each case for consistency concern. Examples of our dataset are shown in Figure 3.

3.2. Intensity Normalization. We found that the intensity distributions of the same patient are different for the three phases, as illustrated in Figure 4. This is caused by the contrast enhancement injection. Though the lesion regions are obviously enhanced, it is a fact that the other tissues, such as health liver parenchyma, are also enhanced at the same time. This breaks down the intensity consistency when combining multiphase data together. Considering this situation, a simple preprocessing technique is employed: the intensities are normalized by the average intensity of the surrounding healthy liver parenchyma. The preprocessing method can

NC phase   ART phase   PV phase
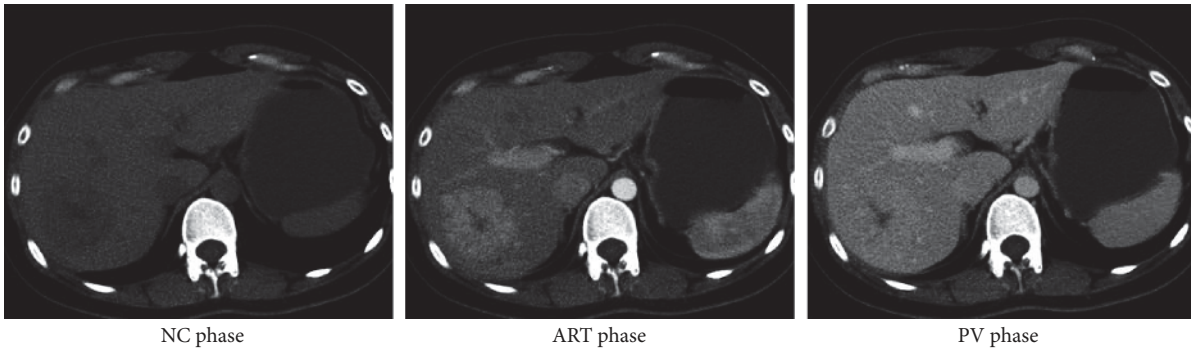
FIGURE 4: Illustration of intensity inconsistency among different phases.

not only avoid the inconsistency of intensity distribution among different phases, but also address the problems caused by images captured from different equipment and other intensity-shift problems. The following experiments are all based on the intensity normalized images.

3.3. Experiments Settings (Image Representations Based on BoVW). The image representation based on Bag-of-Visual-Words technique regards an image as a distribution of local descriptors, which can be viewed as discrete visual proto-types. Given a codebook, the occurrence of the descriptors from an image on the codewords is used to represent the image, in a histogram-like style. Since the medical images are commonly taken under standard situations, no meaningful key points or structures appear in a liver lesion in CT images, and the intensity plays an important role in diagnosis. In this work, we use raw patches, which are densely sampled from every pixel in the lesion region, as the local descriptors of a medical image.

Then it is important to learn an effective codebook from the training samples. The simple and efficient algorithm, K-means, is widely used in this task for natural images. The proposed codebook learning method based on OMP and K-SVD algorithms [18, 19] is used in this study instead of K-means.

We first validated the representation accuracy of sparse coding comparing to K-means method. In this evaluation, we used 5000 local patches from training ART images and approximated them with K-means and sparse coding meth-ods with codebook size 100 for both. Then the reconstruction errors (RE) can be calculated for all the selected patches, and the distributions of the RE of the samples using both methods are plotted in Figure 6. As illustrated in Figure 6, sparse coding method achieves smaller reconstruction errors than using K-means method, which means more accurate approx-imations, and thus more accurate retrieval performance can be expected.

Considering the multiphase medical images, codebooks for each phase are learned separately. Examples of the learned codebook model for ART phase CT scans are shown in Figure 5 and the contrast is enhanced for easy view. With the codebook model for single phase, the retrieval performance will be discussed in the Results section to emphasize the usefulness of contrast enhancement for detection of FLLs.
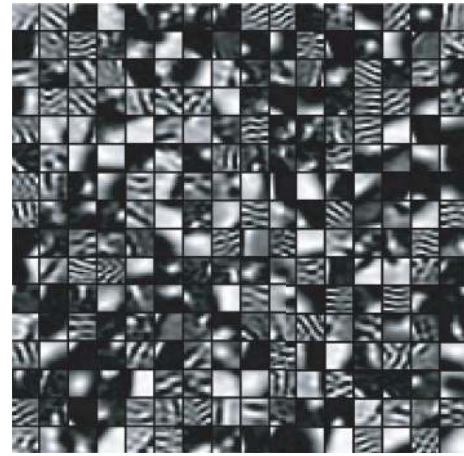

