Int J CARS (2018) 13:151–164
https://doi.org/10.1007/s11548-017-1671-9

ORIGINAL ARTICLE

Texture-specific bag of visual words model and spatial cone matching-based method for the retrieval of focal liver lesions using multiphase contrast-enhanced CT images

Yingying Xu[1] · **Lanfen Lin**[1] · **Hongjie Hu**[2] · **Dan Wang**[2] · **Wenchao Zhu**[2] · **Jian Wang**[3] · **Xian-Hua Han**[4] · **Yen-Wei Chen**[3]

Received: 17 August 2016 / Accepted: 25 September 2017 / Published online: 5 November 2017
© CARS 2017

Abstract

Purpose The bag of visual words (BoVW) model is a powerful tool for feature representation that can integrate various handcrafted features like intensity, texture, and spatial information. In this paper, we propose a novel BoVW-based method that incorporates texture and spatial information for the content-based image retrieval to assist radiologists in clinical diagnosis.

Methods This paper presents a texture-specific BoVW method to represent focal liver lesions (FLLs). Pixels in the region of interest (ROI) are classified into nine texture categories using the rotation-invariant uniform local binary pattern method. The BoVW-based features are calculated for each texture category. In addition, a spatial cone matching (SCM)-based representation strategy is proposed to describe the spatial information of the visual words in the ROI. In a pilot study, eight radiologists with different clinical experience performed diagnoses for 20 cases with and without the top six retrieved results. A total of 132 multiphase computed tomography volumes including five pathological types were collected.

Results The texture-specific BoVW was compared to other BoVW-based methods using the constructed dataset of FLLs. The results show that our proposed model outperforms the other three BoVW methods in discriminating different lesions. The SCM method, which adds spatial information to the orderless BoVW model, impacted the retrieval performance. In the pilot trial, the average diagnosis accuracy of the radiologists was improved from 66 to 80% using the retrieval system.

Conclusion The preliminary results indicate that the texture-specific features and the SCM-based BoVW features can effectively characterize various liver lesions. The retrieval system has the potential to improve the diagnostic accuracy and the confidence of the radiologists.

Keywords Content-based image retrieval · Texture-specific · Bag of visual words · Spatial cone matching

Introduction

Computer-aided diagnosis (CAD) systems can assist radiologists in clinical diagnoses based on image analysis [1]. Studies of CAD systems development primarily follow two routes. Classification-based CAD systems have been extensively investigated for decision support employing machine learning methods, such as the support vector machine (SVM) [2,3]. The other route is to construct a content-based image retrieval (CBIR) system. Given an image dataset with diagnosis information, images in the repository with the most similar appearance and morphological characteristics to the query image are retrieved and rendered to support diagnostic decision making [4,5].

Currently, contrast-enhanced computed tomography (CT) is the most crucial imaging modality employed to detect and characterize focal liver lesions (FLLs) [6–8]. Contrast-enhanced CT scans are divided into four phases before and after contrast injection. An examination is performed before contrast injection to obtain a noncontrast-enhanced (NC)

✉ Lanfen Lin
llf@zju.edu.cn

[1] College of Computer Science and Technology, Zhejiang University, Hangzhou, China

[2] Radiology Department, Sir Run Run Shaw Hospital, Medical School, Zhejiang University, Hangzhou, China

[3] College of Information Science and Engineering, Ritsumeikan University, Kusatsu, Japan

[4] Faculty of Science, Yamaguchi University, Yamaguchi, Japan

scan. After-injection phases include the arterial (ART) phase (30–40 s after contrast injection), portal venous (PV) phase (70–80 s after contrast injection), and delay (DL) phase (3–5 min after contrast injection). Several CAD systems based on CT images have been proposed to identify different types of liver lesions [1,2,4–9]. In most previous studies, only one single-phase CT scan was used for feature extraction [3,4], which neglects the pivotal information conveyed by multiphase scans. Several studies were conducted verifying that the features derived from multiphase contrast-enhanced images are more effective than the original features derived from nonenhanced images or single-phase scans [1,7,8].

Some published studies reported the characterization of FLLs using multiphase images to capture the temporal information among phases. Roy et al. [7] proposed a CBIR framework for liver lesions based on spatiotemporal features derived from multiphase CT volumes. Density features (the normalized average intensity of a lesion), texture features (the gray-level co-occurrence matrix [GLCM]), temporal density features (the intensity enhancement over the three enhancement phases compared to the noncontrast phase), and temporal texture features (the normalized difference in texture at the three enhancement phases) were used for image retrieval. Chi et al. [1] extracted multiphase features (density features, density derivative features, texture features and texture derivative features) to represent the FLL. In our previous work [9], we proposed a temporal density feature and a novel 3D shape feature for CBIR of liver lesions in addition to conventional density and texture features (3D uniform local binary pattern [LBP] [10]). All the methods mentioned above used low-level features.

There are several ways to generate mid-level or high-level image representations in various fields including the medical field. The bag of visual words (BoVW) has been a popular strategy to represent medical images for classification or CBIR [2,3,6,8] and achieves promising results in image analysis [8,11–13]. The BoVW method clusters a low-level feature space into a number of regions that potentially correspond to visual concepts that are called visual words. Vocabulary-pruning strategies such as probabilistic latent semantic analysis (pLSA) [14] can be used to remove nonmeaningful visual words [15]. The Riesz transform is a multiscale method that can analyze texture at multiple scales and yield an overcomplete representation of local orientation properties [16,17]. The features extracted by the Riesz transform can also be merged with the BoVW [16]. In [16], a texture-based organ classification algorithm was developed by combining the Riesz transform and the BoVW. The Riesz transform has been applied in lung texture analysis [17,18]. The Fisher vector [19] was proposed as an extension of the BoVW approach. It has been adopted in CAD-based celiac disease classification using endoscopic image data [20,21] and used in recognition of the fetal facial standard plane using ultrasound imaging [22]. Deep learning is emerging as a state-of-the-art method in the computer vision domain, and its performance is improved when applied using large training sets [23]. However, in the medical domain, such large datasets are not always available. Although challenging, there are some recent applications of deep learning to medical tasks [23], such as lesion detection [24,25] and object segmentation [26,27]. There is some work applying the deep neural networks trained on nonmedical training sets for chest pathology retrieval [28] and identification [29,30]. Though there has also been some work applying the deep neural networks on liver tissues such as classification of age or gender based on liver images [31] and detection or segmentation of liver lesions [32–35], there is little work on liver lesion classification or retrieval in CT images.

Several studies were conducted to show that the BoVW is a powerful feature representation method in liver CT image analysis and is an extensively used feature modeling approach that can integrate various handcrafted features like intensity, texture, and spatial information. Diamant et al. [2] employed the BoVW model for automatic classification of liver lesions in four-phase images based on raw intensity. They generated a different vocabulary for each phase and concatenated the histograms for multiple phases to represent the lesions. Yu et al. [6] divided the lesions into distinct regions using a distance transformation technique and extracted BoVW-based features from each subregion. Some handcrafted features like intensity, texture, and shape features were also computed and combined with the histograms for image representation. Yang et al. [8] employed the BoVW model to construct a visual category-specific vocabulary for each pathological type. The histograms of each single phase were merged together to represent multiphase images. Besides, there is still other work applying the BoVW approach modeling other handcrafted features like texture. Burner et al. [11] proposed texture bags based on local 3D extension of the LBP for anomaly retrieval in lung images. In the research field of computer vision, a method known as bag of LBP words (BoWL) [12] has been proposed to incorporate texture features into the BoVW model. However, our experiments indicated that the BoWL model was not very effective in representing FLLs (this is shown in Sec. 4). Several strategies were introduced to embed spatial information into the BoVW model [36]. The most typical approach is the spatial pyramid matching (SPM) method proposed by Lazebnik et al. [37], which repeatedly subdivides the image and computes histograms for each subregion. Some improvements to the SPM method were suggested [38]. The SPM was proven to be effective in characterization. The regular grid partition strategy is not appropriate for liver CT images because the general shape of a liver lesion in a slice is always elliptical or circular.

Table 1 Distribution of the different diagnosis types

Diagnosis type	Cyst	FNH	HCC	HEM	METs
Quantity	36	22	27	27	20

In this paper, we propose two methods to improve the conventional BoVW model for effective FLL representation by integrating texture and spatial information and apply the novel methods to develop a retrieval system to assist radiologists in clinical diagnosis. Specifically, we make the following contributions: (1) A feature extraction method called texture-specific BoVW is proposed by incorporating texture information in the conventional BoVW model, which can effectively improve the representation ability. (2) Inspired by the SPM [37], a spatial cone matching (SCM) strategy which is specific for circular objects such as FLLs is proposed to describe the spatial information of the visual words in the ROI. (3) A dataset of 132 multiphase contrast-enhanced liver CT images that contains five types of lesions was built. (4) Combining the texture-specific BoVW features and the SCM-based BoVW with our previously proposed shape and temporal features [9], we developed a CBIR system to assist radiologists in clinical diagnosis. (5) A pilot trial is conducted to evaluate the influence of the system in clinical diagnoses. The details of the proposed methods, dataset, experiments and the pilot trial are described later.