FIGURE 5: The codebook learned for CT scans on ART phase. The patch size is 11×11, and codebook size is 256.
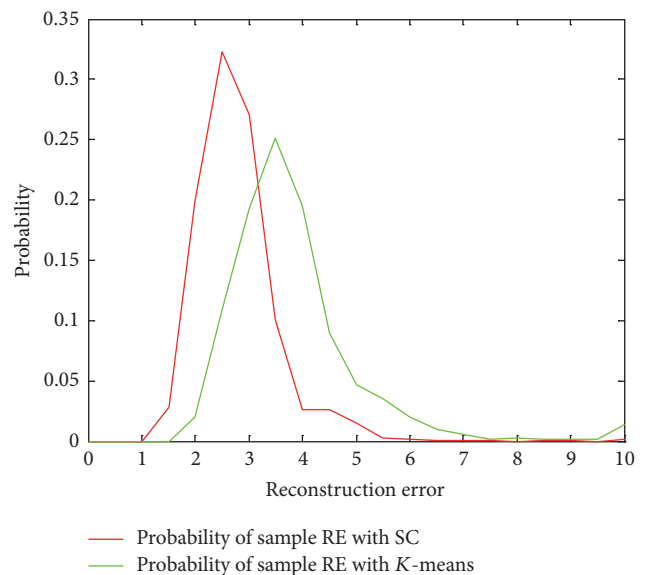

FIGURE 6: Distribution of sample RE using the proposed method and K-means method.

Table 1: Construction of confusion matrix.

	Actual positive	Actual negative
Predicted positive	TP	FP
Predicted negative	FN	TN

Precision = TP/(TP + FP).
Recall = TP/(TP + FN).

Due to large amount of descriptors which will cost too much time for codebook learning, we randomly select 50000 descriptors from training samples for codebook learning. Since the codebook learning is an unsupervised learning procedure, the selected descriptor number usually does not affect the final performance greatly. Then the coefficients of patches are calculated according to the learned codebooks responding to the image's phase in single-phase condition. The coefficients of patches that belong to the same image are pooled together to form a feature vector to represent the image. In our experiments the mean-pooling technique is used due to its simplicity and efficiency. The feature vectors of images in each phase are concatenated to the final representation of a patient case to engage the multiphase features.

3.4. Performance Evaluation. To evaluate the performance of the proposed codebook model for retrieval of FLLs, confusion matrices, Table 1, are built and precision and recall values are calculated correspondingly. Precision versus recall curves [20] are used to represent the precision and recall values, which summarize the pairs of precision and recall pairs when varying the number of retrieved top similar samples.

The leave-one-out cross validation method is used in evaluation. Confusion matrices, as shown in Table 1, are constructed when t $(1 \leqslant t \leqslant T - 1)$ most similar samples are retrieved, where T is the total number of medical cases in the dataset. Then precision and recall are calculated correspondingly. A precision versus recall curve is obtained by averaging precision and recall values of all validated samples at position t $(1 \leqslant t \leqslant T - 1)$.

3.5. Results. Experiments are applied on single-/multiphase medical CT scans. The retrieval performances using single-phase data and multiphase data are shown in Figure 7(a). As shown before in Figure 3, the contrast enhanced phases (ART and PV) show more distinguishable characteristics than the non-contrast enhanced phase (NC). Thus it is reasonable that ART and PV phase have better retrieval performance than NC, while the two contrast enhanced phases have similar performance, as shown in Figure 7(a). The proposed sparse codebook model using multiphase data provides the best performance with the help of temporal information in retrieval of liver lesions. There are two ways to learn codebooks for representation of multiphase data: joint learning and separate learning. Joint learning means that the corresponding local descriptors of data in each phase are combined before learning codebooks. Only one codebook is learned for multiphase data in joint learning. On the other hand, separate learning learns codebooks for each phase and the features of CTs in each phase are concatenated for the final representation of the multiphase data. The compared performance is also given in Figure 7(a). Considering the situation that, in medical practice, the multiphase data are only roughly aligned, and small translations and/or rotations in soft tissues, such as liver, unavoidably exist, which can be caused by factors such as the patient's breath during CT scanning. Because the multiphase data are not exactly registered, the joint pixels from different phases may represent different tissues and thus lead to mismatching combination in the joint learning. As shown in Figure 7(a), the joint learning has comparable performance for a few retrieved cases with separate learning, while the performance reduces fast when more similar cases are retrieved. Since the separate learning outperforms the joint learning, separate learning is used in the rest of the experiments for multiphase data.

Furthermore, the retrieved performance is quiet bad in using the proposed sparse codebook model with the NC phase only. In order to investigate the applicability of the proposed sparse codebook model, we also conducted experiments with state-of-the-art methods on the NC phase only, and the compared retrieval results are provided in Figure 7(b), which illustrated that the proposed patch-based sparse coding (SC) method gets the best performance while all the methods have very poor performance on NC phase data only.

We compared our method to other alternative methods. We implemented the intensity-based method, which means that we used the mean intensity of the focal liver lesion region to represent the image. Then each medical case is featured by a vector that has 3 elements, one from each phase (NC, ART, and PV). We also realized the codebook learning by K-means clustering technique, which refers to K-means-based method, with one hundred clustering centroids. In this realization, the training samples are raw patches, which are the same as the training samples in our proposed method. The retrieval performance comparison is shown in Figure 8: the intensity-based method has comparable results which emphasizes the importance of intensity in focal liver lesion recognition; the proposed codebook model has the best performance while the K-means-based method only performs slightly better than the intensity-based method.