Materials and methods

Materials

Though ImageCLEF provides a benchmark for liver lesion annotation [39], the CT scans contain only one single phase and are not suitable for our study. We therefore constructed a multiphase CT image database for our study. The contrast-enhanced multiphase CTs were performed using a multidetector helical CT scanner between 2011 and 2015. The multiphase CT scans were collected from 132 patients and included five types of lesions with confirmed pathology/diagnosis, i.e., cysts, focal nodular hyperplasia (FNH), hepatocellular carcinoma (HCC), hemangiomas (HEM), and metastasis (METS). Table 1 displays the distribution of different lesions. One lesion per patient was analyzed and outlined by experienced radiologists. The major lesion was selected for patients with more than one lesions. The CT scans were acquired with a slice collimation of 5/7 mm, a matrix of 512 × 512 pixels and an in-plane resolution of 0.57–0.89 mm. For each patient, all slices containing lesions were used to comprise the image dataset. We extracted image features from all slices containing lesions for the task. Figure 1 shows typical images of different kinds of lesions over multiple phases.

Methods

A multiphase CT retrieval system is developed in this study. The flowchart of the retrieval system is shown in Fig. 2. Data are preprocessed prior to feature extraction. The region of interest (ROI) of each case is first outlined by experienced radiologists. The liver parenchyma and hepatic lesions are segmented synchronously using a random walk-based interactive segmentation algorithm [40] according to the manually drawn outlines. Then, features including the proposed texture-specific BoVW, SCM-based BoVW, shape and temporal features are extracted from the lesion regions for multiple phases. A feature database containing all the features extracted from the images with their confirmed pathology type labels was constructed. Histogram intersection is used to compute similarity. In this section, we mainly focus on representations of FLLs and evaluation methodology.

BoVW model

The BoVW model is adapted from the bag of words (BoW) methodology, which was first proposed for text documents [41] and involves, in our application, using sets of visual words to represent images. For the patch-based BoVW model, patch extraction is the first step in the procedure. Feature vectors are extracted from these patches to construct the visual vocabulary. Clustering algorithms, such as k-means, are commonly used to generate clusters of visual words that comprise the visual vocabulary. The image is represented by a histogram of the generated visual words.

Texture-specific BoVW

The proposed texture-specific BoVW model is implemented on the basis of the rotation-invariant uniform LBP, which is an effective texture representation strategy introduced by Ojala et al. [42]. Figure 3 illustrates a computation of the classic LBP code. A LBP is uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary code is considered circular. The measure of the uniform patterns is as follows:

$$\mathrm{U} = \sum_{i=1}^{P} |s(g_{\mathrm{i}} - g_{\mathrm{c}}) - s(g_{i-1} - g_{\mathrm{c}})|, \quad g_P = g_0 \tag{1}$$

Here, P is the number of neighbors. g_{c} represents the gray value of the central pixel, and g_{i} represents the gray value of the encircled neighborhoods. It is a uniform pattern when $U \leq 2$. For classic LBP codes, the number of patterns satisfying $U \leq 2$ is $P(P-1)+2$. If these patterns are rotated to obtain the minimum value, the number of patterns satisfying the rotation-invariant uniform pattern is $P+1$. Therefore,

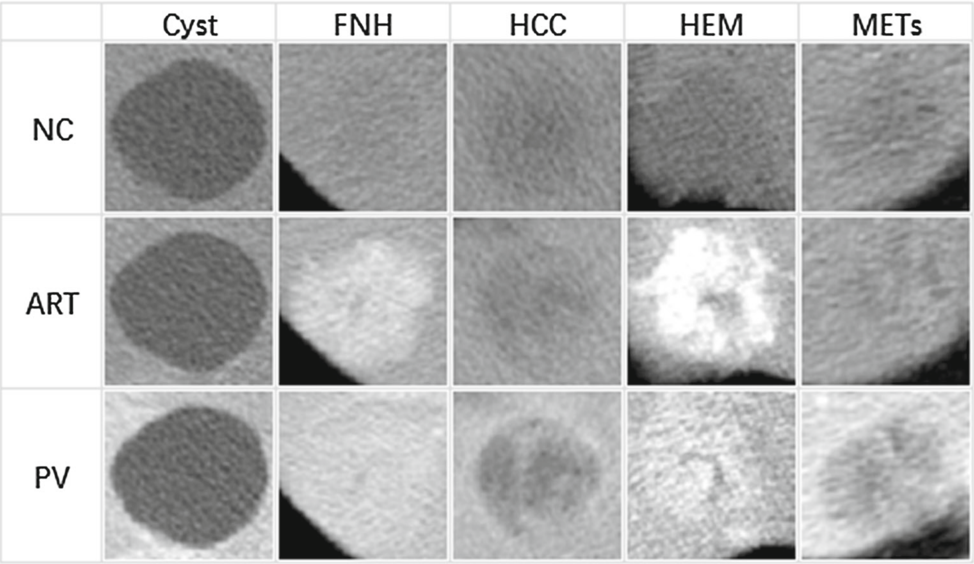


Fig. 1 Typical images of five liver lesions over the three phases

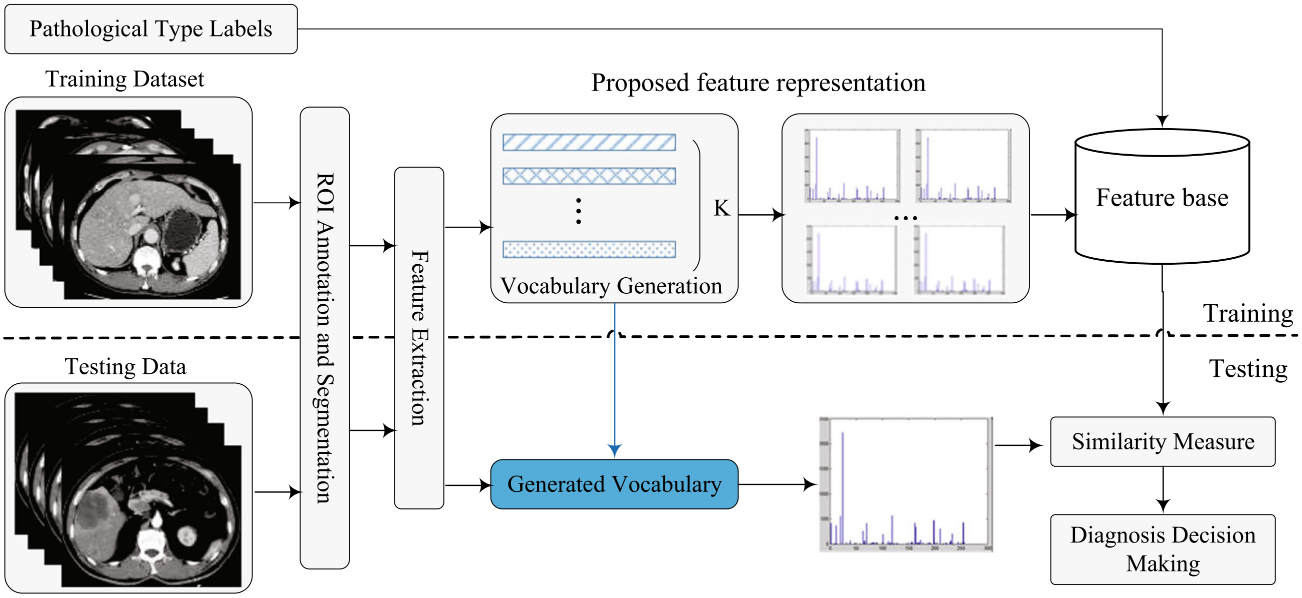


Fig. 2 Framework of our proposed CBIR system

$$\mathrm{LBP}_{P,R}^{\mathrm{riu2}} = \begin{cases} \sum_{i=0}^{p-1} s(g_{\mathrm{i}} - g_{\mathrm{c}}), & U \leq 2 \\ P + 1, & \text{otherwise} \end{cases} \tag{2}$$

where the superscript$^{\mathrm{riu2}}$ reflects the use of a rotation-invariant uniform pattern that has a U value of at most 2. In our study, we used eight neighbors of the central pixel to extract the binary code and to calculate the rotation-invariant uniform LBP. Therefore, the number of texture categories in our study is $8 + 1 = 9$.

For each pixel in the ROI, the LBP is calculated and the pixel is classified into one texture category using the rotation-invariant uniform LBP. All pixels in the ROI are classified into a total of nine categories. The BoVW-based features are calculated for each texture category. Patches of pixels belonging to the same texture category are extracted from the liver image. Raw-intensity-based descriptors are generated

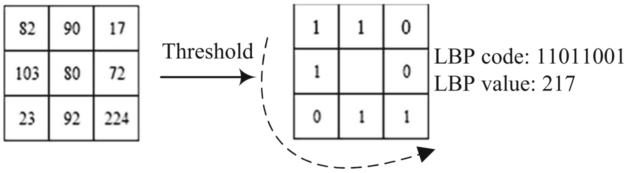


Fig. 3 Illustration of the computation of the classical LBP code

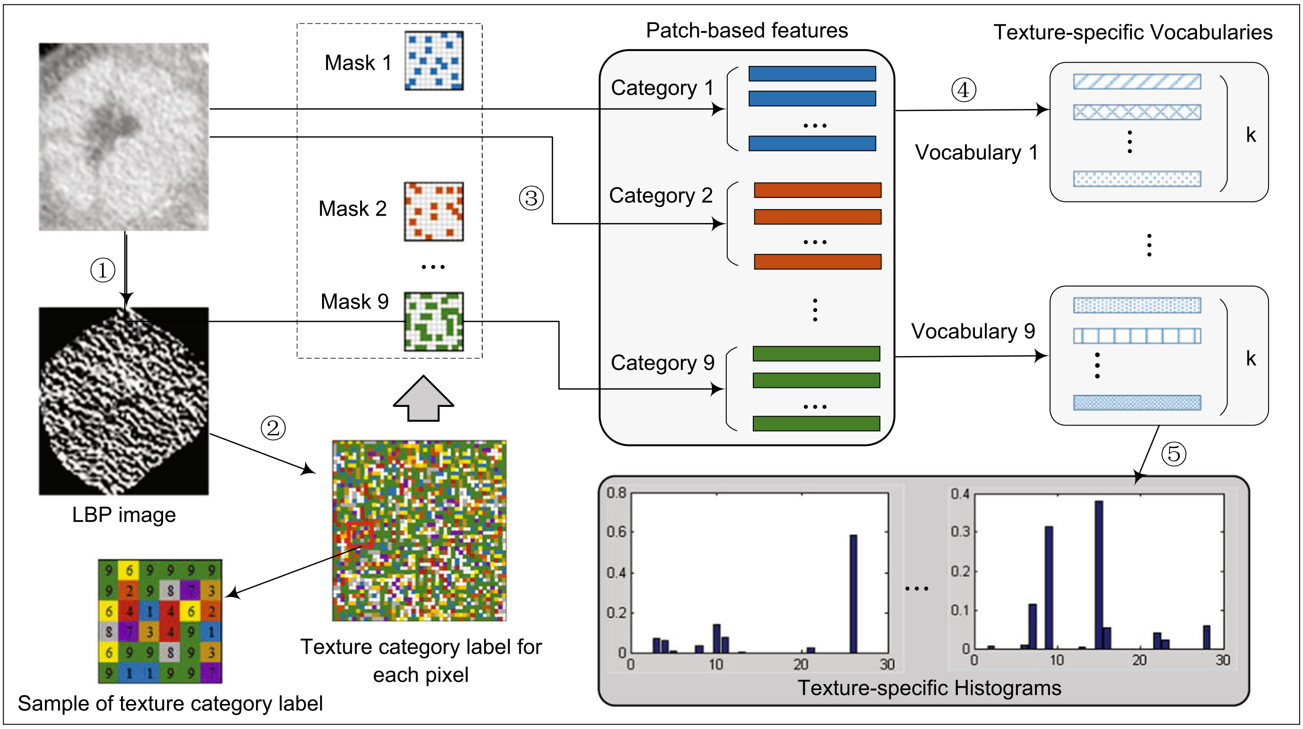


Fig. 4 Illustration of the proposed texture-specific BoVW method. Step 1, LBP image calculation. Step 2, texture classification based on rotation-invariant uniform LBP, assigning a texture category label to each pixel. Step 3, patch-based feature extraction for each texture category. Step 4, vocabulary generation. Step 5, quantization

from these patches to extract the texture-specific vocabularies $V = \{V_1, V_u, \ldots, V_9\}$. Each vocabulary $V_u = \{w_1, \ldots, w_k\}$ is determined by

$$\underset{V_u=\{w_1,\ldots,w_k\}}{\arg\min} \left\{ \sum_i \min_j \left\| \mathbf{x}_i - \mathbf{w}_j \right\|^2 \right\}, \quad \mathbf{x}_i \in \mathrm{T}_u$$
$$\mathrm{T}_u = \{\mathbf{x} | \mathbf{x} \text{ belongs to the } u\text{th texture category}\} \qquad (3)$$

where $\mathbf{x}_i$ is the ith feature vector (patch), $\mathbf{w}_j$ is the jth center vector (visual word), k is the number of visual words of each texture category and the total dimension of histogram $K = 9 \times k$. Figure 4 illustrates the proposed texture-specific BoVW method. Figure 5 shows two texture-specific vocabularies (Category 1 and Category 7) trained by the k-means clustering algorithm in the ART phase.

Spatial cone matching (SCM)-based method

As an improvement in the SPM, we propose a spatial cone matching (SCM) method for circular object representation such as FLLs. We partitioned the FLLs into increasingly fine concentric circular regions and computed histograms for each subregion. For the segmented lesion region of the CT slices, we first calculated the centroid of the pixels in the ROI. Let I be the set of coordinates (x, y) of all pixels in

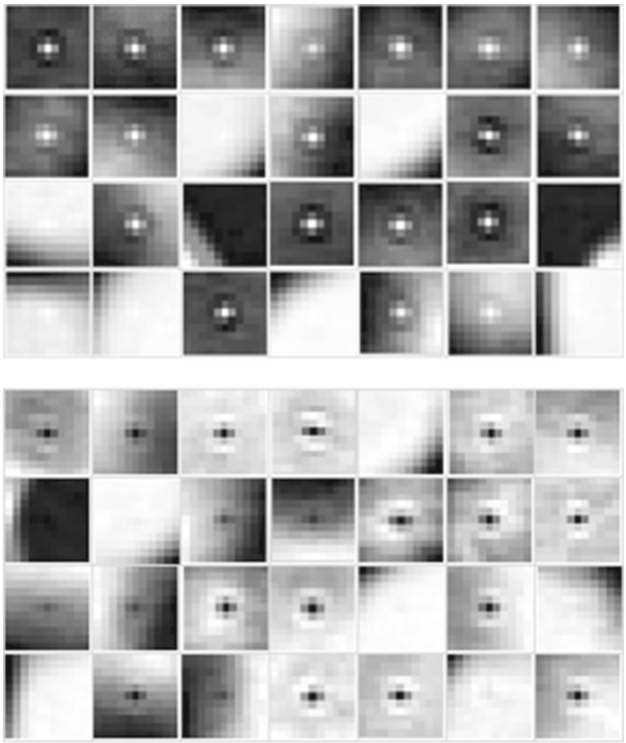


Fig. 5 Two texture-specific vocabularies trained by k-means clustering algorithm of the ART phase

the ROI and denote the centroid as $C = (x_c, y_c)$. The distance between the pixels in the ROI and the centroid can be calculated via Euclidean distance as shown here:

$$D_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}, \quad (x_i, y_i) \in I, \tag{4}$$

where D_i denotes the distance between the ith pixel in the ROI and the centroid. The ROI of the liver images can be outlined by a circle with the centroid at the center. Let r be the radius of the new ROI, which is defined as the maximum value of$D$. We construct a sequence of annular subregions with resolutions of $0, \ldots, L$. The resolution level l corresponds to the number of concentric circles. The ROI at level l is divided into l annular subregions by uniformly dividing the radius into l equal portions. Then, we draw l circles with C at the center and $nr/l(1 \leq n \leq l)$. The k-means clustering technique is applied to the vectors in the ROI to generate the vocabulary, and all feature vectors are quantized into K types based on the visual words. The matching kernel for calculating the total matching degree is the sum of each separate kernel:

$$M^L(X, Y) = \sum_{m=0}^{K} \kappa^L(X, Y). \tag{5}$$

The match kernel for each separate level is defined as

$$\kappa^L(X, Y) = \sum_{l=0}^{L} w_l \times N_l, \tag{6}$$

where w_l is the weight associated with level l, which is set to $\frac{1}{2^l}$, and N_l represents the number of matches at level l.

$$N_l = I(H_l(X), H_l(Y)) - I(H_{l-1}(X), H_{l-1}(Y)) \tag{7}$$

$$I(H_l(X), H_l(Y)) = \sum_{i=1}^{D} \min(H_l^{(i)}(X), H_l^{(i)}(Y)),$$
$$D = L - l + 1 \tag{8}$$

$H_l^{(i)}(X)$ denotes the proportion of the total number of points in the ROI that fall into the ith subregion at level l. As the sizes of the lesions are always different, we represent the match degree at level l via the histogram intersection [43] using the normalized value instead of the true number of points. An illustration of the construction of a spatial histogram for visual words based on the SCM is shown in Fig. 6.