We have also used the Local Binary Pattern (LBP) to capture the local texture features of focal liver lesions. Each image is represented by a histogram with 256 bins from calculating the LBP patterns. And the medical case is featured by a 768-dimensional vector, which is a concatenation of the three 256-bin histograms for each phase of medical image.

We also used the retrieval accuracy at top t retrieved samples to evaluate the performance of different methods. Table 2 shows the retrieval accuracy of the proposed method; comparing with that of three other methods, when t is at the top $(t = 1, 3, 5, 10, 20)$, most similar medical cases are retrieved. The K-means-based method gets best performance when only one case is retrieved and the proposed sparse codebook model can achieve best performance when more cases are retrieved. The retrieval precisions are very low by using the LBP-based methods, due to the fact that the local
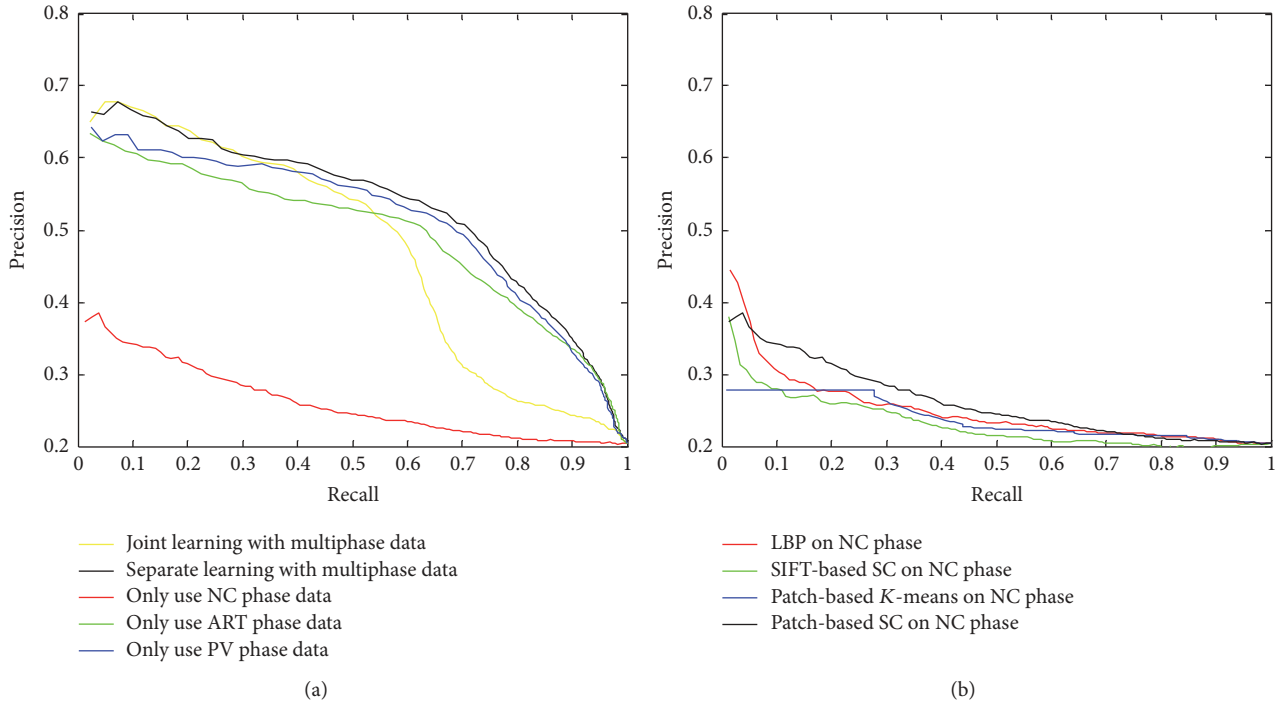
(a)

(b)

FIGURE 7: (a) Retrieval performance using single-/multiphase data. (b) Retrieval performance of different methods by using only NC phase data.

TABLE 2: Retrieval accuracy of different methods at various most similar numbers of cases being retrieved.

Methods	Top 1	Top 3	Top 5	Top 10	Top 20
Proposed method	0.66	0.68	0.65	0.62	0.59
Mean intensity method	0.61	0.6	0.6	0.6	0.56
K-means based on local patch	0.72	0.65	0.64	0.61	0.56
LBP method	0.34	0.31	0.32	0.28	0.25

texture feature is not discriminative for different types of focal liver lesions.

In addition, we also conducted FLL retrieval experiment with the proposed sparse codebook model using Scale-Invariant Feature Transform (SIFT) as local descriptors, and the result is manifested in Figure 8. Figure 8 shows that the performance of SIFT-based method is very poor and is not acceptable, which explains that retaining all detail structure with local patch is more important than the quantized-based SIFT for the fine-grained medical data representation.

3.6. Parameter Optimization. Previous studies suggested that the codebook size and the patch size have significant influence on the performance of image retrieval and categorization for natural images. We evaluated the performance for retrieval of focal liver lesions when varying the patch size, codebook size, and sparsity measure. In previous section, we have proved that retrieval based on multiphase medical data
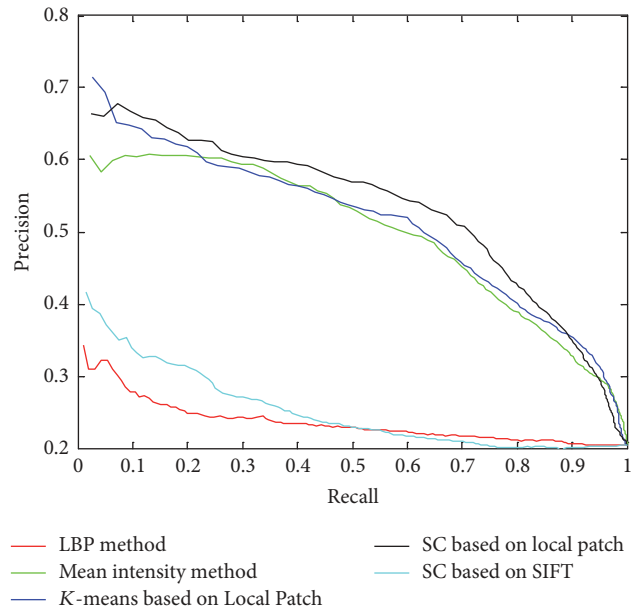

FIGURE 8: Comparison of the proposed codebook model with state-of-the-art methods.

will have the best performance. The following experiments were performed on multiphase CT scans.

3.6.1. Patch Size. To evaluate the influence of patch size on the performance of FLLs retrieval, we fixed the other parameters based on our experience and preliminary results (codebook
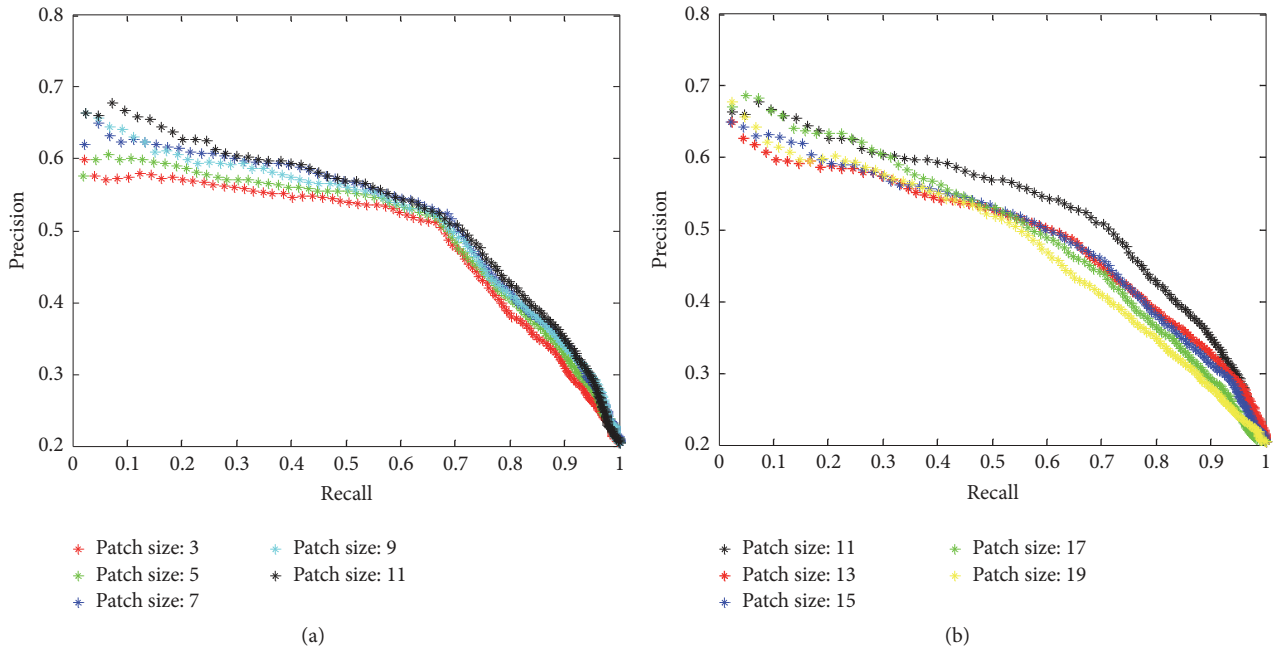
Patch size: 3   Patch size: 9
Patch size: 5   Patch size: 11
Patch size: 7

(a)

Patch size: 11   Patch size: 17
Patch size: 13   Patch size: 19
Patch size: 15

(b)

FIGURE 9: Retrieval performance when varying patch size from 3 to 19.