Shape feature

Shape features are important for distinguishing different types of lesions. In this paper, the sphericity of a lesion extracted by principle component analysis (PCA) is used as the shape feature of the lesion. Considering that the shape of the lesion shows no obvious change in different phases, we extract the shape features from the ART phase images. The shape feature F is denoted by three eigenvalues $\lambda_1, \lambda_2, \lambda_3 (\lambda_1 > \lambda_2 > \lambda_3)$ calculated via PCA as follows:

$$F = \{\lambda_2/\lambda_1, \lambda_3/\lambda_1\} \tag{9}$$

Temporal feature

The temporal feature, TF, is designed to discriminate the evolution patterns after the injection by denoting the enhancement of density in the ART and PV phases compared with the density in the NC phase. Some typical images of the evolution pattern of different lesions are shown in Fig. 1. The calculation of TF is

$$\mathrm{TF} = \left\{ \left(d_{\mathrm{lesion}}^{\mathrm{ART}} - d_{\mathrm{lesion}}^{\mathrm{NC}}\right) / d_{\mathrm{lesion}}^{\mathrm{NC}}, \left(d_{\mathrm{lesion}}^{\mathrm{PV}} - d_{\mathrm{lesion}}^{\mathrm{NC}}\right) / d_{\mathrm{lesion}}^{\mathrm{NC}} \right\}, \tag{10}$$

where $d_{\mathrm{lesion}}^{\mathrm{NC}}$ and $d_{\mathrm{liver}}^{\mathrm{NC}}$ are, respectively, the average voxel intensity of the lesion and the normal liver tissue in the NC phase. $d_{\mathrm{lesion}}^{\mathrm{ART}}$, $d_{\mathrm{liver}}^{\mathrm{ART}}$, $d_{\mathrm{lesion}}^{\mathrm{PV}}$, and $d_{\mathrm{liver}}^{\mathrm{PV}}$ have similar definitions for different phases.

Evaluation methodology

The total of 132 cases is divided into two parts: 32 cases are chosen as test data, and the remaining 100 cases are used as training data. For each lesion type, about one-quarter of cases are randomly selected to constitute the test data. The optimal parameters are determined based on the training data using the fivefold cross-validation, and the test data are only used to evaluate the system performance. Each case in the test data is a query case (an unlabeled case), and the FLLs in the training data are used as labeled cases for retrieving similar lesions. Twenty cases were selected from the test data for the pilot trial, and the remaining cases including the other twelve cases of the test data were used to construct the database for retrieving similar lesions in the pilot trial. To estimate the effectiveness of the retrieval system, several performance evaluation measures are proposed. The results are evaluated based on a precision–recall curve, precision of the top M retrieved FFLs (Prec@M), and the mean average precision (mAP). Precision and recall are defined as follows:

$$\text{precision} = \frac{\text{Number of relevant samples retrieved}}{\text{Total number of samples retrieved}} \tag{11}$$

$$\text{recall} = \frac{\text{Number of relevant samples retrieved}}{\text{Total number of relevant samples}} \tag{12}$$

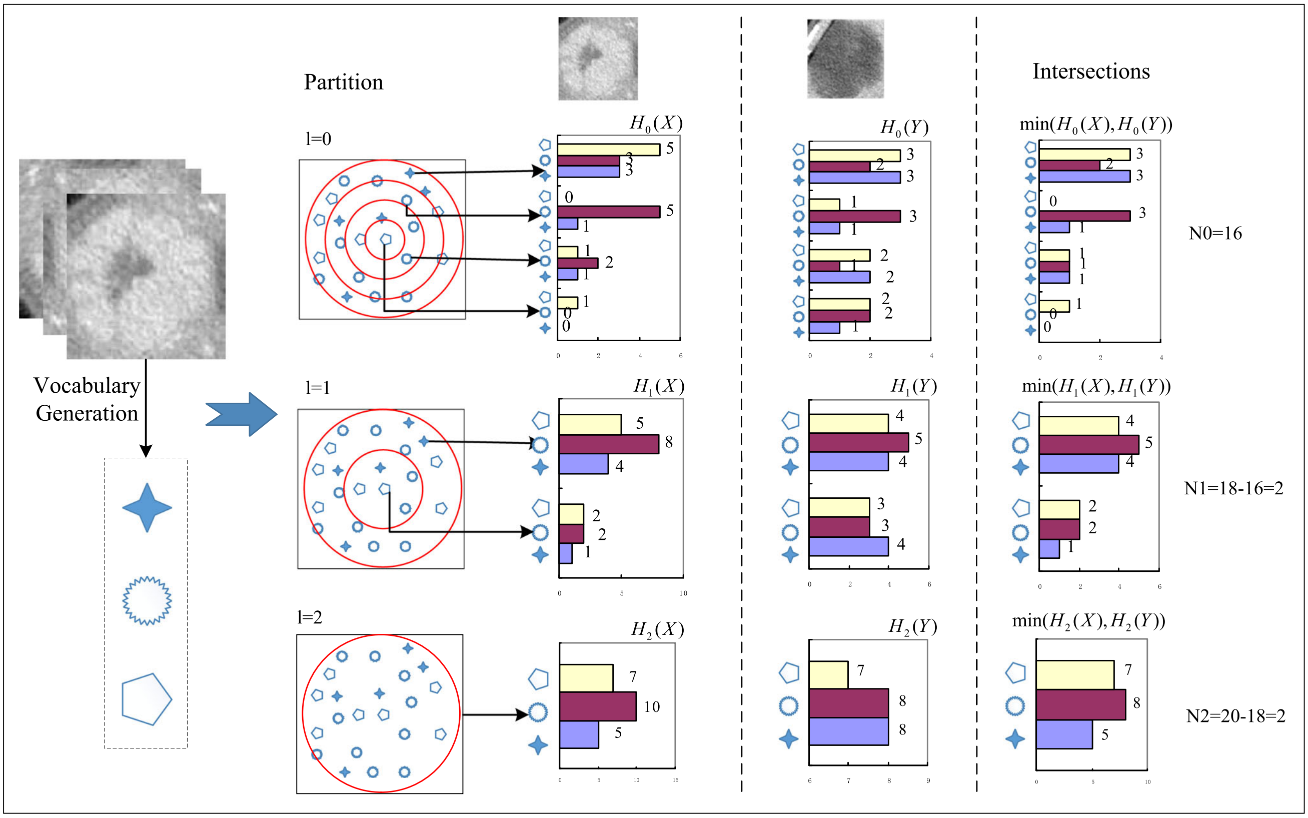


Fig. 6 Illustration of the construction of the spatial histogram for visual words based on the SCM

Prec@M represents the proportion of relevant FLLs in the top M results. Average precision (AP) is the average of the precisions for the positions where there is a relevant sample. AP is defined as follows:

$$\mathrm{AP} = \frac{1}{\mathrm{N}^{+}} \sum_{j=1}^{N} \mathrm{rel}(j) \times \mathrm{Pr}\,ec@\,j \qquad (13)$$

where N is the total number of samples and $\mathrm{rel}(j)$ is a binary function on the relevance of the sample at the position j (1 for relevant, and 0 for irrelevant). mAP is the mean of the AP over all queries.

Experiment and results

A series of experiments were conducted to verify the effectiveness of the texture-specific BoVW and the SCM-based method. The detailed experimental designs and results are described in the following subsections. All the experiments were conducted with a MATLAB® implementation, on an Intel(R) Core(TM) i7-6700K, 4-GHz CPU with 32 GB of RAM.