Codebook size: 20   Codebook size: 80
Codebook size: 40   Codebook size: 100
Codebook size: 60

(a)

Codebook size: 100   Codebook size: 400
Codebook size: 200   Codebook size: 500
Codebook size: 300

(b)
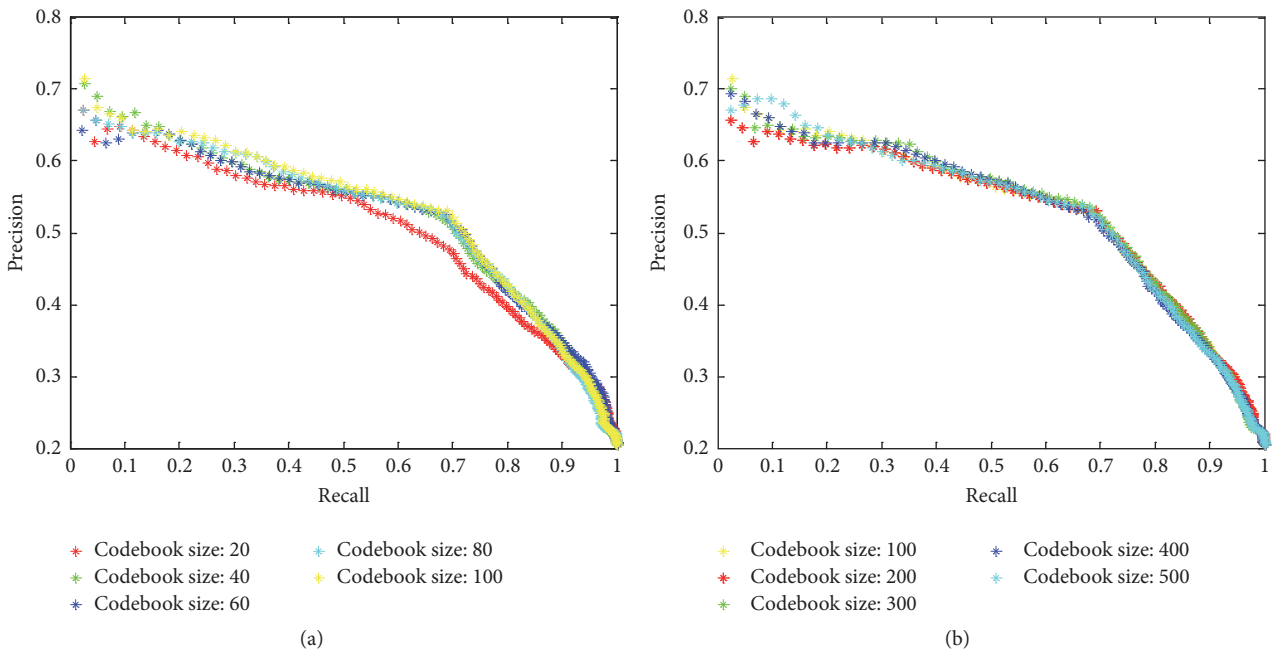
FIGURE 10: Retrieval performance when varying codebook size.

size is set to 100 and sparsity measure is set to 0.1). We have varied the patch size from $3 * 3$ to $19 * 19$ with step size 2, and the results are shown in Figure 9. The retrieval performance will increase as patch size enlarges, Figure 9(a), when using small patch size. This might be explained by the fact that large patches can capture more texture information and thus represent an image more effectively. The performance approaches its optimal at patch size $11 * 11$. When increasing the patch size even further, the performance gets lower,

since too large patches will capture tissues from surrounding region, which are regarded as noise and influence the retrieval accuracy. From this evaluation, the patch size will be set to $11 * 11$ in future experiments to achieve the best performance.

3.6.2. Codebook Size. The codebook size is varied in range from 20 to 100 every 20, when the patch size and sparsity measure are fixed to $11 * 11$ and 0.1, separately. We can see from Figure 10(a) that there is small improvement in the retrieval
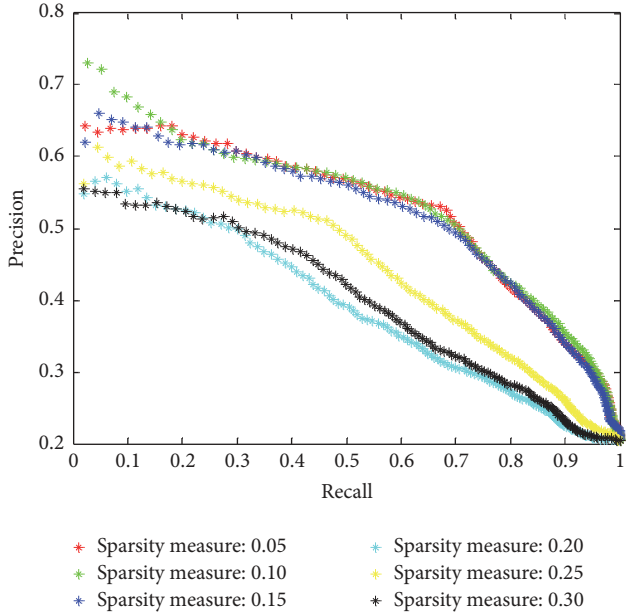
FIGURE 11: The retrieval performance with various sparsity measures, when patch size is $11 * 11$ and codebook size is 32.
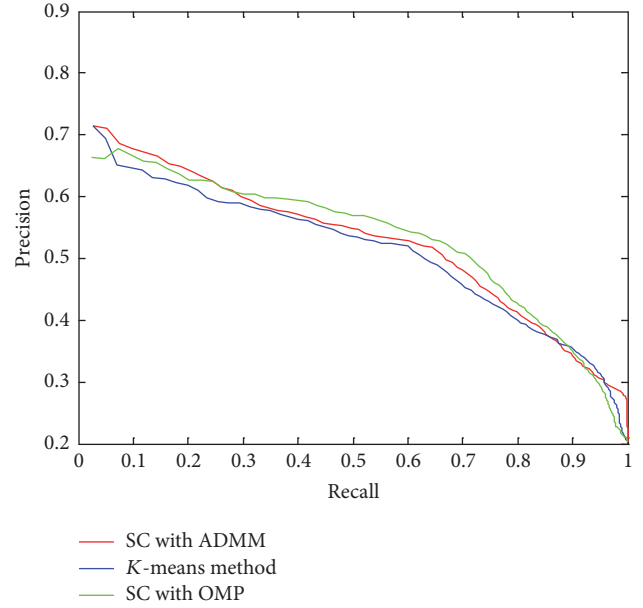

FIGURE 12: Retrieval performance comparison of ADMM implementation method with the proposed method and K-means method.

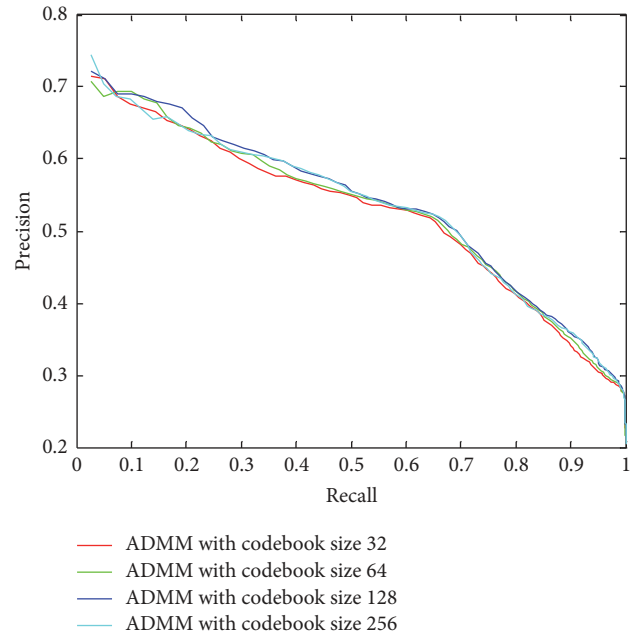FIGURE 13: Retrieval performance of ADMM implementation method with different codebook sizes.

performance when increasing the codebook size. Similar situation occurs in the performance of ADMM implementation method: no obvious improvement with larger codebook size. We have also explored even larger codebook size, from 100 to 500 every 100, while no better results can be achieved, shown in Figure 10(b). Considering getting a balance between retrieval performance and computation time, the codebook size used in other experiments is determined to be 32.

3.6.3. Sparsity Measure. The sparsity measure, α, in formula (2), in this work refers to the ratio between the number of nonzero entries of a sparse approximation and the total number of codewords in the learned codebook. $\alpha * K$ controls the maximum number of codewords that can be used for approximation of the input signal; K is the size of codebook. We evaluate retrieval performance with different sparsity measure α, which varies in $\{0.05, 0.1, 0.15, 0.2, 0.25, 0.3\}$, when patch size is fixed to $11 * 11$ and codebook size is fixed to 100 according to our previous experiments. The experimental results are shown in Figure 11.

3.7. Evaluation of ADMM Implementation Method in Sparse Coding. All the above experiments of our proposed sparse codebook model were conducted with the implementation methods: OMP for coefficient calculation and K-SVD for codebook updating. We also implemented the sparse codebook model using ADMM for solving sparse coding optimization problem. The retrieval performance with K-means, and our proposed sparse codebook model with K-SVD/OMP and ADMM solving methods using codebook size 32 only, respectively, are provided in Figure 12. As shown in Figure 12, the sparse coding method based on either ADMM or K-SVD/OMP optimization methods outperforms the K-means method. The sparse coding with ADMM implementation is

better than that with K-SVD/OMP when only several similar cases to the query case are retrieved, while the SC with K-SVD/OMP performs better when more similar cases are retrieved.