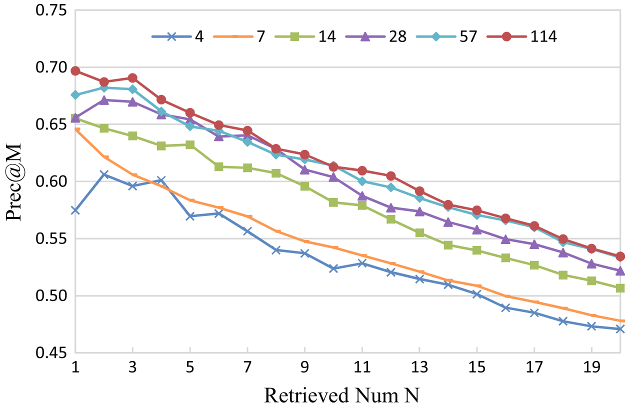


Fig. 7 System performance with various vocabulary sizes

Experiments on parameter selection

We conducted a group of experiments to analyze the impact of the vocabulary size and the patch size of our algorithm on the retrieval performance. The performance of different parameters was evaluated on the 100 training data using the fivefold cross-validation. The average result of fivefold is used as a measure of performance for a given parameter setting. The optimal parameters are selected based on the best performance. The patch size was set to 9 × 9 when assessing

Table 2 System performance with various patch sizes

Patch size	Prec@2	Prec@5	Prec@10	Prec@15	Prec@20
3 × 3	0.6621	0.6504	0.5854	0.5444	0.5119
5 × 5	0.6821	0.6565	0.6057	0.5585	0.5244
7 × 7	0.6721	0.6563	0.6128	0.5632	0.5330
9 × 9	0.6821	0.6482	0.6139	0.5707	0.5335
11 × 11	0.6766	0.6522	0.6070	0.5645	0.5264
13 × 13	0.6663	0.6361	0.5999	0.5613	0.5203

the impact of the vocabulary size. We varied the vocabulary size (k) from 4 (32/9) to 114 (1024/9) for each texture category. Figure 7 shows that larger vocabularies lead to better results. However, a larger vocabulary results in higher computational cost. We can see from the results in Fig. 7 that the gain on performance is not significant when the vocabulary size is greater than 57 (512/9). We can see that a vocabulary size of $k = 57$ is an appropriate value for our texture-specific method in balancing the retrieval performance and the computational cost. The vocabulary size was set to 57 when evaluating the impact of the patch size. Table 2 lists the experimental results and shows that the system gets a relatively better performance when the patch size is set to 9 × 9. In the following experiments, the vocabulary size of the texture-specific BoVW for each texture category is set to 57, which produces a total vocabulary size of $57 \times 9 = 513$ and the patch size is set to 9 × 9.

System performance comparison for different texture representation methods

The performance of different texture description strategies is compared in this subsection. We first compared our texture-specific BoVW with three texture descriptors, i.e., the 3D GLCM applied in [7], the classic LBP [42,44], and the previously proposed improved LBP (imLBP) [9] based on 3D lesion volumes. The offset was given as 1, and 13 directions were considered when computing the GLCM. Six texture coefficients, including energy, entropy, inverse difference moment, inertia, cluster shade and correlation, were derived from the matrixes. The results are shown in Fig. 8 using precision versus recall curve. We also compared our method with another LBP-based BoVW representation model known as BoWL [11], which has been reported to be effective in scene image classification. The BoWL was introduced by Banerji et al. based on multineighborhood LBP for small patches. The total size of the vocabulary for each phase was set to 128, which performed best on our dataset. The results are exhibited in Fig. 8 and Table 3. The results indicate that our method significantly outperforms the low-level features and is superior to the BoWL method. These results also show that LBP-based methods are more effective than GLCM as a texture descriptor.

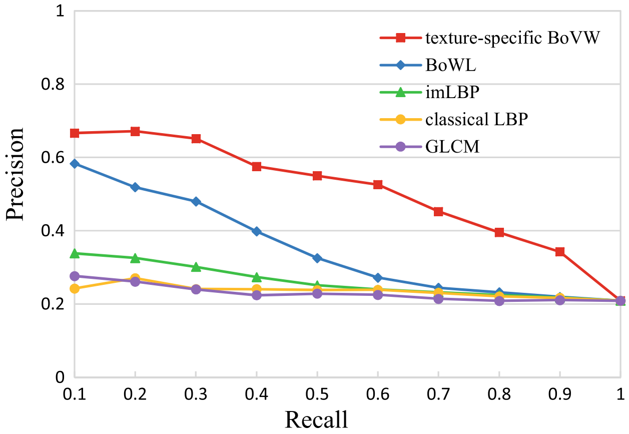


Fig. 8 Results of different texture representations

Table 3 Retrieval performance of different image descriptors

Descriptor	mAP	Pre@6	Pre@10
Low-level descriptors			
GLCM [7]	0.3790	0.2727	0.2697
LBP histogram [21]	0.4478	0.2576	0.2424
imLBP [9]	0.5020	0.3384	0.3242
Mid-level descriptors			
Fisher vector [30]	0.6759	0.5960	0.5576
Global BoVW [8]	0.6746	0.5758	0.5273
Category-specific BoVW [8]	0.7045	0.5808	0.5303
BoWL [11]	0.6551	0.5505	0.5030
Texture-specific BoVW	0.7563	0.6717	0.6515

Comparison of the texture-specific BoVW method and the conventional BoVW methods and the Fisher vector

Several BoVW methods were implemented in the experiments, i.e., the category-specific BoVW and global BoVW methods proposed by Yang et al. [8], and the BoWL method proposed by Banerji et al. [11]. For fair comparison, we first conducted a series of experiments to determine the optimal parameters (the vocabulary size and the patch size for the

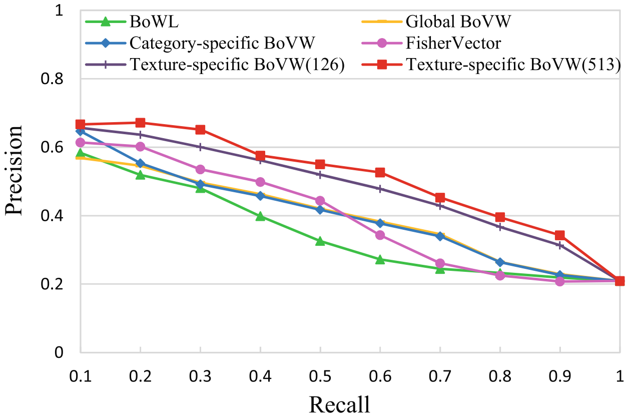


Fig. 9 Comparison of BoVW-based methods and the Fisher vector

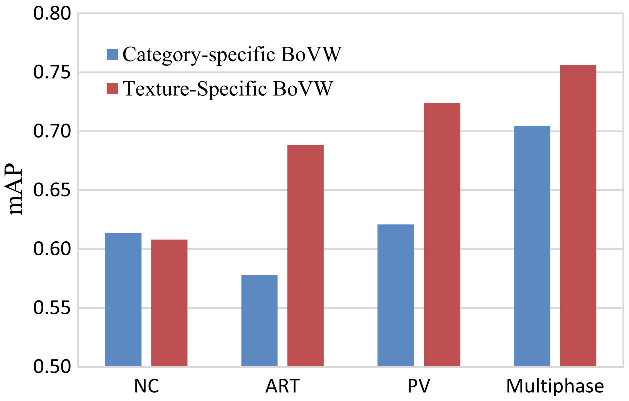


Fig. 10 mAP of the category-specific BoVW and the texture-specific BoVW at single phase and multiple phases

BoVW-based methods and the number of Gaussian components for the Fisher vector) for each method on our training data using the fivefold cross-validation. For the global BoVW and the BoWL, the vocabulary size of each phase was set to 128. The category-specific BoVW was performed with a vocabulary size of 26 (128/5) for each pathological category. The patch size was determined to be 11 × 11 for the BoWL and the category-specific BoVW and 9 × 9 for the global BoVW. For our proposed texture-specific BoVW method, the vocabulary size of each texture category was set to 14 (128/9) and 57 (512/9) for comparison. As for the Fisher vector, we exploited patch-based features as local descriptors and the number of Gaussian components was fixed at 10. The results are shown in Fig. 9. Table 3 lists the retrieval performance of different low-level features and mid-level features in terms of mAP, Prec@6, and Prec@10 for multiple phases. As shown, all mid-level features in the experiments yield better results than the low-level descriptors, and our proposed texture-specific BoVW approach outperforms the other three BoVW methods and the Fisher vector.

Comparison of results based on single-phase and multiphase images

A group of experiments were conducted comparing the system performance on multiphase and single-phase features extracted by the proposed texture-specific BoVW method and the category-specific BoVW methods. The results are shown in Fig. 10. We can see that the features extracted from multiphase images have stronger discriminative capability in retrieval of similar lesions than those extracted from single-phase images.