We have also evaluated the performance of the ADMM method with different codebook sizes, as shown in Figure 13. The retrieval performance improvement is not obvious when increasing the codebook size. Considering the time efficiency,

(a) Cyst

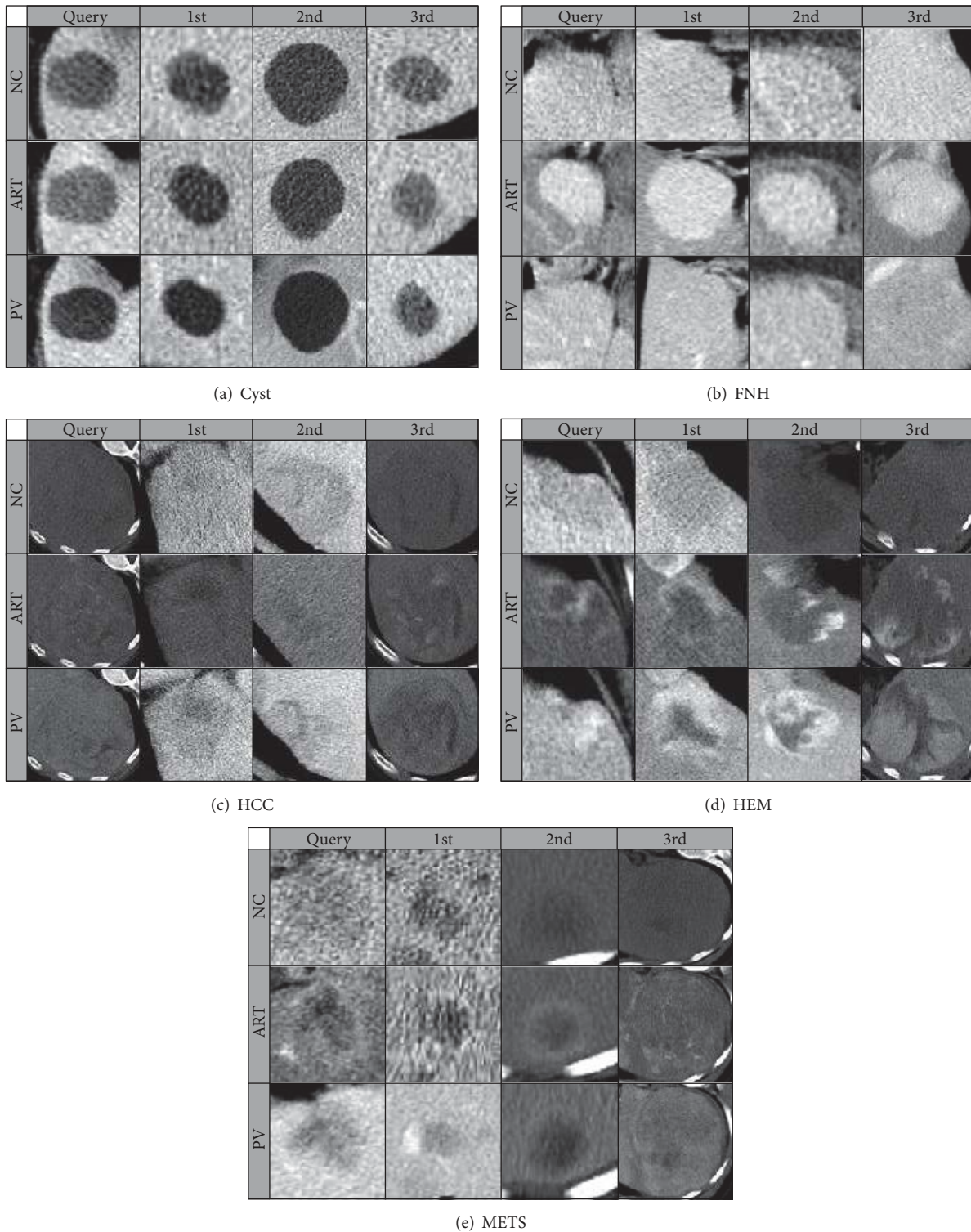(b) FNH

(c) HCC

(d) HEM

(e) METS

FIGURE 14: Illustration of retrieval results (top three retrieved examples) for one example from each class.

the codebook size used in performance comparison with other methods, Figure 12, is determined to be 32.

3.8. Illustration of Top Retrieved Results. The top three retrieved samples for one example from each focal liver lesion class are shown in Figure 14. The retrieval is performed

using the proposed codebook model with parameters that are all optimal, which have been proved in previous section. The three rows of images show the representative slices of the corresponding focal liver lesion on different phases. The first columns are the query samples while the following three columns are the top three retrieved samples using the

leave-one-out cross validation. We can see that the retrieved samples have very similar appearance with query samples. What is more, similar enhancement characters are shown between the query and retrieved samples. However, there are some cases where the top retrieval sample does not belong to the same class as the query sample, for example, the 3rd retrieved sample belongs to HCC for the query example from METS class.