Experiments on the validation of the SCM-based method

To validate the effectiveness of our proposed SCM method, we combined the SCM processing with the conventional BoVW and category-specific BoVW methods proposed by Yang [8] and compared the results with the original methods without the SCM processing. The results are shown

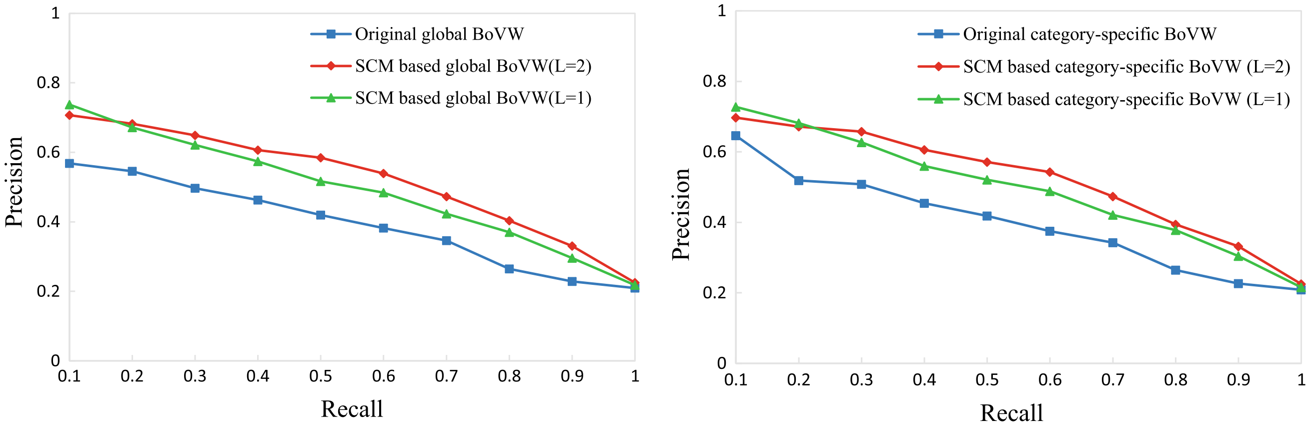


Fig. 11 System performance of applying the SCM to BoVW-based methods. **a** Applying the SCM to conventional BoVW. **b** Applying the SCM to category-specific BoVW

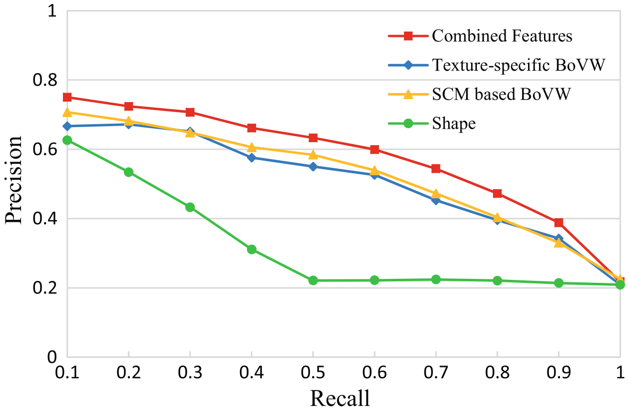
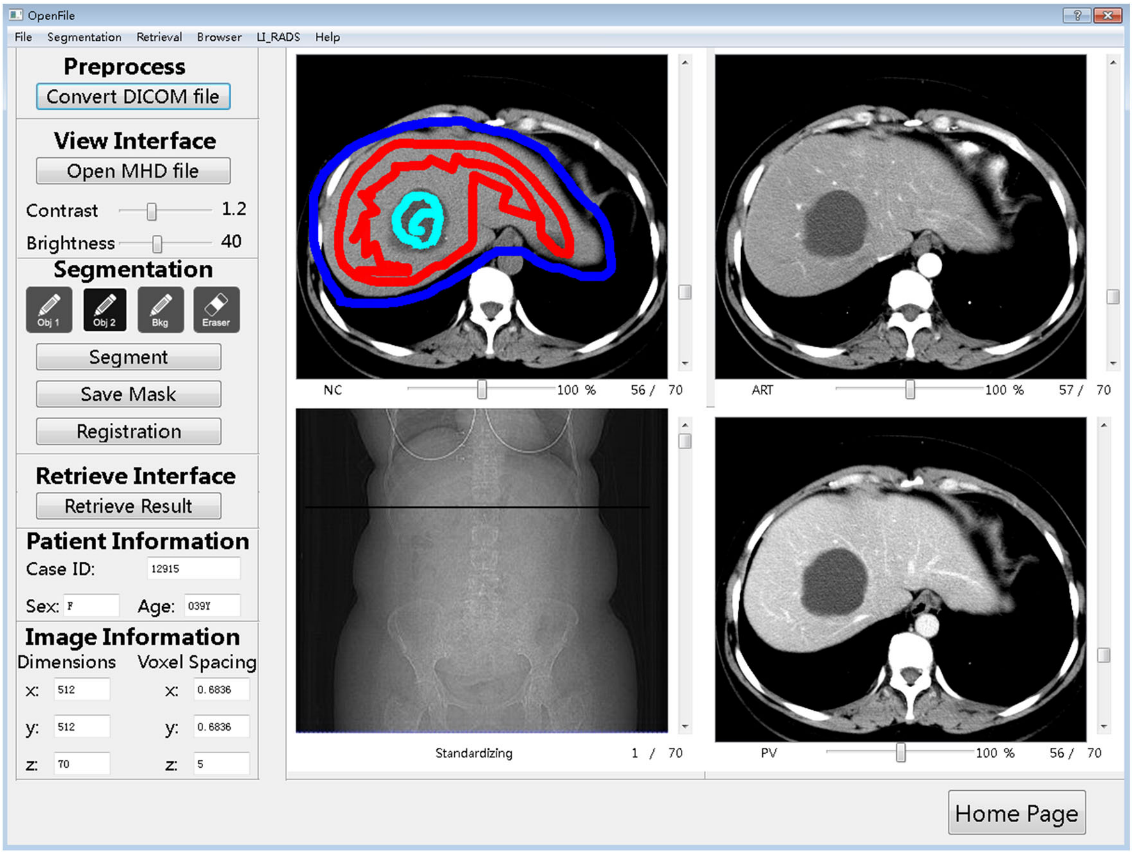


Fig. 12 Retrieval performances using combined features and each single feature

in Fig. 11. We can see from the figure that the application of the SCM effectively improved the retrieval performance. We also compared the SCM methods using different resolutions with L to be 1 and 2. The results displayed in Fig. 11 show that higher resolution performed better than low resolution. In our research, we have tried to apply the SCM method with our proposed texture-specific BoVW, but the SCM-based texture-specific BoVW was not effective in representing FLLs compared to the texture-specific BoVW. We have analyzed the results and suppose that it is probably because processing of the texture classification can be regarded as a kind of partitioning approach to a certain extent.

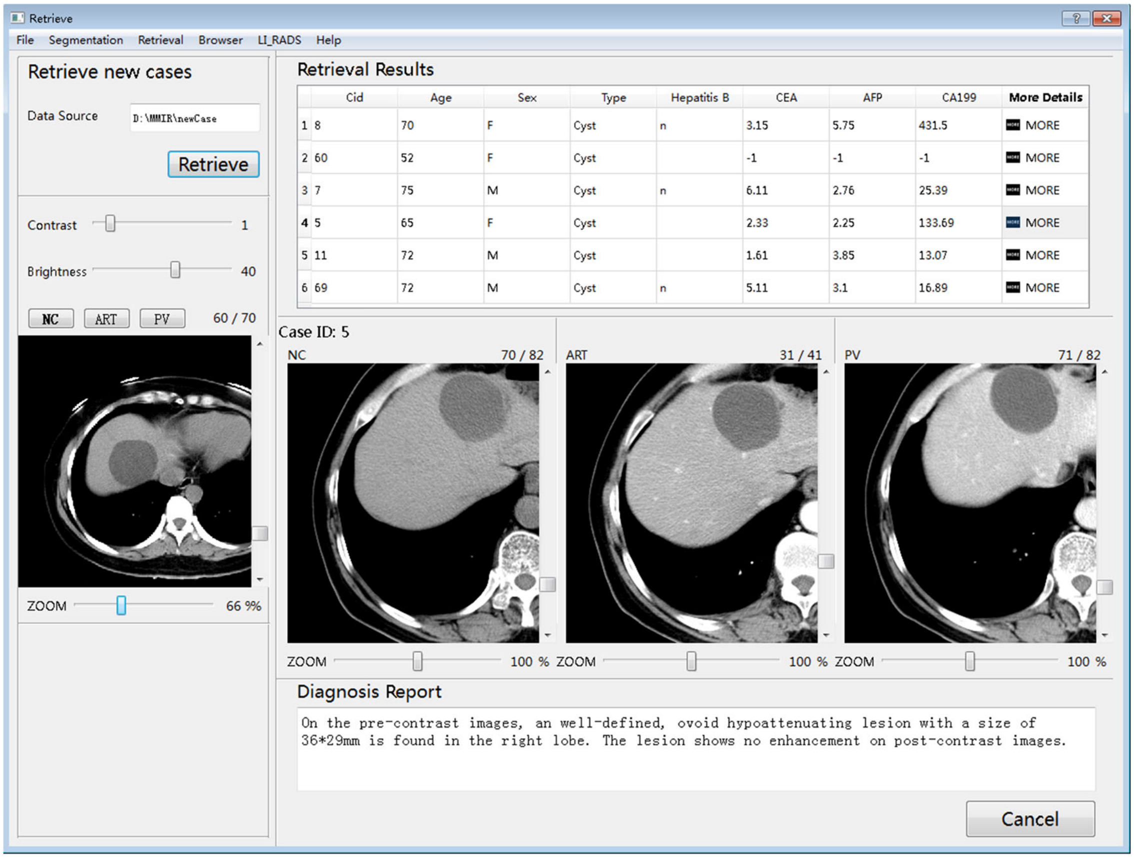
System construction

We developed a CBIR system based on our proposed texture-specific BoVW and SCM-based descriptor. In addition to the features in this study, we also employed the shape and temporal features, which were introduced in our previous study for image retrieval [9]. The retrieval performance using the combined features and each single feature is presented in Fig. 12. The combination of proposed methods can significantly improve the system performance. Figures 13 and 14 show the retrieval system interface. We can see from Fig. 14 that in a retrieval task, the query case information is shown on the left side of the interface. Radiologists click the Retrieve button to retrieve similar cases. The information of the six most similar results retrieved by the system is displayed in the right side of the interface. When one of the retrieval results is chosen by the radiologist, its multiphase CT images and corresponding diagnosis report are shown. More details of the case are shown when the More button is clicked. In the



Contrast 1.2
Brightness 40
Case ID: 12915
Sex: F
Age: 039Y
Dimensions x: 512 y: 512 z: 70
Voxel Spacing x: 0.6836 y: 0.6836 z: 5
NC 100 % 56 / 70
ART 100 % 57 / 70
Standardizing 1 / 70
PV 100 % 56 / 70

Fig. 13 Interface of segmentation



Data Source D:\MMIR\newCase

	Cid	Age	Sex	Type	Hepatitis B	CEA	AFP	CA199	More Details
1	8	70	F	Cyst	n	3.15	5.75	431.5	MORE
2	60	52	F	Cyst		-1	-1	-1	MORE
3	7	75	M	Cyst	n	6.11	2.76	25.39	MORE
4	5	65	F	Cyst		2.33	2.25	133.69	MORE
5	11	72	M	Cyst		1.61	3.85	13.07	MORE
6	69	72	M	Cyst	n	5.11	3.1	16.89	MORE

Case ID: 5

On the pre-contrast images, an well-defined, ovoid hypoattenuating lesion with a size of 36*29mm is found in the right lobe. The lesion shows no enhancement on post-contrast images.

Fig. 14 Interface of retrieval

pilot trial, radiologists determined the diagnosis based on the reference information.

Pilot trial

To evaluate the influence of the retrieval system for liver lesions in clinical diagnosis, a pilot trial was performed. In the pilot study, eight radiologists were invited to perform diagnoses with and without the top six similar retrieved results identified by the retrieval system.

The participants were divided into two groups: group A and group B, reflecting the same seniority and similar diagnostic ability. A total of 20 cases (4 cases for each class) were selected by a radiologist with more than 10 years of experience as the query data. The selected query cases were separated from the dataset, and we used the remaining cases as the retrieval dataset for the retrieval system. Determining the number of images used for query is a trade-off problem. The user test reliability could be raised by increasing the number of query images, but we had to consider the workload of the radiologists. The user test is a time-consuming and labor-intensive task, and we determined the number of query images based on the radiologists' suggestions.

We randomly arranged the query data as No. 1, No. 2 …, to No. 20, respectively, in the first round of diagnoses. Only the CT ID reallocated by the engineer and the CT images of the query cases were provided for the radiologists in group A. For the radiologists in group B, both the CT images of the query cases and their corresponding top six similar retrieved results rendered by our retrieval system were provided. The radiologists performed the first round of diagnoses from No. 1 to No. 20 and gave their diagnosis results (the type of lesion) and their confidence level (1–10: 1 indicates no confidence, and 10 indicates fully confident) of the corresponding decision. Two weeks later, we rearranged the query data randomly as No. 1 through No. 20. The radiologists were requested to do a second round of diagnoses on the same query cases. In this trial, the radiologists in group A were provided with the references from the retrieval system and the radiologists in group B performed diagnoses without use of the decision support. The radiologists again gave their diagnoses and confidence levels.

The results of the two rounds of trials were analyzed in terms of their diagnostic accuracy, which is defined as the ratio of the number of true positives to the total number of query cases. The diagnostic confidence level was assigned a value of 0 when the diagnosis was wrong. The average

Table 4 The results of the diagnoses with and without the top six retrieved results

Participant	Diagnoses	Average diagnostic accuracy	Mean average confidence value
Group A	Round 1 (without retrieved results)	0.66	5.96
	Round 2 (with retrieved results)	0.80	6.75
Group B	Round 1 (with retrieved results)	0.78	5.33
	Round 2 (without retrieved results)	0.71	4.86

confidence values of each radiologist in a round of test (20 diagnoses) were calculated as the confidence level estimation. The average diagnostic accuracy and the mean average confidence value of the radiologists in each group were calculated. The comparison of the two rounds of diagnoses is shown in Table 4. From the results, we can see that with the help of the retrieval system, the average diagnosis accuracy and the confidence levels of the radiologists in group A achieve pronounced improvements. Without the references of the retrieval system information, the average diagnosis accuracy and the confidence levels of radiologists in group B decreased compared to the first round test. The preliminary results demonstrate that the retrieval system has the potential to improve the diagnostic accuracy and the confidence of radiologists.

Conclusion and future work

In this paper, we developed a CBIR system for FLLs based on multiphase contrast-enhanced CT images to assist radiologists in clinical diagnosis. We have proposed a texture-specific BoVW and an SCM-based method for the effective representation of FLLs incorporating texture and spatial information. We also combined the proposed methods with our previously introduced shape and temporal features to achieve more effective image retrieval. The preliminary results indicate that the proposed texture-specific and SCM-based features extracted from multiphase images can effectively characterize various liver lesions. The retrieval system has the potential to improve the diagnostic accuracy and the confidence of radiologists. The precision at the top six retrieved similar cases is 0.74, which can be improved by increasing the size of the database. Our texture-specific and SCM-based methods are proposed as frameworks for integrating texture or spatial information into the BoVW model. As an extension, our proposed frameworks can also be combined with, and thus improve, the Fisher vector and the super vector methods, which will be done in our future work. Furthermore, the texture-specific method with texture information and the SCM-based method with spatial information cannot complement each other in the current research. In future work, we will develop a more effective SCM-based texture-specific descriptor for the retrieval task. Currently, the BoVW-based features in our work are independently extracted from the images of each phase and linearly combined, which neglects the temporal information and relationship among phases. We will explore the temporal co-occurrence information of multiple phases in our future work and will increase the dataset and apply deep learning to feature extraction.

Acknowledgements This research was supported in part by the National Key Basic Research Program of China under the Grant No. 2015CB352400, in part by the National Key Research and Development Program of China under the Grant No. 2016YFB1200203-03, in part by the Recruitment Program of Global Experts (HAIOU Program) from Zhejiang Province, China, in part by the Grant-in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under the Grant No. 15H01130, No. 15K00253 and No.16H01436, and in part by the MEXT Support Program for the Strategic Research Foundation at Private Universities (2013–2017).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standard All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. For this type of study, formal consent is not required.