4. Conclusion

In this work we present a codebook model for retrieval of focal liver lesions using three-phase contrast enhanced CT images. The proposed method is evaluated on a dataset with 137 medical cases comprising five types of focal liver lesions. The leave-one-out cross validation is applied in construction of confusion matrices for retrieval of each medical case. Codebook is learned for medical cases in each phase separately. And thus the middle-level features based on the proposed sparse codebook model can be extracted from each phase data and the concatenated vector of all phase features is used as the representation of the multiphase medical data. Experimental results show that retrieval based on multiphase data achieved best performance while using only one contrast enhanced CT image can also get reasonable results. In future works, we will extend the linear analysis to multilinear analysis of the medical data to capture core information. Since the vectorization of local descriptors is unavoidable in linear analysis, the spatial structure is always destroyed. We want to apply the tensor technique which analyzes the input signals without unfolding and thus keep the structure for better representation of high-dimensional medical data. With improved accuracy in future works, we intend to develop a content-based image retrieval system to assist in medical education and clinical practice.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.
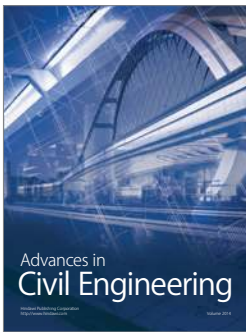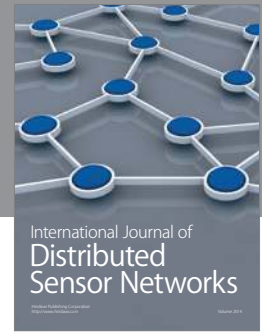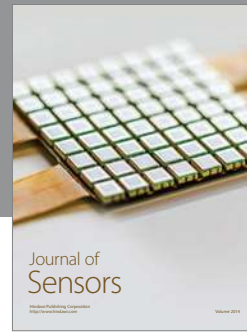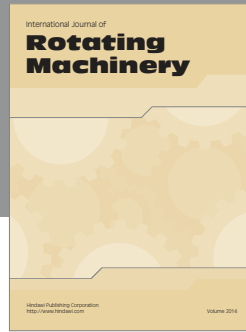
Acknowledgments

References

[1] M. Yu, Z. Lu, Q. Feng, and W. Chen, "Liver CT image retrieval based on non-tensor product wavelet," in *Proceedings of the International Conference of Medical Image Analysis and Clinical Application (MIACA '10)*, pp. 67–70, IEEE, Guangzhou, China, June 2010.

[2] J. Xu, S. Napel, H. Greenspan, C. F. Beaulieu, N. Agrawal, and D. Rubin, "Quantifying the margin sharpness of lesions on radiological images for content-based image retrieval," *Medical Physics*, vol. 39, no. 9, pp. 5405–5418, 2012.

[3] S. Roy, Y. Chi, J. Liu, S. K. Venkatesh, and M. S. Brown, "Three-dimensional spatiotemporal features for fast content-based retrieval of focal liver lesions," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 11, pp. 2768–2778, 2014.

[4] J. Winn, A. Criminisi, and T. Minka, "Object categorization by learned universal visual dictionary," in *Proceedings of the 10th IEEE International Conference on Computer Vision0 (ICCV '05)*, vol. 2, Beijing, China, October 2005.

[5] H. Jégou, M. Douze, and C. Schmid, "Improving bag-of-features for large scale image search," *International Journal of Computer Vision*, vol. 87, no. 3, pp. 316–336, 2010.

[6] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, "Evaluating bag-of-visual-words representations in scene classification," in *Proceedings of the International Multimedia Conference (MM'07)—9th ACM SIG Multimedia International Workshop on Multimedia Information Retrieval (MIR'07)*, pp. 197–206, Augsburg, Germany, September 2007.

[7] F.-F. Li and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 524–531, Madison, Wis, USA, June 2005.

[8] G. Csurka, C. Dance, C. Bray, and L. Fan, "Visual categorization with bags of key points," in *Proceedings of the Workshop on Statistical Learning in Computer Vision (ECCV '04)*, Prague, Czech Republic, May 2004.

[9] M. Yu, Q. Feng, W. Yang, Y. Gao, and W. Chen, "Extraction of lesion-partitioned features and retrieval of contrast-enhanced liver images," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 972037, 12 pages, 2012.

[10] W. Yang, Z. Lu, M. Yu, M. Huang, Q. Feng, and W. Chen, "Content-based retrieval of focal liver lesions using bagof-visual-words representations of single- and multiphase contrast-enhanced CT images," *Journal of Digital Imaging*, vol. 25, no. 6, pp. 708–719, 2012.

[11] I. Diamant, A. Hoogi, C. F. Beaulieu et al., "Improved patch-based automated liver lesion classification by separate analysis of the interior and boundary regions," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 6, pp. 1585–1594, 2016.

[12] M. Varma and A. Zisserman, "Texture classification: are filter banks necessary?" in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03)*, vol. 2, pp. II-691–II-698, Madison, Wis, USA, June 2003.

[13] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, *Computer Vision Using Local Binary Patterns*, Springer, 2011.

[14] S. P. Awate, T. Tasdizen, and R. T. Whitaker, "Unsupervised texture segmentation with nonparametric neighborhood statistics," in *Proceedings of the 9th European Conference on Computer Vision*, Graz, Austria, May 2006.

[15] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[16] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE*

Transactions on Image Processing, vol. 19, no. 6, pp. 1635–1650, 2010.

[17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[18] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit," Tech. Rep., 2008.

[19] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[20] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, pp. 233–240, New York, NY, USA, June 2006.