References

1. Chi Y, Zhou J, Venkatesh SK, Tian Q, Liu J (2013) Content-based image retrieval of multiphase CT images for focal liver lesion characterization. Med Phys 40(10):103502
2. Diamant I, Goldberger J, Klang E, Amitai M (2015) Multiphase liver lesions classification using relevant visual words based on mutual information. In: 2015 IEEE 12th international symposium on biomedical imaging (ISBI). IEEE, pp 407–410
3. Diamant I, Hoogi A, Beaulieu CF, Safdari M, Klang E, Amitai M, Rubin DL (2016) Improved patch-based automated liver lesion classification by separate analysis of the interior and boundary regions. IEEE J Biomed Health Inform 20(6):1585–1594
4. Napel SA, Beaulieu CF, Rodriguez C, Cui J, Xu J, Gupta A, Korenblum D, Greenspan H, Ma Y, Rubin DL (2010) Automated retrieval of CT images of liver lesions on the basis of image similarity: method and preliminary results 1. Radiology 256(1):243–252

5. Akgül CB, Rubin DL, Napel S, Beaulieu CF, Greenspan H, Acar B (2011) Content-based image retrieval in radiology: current status and future directions. J Digit Imaging 24(2):208–222
6. Yu M, Feng Q, Yang W, Gao Y, Chen W (2012) Extraction of lesion-partitioned features and retrieval of contrast-enhanced liver images. Comput Math Methods Med 2012
7. Roy S, Chi Y, Liu J, Venkatesh SK, Brown MS (2014) Three-dimensional spatiotemporal features for fast content-based retrieval of focal liver lesions. IEEE Trans Biomed Eng 61(11):2768–2778
8. Yang W, Lu Z, Yu M, Huang M, Feng Q, Chen W (2012) Content-based retrieval of focal liver lesions using bag-of-visual-words representations of single-and multiphase contrast-enhanced CT images. J Digit Imaging 25(6):708–719
9. Xu Y, Lin L, Hu H, Yu H, Jin C, Wang J, Han X, Chen YW (2016) Combined density, texture and shape features of multi-phase contrast-enhanced CT images for CBIR of focal liver lesions: a preliminary study. In: Chen YW, Toro C, Tanaka S, Howlett RJ, Jain LC (eds) Innovation in medicine and healthcare 2015. Springer, Berlin, pp 215–224
10. Ojala T, Pietikainen M, Harwood D (1994) Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In: Proceedings of the 12th IAPR international conference on Pattern recognition, 1994. Vol. 1-conference a: computer vision & image processing. IEEE, pp 582–585
11. Burner A, Donner R, Mayerhoefer M, Holzer M, Kainberger F, Langs G (2011) Texture bags: anomaly retrieval in medical images based on local 3D-texture similarity. In: MICCAI international workshop on medical content-based retrieval for clinical decision support. Springer, Berlin, pp 116–127
12. Banerji S, Sinha A, Liu C (2013) A new bag of words LBP (BoWL) descriptor for scene image classification. In: International conference on computer analysis of images and patterns. Springer, Berlin, pp 490–497
13. Asherov M, Diamant I, Greenspan H (2014) Lung texture classification using bag of visual words. In: SPIE medical imaging. International society for optics and photonics, pp 90352K–90352K
14. Hofmann T (2001) Unsupervised learning by probabilistic latent semantic analysis. Mach Learn 42(1):177–196
15. Foncubierta-Rodríguez A, García Seco de Herrera A, Müller H (2013) Medical image retrieval using bag of meaningful visual words: unsupervised visual vocabulary pruning with PLSA. ACM international workshop on multimedia indexing and information retrieval for healthcare, pp 75–82
16. del Toro OAJ, Foncubiertarodríguez A, Depeursinge A, Müller H (2015) Texture classification of anatomical structures in CT using a context-free machine learning approach. In: SPIE medical imaging, vol 9414, pp 94140W–94140W-14
17. Depeursinge A, Foncubierta–Rodriguez A, Van de Ville D, Müller H (2012) Multiscale lung texture signature learning using the Riesz transform. Medical image computing & computer-assisted intervention: Miccai international conference on medical image computing & computer-assisted intervention, vol 15, p 517
18. Depeursinge A, Foncubierta-Rodriguez A, Van de Ville D, Müller H (2011) Lung texture classification using locally–oriented Riesz components. In: Medical image computing and computer-assisted intervention (MICCAI 2011). Springer, Berlin
19. Csurka G, Perronnin F (2010) Fisher vectors: beyond bag-of-visual-words image representations. In: International conference on computer vision, imaging and computer graphics. Springer, Berlin pp 28–42
20. Gadermayr M, Kogler H, Uhl A, Vécsei A (2015) Comparing endoscopic imaging configurations in computer-aided celiac disease diagnosis. In: 2015 International conference on image processing theory, tools and applications (IPTA). IEEE, pp 446–451
21. Gadermayr M, Kogler H, Karla M, Merhof D, Uhl A, Vécsei A (2016) Computer-aided texture analysis combined with experts' knowledge: improving endoscopic celiac disease diagnosis. World J Gastroenterol 22(31):7124
22. Lei B, Tan EL, Chen S, Zhuo L, Li S, Ni D, Wang T (2015) Automatic recognition of fetal facial standard plane in ultrasound image via fisher vector. PLoS ONE 10(5):e0121838
23. Greenspan H, van Ginneken B, Summers RM (2016) Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans Med Imaging 35(5):1153–1159
24. Roth HR, Lu L, Seff A, Cherry KM, Hoffman J, Wang S, Liu J, Turkey E, Summers RM (2014) A new 2.5 D representation for lymph node detection using random sets of deep convolutional neural network observations. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin, pp 520–527
25. Akselrod-Ballin A, Karlinsky L, Alpert S, Hasoul S, Ben-Ari R, Barkan E (2016) A region based convolutional network for tumor detection and classification in breast mammography. In: International workshop on large-scale annotation of biomedical data and expert label synthesis. Springer, Berlin, pp 197–205
26. de Brebisson A, Montana G (2015) Deep neural networks for anatomical brain segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 20–28
27. Brosch T, Tang LY, Yoo Y, Li DK, Traboulsee A, Tam R (2016) Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. IEEE Trans Med Imaging 35(5):1229–1239
28. Anavi Y, Kogan I, Gelbart E, Geva O, Greenspan H (2015) A comparative study for chest radiograph image retrieval using binary texture and deep learning classification. In: 2015 37th Annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, pp 2940–2943
29. Bar Y, Diamant I, Wolf L, Greenspan H (2015) Deep learning with non-medical training used for chest pathology identification. In: SPIE medical imaging, international society for optics and photonics, pp 94140V–94140V
30. Bar Y, Diamant I, Wolf L, Lieberman S, Konen E, Greenspan H (2015) Chest pathology detection using deep learning with non-medical training. In: 2015 IEEE 12th international symposium on biomedical imaging (ISBI). IEEE, pp 294–297
31. Andrearczyk V, Whelan PF (2016) Deep learning for biomedical texture image analysis. In: Proceedings of the 18th Irish machine vision and image processing conference (IMVIP2016)
32. Li W, Jia F, Hu Q (2015) Automatic segmentation of liver tumor in CT images with deep convolutional neural networks. J Comput Commun 3:146–151
33. Vivanti R (2015) Automatic liver tumor segmentation in follow-up CT studies using convolutional neural networks. In: Proceedings of the MICCAI2015
34. Todoroki Y, Han X, Iwamoto Y, Lin L, Hu H, Chen YW (2017) Detection of liver tumor candidates from CT images using deep convolutional neural networks. In: Proceedings of the international conference on innovation in medicine and healthcare, pp 140–145
35. Hu P, Wu F, Peng J, Liang P, Kong D (2016) Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution. Phys Med Biol 61(24):8676
36. Zhao LJ, Tang P, Huo LZ (2014) Land-use scene classification using a concentric circle-structured multiscale bag-of-visual-words model. IEEE J Sel Top Appl Earth Obs Remote Sens 7(12):4620–4631
37. Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: spatial pyramid matching for recognizing natural scene categories.

In: 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR’06). IEEE, vol 2, pp 2169–2178

38. Zhou L, Zhou Z, Hu D (2013) Scene classification using a multi-resolution bag-of-features model. Pattern Recognit 46(1):424–433
39. Marvasti NB, Kökciyan N, Türkay R, Yazici A, Yolum P, Üsküdarli S, Acar B (2014) ImageCLEF liver CT image annotation task 2014. In: CLEF (working notes), pp 329–340
40. Dong C, Chen YW, Lin L, Hu H, Jin C, Yu H, Han X, Tateyama T (2016) Simultaneous segmentation of multiple organs using random walks. J Inf Process 24(2):320–329
41. Joachims T (1998) Text categorization with support vector machines: learning with many relevant features. In: European conference on machine learning. Springer, Berlin, pp 137–142
42. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987
43. Barla A, Odone F, Verri A (2003) Histogram intersection kernel for image classification. In: Proceedings of the 2003 international conference on image processing (ICIP 2003), vol 3. IEEE, p III-513
44. Shan C, Gong S, McOwan PW (2005) Robust facial expression recognition using local binary patterns. In: IEEE international conference on image processing 2005, vol 2. IEEE, p II-370